

PRUEBAS NO PARAMETRICAS

Hasta el momento nos hemos dedicado a la estadística paramétrica, que exige la estimación de parámetros y de la comprobación de supuestos sobre las distribuciones de las variables, como por ejemplo que se distribuyan normal.

Note que la distribución normal está relacionada con las distribuciones t-student, chi-cuadrado y F entre otras.

CUANDO UTILIZAR PRUEBAS NO PARAMETRICAS

- *Cuando no se cumplen los supuestos como*
 - » *Normalidad*
 - » *Tamaños mínimos de muestra*
 - » *Numero igual de elementos en cada muestra*
 - » *Homogeneidad de varianza, etc*
- *Cuando se usan tamaños de muestra pequeños*
 - » *Menores a 30 que no permiten comprobar supuestos sobre la población.*
- *Cuando se convierten datos cualitativas (escalas nominales u ordinales) a información útil para la toma de decisiones (escala de intervalo)*
 - » *Utilizado en estudios mercadeo para medir variables como gustos, satisfacción, nivel de necesidad etc.*

- **VENTAJAS**
 - *Son fáciles de usar*
 - *No se requieren comprobar supuestos*
 - *Se pueden usar con muestras pequeñas*
 - *Se pueden usar con variables cualitativas*
- **DESVENTAJAS**
 - *Ignoran información*
 - *No son tan eficientes como las pruebas parametricas, tienen menor potencia.*
 - *Llevan a una mayor probabilidad de cometer error tipo II (no rechazar H_0 falsa)*

PRUEBAS NO PARAMETRICAS

- *Prueba Chi-Cuadrado X^2 de Independencia*
- *Prueba Chi-Cuadrado X^2 de Bondad de Ajuste*
- *Prueba de Signos*
- *Prueba de Rachas*
- *Prueba Wilcoxon*
- *Prueba de Mann-Whitney*
- *Prueba de Kruskal-Wallis*
- *Correlación de Rangos de Sperarman*

Prueba de Signos

No. de grupos de datos

1 o 2

Variable dependiente

En escala al menos ordinal

Objetivo

Esta prueba puede ser utilizada para determinar si la diferencia entre el numero de veces que los datos caen a un lado de la media verdadera es significativamente diferente al número de veces que cae en el otro lado .

Determinar si la diferencia entre el numero de veces en que el valor de una variable es mayor que el de la otra y el numero de veces que es menor es estadísticamente significativa.

Versión no parametrica de la prueba t para una muestra o de la prueba t para muestras pareadas.

Esta prueba se realiza sobre la mediana de los datos Me

- **Ejemplo 1 :**

Carlos y Ángela, Ingenieras de planta de una firma de artículos deportivos tienen la creencia que el deporte afecta la imagen que cada persona tiene de si misma. Para investigar esta posibilidad eligieron a 18 personas de manera aleatoria, para participar en un programa de ejercicios. Antes de empezar el programa las personas respondieron un cuestionario para medir su propia imagen.

Un nivel de 15 puntos en la prueba establece que la persona tiene un concepto indiferente frente a la afirmación, valores menores de 15 que la afecta en forma negativa y valores por encima de 15 que afectan su imagen en forma positiva. Los siguientes son resultados obtenidos :

16	15	12	17	18	14	16	14	16
17	19	16	14	21	20	16	16	16

- **Ejemplo1 (Continuación)**

Hipótesis Nula $H_0 : Me = 15$

Hipótesis Alterna $H_a : Me \neq 15$ ó $H_a : Me > 15$

Estadístico de Prueba :

M^+ : Número de signos positivos

12	14	14	14	15	16	16	16	16	16	17	17	18	19	20	21
-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+

$$M+ = 13$$

Regla de Decisión (Distribución Binomial $p=0.5$ $n=17$)

$$P(X \geq 13) = 1 - 0.9755 = 0.0245 \quad (\text{valor } p)$$

Tabla B Distribución Binomial – Bioestadística – Wayne W. Daniel psg. A-22

Conclusión :

Se rechaza la hipótesis nula, se acepta la hipótesis alterna. Se puede afirmar que las personas tienen la creencia que el deporte afecta positivamente su imagen.

Ejemplo2 : Prueba de Signos para dos poblaciones

Santiago, investigador del C.I.P.U.J. esta interesado en determinar si la acupuntura afecta la tolerancia al dolor. Realiza un experimento en el cual elige al azar a 10 estudiantes universitarios. Cada uno de ellos es utilizado en dos condiciones y recibe un choque eléctrico de corta duración en la pulpa de un diente, la intensidad del choque eléctrico es corta pero produce un dolor moderado al sujeto. Después de cada choque percibido, cada individuo califica el nivel de dolor percibido en una escala de 0 a 10 (10 dolor máximo) Los resultados obtenidos son:

Con acupuntura : 3 4 6 2 3 4 5 7 7 6

Sin acupuntura : 8 9 4 7 9 4 5 10 7 4

Ejemplo 2 : (Continuación)

Los datos obtenidos son :

Con acupuntura : 3 4 6 2 3 4 5 7 7 6

Sin acupuntura : 8 9 4 7 9 4 5 10 7 4

Signo + + - + + + - -

$$H_0: \mu_1 - \mu_2 = 0$$

Se asigna un signo mas cuando la diferencia esta en el sentido que nos interesa (>)

$$H_a: \mu_1 - \mu_2 > 0$$

Estadístico de Prueba $M^+ = 5$

$$M^- = 2$$

$$n = 7 \quad p = 0.5$$

$$P(X \geq 5) = 0.2266 \text{ (valor } p) \text{ Tabla B - pag A-6 Bioestadistica- W. Daniel}$$

Decisión : No se rechaza la H_0 , no existe evidencia en los datos para decir que la acupuntura reduce el dolor.

Cuando n es grande , se puede realizar aproximación de la Binomial a la distribución Normal

$$E[M^+] = \mu_{M^+} = \frac{n(n+1)}{4}$$

$$V[M^+] = \frac{n(n+1)(2n+1)}{24}$$

Estadístico de Prueba

$$Z = \frac{M^+ - \mu_{M^+}}{\sqrt{V[M^+]}} = \frac{M^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$$

Prueba U de Mann-Whitney

- *Numero. de grupos* 2
- *Variable dependiente* *En escala al menos ordinal*
- **Objetivo** *Determinar si la diferencia entre el numero de veces en que el valor de la variable en un grupo es mayor que el otro y el numero de veces en que es menor es estadísticamente significativa*
- *Versión no parametrica de la prueba t para muestras independientes*

Ejemplo 3 :

Los siguientes datos se tomaron de un estudio de comparación de adolescentes sanos (G_1) y adolescentes con bulimia (G_2). El primero esta conformado por 15 estudiantes y el segundo grupo (G_2) por 14 estudiantes, corresponden al consumo diario de calorías.

Los datos obtenidos son los siguientes :

GRUPO I	39 28	10 82	21 30	29 27	11 50	26 19	7 24	12
GRUPO II	68 32	32 61	58 29	16 50	23 64	53 67	55 37	

Nota: Antes de realizar la comparación de medias se debe realizar una prueba chi-cuadrado de bondad de ajuste para determinar si los datos siguen una distribución normal o no. En caso de que los datos no tengan una distribución aproximadamente normal entonces se debe realizar la prueba de signos.

Grupo	X	Rango
1	7	1
1	10	2
1	11	3
1	12	4
2	16	5
1	19	6
1	21	7
2	23	8
1	24	9
1	26	10
1	27	11
1	28	12
1	29	13,5
2	29	13,5
1	30	15
2	32	16,5
2	32	16,5
2	37	18
1	39	19
1	50	20,5
2	50	20,5
2	53	22
2	55	23
2	58	24
2	61	25
2	64	26
2	67	27
2	68	28
1	82	29

Ejemplo3 (continuación)

1. Se ordenan los datos como si se pertenecieran a una sola muestra y se les asigna unos rangos (orden)

En caso de tener números iguales, se suman los rangos correspondientes y el resultado se divide entre el numero de datos iguales, como por ejemplo :

X Rango

29 13

29 14

$$\text{Rango} = (13+14)2 = 13.5$$

Ejemplo 3 (continuación)

2. Se suman los rangos de cada grupo

X	Rango	X	Rango
7	1	16	5
10	2	23	8
11	3	29	13,5
12	4	32	16,5
19	6	32	16,5
21	7	37	18
24	9	50	20,5
26	10	53	22
27	11	55	23
28	12	58	24
29	13,5	61	25
30	15	64	26
39	19	67	27
50	20,5	68	28
82	29		
Suma R1		Suma R2	
162		273	

- *Ejemplo 3 (continuación)*

H_0 : Ambas muestras provienen de la misma población

H_a : Las poblaciones no son idénticas

Estadístico de Prueba

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1 \quad n > 15$$

$$U_1 = 15 * 14 + \frac{15(15 + 1)}{2} - 162 = 168$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum R_2$$

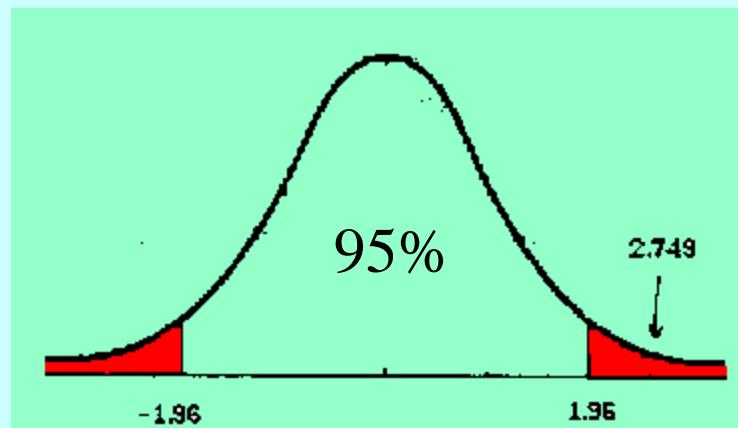
$$U_2 = 15 * 14 + \frac{14(14 + 1)}{2} - 273 = 42$$

$$\mu_u = \frac{n_1 n_2}{2} = \frac{15 * 14}{2} = 105$$

$$\sigma_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{14 * 15(14 + 15 + 1)}{12}} = 22.913$$

$$Z = \frac{u_1 - \mu_u}{\sigma_u} = \frac{(168 - 105)}{22.913} = 2.749$$

Región de Rechazo



Conclusión :

Se rechaza H_0 , se acepta H_a . Se puede concluir que las dos poblaciones son diferentes. por lo tanto las medias son diferentes.

Para analizar cual de los grupos presenta un mayor consumo de calorías, se observa que el Grupo I tiene suma de rangos de 168 y el Grupo II tiene suma de rangos de 273, lo que indica que los datos del Grupo I tienen valores menores en que los del Grupo II

Por lo tanto los estudiantes sanos presentan consumo de calorías menores que los estudiantes que padecen de bulimia.

Prueba de Kruskal-Wallis

Numero de grupos k

Variable dependiente En escala al menos ordinal

Objetivo Determinar si las diferencias entre las medias de los rangos (asignados a las observaciones ordenadas) en los k grupos son estadísticamente significativas

Versión no parametrica del ANOVA

Ejemplo 4

Para verificar si la memoria cambia con la edad, Mónica, investigadora del CIPCE realiza un estudio en el cual hay tres grupos: G1: 60 años G2: 50 años G3: 40 años

Se le presenta a cada persona un grupo de palabras. La serie es presentada dos veces y se cuenta el numero de palabras que pueden recordar.

	G1	G2	G3
1	28	26	37
2	19	20	28
3	13	11	26
4	28	14	35
5	29	22	31
6	22	21	
7	21		

H_0 : Las poblaciones son idénticas

H_a : Las poblaciones no son idénticas

Estadístico de Prueba

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

tiene distribución $\chi^2_{gl:k-1}$

- i. Se ordenan los datos
- ii. Se asignan rangos
- iii. Se suman los rangos por grupo

X
28
19
13
28
29
22
21
26
20
11
14
22
21
37
28
26
35
31

i

Grupo	X	Rango
2	11	1
1	13	2
2	14	3
1	19	4
2	20	5
1	21	6,5
2	21	6,5
2	22	8,5
1	22	8,5
2	26	10,5
3	26	10,5
3	28	13
1	28	13
1	28	13
1	29	15
3	31	16
3	35	17
3	37	18

ii

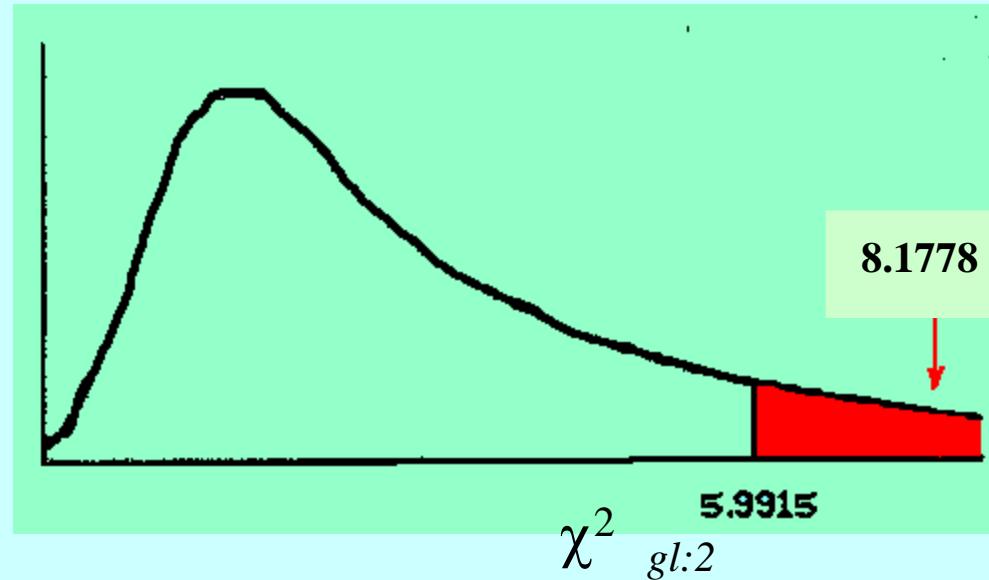
Grupo	X	Rango
1	13	2
1	19	4
1	21	6,5
1	22	8,5
1	28	13
1	28	13
1	29	15
R1		62
2	11	1
2	14	3
2	20	5
2	21	6,5
2	22	8,5
2	26	10,5
R2		34,5
3	26	10,5
3	28	13
3	31	16
3	35	17
3	37	18
R3		74,5

iii

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) \sim \chi^2_{gl:k-1}$$

$$H = \frac{12}{18(18+1)} \sum_{i=1}^3 \frac{R_i^2}{n_i} - 3(18+1)$$

$$H = \frac{12}{18 * 19} \left(\frac{62^2}{7} + \frac{34,5^2}{6} + \frac{74,5^2}{5} \right) - 3 * 19 = 8.1778$$



Conclusión : Se rechaza H_0 , se acepta H_a . Existe evidencia estadística para decir que hay diferencia entre los promedios de los grupos seleccionados

- **Comparaciones Múltiples**

$$\bar{R}_1 = \frac{R_1}{n_1} = \frac{62}{7} = 8.857$$

$$\bar{R}_2 = \frac{R_2}{n_2} = \frac{34.5}{6} = 5.75$$

$$\bar{R}_3 = \frac{R_3}{n_3} = \frac{74.5}{5} = 14.9$$

Valor critico para comparaciones en KW

$$C_k = \sqrt{\chi^2_{\alpha/2, gl:k-1} \left[\frac{n(n+1)}{12} \right] \left[\frac{1}{n_i} + \frac{1}{n_j} \right]}$$

$$C_k$$

$$|\bar{R}_1 - \bar{R}_2| = |8.857 - 5.75| = 3.107$$

$$C_k = \sqrt{5.99 \left[\frac{18.19}{12} \right] \left[\frac{1}{7} + \frac{1}{6} \right]} = 7.27$$

$$|\bar{R}_1 - \bar{R}_3| = |8.857 - 14.9| = 6.043$$

$$C_k = \sqrt{5.99 \left[\frac{18.19}{12} \right] \left[\frac{1}{7} + \frac{1}{5} \right]} = 7.65$$

$$|\bar{R}_2 - \bar{R}_3| = |5.75 - 14.9| = 9.15$$

$$C_k = \sqrt{5.99 \left[\frac{18.19}{12} \right] \left[\frac{1}{6} + \frac{1}{5} \right]} = 7.91 \quad *$$

• *Regla de decisión :*

Se concluye que hay diferencias entre los grupos cuando

$$|\bar{R}_i - \bar{R}_j| > C_k$$

\bar{R}_2	\bar{R}_1	\bar{R}_3
5.75	8.86	14.9

Prueba de Wilcoxon

No. de grupos 2

Variable dependiente En escala al menos ordinal

Objetivo Determinar si la diferencia entre la magnitud de las diferencias positivas entre los valores de las dos variables y la magnitud de las diferencias negativas es estadísticamente significativa.

Versión no paramétrica de la prueba t para muestras pareadas

Ejemplo 5:

Suponga que le interesa analizar los efectos de la transición de la circulación fetal a la circulación postnatal entre niños prematuros. Para cada uno de 12 niños recién nacidos saludables, se midió el ritmo respiratorio en dos momentos diferentes: cuando el niño tiene menos de 15 días de nacido y de nuevo cuando tiene mas de 25 días de nacido.

niño	< 15 días	> 25 días
1	42	40
2	57	60
3	38	38
4	49	47
5	63	65
6	36	39
7	48	49
8	58	50
9	47	47
10	51	52
11	83	72
12	27	33

H_0 : Las poblaciones son idénticas

H_a : Las poblaciones No son idénticas

Para el calculo del estadístico de prueba se requiere calcular

1. Def : Diferencias entre los dos grupos (Antes – Despues)
2. $|Dif|$: Valor absoluto de las Dif.
3. Se ordenan los datos y se les asigna un rango de acuerdo con este orden

niño	< 15 días	> 25 días
1	42	40
2	57	60
3	38	38
4	49	47
5	63	65
6	36	39
7	48	49
8	58	50
9	47	47
10	51	52
11	83	72
12	27	33

nño	< 15 días	> 25 días	Dif	Dif	Rangos
1	42	40	-2	2	6
2	57	60	3	3	8,5
3	38	38	0	0	1,5
4	49	47	-2	2	6
5	63	65	2	2	6
6	36	39	3	3	8,5
7	48	49	1	1	3,5
8	58	50	-8	8	11
9	47	47	0	0	1,5
10	51	52	1	1	3,5
11	83	72	-11	11	12
12	27	33	6	6	10

0	1,5
0	1,5
1	3,5
1	3,5
2	6
2	6
2	6
3	8,5
3	8,5
6	10
8	11
11	12

Estadístico de prueba

$$T^+ : 8.5 + 1.5 + 6 + 8.5 + 3.5 + 1.5 + 3.5 + 10 = 43 \quad \text{Suma rangos positivos}$$

$$T^- : 6 + 6 + 11 + 12 = 35 \quad \text{Suma rangos negativos}$$

$$T = \text{Min.} (T^+, T^-) = 35$$

Tabla para valores críticos de la prueba de Wilcoxon

Pág. .A-86 Tabla K Bioestadistica . Wayne W.Daniel .

Regla de decisión : Se rechazara H_0 si el valor de T es menor o igual que el valor critico de la tabla para $n=15$ y el valor de $\alpha/2= 0.025$

$$n = 15 \quad \alpha = 0.05 \quad T \text{ critico} = 13$$

Conclusión : Como $T_{\text{obtenido}} > T_{\text{critico}}$ ($35 > 13$) no se rechaza H_0 , no existe suficiente evidencia para concluir que existen diferencias entre los ritmos respiratorios de los niños

Ji-cuadrado - prueba de bondad de ajuste

Tipo de variable

Cualitativa

$k > 2$ valores

Objetivo *Determinar si la diferencia entre las frecuencias de cada uno de los valores de la variable y unas determinadas frecuencias teóricas son estadísticamente significativas*

Utilizada para comprobar el supuesto de normalidad, o de otras distribuciones.

Ejemplo 6 :

El Dueño de una panadería tiene la posibilidad de controlar los niveles de inventarios de leche para cuatro marcas diferentes. Con el fin de establecer políticas para la realización de nuevos pedidos requiere saber si la demanda de estas marcas son iguales.

Con este propósito tomo la información de un día

Producto	ventas
Parmalat	33
Uno A	22
Colanta	21
Purace	24
	100

H₀ : La demanda es UNIFORME para las cuatro marcas

H_a : La demanda NO es UNIFORME para las cuatro marcas

Estadístico de prueba :

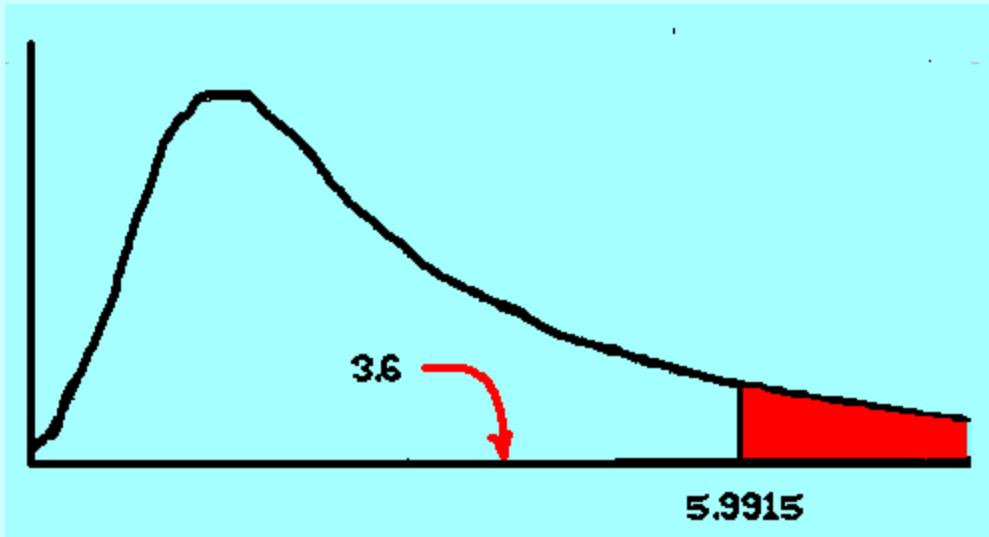
.

$$X^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

tiene distribución χ^2_{k-1}

$$X^2 = \frac{(33-25)^2}{25} + \frac{(22-25)^2}{25} + \frac{(21-25)^2}{25} + \frac{(24-25)^2}{25}$$

$$X^2 = \frac{(64+9+16+1)}{25} = 3.6$$



Conclusión : No se rechaza H_0 , no existe suficiente evidencia para afirmar que la distribución de la venta de leche es diferente a la distribución uniforme (ventas iguales por marca). Se asume que tienen igual demanda.

Queremos ahora comprobar si una variable tiene distribución normal.

Ejemplo 7

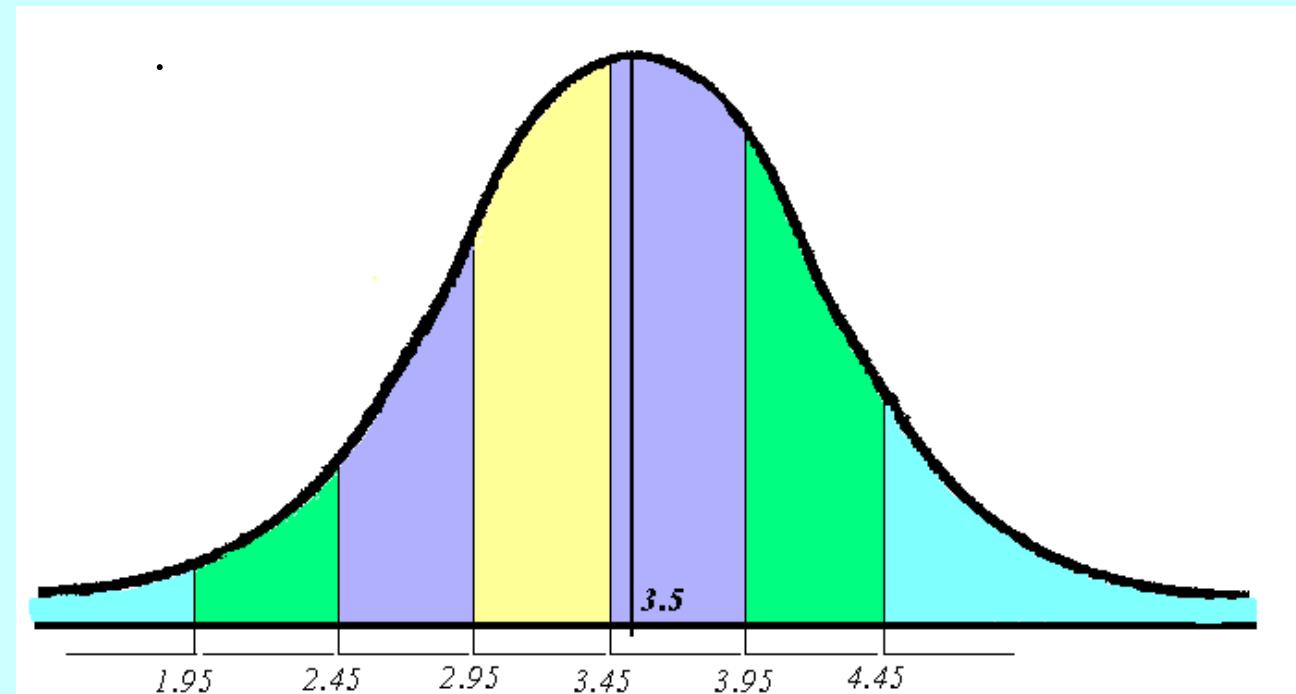
Un director de carrera de una Universidad quiere investigar si los resultados exámenes realizados a los estudiantes de ultimo semestre tienen una distribución normal con media 3.5 y desviación estándar 0.7 puntos.

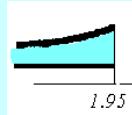
H_0 : *La calificación obtenida tienen una distribución Normal*

H_a : *La calificación obtenida NO tienen una distribución Normal*

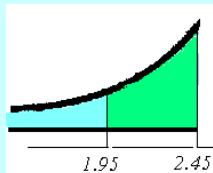
intervalo	frecuencia
1,45 - 1,95	2
1,95 - 2,45	1
2,45 - 2,95	4
2,95 - 3,45	15
3,45 - 3,95	10
3,95 - 4,45	5
4,45 - 4,95	3
Total	40

Calculamos las frecuencias esperadas suponiendo que la distribución de la variable es Normal (H_0 es verdadera) , ubicando los límites superiores en la curva normal estándar y calculando las distribución acumulada

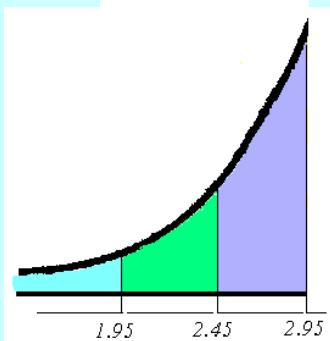




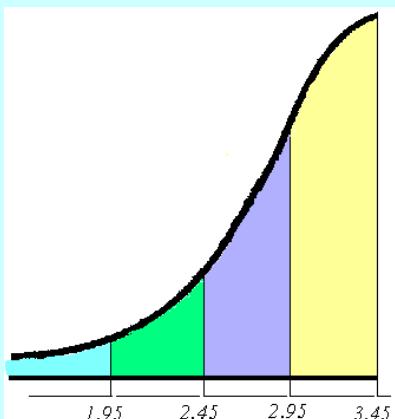
$$P(X < 1.95) = P\left(\frac{X - \mu}{\sigma} < \frac{1.95 - 3.5}{0.7}\right) = P(Z < -2.21) = 0.0136$$



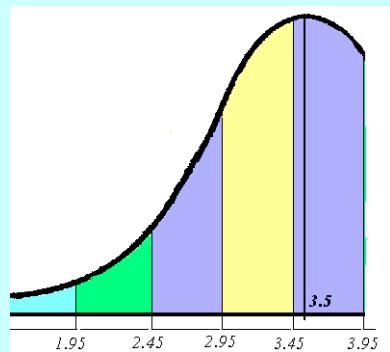
$$P(X < 2.45) = P\left(\frac{X - \mu}{\sigma} < \frac{2.45 - 3.5}{0.7}\right) = P(Z < -1.5) = 0.0668$$



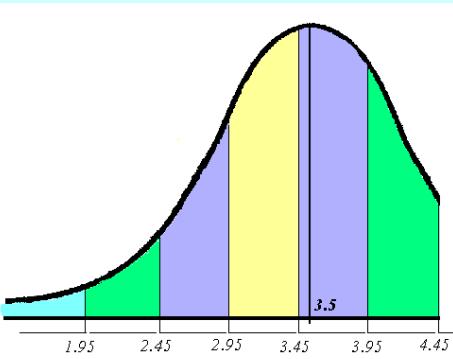
$$P(X < 2.95) = P(Z < -0.79) = 0.2148$$



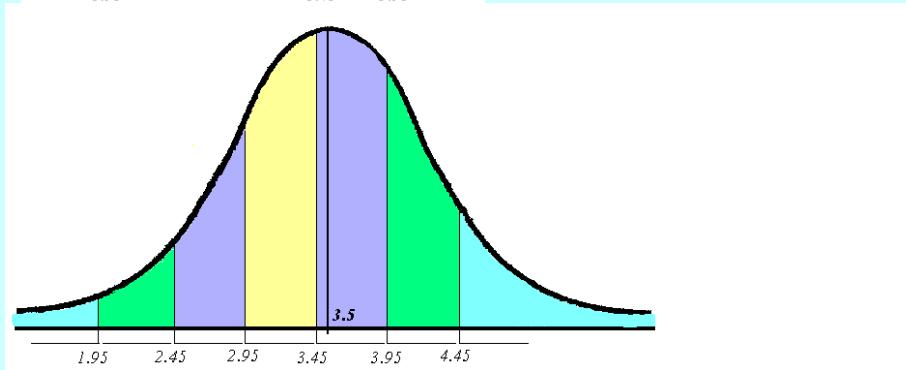
$$P(X < 3.45) = P(Z < -0.07) = 0.4721$$



$$P(X < 3.95) = P(Z < 0.64) = 0.7389$$



$$P(X < 4.45) = P(Z < 1.36) = 0.9131$$



$$P(X < 4.95) = 1$$

intervalo	frecuencia	Z	P(Z<L.sup)
1,45 1,95	2	-2,21	0,0136
1,95 2,45	1	-1,50	0,0668
2,45 2,95	4	-0,79	0,2148
2,95 3,45	15	-0,07	0,4721
3,45 3,95	10	0,64	0,7389
3,95 4,45	5	1,36	0,9131
4,45 4,95	3		1,0000
Total	40		

intervalo	frecuencia	Z	P(Z<L.sup)	Ni
1,45 1,95	2	-2,21	0,0136	0,54
1,95 2,45	1	-1,50	0,0668	0,67
2,45 2,95	4	-0,79	0,2148	8,64
2,95 3,45	15	-0,07	0,4721	18,86
3,45 3,95	10	0,64	0,7389	29,59
3,95 4,45	5	1,36	0,9131	36,51
4,45 4,95	3		1,0000	40,00
Total	40			

Todo valor esperado > 5

Cuando la frecuencia esperada es menor a 5 se juntan intervalos para cumplir con este requisito

.

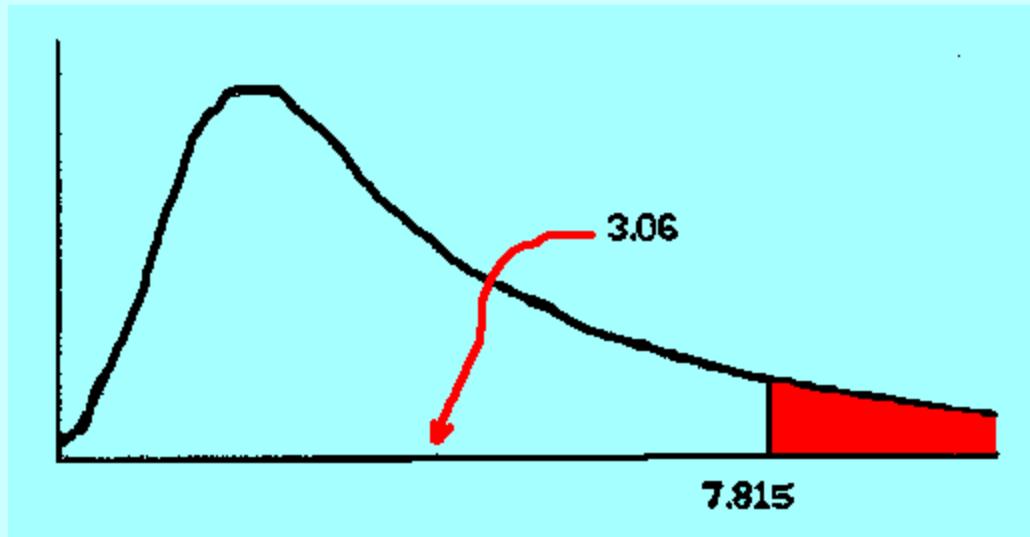
intervalo	frecuencia	Z	P(Z<L.sup)	Ni	f.esperada	
1,45 1,95	2	-2,21	0,0136	0,54	0,544	
1,95 2,45	1	-1,50	0,0668	2,67	2,128	
2,45 2,95	4	-0,79	0,2148	8,59	5,920	8,592
2,95 3,45	15	-0,07	0,4721	18,88	10,292	10,292
3,45 3,95	10	0,64	0,7389	29,56	10,672	10,672
3,95 4,45	5	1,36	0,9131	36,52	6,968	
4,45 4,95	3		1,0000	40,00	3,476	10,444
Total	40					

Estadístico de Prueba

$$X^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

$$X^2 = \frac{(7 - 8.592)^2}{8.592} + \frac{(15 - 10.292)^2}{10.292} + \frac{(10 - 10.672)^2}{10.672} + \frac{(8 - 10.444)^2}{10.444}$$

$$X^2 = 3.06$$



Conclusión

No se rechaza H_0 , no existe suficiente evidencia para decir que los datos no tienen una distribución Normal, Se asume que la distribución de las calificaciones es aproximadamente NORMAL.

Pruebas de Rachas

Tipo de variable

Cualitativa - con 2 valores

Objetivo Determinar si la diferencia entre la secuencia de apariciones de uno y otro valor de la variable y la secuencia de apariciones aleatorias es estadísticamente significativa

Con esta prueba se prueba si un grupo de valores son aleatorios o no. (aleatoriedad en la extracción de una muestra, aleatoriedad en los errores, etc.)

Ejemplo 8

Comprobar que los números aleatorios generados en Excel son realmente aleatorios.

Para generarlos se tomo la función =aleatorio() de la hoja electrónica y la prueba no parametrica de rachas.

H_0 : La muestra es aleatoria

H_a : La muestra no es aleatoria

	Aleatorio	signo
1	0,738	+
2	0,202	-
3	0,357	-
4	0,561	+
5	0,509	-
6	0,146	-
7	0,746	+
8	0,666	+
9	0,133	-
10	0,430	-
11	0,972	+
12	0,999	+
13	0,499	-
14	0,869	+
15	0,821	+
16	0,732	+
17	0,355	-
18	0,189	-
19	0,478	-
20	0,162	-

Se asignan signos mas a los numeros mayores a 0.5 y signo menos a los menores.

Se marcan los cambios de signos y se toma como racha una serie continua de uno o mas signos iguales.

$$n1 = \text{número de signos mas} = 9$$

$$n2 = \text{número de signos menos} = 11$$

$$r = \text{rachas (cambios de “+ a -” ó “- a +”)} = 10$$

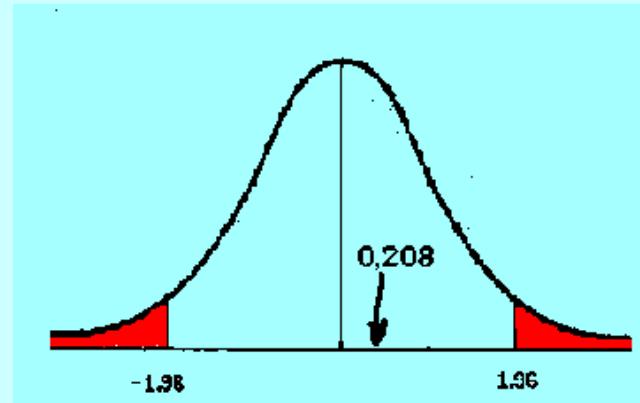
$$\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \frac{2 * 9 * 11}{9 + 11} + 1 = 9.9$$

$$\sigma = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 n_2)^2 (n_1 + n_2 - 1)}} = 0.48095$$

$$Z = \frac{(10 - 9.9)}{0.48095} = 0.2079$$

Conclusión

No se rechaza H_0 , no existe suficiente evidencia para decir que los datos no son aleatorios. Se asume aleatoriedad en los datos.



Coeficiente de correlación de Spearman

Escala de medida

intervalo u ordinal

Observaciones

- *Son medidas de asociación lineal entre dos variables.*
- *Toman valores entre -1 y 1 los cuales indican máximo grado de asociación lineal negativa y positiva*
- *Utiliza rangos para calcular este valor*

Ejemplo 9

Se requiere establecer si existe relación entre el tiempo de estudio y la nota obtenida en un examen. Como la calificación del examen corresponde a una variable cualitativa se debe utilizar una prueba no paramétrica (Coeficiente de correlación de Spearman)

Los datos obtenidos son :

Tiempo	Nota
21	4
18	4
15	4
17	3
18	3
25	5
18	3
4	1
6	1
5	2

Tiempo	Rango	Nota	Rango
21	9	4	8
18	7	4	8
15	4	4	8
17	5	3	5
18	7	3	5
25	10	5	10
18	7	3	5
4	1	1	1,5
6	3	1	1,5
5	2	2	3

Para calcular el estadístico de prueba se deben asignar rangos a cada variable.

Después se restan los rangos para cada uno de los individuos y se calcula la sumatoria de estas diferencias al cuadrado.

Tiempo	Rango	Nota	Rango	Dif	Dif ²
21	9	4	8	1	1
18	7	4	8	-1	1
15	4	4	8	-4	16
17	5	3	5	0	0
18	7	3	5	2	4
25	10	5	10	0	0
18	7	3	5	2	4
4	1	1	1,5	-0,5	0,25
6	3	1	1,5	1,5	2,25
5	2	2	3	-1	1
Suma Dif ²					29,5

Coeficiente de correlación de Spearman

$$r_s = 1 - \frac{6 * \sum Dif^2}{n(n^2 - 1)} = 1 - \frac{6 * 29.5}{10(100 - 1)} = 1 - \frac{177}{990} = 0.8212$$

Para realizar la prueba de hipótesis correspondiente

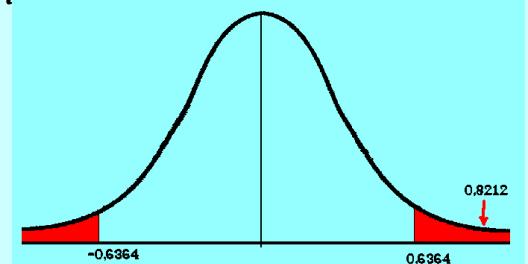
$$H_0 : \rho = 0$$

$$H_a : \rho \neq 0$$

El valor de r_s se compara contra el valor que se encuentra en la tabla P de la pág. A-106 del libro Biestadistica . Wayne W. Daniel

con $n=10$, $\alpha = 0.05$ valor critico 0.6364

Conclusión ρ es diferente de cero , existe una relación directa entre el tiempo dedicado a estudiar y la nota obtenida.



Ji-cuadrado en tablas de contingencia

Objetivo *Determinar si la diferencia entre las frecuencias observadas en la tabla de contingencia correspondiente al cruce de las dos variables y las frecuencias esperadas, suponiendo que las dos variables son independientes, son estadísticamente significativas.*

Ejemplo 10 :

Un investigador desea establecer la relación que puede existir entre la calificación un producto realizada por sus consumidores y su ubicación de residencia. Con este fin recoge información de 100 clientes :

Calificación	Lugar		
	Urbano	Rural	
Bueno	20	11	31
Regular	40	8	48
Malo	15	6	21
	75	25	100

H_0 : La calificación y la ubicación del consumidor son independientes

H_a : La calificación y la ubicación del consumidor no son independientes

En este caso calculamos los valores esperados suponiendo que las variables son independientes, por lo tanto :

$$P(B \cap U) = P(B) * P(U)$$

Calificación	Lugar		
	Urbano	Rural	
Bueno	20	11	31
Regular	40	8	48
Malo	15	6	21
	75	25	100

Calificación	Lugar		
	Urbano	Rural	
Bueno	23,25	7,75	31
Regular	36	12	48
Malo	15,75	5,25	21
	75	25	100

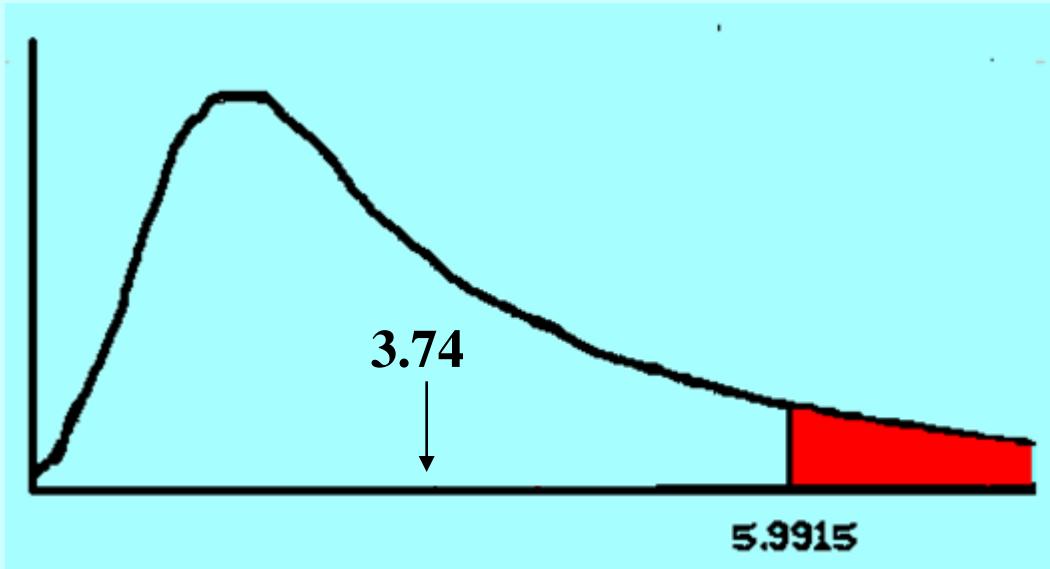
Estadístico de Prueba

$$X^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

con $(f-1)(c-1)$ grados de libertad

$$X^2 = \frac{(20 - 23.25)^2}{23.25} + \frac{(11 - 7.75)^2}{7.75} + \frac{(40 - 36)^2}{36} + \frac{(8 - 12)^2}{12} + \frac{(15 - 15.75)^2}{15.75} + \frac{(6 - 5.25)^2}{5.25}$$

$$X^2 = 3.74$$



Conclusión

No se rechaza H_0 , no existe suficiente evidencia para decir que los datos no son aleatorios. Se asume aleatoriedad en los datos.