

A decorative graphic consisting of three blue circles of varying sizes and two thin blue lines. One line starts from the top left and passes through the center of the largest circle. Another line starts from the top left and passes through the center of the medium-sized circle. A third line starts from the top right and passes through the center of the smallest circle. The circles are arranged in a descending staircase pattern from top to bottom.

# Self-learning and classroom activities

[Research techniques & Statistics  
in Psychology]

RUMEN MANOLOV  
ANTONIO SOLANAS  
DAVID LEIVA  
REGINA M. SELVAN

[2014]

**Rumen Manolov, Antonio Solanas, David Leiva, & Rejina M. Selvam**

Department of Behavioural Sciences Methods

Faculty of Psychology

University of Barcelona

### **Acknowledgements**

This text was revised by Joe Graham working at *Serveis Lingüístics* at the University of Barcelona. The revision was made with the support of the *Serveis Lingüístics*, thanks to a grant “Interlingua” by the *Generalitat de Catalunya*.

The document is a product of the teaching innovation project “Actividades de aprendizaje autónomo y presencial en Metodología de las Ciencias del Comportamiento” (2014PID\_UB/003) carried out by Rumen Manolov, Maribel Però, Antonio Solanas, David Leiva, Joan Guàrdia, José Luis Losada, Rejina Mary Selvam, Lluís Salafranca and Jaume Turbany.

---

## CONTENT

Preface .....	4
Section 1: Writing and looking for reports .....	7
Research in Psychology .....	8
Bibliographic searching .....	10
Section 2: Collecting data .....	16
Observation .....	17
Interview .....	21
Questionnaires & surveys .....	24
Interview & Questionnaires .....	27
Experiments: The Stroop effect .....	30
Social networks .....	41
Section 3: Probability models .....	48
Extrasensory perception .....	49
Omnipresence of the normal distribution .....	56
Application of Normal and $t$ distributions to case-control studies .....	61
Section 4: Descriptive statistic .....	69
Content analysis .....	70
Univariate description .....	76
Diversity .....	79
Bivariate description .....	86
Dyadic concordance in personality .....	95

---

Section 5: Inferential statistic .....	98
Sampling distribution of a statistic .....	99
Statistical process control .....	103
Confidence interval .....	110
One-sample tests .....	116
Association between categorical variables .....	119
The $t$ test .....	122
Nonparametric tests .....	127
Correlation between the Big Five factors .....	132
Inconsistences in correlation coefficients .....	136
Analysis of variance .....	139
Section 6: Modelling .....	142
Linear regression models .....	143
Polynomial regression: The Yerkes-Dodson law .....	147
Multiple regression models .....	150
Nonlinear regression models (1) .....	156
Nonlinear regression models (2) .....	159
Concluding remarks .....	161

## Preface

The current document is a resource accompanying lectures or do-it-yourself learning of topics related to research methods and statistics in their application to behavioural sciences. The main idea underlying its creation is to relate these usually unattractive topics to substantive psychological content. Establishing the relations between psychological knowledge and the way in which it is constructed and assessed aims to illustrate the usefulness of methodology and statistics and, hopefully, to increase the motivation of students to dedicate time and attention to this topics. We consider that being acquainted with how data are gathered and analysed following the scientific method is necessary for all professionals, if they are to be wise consumers of research reports and capable researchers on their own. Full understanding of the actual down-to-earth implications and meaning of the results of a study requires a knowledge base that has to be constructed via consulting theoretical material and practicing – it is this latter aspect which we aim to promote with the current document. “Learning by doing” is also useful for gaining first-hand experience on how to carry out an investigation. This is why a great deal of the practical exercises proposed involves data collection, apart from data analysis, as the two parts of the research process are closely related. We also wanted the students to be familiar with the data, instead of only working with fictitious studies and meaningless numbers. This notwithstanding, performing calculations on already provided data is also useful for understanding better the meaning of the information yielded by each statistical index or test.

The document is organized in six sections. Section 1 is very short (it comprises two practices) and deals with scientific reports: how to write them and how to locate them, emphasizing the importance of consulting databases and not just surf on the Internet, without any control on the content. Section 2 is focused on different data collection techniques: observation, selective methods such as interviews and questionnaires, experiments, and gathering data from the perspective of social networks. Section 3 deals with probability models, including exercises for computing probabilities and some applications. Section 4 is about using descriptive statistics, as applied to the data gathered following the suggestions of the practices included in the second section. Section 5 is about the inferential use of statistics, relating the topics with which Sections 3 and 4 deal. Section 6 presents possibilities to represent the relationship between variables via regression models: the uses of this modelling process can be both descriptive (as in the content of Section 4) or inferential (Section 5).

Given the abovementioned characteristic of the document, we consider that it can be used by two types of collectives. On the one hand, the material created is potentially

useful for teachers in their task to propose meaningful exercises to students willing (or forced) to study methodological and statistical content. On the other hand, the exercises proposed and the questions asked can be dealt with by the students on their own, while they are studying the topics. In that sense, some of the practices included in this document can be used as complementary material and not necessarily be treated in class.

As a very general, how-to-use advice we recommend the students to read carefully the theoretical and contextual explanations provided before engaging in the practical activities and to consult the additional references to the primary sources of information. Additionally, in order to fully benefit from the practical exercises proposed and to be able to answer the questions made, it is necessary to consult the theoretical material of the research methods or statistical course that the student is enrolled in. For that purpose, we suggest the following readings as useful for the topics dealt with in the current document.

### ***Materials in English***

#### Research Methods

Coolican, H. (1999). *Research methods and statistics in psychology* (3rd ed.). London, UK: Hodder & Stoughton.

Meltzoff, J. (1998). *Critical thinking about research: Psychological and related fields*. Washington, DC: American Psychological Association.

Nestor, P. G., & Schutt, R. K. (2012). *Research methods in Psychology*. London, UK: Sage.

#### Statistics

Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example* (4th ed.) London, UK: John Wiley & Sons.

Cohen, J. & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Gravetter, F. J., & Wallnau, L. B. (2009). *Statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth.

Spiegel, M. R., Schiller, J., & Srinivasan, R. A. (2009). *Probability and statistics* (3rd ed.). New York, NY: McGraw-Hill.

---

### ***Materials in Spanish and Catalan***

#### Research Methods

- Ato García, M. (1991). *Investigación en ciencias del comportamiento. I: Fundamentos*. Barcelona: PPU.
- Losada, J.L. y López-Feal, R. (2003). *Métodos de investigación en ciencias humanas y sociales*. Barcelona: Paraninfo-Thomson.
- Meltzoff, J. (2000). *Crítica a la investigación: Psicología y campos afines*. Madrid: Alianza.
- Ponsoda, V. (2009). Metodología al servicio del Psicólogo. *Papeles del Psicólogo*, 31(1), 2-6.  
Disponible en: <http://www.papelesdelpsicologo.es/pdf/1791.pdf>
- Riba, C. (2007). *Metodologia qualitativa en l'estudi del comportament*. Barcelona: UOC.

#### Statistics

- Guàrdia, J., Freixa, M., Però, M. and Turbany, J. (2008). *Análisis de datos en psicología* (2ª ed.). Madrid: DELTA Publicaciones.
- Peró, M., Leiva, D., Guàrdia, J. y Solanas, A. (Eds.) (2012). *Estadística aplicada a las ciencias sociales mediante R y R-Commander*. Madrid: Garceta.
- Solanas, A., Salafranca, Ll., Fauquet, J. y Núñez, M. I. (2005). *Estadística descriptiva en ciencias del comportamiento*. Madrid: Thomson.

## **Section 1:**

### **Writing and looking for reports**

A technical report is the physical written description and summary of an investigation. It is supposed to be a truthful representation of the different stages of the research process and it serves several purposes: a) to inform about new findings and their integration with the knowledge already available; b) to enable interested readers and referees to make their own assessment of the research; c) to make replications possible. In the context of a methodological course it also serves for illustrating concepts such as methodological strategies, variables and their operationalization, hypotheses, etc.

The aim of the first practice included in this section is for students to identify the specific manifestations of these concepts and also detect strengths and weaknesses of the research. The aim of the second practice is related to the second purpose mentioned above – to stress the importance of the peer review process and the need to build the case for an investigation on peer-reviewed reports. Some clues are given about how such reports can be located. Finally, a written report should meet several formal criteria such as the Introduction-Method-Results-Discussion structure (first practice of this section) and the APA citation style (second practice of this section).

Correct reporting enhances correct conduct of research and both make possible the progress of knowledge.



---

## PRACTICE ON RESEARCH IN PSYCHOLOGY

### Aim of the practice

The first aim of the practice is to identify, in the context of a specific study, the main elements of the scientific method in its application to psychological research. Therefore, the objective is to illustrate basic concepts such as variable, hypothesis, validity, etc., and to see how the knowledge of these concepts can be useful when interpreting the information that a scientific work does or does not provide. The second aim is to become familiar with the formal structure of a scientific article – the type of report used to publish the results of an investigation. The organisation of the information into different sections will be discussed, as well as aspects related to style, in order to maintain scientific rigour when writing the report. Finally, a tentative critical appraisal of the investigation will be made. The underlying objective is to show that no scientific work is definitive or perfect and this is why future research should fill in the existing knowledge gaps or methodological flaws. This final aim responds to the type of work that psychologists perform. Although not all will necessarily carry out scientific studies, it is very likely that all will have to consult scientific literature as a basis for their everyday work. Therefore, to perform adequately in a professional context, it is crucial to be able to distinguish evidence with a solid scientific basis from the evidence that lacks such desirable characteristics.

### Description of the tasks

0. Before the practice session: carefully read the article in order to know the main content of each section.
1. During the practice session, discuss the information provided by:
  - a. Abstract;
  - b. Introduction;
  - c. Method, focusing on each of the subsections;
  - d. Results;
  - e. Discussion;
  - f. References;
  - g. Appendix (where available).
2. Identify and comment on the main variables which are relevant to the study:
  - a. Classify them according to the methodological criterion: discuss whether the explanatory or predictor variables can be considered “independent”;

- b. Comment on the operational definitions of all relevant variables, not only for the explained (criterion) variables;
  - c. Classify them according to the statistical criterion: type of variable (qualitative vs. discrete quantitative vs. continuous quantitative) and measurement scale.
3. Comment on the methodological strategy followed in the study and its internal and external validity.
4. Comment on the positive aspects related to: introduction, method, results and discussion.
5. Comment on the aspects for improvement related to: introduction, method, results and discussion.

### **Material necessary for the practice**

- Virtual Campus: Artículo\_cast.pdf (in Spanish).
- Virtual Campus: Article\_eng.pdf (in English).

### **Some questions to think about and answer**

1. What are the main similarities with an article in the press?
2. What are the main differences from an article in the press?
3. Is there a section called “Data analysis”? If so, which is it? If not, where should it be placed in the article?
4. Is there a section called “Instruments”? If so, which is it? If not, where should it be placed in the article?
5. How would you summarise the main research findings, given the limitations of the study?
6. Out of the four main scientific aims, which do you think might be intended in this study?

### **Suggested readings**

Meltzoff, J. (1998). *Critical thinking about research: Psychology and related fields*. Washington, DC: APA.

Meltzoff, J. (2000). *Crítica a la investigación. Psicología y campos afines*. Madrid: Alianza Editorial.

[Back](#)

## PRACTICE ON BIBLIOGRAPHIC SEARCHING

### Aim of the practice

The present practice intends to illustrate how to access scientific knowledge with certain guarantees. This way of accessing information is useful in carrying out the supervised research during the course and also for any psychologist's professional career. The initial idea is that science is a process in continuous progression, using available knowledge as a foundation in order to extend it, refine it and replace it with other knowledge that, currently, seems to provide more plausible descriptions or explanations of the phenomena of interest. Therefore, any scientific work should necessarily be based, to a greater or lesser extent, on previous research in order to replicate findings, explore suggested areas for future research or deal with the same object of study from a different theoretical or methodological perspective. In order to make this possible, it is necessary to get to know the research relevant to the topic of interest, seeking out evidence that is currently useful (from a theoretical or applied perspective) rather than trivial.

In the process of planning the supervised research, the students should take the available knowledge into account so that they can define specific aims for their own work and locate it in the context of existing theories or empirical evidence. According to whether theories exist or not (and thus whether hypotheses can be derived), the course work will follow either a deductive or an inductive approach. It is important to consider whether the *a priori* expectations are theoretically or empirically grounded or, in contrast, they constitute the aim in exploring the topic of interest. The interpretation of the study's results is related to these initial expectations. Therefore, the theories and research reviewed in the section "Introduction of scientific report" are also the ones that have to be considered when making sense of the results and commenting on their theoretical and/or practical implications. This assessment of the findings takes place in the section "Discussion" and it also enables a comparison of the students' results with previous research in order to identify points of convergence and divergence and look for, at least tentatively, possible explanations. In order to test the plausibility of these explanations, further research can be suggested. Thus, it is clear that scientific literature is necessary both to provide a rationale for the research and to locate its results and suggest new research projects. Nevertheless, it should be kept in mind that apart from the conceptual contributions of previous research, the latter is also useful at a methodological level, given that it might provide operational definitions for the variables of interest: previous studies are specific examples of ways of proceeding when carrying out the research. Other methodological aspects on which articles give information include how participants are accessed, what tasks they have to carry out, the features of the instruments used, etc.

At this point, it should be mentioned that the authorship of all the information consulted should be reflected in the reports written by a researcher (or groups of students, in the context of the course). The ideas and results that are a product of other researchers' work should be acknowledged as such. This is important not only from an ethical point of view, but also to provide the interested reader with the possibility of accessing these sources. In the next two sections, we will comment on some aspects that need to be known when accessing scientific works and also to specify in a formally correct way the references of the works used when carrying out one's own research and when writing the report describing this process.

### **Databases**

It is recommended that the initial point of a bibliographic search should be the databases to which the University subscribes, given that these constitute a necessary filter for exclusively accessing works with a positively assessed methodological rigour. In that sense, it ensures that the information used as a basis for the study has been obtained using the scientific method and is not based on common knowledge (e.g., common sense, personal opinions, anecdotes). Another advantage of these databases is that it is currently possible to search in several at once, for instance, PsycInfo, PsycArticles and ERIC, which is an educational database. They also permit searches for studies relevant to Psychology, given the unclear frontiers between disciplines. Additionally, PsycArticles not only helps in the identification of articles of interest, but also obtains their digital version directly via the University.

When performing a bibliographic search, the keywords used in locating relevant works are essential. In order to know these terms, use can be made of the Thesaurus option, which shows the equivalence between words and also their relations. It is important to note that the search can be done in an unspecific way through a whole text or by focusing on the title of a work, a journal name or articles by selected authors. The search can also be restricted to specific populations (according to diagnosis, age group, country of birth) or using certain types of methodologies (e.g., only experiments or only longitudinal studies). Finally, it is commonly recommendable to limit the search to journals that use peer review as part of their assessment of the scientific merit of the work before publishing it. Peer review entails experts in the field evaluating the strengths and limitations of a work in terms of its theoretical or practical relevance, the adequacy of the methodology used to obtain and analyse the information and the validity of conclusions derived from the evidence collected. Once a potentially interesting article is located, the next step is to read the abstract, which contains information on each of the article's sections. If the article proves to be relevant for the study that the researcher wants to carry out, s/he can access it digitally directly via the database or, when this is not possible, through the electronic resources or the library of the University. When

books and book chapters are to be located, in almost all cases it is necessary to consult the printed versions in the university library.

## References

As previously mentioned, references are indispensable for making clear which of the ideas in a scientific report do not belong to the researcher, but are rather the product of consulting previous research. Making reference to other authors and previous studies also has the function of providing a rationale for the current research, given that manuscripts with a strong foundation are more readily accepted in the scientific world (i.e., in the peer-review process).

At that point, two important distinctions need to be made. Firstly, a quotation is not the same as a reference, with the former term being reserved for those cases in which an author decides to literally reproduce part of an already published report (e.g., an article or book chapter). Quotations are identified using quotation marks and the page(s) of the original text from which they are extracted; their use is restricted to the occasions in which an author considers that it is important not to interpret or summarise the initial idea of the work consulted. Secondly, a distinction should be made between a “Bibliography” and a “References” section. In the former case, what is included is the literature consulted or considered to be relevant for an in-depth knowledge of the topic. In contrast, the “References” section includes only the works mentioned explicitly in the text of the report, identifying the author to whom each idea/finding belongs. Scientific works, but not textbooks used for didactic purposes, should only have a “References” section.

Below, we present the rules for preparing references according to the [American Psychological Association](#) (2009). This reference style is not the only one accepted in Psychology, but it is the most common one and this is the reason for including it in the present practice. Specifically, we illustrate how the references for three types of texts should be constructed: articles, books, and book chapters in edited books, in which each chapter has different authors. These are the three most frequently used sources of information. Moreover, we have purposefully excluded the rules for referencing other types of text, such as information obtained from the Internet without the mediation of scientific databases, given that there are no guarantees that the opinions expressed in such documents have any solid scientific foundation.

*Article.* The general rule for preparing a reference for the “References” section of a scientific report is presented below, along with an example:

Author, A. A., & Author, A. A. (2013). Article title. *Title of the Journal*, 10 (1), 10-20.

Vierhaus, M., Lohaus, A., & Shah, I. (2010). Internalizing behavior during the transition from childhood to adolescence: Separating age from retest effects. *European Journal of Psychological Assessment, 26* (3), 187-193.

In order to make reference to this work in the text of the report (in any of its sections: “Introduction”, “Method”, “Results” or “Discussion”), only the surnames of the authors and the year of publication are necessary (i.e., Vierhaus, Lohaus, & Shah, 2010). Given that in this case there are more than two authors, all subsequent references in the text will include only the surname of the first author (i.e., Vierhaus et al., 2010).

*Book.* The general rule for preparing a reference for the “References” section of a scientific report is presented below, along with an example:

Author, A. A., & Author, A. A. (2011). *Book title*. City, STATE: Publisher.

Lancy, D. F., Bock, J., & Gaskins, S. (2010). *The anthropology of learning in childhood*. Walnut Creek, CA: AltaMira Press.

In order to make reference to this work in the text of the report (in any of its sections: “Introduction”, “Method”, “Results” or “Discussion”), only the surnames of the authors and the year of publication are necessary (i.e., Lancy, Bock, & Gaskins, 2010). Given that in this case there are more than two authors, all subsequent references in the text will include only the surname of the first author (i.e., Lancy et al., 2010). Therefore, when referring to a book in the text, the rule is identical as for articles.

*Chapter of an edited book.* The general rule for preparing a reference for the “References” section of a scientific report is presented below, along with an example:

Author, A. A., & Author, A. A. (2011). Chapter tittle. In E. E. Editor & E. E. Editor (Eds.), *Book title* (pp. 100-200). City, STATE: Publisher.

Gaskins, S., & Paradise, R. (2010). Learning through observation in daily life. In D. F. Lancy, J. Bock, & S. Gaskins (Eds.), *The anthropology of learning in childhood* (pp. 85-118). Walnut Creek, CA: AltaMira Press.

In order to make reference to this work in the text of the report (in any of its sections: “Introduction”, “Method”, “Results” or “Discussion”), only the surnames of the authors and the year of publication are necessary (i.e., Gaskins & Paradise, 2010). Given that in this case there are more than two authors, all subsequent references in the text will include the surnames of both authors as well (i.e., Gaskins & Paradise, 2010). Therefore, when referring to a book chapter in the text, the rule is identical as for articles and books.

*Number of authors.* It is important to highlight that there are rules regarding the number of authors of a scientific work, regardless of whether it is an article, a book or a

book chapter. The APA style requires that when a text has one or two authors, their names should always appear in the text, each time their work is referenced in it. In contrast, when there are 3 to 5 authors, their names appear in the text only the first time and all subsequent references include only the name of the first author followed by “et al.”. If a work has 6 or more authors, then, starting with the first reference, only the name of the first author is specified and followed by “et al.”. These rules are applicable only to the main text but not to the “References” section, where the names of all authors should appear, unless there are more than 6 (in the latter case “et al.” is used after the name of the sixth author).

*Differences according to language.* Finally, it is important to stress that this way of preparing references corresponds to reports written in English, although it is also acceptable for texts written in Spanish or Catalan. Nonetheless, it is more appropriate that the references be prepared in the same language. There are two main differences of style: a) the names of the penultimate and last authors are separated with an “y” (Spanish) or “i” (Catalan) instead of an ampersand (&); moreover, the comma after the name of the penultimate author is omitted; b) in order to specify the book to which a chapter belongs, “En” is used in Spanish and “A” is used in Catalan instead of the English “In”.

### **Description of the tasks**

1. Access one of the Psychology databases (probably also including related fields) in order to gain familiarity with how to proceed in terms of:
  - a. Defining search criteria and keywords
  - b. Identifying potentially interesting articles
  - c. Accessing the digital version of these articles
  - d. Verifying whether it is necessary to consult scientific literature in English, given the amount of studies available in Spanish and Catalan and also for accessing up-to-date and relevant knowledge, which is generally published in English.
2. Look for articles related to the topic selected for the supervised research: Access the articles of interest.
3. Prepare the references of the articles identified or accessed in order to start writing one of the sections of the written report. Think about how the references should look, according to the number of authors.

### **Material necessary for the practice**

- Databases to which the University of Barcelona subscribes.

**Complementary readings**

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Nestor, P. G., & Schutt, R. K. (2012). *Research methods in Psychology*. London, UK: Sage. (Chapter 2).

[Back](#)



---

## **Section 2:**

### **Collecting data**

There are several ways of gathering empirical evidence in the context of the behavioural sciences. The set of practices included in this section is not intended to be comprehensive; it is rather supposed to focus on the main instruments and procedures. The reader should note that some of the practices included in other sections also entail data collection prior to statistical analysis.

In the present section we include examples of the three methodological strategies: observation, the selective/correlation approach, and experimentation. We have also included a practice related to social networks analysis, as one of the possible data collection procedures focusing on groups. We hope that the reader will clearly see that each of these techniques can be the optimal choice according to the research aims. Moreover, integrating the contents of the different sections, the reader can verify that the data analysis technique is more closely related to the type of data gathered (i.e., to the measurement scale), whereas internal validity is related to the methodological strategy followed, external validity to the sampling technique, and ecological validity to the setting and the task that the participants were asked to perform.

Finally, this section ends with the stage of the research process that links data collection and data analysis: the organization of the information in data matrices. Clues are given about how data are usually disposed and what initial checks can be carried out in order to avoid incorrect or misleading results due to errors in data entry.

## PRACTICE ON OBSERVATION

### Introduction

The observation to be carried out is focused on a topic inherent to psychology: social skills. Social skills are closely related to assertive behaviour, which as Lazarus (1973) points out includes the ability to say “no”, ask for favours, express positive and negative feelings, and start, maintain and end conversations. Social skills are manifested in the communication patterns considered to be desirable within a specific culture and they take into account factors such as the context (e.g., family and work), the role of the individuals within this context, their age, etc. Therefore, a behaviour is labelled as adequate according to the situation and the characteristics of the people present. Beyond the structural aspects of the behaviour, what is most important is the efficacy of the situation, that is, the consequences of the behaviour, i.e., whether the person whose social competence is being assessed achieves his/her purpose or not. This purpose is likely to include some of the following aims: to maintain or improve the relationship with another person or to enhance self-esteem without causing any harm to the other person.

One topic related to social skills is nonverbal language. This includes communication channels, which are as useful as, or even more useful than, the content of words. Among the aspects of nonverbal communication that help in the transmission of a person’s message, we should highlight facial expressions, gestures and the tone of voice in which words are spoken. These aspects are useful not only to aid in the correct interpretation of the message, but also to detect inconsistencies and any possible intention to deceive or lie.

The present practice deals with the behaviour of people interacting in two different social contexts. It is possible to record how these people behave, both verbally and nonverbally, and, as a function of the aims we can attribute to them, we can assess whether they show social skills in these situations or not.

### Aim of the practice

The aim of the present practice is to carry out an observation, aided by two kinds of recording, in order to obtain empirical information. These two recordings are common in the initial stages of systematic observation, with later stages focusing on the progressive systematisation of the instruments in order to devise a coding scheme. The underlying

idea of the practice is that students follow these stages until the final product, which will be further discussed when working on observational methodology as a way to plan psychological studies (in the subject called “Diseños de investigación”).

### **Situation 1: Description**

A common situation in psychological practice is that one psychologist (especially in the beginning of his/her professional career or when facing a challenging case) asks for supervision by a more experienced colleague. Imagine that such a request has been made by the psychologist in Video 1, with his aim being specifically for an expert colleague to assess his communicative skills in the psychotherapy sessions. The supervisor and the psychologist decide jointly that an adequate way of obtaining information on social skills would be to record on video the sessions with a specific client with whom the psychologist has recently started working.

### **Situation 1: Tasks**

1. Watch Video 1 and carry out a narrative recording of the communicative behaviours of the psychologist, including all the aspects that you consider necessary. (There are no precise indications, given that this is the first time we see the psychologist and his behaviours, which are the aim of this study.)
2. Consider that the psychologist wants the client to start trusting him and convince him of something. Prepare the descriptive recording out of the narrative one, considering how the psychologist responds to or provokes the behaviours of the client.
3. Obtain a “traits list” – all the main observable s, with no repetition.
4. Organise the behaviours in the list into a “catalogue” – grouping them according to the response level (molecular) or according to another substantive criterion, such as the aim assumed to underlie the psychologist’s behaviour (molar).
5. Select and define the categories in such a way that the characteristics of a category system such as a coding scheme are met. This coding scheme should be useful in studying the behaviours of the psychologist objectively.

### **Situation 1: Questions to think about and answer**

7. Why is observation used as a recording technique?
8. What is the reason for video-recording only the sessions with one of the clients? Which kind of validity is affected by this fact? What kind of time sampling for the observation would you suggest in this situation?
9. Considering the way in which you carried out the observation to obtain the descriptive recording, answer the following question: What is the type of observation in terms of degree of participation?

10. Suppose that the situation is as described here. Would there be reactivity in the participant being observed?
11. According to the way in which the behaviours are grouped in the traits list, is the approach followed etic or emic?
12. How can it be controlled so that the information does not exclusively depend on the subjective interpretation of the supervisor?
13. How can you obtain information on whether the social skills of the psychologist have improved after the supervision and intervention? What sources of information could you consult?
14. Is there an ethical aspect that should be considered in this situation?

### **Situation 2: Description**

Imagine that a client asks for a consultation with a psychologist in order to improve his social skills in everyday life situations. The psychologist and the client decide jointly that in order to obtain information, the client should wear a hidden camera to record his interactions. In this way, evidence will be obtained on the initial situation: the strong aspects and the ones that need to be improved, so that the psychologist can make an informed decision on how to intervene. The recordings will also be useful in supervising the progress of the client once the psychologist has shown him techniques directed toward improving his social skills.

### **Situation 2: Tasks**

1. Watch Video 2. Suppose that the underlying client intention is that the female character steps off the train with him. While you are watching the video, prepare the descriptive recording with the male character's behaviour as central and including the antecedent and consequent behaviours of the female character.
2. Considering that a person with social skills would take into account the behaviours of his partner to check whether the aims and desires of both are well-aligned, rate the social skills of the male character using a scale ranging from 1 (minimal) to 5 (maximal).
3. While producing the recording, keep in mind that behaviour can be expressed at different levels of response; comment on their importance for the communication between the characters.

**Situation 2: Questions to think about and answer**

1. Why is observation used as a recording technique?
2. Can an interview be useful in this situation?
3. What kind of time sampling would you suggest for this situation?
4. Considering the way in which you carried out the observation to obtain the descriptive recording, answer the following question: What is the type of observation in terms of degree of participation?
5. Suppose that the situation is as described here. Would there be reactivity in the participant being observed? What effects could this reactivity have (if any does take place)?
6. If the psychologist asks the client about the intentions underlying his behaviours while they are watching the video, which is the approach used for categorising these behaviours – emic or etic?
7. What objective indicators could be used to assess whether the client has improved his social skills?
8. Is there an ethical aspect that should be considered in this situation?

**Materials necessary for the practice**

## Situation 1

- Link in Spanish: <https://www.youtube.com/watch?v=R6eqqFbcknY>
- Link in English: <https://www.youtube.com/watch?v=UYa6gbDcx18>

## Situation 2

- Link in Spanish: <https://www.youtube.com/watch?v=MgIoV4LDQJ0>
- Link in English: <https://www.youtube.com/watch?v=DQ8Q6VuyB74>

**Suggested readings**

Caballo, V. E. (2007). *Manual de evaluación y entrenamiento de las habilidades sociales* (7ª Ed.). Madrid: Siglo XXI.

Lazarus, A. A. (1973). On assertive behavior: A brief note. *Behavior Therapy*, 4, 697-699.

Rodríguez Escanciano, I. y Hernández Herrarte, M. (2010). *Lenguaje no verbal: Como gestionar una comunicación de éxito*. A Coruña: Netbiblo.

[Back](#)

## PRACTICE ON INTERVIEW

### Aim of the practice

The purpose of this practice is to create a template or script for an interview on a predefined topic. As always, the template and the type of interview need to be appropriate for the kind of information that the researcher seeks to obtain. Afterwards, the interview will be carried out as a pilot study in order to practice the whole process from beginning to end. Once the pilot study is done and the definitive template is determined, the students are to collect information from a smoker they are acquainted with. This information will later be analysed and interpreted in the second part of the course (univariate and bivariate statistics). Therefore, it is compulsory to collect and send this information so that it can be used for future practice sessions.

The aim of the interview is to gain familiarity with the factors related to smoking in order to intervene, although the advice is not always to quit smoking, especially considering the possibility of a subsequent increase in weight (Fiore & Baker, 2013). Using the interview, it is necessary to obtain the type and quantity of information which is adequate and sufficient for designing (hypothetically) an individualised therapy focused on smokers in order to convince them and help them to quit smoking.

On the one hand, it is important to get to know the person and his/her characteristics, given that personality traits (Perea-Baena, Oña-Compan, & Ortiz-Tallo, 2009), as well as age, gender, and education level (Sánchez, Olivares, & Rosa, 1998), have been shown to be related to whether anti-smoking therapies are successful or not. This information is also needed because there is no single explanation of smoking behaviour or a single accepted therapy, although some are supported by more evidence than others (Sánchez et al., 1998). Even in the context of the commonly used and frequently effective self-control therapies, the intervention must be adjusted to the specific variables that maintain the behaviour (Rosa, Sánchez, & Olivares, 1998).

It is seen as important that the students take into account the variables related to the smoking habit, for instance, the number of cigarettes smoked per day. Another suggested topic to include in the interview is the relative importance (or severity) of the smoking in comparison to other potentially harmful habits, such as drinking alcohol and gambling. Finally, it would also be relevant to get to know what would motivate or encourage the participant to quit smoking. In order to gather detailed information on the topic, students are allowed to ask any questions they consider appropriate and necessary, focusing on topics such as the smoking habit and the best ways of helping the person to reduce or eliminate this habit.

### Description of the tasks and questions to think on

1. Decide the topics in the interview, including the ones suggested and the ones derived from the texts referenced and from your own experience. However, you should take into account the influence of your experiences when assessing the answers of the participants (i.e., the “reflexivity” concept).
2. According to the degree of structuring and the type of interview: write the questions or the template.
  - For which thematic blocks did you create specific questions? What kind of questions: open, appraisal, multiple choice, etc.?
  - What order do you have in mind for the questions or thematic blocks?
  - In this specific case, what are your expectations about the veracity of the information provided by the interviewee? Why?
  - What strategy would you use to motivate the interviewee?
3. Carry out the interview with a classmate. Take the role of the interviewee so that your classmate can also carry out his/her pilot study.
  - What difficulties did the interviewer come across and how did s/he solve them? What would you change if you had to carry out the interview again?
  - Did the interviewer follow *all* necessary steps or did s/he just focus on asking questions?
  - What was the experience of the interviewee?
  - What type of interview did you carry out according to the classification criteria: degree of structuring, type of content, type of response required from the interviewee, intensive or extensive aim? Why?
  - What recommendations would you give to the interviewer in terms of: motivating participation, maintaining the flow of the interview, achieving the aims of the interview, getting the interviewee to want to participate in a potential follow-up interview?
  - Would there be differences if the same information had to be obtained via a questionnaire?
  - How would you organise the information gathered via the interview?
  - If you interpreted the information obtained in the practice-session interview, would this interpretation be emic or etic? To answer this question, think of the “type of observation” fundamental to this interview.

---

**Material to be used in the practice**

- The interview template: its format depends on the type of interview.
- One option is to take notes on what is considered to be most important in the responses of the participant. In order to avoid losing information and considering that each summary implies an interpretation or evaluation of the answer, an audio recorder can be used.

**References**

- Fiore, M. C., & Baker, T. B. (2013). Should clinicians encourage smoking cessation for every patient who smokes? *The Journal of the American Medical Association*, *309*, 1032-1033.
- Perea-Baena, J. M., Oña-Compan, S., & Ortiza-Tallo, M. (2009). Diferencias de rasgos clínicos de personalidad en el mantenimiento de la abstinencia y recaídas en tratamiento del tabaquismo. *Psicothema*, *21*, 39-44.
- Rosa, A. I., Sánchez, J., & Olivares, J. (1998). Técnicas de autocontrol y adicción al tabaco: Meta-análisis de la literatura española. *Análisis y Modificación de Conducta*, *24*, 829-850. Available at <http://www.um.es/metaanalysis/pdf/7038.pdf>
- Sánchez, J., Olivares, J., & Rosa, A. I., (1998). El problema de la adicción al tabaco: Meta-análisis de las intervenciones conductuales en España. *Psicothema*, *10*, 535-549.

[Back](#)



## PRACTICE ON QUESTIONNAIRES & SURVEYS

### Aim of the practice

The purpose of this practice is to create a questionnaire including the questions that are potentially relevant for collecting information on a specific topic – how to discourage people from smoking. The questions are to be formulated according to the answers required from the participants. Afterwards, the resulting questionnaire will be administered in a pilot study and to practice the opening and closing of the survey process. Therefore, the aim is that students follow, albeit on a small scale, the whole process of a piece of research planned as a survey.

The aim of the questionnaire is to present different pieces of information and/or visual material, bearing in mind the findings that graphic material, especially if it has high emotional content, affects attempts to quit smoking: [Farrelly et al., 2012](#); [Nonnemaker et al., 2013](#)). It is necessary to choose this information and material according to the corresponding search, for instance, taking into account the messages written on packets of cigarettes. It is suggested that students quantify the degree to which these elements are discouraging, as well as exploring improvements to the instrument. Finally, students should take into account whether a discouraging measure has this effect only for participants with some characteristics but not for others. The formulation of the questions and the decision on the thematic blocks relevant for reducing or eliminating the habit of smoking are up to the students' criteria and creativity; possible dimensions include social, professional and health areas.

### Description of the tasks and questions to think about

1. Choose the information that may be useful as a measure to discourage the habit of smoking. The empirical evidence gathered via the interview regarding the main reason that would make participants quit smoking may be relevant.
2. Choose the visual material that may be useful as a measure to discourage the habit of smoking. The empirical evidence gathered via the interview regarding the main reason that would make participants quit smoking may be relevant.
3. Formulate the questions that you consider to be appropriate. Decide the type of question to use in relation to the type of information that the researcher desires to obtain and how this information can be later by organised and analysed (e.g., open with more or less space left for answering; multiple-choice, with alternatives which are or are not exclusive; true/false; assess the degree to which a participant agrees

with a certain statement using a Likert-type scale or a visual analogue scale; assess the frequency; sort according to preference).

- What types of questions did you include in the questionnaire? Why?
  - Is the questionnaire one-dimensional or multi-dimensional?
  - Are all Likert-type scale questions written in the same direction, positive or negative? Which of the two approaches would be more appropriate? How does this affect the analysis of the data?
4. Decide the order of the questions.
  5. Prepare an opening and a closing message for the questionnaire as well, selecting the information that these elements need to include. Decide whether the opening and the will be written according to the way in which the questionnaire is administered to the participants.
    - What were the main difficulties in creating the questionnaire? How did you solve them?
  6. Carry out a pilot study in class: apply the created questionnaire to five classmates in order to detect aspects that require improvement. Collect and keep the information.
    - What do you have to do in order to find out whether there are any problematic aspects/elements of the questionnaire?
    - Did you apply the questionnaire in an oral or a written form? Why? Which is better?
    - What were the main difficulties in applying the questionnaire? How did you solve them?
  7. Think about the way in which the gathered information can be organised.
    - What are the main results that you obtained?
    - Could you have used an interview instead? Which would be the main differences in the information applied?
    - What is the measurement scale for the different pieces of information?
  8. What general rules should you follow when: a) writing questions; b) approaching the participants; c) interpreting the information?
  9. Exercise to do outside of the classroom: administer the questionnaire to three people who smoke.

---

**Materials to be used in the practice**

Apart from a knowledge of questionnaires and surveys, the only requirements are a pen and paper. The remaining material should be sought and located by the students.

**References**

Farrelly, M. C., Duke, J. C., Davis, K. C., Nonnemaker, J. M., Kamyab, K., Willett, J. G., & Juster, H. R. (2012). Promotion of smoking cessation with emotional and/or graphic antismoking advertising. *American Journal of Preventive Medicine, 43*, 475-482.

Nonnemaker, J. M., Farrelly, M. C., Kamyab, K., & MacMonegle, A. J. (2013). Do different styles of antismoking ads influence the types of smokers who call quitlines? *Health Education Research, 28*, 23-30.

[Back](#)

## PRACTICE ON INTERVIEW & QUESTIONNAIRES

### Aim of the experience

The aim of the current experience is to elaborate an instrument that would make possible gathering information regarding the performance of a psychologist in his private practice. Specifically, it is intended to get to know the perception of clients on the attention and help received with the purpose of improving the service provided, introducing modifications in the psychological practice, while also keeping in mind the scientific evidence on how the therapeutic relationship (or rapport; [Travelbee, 1963](#)) has to be established and also acting according to the protocols for the application of the different interventions. In fact, the topic of therapeutic engagement has been specifically studied for adolescent patients ([Oetzel & Scherer, 2003](#)), as it is the case of the situation described in the video for the Practice on Observation.

When constructing the instrument, take into consideration the different aspects that can be assessed (and modified) in the work of a psychologist in his private practice. Keep also in mind that the assessment may be related to the characteristics and situation of the particular client. We propose that the instrument be created in such a way as to be applicable both as an interview and a questionnaire. We also recommend that the students look for information on the common process taking place in assessment and/or therapeutic sessions in order to guide their definition of the thematic blocks of the instruments to be created.

### Description of the tasks

1. Decide the thematic blocks, keeping in mind that demographic and contextual information is also necessary.
2. Create the template of the interview or the summary of the topics that the blocks need to include.
3. Create the questions that you consider relevant and necessary. Decide the type of questions used (e.g., open with more or less space for answering, multiple choice, true/false, assessing the degree of agreement with a statement by means of a Likert scale or a visual analogue scale, assessing the frequency, sorting according to preference) as a function of the type of information that is desired and how this information can later be treated/organized.
4. Decide the order of the questions.

5. Create an opening and a closure for the instrument, selecting the information that has to be included. Decide whether these sections would be available in written form or will be expressed orally.
6. Think how to make the data collection flexible, i.e., how to make possible applying the instrument in written form (to be filled in at home) and oral form (by phone or where the psychological practice takes place).
7. Discussion of the product
  - If we think of the instrument as an “interview”, what type of interview would it be? (according to the classification criteria: degree of structuring, type of content, type of answers required from the interviewee, intensive or extensive purpose).
  - What types of questions did you include in the instrument? Why?
  - Are all Likert-scale statements written in the same sense (positive or negative)? Which way of writing would be more appropriate? How does this affect the analysis of the answers?
  - Is the instrument one- or multi-dimensional?
  - What order of the questions and thematic blocks how you suggested?
  - What are the expectations about the trustworthiness of the information provided by the participants? What does it depend on?
  - What strategy would you use to motivate the person that you wish to interview?
  - In which moment/context would you gather information?
8. Bibliographic search: locate studies that have been carried out on the satisfaction with the psychological practice, attention or services of a professional.
  - What aspects are usually considered?
  - Are there any interviews or questionnaires already created?
  - Are they similar to the ones created in class?
  - Is such information usually gathered in written or oral form?

9. Think on how to organize the information that can be collected.
- If a more experienced psychologist carries out the data collection and the interpretation of the information, would the latter be etic or emic? And what if it is another client the person that does that? What are the advantages and limitations in each case?
  - Which is the measurement scale of the different pieces of information?
  - Would it be necessary to re-design the instrument after a pilot study?

### **Materials required**

- Pen and paper for creating the instrument.
- Computer and internet for searching for information.

### **References**

- Oetzel, K. B., & Scherer, D. G. (2003). Therapeutic engagement with adolescents in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 40, 215-225.
- Travelbee, J. (1963). What do we mean by rapport? *American Journal of Nursing*, 63, 70-72.

[Back](#)

## PRACTICE ON EXPERIMENTS: THE STROOP EFFECT

### Introduction

The experimental study of interferences in tasks requiring cognitive functioning has focused on different kinds of stimuli: habits, syllables, cards, books, etc. As a foundation of studies on interferences between two types of codification (words and colours), we need to mention the evidence showing that the task of naming a colour and the task of reading the names of colours require differing amounts of time. These results have been attributed to practice effects and to the different nature of the associative processes required by each task. The studies by [Stroop](#) (1935/1992) are based on stimuli that consist of words with meanings that refer to colours (e.g., the word “red”, the word “green”) and are written in different colours. The aim of these studies is to compare the conditions in which the word denotes the colour in which it is written (concordant stimuli) with conditions in which the meaning of the word and the colour are in conflict (discordant stimuli). The present practice intends to replicate the available findings, taking into account that the experimental conditions in the university room are not equivalent to the ones in a laboratory, which allow for a high degree of control.

In order to carry out the experiment, a mixed-factorial design is used (more detail is given on the “Diseños de investigación” course). According to this design, two groups of participants are formed at random: one group is given the task of naming the colour and the other group of naming the word. Therefore, task type is one of the factors (independent variables). Both groups are subjected to two conditions according to the type of stimulus (the second independent variable): concordant and discordant stimuli. Whereas the first factor involves creating separate groups, the second factor entails repeated measures, which is why the design is mixed-factorial (with an “inter” variable and an “intra” variable). Regarding the repeated measures, a technique called counterbalancing is used in order to divide the order effect equally: half of each group (selected at random) will follow this order: 1) concordant stimulus; 2) discordant stimulus. The other half of each group will follow the reverse order.

### Description of the experiment

1. Divide the group into two: experimenters and participants. Each student is identified with a number. Given that the total number of participants is equal to  $n$ , half of the students can be randomly selected as experimenters using the

statistical packages **R**. The example below is given for a group of 80 people. Note that sampling without replacement should be used (the option “replace” should be assigned the value “false”):

```
> n <- 80
```

```
> sample(1:n,(n/2),replace=FALSE)
```

2. Each experimenter receives two lists of stimuli: A and B.
3. The experimenters are identified by new numbers, ranging now from 1 to  $n/2$ . The following **R** code selects half of the experimenters who will ask the participants to name colours, whereas the other half will ask them to name the word:

```
> sample(1:(n/2),(n/4),replace=FALSE)
```

4. Using the same numbers between 1 and  $n/2$ , a new random selection takes place using **R**: half of the numbers generated represent the experimenters that follow the order A-B; the remaining ones will follow the order B-A.

```
> sample(1:(n/2),(n/4),replace=FALSE)
```

### Tasks for the experimenter

1. The experimenter reads the instructions to the participant and offers the opportunity to discuss any questions that the participant might have before beginning the experiment.
2. The participants are presented the stimuli, one by one, including the test stimuli and the ones that belong to the experiment itself. The time needed for each trial is measured with the stopwatch on a mobile phone. It is necessary to note whether an answer is an error or not.
3. The number of errors and the average time needed to answer are recorded, with the latter being quantified separately for correct and incorrect answers.
4. The roles of experimenter and participant are reversed.

### Material for carrying out the experiment: Necessary for each experimenter

#### *Data collection via the experiment*

1. Put the name of the participant and the experimenter.
2. Specify whether the task is to name the colour or the word.
3. Read the instructions to the participant.
4. In each case, specify whether the first list of stimuli is A or B; do the same thing for the second list.
5. Administer the 8 test trials.
6. Fill in the time used to give an answer.



7. Mark the error cell only if an error is produced.
8. Administer the 40 trials of the experiment.
9. Fill in the time used to give an answer.
10. Mark the error cell only if an error is produced.
11. Compute the average response time and the total number of errors for each list.

### Questions to think about and answer

1. How would you describe in a single word the basis of the conditions in which the experiment is intended to take place?
2. What contextual confounding variables in the experiment, as carried out in the practice session, may potentially have an influence on the results?
3. What participant confounding variables in the experiment, as carried out in the practice session, may potentially have an influence on the results?
4. What is the basis for the external validity of the experiment?
5. Following the description of the experiment, does the study carried out in class follow an inductive or a deductive approach?

### Additional reading

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 121, 15-23.

[Back](#)

Participant: _____ Experimenter: _____				
Task: name _____				
First list: _____			Second list: _____	
<b>Experiment trials</b>			<b>Experiment trials</b>	
#	Time	Error (X if yes)	Time	Error (X if yes)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
	Average time (correct): _____s Average time (errors): _____s	Number of errors: _____	Average time (correct): _____s Average time (errors): _____s	Number of errors: _____

Participant: _____		Experimenter: _____		
Task: name _____				
First list: _____			Second list: _____	
<b>Test trials</b>			<b>Test trials</b>	
#	Time	Error (X if yes)	Time	Error (X if yes)
1				
2				
3				
4				
5				
6				
7				
8				

*Instructions for the participants: Task – name words*

You will now be presented a series of stimuli consisting in words written in different colours. Your task is to name the word, regardless of the colour in which it is written. Answer as fast as possible, while at the same you should try to avoiding committing errors. Firstly, several test trials will be presented so that you can become familiar with the task. Afterwards, the experiment itself will start. If you have any doubts regarding the task, feel free to ask the experimenter. Remember that your participation in the experiment is voluntary and, therefore, you can leave whenever you like. Thank you for taking part in this study.

Participant: _____ Experimenter: _____				
Task: name _____				
First list: _____			Second list: _____	
<b>Experiment trials</b>			<b>Experiment trials</b>	
#	Time	Error (X if yes)	Time	Error (X if yes)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
	Average time (correct): _____s Average time (errors): _____s	Number of errors: _____	Average time (correct): _____s Average time (errors): _____s	Number of errors: _____

Participant: _____		Experimenter: _____		
Task: name _____				
First list: _____			Second list: _____	
<b>Test trials</b>			<b>Test trials</b>	
#	Time	Error (X if yes)	Time	Error (X if yes)
1				
2				
3				
4				
5				
6				
7				
8				

*Instructions for the participants: Task – name colours*

You will now be presented a series of stimuli consisting in words written in different colours. Your task is to name the colour in which each stimulus is written. Answer as fast as possible, while at the same you should try to avoiding committing errors. Firstly, several test trials will be presented so that you can become familiar with the task. Afterwards, the experiment itself will start. If you have any doubts regarding the task, feel free to ask the experimenter. Remember that your participation in the experiment is voluntary and, therefore, you can leave whenever you like. Thank you for taking part in this study.

**LIST A**

**BLACK**

**GREEN**

**BLACK**

**RED**

**BLACK**

**RED**

**BLACK**

**GREEN**

**RED**

**RED**

**GREEN**

**BLACK**

**RED**

**BLACK**

**RED**

**GREEN**

**BLACK**

**BLACK**

**GREEN**

**BLACK**

**GREEN**

**RED**

**RED**

**GREEN**

**RED**

**BLACK**

**BLACK**

**GREEN**

**RED**

---

**GREEN**

**RED**

**BLACK**

**RED**

**BLACK**

**RED**

**BLACK**

**GREEN**

**GREEN**

**BLACK**

**GREEN**

**GREEN**

**BLACK**

**RED**

**GREEN**

**GREEN**

**RED**

**RED**

**GREEN**

**LIST B**

**GREEN**

**RED**

**RED**

**BLACK**

**BLACK**

**GREEN**

**BLACK**

**GREEN**

**RED**

**RED**

**GREEN**

**GREEN**

**BLACK**

**RED**

**GREEN**

**GREEN**

**BLACK**

**RED**

**GREEN**

**GREEN**

**BLACK**

**BLACK**

**GREEN**

**RED**

**RED**

**BLACK**

**RED**

**BLACK**

**RED**



---

**GREEN**

**BLACK**

**RED**

**BLACK**

**BLACK**

**RED**

**GREEN**

**RED**

**RED**

**BLACK**

**BLACK**

**RED**

**BLACK**

**GREEN**

**RED**

**GREEN**

**GREEN**

**BLACK**

**GREEN**

## PRACTICE ON SOCIAL NETWORKS

### Aims of the practice

The purpose of this practice is to study the relationships in each group that carries out the supervised research, bearing in mind that the groups are formed at random in the context of the “Técnicas de investigación” course. Firstly, the practice deals with obtaining information at the group level, in order to become familiar with a specific technique for collecting data. It has to be taken into account that in Social Networks Analysis, the relationship is the focus of attention, whereas personal attributes have only a secondary role. Secondly, graphical representations and numerical summaries will be used in order to illustrate their meaning in this real (not simulated) situation. These graphs and quantitative indices serve to describe the network in terms of density and centrality and to identify the different roles of the actors in the network.

### Description of the tasks

1. Before the practice session, the Material necessary for the practice should be filled in and sent to the teacher.
2. Using the data collected, several of the groups will be chosen at random in order to work with them during the practice session. Then the data will be presented in such a way that it ensures the anonymity of the people involved.
3. Using the answers from the second column of Part 1 of the Material, construct a matrix that refers to whether each individual would (1) or would not (0) work again with each of his/her group mates.
4. Using the answers from the second column of Part 1 of the Material, construct a matrix based on the perceived affinity, that is, using ordinal measures (i.e., directed graphs or digraphs with scores quantifying intensity).
5. Introduce the two matrices including relationships and the two matrices with attributes (one for “mother tongue” and one for “country of birth”) in Pajek: this software can be downloaded freely from: <http://pajek.imfm.si/doku.php>.

### *Creating the data matrices*

The data matrices containing the relationships between the nodes can be constructed following one of the two strategies outlined below.

#### First option: Reading data files

This option is especially useful for large matrices and it requires the previous creation of a document with a *.net* extension, which can be created using the Notepad. It is necessary for this file to have a very specific structure.

Imagine that the matrix includes, for instance, the following information about a group of 5 people:

$$X = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

An easy way to represent the information in the *.net* document is provided below. It is only necessary to specify that there are 5 vertices (i.e., nodes) – each line is a vertex. Directly beneath this specification, the information is introduced in a matrix format.

```
*Vertices 5
1 "v1"
2 "v2"
3 "v3"
4 "v4"
5 "v5"
*Matrix
0 0 1 1 0
1 0 1 1 1
0 0 0 1 0
0 0 0 0 0
1 0 1 1 0
```

In order to use this matrix in Pajek, the user has to click: File → Network → Read... and locate the data file.

Second option: Creating matrices using the Pajek menu

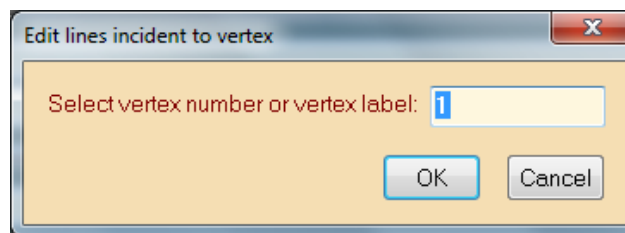
This option is only advisable when the matrix contains few nodes and/or there are few relationships between them. This is so, because the absences of relationship (i.e., the values equal to 0) are not codified. For the same reason, it is not necessary to codify the values of the main diagonal, provided that each participant is not asked about him/herself. The steps to create a matrix with Pajek are:

1) Create an empty network with 5 nodes in order to later introduce the information regarding the arcs manually.

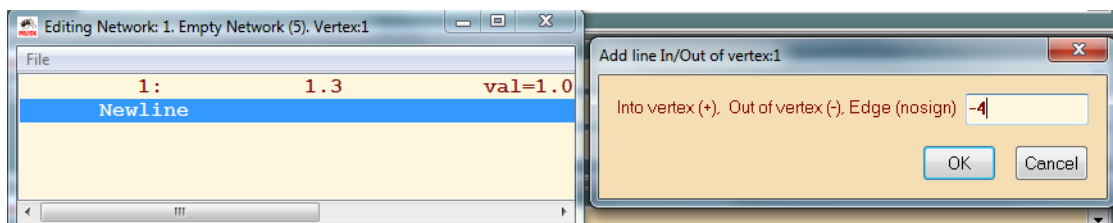
Network → Create new network → Empty network → Enter number of vertices = 4

2) Introduce the data for each node:

File → Network → View/Edit

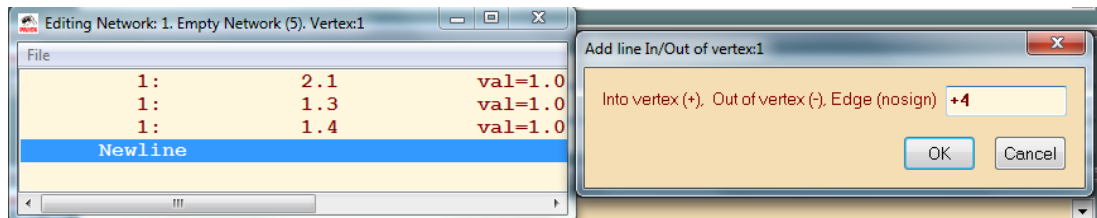


For instance, for vertex 1, the relationships can be introduced without taking into account the direction (i.e., when the matrix is symmetric, using the so called “edges”). This is accomplished specifying only the number of the vertex to which vertex 1 is connected, using no signs. In contrast, when the relationships are directed, the minus sign (-) has to be used to represent a relationship from vertex 1 (emitter) to vertex 4 (receiver), as in the example presented below.



On the contrary, if vertex 1 was the receiver or partner and it were the object of a relationship direction from node 4, then this relationship would be specified using the

plus sign (+), as shown below. Note that it is not necessary to specify anything for those cases in which there is no relation (i.e., the 0s are automatically included).




3) Repeat the same operation for all the nodes. The information can be introduced following the example below. That is, for each node it is necessary only to include the information regarding the nodes with a greater number (e.g., for node 3 there is only information on its relationships with nodes 1 and 2 and, thus, only the relationships with nodes 4 and 5 have to be introduced, if present).

$$X = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

NODO 1  
NODO 2  
NODO 3  
NODO 4  
NODO 5

### *Creating the matrices with attributes*

In order to create the matrices with attributes, the Pajek menu has to be used as follows: Partition → Create constant partition. Afterwards, the students should click on  View/Edit Partition. Using “Val”, the number representing the attribute is defined, and using “Label” it is given a name. Two separate matrices have to be created: one for “mother tongue” and one for “country of birth”.

1. Represent graphically both matrices via Pajek and comment on their differences.

For a graphical representation of a matrix, Draw → Network can be used. In the new window that pops up, the information about the matrices can be completed with all the data available, for instance:

- Options → Mark vertices using → Numbers;
- Options → Lines → Different widths;
- Options → Colours → Partition colours → for vertices (colour according to the attribute).

2. Compute the main numerical indices: degree and betweenness of each node, density, centralisation and  $k$ -cores of the network.

**Degree of each node** (# relationships; allows distinguishing between in-degree and out-degree): Network → Create partition → Degree → All. The result appears in Partitions. With a double-click on the name of the tab, the degree can be seen.

Centrality as **betweenness** for each node: Network → Create vector → Centrality → Betweenness. The result appears in Vectors. With a double-click on the name of the tab, the degree of betweenness can be seen.

**Density of a network and average degree**: Network → Info → General: density2 (no loops / self-relations).

**Centralisation** of a network: Network → Create vector → Centrality → Closeness → All. The closeness centralisation of a network appears in the “report” that is produced automatically, i.e., the closer to 1, the more similar to a network in the form of a star.

**$k$ -core** of the network: Network → Create partition →  $k$ -core → All. The result appears in Partitions. With a double-click on the name of the tab, the core(s) to which each vertex belongs are seen.

3. Interpret the information obtained in meaningful terms.

**Material necessary for the practice:** Information needed to collect data*Part 1: Relationships between the group members*

Definitions:

Do it again?: Regarding each person in the group in which you have carried out the supervised research, you have to answer whether you would work with that person again, if you had the choice. If yes, answer Y. If no, answer no (answer N).

Affinity: using a Likert-type scale ranging from 0 to 5 according to the degree to which you feel close to each groupmate (e.g., in terms of behaviour, ideology, working style or leisure). In order to assess the intensity, take into account your own perception of the relationship, regardless of what the other person might think about it.

Name:

Groupmate*	Do it again? (Y/N)	Affinity (0-5)

\* Anonymity of all students (respondents and their groupmates) will be maintained; random numbers will be assigned to the people when using their data for the practice session.

*Part 2: Personal information*

Name:

Age: \_ \_

Place of residence:

Country of birth:

Mother tongue:

**Questions to think about and answer: Questions on the matrices from the session**

1. Are any of the matrices binary? Are any of the matrices symmetric?
2. Focusing on a group indicator to begin with, to what psychological concept can the density of a network be related?
3. Focusing on another group-level indicator, to what psychological concept can the centralisation of a network be related?
4. Do you think that group-level indicators are good representations of the process in the whole group?
5. At the individual level, to what psychological phenomenon can the betweenness of a node be related?
6. To what psychological phenomenon can the existence of  $k$ -cores be related? To which of the two abovementioned levels of analysis does this indicator belong?
7. What could be the cause of the characteristics of the network, as it is?
8. Considering the information on affinity, where would you intervene in order to improve group dynamics?

**Questions to think about and answer: “Hypothetical” questions**

9. In what kind of group would it be most efficient for a teacher to start a rumour about a book chapter that will be a matter of examination (so that she can be sure that the students read it)? To which of the actors should the teacher tell this information?
10. What type of sampling has to be used in order to reach as many group members as possible?
11. Which actors in a network would be most appropriate for mediating between subgroups?
12. Which relatively isolated actors (with few direct relations) can gain access to group resources?

**Complementary readings**

Hanneman, R. A. (2000). *Introducción al análisis de redes sociales*. University of California, Riverside. Retrieved June 6, 2012 from <http://revista-redes.rediris.es/webredes/> Materiales section.

Hanneman, R. A. & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved July 20, 2012 from <http://faculty.ucr.edu/~hanneman/nettext/>.

de Nooi, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social networks analysis with Pajek*. London, UK: CUP.

[Back](#)



---

### Section 3:

## Probability models

This section deals with one of the most abstract themes, together with the topic on sampling distributions, of any statistical course: probability models. This theme is also the critical one for understanding: a) how information can be obtained from uncertainty and lack of determinism; b) under what conditions can such information be obtained; c) exactly what the quantifications provided by the “moments” of a distribution mean.

In order to illustrate specific applications of the most commonly used probability models we have included three practices that require data collection before proceeding to obtain the quantifications in terms of probabilities. We offer two classical uses of the normal and binomial distributions plus a currently employed application of the  $t$  distribution in neuropsychology. We hope that the reader would be able to think for her/himself of further applications of these and other probability models and the information they could provide regarding substantive topics.

The knowledge of the content of this section is relevant for descriptive statistics (section 4) due to the relation between “moments” of first, second, third, and fourth order and statistical indicators of central tendency, scatter, skewness, and kurtosis, respectively. The content is also necessary for inferential statistics (section 5), given that statistical decisions are made on the basis of  $p$  values, which are derived from probability models assumed to represent (approximately) adequately the behaviour of the test statistics used.

## PRACTICE ON EXTRASENSORY PERCEPTION

### Introduction

Extrasensory perception (ESP) is the acquisition of information using a way that differs from the exteroceptive sensations (i.e., the five classical senses that give us information about the environment), the interoceptive ones (related to the body's internal stimuli), or the proprioceptive ones (through which we obtain information on the position and movement of the body). Sometimes ESP is referred to as the sixth sense, but this use is not widespread.

Telepathy, defined as the transfer of thoughts or feelings between people without the mediation of exteroceptive receptors, is considered to be one of the forms of ESP. There has been a great deal of debate for several years between people who support or reject the possibility of transferring information via telepathy; this lack of agreement is likely to continue into the future. From the perspective of a sceptic, the main point is not whether we are for or against the existence of telepathy; it is rather a question of having empirical proof about its certainty. Therefore, it is reasonable to suggest that experimental studies are called for, given that they ensure controlled laboratory conditions that can enable us to obtain evidence on the existence of telepathy and eliminate the possibility of alternative explanations of the phenomena. Thus, the scientific method and experimental procedures are reasonable ways to study this research problem so that critical thinking, in one direction or the other, can be more solidly grounded.

All physical devices that allow the transfer of information between an emitter and a receiver can present errors when transmitting the signal. This feature is relevant, because errors can also be produced, analogously, in telepathic transmission, although in this latter case, the device used is not obvious. Therefore, given that the portion of reality to be studied presents uncertainty, probabilistic models seem to provide appropriate representations of this reality, as least as initial approximations.

### Description of the experiment

A variety of experiments can be designed to test the assumption that telepathic transfer exists and here we focus on one of the most common examples. The setting requires two laboratory rooms, with no possibility of communication between them, both being isolated from the outside world. In each of these rooms a participant is seated. One of

these people is referred to as the *emitter* and the other is the *receiver*, given that the former concentrates on an image and the latter tries to perceive the thought of the former. The necessary requisite is that there is no possibility of transferring information physically from one room to the other.

The experimental task comprises a fixed number of trials, that is, an established number of tests in which the emitter concentrates on an image and the receiver tries to catch the signal. The number of trials, denoted by  $n$ , is set to 30. Regarding the stimuli that can be transmitted, there are five possibilities and the receiver knows that in each trial only one of these five stimuli is being sent telepathically. It is common to use the five stimuli proposed by Dr Karl Zener (see Figure 1), as they are easy to distinguish from one another. Oftentimes only three of the five cards are used, as will be done in the present practice, focusing on the circle, the square and the plus sign. Thus, in each trial the experimenter selects at random one of the three *Zener cards*. Once the selected card is known for the specific trial, the emitter is invited to concentrate in order to maintain a clear image of the card in his/her mind. Simultaneously, via an acoustic or light signal, the receiver is made aware that the emitter is currently maintaining the image of the stimulus. Finally, a predetermined amount of time is given so that the receiver can write down the stimulus perceived.

In the present experiment the receiver (i.e., one member of the pair) has 15 seconds to indicate the image perceived, following the rule of choosing one of the three cards in each of the trials in the experiment. The other member of the pair takes the role of emitter, that is, s/he concentrates on thinking of the stimulus indicated for each trial. A third person is necessary for the task of experimenter – this is the person who should fill in the recording sheet and who should indicate to the emitter on which card to concentrate (i.e., circle, square or plus sign). The experimenter ought to have prepared a list with the stimuli, in random order, for each of the trials. This process is repeated 30 times, given that  $n = 30$ .



Figure 1. The *Zener cards*

### **Material necessary for the experiment**

In order to organise the information during the experiment, Table 1 will be used to record the results obtained. Later, the results will be analysed.

### Questions to think about and answer

The questions below should be discussed by each group of emitter, receiver and experimenter, except for points 5, 8, 9 and 10, which will be discussed by all the students together.

1. What is the probability model that represents the random variable “number of correct answers” if we suppose that there is no such thing as telepathic transmission? Specify the main reasons for your answer.
2. What is the value (of this random variable) expected for each participant? Do we observe this value for all participants in the experiment? Why?
3. Is the previously identified model symmetric or skewed? What does it entail?
4. For each pair of students (experimenter and participant):
  - a. Compute the probability of getting exactly the  $k$  correct responses actually obtained, if we assume that there is no telepathic transmission (i.e., using the model identified in point 1).
  - b. Compute the probability of obtaining  $k$  or more correct answers. Why can such a probability be useful?
5. If the number of correct answers were equal to 10, would you consider that there is enough empirical evidence in favour of the existence of telepathic transmission? In either case, what is the basis of such a conclusion?
6. What is the minimal number of correct answers you would consider acceptable as strong empirical evidence in favour of the existence of telepathic transmission?
7. If you have found a “not very likely” result, how many cases having such a result should we observe in order to think that it is not just due to random fluctuations?
8. What conclusion do you reach on the existence of telepathic transmission if you consider all the experiments carried out (i.e., for all pairs of students)?
9. Regarding the empirical distribution of the correct answers, does the average coincide with the mathematical expectancy? Are there differences between the model and the empirical distribution in terms of scatter and shape?
10. Finally, what are the five main things you learned from this practice?

### Complementary reading

Paulos, J. A. (1989). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Farrar, Straus, & Giroux. [Spanish translation, Paulos, J. A. (1990). *El hombre anumérico. El analfabetismo matemático y sus consecuencias*. Barcelona: Tusquets Editores].



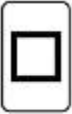




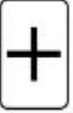
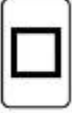









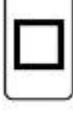
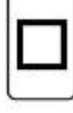


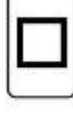


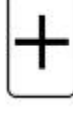

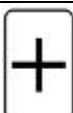
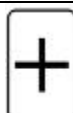

[Back](#)

## EXPERIMENTER'S RECORDING SHEET (Version 1)

Table 1. In this table it is necessary to specify, the name and surname of the participants serving as emitter and receiver. Additionally, it has to be marked in the column called "Result", for each trial, whether the receiver gave the correct answer (using the sign "✓") or not (using the sign "×"). In the column called "Comment" the experimenter would write down the reason for considering a trial as invalid. Finally, the number of correct answers will be counted and written down.

Identity of the emitter:			Identify of the receiver:		
Total number of trials: 30			Number of different stimuli: 3		
Response time available: 15 seconds			Observations: Three Zener cards are being used (circle, square, and plus sign)		
Trial	Result	Comment	Trial	Result	Comment
1			16		
2			17		
3			18		
4			19		
5			20		
6			21		
7			22		
8			23		
9			24		
10			25		
11			26		
12			27		
13			28		
14			29		
15			30		
<b>TOTAL NUMBER OF CORRECT ANSWERS:</b>					

**Stimuli shown to the emitter (Version 1)::** Trials 1 to 30







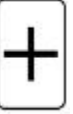




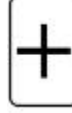


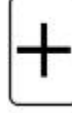















1. 	2. 	3. 
4. 	5. 	6. 
7. 	8. 	9. 
10. 	11. 	12. 
13. 	14. 	15. 
16. 	17. 	18. 
19. 	20. 	21. 
22. 	23. 	24. 
25. 	26. 	27. 
28. 	29. 	30. 

## EXPERIMENTER'S RECORDING SHEET (Version 2)

Table 1. In this table it is necessary to specify, the name and surname of the participants serving as emitter and receiver. Additionally, it has to be marked in the column called "Result", for each trial, whether the receiver gave the correct answer (using the sign "✓") or not (using the sign "×"). In the column called "Comment" the experimenter would write down the reason for considering a trial as invalid. Finally, the number of correct answers will be counted and written down.

Identity of the emitter:			Identify of the receiver:		
Total number of trials: 30			Number of different stimuli: 3		
Response time available: 15 seconds			Observations: Three Zener cards are being used (circle, square, and plus sign)		
Trial	Result	Comment	Trial	Result	Comment
1			16		
2			17		
3			18		
4			19		
5			20		
6			21		
7			22		
8			23		
9			24		
10			25		
11			26		
12			27		
13			28		
14			29		
15			30		
<b>TOTAL NUMBER OF CORRECT ANSWERS:</b>					

## Stimuli shown to the emitter (Version 2): Trials 1 to 30

1. 	2. 	3. 
4. 	5. 	6. 
7. 	8. 	9. 
10. 	11. 	12. 
13. 	14. 	15. 
16. 	17. 	18. 
19. 	20. 	21. 
22. 	23. 	24. 
25. 	26. 	27. 
28. 	29. 	30. 



## PRACTICE ON THE OMNIPRESENCE OF THE NORMAL DISTRIBUTION

### Introduction

Sir Francis Galton (1822-1911), who was a nephew of Charles Darwin (1809-1882), made several important contributions to Statistics. For instance, he proposed the concept of *correlation* and also introduced the notion of *regression toward the mean*. In fact, just like many of his contemporaries, he possessed an extensive training and had knowledge in multiple disciplines, which made it possible for him to contribute to several statistical fields. Among his research topics was genetic inheritance. In this field, Galton thought that inherited characteristics were determined by factors that had a small influence; however, the accumulated or joint action of these factors led to considerable deviations and, thus, to notable differences between individuals. In order to illustrate this conjecture and its conceptual representation, he designed a mechanical device (a physical model), which he called a *quincunx*. It is not precisely known when this device was created, but it is considered to be the year 1873 or 1874 (Stigler, 1986). It is known, however, that the word *quincunx* is derived from Latin and refers to four elements put at the extreme ends of a square or rectangle, with the fifth element being placed at the point at which the two diagonals cross (for instance, the pattern formed by the five points on the fifth side of a die). Its use is related to the organisation of agricultural plantation. Other meanings of the term are related to Roman coins (see Figure 1) and to astrology, where it refers to the disposition of planets.



Figure 1. A Roman coin called a *quincunx*.

What is Galton's quincunx? In Figure 2, you can see the apparatus as it was designed by Galton, where a set of pins are placed in such a way that they form a pyramid, behind the glass. It can also be seen that there is a funnel, through which balls are released and, afterwards, hit the pins and fall into one of the compartments. In order to obtain the result that the *quincunx* is meant to illustrate, it is necessary to release a considerable

number of balls. In Figure 3, you can find a simplified representation of the device, which more clearly shows the fundamental aspects of the physical system.



Figure 2. The original *quincunx* as designed by Sir Francis Galton, although he did not construct it.

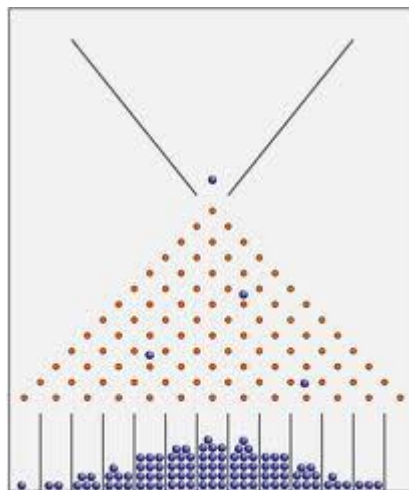


Figure 3. Schematic representation of the *quincunx*.

In order to see what the quincunx does, there are several applications available on the Internet, written in R or as Java applets using animation. The following two URLs are suggested for their simulations of the way in which the *quincunx* works:

<http://yihui.name/en/2008/10/quincunx-in-r/>

<http://www.jcu.edu/math/iseq/quincunx/quincunx.html>

It is also possible to find several YouTube videos showing the physical device and how it works, although they focus on devices that are much more developed than the original. See for instance:

[http://www.youtube.com/watch?v=9xUBhhM4vbM&feature=watch\\_response](http://www.youtube.com/watch?v=9xUBhhM4vbM&feature=watch_response)

How can we statistically represent the events produced by the quincunx? Firstly, each time a ball hits a pin, assuming an identical probability of going to the left or to the right, the event can be described as a *Bernoulli trial* with parameter  $\pi = 0.5$ . Alternatively, this can be represented as a binomial distribution  $B(1; 0.5)$  with a single trial. Secondly, given that each ball has as many impacts as rows of pins in the *quincunx*, it is possible to illustrate the additive property of the binomial distribution. According to this property, for random variables following a binomial distribution with an identical  $\pi$ , the sum of their values follows a binomial distribution with  $B(n; \pi)$ , where  $n$  denotes the sum of trials for all the random variables. In the case of the *quincunx*, the latter means that  $n$  is equal to the number of rows of pins and  $\pi = 0.5$  for each ball that is released. In summary, the repository where the ball ends up represents the number of times that the ball goes to the right (from the observer's perspective), after hitting the pin. Therefore, the random variable following a binomial distribution is the number of times an event occurs, given that there are two possible events (going to the right or to the left). The same reasoning would be valid if the random variable is considered as the number of times that the ball goes to the left. In either case, the random variable can take values ranging from 0 to  $n$ , that is, from a minimum of no displacement to the right to a maximum of all  $n$  displacements in this direction. So what is the relation between this statistical model and genetic inheritance? According to Galton, each random trajectory after hitting the pin represents a factor with a tiny influence and, therefore, the total amount of impacts represents the total deviation observed as a result of the accumulation of several random factors. In this way, the repository represents the deviation of a genetic feature for each person.

So far the explanation has been focused on the probability model describing the passage of a single ball, but what does the fall of several balls represent? One specific ball can show more or less deviation in a genetic characteristic, but what is the distribution for everyone? When several balls are released, the process is actually a *stochastic simulation*, that is, for everyone the aim is to determine the probability of the

deviations of genetic characteristics. The result obtained is, as the device shows, an estimation of the mass probability function of the binomial distribution. It is also an approximation of the *normal law of probability*. Hence, it was through this physical device that evidence was obtained for the conjecture that if there is a set of random factors that have a small influence, the aggregated results can be described by the *normal probability distribution*. This type of scientific law is, thus, probabilistic instead of deterministic and it remains important today as it is common to suppose that, given a set of random components that have a small influence, a normally distributed magnitude will appear. The quincunx is also used to illustrate the *central limit theorem*, but this latter aspect goes beyond the aims of the current practice.

Is the normal distribution so common as to describe, at least approximately, random phenomena? It does not seem so, at least not for all fields of the psychological discipline (Micceri, 1989). In any case, in the present practice we will try to verify that the normal distribution is a reasonable approximation for another kind of variable.

### Description of the practice

The waiting time in an underground station will be measured. This time is defined as the number of seconds that pass from the time a person reaches the platform until the metro in the desired direction comes to a complete stop. In order to obtain this measure, students will start their stopwatch the second they step onto the platform and stop it when the corresponding train stops. They should write down the waiting time in seconds (even if it takes more than one minute for the metro to come). Each student should take 10 such measurements.

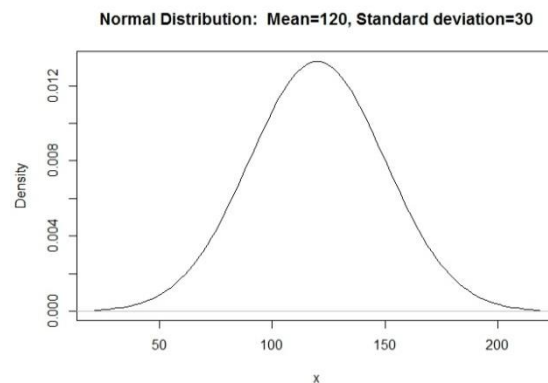
During the practice, all recorded measurements will be put in an R-Commander data matrix, introducing the measurements of 8 students in each matrix. Afterwards, a descriptive analysis will be carried out on the random variable called “waiting time” in order to explore tentatively whether it follows the normal distribution.

### Questions to think about and answer

The following questions should be discussed in each group, unless stated otherwise:

1. If the probability that the ball would go to the right of the pin after hitting it were  $\pi = 0.8$  (and  $\pi = 0.2$  for going to the left), would the approximation to the normal law be reproduced in the *quincunx*? Why?
2. Why can we consider the *quincunx* as an analogy for the deviations of hereditary features?

3. For each student, using his/her 10 measurements of “waiting time”, represent the empirical distribution via a histogram. Does the random variable follow a normal distribution? Why? Repeat the exercise using as a variable the averages for each student and compare the results.
4. Imagine that a reasonable model for representing the waiting time is a normal distribution with a mean of 120 seconds and a standard deviation of 30 seconds:



Answer the following questions, first in an approximate way using knowledge on the properties of the normal distribution and, afterwards, in a more precise way using R-Commander:

- a. What is the probability of having to wait more than 3 minutes?
  - b. What is the probability of having to wait less than half a minute?
  - c. What is the probability of having to wait a period of time somewhere between half a minute and three minutes?
  - d. Obtain the density scores for 2 minutes and for two-and-a-half minutes. How can you interpret these values?
5. For all groups, in what situations can it be of interest to know the waiting time in order to improve the processes?
  6. Finally, what are the five most important things you learned from this practice?

## References

- Micceri, J. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156-166. This article can be found at the following website: <http://www.unt.edu/rss/class/mike/5030/articles/micceri89.pdf>
- Stigler, S. M. (1986). *The history of statistics. The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

[Back](#)

## PRACTICE ON THE APPLICATION OF NORMAL AND $t$ DISTRIBUTIONS TO CASE-CONTROL STUDIES

### **Introduction: case-control studies in Neuropsychology**

Case-control studies consist of comparing the score of an individual with the scores of an appropriate normative group (e.g., considering variables such as age and gender). The normative group is also called a control group, because it contains people who are not diagnosed with any disorder. At the same time, it is common that the person on whom the study is focused is diagnosed with a neuropsychological deficit or is in the process of being assessed or diagnosed. Therefore, this kind of study is intended to verify whether the person has a similar performance to what is observed in people without a disorder or, on the contrary, the person's score in the test(s) applied would not be very likely in people with no diagnosis.

### **Usefulness of probability models**

In order to obtain information on the extent to which an assessed person behaves on a particular test similarly to people with no diagnosis, two probability models can be used: the normal distribution and the  $t$  distribution. In both cases, we look at the statistical distance (in terms of absolute values) between the score of the individual and the average of the control group, considering the amount of variability in the scores of this latter group.

The normal model, in order to be used in this case, requires an assumption that the point estimation obtained in the control sample is equal to the value of the population parameter, in addition to the assumption that the population distribution is normal. If these conditions are met, the score of the individual is transformed into a  $Z$  score with  $Z_{CC} = (X_1 - \bar{X}_2) / s_2$  and we are justified in referring this  $Z$  score to the normal unit distribution. In fact, what is obtained is the probability of obtaining a score as far (or farther) away from the average, using one of the tails of the distribution (usually the tail that implies poorer performance on the test). The use of the normal unit distribution is based not only on the abovementioned assumptions, but also on the idea – as starting point – that the individual belongs to the control group (i.e., people with no diagnosis). The probability value obtained shows the extent to which the result for the individual is rare or infrequent. This quantification can also be conceptualised as an illustration of the fact that the individual does not, in fact, belong to the control population, s/he rather presents a deficit in the cognitive function assessed by the test. Instead of using  $Z$  to

obtain the probability associated with (and the degree of rarity of) the result, Crawford, Garthwaite and Porter (2010) suggest using it as an effect size index.

In contrast with the normal model, the  $t$  distribution can be used in cases in which the size of the control sample is not large, an option that has been considered as more appropriate given that the sample mean and the sample variance are not, by definition, population measures (see Crawford and Garthwaite, 2012, for a review of several alternatives). In that sense, Crawford and Howell (1998) suggest converting the score in the test into a  $t$  score through

$$t = \frac{X_1 - \bar{X}_2}{s_2 \sqrt{\frac{N_2 + 1}{N_2}}},$$

where  $X_1$  is the score of the individual,  $\bar{X}_2$  is the average of the control group,  $s_2$  is the standard deviation of the scores of the control group and  $N_2$  is the number of people in this group. Once the  $X_1$  score is transformed, it is necessary to choose the appropriate  $t$  distribution for quantifying the degree of rarity or infrequency of the score. In contrast with the unique  $Z$  distribution, there are several  $t$  distributions according to the degrees of freedom (for a didactic explanation of this concept, consult Eisenhauer, 2008). For case-control designs and for the  $t$  score presented above, Crawford and Howell (1998) claim that the degrees of freedom are equal to  $N_2 - 1$ . As can be seen, therefore, the sample size of the control group is taken into account in this case. Once the adequate  $t$  distribution is selected, it is necessary to look for the probability of the  $t$  score as far away (or farther) from 0, usually in the direction (distribution tail) indicating a deficient performance in the test.

Note that the use of the  $t$  distribution for obtaining this probability also assumes that the test score in the population is distributed normally. Once again, as is the case for  $Z$ , the underlying idea is that the individual is part of the population with no disorder diagnosed and the quantification provided is the probability of a score as extreme or even more extreme than the one actually obtained. This probability can be used as evidence that the person might show a deficient performance in comparison to what is expected of a person without a disorder (i.e., a person from the control population).

### **Quantification of the degree of rarity without a mathematical-statistical model**

The fact that the distribution of the  $t$  score or the  $Z$  score is known, under the assumptions mentioned above, enables us to quantify the probability of obtaining a control individual (i.e., with no disorder) with a score as far away from (usually below) the average. In terms of frequency, this probability would indicate the proportion of



people in the control population with a score as low as or lower than the one obtained by the person to whom the test is applied. This probability is known as the  $p$  value and it is frequently used in making statistical decisions (i.e., to reject or not the null hypothesis that the individual comes from the control population). The  $p$  value will be a topic extensively treated in the “Estadística” course.

In addition, it should be noted that when a representative sample is available from the reference population, it is also possible to obtain a quantification of the degree of rarity of the result of the individual, without the use of a mathematical-statistical model (and without having to make the corresponding assumptions). For example, in personality inventories, the measures of groups of individuals organised according to relevant demographic variables, such as age, gender, or education level. Once the score of the individual is obtained, it is compared to the reference group (i.e., to those people with whom the individual shares the relevant demographic features). Given that the empirical distribution of the test score obtained by the reference group is sorted, it is possible to know if the individual has obtained one of the highest or lowest scores in comparison to this group. For instance, if a 16-year-old boy obtains 39 points in Openness in the NEO-FFI inventory (Costa & McCrae, 2002), then it can be shown that this score corresponds to percentile 97. That is, 97% of the people in the reference group of male adolescents obtain a score which is lower or equal (they are not as open to experiences as the person studied). Additionally, if the same boy scores 13 in Conscientiousness, this result would correspond to percentile 2 and so 98% of the people in the reference group would obtain a higher score. Therefore, the results closer to the extreme percentiles would be less frequent in the reference group.

### **Quantification of a possible specific deficit**

In case-control studies there is another kind of information that can be obtained and is relevant at a neuropsychological level – this is the possibility of comparing the scores of a person in two neuropsychological tests with the aim of exploring whether there is a specific deficit in the area assessed by the corresponding test. To this end, it is possible to obtain an indicator of the effect size regarding the difference between the score on the two tests and construct a confidence interval around this effect size. For more details the interested reader should consult Crawford et al. (2010); here we present the formula for the effect size of the difference between tasks  $X$  and  $Y$  in order to make it clear that both scores are standardised before being compared (in order to express them in the same metric) and that the correlation between the scores of the control group ( $r_{XY}$ ) is also considered, as is usual in repeated measures analyses (given that it is included in the standard error of the difference between two random variables):



$$Z_{DCC} = \frac{\left[ \frac{(X - \bar{X})}{s_x} \right] - \left[ \frac{(Y - \bar{Y})}{s_y} \right]}{\sqrt{2 - 2r_{XY}}} = \frac{Z_X - Z_Y}{\sqrt{2 - 2r_{XY}}}.$$

In this case, the normal unit distribution is not used for the purpose of obtaining probabilities, because this is not justified; rather the  $Z_{DCC}$  quantification thus obtained is only used for exploring, at a descriptive level, the extent to which this value differs from 0. The father away it is, the more evidence it represents for the existence of a specific deficit in a cognitive function.

### Instrument

To illustrate the use of case-control studies in neuropsychology, an instrument called *Screen for Cognitive Impairment in Psychiatry* (SCIP) will be applied. This instrument is used to evaluate cognitive deficit in patients diagnosed with schizophrenia and bipolar disorder. The instrument has shown adequate psychometric properties both for people with schizophrenia (Pino et al., 2006) and with bipolar disorder (Guilera et al., 2009). Note that, despite the fact that the test has been shown to distinguish acceptably well between people with and without impairment (Rojo et al., 2010), the SCIP should not replace a full neuropsychological assessment. Apart from the version in English (Purdon, 2005), there is one in Spanish (SCIP-S; Pino et al., 2006). Moreover, it has to be pointed out that the test has three alternative versions in order to allow for more than one administration of the test without the results being distorted by the person's recall of the previous testing.

In the present practice, SCIP will be applied for several reasons. Firstly, it is a useful test in the neuropsychological context. Secondly, it requires relatively short training (a video will be shown to fulfil this purpose). Thirdly, the test administration time is brief, requiring 3 minutes for each of the 5 sub-tests (i.e., a total of 15 minutes per person). In the practice, we will focus on two of the sub-tests: 1) *Verbal Learning Test with Immediate Recall* (VLT-I) is a verbal memory test with 3 trials of the same 10 words that the participant listens to and has to recall; the result is the number of words correctly recalled; 2) *Verbal Fluency Test* (VFT) comprises two trials of 30 seconds in which the participant has to generate words starting with the letters of the alphabet, avoiding numbers, people's names and words with the same stem; the result is the number of words correctly generated.

## Situation 1

In order to carry out the comparisons, we will use the information on the control group presented by [Guilera et al. \(2009\)](#). This control group is formed by 45 people with an average score of 21.84 (standard deviation of 3.74) in VLT-I and an average score of 19.73 (standard deviation of 5.71) in VFT. The correlation between these scores is .141 (Georgina Guilera, personal communication, 18 April 2013). In this first situation, we will study the case of a fictitious person to whom form 3 of SCIP has been applied and this person recalled 17 words overall (VLT-I = 17 points) and evoked 2 words with the letter F and 4 words with the letter R (VFT = 6).

### Situation 1: Tasks with the available data

For each of the two sub-tests, VLT-I and VFT:

- 1.1 Obtain the degree of infrequency of the result, according to the  $Z$  distribution (using R-Commander).
- 1.2 Obtain the degree of infrequency of the result, according to the  $t$  distribution (using Singlims\_ES or R-Commander).

Singlims\_ES.exe: Significance test and point & interval estimates of effect size and abnormality for a patient's score

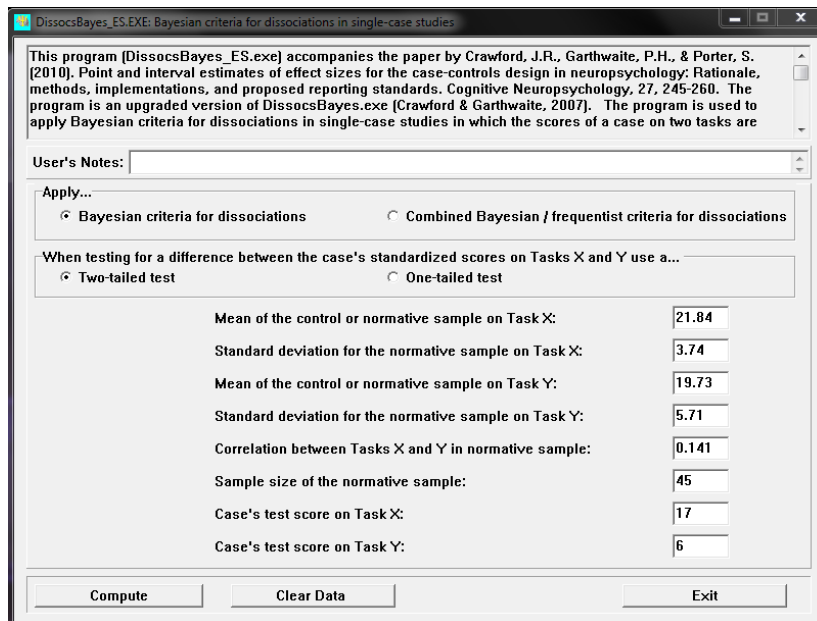
This program (Singlims\_ES.exe) accompanies the paper by Crawford, J.R., Garthwaite, P.H., & Porter, S. [2010]. Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27, 245-260. The program is an upgraded version of Singlims.exe [Crawford & Garthwaite, 2002]. The program tests whether an individual's score is significantly different from a control or normative sample. Unlike Singlims.exe, it also provides a point estimate

User's Notes:

Mean of the control or normative sample:	21.84
Standard deviation for the normative sample:	3.74
Sample size of the normative sample:	45
Case's test score:	17

Compute      Clear Data      Exit

1.3 Obtain the value of the effect size for the difference obtained in the two tests (Dissocs\_ES).



1.4 Comment on the results.

**Situation 2: Available data and further data to be collected**

For this second situation, we will collect data from a more extensive normative group and focus on more specific information, taking into account education level (people with higher education) and age (Georgina Guilera, personal communication, 18 April 2013).

*Verbal Learning Test with Immediate Recall*

Secondary education or higher		
18-39	40-55	>55
N= 177	N= 126	N= 65
Mean=24.67	Mean=22.39	Mean=20.95
SD = 3.23	SD = 3.06	SD = 3.71

*Verbal Fluency Test*

Secondary education or higher		
18-39	40-55	>55
N= 177	N= 126	N= 65
Mean=19.58	Mean=19.88	Mean=17.98
SD = 5.15	SD = 5.47	SD = 6.35

**Situation 2: Tasks with the data collected by the students:**

In order to obtain the score of the individual participant, the task proposed is to collect data on the VLT-I and VFT subtests, working in pairs: researcher-subject.

2.1 Apply form 1 of SCIP to a classmate and obtain the scores for VLT-I and VFT.

For each of the two sub-tests, VLT-I and VFT:

2.2 Obtain the degree of infrequency of the result, according to the  $t$  distribution (using Singlims\_ES or R-Commander).

2.3 Comment on the results.

2.4 Repeat the same process participating as a “subject” instead of “researcher” applying form 2 of SCIP.

**Questions to think about and answer:**

1. Are the scores of the students in Situation 2 below or above the average of the normative group? What could be the reason for such a result?
2. Are the scores of the fictitious participant (Situation 1) and of the students (Situation 2) equally unlikely or unexpected? What could be the reason for such a difference?

**Material necessary for the practice session:**

- To perform the calculations necessary for this practice (and others not mentioned here), the executable files Singlims\_ES y Dissocs\_ES explained in Crawford et al. (2010) can be used. These files can be downloaded from the following web site:
- <http://homepages.abdn.ac.uk/j.crawford/pages/dept/psychom.htm>
- Video recording used for training in the administration of SCIP-S (duration: 15 minutes).
- Copies of SCIP, forms 1 and 2 (see “User rights” below).

## User rights

The rights to use the SCIP-S instrument have been transferred to *TEA Ediciones* by the authors of SCIP and SCIP-S. The SCIP-S and SCIP tests administered in the practice are only used for teaching purposes. Using the instrument outside of the classroom is prohibited; it is also prohibited to make, distribute or sell copies.

## References

- Costa, P. T., Jr., & McCrae, R. R. (2002). *Manual NEO PI-R*. Madrid: TEA Ediciones.
- Crawford, J. R., & Garthwaite, P. H. (2012). Single-case research in neuropsychology: A comparison of five forms of t-test for comparing a case to controls. *Cortex*, *48*, 1009-1016.
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*, 245-260.
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*, 482-486.
- Eisenhour, J. G. (2008). Degrees of freedom. *Teaching Statistics*, *30*, 75-78.
- Guilera, G., Pino, O., Gómez-Benito, J., Rojo, J. E., Vieta, E., Tabarés-Seisdedos, R., et al. (2009). Clinical usefulness of the screen for cognitive impairment in psychiatry (SCIP-S) scale in patients with type I bipolar disorder. *Health and Quality of Life Outcomes*, *7* (28). Available online at <http://www.hqlo.com/content/7/1/28>
- Pino, O., Guilera, G., Gómez, J., Rojo, J. E., Vallejo, J., & Purdon, S. E. (2006). Escala breve para evaluar el deterioro cognitivo en pacientes psiquiátricos. *Psicothema*, *18*, 447-452.
- Purdon, S. E. (2005). *The Screen for Cognitive Impairment in Psychiatry (SCIP): Instructions and three alternate forms*. Edmonton, AB: PNL Inc.
- Rojo, E., Pino, O., Guilera, G., Gómez-Benito, J., Purdon S. E., Crespo-Facorro. B., et al. (2010). Neurocognitive diagnosis and cut-off scores of the Screen for Cognitive Impairment in Psychiatry (SCIP-S). *Schizophrenia Research*, *116*, 243-251.

## Section 4:

### Descriptive statistics

Descriptive statistics include numerical and graphical summaries of the data, with the aim of organizing them and converting them to understandable and useful information. In order to meet this aim, descriptive statistics should be used according to: a) their purpose, that is, the information they provide – central tendency, position, scatter, diversity, skewness, kurtosis; b) the measurement scale of the variables analysed; c) the limitations of the indices (e.g., when data are tied, in presence of outliers, when nominal scale variables have more than two categories).

We have attempted to illustrate this way of using the indices and plots with the practices included in this section. We have also tried to make the analyses more attractive by applying them to data gathered by the students themselves, so that they are aware of the research aim and process.

We decided to include in this section a practice on content analysis, which is typical for qualitative methodology and not usually part of the texts on statistics (focussing the quantitative methodology). The reason for this inclusion is that content analysis provides a description of the (qualitative) data and not an inference to a population: thus it is not suitable for section 5. Moreover, content analysis is initiated in qualitative terms, but usually incorporates quantitative elements such as absolute and relative frequencies or frequency of transition and, of course, frequencies are part of descriptive statistics applied to nominal scale data.

One final remark refers to the distinction between univariate and bivariate analysis. The former is carried out when we are interested in studying one variable at a time (e.g., to describe the degree of satisfaction of the students with a statistical course), whereas the latter is called for when exploring the presence of relationship or association between two variables (e.g., to answer the question whether greater anxiety is related to better or worse performance in a cognitive task).

## PRACTICE ON CONTENT ANALYSIS

### Introduction

Qualitative methodology, which deals with the essence of the object of study, can be carried out from several different perspectives. Not only are objective events important, but also how these events are explained verbally and how they are interpreted by the people who are part of the context. The emphasis is usually put on description, but theories are also gradually formed and fine-tuned through the constant interrelation between data and concepts. Data collection is usually based on participant observation, which provides in-depth knowledge, but also involves greater subjectivity on the part of the researcher. The analysis and interpretation of the information depends on the qualitative approach being followed. In many cases, however, it is possible to give information back to participants, so that they can validate the conclusions drawn by researchers, who can then reconsider whether these conclusions need to be modified. In the present practice, content analysis will only be used to summarise and structure the experience as explained by the person who lived it. Additionally, a brief summary of the story told by the participant is sought, but not with any intention of constructing theories.

### Aim of the practice

The purpose of the present practice is to carry out a first approximation to content analysis, focusing on a real text related to life experiences that can be studied from the perspective of Psychology. The research in which the text is located follows a qualitative methodology and it starts from a question put to Teresa regarding a very unfortunate event in her life. Afterwards, interviews are performed (*triangulation*) in order to complete and clarify the information provided by Teresa (*emic approach*). The next steps of the research depend on the specific approach followed by the researchers. Both the text written initially and the answers to the follow-up questions can be subjected to content analysis; however, we will only focus on the text.

For the current practice, the first steps in the content analysis are:

1. Basic information about the text: research on trauma and resilience using descriptions given by the people who have suffered the trauma. The text is produced by a 26 year-old woman referred to as “Teresa” and the information is collected by a classmate who also interviews her afterwards.

2. The text is a description of a very unfortunate event (with no further details), written by Teresa herself, on demand (i.e., not spontaneously).
3. Sampling: a) intertextual: it is not possible, there is only one text; b) intratextual: we have selected the part of the text after Teresa receives the news that she has cancer, until the end of her self-produced story.
4. Transcription: not necessary, it is a written (not oral) production.
5. Textual unit: the suggestion is to use one or several phrases with the same meaning regarding the events, thoughts and feelings described by Teresa.
6. Contextual unit: all the text; in the etic categorisation the moment in time described by the fragment will be taken into account (i.e., how near it is in time to the moment at which the news is received).

Afterwards, the codification will take place using concrete labels at first and more abstract labels in the end. Finally, the most basic type of quantification is suggested as a support for the interpretation of the information provided by Teresa.

### **Description of the tasks**

1. Read the text in order to know the general content of the event described by Teresa and the context in which it took place.
2. Use as a textual unit the portion of contiguous (adjacent) text that expresses the same idea (e.g., a phrase that is part of a sentence, a whole sentence or several sentences). Perform the segmentation of the text.
3. Assign first-order labels with the aim of summarising or synthesising the text. These labels have to be emic, close to Teresa's own language.
4. Group the first-order labels in more abstract concepts, according to the meaning of the fragments of the text. In this case, you should follow an etic perspective that takes into account the following two aspects (or coding schemes or category systems):
  - a. who is the agent, the active person – Teresa or the people from her environment, and people being passive can also be coded (including Teresa);
  - b. what is Teresa's reaction to the events in her life – does she assess them in a positive way (gain) or in a negative way (loss).

Criterion a) is extracted from the discourse analysis carried out by [McMullen \(2011\)](#), whereas criterion b) is based on the analysis by [Charmaz \(2011\)](#) from the Grounded Theory perspective. It should be kept in mind that other coding schemes and other categories could be established according to the theoretical perspective from which the text is approached when analysed.



5. Obtain the frequency of occurrence of each category and carry out the corresponding comparisons between the categories that can logically be compared.
6. Interpret the results of the categorisation, considering the moment in the discourse at which the different categories appear.

### Material necessary for the practice

- The text written by Teresa: see the Appendix (extracted from [Wertz et al., 2011](#), pp. 104-107).

### Questions to think about and answer

1. If a word has been chosen as the textual unit, what will the contextual unit be?
2. In what situations is it expected that two researchers obtain similar results and in what situations may the content analysis performed diverge to a greater extent? In other words, what element is necessary to unify the analyses?
3. In the current practice, we propose comparing the frequencies of occurrence of two categories of the same coding scheme; what other measure could be used with the same purpose?

### Complementary readings

- Riba Campos, C.-E. (2007). *La metodologia qualitativa en l'estudi del comportament*. Barcelona: Editorial UOC.
- Charmaz, K. (2011). A constructivist Grounded Theory analysis of losing and regaining a valued self. In F. J. Wertz, K. Charmaz, L. M. McMullen, R. Josselson, R. Anderson, & E. McSpadden (Eds.), *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry* (pp. 165-204). London: Guilford Press.
- McMullen, L. M. (2011). A discursive analysis of Teresa's protocol. In F. J. Wertz, K. Charmaz, L. M. McMullen, R. Josselson, R. Anderson, & E. McSpadden (Eds.), *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry* (pp. 205-223). London: Guilford Press.
- Wertz, F. J., Charmaz, K., McMullen, L. M., Josselson, R., Anderson, R. & McSpadden, E. (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry*. London: Guilford Press.

[Back](#)

**Appendix:**

**Instructions:** *Describe in writing a situation when something very unfortunate happened to you. Share I detail what happened, what you felt and did, and what happened after that, including how you responded and what came of this even in your life.*

[Teresa has just received the news that she has cancer, which needs to be removed.]

Then there was silence, and he just sat there, staring at me, waiting for who knows what. I sat back in my chair a little, let out a big breath, and stared back at him. “Okay,” I said. “What’re we going to do about it?” In an instant, he was fumbling around on his desk, grabbing a pen and notepad, clearing a space in the midst of the odd clutter. He drew a picture of what seemed to be the two lobes of the thyroid gland (which looked rather like a weird kind of bow tie), then drew out the incisions and various possible mishaps that could occur. I took it all in very methodically, as though we were talking about someone else entirely that he’d be cutting into the next morning. After that was done, he leaned back in his chair and asked me if I had any questions, and I didn’t. Then, perhaps in an effort to make small talk, he asked, “So you’re a college student... what’s your major?” I told him it was vocal performance, and his face went white. He looked grimmer now than he had at any point in our conversation. “Look”, he said very gently, “because of where this thing is and what we’re going to have to do, there’s a chance you won’t be able to even speak the same way again. You may not be able to be singing anymore after this.”

I froze, I couldn’t breathe, couldn’t move, couldn’t even blink. I felt like I had just been shot. My gut had locked up like I’d been punched in it. My mouth went dry and my fingers, which had been fumbling with a pen, were suddenly cold and numb. Apparently picking up on my shock, the surgeon smiled a little. “We’re going to save your life, though. That’s what counts. And you know what? The other surgeon working with me is a voice guy. We’re going to do everything we can not to be too intrusive.” I started to breathe a little, very little, and I felt myself trembling. I tried to say something meaningful, expressive... all that I could manage was: “Man... I was actually pretty good”.

Then, all of me let loose. I was sobbing, but there was no sound; just a torrent of tears, and the hiss of crying from my open mouth, pushing through the pressure of the accursed mass. The surgeon hastened to my side, armed with a tissue and a firm, reassuring hand on my shoulder. I heard him speaking softly from beside me as I heaved in my silent wailing. “You’re going to beat this. You’re young and you’re going to beat this thing. And you’ll get your voice back, and you’ll be singing at the Met. And I want tickets so don’t forget me”. Slowly, I came back to myself, began to breathe again, and listened to the surgeon as he told me that he was going to use the smallest breathing tubes possible, even make the cleanest possible work of incision. By the end of the visit, I was completely drained, like a ghost of my former self. I felt as though the

biggest and best part of me had died in that office. Cancer wasn't as frightening to me as never being able to sing again. Singing had been my life for as long as I could remember; the one thing I could excel at, the only thing I knew. It had been my solace in all my times of distress, through every hardship ... this would be the most grueling hardship of all, and I wouldn't be able to sing my way out of it. Literally. Worst of all, I still had to tell my mother.

The meeting in the surgeon's office is what, for me, qualifies as my most unfortunate event to date. The next day, I went into surgery, and it went very well. It took a bit longer than expected, since the mass, a large exponentially growing tumor, had already begun spreading to my lymphnodic tract and the muscle tissue on the left side of my neck. When I woke up from surgery, I no longer had any thyroid at all, and had also lost some muscle tissue in my neck and two parathyroids. My voice was indeed changed, and it was very hard to speak for a few weeks. Later, my speaking voice returned, but my singing voice wasn't as quick to reemerge; I was left with no choice but to leave the music school at my university and give up all of my singing projects. I had been a cantor at three area churches, and found that I could no longer bring myself to go to church at all... it was too painful to go if I couldn't sing. All my friends had been fellow singers, and I knew that they couldn't bear the discomfort of being around me under circumstances. My voice teacher, who was like another father to me, greeted me in tears each time he saw me afterwards... he was there for my surgery, and was the last persona I saw before my anesthesia kicked in. Seeing the dreams we had built together go to pieces the way they did was just too much for either of us, and we spoke very little after that.

Many suggested that I take a break from school, that no one would think any less of me, but I was determined to move on as if nothing had happened. When I met new people, I no longer introduced myself as a singer, which was strange for me. Now, I was a psychology major, and I told people this as though I had always been. I suddenly had nonmusical friends, which was also odd, yet strangely refreshing. I was having conversations that I never had the opportunity for in my previous life; my friends were now philosophers, scientists, poets, and historians, and I was learning of a life beyond the hallow catacombs of practice rooms, voice studios, and recital halls. On top of that, I took fencing, motorcycling, rock climbing, and theater acting, and seemed to do pretty well. Frankly, I just wanted to live as much as I possibly could, and do everything imaginable while I was at it. Meanwhile, I had also become acquainted with the intricacies of cancer treatment, undergoing a series of radiation bouts and long days alone in clean isolation rooms so I wouldn't contaminate anyone while eradiating. Just when I could fool myself into thinking I was normal again, I'd be back in the hospital.

It took an extra year to get through my undergrad work due to the change of major, during which I met and married my very nonmusical, very academically inclined husband. I began contemplating what to do with my bachelor's degree in psychology when, three years after my surgery, my singing voice began to come slowly toward getting my voice back in shape, and eventually maintained my own voice studio of around sixty students, serving as my own poster child for the miracles of good voice technique. I sang with two opera choruses, got back into singing at weddings and church services a bit, even visited my old voice teacher a few times for a few lessons. Still I loved my newfound intellectual life, and I didn't want to give it all up and go back to the grind of full-time classical singing. Besides, I had discovered that, while my voice was still misbehaving (and often does, to this day), I could sing other kinds of music pretty well,

particularly rock and blues. I began tinkering with writing my own music, and eventually acquired my own regular gigs at night clubs and live music venues. I continued in my psychology work, as I do now, for I love it dearly, particularly in that it brought forth in me a part of myself I never knew I had, one that seems to hold its own well enough with the more intellectual crowd. The intensive opera chorus work still makes me an opera singer, but that doesn't seem so important to me anymore. I can sing my own music now, so I'm a singer in an entirely new way. I've officially been in remission for over a year now, and, since my type of cancer is an angry sort, I have to go in for scans twice a year. As I see it, though, if I could get through that day in the office with that surgeon (who, by the way, I fully intend to invite to my first breakthrough gig, whatever style of music I'm singing at the time), I suppose I can get through just about anything.

## PRACTICE ON UNIVARIATE DESCRIPTION

### Practice aim

The purpose of this practice is to work with statistical indicators that can be applied to describe variables on which the students have collected information using the different recording techniques. In this practice, emphasis will be put on selecting a graphical representation and a numerical indicator in light of the measurement scale of the variable(s). The same idea will receive further attention when discussing the inferential use of the indices in the “Estadística” course. Another feature of the variables that should be taken into account is their distribution, which makes some indicators more appropriate than others in certain situations (e.g., considering the presence of outliers or asymmetry). Therefore, the second basic idea of the practice is that there are no universally recommendable indicators. Rather, the aim of the statistical analysis and the features of the data have to be taken in consideration. As a third aim, we intend to illustrate that quantification is especially relevant in some, but not all, situations (e.g., when organising and systematising part of the information gathered through an interview, it is not always necessary to quantify). Finally, in some cases, when the amount of data is not large, students are advised to carry out the calculations by hand in order to better understand the operations performed on the data and, thus, the type of information that the indices actually provide. In other cases, the practice will focus on obtaining the numerical indicators and graphical representations via R-Commander, an open-source statistical package that can be useful for students in all their course projects in the degree of Psychology, as well as in their professional tasks outside of the university.

### Description of the tasks

#### Using your data collected via the interview:

##### [Excel file in the Virtual Campus]:

1. Using the information collected via the interview, describe the main features of the sample, specifically, age and gender. For each of these variables, use only the indicators appropriate for the measurement scale and the ones that provide relevant information on the sample. Obtain these quantifications by hand (i.e., using a calculator), without using R-Commander.
2. Provide a graphical representation that shows the main reasons given by participants as motivations that could encourage them to stop smoking. Compute by hand the *odds* for the category “health” (or equivalent) and interpret it.

3. Describe the variable “number of cigarettes” smoked each day, computing at least one indicator of central tendency, one of scatter, and one of shape, in order to obtain information on whether the amount is large or small. Obtain these indicators using the appropriate formulae without the aid of the statistical software.
4. Describe the relative severity of smoking as perceived by the participants, computing at least one indicator of central tendency, one of scatter, and one of shape, in order to obtain information on whether it is considered the most dangerous habit. Obtain these indicators using the appropriate formulae without the aid of the statistical software.

### **Using your data collected via the experiment**

Each student should input his/her data as experimenter or participant following the example of the data matrix shown in the Appendix.

5. Provide an informative graphical representation useful for comparing the response time of the participant for the concordant and discordant conditions.
6. Obtain the main indicators based on position for the distribution of “response time”; also provide the indicators of shape. Comment on the degree of similarity between the concordant and discordant conditions. How do these indicators relate to the graphical representations from the previous question?
7. Obtain a central tendency indicator and a scatter indicator to enable comparisons of the “response time” according to whether the result is for a trial with a correct or an erroneous answer.
8. What is the most frequent value for the variable called “result”? What is the most frequent value for the variable called “condition”?

### **Using everyone’s data gathered via the observation [Excel file in the Virtual Campus]:**

9. Represent graphically and offer at least two adequate quantifications to describe the assessment made by the whole of the group regarding the social skills of the character from the second video shown in the observation practice.

---

**Using everyone's data collected via the experiment [Excel file in the Virtual Campus]:**

10. Represent graphically the response time for correct and erroneous answers. Assess whether the distribution is approximately Normal. Try to explain the results.
11. Focusing on the correct answers for the concordant condition, both numerically and graphically compare the response time according to whether this condition was presented first or second in the experimental procedure.

**Material necessary for the practice**

- The rating of the social skills of the character from the observation practice: the data should be previously sent by the students and available in the Virtual Campus.
- Data matrix from the practice on the Stroop effect – summary information for the whole group: the data should be previously sent by the students and available in the Virtual Campus.
- Data matrix from the interview practice: the data should be previously sent by the students and available in the Virtual Campus.
- Data matrix from the practice on the Stroop effect – the individual sheet that each experimenter should bring to the current practice.

[Back](#)

## PRACTICE ON DIVERSITY

### Introduction

The word *diversity* is widely used in a colloquial sense and it, therefore, has connotations that can entail a wide variety of meanings. In the present text, however, we will focus on diversity in a strict sense, as is necessary in the case of scientific definitions. Moreover, it should be highlighted that a statistical indicator or coefficient does not *per se* help in the theoretical understanding of the underlying idea. In other words, the conceptual content is determined by the definition and not by the measurement obtained. For instance, standard deviation is a well-known statistical indicator of scatter (or, if preferred, variability or heterogeneity). The numerical value obtained when computing the standard deviation for a data set does not have any conceptual meaning, although it is a quantification of the degree of dissimilarity of the measures available. As a matter of fact, in some studies, the standard deviation is used as an indicator of the degree of consensus or agreement among group members. Nonetheless, it should be stressed once again that the meaning of an index, e.g., the degree of agreement among people, depends on the conceptual definition and not on the index itself. Moreover, it has to be noted that any calculation, as a practice, is not in and of itself especially relevant: attention should rather be paid to the representational content of the index and, therefore, to the conceptual meaning derived from the measurement.

Constructs are abstractions, that is, theoretical entities that cannot be observed directly and, thus, their measurement is indirect. Nevertheless, constructs are necessary for organising the knowledge in a certain field and for making possible a theoretical understanding of the object of study. For instance, personality traits are constructs. Indeed, they are theoretical elements that enable the description or explanation of the relations among psychological processes taking place in groups. The similarity-attraction model (Byrne, 1971), for example, states that groups that are naturally formed will most likely be composed of people with similar characteristics. Note that similarity is not measured directly. Rather, it is a construct that makes it possible to describe or explain the relation existing, according to the model, with another construct, namely attraction. It should be added that this model allows predictions of whether the similarity between group members is also related to their effectiveness in performing a group task.

A construct should be defined conceptually, that is, it is necessary to create a constitutive definition by which the concept is defined using other words (and concepts). This definition has to be invariant, i.e., it should remain the same regardless of how the construct is finally measured. Otherwise, the comparability of scientific findings is not guaranteed. This point is relevant when talking about diversity, given that (as it will be



shown) it is important to distinguish three types of diversity. Thus, as diversity is composed of different variants, it cannot be understood as a single unitary construct.

Before defining diversity, it is necessary to consider two aspects. Firstly, this concept does not originate from Psychology. Secondly, it is not always adequate to use a definition created in and designed for a different scientific field without adjusting it to the new field in which it is intended to be applied. Regarding these considerations, the origin of diversity lies in Ecology and it has since been applied in demographic studies. It has to be taken into account that, in some cases, diversity was used as an indicator of the appropriateness of a system, for instance, in Ecology. With this in mind, there are several different definitions of diversity as a construct in the Social Sciences. Which is the most adequate definition? This question does not have an obvious answer. The most convenient definition of diversity, as is usually the case in scientific disciplines, is whichever one is most accepted by researchers. In that sense, the definition must serve to distinguish this concept from other constructs; it should be useful in carrying out scientific studies, and it should lead to interesting results for describing, explaining and predicting the processes of interest. In summary, the most successful concept is the one that is most beneficial in terms of scientific progress.

The need for a constitutive definition is, as indicated, fundamental for guaranteeing the comparability of studies on diversity. The definition of diversity used here is as follows: the total amount of difference between the members of a social unit. Note that this definition expresses the construct as a one-pole dimension, that is, ranging from the inexistence of diversity to its maximal value. However, the abovementioned definition does not distinguish between the types of diversity as a construct. It should be highlighted that the maximal diversity can be obtained if the proportions of the different characteristics are identical (e.g., in a group formed by ten people half are men and half are women), if the opinions of a group are equally distributed at the two extreme scores (e.g., half of the people are totally in favour of carrying out an action and half are totally against it), or if there is an absolutely unequal distribution of a resource, knowledge or status (i.e., in a work group there is only one person assigning tasks and the remaining people simply execute them). Therefore, there are different ways of understanding diversity as a construct and they should not be mixed in empirical studies; different definitions should be used to match the theoretical representation with the empirical measurement.

It has been mentioned that for a construct, it is not possible to obtain a direct measurement in the real world. Therefore, an operational definition of the construct is called for, so that an empirical (albeit indirect) measurement can be obtained. It is crucial to underline that studying the relations between constructs is only possible when there are operational definitions (if the aim is to contrast the scientific statements objectively). Defining diversity operationally is a matter of discussing the numerical indices proposed for its quantification. What properties should the indices possess in

order to offer this quantification of diversity? First, in order to represent diversity, the index should take the value of zero if there is homogeneity, i.e., if there is no dissimilarity in the measurements of the characteristic of interest. That is, given that diversity is a one-pole construct, the index should have a lower bound of zero as a minimal value. Second, higher index values should correspond to increases in diversity. Third, the diversity index should have an upper bound.

It is necessary to distinguish three main types of constructs regarding diversity: variety, separation and disparity (Harrison & Klein, 2007). Each of these different constructs requires its own specific operational definition. When diversity is understood as variety, the members of a social unit differ regarding an attribute (i.e., a qualitative characteristic), such as gender, nationality or ethnicity. In this case, the magnitude of dissimilarity between the members of the social unit is a function of the discrepancy when compared to the uniform distribution of the set of categories (see Figure 1). There are several operational definitions of variety, the most common being *Blau's* and *Teachman's* indices. Regarding diversity as separation, the differences with the social unit can be represented as a continuum (e.g., attitudes, opinions, beliefs, personality traits). In this case, dissimilarity increases as the separation between the locations of the individuals on the continuum increases (see Figure 2). Frequently used operationalisations of separation are the *standard deviation* and the *mean Euclidean distance*. Finally, regarding diversity as disparity, the differences between the individuals are expressed in terms of the degree to which they possess a certain resource or socially valued asset (e.g., income, social influence, skill, prestige). The maximal diversity as disparity entails an asymmetry in the distribution of the social units (see Figure 3). Common operational definitions of this construct are the *coefficient of variation* and *Gini's concentration coefficient*. As far as the measurement scale is concerned, it should be nominal, interval, and ratio, respectively, for the indices of variety, separation, and disparity. It is not uncommon for the latter two types of diversity to use ordinal scale variables. In order to carry out the corresponding calculations, Table 1 includes the indices mentioned above, representing the different operational definitions. For each variant of the concept “diversity”, the indices are related, despite the fact that they are expressed in different metrics. In this sense, if *Blau's* and *Teachman's* indices are computed for the same characteristic of interest in several groups, the result would be that higher values for one index are related to higher values for the other. This is not surprising, given that both indices were designed to quantify the same property.

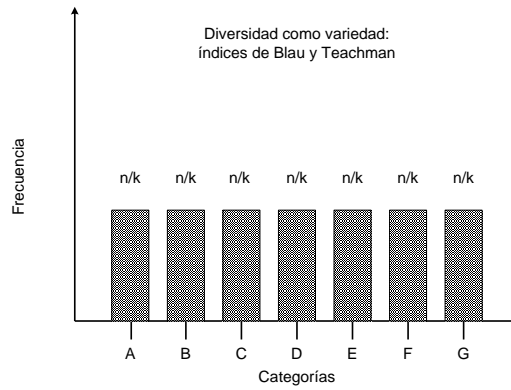


Figure 1. Uniform distribution of the  $n$  members of a group into  $k$  categories. Note that, to achieve the maximal degree of diversity, it is necessary that  $n = mk$ , where  $m$  is a positive integer (abscissa: categories; ordinate: frequency).

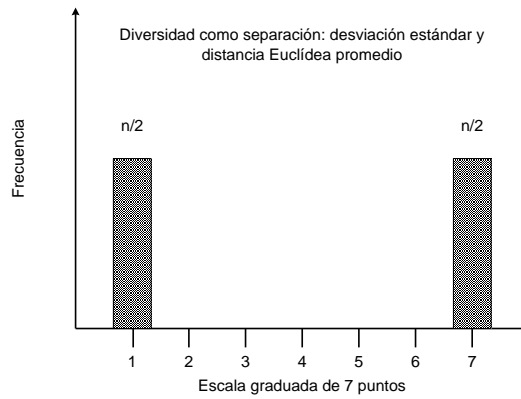


Figure 2. This graphical representation shows that the maximal level of diversity as separation will be obtained if the size of the social unit,  $n$ , is even (abscissa: a 7-point scale; ordinate: frequency).

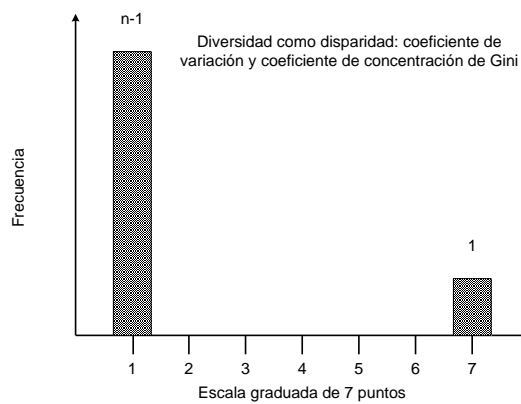


Figure 3. Distribution of scores corresponding to the maximal level of disparity (abscissa: a 7-point scale; ordinate: frequency).

Note that, for all indices, if the result is divided by the maximal value, a normalised index (not to be confused with a standardised one) will be obtained, ranging from 0 to 1, representing minimal and maximal diversity, respectively.

Table 1. For each of the three types of diversity, we provide the two most common indices and their upper and lower bounds. Using  $k$ ,  $p_i$  and  $\ln()$ , respectively, we have denoted the number of categories, the proportion for each one of the them and the natural or Napierian logarithm function. Using  $x_{min}$  and  $x_{max}$ , respectively, we have denoted the minimal and the maximal value that a variable can take, whereas  $n$  represents the group size or, in general, the sample size. Finally,  $x_i$  and  $x_j$  denote the values of the individuals in the group or in the sample, using the conventional symbol to represent the mean.

Diversity	Name of the index	Operational definition	Minimal value	Maximal value
Variety	<i>Blau's index</i>	$1 - \sum_{i=1}^k p_i^2$	0	$\frac{k-1}{k}$
	<i>Teachman's index</i>	$-\sum_{i=1}^k p_i \ln(p_i)$	0	$\ln(k)$
Separation	Standard deviation	$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$	0	$\frac{x_{max} - x_{min}}{2}$
	Mean Euclidean distance	$\frac{\sum_{i=1}^n \sqrt{\sum_{j=1}^n (x_i - x_j)^2}}{\sqrt{n^2}}$	0	$\frac{x_{max} - x_{min}}{\sqrt{2}}$
Disparity	Coefficient of variation	$\frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\bar{x}\sqrt{n}}$	0	$\sqrt{n-1}$
	<i>Gini's coefficient</i>	$\frac{\sum_{i=1}^n \sum_{j=i+1}^n  x_i - x_j }{\bar{x}n^2}$	0	$1 - \frac{1}{n}$

### Description of the study

In order to show how the different diversity indices can be obtained and to illustrate their interpretations, groups of eight people will be formed. In each group, the group members will have to answer questions on socio-demographic characteristics and social relations, and they will also have to use a 6-point scale to quantify their degree of agreement with a statement made (see Table 2).

Once the answers are given, the group members will share their answers and, for each variable, they will compute the two appropriate indices. A calculator will be useful to do so.

### Material necessary for the practice

In order to organise the information during the study, Table 2 will be used to record the results obtained, before analysing them.

Table 2. This table contains questions on socio-demographic variables and statements that convey certain opinions. In the latter case, the participant/student is asked to give a score on a rating scale, where “1”, “2”, “3”, “4”, “5” and “6” correspond, respectively, to “totally disagree”, “disagree”, “partially disagree”, “partially agree”, “agree” and “totally agree”. Mark the degree of agreement or disagreement with a ✕ sign. Note that, among the other questions, the students are asked about the average number of e-mails received per day.

Gender:	Age:	Bachillerato type for admission to university:					
How many e-mails, on average, do you receive from your friends each day?							
Statements	1	2	3	4	5	6	
Active euthanasia should be made legal							
Some crimes justify a sentence of “life imprisonment”							
Defrauding the public treasury is justified							
Scientific research is the only way to attain knowledge that is certain							
Nuclear energy should be prioritised instead of fossil fuels							
Rules exist only to be broken							
There is a divine being that created the universe							
The goodness of human beings is innate							
Financial cutbacks are preferred in Health rather than in Education							
Statistics are useful for psychologists							
I'd reduce my quality of life to help improve the quality of life in the Third World							
Military conflicts may have a justification							

Table 3 will be used to indicate, for each variable, the type of diversity and the numerical value for the two indices designed to quantify this type of diversity. Note that there are a total of 18 variables. Bearing in mind that the groups will be formed of eight people each, each person can compute both indices for two variables. Afterwards, the results will be available for everyone to assess the differences and similarities.

Table 3. For each variable, specify the type of diversity (variety, separation or disparity) and provide the numerical value for the appropriate indices.

Variable	Type of diversity	Value index 1	Value index 2
Gender			
Age			
Bachillerato			
E-mails			
Euthanasia			
Life imprisonment			
Defrauding			
Investigation			
Types of energy			
Rules			
Divinity			
Goodness			
Cutbacks			
Statistics			
Quality of life			
Conflicts			

**Questions to think about and answer:** They need to be answered by each group of students separately (unless stated otherwise):

1. After computing the diversity indices, what is the interest of expressing diversity in terms of normalised indices, that is, ranging between 0 and 1?
2. Once the necessary calculations are performed, sort the features measured for your group in ascending order of diversity. Which features present the most and least diversity?
3. For all groups considered together, which features present the most and least diversity? Are these results consistent across all groups?
4. Given that the diversity results could be heterogeneous for the groups across the different features, what statistical indices can be used to offer a global synthesis for all groups? What graphical representations would you consider useful? Why?
5. In this text there are different examples of characteristics that correspond to each of the types of diversity. Could you indicate two more examples for each type of diversity?

## References

- Byrne, D. E. (1971). *The attraction paradigm*. New York, NY: Academic Press.
- Harrison, D. A. & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, 32, 1199-1228.

[Back](#)

---

## PRACTICE ON BIVARIATE DESCRIPTION

### Practice aim

The purpose of the current practice is to work with the statistical indices design for bivariate description, that is, for quantifying the relation between variables on which the students have gathered information using different recording techniques. Depending on the measurement scale, in some cases we can only describe the strength of the association (qualitative nominal scale variables) and possibly also its direction or sign (ordinal, interval and ratio scale variables). Therefore, with this practice we intend to stress the need to take into consideration the measurement scale of each of the two variables whose association is being quantified or estimated in the population. (The latter aspect is studied thoroughly in the “Estadística” course). Moreover, the practice will illustrate that some indicators are more widely applicable than others (e.g., indicators that are adequate regardless of the number of categories in the nominal or ordinal variables).

For cases in which there are only few data, it is suggested that students carry out calculations by hand in order to better understand the mathematical operations to which the data are subjected. The purpose is to encourage a deeper comprehension of the information that the indices actually provide. In other cases, the practice will focus on obtaining the numerical indicators and the graphical representations by means of R-Commander, an open-source statistical package that can be useful for students in all projects performed in a variety of courses in the degree of Psychology. Additionally, this knowledge is potentially useful for their professional career after completion of higher education.

### Description of the tasks

Keep in mind that some of the questions refer to qualitative and ordinal variables, whereas others focus on quantitative variables. Moreover, it is possible that certain questions require exploring the relation between variables that are not expressed in the same measurement scale. Therefore, it has to be taken into account whether there is an indicator appropriate for variables measured in different scales (and whether this indicator is available in R-Commander) or, on the contrary, it is necessary to choose an alternative way to carry out the bivariate description.

**Using the data gathered via the interview:****[Excel file in Virtual Campus]:**

1. In order to study the strength of relation between gender and the aspect that would motivate participants to quit smoking:
  - a. create the contingency table;
  - b. compute the expected frequencies;
  - c. compute the appropriate quantitative index;
  - d. interpret the information;
  - e. pick an adequate graphical representation.

Use the following example as a model for obtaining expected frequencies: a sample of 136 individuals is available, with measurements in two variables (gender and subtype of phobia). The absolute frequencies for each category are placed in the row and column sums (i.e., the marginal totals in the contingency table):

	agoraphobia	social	specific	total
female				67
male				69
total	40	48	48	<b>136</b>

Example for obtaining the expected frequency for people who are women and are diagnosed with agoraphobia: in the sample, the probability of being a woman is  $67/136 = 0.493$ . The probability of being diagnosed with agoraphobia is  $40/136 = 0.294$ . The probability that a person is in both categories is  $0.493 * 0.294 = 0.145$ . To obtain the frequency out of these probabilities, it is necessary to multiply this result by the number of people in the sample: therefore,  $0.145 * 136 = 19.7$ .

2. In order to study the strength of association between gender and the reasons for quitting smoking, compute also the *odds* between the categories (for instance, health vs. others) of the variable “reasons for quitting smoking” for women and compare it to the *odds* for men, obtaining the *odds ratio*.



Use the following example as a model for how to obtain the *odds ratio*: a sample of 20 people is available, with measures in two variables (type of contract and diagnosis of burnout):

	Burnout: yes	Burnout: no	Total
Temporary contract	9	1	10
Indefinite contract	3	7	10
Total	12	8	<b>10</b>

- *Odds* for “burnout: yes” within “temporary contract” category =  $9/1 = 9$  people diagnosed for each person without diagnosis.
  - *Odds* for “burnout: yes” within the “indefinite contract” category =  $3/7 = 0.43$  people diagnosed for each person without diagnosis.
  - *Odds ratio*:  $9/0.43 = 20.93$  times more at risk for suffering from burnout if the contract is temporary as compared to indefinite.
3. Recode the “age” variable so that it now only has 2 or 3 categories that can be ordered. Assess the strength of association between “age recoded” and the rating of the severity of smoking:
    - f. obtain the number of concordances, discordances and ties for all the comparisons between pairs of participants:
    - g. obtain the value of an appropriate quantitative indicator (consult the Appendix);
    - h. interpret its value in terms of intensity and direction of the relation between the two variables:
    - i. create a dot diagram.
  4. Assess the strength of association between the gender of the participants and their rating of the severity of smoking. Use the Probability of superiority index (explained below) and another appropriate indicator chosen from the ones explained in the theoretical sessions.
  5. Assess the degree to which the participants agree on the severity of the three habits: use Kendall’s coefficient of concordance (explained below).

## Probability of superiority

When studying the effect size or the strength of association between a qualitative dichotomous variable and a quantitative variable, the most common option is to use one of the indices of the family of standardised mean difference indicators (Cohen's  $d$ , its correction called Hedges'  $g$ , or Glass's  $\Delta$ ). Nevertheless, it is possible that the interpretation in terms of standard deviations is not directly intuitive. In order to make comprehensible the information provided by statistics in these cases, the *common language effect size* has been proposed (McGraw & Wong, 1992). This is expressed as the probability that an individual extracted at random from one population has a greater value in the quantitative variable than an individual extracted at random from the other population (e.g., the probability that an autochthonous person has more friends than an immigrant).

In the present practice we will focus on the nonparametric version of the *common language effect size*, the one that only requires ordinal information. This nonparametric index has been referred to as the “probability of superiority” (PS) and has been promoted by Grissom (1994; Grissom & Kim, 2001), although it has also been designated as A in the review article by Ruscio (2008). The definition of this effect size index is  $PS \equiv A = \left[ \#(Y_1 > Y_2) + 0.5\#(Y_1 = Y_2) \right] / n_1 n_2$ . This notation entails identifying  $\#(Y_1 > Y_2)$ , the number of cases in which the members of group 1 have a larger value than the members of group 2, and then adding  $0.5\#(Y_1 = Y_2)$ , the number of cases in which equal values are obtained (note that in this latter case half the weight is given). Finally, after tallying, this sum is divided by the total amount of comparisons made,  $n_1 n_2$ , the product of multiplying the sizes of the two groups.

As an example of the code that could be used to compute this index, consider the case in which we compare the ratings of quality of life made by 4 people who claim to have close friends (“calidad\_sí”) and 4 people who say they do not have such friends (“calidad\_no”).

```
> calidad_si <- c(4,3,5,1)
> calidad_no <- c(1,4,2,4)
> probsup <-function (x1,x2){
+ n1 <- sum(!is.na(x1))
+ n2 <- sum(!is.na(x2))
+ mayor <- sum(unlist(lapply(x1,x2,FUN='>')),na.rm=TRUE)
+ igual <- 0.5*sum(unlist(lapply(x1,x2,FUN='==')),na.rm=TRUE)
+ prob_sup <- (mayor+igual)/(n1*n2)
+ cat('Frecuencia >: ',mayor,'\n')
+ cat('Frecuencia =: ',igual,'\n')
+ cat('Probabilidad superioridad (PS): ', prob_sup,'\n')
+ }
> probsup(calidad_si,calidad_no)
Frecuencia >: 8
Frecuencia =: 1.5
Probabilidad superioridad (PS): 0.59375
```

### Kendall's coefficient of concordance

This index is useful to quantify the strength of association between qualitative and ordinal variables when the data collection design entails repeated measures and not independent groups. In the present section we will focus on Kendall's coefficient of concordance (Kendall & Babington Smith, 1939), designated by  $W$ . It is important to distinguish  $W$  from the tau indices ( $\tau_a$ ,  $\tau_b$ , and  $\tau_c$ ) also devised by Kendall, given that the latter are used to study the relation between two ordinal variables. The tau indices are another topic of the current practices and we would like the student to see the differences in application and information provided in comparison with  $W$ .

Mathematically, the  $W$  index can be applied to cases in which the qualitative variable is dichotomous, but actually its application has focused exclusively on situations in which the categorical variable has three or more levels. In order to compare two sets of ranks, Spearman's rank correlation can be used or, alternatively, Kendall's tau coefficients, given that they, in contrast with  $W$ , can take negative values in the case of an inverse relation. The impossibility of comparing Kendall's  $W$  to the other coefficients is one of the reasons to apply the former only when three or more sets of ranks are available (Sheskin, 2004).

The main application of  $W$  is as an indicator of the degree of agreement between  $m$  judges who rate  $n$  objects or variables. Note that, in contrast with common notation in which the number of participants in a study is designated by  $n$ , we have here preferred to maintain the most frequent notation ( $m$ ) in the context of the  $W$  index. The value of 0 indicates total lack of agreement, whereas 1 refers to total agreement. The expression for computing  $W$  is as follows:

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \sum_{i=1}^n (R_i - \bar{R})^2}{m^2(n^3 - n)}, \text{ where } R_i = \sum_{j=1}^m r_{ij} \text{ and } \bar{R} = \frac{m(n+1)}{2}$$

In the expression presented above,  $r_{ij}$  denotes the rank that a judge  $j$  assigns to an object  $i$ , with  $R_i$  being the sum of the ranks assigned to an object;  $\bar{R}$  is the average of all ranks and  $S$  is the sum of the squared deviations  $R_i - \bar{R}$ . It can be seen that  $\bar{R}$  depends only on the number of judges and objects assessed, given that with this information, the ranks that can be assigned are predetermined: what is not known a priori (i.e., before obtaining the empirical data) is how the judges will distribute the ranks among the different objects.

In order to understand how the  $W$  index is used, we offer the following example. Three participants ( $m=3$ ) sort according to their preference four activities ( $n=4$ ) to which

they can dedicate their spare time: reading, doing adventure sports, watching the TV or sleeping. The following data matrix is obtained:

	Reading	Adventure	Television	Sleeping
Participant 1	$r_{11} = 2$	$r_{21} = 1$	$r_{31} = 4$	$r_{41} = 3$
Participant 2	$r_{12} = 1$	$r_{22} = 2$	$r_{32} = 4$	$r_{42} = 3$
Participant 3	$r_{13} = 2$	$r_{23} = 1$	$r_{33} = 3$	$r_{43} = 4$
$R_i = \sum_{j=1}^m r_{ij}$	5	4	11	10
$(R_i - \bar{R})^2$	6.25	12.25	12.25	6.25

The first row below the data for the participants represents the sum of ranks for each of the  $n$  activities, that is, the  $R_i$  values. The row that contains the deviations from the mean requires, logically, that the mean of all ranks is first computed:  $\bar{R} = \frac{m(n+1)}{2} = \frac{3(4+1)}{2} = 7.5$ . In this way, for instance, for reading, the squared deviation from the mean is equal to  $(5 - 7.5)^2 = (2.5)^2 = 6.25$ . The sum of the deviations for the four activities is  $6.25 + 12.25 + 12.25 + 6.25 = 37$ . Hence, the value of  $W$  is obtained as follows:

$$W = \frac{12 \sum_{i=1}^n (R_i - \bar{R})^2}{m^2(n^3 - n)} = \frac{12 \cdot 37}{3^2(4^3 - 4)} = \frac{444}{9 \cdot 60} = 0.82$$

Considering that  $W$  can take values between 0 and 1, the value obtained indicates a strong agreement on the preferred leisure activities. This agreement can also be seen when inspecting the data matrix, which suggests that adventure and reading (in this order) are considered more highly than sleeping and watching television. It has to be noted that, in this example, the index cannot take the value of 0 (i.e., the deviations from the average rank cannot be equal to zero) if there are no ties in the assigned ranks. That is, if only integer numbers are assigned as ranks (i.e., without the possibility of having two preferred activities that both receive a rank of 1.5), the sum of ranks cannot be a non-integer (e.g., 7.5).

---

**Using everyone's data from the Stroop experiment [Excel file in the Virtual Campus]:**

1. In order to study the strength of association between the number of errors and the response time for correct answers:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.
2. In order to study the strength of association between the type of stimulus and the response time for correct answers:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.
3. In order to study the strength of association between the type of stimulus and the response time for incorrect answers:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.
4. In order to study the strength of association between the type of stimulus and the number of errors:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.
5. In order to study the strength of association between the type of task and the response time for correct answers:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.
6. In order to study the strength of association between the order in which the conditions are presented and the response time within the “concordant” condition:
  - a. Use an appropriate graphical representation
  - b. Compute a quantitative indicator
  - c. Interpret the results.

**Material necessary for the practice**

- The data matrix constructed from the information gathered via the interview.
- The results from the practice on the Stroop effect.

**Questions to think about and answer**

7. Why can it be useful to collapse categories when studying the association between two qualitative variables? What kind of criterion has to be followed when performing such grouping?
8. If the interest is centred on two quantitative variables, what should be done prior to computing a statistical index?
9. How would you interpret a value of 0 for Pearson's correlation coefficient?
10. How would you interpret a value of 3 for the chi-square index?
11. How would you interpret a value of  $-1$  for Cramér's V index?

## References

- Grissom, R. J. (1994). Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology*, 79, 314-316.
- Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods*, 6, 135-146.
- Kendall, M. G., & Babington Smith, B. (1939). The problem of m rankings. *The Annals of Mathematical Statistics*, 10, 275-287.
- McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, 111, 361-365.
- Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, 13, 19-30.
- Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures* (3rd Ed.). London, UK: Chapman & Hall/CRC.

## Recommended readings

- Peró, M., Leiva, D., Guàrdia, J. & Solanas, A. (Eds.) (2012). *Estadística aplicada a las ciencias sociales mediante R y R-Commander*. Madrid: Garceta. Chapter 4.
- Solanas, A., Salafranca, Ll., Fauquet, J. & Núñez, M. I. (2005). *Estadística descriptiva en ciencias del comportamiento*. Madrid: Thomson. Chapters 10, 11 and 12.

## Appendix

( $n$  – number of participants,  $C$  – number of concordances,  $D$  – number of discordances,  $T_X$  – number of ties in the X variable,  $T_Y$  – number of ties in the Y variable,  $T_{XY}$  – number of ties in both variables,  $q = \min\{I, J\}$ ,  $n(n-1)/2 =$  number of comparisons).

$$\hat{\gamma} = \frac{C-D}{C+D} \quad \hat{d}_{Y|X} = \frac{(C-D)}{\left[ \frac{n(n-1)}{2} - T_X \right]} \quad \hat{e} = \frac{(C-D)}{\left[ \frac{n(n-1)}{2} - T_{XY} \right]}; \quad -1 \leq \hat{e} \leq 1$$

$$\hat{\tau}_a = \frac{(C-D)}{\frac{n(n-1)}{2}} \quad \hat{\tau}_b = \frac{(C-D)}{\sqrt{\left( \frac{n(n-1)}{2} - T_Y \right) \left( \frac{n(n-1)}{2} - T_X \right)}} \quad \hat{\tau}_c = \frac{2q(C-D)}{n^2(q-1)}; \quad -1 \leq \hat{\tau}_c \leq 1$$

## PRACTICE ON DYADIC CONCORDANCE IN PERSONALITY

### Introduction

Several studies have verified the similarity between spouses. In this regard, most married couples are people from the same ethnic group, which can be explained by the *similarity-attraction model*. One consequence of this model is that the attraction felt by people is a function of their degree of similarity, that is, when the similarity is greater, stronger attraction is expected. Another finding that supports the model is the age similarity within couples, which can be explained by factors that are not necessarily psychological. A further question is whether spouses are similar in terms of their personality traits: is there a dyadic similarity within the couple regarding the dimensions of personality? It is relatively common to come across the statement that “opposites attract”, but research does not support this. In general, empirical studies have found only slightly positive values of the product-moment correlation coefficient, that is, they provide evidence only for a certain dyadic similarity between spouses. In fact, the strength of association is not especially notable, except in a few studies and only for certain dimensions of personality (Han, Weed, & Butcher, 2003). Therefore, although there is some evidence to support the idea of dyadic similarity, it is not sufficiently solid to hold that dyadic similarity in personality traits is a critical factor in forming couples. Hence, the similarity-attraction model, when applied to personality, does not fully explain how mutual selection occurs within couples. Moreover, it should be taken into account that some features, such as ethnicity or social customs, are more easily perceived than others, e.g., personality.

In the present study, the intention is to provide new evidence on spouses’ dyadic similarity in the case of personality characteristics and to contrast the data with the consequences of the similarity-attraction model.

### Description of the study

Groups of eight students will be formed. Each student should administer the psychological test to measure the personality characteristics of three couples. Specifically, the psychological test called NEO-FFI (Costa & McCrae, 2002) will be applied, because it enables evaluation of the “Big Five” factors: neuroticism, extraversion, openness, agreeableness and conscientiousness. Before focusing on the items of the psychological test, information on each participant’s gender and age should be written in the corresponding part of the answer sheet. Then, the 60 statements are answered. If there are elderly participants, the person administering the test can aid the



participant in order to ensure that s/he correctly understands the statements made in the items and that s/he gives precise responses.

As indicated, the students in each group should administer the test to both members of three couples: according to the age of the spouses, the couple will be labelled as “young”, “mature” or “older”. A young couple is one in which the spouses’ ages are within the interval 25-45 years, inclusive. For mature couples, both members have an age within the interval 46-55 years, inclusive. Finally, older couples are the ones in which the two individuals are aged 66 or more. It would be more appropriate not to include couples in the study when the two members belong to different age groups, as defined here. If this cannot be avoided, then the couple will be classified according to the age of the younger individual. The classification thus established is, of course, arbitrary, but it intends to verify whether there are differences in dyadic similarity in terms of personality across three more or less distinguishable generations.

Once the eight members of each group of students have collected the data via the psychological test, they should introduce the data in an Excel file. Given that part of the practice is to carry out a study with all groups, the same procedure should be followed. Table 1 shows the order in which the variables have to be input in the Excel worksheet. The information for all the groups can be aggregated in the same Excel file and, afterwards, incorporated in R-Commander.

Table 1. This is show the variables should be introduced (and the data matrix created) in Excel. Firstly, each variable should be placed in a separate column. Secondly, note that in the first five columns, the data regarding the personality traits of the woman in the couple are located in the following order: neuroticism, extraversion, openness, agreeableness and conscientiousness. The sixth column is for the age of the woman. Finally, the same order is followed for the six variables corresponding to the man in the couple.

NEU_W	EXT_W	OPE_W	AGR_W	CON_W	AGE_W	NEU_M	EXT_M	OPE_M	AGR_M	CON_M	AGE_M
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

### Material necessary for the practice

The psychological test NEO-FFI (Costa & McCrae, 2002) consists of 60 items. This inventory assesses the five personality dimensions that correspond to the model known as the “Big Five”: neuroticism, extraversion, openness, agreeableness and conscientiousness. The answers to the items are given using a scale of several degrees; the application can be collective and the estimated time necessary for answering is 20 minutes. This material will be given to the students.

**Questions to think about and answer:** should be done by each group of students, unless stated otherwise:

1. Using the information available for each group, are there linear relations between the pairs of variables (i.e., scores on each dimension of personality) for the different couples? Choose an appropriate graphical representation to answer this question.
2. Given the information obtained in the previous question, obtain an adequate measure of the strength of association between the different characteristics of personality within the couple. What are the conclusions regarding the dyadic similarity between the spouses with the data obtained by each group of students?
3. Are there anomalies in the data suggesting that the statistical results might not be representative of all the data available?
4. If correlation coefficients are used in studies on dyadic similarity (as is common in social psychology), are the measures really quantifying dyadic similarity between psychological traits?
5. Considering the information obtained by all groups jointly, what can you say about the likelihood of the similarity-attraction model?
6. What could be the purpose of using the variance-covariance matrix in this case? Is the correlation matrix more useful?
7. Considering the information obtained by all groups jointly, are there differences if the statistical analyses are carried out separately according to the three age strata?
8. Explore whether there is a relation between the age group and any of the personality traits. Assess the strength of association. (Take into consideration whether there are different normative data according to the age of the participants).
9. Explore whether there is a relation between gender and any of the personality traits. Assess the strength of association. (Take into consideration whether there are different normative data according to the gender of the participants). Note that for questions 8 & 9 you might need to re-organise the data.
10. Finally, which are the five most relevant things you learned from this practice?

## References

- Costa, P. T., Jr., & McCrae, R. R. (2002). *Manual NEO PI-R*. Madrid: TEA Ediciones.
- Han, K., Weed, N. C., & Butcher, J. N. (2003). Dyadic agreement on the MMPI-2. *Personality and Individual Differences*, 35, 603-615.

[Back](#)

## Section 5:

### Inferential statistics

Statistical decision making is commonly done referring either to a theoretical sampling distribution (i.e., a probability model, as in Section 3) or to an empirically derived permutation/randomization distribution. The content of this section is based on the former approach. The information about the reference distribution is used to obtain the statistical significance associated with the values of correlation, chi-square,  $t$ -statistic, some of which were dealt with in Section 4. This information also enables interval estimation useful for identifying the most plausible values of the parameter in the population.

The content of the present section is dedicated to a series of statistical significance tests for a variety of situation including one-sample test and multi-sample comparisons. These tests are to be chosen according to the research aim and the measurement scale of the variables of interest, exactly in the same way as the indices and graphs that were included in the (descriptive) Section 4. The emphasis is put on the usefulness and limitations of the techniques, as well as their assumptions that ought to be met for the information ( $p$  values) provided to be meaningful and not biased.

Once again, as was the case for the descriptive statistics, we have included data obtained from actual research, or simulate according to such research, and we also propose an initial data collection stage prior to carrying out the analyses. We consider that the context of use of the statistical techniques and the substantive meaningfulness of the data are necessary for making sense of the statistics and learning how and when to use the different alternatives available.

Finally, we hope that the reader would be cautious with the extent of her/his statistical conclusions when random sampling does not take place. Moreover, we expect that the interpretation (or at least the understanding) of  $p$  values would go beyond the classical statements that an outcome is either “statistically significant” or “not statistically significant”, keeping in mind the actual meaning of such a phrase with respect to the probability of the outcome being due to random fluctuations, but not as a result of real population-level differences, association, etc.

## PRACTICE

### ON THE SAMPLING DISTRIBUTION OF A STATISTIC

#### Introduction

The sampling distribution of a statistic is the foundation for the statistical techniques to be addressed in this course. It will facilitate an understanding of the characteristics of estimators and assist in making interval estimates and statistical decisions based on the relationship between sample statistics and population parameters assumed in the statistical hypothesis. The present practical focuses on studying the sampling distribution of the sample mean. We distinguish two cases:

**Case 1:** Let  $X$  be a random variable with  $f(X) \sim N(\mu, \sigma)$ , and  $X_1, \dots, X_n$  obtained from a random sample of this reference population. Then the sampling distribution of the sample mean ( $\bar{X}$ ) follows a normal distribution with parameters<sup>1</sup>

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{n\mu}{n} = \mu,$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2 X}{n}.$$

And therefore the standard deviation of the sampling distribution of the sample mean, i.e., its standard error, is:

$$SE(\bar{X}) = \frac{\sigma X}{\sqrt{n}}.$$

**Case 2:** When the probability distribution of the random variable  $X$  is not normal but the sample size is large enough, the sampling distribution of the sample mean continues to be normal. Specifically, the **Central Limit Theorem** states that a sum

of i.i.d.  $\left(S_n = \sum_{i=1}^n X_i\right)$  random variables with finite mean and variance converges by law to a normal distribution as  $n$  tends to infinity.

○ Formally:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

<sup>1</sup> In the following expressions we have assumed that the observations of any sample are independent and identically distributed random variables (i.i.d.).

## Tasks

1. Install the package **RcmdrPlugin.TeachingDemos** using R-GUI or the following command:

```
install.packages("RcmdrPlugin.TeachingDemos")
```

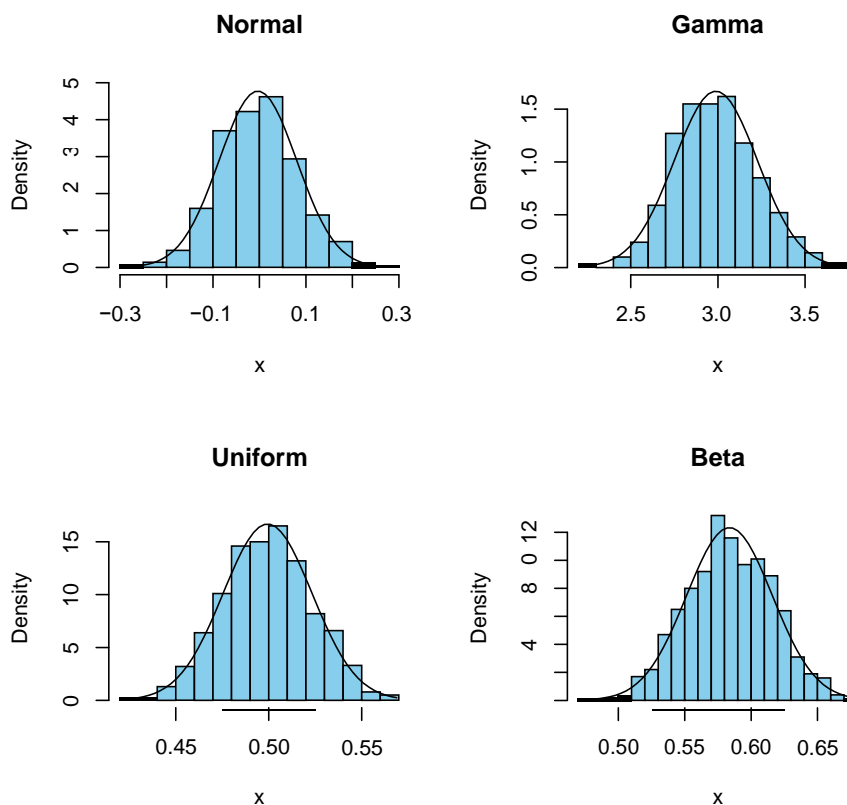
2. Download the plug-in in R-Commander using R-Commander GUI or through the command:

```
library(RcmdrPlugin.TeachingDemos)
```

3. Access the *Central Limit Theorem* module from the menu *Demos* that appears in the menu bar of R-Commander after loading the package.

4. Familiarise yourself with the module and its results:

sample size = 150



## Questions

1. Identify each of the simulated distributions. Determine if they are appropriate for discrete and continuous random variables, as well as their degree of symmetry and expected value (approximate).
2. Try the following values of  $n$ : 1, 5, 10, 50 and 100. Discuss the effect on the different distributions.
3. Try the following values relative to the number of simulated samples and with a fixed sample size of  $n = 10$ : 10, 100, 500, 1000 and 5000. Discuss the effect on the different distributions.
4. Suppose that a random sample of 10 university students is obtained and their CI is measured. Knowing that  $CI \hookrightarrow N(\mu = 100, \sigma = 15)$ , justify whether the following statements are true or false:
  - a) The sampling distribution of the sample mean of CI is not normal because the sample size is small.
  - b) The expected value of the mean of CI is 100.
  - c) The standard error of the sampling distribution is 15.
  - d) If a sample of 20 students was obtained, the standard error would have been smaller.
5. Once the practice is finished, determine if the following statements are true or false and justify your answer:
  - a) The Central Limit Theorem ensures that the sampling distribution of any statistic is normal as long as  $n$  is always sufficiently large.
  - b) The theorem applies only to the case of continuous variables.
  - c) Increasing the size of the samples shifts the expected value of the statistic to the left or right depending on the size of the sample.
  - d) Decreasing the size of the sample also decreases the standard error value.
6. Copy and run the following code in R-Commander and answer the following questions:
  - a) What does this simulation demonstrate?
  - b) Is the same sample size effect observed on the sampling distribution of the sample mean?
  - c) Is the same sample size effect observed on the standard error?

```
tamano <- c(1, 10, 50, 100)
par(mfrow=c(2,2))
muestras <- 1000
param <- list(size=12, prob=0.2)
for (i in 1:length(tamano))
```

```

{
  param$n <- tamano[i]*muestras
  binom.mat <- matrix( do.call(rbinom,param),ncol=tamano[i])
  binom.mean <- rowMeans(binom.mat)

  x <- seq( min(binom.mean),max(binom.mean),length=50)
  binommax <- max(
  dnorm(x,mean(binom.mean),sd(binom.mean))) tmp.hist <-
  hist( binom.mean,plot=FALSE,nclass=16)
  binommax <- max( tmp.hist$density,binommax )*1.05

  hist(binom.mean, main=paste("Binomial n=
  ",tamano[i],
  sep=),xlab="x",col="skyblue",freq=FALSE,
  ylim=c(0,binommax), nclass=16)
  lines(x,dnorm(x,mean(binom.mean),sd(binom.mean)))
}

```

7. Thanks to the *de Moivre-Laplace Theorem*, it is known that a binomial distribution approaches a normal distribution as long as  $n$  tends to infinity.

So,  $B(N, p) \xrightarrow{D} N\left(\mu = Np, \sigma = \frac{\sqrt{Np(1-p)}}{n}\right)$ . Compare the estimation provided by the previous

simulation for the expected value of the sampling distribution of the statistic and its standard error in the case of  $n = 100$  with the values obtained using the normal approximation formulas.

Note that, to avoid confusion in the notation,  $n$  represents the sample size in the simulation and  $N$  is the number of trials in the binomial distribution. Use the following code to obtain the expected value and the standard error estimated from the simulation:

```

esp.matematica <- mean(binom.mean)
error.std <- sd(binom.mean)

```

8. From a general population, 100 people are randomly selected and the variable  $CI \rightsquigarrow N(\mu = 100, \sigma = 15)$ . Answer the following questions:

- Identify the sampling distribution of the sample mean of  $CI$  for characteristics of the samples.
- Find the probability that the average  $CI$  for the sample is less than 108 points assuming an infinite population.
- What average value of  $CI$  would leave 95% of the sample below that value?

## PRACTICE ON STATISTICAL PROCESS CONTROL

### Introduction

From a practical standpoint, there are not many applications for probability intervals, although their interest in *statistical process control* should be emphasised, particularly in the case of so-called control charts. *Control charts* are graphs based on probability intervals, which enable us to determine whether or not a process is appropriately operational according to a previously established criterion. Although control charts were initially linked to the statistical control of production quality, they have since spread to other areas. In the present practical, the usefulness of control charts in the field of human resource management will be seen and a demonstration of how these graphical representations are produced will be explored.

In all processes, there are two types of factors, so-called *systematic factors* and *non-systematic factors*. On one hand, systematic factors or assignable causes are those that introduce a deterministic component in the variability or fluctuations of the measurements. The objective is to identify and control them in order to obtain only the variability that is produced by random uncontrollable factors of low incidence. The assignable causes produce an amount of unpredictable variability whose effect disappears by eliminating its cause. On the other hand, non-systematic factors or non-assignable causes are those that produce an apparent random variability of low magnitude. In any case, the sum of different factors of low impact shows random fluctuation for which it is very difficult to identify the causes. These non-assignable causes produce stable variability that is difficult to eliminate but low in importance.

By *tolerance interval*, we mean a band or range of permissible values for a characteristic of interest, i.e., a set of values that, although they differ from the value of the parameter specified, are not far enough from it to consider that the process is not controlled. Thus, it distinguishes the *nominal value* from the values for *control limits*. The nominal value refers to the value of the parameter established as suitable or maximally acceptable for any characteristic of a product or process in general (for example, proportion of defective products, rate of work absenteeism, percentage of administrative documents with errors, internal diameter of a washer, etc.). In the case of control limits, just as different factors can have a random effect on a production process, the measurements can fluctuate and deviate from the nominal value without necessarily indicating that the process is not in accordance with the specified nominal value. Thus, the control limits (*upper and lower*), which are centred with respect to the nominal value, set the permissible tolerance range in which the process is considered to be under control. If the measurements fall outside the area defined by the control limit, the process is said to be *out of control*. In fact, a single measurement can be outside the control limits, but this should not be considered sufficient to conclude that the process is out of control, though two consecutive measures out of the range defined by



the control limits is an unlikely outcome if the process is under control.

As mentioned above, for different process parameters of interest, control charts may be developed (Figures 1 and 2) to show the nominal value, the control limits and the different measures. In the *control charts for attributes* (Figure 1), the characteristic of interest has two possible values, such “apt” or “inapt”, “pass” or “fail”, “defective” or “non-defective”, “conforming” or “nonconforming”, “satisfied” or “dissatisfied”, “involved” or “uninvolved”, “achieves the goal” or “does not achieve the goal”, among other possibilities. Therefore, the control chart for attributes requires that the variable is nominal or, in any other case, the permitted values are transformed into two discrete values that are mutually exclusive and exhaustive.

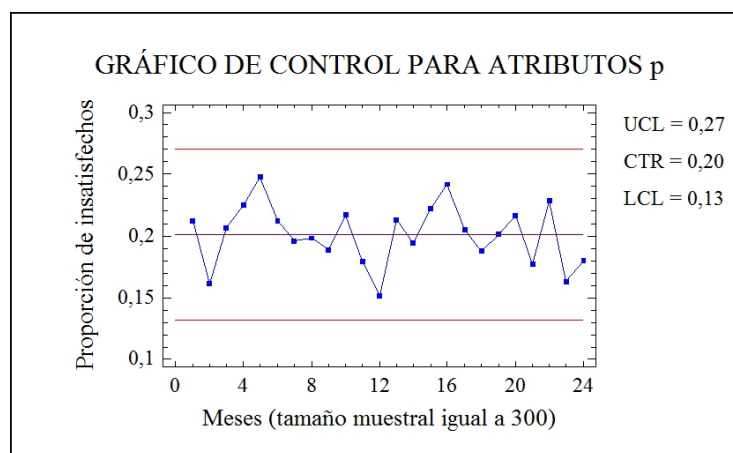


Figure 1: Shows a control chart for attributes, where the horizontal line in the centre corresponds to the nominal value; the lower and upper horizontal lines indicate the lower and upper control limits, respectively. The horizontal axis shows the different points in time at which the random sample of the system (e.g., an organisation) was obtained and the vertical axis shows the proportion of people who felt dissatisfied. *UCL*, *CL* and *LCL* denote the upper control limit, the centre line (nominal value) and the lower control limit, respectively.

*Control charts for variables* are of different types, but one of the most common represents the average value of a numerical characteristic (Figure 2). It is also typical to use control charts for standard deviation and even for the range. Therefore, to construct a control chart for variables, it is necessary to use an interval, ratio or absolute scale.

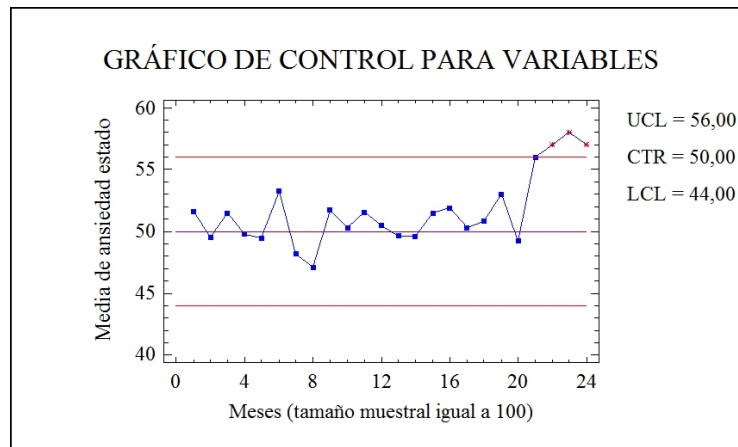


Figure 2: Shows a control chart for variables, particularly for the mean, where the horizontal line in the centre corresponds to the nominal value; lower and upper horizontal lines indicate the lower and upper control limits, respectively. The horizontal axis shows the different points in time at which the random sample of the system (e.g., an organisation) was obtained and the vertical axis shows the average state of anxiety of the people selected in the sample. *UCL*, *CL* and *LCL* denote the upper control limit, central line (the nominal value) and the lower control limit, respectively.

### Control charts for attributes

Table 1 presents the 12 measurements obtained monthly from a sample size of 150. The measurements correspond to the proportion of people who feel involved with the objectives of the organisation. To realise the practical, we assume that the organisation consists of a total of 2000 people and that each month 150 people are randomly selected to form the sample. It is also assumed that a proportion lower than 0.10 of the members of the company feeling involved may have undesirable consequences on the organisational climate. Therefore, the people responsible for the company's human resources conduct a monthly control to determine the proportion of people not feeling involved. To obtain the limits of a probability interval for the proportion, two possible statistical calculations for either infinite or finite populations exist, as shown below:

$$\pi \pm z_{\alpha/2} \sqrt{\frac{\pi(\pi-1)}{n}}$$

$$\pi \pm z_{\alpha/2} \sqrt{\frac{\pi(\pi-1)}{n}} \sqrt{\frac{N-n}{N-1}}$$

In the above expressions,  $\pi$ ,  $N$ ,  $n$  and  $z_{\alpha/2}$  denote the nominal value for the proportion, the size of the population, the size of the sample and a value corresponding to the unit normal

distribution, i.e.,  $N(0, 1)$ , respectively. In statistical quality control, it is common to use  $z_{\alpha/2} = 3$ , because this implies that if the process corresponds to a fixed nominal value with a probability of 0.9973, the measurements will be within the tolerance interval. Therefore, when the proportion is located outside this range, then the evidence is sufficient to consider that the process is out of control. In any case, to decide that a process is out of control, it is common to require the presence of two or more consecutive points outside the control limits or the existence of a series of measures that, although they fall within the tolerance interval, show a monotonously increasing or decreasing pattern.

Register (monthly)	Number of people not feeling involved	Proportion of people not feeling involved
1	19	0.1267
2	12	0.0800
3	10	0.0667
4	23	0.1533
5	15	0.1000
6	14	0.0933
7	29	0.1933
8	9	0.0600
9	12	0.0800
10	22	0.1467
11	11	0.0733
12	8	0.0533

Table 1: Over the year, the number of people in the sample always totals 150 individuals. The middle column shows the number who reported not feeling involved with the organisational objectives and the column to the right indicates the proportion of the total that this number represents.

Guided by a series of questions, students will construct the corresponding control chart for attributes. A sheet of graph paper is required or, alternatively, a paper with grids. Answer the following questions while constructing the control chart for attributes:

1. Draw a horizontal line for the horizontal axis and mark the points corresponding to the number of available measurements over a period of time.
2. Draw a vertical line for the ordinate axis and mark the points corresponding to the significant values for the statistic of interest.
3. What is the nominal value of the characteristic of interest in the study?  
Draw the corresponding horizontal line.
4. What are the values for the control limit? Draw the two corresponding horizontal lines.
5. From a conceptual perspective, is the interpretation identical for a measure obtained above the upper control limit and for one obtained below the lower control limit?
6. In the samples available, what is the sample size? What are the maximum and minimum values that can be obtained for people not feeling involved? Within what two limits, both inclusive, can the possible values of statistical proportion be obtained?

7. In the control charts, circle the different values of the statistic of interest for each and every one of the different points in time at which we obtained a measurement of the process. After this task, join the adjacent points using a straight line.
8. Can the process be considered under control during the registration period?

### Control charts for variables

The *Yerkes-Dodson Law* (Yerkes & Dodson, 1908) states that there is a U-shaped relationship between the level of cortical activity in people and their performance of a task. In other words, increased cortical activity is appropriate and enhances the performance up to a central and optimal point, but if cortical activity continues to increase, it can interfere with cognitive processes such as attention or memory. Sometimes, this law is also linked to the effects of stress. So, for proper performance, an intermediate level of stress is desirable, while low and high levels are inappropriate. Table 2 presents 12 measurements obtained monthly with a sample size of 100, corresponding to an average stress level for the members of an organisation composed of 1000 people. Also, the standard deviation is shown for the 100 values in each sample. It is assumed that the psychological test for quantifying the degree of people's stress can take values between 0 and 100, both inclusive, and therefore the central point is equal to 50. It is known that for the reference population, the scores of the psychological test that measures stress is normally distributed with mean and standard deviation equal to 50 and 10, respectively. To calculate the probability interval limits for the mean statistic for both infinite (upper) and finite (lower) populations, there are two approaches:

$$\mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

The notations of the formulas have already been explained except for  $\sigma$ , which corresponds to the standard deviation of the numerical variable of interest.

<b>Register (monthly)</b>	<b>Mean for stress</b>	<b>Standard deviation for stress</b>
1	52.12	12.21
2	49.45	13.45
3	50.25	9.18
4	51.98	10.01
5	50.02	11.98
6	47.56	10.24
7	52.23	7.99
8	48.92	12.12
9	50.34	12.48
10	52.67	9.67
11	53.21	10.73
12	54.78	9.44

Table 2: The table shows the monthly values corresponding to the mean and standard deviation for stress among a sample of 100, selected randomly from the population of the company.

Guided by a series of questions, students will construct the corresponding control chart for variables. A graph paper or a paper with grids is required. Answer the following questions while constructing the control chart for variables:

1. Draw a horizontal line for the horizontal axis and mark the points corresponding to the number of available measurements over a period of time.
2. Draw a vertical line for the ordinate axis and mark the points corresponding to the significant values of the statistic of interest.
3. What is the nominal value of the characteristic of interest in the study?  
Draw the corresponding horizontal line.
4. What are the values for the control limits? Draw the two corresponding horizontal lines.
5. From a conceptual perspective, is the interpretation identical for a measure obtained above the upper control limit and for one obtained below the lower control limit?
6. In the samples available, what is the sample size?
7. In the control charts, circle the different values of the statistic of interest for each and every one of the different points in time for which we obtained a measurement of the process. After this task, join the adjacent points using a straight line.
8. Can the process be considered under control during the registration period?

## Questions

The questions and insights explored below should be done jointly with the class:

1. Why set a range of admissible values in the control charts and not consider only whether or not the nominal value is obtained in the samples?
2. What are the consequences of modifying through  $1 - \alpha$ : what is the effect on the control limits?
3. If the sample size,  $n$ , is increased, what are the consequences for the control limits?
4. Must the population size,  $N$ , always be known to construct the control charts?
5. Do the control charts have any interest for human resource management?

## Additional references

- Hansen, B. L. & Ghare, P. M. (1987). *Quality control and application*. New Delhi: Prentice-Hall [Translation in Spanish, Hansen, B. L. & Ghare, P. M. (1990). *Control de la calidad. Teoría y aplicaciones*. Madrid: Díaz de Santos].
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.

[Back](#)

---

## PRACTICE ON CONFIDENCE INTERVAL

### Introduction

The purpose of this practical is to illustrate the meaning of confidence intervals as ways of estimating the population parameter, i.e., to obtain information about the population from the empirical data collected in the sample. The practical will examine the various uses of confidence intervals and look at a number of incorrect conceptions of their use.

The practical will focus on the “mean” estimator, one of the estimators for which the sampling distribution is known, provided that certain conditions are met. It should be noted that the characteristics of the sampling distribution arise from the information used to construct the confidence interval around the sample mean. The practical also looks at overlaps, the possibility of constructing a confidence interval to estimate the difference of means between two populations, and the usefulness of this type of interval.

### Tasks

1. Download the ESCI program (Exploratory Software for Confidence Intervals; [Cumming, 2012](#)) from [www.thenewstatistics.com](http://www.thenewstatistics.com). This can be used with Excel 2003, 2007 or 2010: there are different versions for operating systems with 32 and 64bits.
2. The page [CIjumping](#) of the file ESCI chapters1-4 contains a representation of the population from which samples can be extracted and the confidence intervals for each of the samples can be constructed. We will focus on the personality trait called “Extraversion” measured according to the NEO-FFI inventory ([Costa & McCrae, 2002](#)): in the adult reference group mixing men and women, the mean value is 32.56 with standard deviation equal to 6.24. For this exercise these values (rounded off) are taken as population parameters. Draw a sample of 20 people; repeat the process 36 times: a) see how the amplitudes of the confidence interval vary according to the dispersion in each sample; b) check each interval for the inclusion of the population parameter; c) check what percentage of the intervals contains the population parameter; compare this percentage in the case of extracting 10 samples or 300 samples.

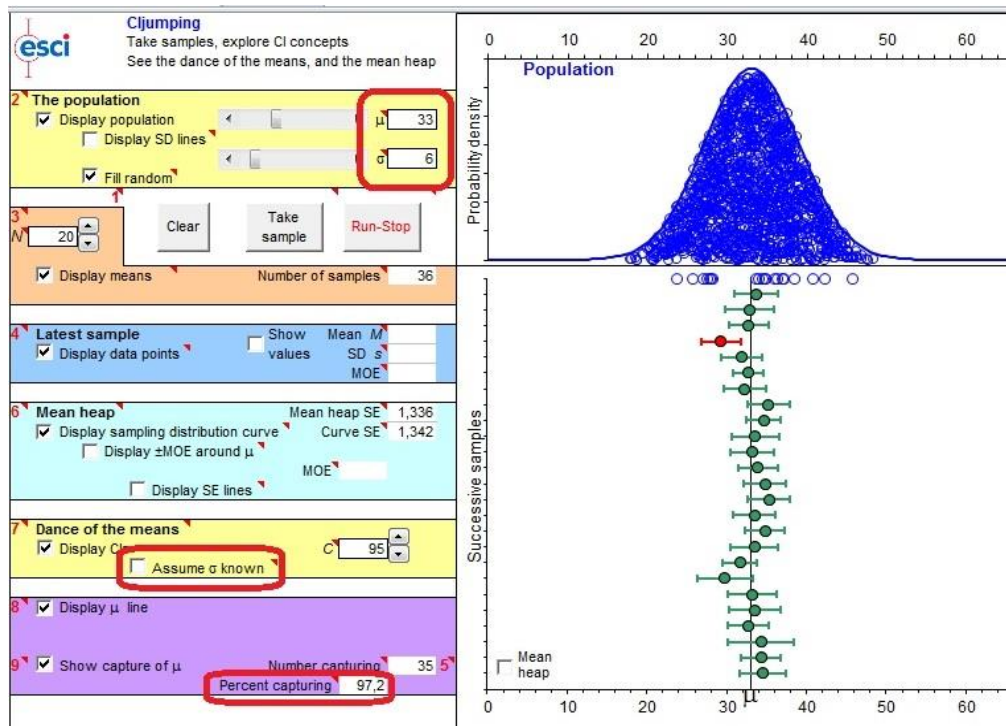


Figure 1: Results obtained from 36 samples.

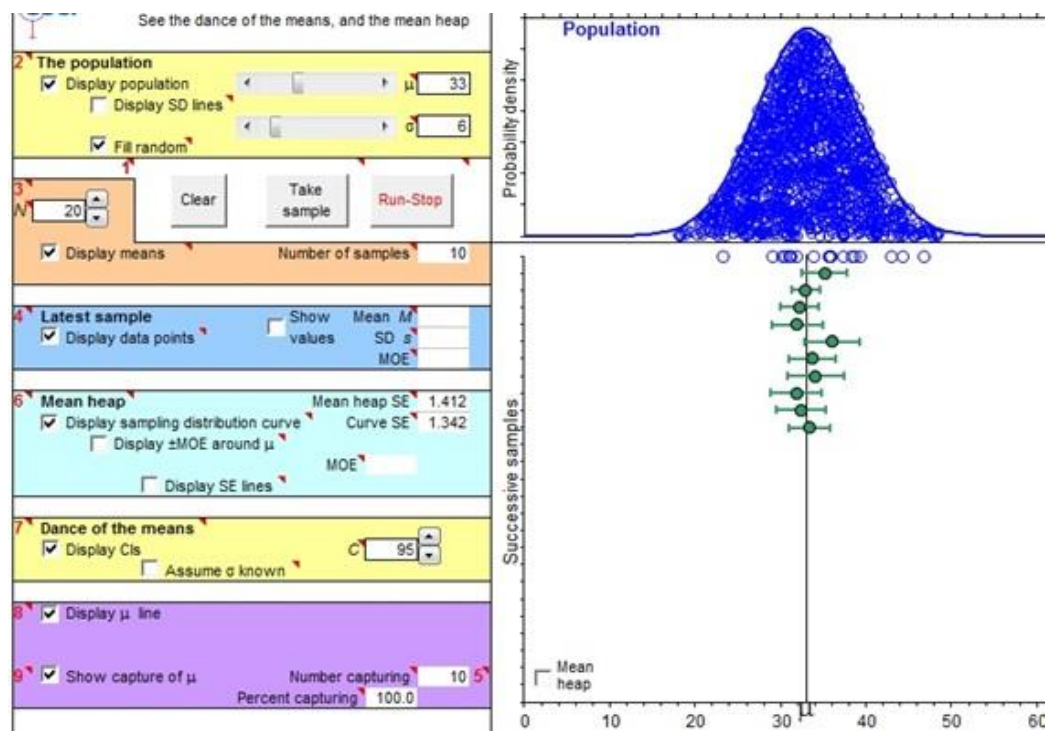


Figure 2: Results obtained from 10 samples.



**Problems to solve**

1. What would happen to the standard error of the mean statistic if the variability ( $\sigma = 12$ ) in the population were twice as great? What would happen to the confidence intervals?
2. What would happen to the standard error of the mean statistic if the size of the sample extracted from the population were twice as great (i.e., 40 individuals), with  $\sigma = 12$ ? Do you find that the variability would also double? What should be done to return to the same precision as when  $\sigma = 6$ ?
3. Between what values does the population parameter from the results of task 2 vary?
4. If another sample of the same size and from the same population is randomly drawn, do you expect the confidence interval constructed to be the same or different?
5. What indicates the (semi-)longitude of the confidence interval? What is preferable for the statistic: narrow or wide intervals, closer or farther limits?
6. What is the probability that a confidence interval constructed with a confidence level of 95% contains the population parameter?
7. Imagine that for a town of 6000 inhabitants the income of each person is known and the mean of the population is equal to 800 euros. It is known that if you take many random samples of 50 people from this population and calculate the average income for each of the samples, the average of these means will be 800. Also, it is expected that if for each sample the confidence interval ( $1 - \alpha = 0.95$ ) is calculated for the mean, 95% of intervals will contain the value 800. Now imagine that the participants of the samples are in fact inhabitants of the same residential area where the researcher lives in this town and that they participate in response to the researcher's invitation because they live nearby. Will the samples selected this way be random? What effect will this kind of sample have on the sample means? How will it affect the confidence intervals?
8. Imagine that two confidence intervals have been constructed ( $1 - \alpha = 0.95$ ) to estimate the population mean for the scores of depression measured according to the BDI (Beck, Steer, & Brown, 1996): one for the population of people diagnosed with cyclothymia and another for people diagnosed with major depression. The first confidence interval lies between the values 13.5 and 21, while the second interval has a lower limit of 23.3 and an upper limit of 39.1. For this exercise, the page COMPARE A B will be used from the Excel file called ESCI chapters 5-6 Jul 4 2011:

Check how such intervals are obtainable by varying the size of the two samples while simultaneously taking the variability in the samples into account? What can you do to counter greater variability?

What can be done to counter the effect of being forced to select a small sample?

In subsequent sessions, we will work with statistical decision-making and the significance of an indicator called a "p-value". For now, we will stick with the idea that if this indicator takes values less than or equal to 0.05, it can be considered that a zero difference is possible between the means for the populations (e.g., for people

with cyclothymia vs. people diagnosed with major depression). Consider that the confidence interval of the population of people with cyclothymia suggests that, in this population, the possible values for the parameter “average score in the BDI” are between 13.5 and 21. Similarly, the possible values for the same parameter for the population of people with major depression are between 23.3 and 39.1 points. Therefore the possible values are not shared. Does this mean that it is possible for the difference of the means for the two populations to be zero? Stated another way, does it mean that the p-value will take a value less than or equal to 0.05? Check with the same Excel file page.

Consider now that the interval for the population of people with cyclothymia is compared with an interval constructed from a sample of people with dysthymia with a lower limit of 19.7 and upper limit of 26.4. Do the possible values share the parameter in both populations? What does the p value indicate? Interpret the information.

9. Continuing with the population values of Extraversion ( $\mu = 33$ ,  $\sigma = 6$ ), what percentage of subsequent sample means will be contained within the confidence interval constructed from an initial sample?

Use the CI page and copy from the Excel file called ESCI 64-bit chapters 5-6 (for 64-bit operating systems) or ESCI chapters 5-6 (for all other operating systems). Specify the population parameters and the size of the samples to be extracted.

Extract 20 samples and observe for each of them the percentage of sample means that are contained in the interval constructed from the data of the first sample.

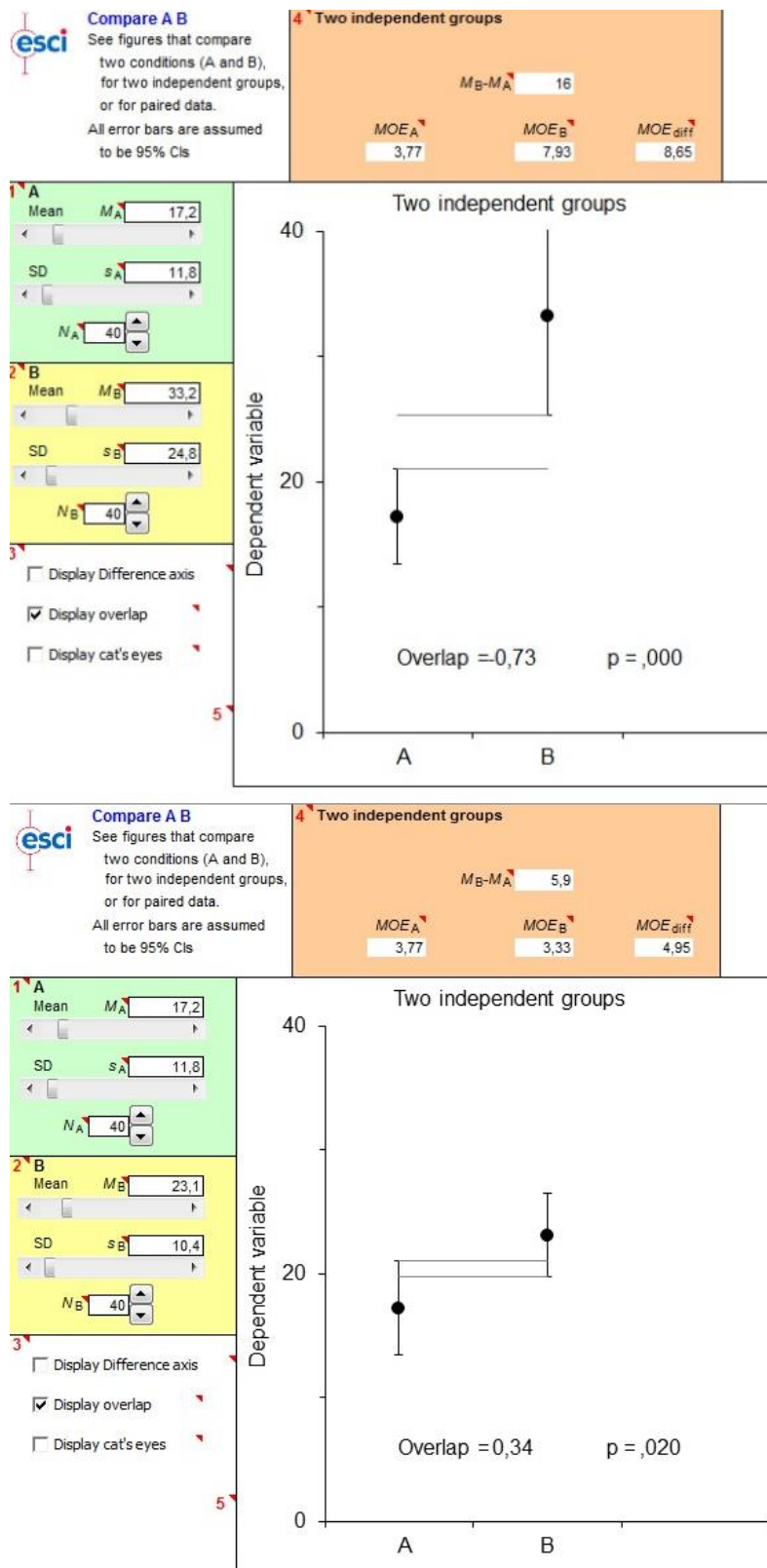


Figure 3: Confidence intervals for various mean comparisons.

**Additional references**

- Beck, A. T., Steer, R. A. & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Costa, P. T., Jr. & McCrae, R. R. (2002). *Manual NEO PI-R*. Madrid: TEA Ediciones.
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. London, UK: Routledge.

[Back](#)

## PRACTICE ON ONE-SAMPLE TESTS

### Introduction

The Big Five Model has become a benchmark for the measurement of personality (Sanz & García-Vera, 2009). According to this model, personality is characterised by five dimensions, namely: neuroticism, extraversion, openness to experience, agreeableness and conscientiousness. Costa and McCrae developed the NEO-PI questionnaire to assess these five personality dimensions, which incorporate a cluster of specific traits. Since then several revisions and versions have been made. Also, both the NEO-PI and its several versions have been adapted to the Spanish system (Costa and McCrae, 2002) using scales and standard groups. In this regard, the representation of such scales has recently been put into question (Sanz and García-Vera, 2009), because although they were made with a large sample size, they were obtained from different personnel selection processes.

The above study states that the scales in the Spanish adaptations are suitable for selection processes in which it is assumed that candidates provide a more positive picture of themselves than is actually the case, i.e., the so-called social desirability effect exists. However, these standard groups may be less useful in other contexts, for example, in studies of groups of volunteers. To test this assumption, the results obtained for the scales of the Spanish adaptation of the NEO-PI were compared with the results obtained from a sample of 682 volunteers. This comparison confirmed that the scales included in the commercial version of the NEO-PI were not suitable for evaluation contexts different from its own selection process. One of the versions of the NEO-PI is an abbreviated form, called the NEO-FFI questionnaire, by which the five dimensions can be assessed quickly (completed in 10-15 minutes). Specifically, the abbreviated version reduces the 181 original items to 60. In the particular case of the scales of the Spanish adaptation of the NEO-FFI (commercial version), a sample of 5847 candidates for various positions in a hospital (technical specialists, staff nurses, management and administration) was used. As a result, it is possible that the scales are not representative and, therefore, of little use in the process of personality evaluation, in which participants, like volunteers, broadly respond honestly and do not have an interest in distorting their image. Therefore, the aim of this practical is to check whether the scales of the NEO-FFI are useful to represent a sample of college students, specifically psychology students. For this purpose, the one-sample test will be used.

## Tasks

1. In a preliminary practical session, each student will self-administer the NEO-FFI questionnaire (maximum 10-15 minutes), filling in the answers to all the items including the information on sex and age.
2. Once administered, each student will correct their questionnaire with the template provided. In this way, the scores for each of the five scales are obtained.
3. Completed and corrected questionnaires will be handed over to the teachers.
4. In a subsequent practical session, a database with all the results for the five dimensions of the NEO-FFI will be provided in order to answer the questions listed below.

## Questions

1. Complete the empty column in the table below with the information obtained from the student responses included in the database.

	Applicants	Students
Sample size	5847	
N Scale	15.63 (7.01)	
E Scale	32.56 (6.24)	
O Scale	29.44 (6.29)	
A Scale	33.13 (5.66)	
C Scale	36.42 (5.64)	

Table 1: Table with comparison of results obtained from the sample for the Spanish adaptation of NEO-FFI and from the sample of students: sample size, mean and standard deviation (in parentheses) on the five scales.

- Once the table is completed, observe the mean values and the indicators of dispersion for each of the scales. If you compare both samples at the descriptive level, what can you conclude about the results in terms of central tendency and variability? Explain whether these results are expected.
2. Assuming an effect similar to the one found in the study by [Sanz and García-Vera \(2009\)](#), formulate the statistical hypothesis for each of the dimensions of the NEO-FFI for this sample.
  3. Perform the one-sample tests using R-commander on the data from the student sample. Take a decision for each of the one-sample tests made. What can be concluded from these results in relation to the usefulness of the scales?
  4. The scores for the five NEO-FFI scales are considered to be distributed according to a normal distribution in the general population. This fact implies that the random variables are associated with the following moments of order 3 and 4, respectively:  $C.As.(X) = \gamma_1 = 0$  y  $C.Ap.(X) = \gamma_2 = 0$ . Perform the one-sample tests for each of the scales on the questionnaire. Based on the results obtained, can it be assumed that the distribution of the scores is normal? Provide a reason for your answer.

**Additional references**

- Costa, P. & McCrae, R. (2002). *Manual NEO PI-R* (2nd ed.). Madrid: TEA Ediciones.
- Sanz, J. & García-Vera, M. (2009). Nuevos baremos para la adaptación española del inventario de personalidad NEO revisado (NEO PI-R): Fiabilidad y datos normativos en voluntarios de la población general. *Clínica y Salud*, 20 (2), 131–144.

[Back](#)

## PRACTICE ON THE ASSOCIATION BETWEEN CATEGORICAL VARIABLES

### Introduction

Following Guerrero, Ávila and Miranda (2008), magical beliefs such as paranormal and extraordinary phenomena, superstition and magical thinking are explanations that contradict the laws of nature and lack scientific support. These researchers have provided some evidence for the association between sociodemographic variables and magical beliefs. In addition to these sociodemographic characteristics, an association has been found between some property of personality, such as neuroticism level, and superstitious beliefs (Wiseman & Watt, 2004).

The present practical examines whether there is evidence for the existence of an association between some of the sociodemographic variables and beliefs generally considered as paranormal or events commonly accepted as highly unlikely. Unlike previous studies, both the sociodemographic characteristics and the answers given by the respondents in relation to their beliefs are treated as categorical variables.

### Questionnaire on beliefs: description and administration

Table 1 consists of a questionnaire on paranormal or low-probability beliefs. The questionnaire includes only ten statements about beliefs that, as noted above, have not been empirically verified and therefore cannot be accepted as facts. Only one of two responses can be ticked for each statement, that is, total agreement or total opposition. Although respondents may have difficulty stating an opinion clearly leaning towards one end or the other, you should suggest that they choose one or the other response option. In fact, it may be suggested to those who reply that they do not totally agree with the statement that they are opposed to it. Each student should contribute ten questionnaires, ensuring that respondents do not all fit the same profile, i.e., their sex, age and educational level should vary.

This diversity of respondents is necessary to carry out a statistical analysis that is minimally representative of the various profiles based on the demographic variables. Minimum intervention is to be made when people respond to the questionnaire, except that, as already mentioned, they should give one of the two answers or, if this is difficult, express a leaning toward one or the other.

Also, participants should be told that there are no right or wrong answers because it is a matter of opinion. Ask for honest responses. Keep in mind that the person administering the questionnaires should not bias the responses of participants. In order to avoid any impact on participant responses, the role should be limited to setting the task.



Sex: Male Female    Age:		
Education: EGB ESO    Bachillerato FP CF    University degree		
Statements	Yes	No
Spacecraft have come from other planets in the terrestrial biosphere		
It is beyond doubt that some people have experienced astral travel		
There are people known as `future seers´ who are able to visualise the future of others		
A divine being designed and created the universe		
The existence of premonitory dreams in some people is unquestionable		
Some living humans can communicate with the dead		
The position of the stars at the time of birth partly determines people's personalities		
There are people who are able to read the thoughts of others through telepathy		
EVP voices are truly recordings of deceased people		
It is true that some people can levitate		

Table 1: Apart from providing the personal data, respondents must state whether they agree or disagree with each of the ten statements. They are suggested to use a cross to indicate whether they agree or disagree with each statement.

## Tasks

The practical is done according to the following steps:

1. How should the information be organised in order to analyse the possible association between responses to the statements and the respondents' sex or level of education?
2. Draw a contingency table in which the rows indicate the three general levels of education, that is, basic (primary, EGB, ESO), intermediate (high school, professional training or vocational training) or advanced (diploma, bachelor's degree, master's or doctoral studies), while the columns specify the two possible responses for each of the statements. Later, for each cell, specify the observed frequency.
3. State the null hypothesis for the specific test, i.e., for both categorical variables considered in the statistical analysis.
4. Next, calculate the row and column marginals and determine the total for the sample and the corresponding percentages.
5. For each and every one of the boxes in the contingency table, obtain the expected frequencies (also known as theoretical frequencies).
6. Analyse the condition required for the validity of the statistical test of the decision on the independence of two categorical variables.
7. Next, calculate the chi-square statistic and an index of association, such as Cramér's V.

8. Finally, take a decision on the null hypothesis and conclude to reject or not. Then interpret the results obtained from a substantive point of view.

Once the above process has been completed for one of the statements, it can be repeated for any of the remaining nine, even for the whole. You can also perform the analysis for other sociodemographic variables and the interaction tables can be illustrated. Regarding the latter, any of the possible contingency tables can be taken (i.e., crossing the level of education and the response associated with one of the statements of the questionnaire), but now two contingency tables will be prepared, one exclusively for females and another exclusively for males. What interest do the interaction tables have? Perhaps, if females and males are mixed in the same contingency table, there might not be any association between the two categorical variables of interest, i.e., the level or grade of education and the response associated with one of the statements. However, when analysing the association separately for females and males, there may be a relationship between level of education and the response associated within only one of the subpopulations. Reaching different conclusions based on whether the data are aggregated or disaggregated is known as Simpson's paradox.

### Questions

9. Indicate an example to use the test of independence between two categorical variables in the field of Psychology.
10. What do the interaction tables suggest?
3. What have you learned about the association between paranormal beliefs and sociodemographic variables?
4. If you have not included the age variable in the analysis, why has it been taken into account in the study? What type of validity does it contribute?

### Additional references

- Guerrero, C., Ávila, R., & Miranda, P. (2008). La correlación entre creencias mágicas y variables sociodemográficas. *Psicología y Ciencia Social*, 10, 5-15.
- Wiseman, R., & Watt, C. (2004). *Measuring superstitious belief: Why lucky charms matter*. Proceedings of the 47th Annual Convention of the Parapsychological Association (pp. 291-298).

[Back](#)

## PRACTICE ON THE *t*-TEST

### Introduction

The purpose of this practical is twofold: methodological and statistical. This is because both aspects are usually present when conducting scientific research. In fact, the methodology used to plan and conduct a study is closely related to the way the information is later treated through the use of descriptive and/or inferential statistics. Additionally, these two components largely determine the scientifically defensible conclusions that can be drawn from a study.

Regarding the methodological aspect, the aim of this practical is to illustrate a way to plan an experimental study in order to obtain evidence on the causal relationship between an independent variable and a dependent variable. In this case, the independent variable is defined as the type of information received on immigration: the variable has two levels, which are carried out by means of two informational videos labelled “A” and “B”. The assignment of participants to each of the two levels (i.e., for each experimental condition) is done randomly. Therefore there is a grouping variable (also known as a “between subjects” variable), resulting from the definition of each group in terms of the condition through which it passes. Because of this feature and the way the two groups are formed, the design could be called a “two-group randomised design”.

The dependent variable is the assessment made by participants (i.e., the students) on immigration. This assessment is performed using a visual analogue scale in which students indicate their views on a spectrum ranging from the maximum negative assessment to the maximum positive one. Participants must provide this assessment both before and after being exposed to different types of information about immigration. This provides a repeated measures variable (also called an intra-subject variable) relating to the time of measurement, which has two levels: “pre” and “post”. Given an independent variable for the cluster and the presence of other repeated measures, the design used for the study can be called a “mixed factorial design” (for more information, see [Ato and Vallejo, 2007](#) in Spanish and [Keppel, 1991](#), in English).

The statistical objective of this practical is to apply two versions of the *t*-test. One of these versions is called the “*t*-test for independent samples”. This test is applied to designs in which data are collected from independent groups in order to compare the mean parameter of the variable of interest in the two populations. Alternatively, this test can be conceptualised as a test of association between a dichotomous qualitative variable (type of information provided: video A and video B) and a quantitative variable (assessment of immigration). Another version is called the “*t*-test for related samples or paired samples”. This test is applied to designs in which the data are collected from a group measured twice or from two groups, the members of which have been paired as a function of some variable of possible confusion (i.e., a variable that can affect response and be controlled to increase confidence in the causal relationship between the main variables of interest). As the situation

in the present practical has been described, this second version of the test can be seen as a test of association between a dichotomous qualitative variable (time of measurement: before and after receiving the information) and a qualitative variable (assessment of immigration). The practical will show the use of t-tests in order to illustrate the type of information they provide and also their limitations.

The combined methodological and statistical objective is to reach a conclusion on whether the type of information provided can relate unambiguously and causally with the assessment of immigration. The illustration shows how the information obtained through descriptive and inferential statistics can be used for the abovementioned purpose.

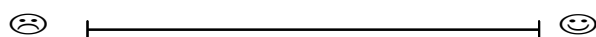
### Tasks: Data collection

- o. Ideally: select participants/students from a list of all students for the first and second semesters.
1. Select randomly from the whole sample (with size  $n$ ) those participants ( $n/2$ ) who will be part of group A. This selection can be done using the following example, which is for a group of 80 people. Note that sampling without replacement must be used (the option `replace` is `FALSE`):

```
> n <- 80
> sample(1:n, (n/2), replace=FALSE)
```

2. Collect the pre-measurement data:

Item: Indicate the extent to which you are against (☹) or in favour of (☺) immigration by putting a vertical line on the spectrum below.



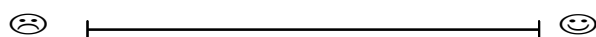
3. Manage the two types of information:

Group A watches video: <http://www.youtube.com/watch?v=Ef2KiU-r6vo>.

Group B watches video: <http://www.youtube.com/watch?v=OWBH-C-LGq4>.

4. Collect the post-measurement data:

Item: Indicate the extent to which you are against (☹) or in favour of (☺) immigration by putting a vertical line on the spectrum below.



5. Construct the data matrix with the following variables: ID of the participant, experimental condition, pre-measurement and post-measurement.

### Questions related to the methodology

1. Why is the variable 'ID of the participant' required?
2. Why is "type of information provided" presented as an active independent variable and not assigned as a predictor or explanatory or independent variable?
3. In order to maintain that the type of video is related to the assessment of immigration, what comparisons have to be made and what results are expected? Formulate the appropriate null hypothesis, based on whether these must be directional or non-directional, and indicate for each of them if the desired result is that the hypothesis is rejected or not.

### Questions related with statistics

4. Obtain the mean difference: descriptive value. How likely is it that the true mean difference in the population will be 0?
5. Discuss the meaning of the formulas of t.
6. Interpret the p value: a) discuss the difference between a directional and a non-direction (one or two-tailed) hypothesis; b) interpret the risk of being wrong in rejecting the null hypothesis and provide another interpretation that can be made.
7. Discuss what it implies if the null hypothesis is not rejected.
8. Discuss what it implies if the null hypothesis is rejected.
9. For each comparison of independent samples, obtain the range of plausible values for the estimated population parameter using the page Data two from the Excel file named ESCI chapters 5-6 or ESCI 64bit chapters 5-6 (depending on the features of your operating system). Introducing the data for the two samples, obtain the following information: a) the confidence intervals for each sample; b) a graphical representation (including overlapping confidence intervals); and c) the level of significance for comparisons of independent groups. Interpret the results.

Note that the results concerning the statistical decision may be obtained through **R-commander**: *Statistics* → *Means* → *Independent samples t-test*. Additionally, the graphical representation of the confidence intervals can be obtained through **R-Commander**: *Graphs* → *Plot of means*: select the option "Confidence intervals".

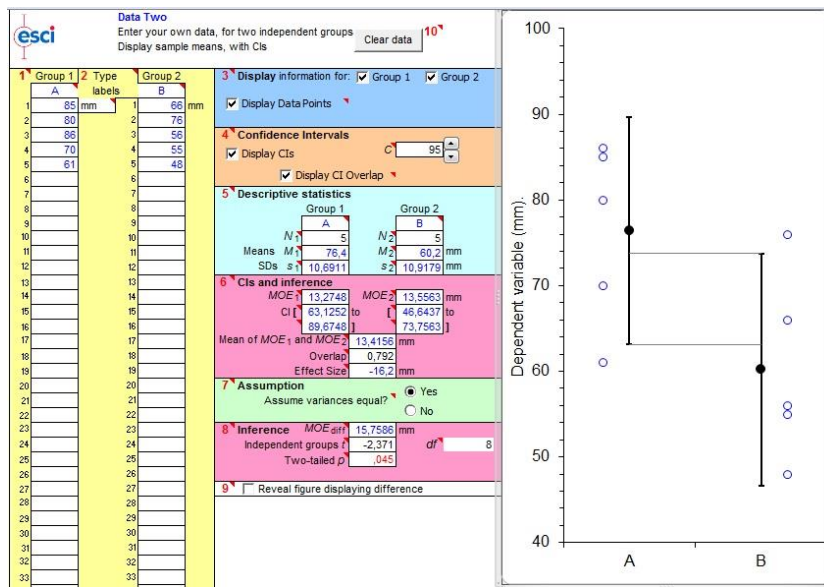


Figure 1: Example: Comparison of post-measurement data (visual analogue scale in mm) between the two groups.

- For each comparison of paired samples, obtain the range of plausible values for the population parameter estimated using the page **Data paired** from the Excel file named **ESCI chapters 5-6** or **ESCI 64bit chapters 5-6** (depending on the features of your operating system). Introducing the data for the two samples, obtain the following information: a) the confidence intervals from each sample and its graphical representation (including overlapping confidence intervals); b) the confidence interval for the mean differences and their graphical representation; and c) the level of significance for comparisons of paired groups. Interpret the results.

Note that the results concerning the statistical decision may be obtained through **R-commander: R-Commander: Statistics → Means → Paired t-test**.

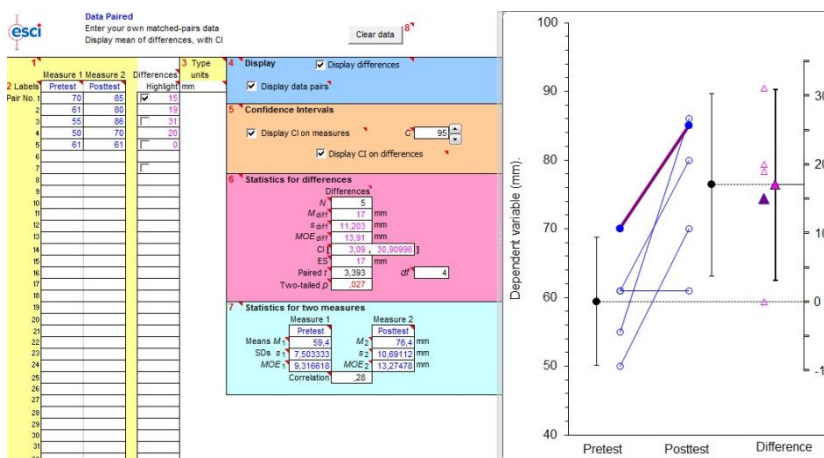


Figure 2: Example: Comparison of pre- and post-measurements (visual analogue scale in mm) for the group that watched the video aimed at influencing perceptions.

8. Discuss the effect size: the suggestion is to use the standardised mean difference assuming that criteria are available for interpretation (see [Cohen](#), 1992).
9. Assess whether evidence has been obtained on the causal relationship between the type of information received and the assessment of immigration.
10. Discuss what population can be extrapolated from the results, taking into account how the participants were selected in this study.
11. Discuss whether the extension of the extrapolation changes, bearing in mind that the same type of study was carried out in all groups registered for the course “Statistics” taught at the Faculty of Psychology.

### **Additional references**

- Ato, M. & Vallejo G. (2007). *Diseños experimentales en psicología*. Madrid: Pirámide. (Chapter 10).
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall. (Chapter 17).

[Back](#)

## PRACTICE ON NON-PARAMETRIC TESTS

### Introduction

Performance in subjects in Statistics and Methodology can be explained in part by how the content is taught and by the areas on which the teacher places greater emphasis. Therefore, to increase academic achievement, we have pursued initiatives such as redesigning the subjects in search of real benchmarks to enhance student understanding of the more abstract content and devise teaching strategies for problem-solving.

However, although effective performance is partly due to teacher performance, there is no denying the fact that academic performance, not only in Statistics but in any other subject, can largely be explained by student motivational aspects. Within this broad field, studies have investigated the effect of so-called self-efficacy on academic performance, that is, the perception that each person has of his/her ability to achieve a specific objective (see [Bandura, 1977](#)). According to the latest research, the initiation of specific goal-directed conducts (e.g., to learn Statistics) and the persistence of coping strategies in the face of potential problems encountered in the attainment of this objective are explained by the perceptions of students in relation to their ability to achieve the above goal.

In a relatively recent study ([Finney & Schraw, 2003](#)), two instruments were developed to measure self-efficacy in Statistics: a test of self-efficacy in Statistics (ESSE, to assess perceived competence to solve specific statistical problems) and a test of self-efficacy in learning (SELS, to assess perceived competence to learn the typical content of a Statistics course). In this practical, we will evaluate self-efficacy in Statistics (through ESSE) among a sample of students in the Statistics course. To do so, we will follow several steps as the course progresses. Our assumptions will be as follows:

1. Negative or similar experiences in the subjects (being required to repeat the subject of Statistics or having failed a Research techniques subject) are associated with lower perceived competence in students before they even start the course.
2. Students who prepare the subject (attend classes and steadily work through the content) increase their perception of self-efficacy.

As an additional assumption, we will consider that the conditions needed to apply parametric tests of mean comparisons are not met and that, therefore, we should use non-parametric tests throughout the practical.



## Development of the practice

In order to test the initial assumptions, it is necessary to follow these steps:

1. On the first day of class, the questionnaire found in Table 1 will be handed out. It is important that students identify themselves with their NIUB because an identifier is needed to match subsequent measurements. The CSSE is composed of 14 items assessing the confidence shown by people in their ability to solve a set of tasks in Statistics. The items are answered using a graded scale with 6 responses ranging from no confidence (1) to total confidence (6). For each item, the respondent should mark the answer that best matches the current state of his/her ability. It is important to answer all the items to get overall scores on the questionnaire.
2. Once the data are collected, we will proceed to encode the database for later statistical analysis. In this respect, you can create text files (e.g., txt file), calculation sheets (e.g., Excel files) or other types of files that include the collected data. After creating the data file, we can obtain the overall scores for each of the participants (the sum of all items). Additionally, a new variable will be created to classify participants as having or not having negative experiences related to the subject of Statistics or Research techniques (having failed Research techniques or been required to repeat the Statistics course). These variables can be obtained using software such as R-Commander that already incorporates routines for the recodification and creation of variables.
3. The questionnaire in Table 2 will be handed out in a subsequent session. Note that this second questionnaire collects information to track the subject of Statistics, in addition to the NIUB and the CSSE items.
4. Subsequent measurements must match the participant identifiers for all this information to be added to the database.
5. Once the coding is done, we will proceed to perform the relevant statistical analysis needed to answer the questions listed in the following section.

## Tasks

1. Formulate the statistical hypothesis for the first assumption, justifying the variables of interest to test this hypothesis.
2. Justify the statistical test of choice for testing the statistical hypothesis formulated in the previous section. Take into account the initial assumption that the conditions needed to apply a parametric test of mean comparison are not met.
3. Interpret the results of the statistical test: value of test statistic, degree of statistical significance, etc.
4. Formulate the statistical hypothesis for the second assumption, justifying the variables of interest to test this hypothesis.
5. Justify the statistical test of choice to test the statistical hypothesis formulated in the previous section. Take into account the initial assumptions that the conditions needed to apply a parametric test of mean comparison are not met.

6. Interpret the results of the statistical test: value of test statistic, degree of statistical significance, etc.
7. Discuss the general conclusions coming from the two statistical tests performed.
8. Finally, graphically represent the joint distribution of the two variables of interest for each of the statistical tests performed. Does the graphical representation confirm what was concluded from the statistical tests?

### **Additional references**

- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Finney, S. J. & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28, 161-186.

[Back](#)

NIUB:						
Repeated Statistics:	YES	NO				
Failed Research Techniques:	YES	NO				
1. Identify the measurement scale of a variable	1	2	3	4	5	6
2. Interpret the p-value of a test statistic	1	2	3	4	5	6
3. Identify whether a distribution is asymmetric from the 3 measures of central tendency	1	2	3	4	5	6
4. Select the appropriate statistical procedure to answer a research question	1	2	3	4	5	6
5. Interpret the results of a statistical procedure in terms of the research question	1	2	3	4	5	6
6. Identify factors influencing the power of a test statistic	1	2	3	4	5	6
7. Explain what is meant by the standard deviation of the variable measured	1	2	3	4	5	6
8. Differentiate between type I error and type II error in the hypothesis test	1	2	3	4	5	6
9. Explain what measures the standard error	1	2	3	4	5	6
10. Distinguish between the purpose of a descriptive and of an inferential procedure	1	2	3	4	5	6
11. Distinguish the differential information provided by the 3 measures of central tendency	1	2	3	4	5	6
12. Identify the differences between a parameter and a statistic	1	2	3	4	5	6
13. Identify the mean, median and mode as indicators of central tendency	1	2	3	4	5	6
14. Explain the difference between a sample distribution and the distribution of a variable in the population	1	2	3	4	5	6

Table 1: Shows some variables of interest for the first data collection of the practical, followed by the 14 items of the ESSE test evaluating self-efficacy in relation to statistical skills.

---

NIUB:  
I have attended class and steadily worked on the subject of statistics constantly up to today: YES NO

---

1. Identify the measurement scale of a variable	1	2	3	4	5	6
2. Interpret the p-value of a test statistic	1	2	3	4	5	6
3. Identify whether a distribution is asymmetric from the 3 measures of central tendency	1	2	3	4	5	6
4. Select the appropriate statistical procedure to answer a research question	1	2	3	4	5	6
5. Interpret the results of a statistical procedure in terms of the research question	1	2	3	4	5	6
6. Identify factors influencing the power of a test statistic	1	2	3	4	5	6
7. Explain what is meant by the standard deviation of the variable measured	1	2	3	4	5	6
8. Differentiate between type I error and type II error in the hypothesis test	1	2	3	4	5	6
9. Explain what measures the standard error	1	2	3	4	5	6
10. Distinguish between the purpose of a descriptive and of an inferential procedure	1	2	3	4	5	6
11. Distinguish the differential information provided by the 3 measures of central tendency	1	2	3	4	5	6
12. Identify the differences between a parameter and a statistic	1	2	3	4	5	6
13. Identify the mean, median and mode as indicators of central tendency	1	2	3	4	5	6
14. Explain the difference between a sample distribution and the distribution of a variable in the population	1	2	3	4	5	6

---

Table 2: Shows some variables of interest for the second data collection of the practical, followed by the 14 items of the ESSE test evaluating self-efficacy in relation to statistical skills.

## PRACTICE ON THE CORRELATION BETWEEN THE BIG FIVE FACTORS

### Introduction

The NEO PI-R (Costa and MacCrae, 2002) is a psychological test developed to measure a set of personality traits known as the Big Five. In the NEO PI-R there are five broad scales corresponding to five personality factors: neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A) and conscientiousness (C). The NEO PI-R also obtains measures for the various facets and elements constituting the abovementioned factors. Table 1 lists the various facets for the factors, so as to better understand what each dimension represents.

The factor called “neuroticism” is bipolar and contrasts balance and emotional stability against imbalance or neuroticism. The “extraversion” dimension corresponds to sociability, a taste for excitement and stimulation. At the other extreme, the introvert, who is not the opposite of the extrovert, is reserved, independent and consistent. The characteristic “openness to experience” refers to the interest of people in the outside world as well as in their interior world. Openness is related to divergent thinking and creativity. The “agreeableness” factor is associated to, at one end, altruistic and cooperative people, while the other end refers to egocentric, suspicious people and people opposed to cooperation. As for the “conscientiousness” dimension, it is associated at one of its poles with the active process of planning, organising and performing tasks. Conscientious people are scrupulous, punctual and reliable. At the opposite pole, people are more hedonistic and less constant in achieving objectives. For a definition of the facets of these five factors, see the manual on psychological testing (Costa and McCrae, 2002).

The NEO PI-R is a psychological test that can be administered individually or collectively. Most people require between 30 and 40 minutes to respond to all the statements. The maximum time needed for the test, however, is not available. Therefore, we will use a reduced version, the NEO-FFI, for this practical. The reduced version does not obtain measurements for the facets, but it does obtain measurements for the broad factors of neuroticism, extraversion, openness to experience, conscientiousness and agreeableness. There is no time limit for answering the reduced version, but most people will require between 10 to 15 minutes to respond to all the statements.

Similar to the NEO PI-R, the NEO-FFI can be administered individually or collectively.

Facets of neuroticism
N1 Anxiety
N2 Hostility
N3 Depression
N4 Self-consciousness
N5 Impulsiveness
N6 Vulnerability
Facets of extraversion
E1 Warmth
E2 Gregariousness
E3 Assertiveness
E4 Activity
E5 Excitement seeking
E6 Positive emotions
Facets of openness
O1 Fantasy
O2 Aesthetics
O3 Feelings
O4 Action
O5 Ideas
O6 Values
Facets of agreeableness
A1 Trust
A2 Straightforwardness
A3 Altruism
A4 Compliance
A5 Modesty
A6 Tender-hearted
Facets of conscientiousness
C1 Competence
C2 Order
C3 Dutifulness
C4 Achievement striving
C5 Self-discipline
C6 Deliberation

Table 1: Facets or constituents corresponding to the different factors of the NEO PI-R.

Referring to the NEO-FFI, the five factors do not seem to be independent of each other, although the intensity of the association may vary significantly according to the pairs of factors considered. Table 2 shows the values of the linear product-moment correlation coefficient for all pairs of factors corresponding to the NEO-FFI (Costa & McCrae, 2002).

	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Neuroticism	—	−0.40	−0.15	−0.32	−0.53
Extraversion	−0.40	—	0.38	0.28	0.38
Openness	−0.15	0.38	—	0.17	0.22
Agreeableness	−0.32	0.28	0.17	—	0.33
Conscientiousness	−0.53	0.38	0.22	0.33	—

Table 2: Values of the linear product-moment correlation coefficient for all pairs of factors of the NEO-FFI. Values are obtained for the general population.

### Development of the practice

Students will be given a file, for example in Excel format or R-commander, with the information corresponding to the five factors, sex and the type of studies they were enrolled in prior to entering the university. Next, the statistical analysis will be performed using a programme (e.g., R-commander). Then the following tasks must be completed:

1. Obtain the corresponding scatter plots for the ten different pairs of variables. Determine, in each case, the convenience of obtaining the linear product-moment correlation coefficient or the correlation coefficient by ranks.
2. Determine the values corresponding to the different pairs of factors and fill in the empty cells in Table 3.

	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Neuroticism	—				
Extraversion	−0.40	—			
Openness	−0.15	0.38	—		
Agreeableness	−0.32	0.28	0.17	—	
Conscientiousness	−0.53	0.38	0.22	0.33	—

Table 3: As in Table 2, the bottom triangular shows the values of the linear product-moment correlation coefficient for all pairs of NEO-FFI factors.

**Questions**

1. Are the personality dimensions independent of the Big Five model?
2. The correlations were obtained for a sample of Psychology students. Are the results obtained for Psychology students similar to those of the general population?

**Additional references**

Costa, P. T., Jr. & McCrae, R. R. (2002). *Manual NEO PI-R*. Madrid: TEA Ediciones.

[Back](#)



---

## PRACTICE ON THE INCONSISTENCIES IN CORRELATION COEFFICIENTS

### Introduction

Whether we are carrying out a descriptive analysis or making a statistical inference, it is not desirable merely to calculate values for the various correlation coefficients from a specific database. The reason for this is contrary to what one might expect. In several specific cases, the values of the various correlation coefficients may not adequately represent reality. In other words, doing a statistical analysis of the association between two variables – in this case, their correlation – is not as immediate as calculating the value of one or a few statistical indices. So what should be done? Undoubtedly, the need is to make a scatter plot, i.e., a graphical representation where the x-axis represents one of the variables and the y-axis the other one. This simple graphical representation will demonstrate if the values we have obtained for the correlation coefficients adequately represent the association between the two variables.

### Database

A variety of examples can illustrate how the values of correlation coefficients sometimes lead to a lack of representativeness of the existing association between two variables. In the context of a simple linear regression model, it may be useful to refer to correlation coefficients, but there are also databases that show the desirability of making graphical representations rather than limiting the statistical analysis exclusively to a calculation of the values of statistical indices ([Anscombe, 1973](#)). These databases are often referred to in scientific publications, on statistics courses, and in works dealing specifically with the importance of graphical representations in statistical analysis ([Tufté, 1983](#)). As a result, we will use these widespread databases to show some of the problems presented by correlation coefficients when representing the association between two variables. Table 1 contains eight variables considered by [Anscombe](#) (1973) to illustrate problems of representing a linear relationship. The eight variables are grouped into four pairs. In each pair, one variable is denoted by  $X$  and the other by  $Y$ . A total of eleven participants were involved in the study. It should also be noted that  $X_1$ ,  $X_2$  and  $X_3$  have the same values, while all the values of  $X_4$  are identical except one.

Participants	X1	Y1	X2	Y2	X3	Y3	X4	Y4
1	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
2	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
3	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
4	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
5	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
6	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
7	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
8	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
9	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
10	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
11	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Table 1: Shows the eight variables proposed by [Anscombe](#) (1973).

### Statistical analysis

Students should complete the steps below in order and examine the results, where appropriate:

1. Enter the data of Table 1 in a database or a computation spreadsheet (e.g., Excel). Once the data has been introduced, they can be imported into a statistical application. Also the data can be entered directly into the data editors available in the statistical applications (e.g., R-commander). Table 1 can be imported using the following steps. First, select Table 1 in the PDF and copy it. Second, type the prompt in the command indicator:

```
Anscombe <- read.table(file="clipboard",header=TRUE, dec=",")
```

Third, and finally, run R-commander and from active data select the database named Anscombe.

2. Using scatter plots, represent graphically the relationships between variables for X 1 - Y 1, X 2 - Y 2, X 3 - Y 3 and X 4 - Y 4. Once this task is finished, it is convenient to discuss the patterns of the observed relationships.
3. For the previous four pairs of variables, calculate the linear product-moment correlation coefficient, also known as the Pearson correlation coefficient. When the values of association are available, it is necessary to compare their values and connect them with the graphical representations obtained in the previous section.
4. Again for the four pairs of variables, obtain the correlation coefficient by ranks, also known as Spearman's correlation coefficient. Similarly, as was done for the linear product-moment correlation coefficient, it is interesting to compare the values of correlation obtained and relate them to the graphical representations shown by the scatter plots.
5. Both for the linear product-moment correlation coefficient and for the correlation coefficient by ranks, specify the null hypothesis for the test of independence between a pair of variables.

6. Both for the linear product-moment correlation coefficient and for the correlation coefficient by ranks, will the null hypothesis that states the independence between the variables for each and every one of the pairs of variables specified above, i.e., for  $X_1 - Y_1$ ,  $X_2 - Y_2$ ,  $X_3 - Y_3$  and  $X_4 - Y_4$ , be rejected or not?

### Questions

Each student must answer the following questions:

1. Why is it convenient to obtain a graphical representation before calculating a correlation coefficient?
2. What implications can be drawn from the statistical decision that the correlation coefficients may not accurately represent reality?

### Additional references

- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27, 17-21.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

[Back](#)

## PRACTICE ON THE ANALYSIS OF VARIANCE

### Introduction

What activities do you do in your free time? Because such activities are highly varied, no survey can include all the possibilities. However, it is possible to find out how most people spend their free time. At the following link, you can find the usual free-time activities that people pursue:

<http://www.csd.gob.es/csd/sociedad/encuesta-de-habitos-deportivos/encuesta-de-habitos-deportivos-2005/2-tiempo-libre-actividades-de-ocio-y-deporte>

As shown in the link, the available information refers to the percentage of people performing different activities. From the survey carried out in 2005, the ten most common activities were as follows (with the percentage of respondents who reported the activity indicated in brackets): *being with the family* (85%), *watching television* (78%), *walking* (67%), *being with friends* (65%), *listening to music* (55%), *reading books, magazines* (53%), *listening to the radio* (47%), *going shopping in shopping centres* (42%), *going to the movies* (39%), *going on outings in the countryside, excursions* (39%). However, that these activities are the most frequently made does not imply that they take up the most time. In order to collect information on this question, the present study asks not only whether the respondents do or do not do each of a series of activities, but also, in the event they do, how much time they spend on the activity. Specifically, the participants are requested to estimate the minutes per week dedicated to the activity.

As the time spent on each activity may depend on both participants' gender and level of education, this information will be requested to determine whether these sociodemographic variables are associated with the amount of time spent on various activities in free time.

### Questionnaire on free-time activities

Table 1 consists of the questionnaire covering the activities in detail in the link provided above. In addition to their sex, age and level of education, participants should indicate whether they do each of the activities or not and if so, they should provide an estimate of the amount of time spent weekly.

Sex: Male Female Age: Education: EGB/ESO Bachillerato/FP/CF University degree		
Free-time activities	Do you do the activity?	Minutes per week?
Reading books and magazines		
Playing a musical instrument		
Watching sports		
Going on an outing with my boyfriend/girlfriend or other boys/girls		
Being with the family		
Doing sports		
Going out for a drink		
Watching TV		
Going to a club or association		
Walking		
Doing nothing special		
Hanging out with friends		
Listening to music		
Listening to the radio		
Going shopping in a shopping centre		
Going to the cinema		
Going on an outing in the countryside or on excursions		
Doing handicrafts		
Going to political meetings		
Gardening		
Participating in operas and concerts		
Frequentering fast food establishments		
Digital entertainment (Browsing the internet, exchanging SMSs, ...)		
Going to the theatre		
Participating in cultural activities (lectures, presentations)		
Going dancing		
Other answers (specify):		

Table 1: Includes twenty six free-time activities. There is an additional option to specify the activity or activities that the respondent does but that do not appear on the list. For each activity, respondents should indicate whether they do the activity or not and, if they do, they should indicate an approximate time in minutes per week dedicated to that activity.

No doubt there are other questionnaires (see [Rodríguez and Agullo, 1999](#)). However, their remoteness in time means that they do not include current leisure activities. For example, the use of free time linked to the latest information technologies was not as widespread in the late

twentieth century as it is in the twenty-first century. This fact calls for the use of questionnaires that have been updated and are consistent with the present day.

### Development of the practice

The practice follows these steps:

1. The information is merged into a common database.
2. For one of the free-time activities, perform an analysis of variance separately for the factors *sex* and *education*. The hours spent on the activity will be treated as a numerical variable. The *R-commander* application may be used to obtain the decomposition table of the sum of squares, degrees of freedom and probability values associated with different values of *F*-statistics. You should also verify the conditions of application.
3. Then for both analyses, take a decision on the existence or non-existence of significant statistical effects for the factors.
4. Now perform an analysis of variance considering both factors (i.e., *sex* and *education*) simultaneously. This will allow you to decompose the interaction effect.
5. Finally, take decisions on the presence or absence of the main effects of statistical significance. Additionally, are the interaction effects statistically significant? If yes, a graph may be made to illustrate this kind of effect visually.

Once the above statistical analysis has been completed, the process will be repeated for all remaining free-time activities in order from the most frequent to least frequent activities.

### Questions

1. Taken together, is there a difference between women and men in terms of time spent on various leisure activities? Also as a whole, are there differences in leisure time in terms of respondents' level of education?
2. If significant interaction effects are identified, is there a pattern seen in the results?
3. For which activities is the relationship between sex and minutes spent intense? What about for level of education? Here, an indicator of effect size may be used.
4. What interest does the analysis of variance have in the context of experiments conducted in the laboratory?
5. Is it of interest to control a variable like age?

### Additional references

Rodríguez, J., & Agullo, E. (1999). Estilos de vida, cultura, ocio y tiempo libre de los estudiantes universitarios. *Psicothema*, 11, 245-259.

[Back](#)

## **Section 6:**

### **Modelling**

The bivariate descriptive statistics presented in Section 4 allow obtaining different kinds of information: 1) whether two variables are related or independent; 2) what is the direction of the relationship, positive or negative (for ordinal, interval and ratio scale variables), 3) whether the relationship is linear or nonlinear (for interval and ratio scale variables), 4) a quantification of the strength of association. The inferential statistical tests which are object of Section 5 can be thought of as a continuation of the descriptive use, adding one other piece of information: whether the association observed is statistically significant or not, i.e., whether it is likely that the actual situation in the population were independence between the variables. The content of Section 6 can be thus conceptualized as one of the tools available in statistics for providing additional information about the relationship between two or more variables – this information refers to the mathematical function that describes the way in which the values of one variable are related to the values of the other(s).

Therefore, modelling can be used for gaining deeper insight on the relationship between the variables, but it also makes possible predicting a response variable from its theoretically or empirically established causes. In the context of the modelling procedures, the analysts also reminded that there is virtually always part of the behaviour of interest that remains unexplained and, thus, the prediction is subjected to error, an error we would like to reduce as much as possible, without including too many (or superfluous ) variables in order to keep with scientific parsimony. Finally, in the process of modelling via regression analysis it is possible to include a third variable with different roles (i.e., as an additional predictor, a mediator, or a moderator).

We hope that the student will learn to use regression models when required by the research aim (as always), considering different modelling possibilities according to the measurement scale of the variables, and choosing predictors and dependent variable on the basis of theory or design, given that thinking and decision-making is always a task required from the analyst and never attributable to the software or the analysis.

## PRACTICE ON LINEAR REGRESSION MODELS

### Introduction

Many Statistics manuals and courses present linear regression by demonstrating the various measures of association and, more specifically, by showing the linear *Pearson product-moment correlation coefficient*, which was conceptually framed before the formal development of the measure of association. Certainly, the work of Sir Francis Galton in the late nineteenth century on *regression toward the mean* paved the way for the development of the linear correlation coefficient by Karl (see [Stanton, 2001](#)) some years later. According to Galton's findings, a characteristic (for example, height) may be inherited by offspring but it will most likely be manifested averagely in comparison to their progenitors. That is, the characteristic will be less extreme. Since then linear regression models and measures based on linear correlation have been used in numerous scientific disciplines. Moreover, these techniques have been the basis of other complex statistical procedures ([Wright & London, 2009](#)).

In the present practical, we are going to work with the key aspects to consider when formalising the relationship between two numerical variables\* by means of a simple linear regression model. In summary, the fundamental aspects are:

1. Specification of a simple linear regression model.
2. Estimation of coefficient models.
3. Validation of the simple linear regression model.
4. Diagnostic testing of the model.
5. Prediction using a simple linear regression model.

Additionally, a multiple regression model may be created and some of the fundamental aspects mentioned in the previous paragraph can be explored.

---

\*Also, the functional relationship can be modelled between one numerical variable and one categorical variable, the latter coded as a fictitious binary variable or *dummy* variable.



## Database

The package `car` (Fox & Weisberg, 2011) includes a database called `Ginzberg` with 82 psychiatric patients. The variables included in this database are `depression`, corresponding to the degree of depression obtained by means of the Beck scale (Beck, Steer, & Brown, 1996); `fatalism`, including a score obtained on a scale of fatalistic thoughts, and `simplicity`, corresponding to the extent of the person's need to classify the world in dichotomous terms (good/bad, black/white, for/against, etc.). The database also contains three variables `adjsimp`, `adjfatal` and `adjdep`, which will not be used in this practical.

## Statistical analysis

To carry out this practical, first estimate the simple regression model to determine up to what point the level of depression can be predicted from the extent of a person's dichotomous thoughts. For this, each student must:

1. Load the database in a working session of *R-commander*, for example, through the *Data* module of the GUI, or via the following code:

```
library(car)
data(Ginzberg)
```

Once executed, you can activate the database directly in *R-commander*.

2. Graphically represent the joint distribution of the variables corresponding to the level of depression and the degree of dichotomous thoughts. Take into account that the axes of the graph should have the variables in the functional relationship that is to be modelled.
3. Specify the linear regression model that formalises the functional relationship between the two variables of interest.
4. Estimate the coefficients of the model and interpret each of them.
5. Validate the simple linear regression model using an F-test indicating the value of the statistic and the degree of significance.
6. Perform some diagnostic tests to check if the assumptions of the model are met.

After studying the previous model, create a new model from which the depression score can be predicted from a person's degree of fatalism. For this, again perform the following tasks:

1. Graphically represent the joint distribution of the variables corresponding to the level of depression and the degree of fatalism. Take into account that the axes of the graph should have the variables of the functional relationship that is to be modelled.

2. Specify the linear regression model that formalises the functional relationship between the two variables of interest.
3. Estimate the coefficients of the model and interpret each of them.
4. Validate the simple linear regression model using an F-test indicating the value of the statistic and the degree of significance.
5. Perform some diagnostic tests to check if the assumptions of the model are met.

Finally, estimate the multiple linear regression models that enable us to predict the level of depression from the level of fatalism and dichotomous thoughts.

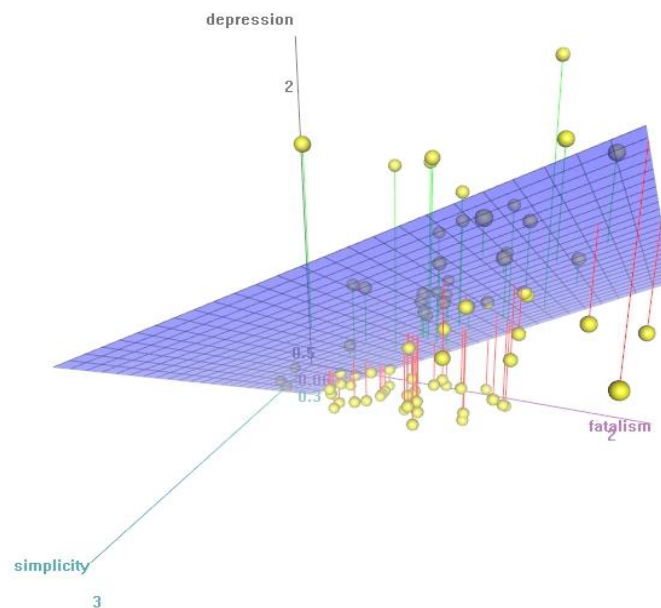


Figure 1: Three-dimensional scatter plot of the variables **fatalism**, **simplicity** and **depression**. Note that in this case the regression plane adjusted by the least squares methods is shown.

## Questions

1. Make a prediction for a person with a score of .5 points on the scale of fatalism and dichotomous thinking for each of the models studied.
2. Once the analysis described in the previous paragraph is made, what would you recommend to the researchers of a similar study? Should a simple linear regression model or a multiple regression model be recommended? Justify your answer.
3. Determine the adequacy or otherwise of the following conclusions in the case of this study:

The regression model number 1 better helps to predict the level of depression.

It can be concluded that fatalism is the main cause of depression.

---

**Additional references**

- Beck, A. T., Steer, R. A. & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.
- Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. *Journal of Statistics Education*, 9 (3).
- Wright, D. B., & London, K. (2009). *Modern regression techniques using R. A practical Guide for students and researchers*. London, UK: Sage.

[Back](#)

## PRACTICE ON THE YERKES-DODSON LAW

### Introduction

There is no doubt that in terms of the accuracy and efficiency with which a person performs a task, there are closely associated cognitive aspects. Thus, for example, processing capacity and practice are related to the execution of tasks. However, physiological state also affects the efficiency with which a task is performed. Several models have been proposed to describe the relationship between physiological states and the level of accomplishment in the performance of tasks (see [Eysenck, 1982](#)). This practical will look at the *Yerkes-Dodson law*, whose origin is a study of motivation and discrimination ([Yerkes & Dodson, 1908](#)).

The Yerkes-Dodson law establishes an empirical relationship in an inverted-U shape between the level of tension, motivation and activation on one side and performance on the other. That is, between a physiological state and the result of a cognitive process, it is supposed that there is a curvilinear relationship, specifically a functional relationship according to a second-degree polynomial function. The law has another assumption: that the intensity of motivation or activation required to optimally perform a task increases with the ease of the task. So, a lesser degree of activation is required to achieve optimum results for difficult tasks than for easy tasks.

### Database

Table 1 presents the results for an assumed experiment where we verify the functional relationship between the degree of stress and the realisation and execution of a task following the Yerkes-Dodson law. It is assumed that for the 20 participants in the experiment the stress scores have been obtained through a standardised psychological test. The assumed experiment involves showing participants a sequence of 30 pairs of words in an initial list of words on a computer monitor. In all pairs, the two words have the same number of syllables, that is, the 60 words are disyllabic. Five minutes later, participants are again shown 30 pairs of two-syllable words on the computer screen. Each of the pairs may or may not coincide with any of the previously presented pairs. The task of participants consists of deciding for each pair of words that appear together in the test list whether they were presented together in the initial list, i.e., they were associated in the initial 30 pairs. The participants are given five seconds to respond. An answer is considered wrong if the time elapses before a response is given. Before submitting the lists of words, the participants are asked to focus fully on the task.

Participants	EXECUTION	STRESS <sup>1</sup>	STRESS <sup>2</sup>	STRESS <sup>3</sup>
1	7	11	121	1331
2	19	32	1024	32768
3	28	25	625	15625
4	25	24	576	13824
5	17	18	324	5832
6	21	22	484	1064
7	11	33	1089	35937
8	8	40	1600	64000
9	27	28	784	21952
10	14	35	1225	42875
11	20	31	961	29791
12	12	17	289	4913
13	26	23	529	12167
14	25	22	484	10648
15	20	30	900	27000
16	15	18	324	5832
17	26	26	676	17576
18	17	32	1024	32768
19	27	24	576	13824
20	24	20	400	8000

Table 1: The data correspond to the extent of the task execution and the scores of participants in the experiment. While the level of efficacy in the execution of the task can vary from 0 to 30, both inclusive, the degree of stress can take values from 0 to 48 for the lowest and highest degree of stress, respectively. The data are simulated and not obtained in a real experiment. In the third, fourth and fifth columns appear the stress scores for each participant raised to the first, second and third power, respectively.

### Statistical analysis

Each student must complete the following tasks in the order listed below and, where appropriate, discuss the results:

1. Enter the data from Table 1 into a database or spreadsheet, for example, *Excel*. Once the data are input, they can be imported into a statistical application. The data can also be entered into the editors available in statistical applications (e.g., *R-commander*).
2. Determine the value for the coefficient of determination for the simple linear regression model ( $y = \beta_0 + \beta_1x + \varepsilon$ ), where stress is the predictor variable for task execution. What can be concluded from the results?
3. Make a scatter chart where the x-axis shows the degree of stress and the vertical axis is the level of the task execution. Does this graph help to understand the value of the coefficient of determination obtained for the simple linear regression model? Which model seems most appropriate?
4. Obtain the level of the coefficient of determination, but now for a second-order polynomial model ( $y = \beta_0 + \beta_1x + \beta_2x^2 + \varepsilon$ ). Provide an estimate of the parameters and their standard errors, in addition to the level of significance of the quadratic term. Also, analyse the assumptions for which the following functions of **R** can be used: **dwtest(reg1)** or

**bptest(reg1)** from the **Intest** package for autocorrelation and heteroscedasticity, respectively, as well as the option **shapiro.test(reg1res)**.

5. Calculate the coefficient of determination for a third-order polynomial regression model ( $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \varepsilon$ ). What consequences are extracted from the results obtained for the cubic term?

## Questions

The questions below must be answered by each student:

1. Why is it desirable to obtain a graphical representation before estimating the parameters of a regression model?
2. Is it true that a polynomial model of order  $n - 1$ , where  $n$  denotes the total number of individuals in the sample, leads us to obtain a coefficient determination equal to 1? If so, should the model of order  $n - 1$  always be used? Whether your answer is yes or no, you must provide an argument.
3. What theoretical and applied implications can be derived from the Yerkes-Dodson law?

## Additional references

- Eysenck, M. W. (1982). *Attention and arousal. Cognition and performance*. Berlin: Springer-Verlag. [Traducción en español, Eysenck, M. W. (1985). *Atención y activación. Cognición y realización*. Barcelona: Herder].
- Yerkes, R. M., & Dodson J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

[Back](#)

## PRACTICE ON MULTIPLE REGRESSION MODELS

### Introduction

Stimulus-response (SR) models have had a significant influence on Psychology. In this type of model, behaviour is understood as a response to a stimulus. Now, with the aim of highlighting the impact on organisms, stimulus-organism-response (S-O-R) models have been proposed. S-O-R models presuppose the existence of mediator mechanisms between stimuli and responses and these mechanisms, such as mediating variables, can be understood as not directly observed. As a result, we find a clear connection between statistical models for the analysis of measurements and a number of conceptual representations in the discipline of Psychology.

If stimulus and response are represented as two variables,  $X$  and  $Y$ , and there is no other intervening variable, there are only four possibilities. First, both variables are not related (lack of effect). Second,  $X$  causes  $Y$  (asymmetric or directional effect). Third,  $Y$  causes  $X$  (asymmetric or directional effect). Fourth, and lastly, both variables are associated (reciprocal effect).

Now, we may discuss the case where there are three variables,  $X$ ,  $Y$  and  $Z$ , assuming no other variable is involved. First, the case may involve a relationship between pairs of internal variables. In this situation, asymmetric effects can be seen, i.e., a variable is the cause in the six possible directional effects between the pairs of variables. The effect can also be reciprocal. There may even be no effect between the variables. Second, if we consider the third variable, additional possible effects other than asymmetric or reciprocal ones include: confusion, covariance and mediator. The effect of confusion occurs when, for a pair of variables  $X$  and  $Y$ , the variable  $Z$  modifies the relationship between the two as it relates to both of them. This can be noted both in causal and reciprocal relationships. The covariance effect means that, if  $X$  is an independent variable and  $Y$  is a dependent variable, the variable  $Z$  can be another predictor of  $Y$  and, therefore, increase the accuracy of the predictive model. If  $Z$  is not related to  $X$ , there will not be any change in the relationship between  $X$  and  $Y$ . Thus, the difference between a variable of confusion and covariance effect is that in the former case variable  $Z$  modifies the relationship between  $X$  and  $Y$ . An effect of mediation is defined as the intermediation of  $Z$  in the causal relationship between  $X$  and  $Y$ , i.e.,  $X$  causes  $Z$  and, in turn,  $Z$  causes  $Y$ .

In other words, the mediator variable explains the relationship between  $X$  and  $Y$  because it conveys the effect of  $X$  on  $Y$  through the mediation of  $Z$ . Therefore,  $Z$  is called the mediator variable and it is common to represent variables of this kind by the letter  $M$ . Also, note that mediation implies asymmetry or directionality in the relationships. It is common to refer to mediator variables as intervening variables or intermediaries. What is the difference between a mediator variable and a moderator variable? While mediator variables refer to the mechanisms by which an impact occurs, moderator variables point to factors producing a special effect, modulating the magnitude thereof (Judd, Kenny, & McClelland, 2001). An

example might be useful to differentiate between mediator and moderator variables. The number of hours devoted to studying a subject can be increased and this may lead to better performance on academic tests of knowledge, resulting in a mediation effect. However, the impact of academic knowledge tests on performance can differ by type of school, showing a moderating effect.

These ideas about mediation can be extended to cases where there are four or more variables.

### Fundamentals of the statistical analysis of mediation

To illustrate statistical mediation analysis, note that figure 1 shows a simple regression model, where the independent variable has a clear effect on the dependent variable, denoted by  $a_1$ , while the simple linear regression model establishes a functional relationship between predictor and target variables.

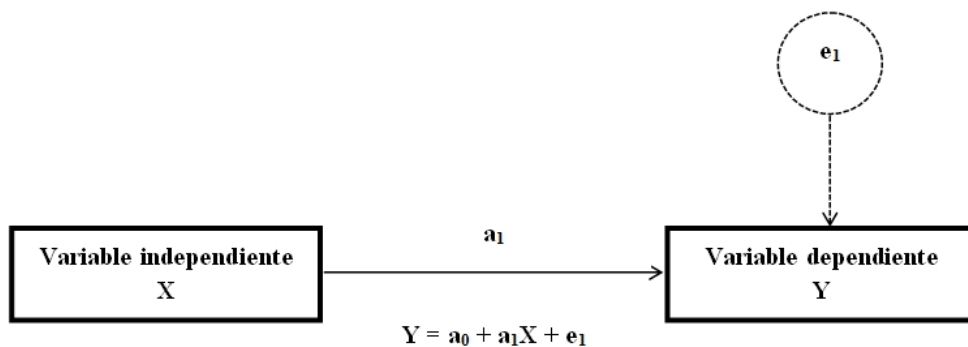


Figure 1: A continuous solid arrow and a single point indicate the asymmetry of the relationship, i.e., the variable  $X$  predicts variable  $Y$ . Therefore, the independent variable permits prediction of the dependent variable:  $a_1$  denotes the relationship between the independent and dependent variables, while  $a_0$  and  $e_1$  correspond to the constant and the error term, respectively, of the simple linear regression model.

Figure 2 shows a graphical representation with one independent variable, one dependent variable and a mediating variable. Note that the effect of the independent variable on the dependent variable can be both direct and indirect; in the second case, it is intermediated by a mediator variable. Therefore, a statistical analysis of the mediation based on linear regression models allows us to determine whether or not a mediation effect exists in the functional relationship between the variables.



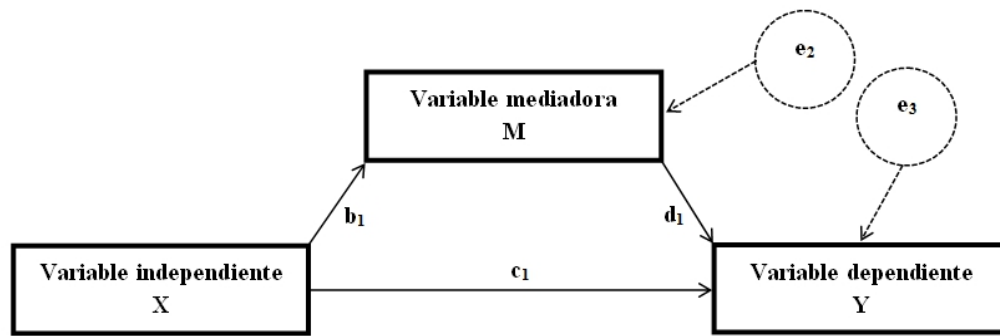


Figure 2: The figure shows an independent variable, a dependent variable and a mediator variable. The effect of the independent variable on the dependent variables arrives in two ways. There is a direct effect of  $X$  on  $Y$  noted by  $c_1$ , but there is also a secondary effect, an indirect effect in this case, which is denoted by  $d_1$ , of the independent variable over the dependent variable generated through the mediating variable  $M$ .

From the graph of figure 2 the following pair of linear regression models can be easily specified, the first simple and second multiple:

$$M = b_0 + b_1 X + e_2$$

$$Y = c_0 + c_1 X + d_1 M + e_3.$$

Note that  $c_1$  is the partial effect of variable  $X$  on variable  $Y$ , i.e., eliminating the contribution of the independent variable on the dependent variable conveyed through  $M$ , as shown in the linear multiple regression model in the last expression. In multiple regression models, the letter  $M$  represents the original scores for the mediating variable and not the values predicted from the scores for  $X$ . If  $a_1 = c_1$ , there will not be any mediation effect. In summary, the total effect of  $X$  on  $Y$ , denoted by  $a_1$ , can be decomposed into the direct effect  $c_1$  and the indirect effect  $d_1$ . Note also that  $a_1 = c_1 + a_1 - c_1$ , where  $a_1 - c_1 = b_1 \times d_1$  corresponds to the effect of mediation.

Various methods have been proposed to determine whether there is a mediation effect, but this point is beyond the scope of the practical.

For this reason, the only approach here refers to the criteria advocated by [Baron and Kenny \(1986\)](#). Based on the above linear regression models, we expect that  $b_1$  and  $d_1$  are statistically significant when testing the null hypothesis that establishes lack of effect. On the other hand, the expectation would be not to reject the null hypothesis that states that  $c_1$  is equal to zero if there is only a mediation effect. Now, that does not occur often; therefore, it is very likely that other mediator variables mediate between  $X$  and  $Y$ . For the three mentioned statistical parameters  $b_1$ ,  $c_1$  and  $d_1$ , the statistical null hypothesis corresponds to the general form  $H_0: \beta_k = 0$ , which is highlighted to illustrate the strong connection between the analysis of mediation and the procedures of regression analysis. Similarly,  $R^2$  values

may also be considered in the analysis of mediation in order to quantify the effect size; in any case, there are several indices to quantify the strength of the association or the effect size, and each provides specific information (see [MacKinnon, 2008](#)).

As is common practice in linear regression models, it is possible to make an estimate for the range of parameters, i.e., building confidence intervals when carrying out the mediation analysis. Confidence intervals are of interest because of the error in the estimation and the plausible range of values for the parameter corresponding to the effect of mediation. To obtain each confidence interval, it is necessary to determine the lower and upper limits. The a priori confidence level,  $1 - \alpha$ , must be established. We also need to determine the standard error to construct the confidence intervals, although different expressions can be obtained through the following mathematical formula:

$$SE(\hat{b}_1 \times \hat{d}_1) = \sqrt{\hat{b}_1^2 \times s_{d_1}^2 + \hat{d}_1^2 \times s_{b_1}^2}$$

In the above expression,  $SE()$  denotes the standard error, in this case for the effect of mediation with circumflex  $b_1$  and  $d_1$  representing the estimations of the effects  $b_1$  and  $d_1$ , and  $s$  refers to the estimation of the standard error for both parameters  $b_1$  and  $d_1$ , information that is obtained from the regression analysis. Finally, we obtain the confidence interval (i.e., the upper and lower limits) for the mediating effect through:

$$\hat{b}_1 \times \hat{d}_1 \pm z_{\alpha/2} \times SE \hat{b}_1 \times \hat{d}_1$$

It is also very common to use the confidence interval to verify whether or not there is a mediation effect. The answer is affirmative only if the interval does not contain the value of zero.

Regarding the assumptions, the statistical analysis of mediation requires the specific conditions of analysis relating to linear regression models. In other words, it assumes that the model correctly specifies the linear relationships between variables (i.e., no relevant variables are omitted). It also assumes the homogeneity of variances across the entire range of values, independence between errors, and that the remainders are distributed according to the normal distribution model.

## Database

To carry out a statistical analysis of mediation, a fictitious example will be used. The study involves in an assumed group of 30 students. The independent variable will be the number of hours present in class, either practical or theory classes, in a compulsory subject of a university degree. This variable has a maximum value of 60 hours and a minimum value of zero. The dependent variable will be their results on a knowledge test conducted at the end of the semester, which may have a score between 0 and 55. Finally, the 30 students are given a questionnaire about motivation towards learning content in the various topics covered by the course, providing a mediator variable that may take values between 0 and 50. The main objective of the study is to determine whether motivation has a mediator effect between the hours of classes attended and the students' academic performance. Table 1 shows the data corresponding to the assumed investigation.

## Development of the practice

1. Develop a graphical representation of the model.
2. Determine the coefficients  $a_1$ ,  $b_1$ ,  $c_1$  and  $d_1$ .
3. For all the above factors, take a decision on the statistical null hypothesis which states that its effect is not present, i.e., the effect is zero.
4. Make an interval estimate of the mediation effect.
5. Finally, is there a mediation effect? Controlling the mediation effect, is there a direct effect? If so, can this direct effect be explained by other mediating variables?

## Questions

1. First, draw a graph representing an independent variable, a dependent variable and two mediator variables. Denote each arrow with a letter corresponding to the effect of the variable over the other variable and find the different error terms.
2. Next, propose a psychological model that corresponds to the graphical representation of the previous point.

ID	Attendance	Motivation	Performance
1	43	20	25
2	51	30	45
3	50	50	55
4	51	48	55
5	51	40	55
6	42	20	25
7	50	40	45
8	49	30	55
9	52	30	45
10	51	20	35
11	51	30	45
12	44	10	20
13	47	40	25

14	51	30	45
15	51	30	25
16	50	20	45
17	50	30	35
18	49	40	45
19	49	30	42
20	50	20	35
21	41	10	15
22	44	20	35
23	45	20	30
24	46	20	15
25	51	40	35
26	48	20	15
27	52	40	35
28	49	30	42
29	50	30	42
30	48	30	25

Table 1: For a sample of thirty people, the tables shows the measures for the variables of *attendance (hours in class)*, *motivation toward the subject* and *academic performance on the knowledge test*. These variables correspond to the independent variable, mediator variable and dependent variable, respectively.

### Additional references

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychology research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*, 115-134.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis Group.

[Back](#)

## PRACTICE ON NON-LINEAR REGRESSION MODELS (1)

### Introduction

Measurement in psychology has its origins in classical psychophysics founded by Weber and Fechner in the mid-nineteenth century (see [Aznar, 2009](#)). The core of psychophysics research can be summarised in the operationalisation of the relationship between the physical characteristics of a given stimuli and the magnitude of the response by the subject affected by such stimuli. Put another way, it focuses on the study of the relationships between physical stimuli and psychological reactions. According to the proposed model, the objects of study are the relationships established between the physical plane (outer and physical) and the psychological plane (inner and psychophysical). In addition, a paramount interest of radical classical psychophysics is the study of different perceptual thresholds.

In this regard, Weber (1795-1878) demonstrates that people are more skilful in the detection of relative changes than in the detection of absolute changes. From experiments, it can be concluded that the ratio between the change from baseline stimuli that we may be able to perceive ( $\Delta S$ ) and the reference stimuli intensity ( $s$ ) is a constant term ( $\kappa$ ), which is also called the *Weber Fraction*. Therefore,  $\kappa = \frac{\Delta S}{s}$ . For example, suppose that a person holding a weight of 500 grams in one hand is given another heavier weight in the other hand. From which weight will the difference be noted? Assume that in this case the minimum increase to be perceived is 10 grams, that is, the heavier weight is 510 grams. Therefore, the constant in this case will be equivalent to 0.02.

Based on the *Weber Fraction*, Fechner (1801-1877) developed the *Fechner-Weber Law*. Fechner differentiated between the absolute threshold, i.e., the lowest intensity at which a stimulus can be perceived, and the differential threshold, i.e. the increase needed to perceive switching between two magnitudes of stimuli. The latter is precisely the unit of measurement of a subjective scale, in other words, the psychological scale unit. Assuming that differential thresholds are equivalent and, therefore, the increasing sensation can be calculated using the Weber fraction:  $\Delta S = C \frac{\Delta S}{S}$ , where  $C$  is a constant of proportionality. From this expression the relationship between stimuli and the sensation associated with the stimuli is known as *Fechner's Law*:  $S = K \log(S)$ . That is, the experienced intensity increases proportionally to the logarithm of the physical stimuli.

Subsequently, Stevens articulated *Stevens' Law* or the *power law*, which predicts a change in sensation equivalent to the change in the intensity of stimuli ([Stevens, 1957](#)). The general expression is:  $R = CS^r$ , where  $R$  is a subjective response to the stimuli,  $S$  is the intensity of stimuli,  $r$  is the slope that varies in function of the sensory modality and  $C$  is an arbitrary constant that determines the measurement unit of the scale. From the above, it follows that both scales grow on a logarithmic scale. Note also that both Weber's law and Fechner's law are special

cases of the power law (see Figure 1).

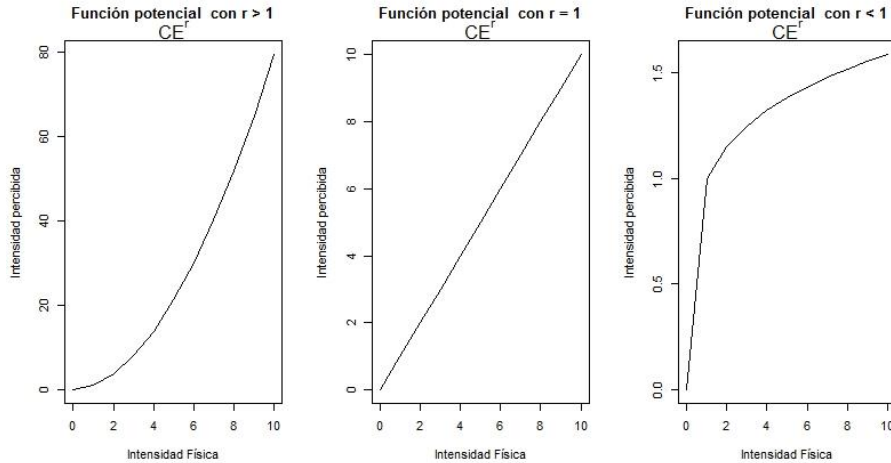


Figure 1: Power function according to different values of  $r$ . Note that when  $r > 1$ , it is an exponential function; where  $r = 1$ , it corresponds to a linear function, and in the case of  $r < 1$ , it is a logarithmic function.

In this practical, we will proceed to provide data for a non-linear model corresponding to a power model according to Stevens' law. To do so, we will estimate the model of coefficients, take decisions about them, and lastly make predictions.

### Database

- In a group of 30 people, perform the following experiment:
- Each participant will be asked to touch a small metal plate with one hand. The plate can be different temperatures.
- After being exposed to the stimuli, participants are asked for an approximate value of the temperature.
- The temperature of the metal plate can vary by 1 degree within the range of 15 to 55 degrees Celsius. In this way 41 conditions are calculated.
- Once the information is collected from all participants, calculate the means for each condition.

The **Stevens** database, which is available on the Virtual Campus, contains the averages for subjective temperatures provided by the subjects (variable **sensación**) for each of the experimental conditions relating to the intensity of physical stimuli (variable **intensidad**). The database also includes the logarithmic transformation (base 10) of the above variables. The transformations are **logsens** and **logint**, respectively.

## Statistical analysis

For this practical, each student should perform the following tasks:

1. Load the **Stevens** database in the work session of **R**.
2. Perform a scatter plot diagram for the variables **sensación** and **intensidad**. Take into account the arrangement of the variables on the axes of the graph based on the functional relationship you want to model.
3. Once the scatter plot diagram is made, determine the type of model adequate for this case.
4. Specify the non-linear regression model that seems appropriate in view of the scatter plot diagram.
5. Determine if it is a linearisable non-linear regression model. If yes, specify the mathematical transformation that enables you to express the linear relationship between the parameters of the model.
6. Estimate the coefficients of the model from the sample data and perform the test of statistical decision for each of them.
7. Validate the regression model using an F-test indicating the value of the statistics and the degree of significance. Also, determine the model's predictive ability.
8. Using the model, determine the thermal sensation of a person exposed to contact with a piece of metal at 56°C.

## Questions

The students should carefully consider and answer the following questions:

1. Why is it important to make a scatter plot diagram before modelling the functional relationship between sensation and intensity of stimuli?
2. Is Stevens' law adequate to explain the relationship between psychological intensity and physical intensity?

## Additional references

Aznar Casanova, J. A. (2009). *Introducción a la psicofísica*. Barcelona: Universitat de Barcelona.  
Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

[Back](#)

## PRACTICE ON NON-LINEAR REGRESSION MODELS (2)

### Introduction

Sometimes a linear regression model is not suitable for modelling the relationship between two or more variables. On these occasions, non-linear models can be useful.

There are two types of non-linear models: linearisable non-linear models (inherently linear) and non-linearisable non-linear models (inherently non-linearisable). The first type involves a transformation that enables us to express the relationship between the parameters in a linear form. The second type permits no transformation.

In this practical, we will proceed to apply data to a non-linear model. To do so, we will estimate the model of coefficients, take decisions about them, and lastly make predictions.

### Database

Suppose that a researcher from the animal learning area is interested in studying the relationship between number of trials and the average length of time (measured in seconds) between successive pressing of a lever by five lab rats in a Skinner box:

- Log the average length of time between a lab rat's pressing of the lever as a function of the number of trials.
- Once the information is collected for all animals, calculate the means for each of the conditions (number of trials).

The database *pulsaPalanca*, which is available on the Virtual Campus, contains the averages of the time lapses by animal (variable *lapso*), by number of trials performed (variable *ensayo*) and by the natural logarithm of the time lapse (variable *lnlapso*).

### Statistical analysis

For this practice, the student should perform the following tasks:

1. Load the *pulsaPalanca* database in the work session of R.
2. Make a scatter plot diagram for the test variables and time. Take into account the arrangement of the variables on the axes of graphs based on the functional relationship that you want to model.
3. Once the scatter plot diagram is made, determine the type of model for this case.
4. Specify the non-linear regression model that seems appropriate in view of the scatter plot diagram.
5. Determine if it is a linearisable non-linear regression model. If yes, specify the mathematical transformation that enables you to express the linear relationship between the parameters of the model.



6. Estimate the coefficients from the sample data and perform the test of statistical decision for each of them.
7. Validate the regression model using an F-test indicating the value of the statistics and the degree of significance. Also determine the model's predictive ability.
8. Using the model, determine the variable "lapso entre 2 pulsaciones" after making 10 runs.

### Questions

The students should carefully consider and answer the following questions:

1. Why is it important to make a scatter plot diagram before modelling the functional relationship between the number of trials and the time between each pressing of the lever?
2. Is the non-linear regression model obtained from this study of animal learning adequate?

[Back](#)

## Concluding remarks

We would like to share our thoughts about methodology and statistics as a wrap-up to this practical document. We hope that the reader has grasped and accepted the idea that scientific studies are useful as an objective, reliable, and valid way of gathering empirical information about the phenomena of interest, but that they are also imperfect. Two of the aspects that contribute to the validity of the conclusions reached are the way in which the data are gathered and analysed. Nevertheless, these two stages of the research process may also be the main issues undermining the credibility of the information or the interpretations presented in the report of a study. Therefore, we consider that the greater theoretical and practical knowledge on methodological and statistical topics is an invaluable resource that enables each student, professional, or researcher to be a critical assessor of reports and a skilled aware-of-the-limitations investigator.

Usually the statistical analysis is regarded as the least attractive and most cumbersome part of the research process. We would be glad if the practices included in Sections 4, 5, and 6 have contributed to show that statistics are not *the* problem, given that they can be understood and applied correctly with moderate training (especially if these training includes probability models, such as the ones illustrated in Section 3). However, we also hope that the reader has seen that statistics are also not the *answer*, as the data collection procedure (i.e., sampling, definition and allocation to different conditions, instruments used, context of application of the instruments) is usually the critical aspect of the research. Section 2 has illustrated only the most frequently used ways of gathering data, but the design of study requires much more attention. Finally, the report of a research can be both an appropriate answer to study limitations and a catalyser of further problems. In that sense, Section 1 was intended to show that a study needs solid bases, build on the currently available scientific knowledge, as summarized in peer review articles and books, whereas the description of an already performed study needs to reflect honestly the whole process. Therefore, the report can be used to make clear the strengths and limitations of a study in order to enable the reader to understand the extent to which its conclusions can reasonably be applied. At the same time, the report can deepen into the methodological and analytical weaknesses by not making the reader aware of the way in which the information was gathered or the specific meaning of the numerical values yielded by the statistical techniques.

We hope that the practices included in this document and further readings on the topic would enable avoiding the problems in the eternal look for better answers.