anales de psicología
2012, vol. 28, nº 1 (enero), 281-288

ISSN edición impresa: 0212-9728. ISSN edición web (http://revistas.um.es/analesps): 1695-2294

When the mean fails, use an M-estimator

Berta Cajal, Elena Gervilla, and Alfonso Palmer*

University of the Balearic Islands

Título: Cuando falle la media, utilice un M-estimador.

Resumen: En el campo de las adicciones en muchas ocasiones se tiene que trabajar con variables cuantitativas, siendo la media aritmética el índice de localización utilizado mayoritariamente. No obstante, el uso de este índice debería limitarse a aquellas situaciones en las que las distribuciones de las variables sean simétricas. El objetivo de este trabajo es ejemplificar la importancia de recurrir a estadísticos descriptivos adecuados para resumir variables cuantitativas, mediante el estudio de la cantidad de consumo de sustancias adictivas en la adolescencia.

La muestra está formada por 9300 estudiantes con edades entre los 14 y los 18 años (47.1% chicos y 52.9% chicas) que contestaron de forma anónima un cuestionario sobre consumo de sustancias.

Se describe la cantidad de consumo semanal de diferentes sustancias mediante índices de localización clásicos y pertenecientes al Análisis Exploratorio de Datos (EDA). Se puede observar cómo los resultados varían notablemente en función del estadístico elegido, siendo el *M*-estimador de Huber un índice con valores más "reales".

La media aritmética no es un buen índice para acercarnos debidamente a la realidad del consumo de drogas cuando las distribuciones son asimétricas, siendo necesario utilizar índices resistentes, tal como, entre otros, el *M*-estimador de Huber.

Palabras clave: estimadores robustos; media; sustancias adictivas; consumo de drogas; *M*-estimador.

Abstract: In the field of addictions on many occasions one has to work with quantitative variables, and the arithmetic mean is the most used location index. Nevertheless, the use of this index should be limited to those situations in which the distributions of the variables are symmetrical. The aim of this work is to exemplify the importance of recurring to adequate descriptive statistics in order to summarize quantitative variables, through the study of the quantity of consumption of addictive substances in adolescence.

The sample is made up of 9300 students between 14 and 18 years (47.1% boys and 52.9% girls) who anonymously answered a questionnaire on consumption of substances.

The quantity of weekly consumption of different substances is described using classical location indexes belonging to Exploratory Data Analysis (EDA). It can be seen how the results vary noticeably according to the statistics selected, with the Huber *M*-estimator as the index giving more "real" values.

The arithmetic mean is not a good index in order to duly approach the reality of drug consumption when the distributions are asymmetrical, in which cases it becomes necessary to use resistant indexes such as, among others, Huber's *M*-estimator.

Key words: robust estimators; mean; addictive substances; drug consumption; *M*-estimator.

Introduction

In the field of addictions it is necessary to have information concerning the pattern of use of addictive substances and on many occasions this means having to describe and summarize quantitative variables like the quantity of consumption of these substances. In these cases, the arithmetic mean is the location index used in the great majority of research studies.

The arithmetic mean provides a value of the variable which represents the centre of gravity of the distribution (in which the distribution of observations is balanced). Nevertheless, it is worth remembering that the use of this classical location index should be limited only to those occasions on which the distribution of the variable is symmetrical.

As opposed to the arithmetic mean, there are other more appropriate measures of central tendency to describe data when dealing with asymmetrical distributions and/or with outlier values. In this sense, we have indexes such as the trimmed mean which consists of eliminating a proportion of the data from each extreme and calculating the mean of the remaining values, or the Winsorized mean, which instead of eliminating a whole number of cases from each extreme, substitutes them for the last value, at each extreme, which is part of the analysis.

*** Dirección para correspondencia [Correspondence address]**:
Alfonso Palmer. Dpto de Psicología. Ctra de Valldemossa, km 7.5. Universitat de les Illes Balears. 07122 Palma (Baleares, Spain).
E-mail: alfonso.palmer@uib.es

On the other hand, Exploratory Data Analysis, generally known as EDA (Tukey, 1977) offers a set of simple, resistant and clear techniques. EDA, contrary to traditional descriptive analysis, places more relevance on resistant measures and on graphic information (Palmer, 1999). As a result, EDA incorporates indexes and graphics that overcome the problems presented by classical descriptive statistics, when facing non symmetrical distributions and the presence of outlier values.

Among the indexes included in EDA we find the median and the M-estimators (Huber, 1964). The median is defined as the value of the variable that divides the distribution into two equal parts, each of which contain 50 per 100 of the observations, whereas the M-estimators look for a location index from the total set of observations, pondering these depending on how near or far they are from the centre of the data.

An estimator of location or of scale is said to be resistant if slight changes in the distribution of the data have hardly any effect on its value. From this point of view, it is obvious that the introduction of only one extreme value in the distribution means there is a change in the arithmetic mean of that distribution. Thus, the arithmetic mean is not a resistant index.

However, in order to be able to talk about resistance there is a series of properties from which it is possible to establish what the best estimator for a certain distribution could be. In this work we will focus on the search for the best location index, basically in the presence of extreme values, that is, the so-called "outliers".

Below we briefly define the properties to be taken into account (Huber, 1981; Hampel et al., 1986; Wilcox, 2005; García Pérez, 2005):

1.- The *influence function* determines the influence an anomalous value has on the value of the estimator. If the influence function is not bounded, it means that the further away the anomalous datum is, the greater the influence exerted on the estimator. This is what happens in the arithmetic mean, whose influence function is linear and is not therefore bounded.
2.- The *gross-error sensitivity* measures the influence exerted by a certain quantity of contamination (anomalous values) in the data on the value of the estimator. If this value is finite, the estimator is said to be B-robust.
3.- The *local-shift sensitivity* is the one determined by small fluctuations in the data, and it is desirable for it to be small and finite.
4.- Under the strategy that it is convenient to eliminate clearly anomalous values, the influence function must be zero from a certain value. For symmetrical distributions around zero, the *rejection point* is the value from which the data must be rejected. It is desirable for the estimator to have a finite rejection point.
5.- The *breakdown point* of an estimator is the percentage of outliers the estimator can stand before breaking down, that is, before ceasing to be valid, and this defines the quantitative robustness. An estimator is resistant only if its breakdown point is greater than zero.

The M-estimators of location weight the observations on the basis of their relative distance from the centre of the distribution, whereas a winsorized mean replaces a predetermined alpha percentage of observations with a non-outlier value, and what a trimmed mean does is to eliminate this percentage of observations. Why use an M-estimator and not a winsorized mean, the median or a trimmed mean, instead of an arithmetic mean?

The sample median, is B-robust, its gross-error sensitivity is finite, its local-shift sensitivity is infinite, its rejection point is infinite, it is qualitatively robust and its breakdown point is ½.

The winsorized mean is B-robust; however it has an infinite value for local-shift sensitivity, when what is desirable is that the estimator has the smallest possible finite value. Likewise, it has a non finite rejection point and an *a* value breakdown point, which means that it stands a maximum proportion of alpha value of outliers in order to continue making sense as a location index. Lastly, the winsorized mean is not a qualitatively robust estimator.

The trimmed mean is B-robust, therefore its gross-error sensitivity is bounded and its local-shift sensitivity is finite, depending on the alpha value, in which case in this sense it is an improvement on the winsorized mean. However, the rejection point is still infinite and the breakdown point is alpha, just like the winsorized mean.

Huber's *M*-estimator is an estimator with good properties, both in terms of resistance and efficiency, since, amongst others, it is qualitatively robust and reaches the maximum possible breakdown point and is the best optimum B-robust estimator, that is to say, its gross-error sensitivity is bounded.

Despite the fact that Huber's estimator does not have a finite rejection point, as Hampel's (three part redescending) estimator may have, this does not have the efficiency properties that Huber's estimator has.

Hence, of the different *M*-estimators, Huber's is one of the ones that provides values nearest the arithmetic mean, due to the fact that it weights with a value of 1 a greater quantity of central data than other *M*-estimators, Tukey or Andrews types, which weight all observations below 1, in which case their comparison with the arithmetic mean is one of the most conservative.

One advantage of the M-estimator over a trimmed mean is that this will always eliminate data from both extremes of the distribution, also eliminating possible valid data, whereas the M-estimator will focus more on the extreme with most outlier values, and could even leave the other extreme of the distribution without modification.

The sample mean, *m*, as an estimator of central tendency, is not B-robust, its gross-error sensitivity is infinite, its local-shift sensitivity is 1, its rejection point is infinite, it is not qualitatively robust and its breakdown point is zero. Automatically choosing the arithmetic mean as the best index of location, and efficiency only makes sense when the population distribution is the normal distribution, but it does not show any protection against the presence of outliers.

Despite the drawbacks of the arithmetic mean in asymmetrical distributions and in the presence of outlier values, and in spite of the existence of simple procedures with the advantages already outlined, this is still used practically exclusively as if it were the only available location estimator. In this way, and even though it is normal to find asymmetrical distributions in this field, the arithmetic mean is mainly used to summarize the state of consumption of substances in different populations, both in research studies by individual authors and also in official reports produced by institutions of recognised prestige (Best, Manning, Gossop, Gross, & Strang, 2006; Government Delegation for the National Plan on Drugs, 2008; Gómez-Talegón, Prada, Del Río and Álvarez, 2005; López-Torrecillas, Peralta, Muñoz-Rivas and Godoy, 2003; Sáiz et al., 2001), documents which will act as the basis for the subsequent elaboration of prevention and intervention programmes.

The study we are presenting has the aim of exemplifying the importance of turning to appropriate descriptive statistics in order to duly approach the reality of drug consumption using the study of the quantity of consumption of addictive substances in adolescence.

Method

Participants

A random sample was carried out in education centres (conglomerates) on the island of Mallorca, using 47 centres out of a total of 122. The sample is made up of 9300 students with ages between 14 and 18 years (47.1% boys and 52.9% girls). Regarding the sample size, it is worth noting that this constitutes 41.16% of the Population size (N = 22593) from which it was extracted.

Before calculating the different statistical indexes, a previous study of the data matrix was carried out, eliminating the subjects who offered answers that were outside the range.

Procedure

The adolescents had to anonymously answer a questionnaire asking about the use of different addictive substances as well as a series of psychosocial variables (environmental, personal, family and school).

In this study we analyse the data using the SPSS programme 15.0.

Demographic variables were taken into account (sex, age and place of residence) as well as the quantity of weekly consumption of different legal and illegal addictive substances: alcohol, tobacco, cannabis, cocaine and extasis.

Results

First of all, we offer a description of the quantity of weekly consumption of the different substances in adolescence using classical location statistics and those belonging to Exploratory Data Analysis (EDA): arithmetic mean, trimmed mean, Winsorized mean, median and Huber's M-estimator. Thus, in Table 1, you can see how the estimate of quantity of weekly consumption of the different substances varies notably depending on the statistics chosen.

Table 1. Quantity of weekly consumption of addictive substances among consumers using the arithmetic mean, the median, the trimmed mean, the Winsorized mean and Huber's M-estimator.

	Arithmetic mean	Trimmed mean	Winsorized mean	Median	*Huber's M*-estimator
SDU (Standard Drink Unit)	9.06	7.72	7.37	6.00	6.95
Fermented SDUs	2.73	2.18	1.88	2.00	1.94
Distilled SDUs	8.06	6.87	6.53	6.00	6.08
Tobacco (cigarettes/week)	32.62	27.81	25.13	20.00	21.36
Cannabis (joints/week)	8.97	7.15	6.35	4.00	4.94
Cocaine (mg/week)	8.63	5.87	4.49	3.00	3.61
Extasis (pills/week)	3.52	2.94	2.77	2.00	2.58

In Table 2 we show the quantity of weekly consumption, taking into account the place of residence (urban or rural), through two indexes: arithmetic mean and Huber's M-estimator. It can be seen how the results obtained using both statistics in the weekly consumption of tobacco and cocaine are especially conflicting.

Table 2. Description of the quantity of weekly consumption of addictive substances among consumers according to the place of residence using the arithmetic mean and Huber's M-estimator.

	Urban		Rural	
	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator
SDU (Standard Drink Unit)	8.77	6.75	9.17	7.02
Fermented SDUs	2.38	1.80	3.08	2.06
Distilled SDUs	7.92	5.92	8.03	6.17
Tobacco (cigarettes/week)	30.85	18.26	34.46	23.03
Cannabis (joints/week)	8.97	4.94	9.44	4.94
Cocaine (mg/week)	12.16	5.43	6.17	3.36
Extasis (pills/week)	3.27	2.39	2.67	2.46

In Table 3 we present the quantity of weekly consumption according to gender of the adolescent analyzed using the arithmetic mean and Huber's M-estimator.

Here, we can also observe important differences in the quantity of consumption summarized by the statistics.

The case of weekly cocaine consumption is particularly notable. In the first data cleansing the subjects who reported excessive consumption (e.g. 1000 milligrams a week) were eliminated. Afterwards, in a second cleansing, it was also decided to omit the answers of all those subjects who indicated a consumption equal to or greater than 250 milligrams a week. All the results expressed in the tables took into account both cleansings in the variable relating to quantity of weekly cocaine consumption.

In the case of the boys, 78.6% reported a weekly cocaine consumption lower than 30 milligrams, 3.6% indicated a weekly consumption of 100 milligrams and the remaining 17.8% reported consuming over 250 milligrams a week. In the case of the girls the situation was parallel: 72.7% with a

consumption below 26 milligrams a week, and 27.3% with a consumption of 250, or more, milligrams a week.
These data show once again that, even alter cleansing the data, when the distribution has outlier values, the mean does not offer a representative value of the whole set of subjects and, therefore, does not draw an accurate view of reality.

Table 3. Description of the quantity of weekly consumption of addictive substances among consumers according to gender using the arithmetic mean and Huber's M-estimator.

	Boys		Girls	
	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator
SDU (Standard Drink Unit)	10.18	7.10	8.13	6.25
Fermented SDUs	3.19	2.11	2.24	1.77
Distilled SDUs	8.79	6.50	7.45	5.80
Tobacco (cigarettes/week)	33.61	21.91	32.16	21.09
Cannabis (joints/week)	10.31	6.29	7.71	3.84
Cocaine (mg/week)	10.37	3.58	6.13	4.16
Extasis (pills/week)	4.44	2.71	2.53	2.46

Finally, in Table 4 the quantity of weekly consumption of the different substances studied are compared according to the adolescents' age. We can observe the discrepancies shown by both statistics when it comes to describing the same variable. For instance, in the weekly consumption of tobacco, the statistics differ up to 14 points at 18 years of age; that is, depending on the statistics chosen we can conclude that adolescents smoke 14 cigarettes more (or less) per week. In the following paragraphs we analyze this variable graphically in order to find out why these differences are produced.

Table 4. Description of the quantity of weekly consumption of addictive substances among consumers according to age using the arithmetic mean and Huber's M-estimator.

	14		15		16		17		18	
	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator	Arithmetic mean	*Huber's M*-estimator
SDU (Standard Drink Unit)	7.66	5.84	8.48	6.59	8.95	6.80	9.53	7.36	10.91	8.68
Fermented SDUs	2.11	1.00	2.35	1.81	2.82	1.93	3.06	2.07	3.59	2.47
Distilled SDUs	7.04	5.09	7.64	5.86	8.02	6.01	8.34	6.32	9.33	7.04
Tobacco (cigarettes/week)	22.45	12.50	27.12	14.61	34.35	21.37	36.36	24.52	42.08	28.34
Cannabis (joints/week)	9.47	4.61	7.87	4.77	9.73	5.43	8.76	4.29	8.88	5.69
Cocaine (mg/week)	3.67	5.00	9.30	5.60	8.25	6.10	13.89	1.00	1.70	1.00
Extasis (pills/week)	4.00	4.00	2.80	2.67	4.11	2.24	2.70	2.63	5.00	4.51

The graphical description of a quantitative variable in classical statistics is carried out using a histogram, which provides the frequency observed in each of the intervals the data distribution is divided into. Figure 1 shows the histogram relating to the number of cigarettes consumed per week between 14 and 18 years of age. The line superimposed in the histograms represents the normal distribution. It can be seen how, throughout the age range analyzed, the distributions observed lie notably far away from the normal distribution, and therefore from the symmetry.

Figures 2 and 3 present, respectively, the stem and leaf graphs and the boxplots of the distributions observed between 14 and 18 years of age. These graphs make it possible to appreciate the asymmetry presented by these distributions as well as the existence of outlier values.

The asymmetry of the distributions as well as the presence of outlier values show the discrepancies observed when we analyze the number of cigarettes smoked per week using the arithmetic mean or Huber's M-estimator, while the latter is robust for these two characteristics.

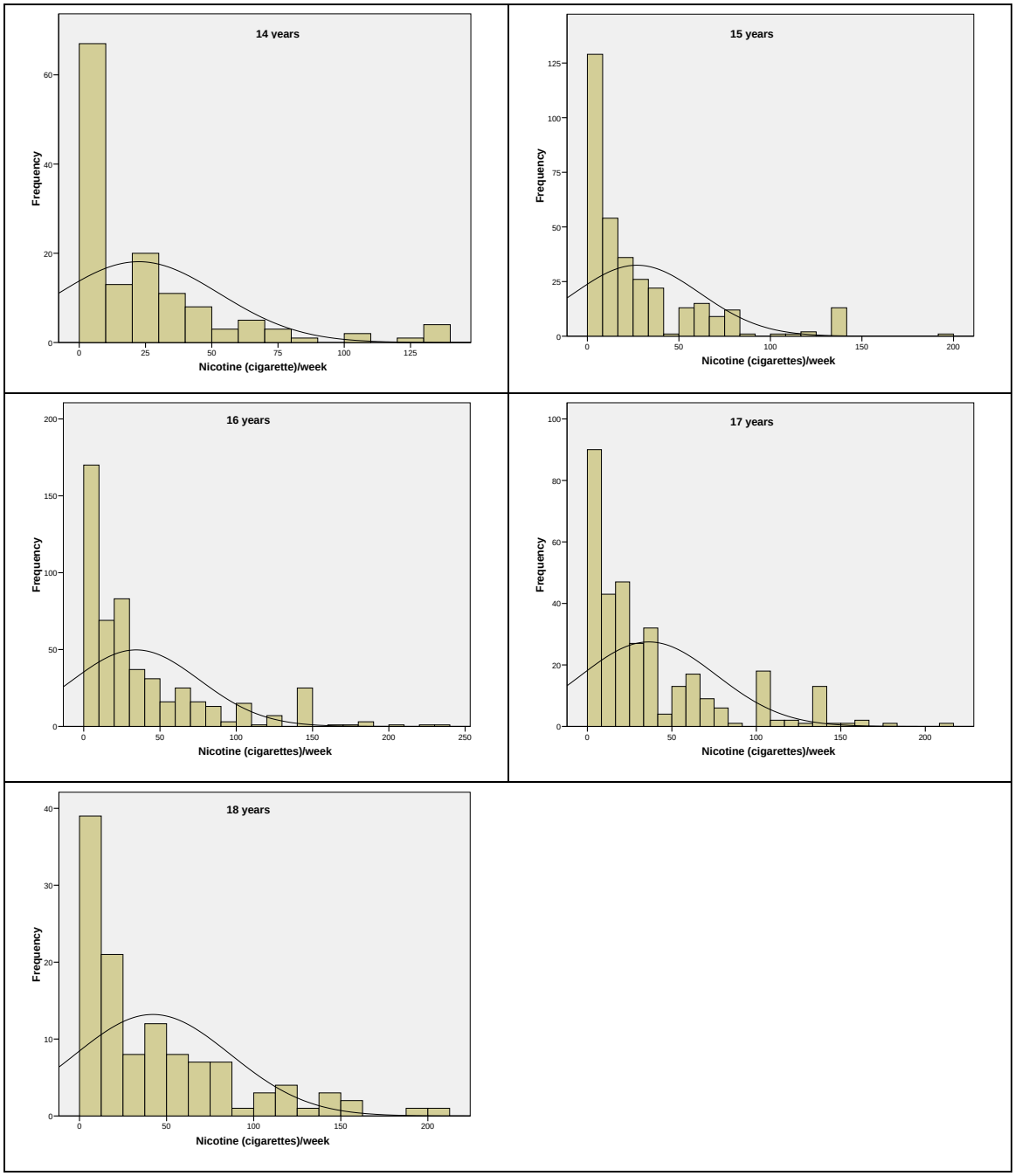

Figure 1. Histogram of the number of cigarettes consumed per week among smokers according to age.

Nicotine (cigarettes/week) Stem-and-Leaf Plot for
Age= 14

Frequency	Stem &	Leaf
47,00	0 .	11111111111111122222222233333333333333334444444
20,00	0 .	55555555556666777888
12,00	1 .	000000001244
1,00	1 .	5
17,00	2 .	00000000000000113
3,00	2 .	578
7,00	3 .	0000000
4,00	3 .	5555
7,00	4 .	0000000
1,00	4 .	9
3,00	5 .	000
,00	5 .	
5,00	6 .	00000
,00	6 .	
3,00	7 .	000
8,00 Extremes		(>=80)

Stem width: 10
Each leaf: 1 case(s)

Nicotine (cigarettes/week) Stem-and-Leaf Plot for
Age= 15

Frequency	Stem &	Leaf
89,00	0 .	11111111111112222222222223333333333333334444
43,00	0 .	55555666666777778889
44,00	1 .	00000000000000000024
10,00	1 .	5558&
33,00	2 .	0000000000000000
9,00	2 .	5558
17,00	3 .	00000000&
5,00	3 .	55
17,00	4 .	00000000
1,00	4 .	&
12,00	5 .	00000&
1,00	5 .	&
14,00	6 .	0000000
1,00	6 .	&
9,00	7 .	0000
,00	7 .	
13,00	8 .	000000&
18,00 Extremes		(>=105)

Stem width: 10
Each leaf: 2 case(s)

Nicotine (cigarettes/week) Stem-and-Leaf Plot for
Age= 16

Frequency	Stem &	Leaf
170,00	0 .	11111122222222333333445555556667778&
69,00	1 .	00000000024555&
83,00	2 .	0000000000000058&
37,00	3 .	0000005
31,00	4 .	000005
16,00	5 .	000&
25,00	6 .	00000&
16,00	7 .	00&
13,00	8 .	000
3,00	9 .	0
15,00	10 .	000&
41,00 Extremes		(>=110)

Stem width: 10
Each leaf: 5 case(s)

Nicotine (cigarettes/week) Stem-and-Leaf Plot for
Age= 17

Frequency	Stem &	Leaf
92,00	0 .	11111122222333333334455555667889
43,00	1 .	0000000023455558&
58,00	2 .	00000000000000045555&
17,00	3 .	000005
33,00	4 .	000000000005&
13,00	5 .	0000
17,00	6 .	000000
9,00	7 .	000
7,00	8 .	00&
,00	9 .	
18,00	10 .	000000&
2,00	11 .	&
22,00 Extremes		(>=120)

Stem width: 10
Each leaf: 3 case(s)

Nicotine (cigarettes/week) Stem-and-Leaf Plot for
Age= 18

Frequency	Stem &	Leaf
31,00	0 .	11112222333333444455555566777777
14,00	1 .	00000002444455
17,00	2 .	00000000000000455
6,00	3 .	000555
12,00	4 .	000000000009
3,00	5 .	000
5,00	6 .	00000
7,00	7 .	0000000
7,00	8 .	0000000
1,00	9 .	0
2,00	10 .	00
1,00	11 .	0
5,00	12 .	00005
,00	13 .	
3,00	14 .	000
1,00	15 .	0
1,00	16 .	0
2,00 Extremes		(>=190)

Stem width: 10
Each leaf: 1 case(s)

Figure 2. Stem and leaf graph of the number of cigarettes consumed per week among smokers according to age.

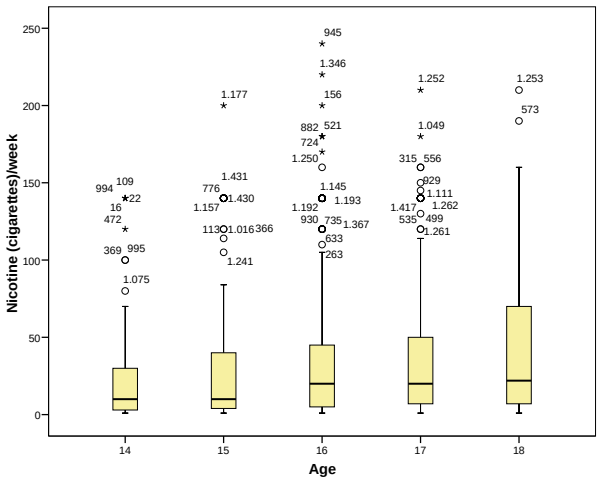

Figure 3. Boxplot of the number of cigarettes consumed per week among smokers according to age.

Discussion

The aim of this work is to exemplify the importance of using appropriate descriptive statistics in order to properly approach the reality of substance consumption in adolescence, where the data distributions people normally work with are asymmetrical and have outlier values. Therefore, the results obtained using different classical statistical methods and exploratory data analysis were compared.

The data obtained show the discrepancy observed in the results offered by classical statistics and EDA when trying to summarize one and the same variable, with the latter showing the advantage of being resistant. In this sense, it has been shown that the arithmetic mean is not a good location index when dealing with distributions that show asymmetry and/or outlier values.

It is worth highlighting that in order to be able to correctly choose the statistics to be used it is first of all necessary to study the shape of the distribution. In this sense, in order to describe the characteristic shape of a distribution, visual representations are better than purely numerical representations. And when the distributions deviate from the normal, resistant measures are preferable (Palmer, 1999).

In other words, the arithmetic mean should not be used in asymmetrical distributions as this index is affected by the extreme values of the distribution; in this case, it is not a very representative location measurement of the consumption of substances in adolescence. On the other hand, in these cases it is important to use statistics, such as M-estimators, which are robust and carry out ponderations of the extreme observations in the distribution, which makes them more representative of the central body of data.

Besides, in the case of drug consumption in adolescence, resistant measurements are preferable as they do not take into account the answers of some adolescents who may have sincerity problems, thereby excessively inflating their consumption, an aspect which can make the results obtained vary considerably, as was seen in the case of cocaine. In this sense, and despite previous cleansing of the data base, by eliminating the subjects who offered answers outside the range, the mean still offers values that are higher than reality.

If the distribution of the data does not show outliers, and it is symmetrical, the Huber estimator will give as the result the value of the sample mean which, in this case, is the most efficient one. Therefore, if you want to use a location estimator which is valid, both in the presence and absence of outliers, you should not, indiscriminately, use the arithmetic mean but rather a resistant estimator, such as, for instance, Huber's M-estimator.

The ideas put forward in this work should be extended to the study of other variables of a quantitative character that are frequently used in research into drug dependency, such as the age of initiation in the consumption of substances.

Finally, it is worth remembering that the advantages of exploratory statistics as opposed to classical location indexes were put forward a long time ago as a mean that allows researchers in Behaviour and Health Sciences, to work adequately with the types of data they normally handle (Palmer, 1993; Palmer, Amengual and Calafat, 1992). Nevertheless, perhaps due to tradition, the use of these non resistant descriptive statistics in the field of addictive substances is still majority, which lacks all sense from the method point of view and, also, taking into account the fact that the statistical packs used incorporate the calculation of resistant measurements. On the other hand, it is worth pointing out that the problem presented in this study is not exclusive to the field of addictions (Palmer, Beltrán and Cortiñas, 2006).

What is more, the interest of this work lies in offering an adequate vision of the quantity of weekly consumption of addictive substances in adolescence, as there is a large sample available made up of almost 50% of the population studied and the variables have been studied using resistant statistical indexes.

In conclusion, the use of one statistic or another may offer, as we have seen, very different views of the same reality. This has important repercussions concerning the elaboration of prevention or intervention programmes as well as when it comes to setting the objectives for these programmes. In this sense, this work aims to insist on the importance of using appropriate statistical methods if we want to obtain an adequate summary of the variables related to consumption of substances and to have a real picture of what happens in the field of drug consumption.

Acknowledgements: This work was carried out, partially, thanks to the help of the National Plan on Drugs (INT/2012/2002), from a three-year research project, for which we had all the necessary permissions in each case, both from the institutions, teachers, and also the students who participated in the research voluntarily.

References

Best, D., Manning, V., Gossop, M., Gross, S., & Strang, J. (2006). Excessive drinking and other problem behaviours among 14-16 year old school children. *Addictive Behaviors, 31*, 1424-1435.

Government Delegation for the National Plan on Drugs (2008). *Encuesta estatal sobre uso de drogas en enseñanzas secundarias (ESTUDES), 1994-2008. Retrieved 8, November, 2008 from http://www.pnsd.msc.es/Categoria2/observa/pdf/Estudes2008_Web.pdf*

García Pérez, A. (2005). *Métodos avanzados de estadística aplicada. Métodos robustos y de remuestreo*. Madrid: UNED.

Gómez-Talegón, M.T., Prada, C., Del Río, M.C., & Álvarez, F.J. (2005). Evolución del consumo de alcohol de los españoles entre 1993, 1995 y 1997, a partir de los datos de la encuesta nacional de salud. *Adicciones, 17*(1), 17-28.

Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1986). *Robust Statistics*. New York: Wiley.

Huber, P. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 36*, 73-101.

Huber, P.J. (1981). *Robust Statistics*. New York: Wiley.

López-Torrecillas, F., Peralta, I., Muñoz-Rivas, M.J., & Godoy, J.F. (2003). Autocontrol y consumo de drogas. *Adicciones, 15*(2), 127-136.

Palmer, A. (1993). M-estimadores de localización como descriptores de las variables de consumo. *Adicciones, 5*(2), 171-184.

Palmer, A. (1999). *Análisis de datos. Etapa exploratoria.* Madrid: Pirámide.

Palmer, A., Amengual, M., & Calafat, A. (1992). ¿Cuánto alcohol consumen realmente los jóvenes?: una técnica de análisis. *Adicciones, 4*(4), 315-338.

Palmer, A., Beltrán, M., & Cortiñas, P. (2006). Robust estimators and bootstrap confidence intervals applied to tourism spending. *Tourism Management, 27*(1), 42-50.

Sáiz, P.A., González, M.P., Paredes, B., Delgado, J.M., López, J.L., Martínez, S., & Bobes, J. (2001). Consumo de MDMA (éxtasis) en estudiantes de secundaria. *Adicciones, 13*(2), 159-171.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Wilcox, R.R. (2005, 2 ed.). *Introduction to Robust Estimation and Hypothesis Testing*. San Diego, CA: Elsevier.

(Article received: 26-9-2010; reviewed: 3-3-2011; accepted: 05-3-2011)