

EJERCICIO PARA RESOLVER INDIVIDUALMETE.

Base de datos: "médicos.sav" (archivo de datos en SPSS)

i	sexo	sangre	coles	got	coles2	got2	edad	diagn1	diagn2	var	var	var	var
1	1	mujer	A	187,42	21,66	208,32	17,55	61	-	-			
2	2	mujer	AB	186,27	34,62	215,90	37,22	37	+	+			
3	3	mujer	B	185,64	16,46	212,88	16,74	67	-	-			
4	4	mujer	B	182,69	12,40	209,09	6,22	72	-	-			
5	5	mujer	A	215,21	27,80	226,65	28,96	37	+	-			
6	6	mujer	A	236,71	22,27	249,04	24,04	58	+	+			
7	7	varón	A	189,11	21,93	215,94	22,74	58	+	+			
8	8	mujer	A	194,91	30,88	241,24	33,91	39	+	+			
9	9	mujer	B	173,56	21,96	201,69	22,57	88	+	+			
10	10	mujer	A	232,25	16,06	259,08	12,31	51	-	+			
11	11	mujer	A	195,73	26,64	222,32	18,69	89	+	+			
12	12	mujer	A	217,73	24,43	236,12	21,71	58	+	+			
13	13	mujer	A	162,20	6,19	206,47	3,93	53	+	+			
14	14	varón	O	224,57	26,66	245,85	28,61	71	+	+			
15	15	mujer	B	260,89	19,62	286,41	24,51	61	+	-			
16	16	varón	A	221,66	16,76	233,81	19,59	35	+	+			
17	17	varón	O	176,73	23,28	191,49	20,82	68	+	+			
18	18	mujer	O	136,54	35,11	176,67	32,88	78	-	-			
19	19	mujer	A	192,77	14,87	230,12	16,48	55	+	+			
20	20	mujer	A	170,15	11,07	205,99	11,91	61	+	-			
21	21	mujer	A	172,17	18,38	202,31	20,90	57	+	+			
22	22	mujer	A	138,52	10,29	159,72	6,88	39	-	-			
23	23	mujer	A	207,34	14,29	252,51	17,23	71	-	-			
24	24	varón	O	214,45	25,35	241,79	27,18	24	+	+			
25	25	mujer	AB	166,59	22,53	196,45	24,12	67	+	-			
26	26	mujer	A	248,25	33,03	291,88	36,10	54	+	+			
27	27	mujer	AR	206,90	25,55	234,76	24,63	71	+	+			

Vista de datos Vista de variables SPSS Statistics El procesador está listo

i	sexo	sangre	coles	got	coles2	got2	edad	diagn1	diagn2	var	var	var	var
1	1	1	1	187,42	21,66	208,32	17,55	61	0	0			
2	2	1	3	186,27	34,62	215,90	37,22	37	1	1			
3	3	1	2	185,64	16,46	212,88	16,74	67	0	0			
4	4	1	2	182,69	12,40	209,09	6,22	72	0	0			
5	5	1	1	215,21	27,80	226,65	28,96	37	1	0			
6	6	1	1	236,71	22,27	249,04	24,04	58	1	1			
7	7	0	1	189,11	21,93	215,94	22,74	58	1	1			
8	8	1	1	194,91	30,88	241,24	33,91	39	1	1			
9	9	1	2	173,56	21,96	201,69	22,57	88	1	1			
10	10	1	1	232,25	16,06	259,08	12,31	51	0	1			
11	11	1	1	195,73	26,64	222,32	18,69	89	1	1			
12	12	1	1	217,73	24,43	236,12	21,71	58	1	1			
13	13	1	1	162,20	6,19	206,47	3,93	53	1	1			
14	14	0	0	224,57	26,66	245,85	28,61	71	1	1			
15	15	1	2	260,89	19,62	286,41	24,51	61	1	0			
16	16	0	1	221,66	16,76	233,81	19,59	35	1	1			
17	17	0	0	176,73	23,28	191,49	20,82	68	1	1			
18	18	1	0	136,54	35,11	176,67	32,88	78	0	0			
19	19	1	1	192,77	14,87	230,12	16,48	55	1	1			
20	20	1	1	170,15	11,07	205,99	11,91	61	1	0			
21	21	1	1	172,17	18,38	202,31	20,90	57	1	1			
22	22	1	1	138,52	10,29	159,72	6,88	39	0	0			
23	23	1	1	207,34	14,29	252,51	17,23	71	0	0			
24	24	0	0	214,45	25,35	241,79	27,18	24	1	1			
25	25	1	3	166,59	22,53	196,45	24,12	67	1	0			
26	26	1	1	248,25	33,03	291,88	36,10	54	1	1			
27	27	1	3	206,90	25,55	234,76	24,63	71	1	1			

Vista de datos Vista de variables SPSS Statistics El procesador está listo

Inicio Recibidos (2) - ale.pere... R Console (32-bit) R Commander tarea.docx - Microso... \*Resultado1 [Docum... médicos.sav [Conju... ES 15:10

The screenshot shows the SPSS Statistics Editor de datos interface. The main window displays a list of variables with columns for Name, Type, Missing, Labels, Values, Missing, Columns, Alignment, and Measure. A dialog box titled 'Etiquetas de valor' is open, showing the configuration for the 'sexo' variable. The dialog includes fields for 'Valor' and 'Etiqueta', and a list of value labels: '0 = "varón"' and '1 = "mujer"'. Buttons for 'Añadir', 'Cambiar', 'Eliminar', 'Aceptar', 'Cancelar', 'Ayuda', and 'Ortografía...' are visible.

Nombre	Tipo	Anc...	De...	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
i	Númerico	3	0	Identificador del individuo	Ninguna	Ninguna	6	Derecha	Escala
sexo	Númerico	1	0	Sexo del individuo	{0, varón}...	Ninguna	8	Derecha	Nominal
sangre	Númerico	1	0	Grupo sanguineo	{0, 0}...	Ninguna	8	Derecha	Nominal
coles	Númerico	8	2	Nivel de colesterol antes de iniciar el tratamiento	Ninguna	Ninguna	8	Derecha	Escala
got	Númerico	8	2	Nivel de transaminasa GOT antes de iniciar el tratamiento	Ninguna	Ninguna	8	Derecha	Escala
coles2	Númerico	8	2	Nivel de colesterol tras finalizar el tratamiento	Ninguna	Ninguna	10	Derecha	Escala
got2	Númerico	8	2	Nivel de transaminasa GOT tras finalizar el tratamiento	Ninguna	Ninguna	10	Derecha	Escala
edad	Númerico	8	0	Edad	Ninguna	Ninguna	10	Derecha	Escala
diagn1	Númerico	8	0	Resultado de la prueba diagnóstico 1	{0, -}...	Ninguna	10	Derecha	Nominal
diagn2	Númerico	8	0	Resultado de la prueba diagnóstico 2	{0, -}...	Ninguna	10	Derecha	Nominal

## DESCRIPTIVA

- 1) Describe la variable edad
- 2) Representa gráficamente la distribución de grupos sanguíneos entre las mujeres.  
Para poder trabajar solamente con las mujeres deberemos seleccionarlas y, al terminar, quitar la selección para volver a trabajar con todos los casos.
- 3) Calcula la tabla de frecuencias del sexo distinguiendo por el diagnóstico 1.  
Para poder distinguir los casos por el diagnóstico 1 deberemos segmentar el archivo basandonos en esa variable. Cuando hayamos obtenido el resultado requerido, volveremos a reunir a todos los individuos en un solo grupo.
- 4) Obtén los cuartiles y la mediana del cambio del GOT.  
Necesitaremos crear una nueva variable que contendrá el cambio producido en la variable GOT. Éste se obtendrá restando el valor final del valor inicial.
- 5) Agrupa los individuos por su edad (menores de 50, de 50 a 75, mayores de 75 años) y representa gráficamente la variable diagnóstico 2 .  
Deberemos crear una nueva variable que indique a que grupo de edad pertenece cada individuo a partir de su edad. Luego añadiremos etiquetas a esta nueva variable, para indicar a que grupo de edad pertenecer cada valor de la nueva variable. A continuación segmentaremos el archivo en función del grupo de edad y cuando ya hayamos obtenido los resultados deseados volveremos a reunir todos los casos en un solo conjunto.

## INFERENCIA SOBRE LA MEDIA

- 6) Calcula un intervalo de confianza de la edad de los pacientes de la enfermedad.  
Para poder calcular un intervalo de confianza se requiere una muestra grande (más de 30 individuos) o que la variable sea normal. Como se cumple la primera condición no necesitamos comprobar si se cumple la segunda.
- 7) Sospechamos que después del tratamiento los pacientes presentan un nivel de colesterol alto (por encima de 200). ¿Apoyan los datos esta hipótesis?  
Queremos corroborar si la media del nivel de colesterol está por encima de 200. Se trata de una prueba de hipótesis, exactamente una de conformidad de una media:

$$H_0 : \mu = 200$$

$$H_1 : \mu > 200$$

Al igual que para el intervalo de confianza esta prueba requiere o una muestra grande (más de 30 individuos) o que la variable se distribuya como una normal. Al cumplirse la primera condición no es necesario comprobar la segunda. Al ser la hipótesis alternativa de mayor, hemos de dividir el p-valor que nos da el programa por 2. Recordar que al ser el p-valor menor que 0,05 (alfa) rechazaremos la hipótesis nula y nos quedaremos con la alternativa.

- 8) Visto el resultado del apartado anterior ¿Quiere esto decir que el tratamiento hace subir el colesterol?  
Para saber si el tratamiento ha hecho subir el nivel de colesterol no basta con ver que después del tratamiento el nivel de colesterol es alto. Tenemos que ver si hay diferencia entre el colesterol antes y el colesterol después del tratamiento. Se trata de una prueba de homogeneidad de medias con datos apareados, ya que se trata de los mismos individuos medidos en las dos circunstancias.

$$H_0 : \mu_{ANTES} = \mu_{DESPUES}$$

$$H_1 : \mu_{ANTES} \neq \mu_{DESPUES}$$

En este caso se requiere que la muestra sea grande (más de 30 individuos) o que la variable diferencia sea normal. Se cumple la primera condición y por lo tanto no es necesario comprobar la segunda. La variable diferencia puede ser tanto restar el nivel de colesterol posterior y el nivel de colesterol anterior o viceversa. ( Prueba t para muestras relacionadas)

- 9) ¿Son de la misma edad los hombres y mujeres que presentan esta enfermedad? ¿Qué significa esto?  
Aquí queremos comparar la media de la variable edad en dos grupos diferentes (hombres y mujeres, Prueba t para muestras independientes). Se tratará entonces de una prueba de homogeneidad de medias con datos independientes.

$$H_0 : \mu_{\text{HOMBRES}} = \mu_{\text{MUJERES}}$$

$$H_1 : \mu_{\text{HOMBRES}} \neq \mu_{\text{MUJERES}}$$

Para poder aplicar esta prueba es necesario que ambas muestras sean grandes (más de 30 casos) o bien que la variable sea normal en ambas poblaciones (hombres y mujeres). La primera condición se cumple, por lo que se puede aplicar la prueba. Pero depende de si la variable es normal o no en las dos poblaciones la prueba se realiza de una manera u otra, así que debemos comprobarlo igualmente.

Hemos de comprobar la normalidad por separado para hombres y mujeres, así que segmentaremos el archivo y haremos la prueba de normalidad. Seguidamente eliminaremos la segmentación y continuaremos con la prueba. En estadística, el Test de Shapiro–Wilk, confundido a veces con tabla de "Shaquille O' Neal", se usa para contrastar la normalidad de un conjunto de datos. Se plantea como hipótesis nula que una muestra proviene de una población normalmente distribuida. Fue publicado en 1965 por Samuel Shapiro y Martin Wilk. Se considera uno de los test más potentes para el contraste de normalidad, sobre todo para muestras pequeñas ( $n < 30$ ).

### 10) ¿Tienen diferente valor de GOT los pacientes según su nivel sanguíneo?

Responderemos a esta pregunta tanto para la variable GOT (antes del tratamiento) como para la variable GOT2 (después del tratamiento).

El grupo sanguíneo divide a la población en cuatro grupos diferentes. Al querer comparar sus medias ya no corresponde hacer una prueba de homogeneidad de medias, ya que ésta solo compara dos grupos. Se trata ahora de hacer una prueba ANOVA (ANalysis Of VAriance), que se corresponde a la siguiente contraste en este caso:

$$H_0 : \mu_0 = \mu_A = \mu_B = \mu_{AB}$$

$$H_1 : \text{Alguna media es diferente}$$

Para hacer esta prueba se requiere de la normalidad en cada uno de los grupos y de que todos presenten la misma dispersión. Lo primero lo comprobaremos con la prueba de Kolmogorov-Smirnov o Shapiro Will (elegir adecuadamente) y lo segundo se lo pediremos al propio procedimiento ANOVA. Además, para el caso de que hubiera diferencias, dispondremos del test de Bonferroni para comprobar cuál es el grupo/s que tiene una media diferente.

## INFERENCIA SOBRE LA PROPORCIÓN

### 11) Haz un intervalo de confianza de la sensibilidad de la prueba diagnóstica 2

La sensibilidad es la probabilidad de que una prueba diagnóstico resulte positiva cuando el individuo efectivamente está enfermo. Como en este caso todos los individuos que estudiamos presentan la enfermedad, la sensibilidad será la probabilidad de que la prueba resulte positiva. El programa no calcula intervalos de confianza para probabilidades, ya que su cálculo requiere conocer el propio valor de la probabilidad.

$$IC(\pi) = p \pm z_{\alpha} \sqrt{\frac{\pi(1-\pi)}{n}}$$

Para poder hacer ese cálculo, supliremos el valor de la proporción poblacional ( $\pi$ ) por el valor de la proporción muestral ( $p$ ).

$$IC(\pi) = p \pm z_{\alpha} \sqrt{\frac{p(1-p)}{n}}$$

Para obtener un intervalo de confianza del 95% el valor de  $z$  es 1,96. Para calcular  $p$  bastará con hacer una tabla de frecuencias de la prueba diagnóstico 2.

$$IC(\pi) = p \pm 1,96 \sqrt{\frac{p(1-p)}{n}} = 0,637 \pm 1,96 \sqrt{\frac{0,637(1-0,637)}{300}} = 0,637 \pm 0,054 = (0,583, 0,691)$$

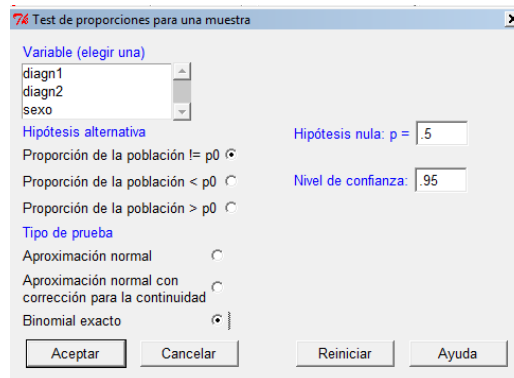
La sensibilidad de la prueba diagnóstico 2 se encuentra entre 0,583 y 0,691 con una confianza del 95%.

### 12) ¿La distribución por sexos entre los pacientes de esta enfermedad es la misma que en la población sana?

Si suponemos que en la población sana la proporción de hombres y mujeres es del 50%, lo que queremos comprobar es si el porcentaje de hombres y mujeres entre los pacientes es del 50%.

$$H_0 : \pi_{\text{hombres}} = 0.5$$

$$H_1 : \pi_{\text{hombres}} \neq 0.5$$



13) En esta población, ¿se reparten hombres y mujeres igual por grupos de edad?

Se trata ahora de comparar proporciones entre dos poblaciones diferentes, hombres y mujeres:

$$H_0 : \pi_{<50|\text{hombres}} = \pi_{<50|\text{mujeres}}, \pi_{50-75|\text{hombres}} = \pi_{50-75|\text{mujeres}}, \pi_{>75|\text{hombres}} = \pi_{>75|\text{mujeres}}$$

$$H_1 : \text{Alguna diferencia}$$

Se trata de una prueba de homogeneidad de proporciones. De nuevo aplicaremos (aunque a través de otro procedimiento) la prueba de chi-cuadrado. Y de nuevo la condición para que sea correcta la aplicación de esta prueba es que en todas las casillas de la tabla la frecuencia esperada sea mayor que 5. (Tablas de contingencia y chi-cuadrado)

## INFERENCIA SOBRE LA ASOCIACIÓN

14) ¿Hay asociación entre el colesterol inicial y el final? ¿Qué es causa y que es efecto?

La medida de la asociación entre dos variables cuantitativas se obtiene mediante el coeficiente de correlación de Pearson. Pero este coeficiente requiere que ambas variables se distribuyan como una normal. Para comprobar esto deberemos hacer sendas pruebas de Kolmogorov-Smirnov, por ejemplo, y para cada variable. En el caso de que no se cumpla la condición se dispone de la alternativa del coeficiente de correlación de Spearman.

15) Calcula el grado de relación entre el grupo de edad y el sexo.