

Examen de la asignatura "Estadística aplicada a las ciencias sociales"

Profesor Josu Mezo. 17 de enero de 2008.

Instrucciones:

- Puedes responder a las preguntas en el orden que quieras, pero indica claramente a cuál estás respondiendo en cada caso.
- En los problemas donde debes hacer un cálculo, explica el procedimiento que sigues, la fórmula que aplicas (expresada matemáticamente o con palabras), los pasos que vas dando al elaborar tu respuesta, etc.. Así podré comprobar que sabes cómo se hacen las cosas (que es lo que importa), aunque, en su caso, puedas cometer algún pequeño error de cálculo.
- El examen tiene 9 preguntas, en cuatro páginas, con una puntuación total de 100 puntos. En cada pregunta te indico su peso en la puntuación total, para que puedas distribuir tu tiempo y atención teniendo en cuenta la importancia relativa de cada una.

Recordatorio de fórmulas (no todas son necesarias)

$$\bar{x}_c = \sum c_i f_i \quad \bar{x} = \frac{\sum x_i}{N} \quad s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \quad s_c = \sqrt{\sum (c_i - \bar{x}_c)^2 f_i}$$

$$ET = \sqrt{\frac{pq}{n}} \quad ET = \frac{s_x}{\sqrt{n}} \quad ET = \frac{\hat{s}_x}{\sqrt{n}} \cdot \sqrt{1-f} \quad ET = \sqrt{\frac{pq}{n}} \times \sqrt{1-f}$$

Pregunta nº 1 (5 puntos).

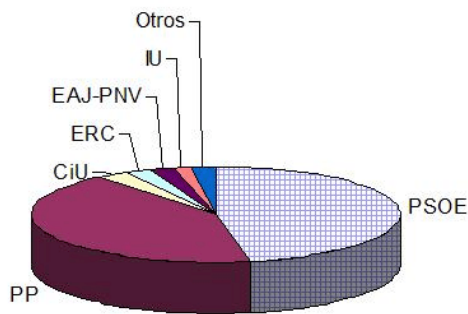
Clasifica las siguientes variables según sean de escala nominal, ordinal o de intervalo, y en el último caso, según sean discretas o continuas (en el sentido “práctico”, no en el sentido teórico de la expresión):

- La opinión de un encuestado sobre la situación económica (muy buena, buena, regular, mala, muy mala). **Escala ordinal**
- El problema que un encuestado considera el más grave al que se enfrenta España en este momento (respuestas: el paro, el terrorismo, la inmigración...). **Escala nominal**
- El número de bombillas de bajo consumo que tiene en uso en su casa el entrevistado **Escala de intervalo; continua**
- El número de conductas ambientales (de una lista de siete) que el entrevistado realiza habitualmente **Escala de intervalo; discreta**
- La edad del entrevistado **Escala de intervalo; continua**

Pregunta nº 2 (10 puntos)

El siguiente gráfico representa a los miembros del Congreso de los Diputados recién disuelto, elegido en ~~2004~~ 1984, según el partido al que pertenecen. Responde a las siguientes preguntas:

- ¿Cómo se llama este tipo de gráfico? **Pictograma**
- ¿Qué aprende el lector a través del gráfico? ¿Qué información transmite? **Lo que se ve es que en el Congreso había dos partidos principales, el PSOE y el PP, y otros cuantos partidos pequeños (CiU, ERC, PNV...). Da la impresión de que el principal partido era el PSOE, que tendría algo menos de la mitad de los diputados, y el PP tenía algunos diputados menos que el PSOE. El conjunto de los partidos pequeños tendría muchos menos diputados que cualquiera de los partidos grandes.**



c) ¿Qué críticas crees que pueden hacerse a la presentación de la información en ese gráfico? ¿Se te ocurren cambios en el mismo u otros tipos de gráficos que podrían presentar mejor la información? **Se puede criticar el gráfico por varias razones. En primer lugar, el efecto 3-D que le ha dado el autor es totalmente inútil, y no ayuda en absoluto a visualizar las proporciones entre los diferentes segmentos. Por otra parte, el gráfico dificulta la visión de la proporción entre los grandes partidos, y de estos respecto a los pequeños.**

Sería probablemente mucho más fácil de interpretar y más informativo un simple diagrama de barras.

Pregunta nº 3 (21 puntos)

La siguiente tabla presenta la distribución de frecuencias absolutas de la variable “Aumento medio de los precios de los pisos entre 2003 y 2007” en los 150 municipios más poblados de España (cifras inventadas). Calcula (utilizando la aproximación basada en las marcas de clase),

Aumento de precios de los pisos 2003-2007 (%)	Número de municipios
0-10	5
10,1-20	23
20,1-30	35
30,1-40	54
40,1-50	25
50,1-60	8

- a) La media
- b) La desviación típica
- c) El coeficiente de variación
- d) La mediana
- e) El primer y tercer cuartil
- f) El rango intercuartílico
- g) La moda

Aumento de precios de los pisos 2003-2007 (%)	Marca de clase (c)	Número de municipios	Frecuencia relativa (f)	Frecuencia acumulada (F)	f*c	(c-media)^2 (A)	A*f (B)
0-10	5	5	0,03	0,03	0,15	683,8	20,51
10,1-20	15	23	0,15	0,18	2,25	260,8	39,12
20,1-30	25	35	0,23	0,41	5,75	37,8	8,69
30,1-40	35	54	0,36	0,77	12,6	14,8	5,33
40,1-50	45	25	0,17	0,94	7,65	191,8	32,61
50,1-60	55	8	0,05	1	2,75	568,8	28,44
Sumas		150	1	1	31,15		134,7

a) La media Es la suma de la multiplicación de las frecuencias relativas por las marcas de clase: 31,15

b) La desviación típica La varianza es la suma de los productos de las desviaciones entre las marcas de clase y la media, multiplicadas por sus frecuencias relativas (columna B). Es decir la varianza es 134,7

Y la desviación típica es su raíz, es decir: 11,6

c) El coeficiente de variación Es la desviación típica partida por la media, es decir: $11,6/31,15=0,372$

d) La mediana Es el valor del caso que ocupa la posición central, una vez ordenados todos los casos de mayor a menor. Tratándose de clases de valores, será la marca de clase que "contenga" la distribución acumulada 0,50. En este caso es la clase 30,1-40, cuya marca de clase es 35.

e) El primer y tercer cuartil. Son las clases (o sus marcas) que tienen la frecuencia relativa acumulada 0,25 y 0,75. En este caso: el primer cuartil es 25 (o la clase de 20,1 a 30) y el tercer cuartil coincide con la mediana, es también 35 (30,1-40).

f) El rango intercuartílico. Es la diferencia entre los valores del tercer y el primer cuartil, es decir, en este caso, 10 (35-25).

g) La moda. Es el valor (o la clase) más frecuente. En este caso coincide con la mediana, ya que la clase más frecuente es la clase 30,1-40, que tiene 54 casos, el 0,36 del total, más que ninguna otra clase.

Pregunta nº 4 (4 puntos)

La media obtenida en la pregunta anterior ¿representaría adecuadamente la subida media de los precios de todos los pisos de esos 150 municipios? Justifica tu respuesta y explica, en su caso, qué herramienta estadística sería útil para averiguar esa subida media de precio, y qué datos serían necesarios para poder utilizarla.

No lo sería necesariamente, ya que en esa media (y en las demás estadísticas del ejercicio anterior), damos el mismo peso a cada municipio, independientemente de su tamaño, y del número de pisos que haya (o se haya vendido) en cada uno.

Por ello, para poder calcular la verdadera media de la subida de los pisos (no la media entre municipios) habría que hacer una media ponderada, en la que supiéramos cuantos pisos hay en cada municipio, o al menos, en cada clase de municipios; o supiéramos al menos qué proporción del total de pisos hay en cada municipio o clase de municipios; de forma que pudiéramos dar diferente peso en la media a los municipios según el número de pisos en cada uno.

Pregunta nº 5 (20 puntos)

El gráfico siguiente representa la evolución de los precios de los pisos en el municipio de Madrid, por trimestres, entre el último trimestre de 2000 y el último trimestre de 2007.

También se representa la línea de tendencia (en línea recta más fina), acompañada de la ecuación que expresa la tendencia ($y = 91,452x + 2048$).

a) Explica la información que obtienes al ver el gráfico sobre la evolución de los precios de los pisos en Madrid



a) Lo que podemos aprender viendo este gráfico es que los precios de los pisos en Madrid entre el último trimestre de 2000 y el último trimestre de 2007 han subido mucho, se han multiplicado por más de dos, ya que al comienzo eran de menos de 2.000 €/m²- y ahora están en torno a los 4.300 €/m².

También podemos ver que esa subida ha sido casi uniforme a lo largo de esos siete años, ya que hasta comienzos de 2007 todos los trimestres el precio era superior al del trimestre anterior. Dentro de esos 6 años de crecimiento, pueden señalarse dos periodos: hasta finales de 2003, con un ritmo de crecimiento muy rápido (pasó de 1800 a 3.500, más o menos, en 3 años) y desde finales de 2003 a finales de 2006, en el que el ritmo de crecimiento fue mucho más suave (de 3.500 a 4.300 en otros tres años). Finalmente, como he señalado, en los últimos trimestres, los que corresponden al año 2007, hay un cambio de tendencia, ya que los precios ya no suben, sino que se han estabilizado o han empezado a bajar ligeramente (quizá de 4.400 a 4.300 o algo similar).

b) Teniendo en cuenta que el trimestre 0 sería el tercer trimestre de 2000; ¿cuál sería la predicción de la línea de tendencia para cuando acabe 2008? ¿Y para cuando acabe 2010?

Si $0 = 3t$ 2000 y $1 = 4t$ 2000

4t 2008 haría el trimestre 33 ((4*8) +1)

Y el 4t de 2010 serían 8 trimestres más, por tanto el t 41

Aplicamos la fórmula y calculamos la predicción:

Para el último trimestre de 2008:

$y = 91,452x + 2048 = 91,452 * 33 + 2048 = 3.018 + 2.048 = 5.066$ €/m² sería el precio previsto para el último trimestre de 2008

$y = 91,452x + 2048 = 91,452 * 41 + 2048 = 3.750 + 2.048 = 5.798$ €/m² sería el precio previsto para el último trimestre de 2010

c) ¿Te parece que las predicciones derivadas de la ecuación se cumplirán? Razona tu respuesta. Es imposible saberlo con certeza, claro, pero más bien es probable que no se cumplan porque vemos que en los últimos trimestres ha habido un cambio de tendencia, de forma que los

precios ya no crecen, o incluso han bajado un poco. Lógicamente, ese cambio de tendencia final no puede ser recogido por la ecuación de la recta de tendencia, que resume los datos de todos los trimestres. Por eso la recta de tendencia sigue prediciendo crecimientos. Si el cambio de tendencia se consolida, y los precios ya no suben, o incluso bajan, a finales de 2008 los precios estarán lejos de los 5.066 €/m² previstos por la recta. Más bien andarían por los 4.200 ó 4.100. Y a finales de 2010, en lugar de ser 5.798€/m², a lo mejor serían 3.900 ó 3.800. Por lo tanto, la predicción de la recta se equivocaría bastante.

Por otra parte, pudiera ser que fuera sólo un cambio temporal, y que los precios volvieran a crecer a ritmos similares en el futuro, y en ese caso quizá la predicción no estuviera del todo mal.

La interpretación sobre si el cambio de tendencia es un “bache” o es algo que se va a consolidar no es algo puramente matemático o estadístico. Es algo que tiene que ver con nuestro conocimiento previo o contextual sobre la variable estudiada. En este caso, por ejemplo, por lecturas de periódicos, por lo que dicen los expertos... todos llevamos tiempo esperando un cambio de tendencia. Por eso es más “sensato” suponer que el cambio de tendencia no es un bache, sino que va a durar. Pero también es cierto que hace años que todo el mundo pensaba que los precios “no podían subir más”. Así que sería posible equivocarse, claro.

Pregunta nº 6 (5 puntos)

Explica con ejemplos qué son los sesgos de respuesta y de no respuesta.

Sesgo de no respuesta: se produce porque algunas personas se niegan a participar en las encuestas; o bien, porque entre los participantes, algunos se niegan a responder a algunas preguntas en particular. Por ejemplo, en las encuestas de temas políticos, hay personas que se niegan a decir a quién van a votar, o a quién han votado. El hecho de que unas personas respondan y otras no, y que esa disposición no esté (seguramente) repartida al azar, hace que los resultados de una muestra se desvíen de los que obtendríamos con una muestra puramente aleatoria.

Sesgo de respuesta: se produce cuando los entrevistados responden, pero no lo hacen de manera sincera. Por ejemplo, porque ocultan opiniones que consideran que son poco populares o políticamente incorrectas (machismo, racismo); o porque responden lo que creen que al encuestador le gustaría oír, según cuál sea el tema de la encuesta...

Pregunta nº 7 (15 puntos)

El tiempo medio que tarda en venderse una vivienda en España es de 112 días, con una desviación típica de 45. Suponiendo que la distribución de esa variable fuera normal, y utilizando, cuando sea necesario, la tabla de probabilidades de los valores de z en una distribución normal estándar, que tienes reproducida al final del examen, calcula

a) ¿Cuál es el periodo de tiempo en el que se venden el 75% de las viviendas?

Tendremos que buscar en la tabla de valores z el valor z tal que el 75% de los casos tienen un valor menor o igual.

El valor más cercano a 0,75 es 0,7486, que corresponde al valor $z=0,67$

El valor $z=0,67$ corresponde a $112+(0,67*45)=142,15$ días.

Por tanto, el 75% de las casas se venderían antes de los 142,15 días

b) ¿Qué proporción de las viviendas se vende en menos de un mes?

Un mes son 30 días.

El valor z correspondiente a 30 días sería:

$$(30-112)/45 = -1,82$$

Buscamos en la tabla cuál es la proporción de casos con un valor de -1,82 o menor. Como la tabla es sólo de valores positivos, buscamos el valor 1,82, y encontramos que la proporción de casos hasta ese valor z es 0,9656. Por tanto, los valores con un valor z mayor que 1,82 serían $(1-0,9656)=0,0344$, y esa sería también, lógicamente, la proporción de los casos con valor z menor que -1,82.

Por tanto, la respuesta es: se vende en menos de un mes un 0,0344 de los pisos, o dicho de otra forma el 3% de los pisos

c) ¿Qué proporción de las viviendas tarda más de 6 meses en venderse?

6 meses son 180 días

$$\text{Convertido en valores } z = (180-112)/45 = 1,51$$

Buscamos en la tabla la proporción de casos con valor MENOR que 1,51 y es: 0,9345.

Pero como nos han preguntado cuántos pisos tardan MÁS de seis meses en venderse, la proporción que buscamos es $(1-0,9345)=0,0655$.

El 0,0655 (ó el 6,55%) de los pisos tardan MÁS DE 6 meses en venderse

d) ¿Cuál es el periodo de tiempo en el que venden la mitad de las viviendas?

Por definición, si es una distribución normal, la media es igual a la mediana.

Por tanto, la mitad de las viviendas se venden en el tiempo medio, es decir en 112 días

e) ¿Qué proporción de las viviendas tarda entre 2 y 4 meses en venderse?

Tenemos que buscar la proporción de viviendas que se vende en los dos primeros meses, la que se vende en los primeros 4 meses, y a la segunda restarle la primera.

$$2 \text{ meses} = 60 \text{ días. Valor } z = (60-112)/45 = -1,16$$

$$4 \text{ meses} = 120 \text{ días Valor } z = (120-112)/45 = 0,18$$

Proporción de casos con valor z hasta -1,16: Buscamos en la tabla la proporción hasta 1,16 y la restamos de uno.

$$\text{Proporción de casos hasta valor } z \text{ 1,16: } 0,8554$$

$$\text{Proporción de casos hasta valor } z \text{ -1,16} = (1-0,8554) = 0,1446$$

Proporción de casos hasta valor z 0,18: lo vemos en la tabla y es 0,5714

Por tanto, la proporción de pisos que se vende entre 2 y 4 meses es $0,5714-0,1446=0,4268$, o el 42,68% de los pisos

Pregunta nº 8 (6 puntos)

En una encuesta a una muestra aleatoria de 1.300 mayores de 16 años, la proporción de ellos que decía que hacía ejercicio o deporte habitualmente era del 38,2%. Calcula:

a) El error típico

b) El intervalo de confianza, con un nivel de confianza del 95,5%, para el porcentaje de los mayores de 16 años hacen ejercicio o deporte habitualmente.

a) El error típico es $RAIZ(p*q/N)$.

Usando el parámetro como aproximación al valor de p, tenemos que la fórmula se convierte en $RAIZ(0,382*0,618/1.300)=$

$$RAIZ(0,236076/1.300) = RAIZ(0,00018159) = 0,0135$$

El error típico es 0,0135 (ó 1,35%)

b) El intervalo de confianza, con un 95,5% de confianza es el valor del parámetro ± 2 veces el

error típico, por tanto $38,2\% \pm(2*1,35)=38,2\% \pm 2,7\% =$ un intervalo entre 35,7% y 40,9%
 Con un 95,5% de seguridad podemos decir que el porcentaje de los españoles mayores de 16 años que hace deporte o ejercicio habitualmente está entre el 35,7% y el 40,9%

Pregunta nº 9 (9 puntos)

En una muestra aleatoria de 500 pisos vendidos en España en 2007 encontramos que el precio medio de venta (por metro cuadrado) fue de 2.892 euros. La desviación típica muestral era de 553 euros. Calcula:

- a) El error típico
- b) El intervalo de confianza dentro del cual estará el precio medio por metro cuadrado de las viviendas vendidas en España en 2007, con un nivel de confianza del 95,5%
- c) Si la muestra no fuera de toda España, sino de una provincia en la que se han vendido 1.300 viviendas ¿cómo se modificarían las respuestas a las dos preguntas anteriores?

a) Error típico: $S_x/\text{raiz}(N) = 553/\text{RAIZ}(500) = 553/22,36 = 24,73$

b) Con un 95,5% de confianza, el intervalo de confianza es la media ± 2 veces el error típico, es decir $2.892 \pm(2*24,73) = 2.892 \pm 49,46 =$ entre 2.842,54 y 2941,46

Con un 95,5 de confianza podemos decir que el precio medio por metro cuadrado de todas las viviendas vendidas en España ese año estaría entre 2.842,54 y 2941,46 euros/metro cuadrado

c) El error típico habría que calcularlo de nuevo, corrigiéndolo teniendo en cuenta la fórmula:
 Error típico = $(S_x/(\text{RAIZ}(N)) * \text{Raiz}(1-f)$

Donde f es N/tamaño de la población

$f = (500/1300) = 0,385$

$\text{RAIZ}(1-f) = \text{RAIZ}(1-0,385) = \text{RAIZ}(0,615) = 0,785$

Error típico sería: $24,73 * 0,785 = 19,41$

Y entonces el intervalo de confianza, con un 95,5% de confianza, sería:
 $2.892 \pm(2*19,41) = 2.892 \pm 38,82 =$ un intervalo entre 2.853,18 y 2.930,82

Áreas bajo la curva normal estándar. Los valores de la tabla que no se muestran en negrita representan la probabilidad de observar un valor menor o igual al valor correspondiente de z

Segunda cifra decimal del valor de z										
z	0.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
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	.6103	.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	.7704	.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	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.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	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.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	.9875	.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	.9987	.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	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998