Tema 17 : CONTRASTE DE UNA VARIABLE CUALITATIVA Y OTRA CUANTITATIVA

Se concreta en un contraste de 2 ó más medias. Los datos pueden ser independientes, en los que los problemas de comparación y relación se resuelven por las mismas fórmulas, o apareados, en cuyo caso hay que distinguir muy bien si se trata de una comparación o de una relación, ya que las fórmulas a utilizar son distintas.
Hay que plantearse la pregunta: ¿Me piden que busque si hay diferencias entre los grupos o muestras contrastados o bien si hay una relación, una dependencia entre ellos?. En la tabla guía del tema 15 pueden verse las diversas situaciones que se nos pueden plantear y la forma de abordarlas.

1) La variable cualitativa tiene dos modalidades y los datos son independientes.

Se trata de un contraste de dos medias. Para resolverlos se dispone de una prueba paramétrica ,que llamamos fórmula nº 6, y de otra no paramétrica, la prueba de Mann-Whitney

Fórmula nº 6

$$\mathbf{Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s^2}{N_1} + \frac{s^2}{N_2}}}\ ,\ siendo\ s^2 = \frac{s_1^2(N_1 - 1) + s_2^2(N_2 - 1)}{N_1 + N_2 - 2}}$$

s^2 es la varianza común

H_0: no hay diferencias significativas entre las medias contrastadas ; las diferencias numéricas observadas se explican por el azar.

Condición de aplicación para muestras pequeñas : que el cociente da varianzas, V, obtenido al dividir la varianza mayor por la menor, no supere el valor de referencia de F. Con independencia del orden con que nos den los datos, la muestra nº 1 será la de varianza mayor y la de varianza menor será la nº 2. $V < F(N_1\text{-}1 ; N_2\text{-}1; 0'05)$ .Si no cumple la condición, hay que pasar de oficio a la prueba no paramétrica.

Valoración: si ambas muestras son grandes por la DN ; si alguna es pequeña por $t(N_1+ N_2\text{-}2 ; \propto)$
Si Z < valor de referencia : no puede rechazarse H_0 , no se han encontrado diferencias significativas. (suele escribirse :N.S. ó n.s.)
Si $Z \geq$ valor de referencia ; se rechaza H_0 al nivel de significación probado (y suele escribirse p <0,05 ó p<0,01 ó p< 0,001) y se acepta la hipótesis alternativa, H_1 Hay que dar el sentido.
Recuerdo que, de no decirse lo contrario, si se supera un nivel de siginificación, hay que probar con el siguiente...

Ejercicio 17-1
Se mide la talla en muestras de adultos jóvenes de los pueblos A y B
En A obtenemos: x_1 = 169 cm , s_1 = 5 cm , N=100
En B : x_2 = 166 cm , s_2^2 = 16 cm^2 , N = 80.
¿Puede afirmarse que los de son más altos que los de B?

**Se trata de una prueba de comparación entre una variable CL, PUEBLO, con dos modalidades, A y B, y otra variable CT, TALLA, medida en los individuos de las muestras de A y B. Los datos son independientes. Contraste de dos medias. A resolver en principio por la fórmula nº 6.
** H_0 : no hay diferencias significativas entre las tallas de A y B = los de A no son más altos que los de B
**Las muestras son grandes y no hay condición de aplicación que comprobar.

**Se calcula la varianza común s^2

$$s^2 = \frac{(25 * 99) + (16 * 79)}{100 + 80 - 2} = 21$$

**Se calcula Z

$$Z = \frac{169 - 166}{\sqrt{\frac{21}{100} + \frac{21}{80}}} = 4,36$$

** <u>Valoración</u>: por los valores de c de la DN correspondientes a los niveles de significación habituales

Z = 4'36 es > que c0'05 = 1'96 y también a c0'01 = 2'58 y a c0'001 = 3,30

Por tanto se rechaza H_0 al nivel de significación de 0'001 y se acepta H_1: las tallas no son iguales; hay diferencias significativas entre ellas. <u>Sentido</u> : la media de A es más alta que la de B.

**Y contestando a la pregunta que nos han hecho : Sí

En un examen hay que seguir fielmente los pasos del ejercicio anterior En los siguientes, por ahorro de espacio, se hará de forma más telegráfica

<u>Ejercicio 17-2</u>

En 15 soldados se mide la concentración de la proteína P en la sangre (en mg/dl). En 5, oriundos de la provincia A, obtenemos lo siguiente: 5 , 7 , 6 , 7 , 5. En los 10 restantes, que proceden de la provincia B: 8 , 10 , 11 , 8 , 8 , 7 , 7 , 6 , 7 , 8 . ¿Hay diferencias entre ambas provinicas? ¿Puede decirse que las diferencias se deben a la excelente calidad del agua de B?

Contraste de una Vble. CL , Provincia, con 2 modalidades, A y B, y otra CT, concentración sanguínea de P. Datos independientes.

→ fórmula nº 6 . Al ser muestras pequeñas hay que comprobar si cumple la condición de aplicación.

H_0 : no hay diferencias significativas entre A y B

Como nos dan los datos originales, hay que calcular la media y la varianza de cada grupo.

	Media	Varianza	N
A	6	1	5
B	8	2'22	10

Como la varianza de B es mayor que la de A, la muestra 1 será B y la 2 será A

V=2'22/1=2'22 que es < F(9 ; 4 ; 0'05)= 6'00 y por tanto cumple la condición y podemos seguir

s^2 = 1'84 y

$$Z = \frac{6 - 8}{\sqrt{\frac{1,84}{5} + \frac{1,84}{5}}} = -2,69$$

|Z| > t(13 ; 0'05) = 2,160, por lo que se rechaza H_0 al nivel de significación de 0'05: hay diferencias entre los soldados de A y B ; sentido: los soldados de B tienen la proteína P significativamente más alta.

Y contestando a la otra pregunta: no lo podemos saber....

<u>Ejercicio 17-2 bis</u>

Resolver el problema anterior por una prueba no paramétrica.

La prueba es la **nº 7 , de Mann-Whitney**.
Como prueba no paramétrica no tiene condiciones de aplicación.
Consiste en ordenar a todos los individuos en conjunto, asignándoles un nº de orden. La ordenación se puede hacer de mayor a menor o de menor a mayor. En caso de empate a cada individuo se le asigna la media de los números de orden que habría que repartir entre ellos.
El nº de orden que se adjudica a cada dato se anota en la columna de R que le corresponde. A la suma de las columnas de R, las llamamos, respectivamente, R_1 y R_2

Para aplicar la fórmula se toma para R el valor de la menor de R_1 y R_2 , con su n correspondiente.

$N = n_1 + n_2$.
Se valora por la DN (si $N \geq 30$) ó por $t(N-2 , \alpha)$
Una forma práctica de resolverlo es utilizar una plantilla como la que se ofrece a continuación

Se ordenan todos los datos a la vez

Individ.		**X_1**	**R**		**X_2**	**R**
1						
2						
3						
4						
5						
6						
.....						
n						
		Suma			**Suma**	
			R_1			**R_2**

$N = n_1 + n_2$

$$Z = \frac{R - \frac{n(N+1)}{2}}{\sqrt{\frac{n_1 n_2 (N+1)}{12}}}$$

R es la menor de R1 y R2 ; **n** es el tamaño de la muestra que corresponde a esa R

Prueba de que se han calculado bien las R : **$R_1 + R_2 = N(N+1)/2$**

Si hay diferencias significativas, hay que dar el sentido: la media más alta es la del grupo con R mayor (si hemos ordenado de menor a mayor)

****en el problema propuesto:
Es un contraste de una Vble. CL , Provincia, con 2 modalidades, A y B, y otra CT, concentración sanguínea de P. Datos independientes
H_0 : no hay diferencias significativas entre A y B

Individ.		**X_1 = B**	**R**		**X_2 = A**	**R**
1		**8**	***11'5***		**5**	***1'5***
2		**10**	***14***		**7**	***7***
3		**11**	***15***		**6**	***3'5***
4		**8**	***11'5***		**7**	***7***
5		**8**	***11'5***		**5**	***1'5***
6		**7**	***7***			
7		**7**	***7***			
8		**6**	***3'5***			
9		**7**	***7***			
10		**8**	***11'5***			
		Suma	***99'5***		**Suma**	***20'5***
			R_1			**R_2**

$N = n_1 + n_2 = 15$

$$Z = \frac{20{,}5 - \frac{5*16}{2}}{\sqrt{\frac{10*5*16}{12}}} = -2{,}388$$

Prueba:
R1+R2: 99,5+20,5 = 120
N(N+1)/2: 15*16/2 = 120

<u>Se valora por</u> t con g.l. de 13 : |Z| > t(13, 0'05) = 2'160
Por tanto se rechaza H_0 al nivel de significación de 0'05 y se acepta H_1 : sí hay diferencias , p<0,05

La prueba no paramétrica, aunque menos potente, también ha logrado descubrir las diferencias
Las preguntas se responden como en el ejercicio anterior

Nota:
** Hay un <u>procedimiento clásico</u> de resolver el Mann-Whitney. Se calculan dos posibles resultados:
$Z_1 = n_1n_2 + n_1 (n_1+1)/2 - R_1$ y $Z_2 = n_1n_2 + n_2 (n_2+1)/2 - R_2$
Se toma como resultado final, Z, el menor de los dos y se compara con un valor de referencia en una tabla especial, la tabla de la U, para tomar la decisión estadística. no vemos aquí este método.
** La fórmula que utilizamos, la nº 7, es válida a partir de un tamaño muestral pequeño, que algunos cifran en 5, y tiene la ventajas obre el procedimiento clásico de poder ser valorada por la DN o la t de Student.
** Hay una <u>variante de nuestra fórmula 7</u>, que tiene en cuenta el menor de Z_1 y Z_2, y se valora también por la DN o la t . Sólo cambia el numerador, que es : Zmenor - $(n_1n_2 / 2)$

---

2) La variable cualitativa tiene más de dos modalidades con datos independientes

<u>Es un contraste de 3 o más medias</u>, cuyo método paramétrico es el <u>análisis de la varianza</u>, más conocido como **ANOVA** (de su nombre en inglés: <u>AN</u>alyisis <u>O</u>f <u>Va</u>riance). Hay varios ANOVAs ; aquí utilizaremos el **ANOVA-1** (también conocido como One Way ANOVA). Se analiza un factor (más adelante se verá el ANOVA-2, que analiza dos factores) utilizando las varianzas.
Se necesitan los datos originales para el cálculo clásico, que es bastante farragoso y que se facilita utilizando la plantilla siguiente :

Muestras →	1		2		3		……		k		
Individuos ↓	**X**	**X^2**	**X**	**X^2**	**X**	**X^2**	**X**	**X^2**	**X**	**X^2**	**Valoración**: **por F(k-1 , N-k , α)**
1											
2											
3											
4											
5											
6											
………											**Si se rechaza H_0 hay que aplicar prueba de Scheffé, de dos en dos, ordenados por su media**
ΣX											ΣΣX = B
$(\Sigma X)^2$											
n											**Σn = N**
$(\Sigma X)^2/n$											**$\Sigma[(\Sigma X)^2/n] = A$**
ΣX^2		$\boxed{}$		$\boxed{}$		$\boxed{}$		$\boxed{}$		$\boxed{}$	**$\Sigma(\Sigma X^2) = C$**
$\bar{X}$											

$$C_A = A - \frac{B^2}{N} = \qquad V_A = \frac{C_A}{k-1} =$$

$$C_T = C - \frac{B^2}{N} = \qquad Z = \frac{V_A}{V_R} =$$

$$V_R = \frac{C_R}{N-k} =$$

$$C_R = C_T - C_A =$$

En la mayoría de los programas estadísticos se utiliza una nomenclatura distinta a la usada aquí :
C_A es llamada variación inter ó entre grupos ("between") , la que procede del objeto de estudio
C_R es llamada variación intra o variación residual (·within") , la que procede de los individuos
C_T es la variación total, suma de las otras dos
Los números suelen ir bajo el epígrafe "suma de cuadrados" o "ssq" ´o "msq"
V_A es la varianza inter ; V_R es la varianza intra En vez de Z ponen F

¿Y si no se conocen los datos originales?
Conociendo la media, la varianza y el tamaño de cada uno de los grupos se pueden calcular sus respectivos ΣX y ΣX_2, por las fórmulas siguientes, que están el página 15 del Formulario:

$$\sum X = n\overline{X}$$

$$\sum X^2 = \frac{s^2 n(n-1) + \left(\sum X\right)^2}{n}$$

y pueden colocarse en su sitio en la plantilla anterior

El ANOVA-1 es una prueba muy robusta y no es preciso comprobar condiciones de aplicación.
Si la prueba lleva a rechazar H_0 , la conclusión es que los grupos, las k medias, difieren entre sí. Pero ésto no quiere decir que estas diferencias existan en todos los casos cuando las tomamos dos a dos. Puede ocurrir que sólo alguna o algunas de las medias sean las responsables de las diferencias. Para averiguar ésto se dispone de varios métodos. El aquí elegido es el **método de Scheffé**, cuya metódica se verá más adelante.

Ejercicio 17-3

A 4 grupos de cobayas se les alimenta con dietas distintas (cada grupo dieta distinta). Al cabo de unos días se comprueba su ganacia de peso en gramos :

Dieta A : 32 , 37 , 34 , 30 , 33
Dieta B : 36 , 38 , 37 , 30 , 34 , 39
Dieta C : 35 , 30 , 36 , 29 , 31 , 29
Dieta D : 29 , 31 , 39 , 39 , 28
Valorar el resultado

*Problema de contraste entre una variable CL, DIETA, con 4 modalidades, A – B – C – D, y otra CT, ganacia de peso.
Los datos son independientes. Por tanto es un contraste de k medias, a resolver por ANOVA-1
H_0 : no hay diferencias significativas entre las medias de los grupos contrastados; las variaciones de las medias se deben al azar
Para los cálculos utilizaremos la plantilla correspondiente

Muestras →	**1 A**		**2 B**		**3 C**		**4 D**		
Individuos ↓	**X**	**X^2**	**X**	**X^2**	**X**	**X^2**	**X**	**X^2**	**Valoración**: por **F (k-1 , N-k , α)**
1	32	1024	36	1296	35	1225	29	841	
2	37	1369	38	1444	30	900	31	961	
3	34	1156	37	1369	36	1296	30	900	
4	30	900	30	900	29	841	30	900	
5	33	1089	34	1156	31	961	28	784	
6			39	1521	29	841			
………									
									Si se rechaza H_0 hay que aplicar prueba de Scheffé, de dos en dos, ordenados por su media
ΣX	166		214		190		148		$\Sigma\Sigma X$ = B **718**
$(\Sigma X)^2$	27556		45796		36100		21904		
n	5		6		6		5		**Σn = N 22**
$(\Sigma X)^2/n$	5511’2		7632’67		6016’67		4380’6		**$\Sigma[(\Sigma X)^2/n]$ = A 23541’33**
ΣX^2		$\boxed{5538}$		$\boxed{7686}$		$\boxed{6064}$		$\boxed{4386}$	**$\Sigma(\Sigma X^2)$= C 23674**
$\bar{X}$	**33’20**		**35’67**		**31’67**		**29’60**		

$C_A = A - \frac{B^2}{N}$ = $\boxed{108'4}$ $V_A = \frac{C_A}{k-1}$ = $\boxed{36'14}$

$C_T = C - \frac{B^2}{N}$ = $\boxed{241'09}$ $Z = \frac{V_A}{V_R} =$ $\boxed{4'90}$

$V_R = \frac{C_R}{N-k} =$ $\boxed{7'37}$

$C_R = C_T - C_A =$ $\boxed{132'67}$

Valoración: por F(3 ; 18 ; ∝) : para 0’05 vale 3’16 y para 0’01 vale 5’09; $Z > F_{0'05}$ y por tanto se rechaza H_0 y se acepta H_1 : hay diferencias significativas entre el conjunto de las medias contrastadas. Esto nos obliga a realizar la prueba de Scheffé, fórmula 8 bis

Prueba de Scheffé

Pasos

1) ordenar las medias, de mayor a menor o viceversa

2 compararlas por parejas, empezando por las màs dispares, las de los extremos

3) aplicar la fórmula 8 bis

$$\boldsymbol{Z_{sch} = \frac{(\overline{X}_i - \overline{X}_j)^2}{V_R(k-1)(\frac{1}{n_i} + \frac{1}{n_j})}}$$

Valoración por $F_{k-1,\, N-k}$

Los datos los tomamos del cálculo del ANOVA-1 . En el numerador están las medias de los dos grupos. En el denominador aparte de V_R están el nº de grupos o muestras (k) y los tamaños de las dos muestras que estamos contrastando (n_i y n_j).

4) la Z obtenida se contrasta con la F de referencia y se toma la decisión estadística

En el problema que nos ocupa:

El orden es:	muestra →	B	A	C	D
	Media →	35’67	33’20	31’67	29’60

*** comparamos B y D

$$Z = \frac{(35'67 - 39'60)^2}{7'37 * 3 * (\frac{1}{6} + \frac{1}{5})} = 4'54$$

Contrastamos Z con F . Es mayor que la F(3 ; 18 ; 0’05)=3’16 y por tanto se rechaza la hipótesis nula y se acepta la alternativa en el sentido de que **B es superior a D**.

*** esto obliga a seguir probando, ahora con B y C

$$Z = \frac{(35'67 - 31'67)^2}{7'37 * 3 * (\frac{1}{6} + \frac{1}{6})} = 2'17$$

Aquí Z es menor que la F de referencia y por tanto **no hay rechazo de H_0**

*** no hace falta probar con B y A, pues nos darán una Z aún más baja

*** sí que hay que probar A y D

$$Z = \frac{(33'20 - 29'60)^2}{7'37 * 3 * (\frac{1}{5} + \frac{1}{5})} = 1'47$$

Z también es menor que la F de referencia y por tanto no hay rechazo de H_0

*** no hace falta seguir probando, ya que las Z que obtengamos serán aún menores.

Conclusión final : La prueba de ANOVA-1 nos dice que las ganancias de peso conseguidas con las cuatro dietas son significativamente distintas en su conjunto. La prueba de Scheffé nos aclara que ello se debe fundamentalmente a la superioridad de B sobre D.

PRUEBA DE KRUSKAL-WALLIS

Como el ANOVA-1 es una prueba muy robusta y no comprobamos condiciones de aplicación, no se nos remite de oficio a la prueba no paramétrica correspondiente, que es la de Kruskal-Wallis.

Esta prueba al ser no paramétrica no tiene condiciones de aplicación. Funciona de forma similar al Mann-Whitney. Se ordenan todos los individuos en conjunto, asignándoles un nº de orden. La ordenación se puede hacer de mayor a menor o de menor a mayor. En caso de empate a cada individuo se le asigna la media de los números de orden que habría que repartir entre ellos.

Una forma práctica de resolverlo es utilizar una plantilla como la que se ofrece en el Formulario y que vemos ahora para resolver el problema anterior por la prueba de Kruskal-Wallis.

Ejercicio 17-3 bis

Resolver el ejercicio anterior por una prueba no paramétrica.

Para las variables de este supuesto la prueba adecuada es la de Kruskal-Wallis

Individ.	**Muestras**							
	1 A		**2 B**		**3 C**		**4 D**	
	x_1	**R**	**x_2**	**R**	**x_3**	**R**	**x.**	**R**
1	32	*12*	36	*17'5*	35	*16*	29	*3*
2	37	*14'5*	38	*21*	30	*7*	31	*10'5*
3	34	*14'5*	37	*19'5*	36	*17'5*	30	*7*
4	30	*7*	30	*7*	29	*3*	30	*7*
5	33	*13*	34	*14'5*	31	*10'5*	28	*1*
6			39	*22*	29	*3*		

***** Se ordenan los datos de todas las muestras en conjunto.**

***** Valoración por χ2 (k-1 , α)**

***** Si se rechaza H_0 hay que aplicar la prueba de Mann-Whitney de dos en dos, ordenados por su T**

						Suma
T = ΣR	66	101'5	57	28'5		
T^2	4356	10302'25	3249	812'25		
n	5	6	6	5	**→ N = Σn →**	**22**
T^2/n	871'2	117'04	541'5	162'25	**→ Σ(T^2/n) →**	**3292'19**

$$Z = \left[\frac{12}{N(N+1)} \sum \left(\frac{T^2}{n} \right) \right] - 3(N+1)$$

Z > χ2(3 ; 0'05)=7'81 y por tanto se rechaza H_0 a ese nivel de significación y la conclusión es que los grupos en conjunto difieren significativamente. Para saber que grupos son los que más contribuyen a estas diferencias se aplica la prueba de Mann-Whitney de dos en dos, ordenados por su T

El orden es B -A–C–D. Se empieza comparando los grupos más dispares y se sigue así en orden decreciente.

Resumiendo:

	R_1	R_2	n_1	n_2	Z	t(N-2 ; 0'05)	¿significativo?
B y D	49	17	6	5	-2'37	2'262	si
B y C	49	29	6	6	-1'60	2'228	no
....							no
A y D	38	17	5	5	-2'19	2'306	no

La conclusión es la misma que en la prueba de Scheffé: las diferencias se deben fundamentalmente a la superioridad de B sobre D

3) La variable cualitativa tiene dos modalidades y los datos son apareados.

Se trata de un contraste de dos medias. Al ser los datos apareados hay que distinguir muy bien si es un problema de **comparación**, en cuyo caso se toman las fórmulas 10 u 11, o bien si es un problema de **relación**, a resolver por las fórmulas 14 ó 15

3-a : problema de comparación

Primero hay que calcular las diferencias entre los pares de valores y luego calcular la media y la varianza de estas diferencias (para la varianza necesitamos también los cuadrados de las diferencias). Con ello ya se puede aplicar la fórmula nº 10

$$Z = \bar{X}_d \sqrt{\frac{N}{s_d^2}}$$ **Valoración: muestra grande por DN; si pequeña por t_{N-1}**

H0 : no hay diferencias entre los datos comparados

Es útil disponerse una tabla auxiliar cuyos encabezados sean: X Y x_d x^2_d

Ejercicio 17-4

Probamos el efecto de un somnífero en 15 personas midiendo las horas que duermen tomándolo y sin tomarlo**.**

Individuos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
horas CON	12	5	13	10	13	10	8	8	7	6	9	8	7	7	5	
horas SIN	8	6	8	6	10	9	4	7	6	6	8	6	9	7	6	¿Es efectivo?

Solución: Problema de contraste de una variable cualitativa , TIPO DE SUEÑO (CON, SIN) y otra cuantitativa, HORAS DE SUEÑO . Datos apareados. Es un problema de comparación, a resolver por la fórmula nº 10. H0 : no hay diferencia entre las horas dormidas en ambas situaciones

Indiv	X	Y	x_d	x^2_d
1	12	8	4	16
2	5	6	-1	1
3	13	8	5	25
4	10	6	4	16
5	13	10	3	9
6	10	9	1	1
7	8	4	4	16
8	8	7	1	1
9	7	6	1	1
10	6	6	0	0
11	9	8	1	1
12	8	6	2	4
13	7	9	-2	4
14	7	7	0	0
15	5	8	-1	1
Suma			22	96

$\Sigma d = 22$, $\Sigma d^2 = 96$, $\bar{X} = 1,47$, $s^2 = 4,552$

z = 1,47 √ 15 / 4,552 = 2,67 que es mayor que t (14, 0'05) = 2,145

Se rechaza H_0 a ese nivel de significación y se acepta H1: hay diferencias significativas entre las horas dormidas tomando y sin tomar el medicamento.

Sentido: tomándolo se duerme más

.Ejercicio 17-4 bis
----Resuelva el ejercicio anterior con una prueba no paramétrica

La prueba no paramétrica es el **test de los signos**.
Se compara el par de valores de cada individuo y se anota un signo (+ , - , 0) según el criterio que se adopte: por ejemplo, "+" si el primer dato es mayor , "-" si es menor y "0" si son iguales. También puede hacerse todo lo contrario, ya que el resultado no variará, pues se toma siempre el signo mayoritario.
Se cuentan los signos "+" y "-". Uno de ellos, cualquiera, se asigna a N_1 y el otro a N_2 . $N=N_1+N_2$
Para la fórmula se toma la la mayor de N_1 y N_2 y para evitar confusiones con las" enes" la llamamos x.

Fórmula nº 11 (Test de los signos)

$$Z = \frac{(2x - N)}{\sqrt{N}}$$

siendo x el mayor de N_1 y N_2
***valorar* por t_{N-1} ó DN (si N ≥ 30)**

	Individuos														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
horas CON	12	5	13	10	13	10	8	8	7	6	9	8	7	7	5
horas SIN	8	6	8	6	10	9	4	7	6	6	8	6	9	7	6
SIGNO	+	-	+	+	+	+	+	+	+	0	+	+	-	0	-

N1 (+) = **10** ; N2 (-) = 3 ; N = 10 + 3 = 13 ; por tanto x = 10 ,
y Z= (2*10 – 13)/ √13 = 1'941 que es < t(12, 0'05)=2'179 y por tanto no se puede rechazar H_0. No se han encontrado diferencias.
Las pruebas no paramétricas son menos potentes que las paramétricas. El test de los signos no ha podido encontrar las diferencias que evidenció la prueba anterior.

Ejercicio 17-5
Un sociólogo quiere investigar si una determinada película sobre la delincuencia juvenil puede cambiar la opinión de las personas adultas de la población X. Para ello estudia una muestra de 100 adultos que han visto la película. Les pregunta si ha cambiado su opinión sobre estos chicos. 15 dicen que siguen opinando lo mismo, 59 los ven con más benevolencia que antes y 26 dicen que los ven peor que antes y que hay que castigarlos con más severidad. ¿Cual es la conclusión?

--- Aquí se puede aplicar el test de los signos, ya que tenemos una opinión después de ver la película, que se contrasta con la que tenían antes de verla.Nos dan los signos ya calculados.
Y tenemos como resultados: 59 + , 26 - , 15 0
Por tanto N_1= 59 , N_2= 26 y N=85 (los 15 que piensan igual no cuentan). X vale pues 59
Ho = no hay cambios de opinión

Z = (2 * 59 - 85) / √85 = 3'58 > $c_{0'001}$ = 3'30 y por tanto se rechaza H_0 a ese nivel de significación. La opinión sobre este asunto ha cambiado significativamente, sobre todo en una mayor tolerancia, pero también, aunque menos, en sentido contrario.

3-b : problema de relación
se resuelven como si ambas variables fueran CT por la fórmula nº 14, que veremos más adelante

4) La variable cualitativa tiene más de dos modalidades y los datos son apareados.

Es un problema de contraste de k medias, que se resuelve por ANOVA-2 (cuyo equivalente no paramátrico es el test de Friedman).
ANOVA-2 permite valorar a la vez dos factores. Dos factores sin repeticiones, ya que hay otros modelos de ANOVA en los que para cada combinación de ambos factores hay más de un dato, “repeticiones”, y que no veremos en esta asignatura.
En muchas ocasiones sólo uno de los factores es interesante. El otro, que suelen ser los individuos, pocas veces es objeto de estudio, ya que se sabe de sobras que los individuos difieren bastante entre ellos. Aunque a veces sí puede ser de interés. En todo caso el análisis conjunto es esencial, pues tiene en cuenta la interacción entre ambos factores. Si no se tiene en cuenta esta interacción, el análisis del factor “principal” puede resultar falseado.
Como siempre H_0 dice que no hay diferencias entre las *k* muestras comparadas ni entre los *n* niveles del otro factor, que suelen ser los individuos.
La decisión estadística se toma tras contrastar Z con una F de referencia.
El ANOVA-2 se puede calcular con más facilidad utilizando la siguiente plantilla

Factor A (muestras)

A→	1		2		3		……		k			
Factor B Individuos↓ o bloques	X	X^2	X	X^2	X	X^2	X	X^2	X	X^2	ΣX_A	$(\Sigma X_A)^2$
1												
2												
3												
4												
…..												
n												
											↑ $\Sigma\Sigma X = B$ ↓	↑ $\mathbf{\Sigma(\Sigma X_A)^2 = C}$
ΣX_B												
n												← $\Sigma n = kn = N$
$\bar{\mathbf{X}}$												
$(\Sigma X_B)^2$												← $\mathbf{\Sigma(\Sigma X_B)^2 = A}$
ΣX^2												← $\mathbf{\Sigma\, \Sigma X^2 = D}$

$$C_A = \frac{A}{n} - \frac{B^2}{N} = \boxed{} \qquad C_T = D \; \boxed{}$$

$$C_B = \frac{C}{k} - \frac{B^2}{N} = \boxed{} \qquad C_R = C_T - (C_A + C_B) = \boxed{}$$

$$V_A = \frac{C_A}{k-1} = \boxed{} \qquad V_B = \frac{C_B}{n-1} \boxed{} \qquad V_R = \frac{C_R}{(k-1)(n-1)} = \boxed{}$$

$$Z_A = \frac{V_A}{V_R} = \boxed{} \qquad Z_B = \frac{V_B}{V_R} = \boxed{}$$

Valoración de A : por F(k-1 ; (k-1)(n-1)). **Valoración de B** : por F(n-1 ; (k-1)(n-1))

El Anova-2 es una prueba muy robusta, por lo que no comprobamos condiciones de aplicación. De oficio no se nos planteará utilizar la prueba no paramétrica correspondiente, que es el test de Friedman.

En el test de Friedman también es conveniente utilizar una plantilla para hacer los cálculos. Esta plantilla tal cual está diseñada sirve para valorar el factor A, muestras. Si se quiere valorar el otro factor, que llamaremos B, intercambiaremos A y B. Es decir, A lo que antes llamábamos "A" le ponemos el nombre de "B" y viceversa. Los datos se introducen ahora en un orden distinto. Y así podremos estudiar lo que inicialmente era "B".
La prueba de Friedman se valora por Chi-cuadrado, con grado de libertad k-1 (muestras-1).

Si se rechaza H_0 hay que aplicar la prueba de los signos. Se ordenar las muestras, de mayor a menor o viceversa
Y se comparan por parejas, empezando por las más dispares, las de los extremos, de forma similar a como veíamos en el Kruskal—Wallis.

A continuación viene un ejercicio que se resolverá tanto por el ANOVA-2 como por la prueba no paramétrica de Friedman.

Ejercicio 17-6

Queremos probar dos productos estimulantes de la memoria, M1 y M2. Diez personas toman en un orden establecido por el azar M1 , M2 y P (placebo) y cada vez se hace un test de memoria. Se obtienen las siguientes puntuaciones:

M1	M2	P
30	31	26
29	21	19
36	35	37
33	32	27
34	31	26
32	29	30
31	38	35
39	21	14
32	23	19
29	26	29

¿Que producto es el mejor?

**Es un problema de contraste de una Vble. CL, PRODUCTO, con 3 modalidades, M1, M2 y P, y otra CT, que es la PUNTUACION en el test de memoria, que se ha obtenido en cada una de estas tres modalidades. Los datos son apareados. La prueba correspondiente es ANOVA-2. Pero a efectos didácticos se resolverá también por el test de Friedman.

H_0 : no hay diferencias entre las 3 muestras comparadas ni entre los 10 niveles del otro factor, los individuos. En este problema el factor interesante son los productos.

Resolución por ANOVA-2 : Utilizaremos la plantilla de que disponemos.

ANOVA-2 Factor A (muestras)

A→	1		2		3			
Factor B Individuos↓ o bloques	X	X^2	X	X^2	X	X^2	ΣX_A	$(\Sigma X_A)^2$
1	30	900	31	961	26	676	87	7569
2	29	841	21	441	19	361	69	4761
3	36	1296	35	1225	37	1369	108	11664
4	33	1089	32	1024	27	729	92	8464
5	34	1186	31	961	26	676	91	8281
6	32	1024	29	841	30	900	91	8281
7	31	961	38	1444	35	1225	104	10816
8	39	1521	21	441	14	196	74	5476
9	32	1024	23	529	19	361	74	5476
10	29	841	26	676	29	841	84	7056
							874	77844
							↑ $\Sigma\Sigma X$= B ↓	↑ **$\Sigma(\Sigma X_A)^2$= C**
ΣX_B	325		287		262		874	
n	10		10		10		30	← Σn=kn= N
$\bar{X}$	32’5		28’7		26’2			
$(\Sigma X_B)^2$	105625		82369		68644		256638	← **$\Sigma(\Sigma X_B)^2$= A**
ΣX^2		10653		8543		7334	26530	← **$\Sigma\,\Sigma X^2$= D**

$$C_A = \frac{A}{n} - \frac{B^2}{N} = \boxed{201'2667} \qquad C_T = D - \frac{B^2}{N} = \boxed{1067'4667}$$

$$C_B = \frac{C}{k} - \frac{B^2}{N} = \boxed{485'4667} \qquad C_R = C_T - (C_A + C_B) = \boxed{380'7333}$$

$$V_A = \frac{C_A}{k-1} = \boxed{100'6335} \qquad V_B = \frac{C_B}{n-1} \boxed{53'9407} \qquad V_R = \frac{C_R}{(k-)(n-1)} = \boxed{21'1519}$$

$$Z_A = \frac{V_A}{V_R} = \boxed{\mathbf{4'76}} \qquad Z_B = \frac{V_B}{V_R} = \boxed{\mathbf{2'55}}$$

Valoración de A : por F(k-1 ; (k-1)(n-1)). **Valoración de B** : por F(n-1 ; (k-1)(n-1))

Sólo nos interesa valorar el factor A, los 3 productos : $Z_A > F(2 ; 18; 0'05)=3'65$ y por tanto se rechaza H_0 a ese nivel de significación: en su conjunto las 3 muestras se comportan de manera significativamente distinta.
Esto nos obliga a realizar la prueba de Scheffé, fórmula 8 bis

El orden es	muestras	1	2	3
	Medias	32’5	28’7	26’2

Comparando 1 y 3 : $Z_{SCH} = (32'5-26'2)^2 / 21'1519*2*0'2=4'69 > F_{0'05}$ .Rechazo de H_0 : 1 es mejor que 3
Comparando 1 y 2 : Z_{SCH} =1’70 <$F_{0'05}$ y no hay rechazo de H_0
Comparando 2 y 3 : Z_{SCH} =0’74 <$F_{0'05}$ y no hay rechazo de H_0
<u>Conclusión final</u> : La prueba de ANOVA-2 nos dice que las puntuaciones de memoria son significativamente distintas en su conjunto. La prueba de Scheffé nos aclara que ello se debe fundamentalmente a la superioridad del producto 1 sobre el 3.

Ahora, el mismo ejercicio resuelto por la prueba no paramétrica, utilizando su plantilla

TEST DE FRIEDMAN

Valoración del factor A

los datos se ordenan por filas

Factor A (muestras)

A→	1		2		3	
B↓ Individuos o bloques	X	R	X	R	X	R
1	30	2	31	3	26	1
2	29	3	21	2	19	1
3	36	2	35	1	37	3
4	33	3	32	2	27	1
5	34	3	31	2	26	1
6	32	3	29	1	30	2
7	31	1	38	3	35	2
8	39	3	21	2	14	1
9	32	3	23	2	19	1
10	29	2'5	26	1	29	2'5
Σ R	25'5		19		15'5	
(Σ R)^2	650'25		361		240'25	

$\boxed{1251'5}$ **Σ (Σ R)^2**

Fórmula: $$\mathbf{Z = \frac{12\sum(\sum R)^2}{nk(k+1)} - 3n(k+1)}$$

Valoración de A : por $\chi 2$ con g.l. k-1

Z=5'15 < $\chi 2$ (2 ; 0'05)=5'99 y por tanto no hay rechazo de H_0 . La prueba no paramétrica, menos potente, no ha podido descubrir las diferencias que sí encontró el ANOVA-2

Valoración de B : (aquí no interesa) ; si interesara, se intercambian los nombres de A y B, es decir, que lo que antes era A pasa a ser B y viceversa y se ponen los datos en la tabla