

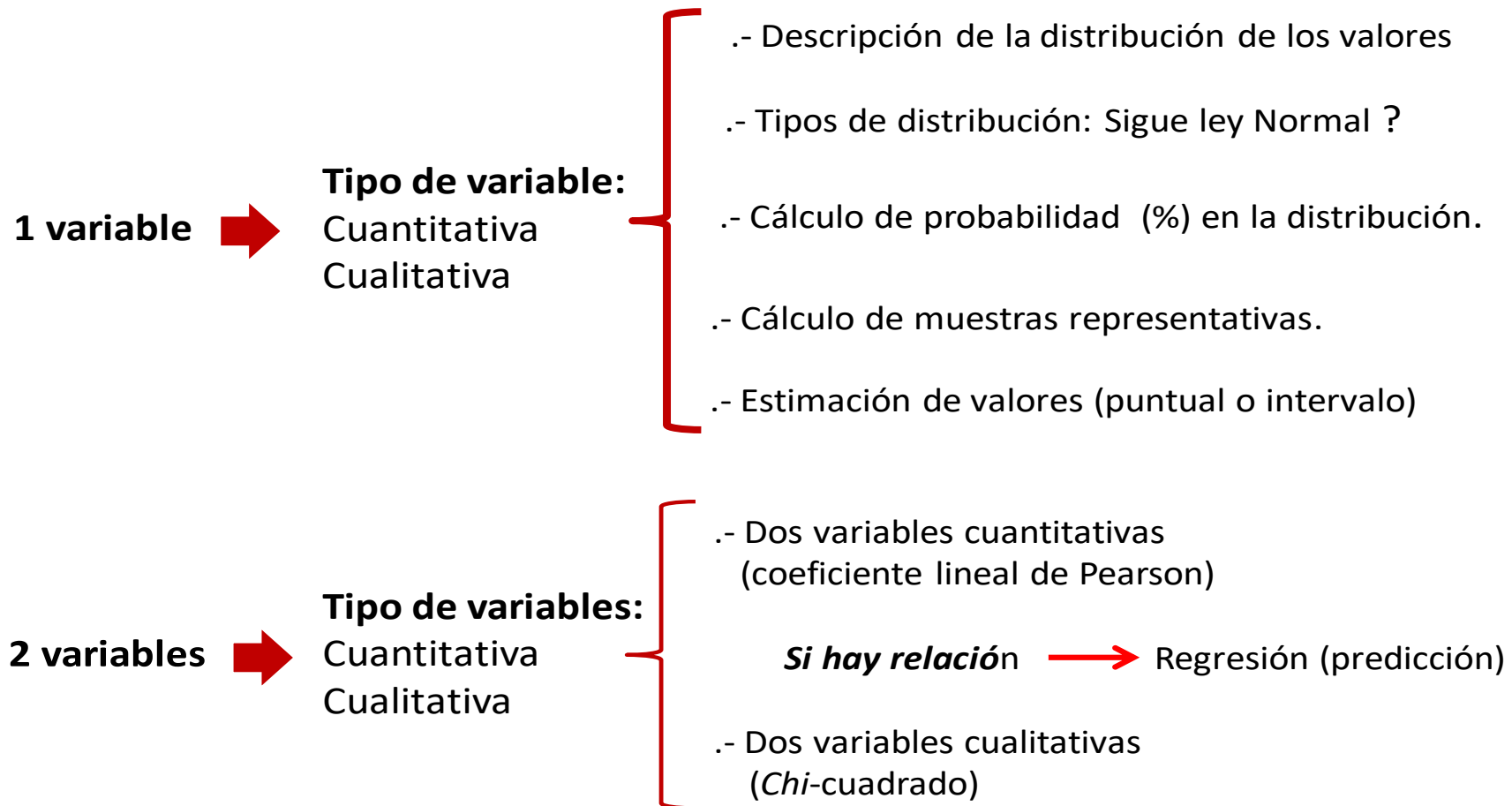
ESTADÍSTICA BÁSICA

GLOSARIO Y MAPAS CONCEPTUALES





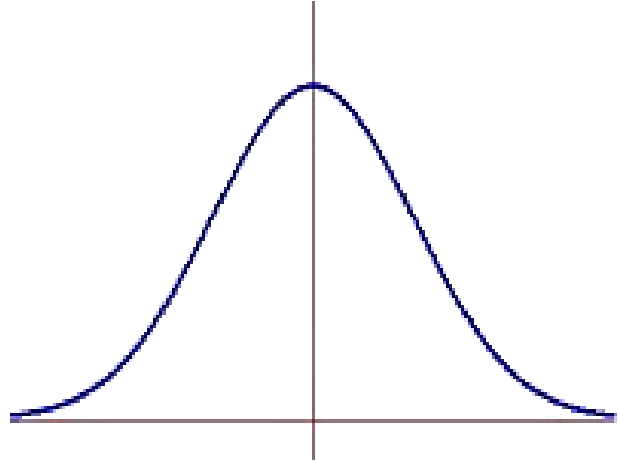
Antoni Ruiz Bueno_2018

Estrategias para responder a una pregunta de ESTADÍSTICA:

Nº de variables implicadas en la interrogación:



1.- Para resumir una distribución se debe responder a lo siguiente:

TIPOS DE INDICADORES	INDICADORES	CÁLCULO	INTERPRETACIÓN																		
<p><u>Indicadores de posición:</u></p> <p>Cómo se encuentran los valores ordenados</p>	<p>Centiles</p> <p>Percentiles</p> <p>Deciles</p> 	$C_i = li + \frac{\frac{i \times n}{100} - fa_{i-1}}{fi} \times i$	<p>Quiere decir que hay un porcentaje por encima o debajo del valor del índice de puntuaciones. Indican el lugar que le corresponde a un valor concreto respecto al grupo.</p> <p>Ejemplo: Si obtenemos el valor del percentil 80 de una distribución, a partir del valor obtenido por encima se encuentra el 20% de las puntuaciones.</p> <p>¿Qué porcentaje de alumnos están aprobados (valor que se debe obtener para aprobar). Porcentaje de suspendidos.</p>																		
<p><u>Indicadores de tendencia central:</u></p> <p>Dónde se concentran los valores</p>	<p>Media aritmética </p> <p>Mediana </p> <p>Moda </p>	<p>Es el valor promedio. El valor que representa a todos los demás.</p> <p>Es aquel valor que ocupa el valor central de la distribución. Que es lo mismo que el percentil 50, el Cuartil 2 y el decil 5 (previa ordenación creciente de los valores)</p> <p>Valor que se presenta con mayor frecuencia en una distribución. Puede ser unimodal, bimodal o multimodal</p> <p>Statistics</p> <table border="1"> <tr> <td colspan="3">Edat Edat</td> </tr> <tr> <td>N</td> <td>Valid</td> <td>20</td> </tr> <tr> <td></td> <td>Missing</td> <td>0</td> </tr> <tr> <td></td> <td>Median</td> <td>42,50</td> </tr> <tr> <td></td> <td>Mode</td> <td>32^a</td> </tr> <tr> <td>Percentiles</td> <td>50</td> <td>42,50</td> </tr> </table> <p>a .-Existen múltiples modas. Se muestra el valor más pequeño</p>	Edat Edat			N	Valid	20		Missing	0		Median	42,50		Mode	32 ^a	Percentiles	50	42,50	
Edat Edat																					
N	Valid	20																			
	Missing	0																			
	Median	42,50																			
	Mode	32 ^a																			
Percentiles	50	42,50																			

TIPOS DE INDICADORES	INDICADORES	CÁLCULO	INTERPRETACIÓN
<p><u>Indicadores de dispersión:</u></p> <p>Que variabilidad existe entre los datos</p>	<p>Rango, recorrido o Amplitud (R)</p> <p>Rango intercuartil (RI) o (IQR) →</p> <p>Varianza →</p> <p>Desviación Estándar/Típica →</p> <p>Coefficiente de variación de Pearson. →</p>	<p>Con intervalos: $R = V_{\max} - V_{\min} + 1$ Sin intervalos: $R = V_{\max} - V_{\min}$</p> <p>Es la diferencia entre los cuartiles tercero y primero. Nos da básicamente, información sobre el 50% central de la distribución $RI = Q_3 - Q_1$</p> <p>La varianza es el promedio de las diferencias al cuadrado de cada valor respecto a la media de la distribución</p> <p>Las unidades de la varianza no son las mismas que las de la muestra, ya que estamos elevando las diferencias al cuadrado. Esto implica ciertas dificultades en interpretar la magnitud de la dispersión del grupo. Para superar esta dificultad, disponemos de la desviación típica</p> <p>Es una unidad de medida relativa, no depende de las unidades de medida de la variable, por lo tanto, permite la comparación entre diferentes muestras. Se trata de saber qué media de dos muestras es más representativa, en relación a sus desviaciones típicas</p> $CV = \frac{Sd}{X} \times 100$	
<p><u>Indicadores de forma:</u></p> <p>El tipo de forma que presenta el conjunto de datos de la distribución</p>	<p>Asimetría →</p> <p>Curtosis →</p>	<p>Indica la distribución de los valores en torno a la media aritmética (a su derecha e izquierda) respecto a la moda o mediana.</p> <p>Es el estudio del apuntamiento de la curva en la parte central de la distribución, en relación a la curva normal</p>	<p>La interpretación del resultado es el siguiente: $As > 0,5 \rightarrow$ La distribución es asimétrica a la izquierda (asimetría positiva) $As < -0,5 \rightarrow$ La distribución es asimétrica a la derecha (asimetría negativa) $-0,5 < As < 0,5 \rightarrow$ Se considera simétrica, porque en muy pocas ocasiones el coeficiente será igual a cero.</p> <p>$Cu > 0'363 \rightarrow$ Leptocúrtica (Levantada) $0'263 < Cu < 0'363 \rightarrow$ Mesocúrtica (normal) $Cu < 0'263 \rightarrow$ Platicúrtica (plana)</p>

2.- Probabilidad

Formulas a utilizar:

a.- Para conocer la probabilidad, si se tienen los valores de la variable:

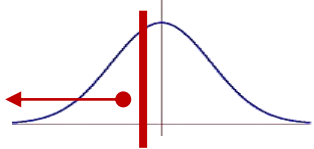
$$Z = \frac{x_i - \bar{x}}{sd}$$

b.- Para conocer los valores de una probabilidad dada:

$$x_i = \bar{x} \pm Z.sd$$

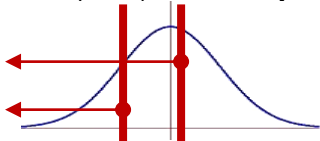
c.- Si nos dan el área o el porcentaje, el proceso consiste en ver la probabilidad en la tabla (página, 8 de este documento) y buscar el valor de Z correspondiente.

c-1) Probabilidad de que, si elegimos un alumno al azar, tenga una puntuación igual o menor que 6,8



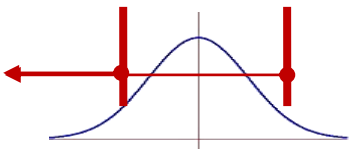
$$Z = \frac{x_i - \bar{x}}{sd}$$

c.2) Id. (Para una probabilidad,) puntuación entre 4,2 y 5,6



$$Z = \frac{x_i - \bar{x}}{sd}$$

c.3) ¿Entre qué valores está comprendido el 95% de la distribución?

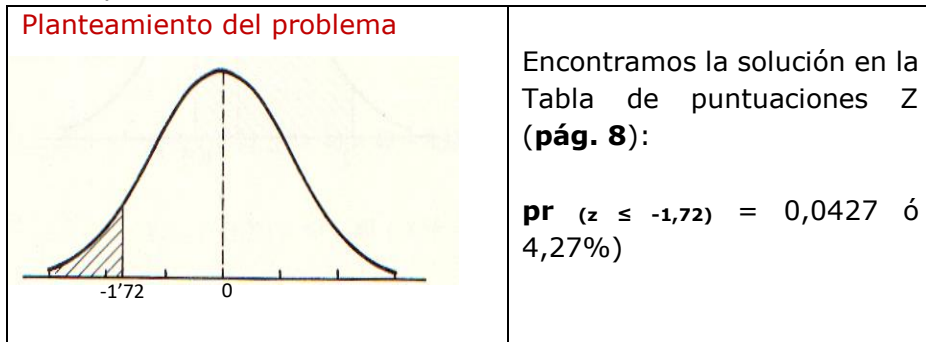


$$x_i = \bar{x} \pm Z.sd$$

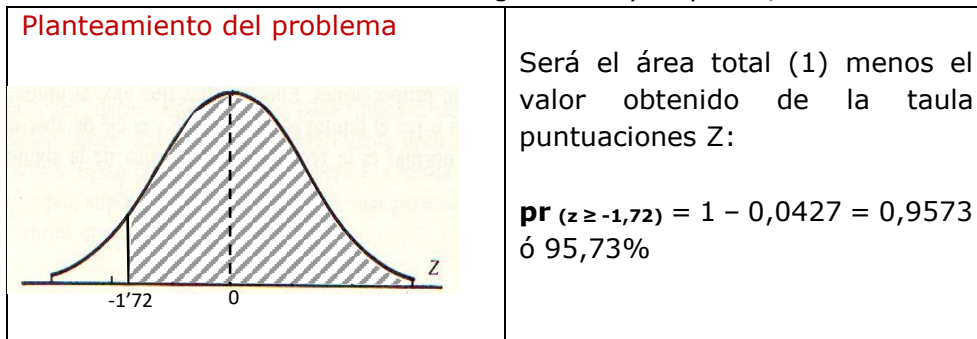
Sd = Desviación Estándar; \bar{x} Media aritmética;
 x_i Valor de la Variable

Ejemplos: (Fuente: Ormazábal, F.J. (coord.) Vila, R.; Mateo, M.; Torrado, M.; Berlanga, V.; el Barrio, J.; Ruiz, A. (2012) Probabilitat i estadística aplicada a l'educació. Barcelona: Universitat de Barcelona. Dipòsit Digital: <http://hdl.handle.net/2445/21384>)

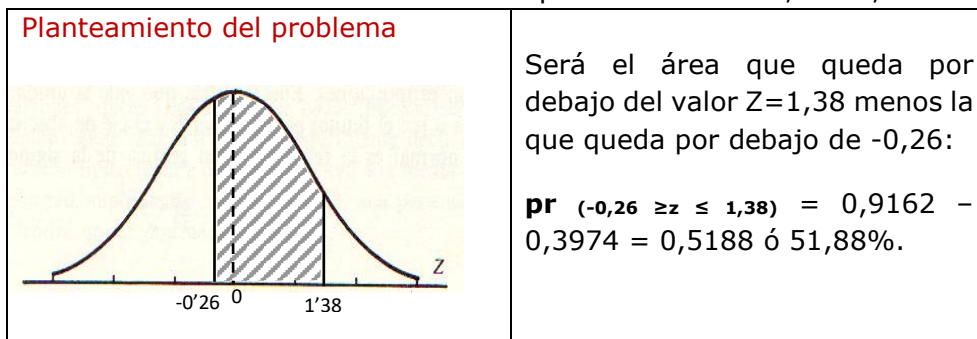
a.- Probabilidad de obtener un sujeto al azar tanga una puntuación **Z** igual o menor que -1,72



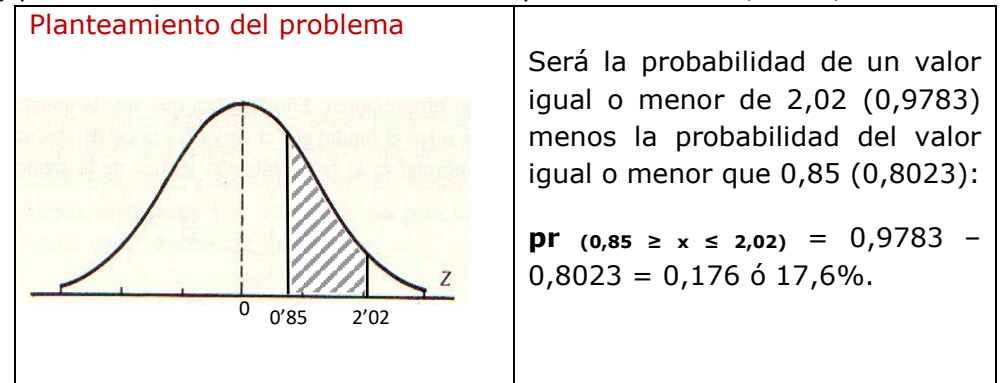
b) Probabilidad de obtener un valor **Z** igual o mayor que -1,72



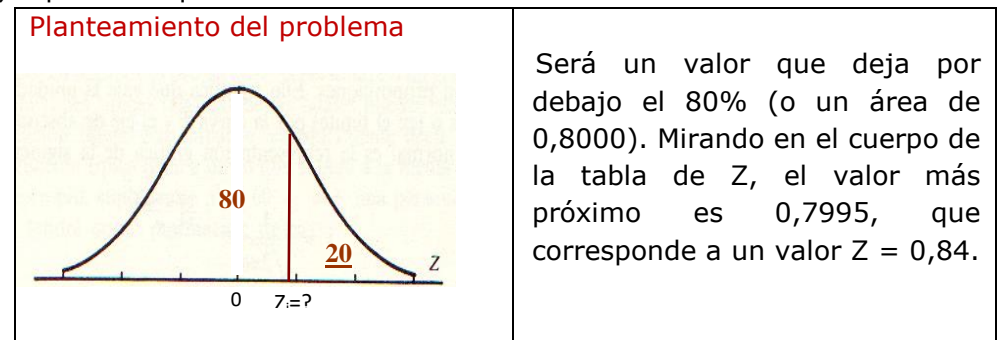
c) Probabilidad de obtener un valor **Z** comprendido entre - 0,26 i 1,38



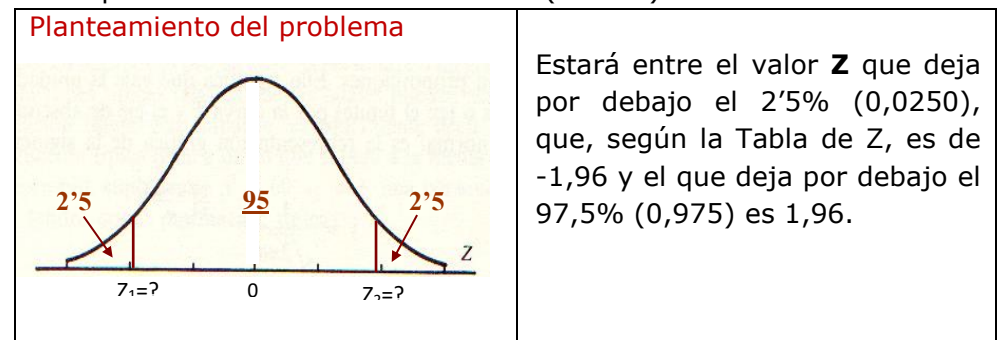
d) probabilidad de obtener un valor comprendido entre 2,02 i 0,85



e) A partir de que valor Z estará el 20% más alto?



f) Entre que valores se encuentra el 95% (central) de la distribución?



2.1 Comprobar si una variable sigue ley normal: (prueba de kolmogorov)

(EN EL SPSS, LA HIPÓTESIS NULA: ES QUE LA VARIABLE SIGUE LEY NORMAL)

1.- Ver cuál es la diferencia máxima obtenida en la muestra (**Diferencia máxima observada**) para la variable. Este dato lo proporciona el output de Spss.

		B1 Difícil	B2 Interessant	B12 Pesada	P3 Edat anys	P4 Temps arribar Universitat
N		108	109	108	109	109
Parámetros normales ^{a,b}	Media	5,60	4,76	4,68	20,13	49,93
	Desviación típica	1,119	1,121	1,259	2,216	22,812
Diferencias más extremas	Absoluta	,269	,183	,176	,282	,127
	Positiva	,166	,183	,130	,282	,127
	Negativa	-,269	-,153	-,176	-,177	-,089
Z de Kolmogorov-Smirnov		2,792	1,907	1,825	2,943	1,324
Sig. asintót. (bilateral)		,000	,001	,003	,000	,060

Diferencia máxima observada

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

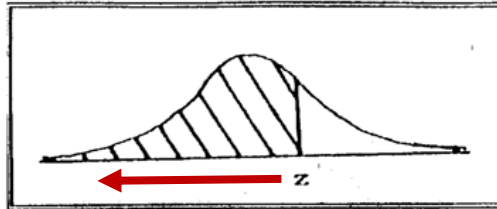
2.- **Cálculo de la diferencia máxima teórica** (en las tablas *LILLIEFORS* – página 9 de este documento- se aplica cuando se desconocen los parámetros) si la muestra es superior a 35, hemos de calcular:

$$\frac{0,886}{\sqrt{n}}$$

3.- **Decisión: En valor absoluto** (sin tener en cuenta el signo (+ ó -))

Sigue Ley Normal
Si la diferencia máxima Observada (la calculada en la muestra) es INFERIOR O IGUAL que la Diferencia máxima Teórica obtenida con la tabla de <i>LILLIEFORS</i> .
Sigue ley Normal Si $ d_i \text{ máx.} \text{ obs.} \leq d_i \text{ máx.} \text{ teórica}$ (INFERIOR O IGUAL)
No sigue Ley Normal
Si la diferencia máxima Observada (la calculada en la muestra) es SUPERIOR que la Diferencia máxima Teórica obtenida con la tabla de <i>LILLIEFORS</i> .
NO Sigue ley Normal $ d_i \text{ máx.} \text{ obs.} > d_i \text{ máx.} \text{ teórica}$ (SUPERIOR)

Ley Normal Estándar



	0	1	2	3	4	5	6	7	8	9
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0005	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0014	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0238	.0233
-1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0570	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4620	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4268	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6113	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9986	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

PROVA DE KOLMOGOROV

Tamany de la mostra (n)	Massey		Lilliefors	
	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
1	0,975	0,995	—	—
2	0,812	0,929	—	—
3	0,708	0,828	—	—
4	0,624	0,733	0,381	0,417
5	0,565	0,669	0,337	0,405
6	0,521	0,618	0,319	0,364
7	0,486	0,577	0,300	0,348
8	0,457	0,543	0,285	0,331
9	0,432	0,514	0,271	0,311
10	0,410	0,490	0,258	0,294
11	0,391	0,468	0,249	0,284
12	0,375	0,450	0,242	0,275
13	0,361	0,433	0,234	0,268
14	0,349	0,418	0,227	0,261
15	0,338	0,404	0,220	0,257
16	0,328	0,392	0,213	0,250
17	0,318	0,381	0,206	0,245
18	0,309	0,371	0,200	0,239
19	0,301	0,363	0,195	0,235
20	0,294	0,356	0,190	0,231
25	0,270	0,320	0,180	0,203
30	0,240	0,290	0,161	0,187
Superior a 35	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

MASSEY Se aplica cuando los parámetros son conocidos.

LILLIEFORS Se aplica cuando se desconocen los parámetros

3.- Estadística Inferencial: Pautas resolución problemas estadísticos

Pasos a seguir y diagrama de decisión

- 1.- Qué tipo de variable presenta el problema (cualitativa o cuantitativa)?
- 2.- Decidir que nos demanda el problema, formula a utilizar y condiciones

Tipo de variable	Qué nos demanda el problema	Formulas a utilizar	Valores de alfa, Z y Condiciones de aplicación								
Cuantitativa	Estimación puntual o De Intervalo (Estima la MEDIA de la población) (Cuantos alumnos se podrían esperar.....,Cuantos sujetos se pueden esperar)	La puntual es el estadístico de la muestra (media aritmética) De intervalo: $\bar{X} \in \mu \pm Z_{\alpha/2} \cdot \sigma_{\bar{x}} = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	Valores $Z_{\alpha/2}$ (bilateral) <table border="1"> <tr> <td>0.10 (10%)</td> <td>0.05 (5%)</td> <td>0.02 (2%)</td> <td>0.01 (1%)</td> </tr> <tr> <td>Z = 1.645</td> <td>Z = 1.96</td> <td>Z = 2.33</td> <td>Z = 2.58</td> </tr> </table>	0.10 (10%)	0.05 (5%)	0.02 (2%)	0.01 (1%)	Z = 1.645	Z = 1.96	Z = 2.33	Z = 2.58
	0.10 (10%)	0.05 (5%)	0.02 (2%)	0.01 (1%)							
Z = 1.645	Z = 1.96	Z = 2.33	Z = 2.58								
Calcular una muestra (Tamaño de la muestra, individuos o sujetos necesarios, tamaño necesario de la muestra etc..)	Poblaciones infinitas (No se conoce N) $e = \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2} \Rightarrow n = \frac{Z^2_{\alpha/2} \cdot \sigma^2}{e^2} \Rightarrow n = \left \frac{Z_{\alpha/2} \cdot \sigma^2}{e} \right ^2$ Poblaciones finitas (Se conoce N) $e = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \cdot Z_{\alpha/2} \Rightarrow n = \frac{N \cdot Z^2_{\alpha/2} \cdot \sigma^2}{Z^2_{\alpha/2} \cdot \sigma^2 + e^2(N-1)}$	N = Es el parámetro de la población. n = Es el estadístico de la muestra									
Cualitativa (categorial)	Estimación puntual o Intervalo (Estima la proporción o el porcentaje ...) (Cuantos sujetos se pueden esperar.....)	La estimación puntual es el porcentaje o la proporción obtenida en la muestra De intervalo $\sigma_{p_o} \in p \pm Z_{\alpha/2} \cdot \sigma_{p_o} = p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n}}$	Condiciones de aplicación: Muestras grandes, Todos los productos n·p, n·q, n·p _o , n·q _o .. deben tener un valor mínimo de 5.								
	Calcular una muestra. (Tamaño de la muestra, individuos o sujetos necesarios, Tamaño necesario de la muestra)	Poblaciones infinitas (Se desconoce N) $e = Z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n}} \rightarrow n = \frac{Z^2_{\alpha/2} \cdot p \cdot q}{e^2} \rightarrow n = p \cdot q \left \frac{Z_{\alpha/2}}{e} \right ^2$ Poblaciones finitas (Se conoce N) $e = Z_{\alpha/2} \cdot \sqrt{\frac{p \cdot q}{n} \frac{N-n}{N-1}} \rightarrow n = \frac{N \cdot Z^2_{\alpha/2} \cdot p \cdot q}{e^2(N-1) + (Z^2_{\alpha/2} \cdot p \cdot q)}$	Si se desconoce P -Se estiman a partir de los datos obtenidos en estudios anteriores. - Si es desconoce p i q - Es realiza un sondeo previo (un piloto) - Se escoge el caso más desfavorable o máximo: p = q = 0'5 (Ya que el producto p·q es máximo: p·q = 0'25) - Se utilizan los valores observados (p _o y q _o).								

μ (Media aritmética); σ (Desviación estándar); $Z_{\alpha/2}$ (Valor Z-bilateral- del error considerado)

e (Precisión o error muestral).

4.- Correlación y Regresión: Pruebas estadísticas según tipo de variables

Coeficiente de correlación	Variable X	Variable Y	SPSS
Pearson (r_{xy})	Cuantitativa continua (o discreta)	Cuantitativa continua (o discreta)	Analizar > Correlaciones > Bivariadas... > Pearson
Spearman (r_s ó ρ_{xy})	Medida con una escala ordinal	Medida con una escala ordinal	Analizar > Correlaciones > Bivariadas... > Spearman
Chi-cuadrado (χ^2)	Cualitativa	Cualitativa	Analizar > Estadísticos descriptivos > Tablas de contingencia... > Estadísticos... > Chi-cuadrado
Biserial (r_b)	Cuantitativa continua (o discreta)	Dicotomizada (cualitativa artificial)	--
Biserial puntual (r_{bp})	Cuantitativa continua (o discreta)	Cualitativa dicotómica	--
Tetracòrica (r_t)	Dicotomizada (cualitativa artificial)	Dicotomizada (cualitativa artificial)	--

ELEMENTS FONAMENTALS DE CORRELACIÓ Y REGRESIÓ:

1.- Correlaci3: Relaci3 entre dos variables de tipo cuantitativo.

Tabla de Interpretaci3 coeficientes de la correlaci3 lnea de Pearson.

Valor coeficiente	Tipo de relaci3	Direcci3
$r = 0$	Relaci3 lineal NULA	Creciente Si $r = 1$ Decreciente Si $r = -1$
$ r = 1$	Relaci3 lineal PERFECTA	Creciente Si $r > 0$ Decreciente Si $r < 0$
$0 < r \geq 0.30$	Relaci3 lineal MUY D3BIL	Creciente Si $r > 0$ Decreciente Si $r < 0$
$0.30 < r \geq 0.7$	Relaci3 lineal D3BIL	Creciente Si $r > 0$ Decreciente Si $r < 0$
$0.70 < r \geq 0.87$	Relaci3 lineal MODERADA	Creciente Si $r > 0$ Decreciente Si $r < 0$
$0.87 < r < 1$	Relaci3 lineal FUERTE	Creciente Si $r > 0$ Decreciente Si $r < 0$

$|r|$ (En valor

En la prueba de correlaci3, La **Hip3tesis nula** de inicio de la prueba es que **No hay relaci3** entre las dos variables.

Ejemplo Interpretaci3, a partir output Spss:

Correlaciones			
		Test verbal (x)	Examen de llengua (y)
Test verbal (x)	Correlaci3 de Pearson	1	,824**
	Sig. (bilateral)	,	,003
	N	10	10
Examen de llengua (y)	Correlaci3 de Pearson	,824**	1
	Sig. (bilateral)	,003	,
	N	10	10

** . La correlaci3 es significativa al nivel 0,01 (bilateral).

Observando la Tabla anterior que nos ofrece el programa Spss, vemos como nos muestra lo siguiente: Es un Tabla simétrica (contiene información del cruce las dos variables), Nos proporciona el tamaño muestra (N), el coeficiente de correlación de Pearson y la significación estadística de la relación entre dos variables (siempre es de dos en dos). Con estos datos podemos decir lo siguiente:

- Como la significación bilateral (,003), es inferior a 0,05, por tanto podemos rechazar la H_0 con un riesgo α de 0,05 (0,01) y aceptamos la H_1 , con lo que podemos concluir que la correlación entre las variables (0,824) es significativa (por tanto, no puede ser explicada únicamente por la influencia del Azar).
- Hay una relación positiva (signo positivo del coeficiente de correlación de Pearson), por consiguiente, a medida que aumenta una variable la otra también aumenta. A mayor puntuación test verbal, mayor puntuación en el examen de lengua. Además, la relación encontrada la podemos calificar de moderada (0,824, en el coeficiente de correlación).

Correlación de Spearman:

Cuando las condiciones de aplicación de Pearson no se cumplen (Nivel de medida de las variables de intervalo o continuas, variables siguen Ley normal, Relación lineal, **Homocedasticidad** (homogeneidad de varianzas), Tamaño muestral grande. La interpretación de los resultados es del mismo tipo que el que se ha lleva acabo con la correlación de Pearson.

2.- Regressió:

Cuando existe correlación entre dos variables cuantitativas, esto nos permite predecir valores de una variable sobre la otra, esto se realiza mediante una prueba de regresión. Por tanto, nos permite predecir el valor de la variable **dependiente** a partir de los valores de la variable **independiente**.

La prueba de regresión nos permite obtener una recta que se denomina de regresión de una variable Y sobre X. Dicha recta se utiliza para estimar o predecir los valores de la variable Y a partir de los valores de la variable X.

$$Y' = a_{yx} + b_{yx} \cdot X$$

Ordenada en el origen

Pendiente de la recta

Variable DEPENDIENTE
Es en la que obtendré un resultado en función de los valores de **X**.

Variable INDEPENDIENTE
Es a partir de la que se realiza la predicción.
Si se conoce el valor de X podré calcular el valor que tendrá Y

Ejemplo: Predecir la puntuación de un **examen** a partir de la nota de un **parcial**.

X es la variable que hace la predicción, la explicativa o la independiente. (Nota parcial)

Y es la variable a predecir, la explicativa o la dependiente (puntuación examen)

Ejemplo interpretación output SPSS: Variables Test Verbal y Examen de Lengua.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,824 ^a	,679	,639	2,505

a. Variables predictoras: (Constante), Test verbal (x)

R = Coeficiente de correlación.

R² = Coeficiente de determinación.

R² corregida = Índice que tiene en cuenta el número de variables introducidas en el modelo.

Error típ. de la estimación = Es el valor de la raíz positiva de la variancia residual.

$$S_{xy} = S_x \sqrt{1 - r_{xy}^2}$$

R²

Es una valoración de la cualidad del ajuste del modelo.

En nuestro ejemplo: el valor es de 0.679. Esto significa que el modelo explica el 67.9 % de la variación de la variable dependiente. En otras palabras, indica que el 67.9% de los cambios que se dan en el examen de lengua es explicada por las diferencias que se dan en el test verbal.

Coefficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	
	B	Error típ.	Beta			
1	(Constante)	a = -9,661	4,222		-2,288	,051
	Test verbal (x)	b = ,958	,233	,824	4,114	,003

a. Variable dependiente: Examen de lengua (y)

Utilizamos la recta de regresión construida a partir de la función de relación lineal (**y = a + bx**) donde X e Y son las variables y **a** (ordenada en el origen) y **b** (pendiente de la recta) **son** constantes.

Si tomamos los valores de la salida que nos proporciona el Spss, tenemos la siguiente función:

$$Y' = -9,661 + 0,958 \cdot X$$

De esta manera, si una persona tiene, por ejemplo, una puntuación en el test verbal, el valor más probable en el examen de lengua es: (Substituyendo la **X**, por el valor 20).

$$Y' = -9,66 + 0,958 \cdot 20 = \mathbf{9,51}.$$

Cálculo del intervalo de regresión (con un determinado error):

$$S_{.xy} = S_x \sqrt{1 - r_{xy}^2} = \mathbf{2.505} \quad (\text{Error típ. de la estimación}) \text{ ver output}$$

Intervalo de confianza será: (tomando un riesgo 0.05).

Recordar aquí que tomaremos la fórmula de la estimación de una variable cuantitativa por intervalo (**ver sección 3, estadística inferencial**)

$$Y \in Y \pm Z_{\alpha/2} \cdot S_{yx} = 9'51 \pm 1'96 \cdot 2'24 = 9'51 \pm 4'39 \prec \begin{matrix} 13'90 \\ 5'12 \end{matrix}$$

Este resultado nos indica que para un individuo que ha obtenido 20 puntos en X (test verbal) lo más probable (Y) es que obtenga 9,51 en Y, pero en el 95% de los casos (con un nivel de confianza del 95 % o un riesgo de error del 5%) su puntuación oscilará entre un máximo de 13'90 y un mínimo de 5'12.

3.- Relación entre dos variables cualitativas (categoriales)

La prueba de Chi-cuadrado de Pearson

Se puede aplicar con cualquier tipo de variables cualitativas (categoriales). No tiene ningún tipo de restricción (no hay una **n** mínima, no es necesario que las distribuciones se ajusten a la curva normal, etc..). El valor del coeficiente de chi-cuadrado indica si hay o no correlación pero no el sentido ni la intensidad de la correlación.

Coeficientes que indican la intensidad de la correlación a partir de chi-cuadrado:		
Tablas de 2 per 2 (V. Dicotómicas) Coeficiente phi	$2 \times 2 \longrightarrow \Phi = \sqrt{\frac{\chi^2}{n}}$	Toma valores en el intervalo: $-1 \leq \phi \leq 1$ Se interpreta igual que el coeficiente de correlación Pearson para variables cuantitativas,
Tallas de k per l (V. politómicas) Coeficiente de contingencia	$k \times l \longrightarrow C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$	

El cálculo de chi-cuadrado tiene una condición de aplicación que consiste en que como máximo un 20% de las frecuencias teóricas (o esperadas) pueden ser inferiores a 5. Si esta condición no se cumple entonces se deben reagrupar alguna categoría y volver a hacer el cálculo de la chi-cuadrado desde el principio.

Si la tabla de contingencia es de 2x2 y hay alguna frecuencia esperada inferior a 5, entonces como que ya no se puede reagrupar ninguna categoría, en vez de consultar la "Chi-cuadrado de Pearson" que aparece en el output se ha consultar la "Corrección por Continuidad". Esto únicamente para las tablas de contingencia de 2 x 2.

El output del Chi cuadrado en Spss:

Los descriptivos de la contingencia:

Tabla de contingencia NIVELL Nivell (curs) * SEXE Sexe

		SEXE Sexe		Total	
		1 Home	2 Dona		
NIVELL Nivell (curs)	1	Recuento	186	136	322
		Frecuencia esperada	202,7	119,3	322,0
		% dentro de NIVELL Nivell (curs)	57,8%	42,2%	100,0%
		% dentro de SEXE Sexe	54,7%	68,0%	59,6%
		% del total	34,4%	25,2%	59,6%
	2	Recuento	154	64	218
		Frecuencia esperada	137,3	80,7	218,0
		% dentro de NIVELL Nivell (curs)	70,6%	29,4%	100,0%
		% dentro de SEXE Sexe	45,3%	32,0%	40,4%
		% del total	28,5%	11,9%	40,4%
Total	Recuento	340	200	540	
	Frecuencia esperada	340,0	200,0	540,0	
	% dentro de NIVELL Nivell (curs)	63,0%	37,0%	100,0%	
	% dentro de SEXE Sexe	100,0%	100,0%	100,0%	
	% del total	63,0%	37,0%	100,0%	

Nos proporciona Tabla descriptiva del recuento del cruce de categorías de las variables. Se muestra la frecuencia, el porcentaje por fila, porcentaje por columna y la frecuencia esperada.

Las pruebas de relación:

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	9,245 ^a	1	,002		
Corrección por continuidad ^b	8,701	1	,003		
Razón de verosimilitudes	9,372	1	,002		
Estadístico exacto de Fisher				,003	,002
Asociación lineal por lineal	9,228	1	,002		
N de casos válidos	540				

a. 0 casillas (0,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 80,74.

b. Calculado sólo para una tabla de 2x2.

En primer lugar se comprueba que el porcentaje de frecuencias esperadas (o teóricas) que hay con valores inferiores a 5: en este ejemplo hay un 0%, por lo tanto se cumple la condición de aplicación.

El valor de Chi-cuadrado es 9.245 y su significación de 0,002; como esta significación es menor que el alfa a 0,05 podemos concluir que el Sexo del alumnado tiene relación significativa con el curso en que están los alumnos.

Ahora bien, si queremos saber cuan intensa es esta relación, nos fijaremos en el coeficiente Phi, que en el ejemplo da -1.131, con una significación estadística de 0,002.

Medidas simétricas

		Valor	Sig. aproximada
Nominal por nominal	Phi	-,131	,002
	V de Cramer	,131	,002
	Coeficiente de contingencia	,130	,002
N de casos válidos		540	

Se encontró una relación estadísticamente significativa, baja y directamente proporcional ($r \phi = 0.131$, $p < 0.05$), entre el sexo y el curso de los estudiantes.