Contraste de Hipótesis

1. Se quiere comprobar si una muestra de tamaño 20 con media 10 procede de una población N(14,3) con el **nivel de significación** 0,05.

Solución

2.- En una propaganda se anuncia que unas determinadas pilas proporcionan más horas de luz que las normales. Doce personas deciden comprarlas y los resultados obtenidos son los siguientes:

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Variación	0,2	0	1	0,6	-0,5	-0,6	-1	0,6	1	0,5	-0,4	-0,5

¿Se puede admitir, teniendo en cuenta estos datos, que el anuncio es correcto?

Solución

3.- Se intenta comparar dos métodos de enseñanza y para ello se selecciona una muestra de 50 individuos. A 20 personas se les aplica el método A y a las otras 30 el B. Al cabo de un año los aumentos en conocimiento han sido los siguientes $n_A = 20; \bar{x}_A = 5; S_A = 1,2; n_B = 30; \bar{x}_B = 3; S_B = 0,9$. Contrastar la **hipótesis de igualdad de varianzas y de igualdad de medias** al 95%.

Solución

4.- De 1000 familias con 5 hijos cada una se ha analizado la distribución de la variable X =”número de hijos varones, obteniéndose los siguientes resultados:

X	0	1	2	3	4	5
Nº de familias	26	162	309	321	153	29

¿Se puede afirmar que la variable X sigue una **distribución binomial** B(5,p)?

Solución

5.- Un tipógrafo asegurar que el número medio de errores por página que comete es 2, mientras el editor sospecha que es mayor. Suponiendo que el número de errores por página sigue una distribución de Poisson y que en una muestra de 200 páginas se encontraron 450 errores, especificar las **hipótesis nula** y **alternativa** del contraste y realizar el contraste con un **nivel de significación** del 5%.

Solución

6.- Un fabricante de bolsas de plástico asegura que el 95% de sus bolsas resisten 6 kg o más. Se toman 40 de estas bolsas al azar y se llenan con 6 kg rompiéndose 6 de ellas. ¿Existe evidencia estadística para rechazar la afirmación del fabricante?

Solución

7.- Se sabe que la renta anual de los individuos de un país sigue una distribución normal de media desconocida y de desviación típica 1.400. Se ha observado la renta anual de 16 individuos escogidos al azar y se ha obtenido un valor medio de 19.500 euros. Contrastar, a un **nivel de significación** del 5%, si la media de la distribución es de 18.000 de euros.

Solución

8.- Se aplica un cierto test de memoria a estudiantes de dos universidades diferentes. Los resultados siguen una distribución normal con desviación típica 33,5 para la primera universidad y para la segunda 38,2. Se toma una muestra aleatoria simple de 18 alumnos de la primera universidad y de 25 alumnos de la segunda; las medias muestrales obtenidas son, respectivamente, 183,7 y 165,4.

a) ¿Son estos datos significativos de una diferencia de memoria entre los alumnos de ambas universidades? (Considerar un **nivel de significación** del 0,05)

b) Calcular el **p-valor** e interpretarlo.

Solución

9.- Se han recogido los siguientes datos de una población:

X	0	1	2	3	4	5	6	7
n_i	8	19	22	24	23	15	5	4

¿Podemos admitir que siguen una **distribución de Poisson**?

Solución

10.- Se quiere comprobar si existe alguna relación entre las calificaciones de los alumnos en el tipo de asignaturas superadas. Para ello se ha tomado una muestra con alumnos de las tres asignaturas siguientes:

	Aprobado	Notable	Sobresaliente	Total
Geomática	152	99	29	280
Matemáticas	35	28	17	80
Física	12	18	10	40
Total	199	145	56	400

Contrastar la independencia con el **test de la Chi-cuadrado**. ¿Qué ocurre si consideramos sólo las asignaturas de Matemáticas y Física?

Solución

11.- El nivel medio de errores en nivelación es de 20 errores/100 tramos. Se estudian 40 nivelaciones en las que se sabe se han cometido errores. Los resultados son: $\overline{x} = 18.5$ err./100 tr. y s = 4 err/100 tr.

a) **Intervalo de confianza** del 95% del nivel medio de errores en las 40 nivelaciones.

b) ¿Es la muestra comparable con la población con un **nivel de significación** $\alpha = 0.05$?

Solución

12.- Queremos contrastar la **hipótesis** H_0: "la duración media de los motores de una determinada marca es 180.000 km". Para ello se toma una muestra de 9 motores de automóviles, obteniéndose una media de 150.000 km y una varianza muestral $S^2=(30.000)^2$. Estudiar si debemos aceptar la hipótesis H_0 con un **nivel de confianza** del 95%.

Solución

13.- Si el coeficiente medio de inteligencia de la población universitaria de la U.P.M. es $\mu = 95$ y $\sigma=14$, y se extrae una muestra de 49 estudiantes de esa población.

a) ¿Qué probabilidad hay de que resulte una media muestral igual o inferior a 92?

b) Hallar un **intervalo de confianza** al 95% para la media de la población universitaria, para la muestra de media 92.

c) ¿Podemos aceptar la **hipótesis** de ser la media $\mu = 95$,

Solución

14.- Se prueban 12 piezas de un material A resultando un desgaste medio de 85 unidades con una desviación típica de 4. Por otra parte 10 piezas de un material B produce unos resultados de 81 unidades y una desviación típica de 5.

a) Comprobar que las varianzas poblacionales son iguales a un **nivel de significación** de 0,1.

b) ¿Podemos concluir con un **nivel de significación** del 0,05 que el desgaste del material A excede el del material B en más de 2 unidades? Suponga que las poblaciones son normales.

Solución

15.- En medidas de ángulos con un cierto teodolito, un topógrafo asegura que la varianza que obtiene es igual o menor que 5. Se le pone a prueba y se le hacen 20 determinaciones, obteniéndose una varianza de 6. Si la variable medida del ángulo es normal. Se pide:

a) **Intervalo de confianza** del 99% para la varianza obtenida.

b) ¿Aceptamos su aseveración a un **nivel de significación** 0,01?

Solución

16.- Se tabulan los errores de cierre en nivelación obtenidos en 1000 polígonos. ¿Se puede admitir que el error de cierre se distribuye normalmente? Contrastar la **bondad del ajuste** mediante la χ^2_n con α=0,05.

error de cierre	n° de polígonos
0 -0,1	64
0,1-0,2	242
0,2-0,3	390
0,3-0,4	238
0,4-0,5	66

Solución

17.- La siguiente tabla contiene los datos de dos muestras de tamaño 50. ¿Podemos admitir que tienen la misma distribución utilizando el **estadístico de Pearson** y calculando el **p-valor**?

1ª muestra	2ª muestra
3	1
2	4
4	9
11	5
11	10
9	9
5	4
3	4
2	4

Solución

18.- A la vista de los datos que aparecen en la tabla adjunta, obtenidos después de una encuesta realizada para ver si la edad tiene influencia en los errores cometidos por los observadores. ¿Se puede admitir que el cometer errores es independiente de la edad?

Edad	Cometer error	No cometer error
<30	38	44
30-40	45	28
40-50	30	54
50-60	22	62
>60	20	57

Solución

19.- Se quieren comparar la vida útil de dos tipos de baterias para moviles, para ello obtenemos muestras de tamaño 9 de cada tipo, obteniendo los siguientes resultados $n_A = 9; \bar{x}_A = 5; S_A = 1,2; n_B = 9; \bar{x}_B = 3; S_B = 0,9$. Contrastar la hipótesis de igualdad de medias al 95% suponiendo que la varianza es desconocida pero igual para las dos muestras.

Solución

20.- Al calcular cinco veces la distancia entre dos puntos, obtenemos los siguientes valores*:*

170,13m; 170,12m; 170,2m; 170,65m; 170,4

Se pide:

a) Intervalo de confianza del 80% para la media.

b) ¿Cuál será el número de mediciones necesaria para que el error sea inferior a 0,1 m con un nivel de significación de 0,2?

c) Intervalo de confianza del 90% para la varianza obtenida.

d) Si la variable medida del ángulo es normal ¿aceptamos una desviación típica inferior a 0,1 m con un nivel de significación 0,1?

Solución

21.- Se ha observado un ángulo cinco veces, obteniéndose los siguientes valores:

65º25'; 65º33'; 65º32'; 65º28'; 65º27'

Se pide:

a) Intervalo de confianza del 95% para la media del ángulo observado.

b) ¿Cuál será el número de observaciones necesaria para que el error sea inferior a 1 minuto con un nivel de significación de 0,05?

c) Intervalo de confianza del 99% para la varianza obtenida.

d) Si la variable medida del ángulo es normal ¿aceptamos una desviación típica inferior a 2 minutos a un nivel de significación 0.01? Escribir la región crítica.

Solución

22.- Se quiere comprobar si es cierto que un nuevo modelo de automóvil consume 5,7 litros por cada 100 km a 90 km/h tal y cómo afirma en su propaganda. Para ello se realizan 10 recorridos de 100 km a dicha velocidad, con los siguientes consumos:

5,8; 6,3; 6; 5,5; 6,1; 6,5; 5,5; 7; 6,8; 5,7

Se pide:

a) Intervalo de confianza para la media obtenida con un nivel de significación del 0,05.

b) ¿Se puede aceptar la afirmación del fabricante al 95%?

Solución

23.- Un jugador afirma que los dos dados que utiliza no están trucados. Para comprobar esta afirmación, se lanzan los dos dados 360 veces y se anota la suma de los resultados. Dichas sumas se muestran en la siguiente tabla, junto con su probabilidad correspondiente (para facilitar la comparación).

Suma	FRECUENCIA OBSERVADA	PROBABILIDAD
2	11	1/36
3	18	2/36
4	33	3/36
5	41	4/36
6	47	5/36
7	61	6/36

8	52	5/36
9	43	4/36
10	29	3/36
11	17	2/36
12	8	1/36

Para poder decidir si los dados están trucados o no, realizar un contraste χ^2. ¿Deberíamos jugar con estos dados?

Solución

24.- Durante 50 días se ha observado la variable "número diario de cancelaciones de cuentas en una sucursal bancaria" anotándose la siguiente tabla de frecuencias:

Nº cancelaciones	Frecuencias observadas n_i
0	16
1	23
2	8
3	3
Sumas	50

Contrastar el hecho de que la distribución es de Poisson de parámetro λ=1.

Solución

25.- Para curar una enfermedad se sabe que existen cuatro tratamientos diferentes. Aplicados por separado a unos grupos de enfermos se han observado los resultados siguientes:

Tratamiento	Curados	No Curados
A	83	44
B	45	28
C	60	54
D	82	62
Total	270	188

¿Se puede asegurar que la eficacia de los cuatro tratamientos es la misma, con un nivel de significación del 95%**?**

Solución

26.- Si el coeficiente medio de inteligencia de la población universitaria de la U.P.M. es $\mu = 95$ y σ=14, y se extrae una muestra de 49 estudiantes de esa población y el resultado de la varianza muestral S^2=100. ¿Podemos concluir con un nivel de significación del 0,05 que la varianza es menor de 14^2?

Solución

27.- Se lleva a cabo un experimento para comparar el desgaste por abrasivo de un material laminado. Se prueban 12 piezas del material mediante la exposición de cada pieza a una máquina para medir la profundidad del desgaste. La muestra da un desgaste medio de 53 unidades con una desviación estándar muestral de 4. Supongamos que la población es aproximadamente normal con varianza desconocida.

a) ¿Podemos concluir con un nivel de significación del 0,05 que el desgaste abrasivo medio del material es menor de 50 unidades?

b) Se justifica la suposición de varianza igual a 10 con un nivel de significación de 0,1.

Solución

28.- Al realizar un muestreo aleatorio con 100 alumnos de la U.P.M para estudiar la estatura de los alumnos de esta Universidad, se obtuvieron los siguientes resultados agrupados en intervalos de amplitud 10 cm, así como los parámetros de la muestra $\overline{X} = 1.772$ y $S = 0.1168$.

	n_i
menos de 1.5	0
1.5 - 1.6	6
1.6 - 1.7	21
1.7 - 1.8	35
1.8 - 1.9	24
1.9 - 2.0	11
2.0 - 2.1	3
más de 2.1	0

Contrastar la hipótesis,

La muestra obtenida sigue una distribución normal $N(\overline{X} = 1.772, S = 0.1168)$

a) Obtener el p-valor y dar su interpretación.

b) Para un nivel de significación $\alpha = 0.05$ calcular el valor crítico.

c) ¿Se acepta el ajuste para un nivel de significación $\alpha = 0.05$?

Solución

29.- Para contrastar que la esperanza matemática μ de una distribución Normal es 10, se toma una muestra de tamaño 16 y se rechaza la hipótesis en el caso en que la media muestra sea mayor que 10, aceptándose en caso contrario. Sabiendo que la desviación típica de la población es σ=2 y habiendo obtenido una media muestral de 11. Calcular el p-valor y decidir si se acepta la hipótesis de que la media de la población es μ<10

Solución

30.- Para los siguientes datos

x_i	n_i
0	10
1	30
2	50
3	10

Se pide:

a) Calcular las respectivas frecuencias teóricas correspodientes a una distribución Binomial de parámetros n=3 y p=0,53, es decir, $p_i=P(X=x_i)$

b) Hallar el valor del estadístico de contraste:

$$D=\sum_i \frac{(n_i-np_i)^2}{np_i}$$

c) Estudiar la Bondad del ajuste para un nivel de significiación α=0,05

Solución

31.- Suponiendo una población normal con media 50 y desviación típica 8. Para una muestra de tamaño 9, ¿cuál es la probabilidad de que:
a) La media muestral sea superior a 55?
b) La desviación típica muestral sea superior a 10?

Solución

32.- Según un estudio sobre los hábitos de consumo de las familias de la Comunidad de Madrid, el gasto mensual en productos de marcas blancas de las familias puede considerarse según una variable aleatoria *X* con ley Normal de desviación típica 50 euros. Los datos siguientes corresponden a una muestra aleatoria simple de tamaño *n* = 10 familias:

Familia	1	2	3	4	5	6	7	8	9	10
Gasto	325	193	203	305	251	240	233	238	354	311

a) ¿Qué estimación del gasto medio mensual se obtiene con un estimador insesgado?

b) Calcula un intervalo de confianza al 95% para el gasto medio mensual en productos de marcas blancas de las familias de la Comunidad de Madrid.

c) Si queremos trabajar con niveles de confianza del 90%, ¿obtendríamos un intervalo más o menos) amplio que el calculado anteriormente?

Solución

33.- Se sabe que la demanda diaria de un producto fabricado por cierta empresa sigue una distribución normal. Se toma una muestra aleatoria simple de la demanda registrada en 10 días distintos, obteniéndose una media muestral igual a 3,5 y una desviación típica muestral (o cuasidesviación típica) igual a 0,9. Se pide:
a) Un intervalo de confianza al 95% para la media de la demanda diaria de dicho producto.
b) El plan de producción de la empresa está basado en la suposición de que la demanda media diaria del citado producto es como mínimo igual a 4. Contrastar esta hipótesis, al nivel de significación de 0,05.

Solución

34.- Se supone que el retraso de los trabajadores de una empresa es, por término medio, de 5 minutos; para contrastar dicha hipótesis se seleccionó al azar una muestra de 10 empleados y se midió su retraso: 0.1, 2, 7, 5.6, 7.4, 5.1, 6.1, 6, 4.5 y 9. Utilizar el p-valor.

Solución

35.- La siguiente tabla de contingencia contiene datos sobre el gasto mensual en llamadas telefónicas (en euros) y el sexo de 500 encuestados:

Gasto	Hombres	Mujeres
[0,6)	59	76
[6,12)	42	62
[12,30)	54	51
[30,60)	67	39
[60,120]	28	22

Justificar si es cierto que el gasto mensual medio de los hombres es superior al de las mujeres con un nivel de significación del 5%.

Solución

36.- ¿Se puede admitir la distribución uniforme de valores angulares en una triangulación de primer orden de un país en la que se ha tomado una muestra de tamaño 100 y se han obtenido los siguientes resultados? Utilizar el p-valor.

Medidas angulares	frecuencias observadas
<40	16
40-50	22
50-60	20
60-70	19
>70	23

Solución

37.- Un proceso industrial fabrica piezas cuya longitud en milímetros se distribuye según una Normal de media 190 y desviación típica 10.
Una muestra de 5 piezas proporciona los resultados siguientes: 185, 200, 186, 197 y 185. Utilizando un nivel de significación del 0,05, se pide:

a) Contrastar la hipótesis de que la media del proceso es 190.
b) Contrastar la hipótesis de la varianza del `proceso es 100.

Solución

38.- Se quiere contrastar la hipótesis nula "una moneda de un euro está bien equilibrada", y que por tanto, cara y cruz son equiprobables; para ello se realizan 100 lanzamientos de una moneda de un euro y se observan 63 caras y 37 cruces. Contrastar la hipótesis utilizando el test de χ^2 con un nivel de significación α de 0,05.

Solución

1.- Se quiere comprobar si una muestra de tamaño 20 con media 10 procede de una población N(14,3) con el nivel de significación 0,05.

Solución:

Se trata de un contraste de hipótesis para la media de una población normal de varianza conocida:

$$H_0 : \mu = 14$$
$$H_1 : \mu \neq 14$$

Sabemos que: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \equiv N(0,1)$

El valor del estadístico Z bajo la hipótesis nula es:

$$Z = \frac{\left|\bar{X} - \mu_0\right|}{\sigma / \sqrt{n}} = \frac{|10-14|}{3/\sqrt{20}} \simeq 5{,}96$$

Para α=0,05 en la N(0,1) tenemos que:

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha \Leftrightarrow P\left(Z < 1{,}96\right) = 0{,}975 \Rightarrow z_{0,025} = 1{,}96$$

Como el valor de nuestro estadístico Z bajo la hipótesis nula cae fuera de la región de aceptación ($|Z| > z_{\alpha/2}$), RECHAZAMOS la hipótesis nula, y por tanto la muestra no procede de dicha población.

2.- En una propaganda se anuncia que unas determinadas pilas proporcionan más horas de luz que las normales. Doce personas deciden comprarlas y los resultados obtenidos son los siguientes:

Individuo	1	2	3	4	5	6	7	8	9	10	11	12
Variación	0,2	0	1	0,6	-0,5	-0,6	-1	0,6	1	0,5	-0,4	-0,5

¿Se puede admitir, teniendo en cuenta estos datos, que el anuncio es correcto?

Solución:

Realizamos la hipótesis nula de que no proporcionan más horas de luz que las normales.

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

Datos: $n = 12; \overline{x} = 0,075$; $S = 0,672$; $\alpha = 0.05$

$$t = \left|\frac{\overline{x}-\mu_0}{S/\sqrt{n}}\right| = \left|\frac{0,075-0}{0,672/\sqrt{12}}\right| \simeq 0.3866184838$$

Calcularemos el p-valor

$p = 1 - P\left(-0,3866 < t_{n-1} < 0,3866\right) = 2P\left(0,3866 < t_{n-1}\right) = 0,7064$.

DERIVE:

```
#1:    2(1-(STUDENT(0.3866, 11))
#2:                                 0.7064262140
```

EXCEL:

=DISTR.T.(0,3866;11;2) 0,70642621

SPSS:

2(1-CDF. T(0.3866,11)) 0,7

Podemos aceptar la hipótesis nula, para cualquier valor de $\alpha < 0,7$ y en consecuencia la propaganda es falsa.

3.- Se intenta comparar dos métodos de enseñanza y para ello se selecciona una muestra de 50 individuos. A 20 personas se les aplica el método A y a las otras 30 el B. Al cabo de un año los aumentos en conocimiento han sido los siguientes $n_A = 20; \bar{x}_A = 5; S_A = 1,2; n_B = 30; \bar{x}_B = 3; S_B = 0,9$. Contrastar la hipótesis de igualdad de varianzas y de igualdad de medias al 95%.

Solución:

Debemos en primer lugar contrastar la hipótesis de igualdad de varianzas

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

$$\frac{S_A^2}{S_B^2} \in \left(F_{1-\frac{\alpha}{2}, n_A - 1, n_B - 1}, F_{\frac{\alpha}{2}, n_A - 1, n_B - 1} \right)$$

$$\frac{S_A^2}{S_B^2} = \frac{1.2^2}{0.9^2} = 1.78 \in \left(F_{0.975,19,29}, F_{0.025,19,29} \right) = (0.42, 2.23)$$

EXCEL:

=DISTR.F.INV(0,975;19;29)0,41633; =DISTR.F.INV(0,025;19;29) 2,231275

SPSS:

IDF.F(0.975,19,29) .42;IDF.F(0.025,19,29) 2.23

0,42 < 1,78 < 2,23 y por tanto **aceptamos la hipótesis de varianzas iguales**.

Contrastamos ahora la igualdad de medias de dos poblaciones normales de varianzas desconocidas pero iguales y muestras pequeñas.

$$H_0 : \mu_A = \mu_B \Leftrightarrow \mu_A - \mu_B = 0$$

$$H_1 : \mu_A \neq \mu_B \Leftrightarrow \mu_A - \mu_B \neq 0$$

El estadístico de prueba es: $\dfrac{|\bar{x}_A - \bar{x}_B|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = t$ siendo

$$S^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)} = \frac{19 \cdot 1.2^2 + 29 \cdot 0.9^2}{19 + 29} = 1.059375 \text{ con lo cual}$$

$$\frac{|\bar{x}_A - \bar{x}_B|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = \frac{5 - 3}{\sqrt{1.059375}\sqrt{\dfrac{1}{20} + \dfrac{1}{30}}} = 6.73$$ y para α=0.05, $t_{0.025,48}$=2

EXCEL: =DISTR.T.INV(0,05;48) 2,01063472

SPSS:IDF.T(0.975,48) 2.01

Como 6.73 > 2 **RECHAZAMOS** la hipótesis de igualdad de medias.

4.- De 1000 familias con 5 hijos cada una se ha analizado la distribución de la variable X =”número de hijos varones, obteniéndose los siguientes resultados:

X	**0**	**1**	**2**	**3**	**4**	**5**
Nº de familias	**26**	**162**	**309**	**321**	**153**	**29**

¿Se puede afirmar que la variable X sigue una distribución binomial B(5,p)?

Solución:

Primeramente debemos estimar el valor del parámetro p, hacemos

$$\bar{X} = np \Rightarrow p = \frac{\bar{X}}{n} = \frac{2,5}{5} = 0,5$$

Observando una distribución B(5,0.5) y multiplicando por 1000, obtenemos el número de familias según la distribución teórica (esperadas).

X	0	1	2	3	4	5
n_i	26	162	309	321	153	29
p_i	$\binom{5}{0}0.5^5$	$\binom{5}{1}0.5^5$	$\binom{5}{2}0.5^5$	$\binom{5}{3}0.5^5$	$\binom{5}{4}0.5^5$	$\binom{5}{5}0.5^5$
np_i	31.25	156.25	312.5	312.5	156.25	31.25
$\frac{(n_i - np_i)^2}{np_i}$	$\frac{(26-31.25)^2}{31.25}$	$\frac{(162-156.25)^2}{156.25}$	$\frac{(309-312.5)^2}{312.5}$	$\frac{(321-312.5)}{312.5}$	$\frac{(153-156.25)^2}{156.25}$	$\frac{(29-31.25)^2}{31.25}$

$$D = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = 1,59$$

utilizando el p-valor:

DERIVE: 1 - CHI_SQUARE(1.59,4)= 0.8105883166> α

EXCEL: = DISTR.CHI(1,59;4) 0,810588> α

SPSS: 1 - CDF.CHISQ(1.59,4) .81> α

Para α = 0,05 $P(\chi_4^2 \geq \chi_{0.05}) = 0.05 \Rightarrow \chi_{0.05} = 9,49$ siendo D = 1,59 menor que $\chi_{0.05}$ aceptamos la hipótesis de ser el ajuste bueno. La diferencia entre la distribución empírica y la ley de la distribución binomial no es significativa. **Aceptamos el ajuste para todo nivel de significación menor que 0,81**.

5.- Un tipógrafo asegurar que el número medio de errores por página que comete es 2, mientras el editor sospecha que es mayor. Suponiendo que el número de errores por página sigue una distribución de Poisson y que en una muestra de 200 páginas se encontraron 450 errores, especificar las hipótesis nula y alternativa del contraste y realizar el contraste con un nivel de significación del 5%.

Solución:

La variable aleatoria X=”número de errores por página”=P(λ)

$$\boxed{\begin{array}{l} H_0 : \lambda \le 2 \\ H_1 : \lambda > 2 \end{array}}$$

Para realizar el contraste se extrae una muestra $\{x_1, x_2, ..., x_n\}$ con n=200.

La variable $Y = X_1 + X_2 + ... + X_n$ que representa el número de errores en n páginas, sigue una distribución aproximadamente Normal (Teorema Central del Límite)

$$Y \equiv N\left(E[Y], \sqrt{V[Y]}\right) \Rightarrow \frac{Y - E[Y]}{\sqrt{V[Y]}} \equiv N(0,1)$$

Entonces:

$$E[Y] = E[X_1 + ... + X_n] = n\lambda = 400$$

$$V[Y] = V[X_1 + ... + X_n] = nV[X] = n\lambda = 400$$

$$Z = \frac{Y - E[Y]}{\sqrt{V[Y]}} = \frac{450 - 400}{\sqrt{400}} = 2,5$$

Para α=0,05 en la N(0,1) tenemos que:

$$P(Z < z_\alpha) = 1 - \alpha \Leftrightarrow P(Z < 1,6448) \approx 0,95 \Rightarrow z_{0,05} = 1,6448$$

Como el valor de nuestro estadístico Z bajo la hipótesis nula cae fuera de la región de aceptación (2,5>1,64), se **RECHAZAMOS** la hipótesis nula y se puede concluir que el número de errores por página es superior a 2.

DERIVE:

```
#1:    NSOLVE(NORMAL(λ) = 0.95, λ, Real)
#2:                              λ = 1.644853651
```

EXCEL: =DISTR.NORM.INV(0,95;0;1)   1.6448536

SPSS: IDF. NORMAL(0.95, 0,1)   1,64

6.- Un fabricante de bolsas de plástico asegura que el 95% de sus bolsas resisten 6 kg o más. Se toman 40 de estas bolsas al azar y se llenan con 6 kg rompiéndose 6 de ellas. ¿Existe evidencia estadística para rechazar la afirmación del fabricante?

Solución:

Establecemos la proporción de bolsas rotas de nuestra muestra en p y n=40, siendo X= número de bolsas que se rompen tendremos una distribución Binomial de parámetros n y p que ajustaremos a una distribución Normal; en nuestro caso:

$$X \equiv N\left(np, \sqrt{np(1-p)}\right) = N\left(40p, \sqrt{40p(1-p)}\right)$$

Planteamos $\boxed{\begin{array}{l} H_0 : p \leq 0,05 \\ H_1 : p > 0,05 \end{array}}$, en cuyo caso la distribución adecuada queda en N(2,1.3784) que podemos ajustar a una distribución Normal que tipificando resulta $Z = \frac{X-\mu}{\sigma} = \frac{X-2}{1.3784} \equiv N(0,1)$ y en particular para X=6 $\Rightarrow Z \approx 2.901915264$

Para α=0,05 en la N(0,1) tenemos que:

$$P\left(Z < z_\alpha\right) = 1-\alpha \Leftrightarrow P\left(Z < 1,6448\right) \approx 0,95 \Rightarrow z_{0,05} = 1,6448$$

Como el valor de nuestro estadístico Z bajo la hipótesis nula cae fuera de la región de aceptación (2,9>1,64), se **RECHAZAMOS la afirmación del fabricante**

7.- Se sabe que la renta anual de los individuos de un país sigue una distribución normal de media desconocida y de desviación típica 1.400. Se ha observado la renta anual de 16 individuos escogidos al azar y se ha obtenido un valor medio de 19.500 euros. Contrastar, a un nivel de significación del 5%, si la media de la distribución es de 18.000 de euros.

Solución:

Se trata de un contraste de hipótesis para la media de una población normal de varianza conocida:

$$\mathrm{H}_0 : \mu = 18.000$$

$$\mathrm{H}_1 : \mu \neq 18.000$$

Sabemos que: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \equiv N(0,1)$

El valor del estadístico Z bajo la hipótesis nula es:

$$Z = \frac{\left|\bar{X} - \mu_0\right|}{\sigma / \sqrt{n}} = \frac{|19500 - 18000|}{1400 / \sqrt{16}} \simeq 4.285714285$$

Para α=0,05 en la N(0,1) tenemos que:

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha \Leftrightarrow P\left(Z < 1,96\right) = 0,95 \Rightarrow z_{0,025} = 1,96$$

Como el valor de nuestro estadístico Z bajo la hipótesis nula cae fuera de la región de aceptación ($|\mathrm{Z}| > \mathrm{z}_{\alpha/2}$), RECHAZAMOS la hipótesis nula, y la media no puede ser 18.000 euros.

8.- Se aplica un cierto test de memoria a estudiantes de dos universidades diferentes. Los resultados siguen una distribución normal con desviación típica 33,5 para la primera universidad y para la segunda 38,2. Se toma una muestra aleatoria simple de 18 alumnos de la primera universidad y de 25 alumnos de la segunda; las medias muestrales obtenidas son, respectivamente, 183,7 y 165,4.
a) ¿Son estos datos significativos de una diferencia de memoria entre los alumnos de ambas universidades? (Considerar un nivel de significación del 0,05)
b) Calcular el p-valor e interpretarlo.

Solución:
a)
Contrastamos la igualdad de medias de dos poblaciones normales de varianzas conocidas.

$$H_0 : \mu_A = \mu_B \Leftrightarrow \mu_A - \mu_B = 0$$
$$H_1 : \mu_A \neq \mu_B \Leftrightarrow \mu_A - \mu_B \neq 0$$

El estadístico de prueba es: $\dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}} = z$ siendo

$\overline{X}_A = 185,7;\ \overline{X}_B = 165,4;\ \sigma_A = 33,5;\ \sigma_B = 38,2;\ n_A = 18;\ n_B = 25$ con lo cual

$$\frac{\overline{X}_A - \overline{X}_B}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}} = \frac{183,7 - 165,4}{\sqrt{\dfrac{33,5^2}{18} + \dfrac{38,2^2}{25}}} = 1,665586501$$

Para α=0,05 en la N(0,1) tenemos que:

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha \Leftrightarrow P\left(Z < 1,96\right) = 0,95 \Rightarrow z_{0,025} = 1,96$$

Como 1,6655 < 1,96 **ACEPTAMOS** la hipótesis de igualdad de medias.

b) utilizando el p-valor:

$$p = 1 - P(-1,66 \leq Z \leq 1,66) = 2P(Z \geq 1,66) = 2\left(1 - P(Z < 1,66\right) =$$
$$= 2\left(1 - F(1,66)\right) \approx 0,0969$$

DERIVE: 2(1 -NORMAL(1.66))= 0.09691445252 > α
EXCEL: = 2(1-DISTR.NORM.ESTAND(1,66)) 0,09691445 > α
SPSS: 2 * (1- CDF.NORMAL(1.66,0,1)) .1 > α
Para CUALQUIER α < 0,1 = P-VALOR aceptamos la hipótesis de igualdad de media.

9.- Se han recogido los siguientes datos de una población:								
X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
n_i	**8**	**19**	**22**	**24**	**23**	**15**	**5**	**4**

¿Podemos admitir que siguen una distribución de Poisson?

Solución:

La distribución de Poisson depende de un único parámetro λ cuyo estimador de máxima verosimilitud es la media muestral $\overline{X}$, que en nuestro caso es $\overline{X} = \frac{\sum_{i=0}^{7} x_i n_i}{n} = \frac{360}{120} = 3$

Calculamos la distribución teórica de tal forma que la última opción sea X= 7 o superior

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{3^k}{k!} e^{-3}$$

X	n_i	p_i	n p_i
0	**8**	0,04978707	5,9744482
1	**19**	0,14936121	17,9233446
2	**22**	0,22404181	26,8850169
3	**24**	0,22404181	26,8850169
4	**23**	0,16803136	20,1637627
5	**15**	0,10081881	12,0982576
6	**5**	0,05040941	6,04912881
≥ 7	**4**	0,033508536	**4,02102426**

Debemos agrupar las dos últimas clases para cumplir el requisito de np_i>5

X	n_i	n p_i	n_i - n p_i	(n_i - n p_i)^2	(n_i-np_i)^2/(np_i)
0	**8**	5,9744482	2,0255518	4,10286008	0,68673456
1	**19**	17,9233446	1,07665539	1,15918682	0,06467469
2	**22**	26,8850169	-4,88501691	23,8633902	0,88760927
3	**24**	26,8850169	-2,88501691	8,32332258	0,30958964
4	**23**	20,1637627	2,83623732	8,04424211	0,39894549
5	**15**	12,0982576	2,90174239	8,42010891	0,69597699
≥6	**9**	**10,0701531**	-1,07015307	1,14522759	0,11372494
sumas	**120**	**120**			**3,15725558**

El estadístico de Pearson es $D = \sum_{i=0}^{k} \frac{(n_i - np_i)^2}{np_i} \approx 3{,}16$

Como quedan 7 clases resulta una distribución chi- cuadrado con 5 grados de libertad. utilizando el p-valor:

DERIVE: 1 - CHI_SQUARE(3.16,5)= 0.6753340614> α

EXCEL: = DISTR.CHI(3,16;5) 0.675334066> α

SPSS: 1 - CDF.CHISQ(3.16,5) .68> α

Aceptamos la hipótesis de ser el ajuste bueno. La diferencia entre la distribución empírica y la ley de la distribución Poisson no es significativa. **Aceptamos el ajuste para todo nivel de significación menor que 0,68**.

10.- Se quiere comprobar si existe alguna relación entre las calificaciones de los alumnos en el tipo de asignaturas superadas. Para ello se ha tomado una muestra con alumnos de las tres asignaturas siguientes:

	Aprobado	Notable	Sobresaliente	Total
Geomática	152	99	29	280
Matemáticas	35	28	17	80
Física	12	18	10	40
Total	199	145	56	400

Contrastar la independencia con el test de la Chi-cuadrado. ¿Qué ocurre si consideramos sólo las asignaturas de Matemáticas y Física?

Solución:

¿Qué debería salir, si fueran independientes?

	Aprobado	Notable	Sobresaliente	Total
Geomática	**199 280/400**	**145 280/400**	**56 280/400**	280
Matemáticas	**199 80/400**	**145 80/400**	**56 80/400**	80
Física	**199 40/400**	**145 40/400**	**56 40/400**	40
Total	199	145	56	400

H_o : Las tres asignaturas son independientes

H_1 : Alguna no es independiente

	Aprobado	Notable	Sobresaliente	Total
Geomática	152**(139,5)**	99**(101,5)**	29**(39,2)**	280
Matemáticas	35**(39,8)**	28**(29)**	17**(11,2)**	80
Física	12**(19,9)**	18**(14,5)**	10**(5,6)**	40
Total	199	145	56	400

Comparamos frecuencias observadas (O_{ii}) y esperadas (e_{ij})

$$D = \sum_{j=1}^{r} \sum_{i=1}^{k} \frac{(O_{ij} - e_{ij})^2}{e_{ij}} \approx 14.92861872$$

Resulta una distribución chi- cuadrado con (3-1)x(3-1)=4 grados de libertad.

utilizando el p-valor:

DERIVE: 1 - CHI_SQUARE(14.9,4)= 0.004913181563 > α
EXCEL: = DISTR.CHI(14,9;4) 0.0049132 > α
SPSS: 1 - CDF.CHISQ(14.9,4) .00 > α

Con un valor de p tan pequeño debemos **RECHAZAR la independencia de los datos.**

Consideraremos solamente dos asignaturas

H_o : Las dos asignaturas son independientes

H_1 : Las asignaturas no son independientes

	Aprobado	Notable	Sobresaliente	Total
Matemáticas	35**(31,3333)**	28**(30,6666)**	17**(18)**	80
Física	12**(15,6666)**	18**(15,3333)**	10**(9)**	40
Total	47	46	27	120

Tenemos un estadístico D=2.149552883 con una distribución chi_cuadrado con (2-1)x(3-1) = 2 grados de libertad.
utilizando el p-valor:

DERIVE: 1 - CHI_SQUARE(2,15,2)= 0.3412977553 > α
EXCEL: = DISTR.CHI(2,15;2) 0.34129776 > α
SPSS: 1 - CDF.CHISQ(2.15,2) .34 > α

No tenemos evidencia para rechazar la independencia de las calificaciones.
Aceptamos la independencia para todo nivel de significación menor que 0,34.

11.- El nivel medio de errores en nivelación es de 20 errores/100 tramos. Se estudian 40 nivelaciones en las que se sabe se han cometido errores. Los resultados son: $\overline{x}$ = 18.5 err./100 tr. y s = 4 err/100 tr.
a) Intervalo de confianza del 95% del nivel medio de errores en las 40 nivelaciones.
b) ¿Es la muestra comparable con la población con un nivel de significación $\alpha = 0.05$?

Solución:
a)
Tenemos una muestra de tamaño grande (n=40) y varianza desconocida de una distribución normal:

$$I_\alpha = \left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Datos: $\overline{X} = 18.5$; $S = 4; \alpha = 0.05$ y en la distribución normal

$$F(Z < z_{\alpha/2}) = \frac{\alpha}{2} + 1 - \alpha = 0.975 \Rightarrow \lambda_{\alpha/2} = 1.96$$

$$\Rightarrow I_{\alpha=0.05} = \left(18.5 - 1.96\frac{4}{\sqrt{40}}, 18.5 + 1.96\frac{4}{\sqrt{40}}\right) = \boxed{(17.867545, 19.132455)}$$

b)
H_0: $\mu = 18.5$.
H_1: $\mu \neq 18.5$

$\Rightarrow \mu = 20 \notin I_{\alpha=0.05} = (17.867545, 19.132455)$ Rechazamos la hipótesis nula H_0
.

Otro método puede ser, $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \dfrac{18.5 - 20}{4 / \sqrt{40}} = -2.3717 \notin (-1.96, 1.96)$ **la muestra no es comparable a la población**.

12.- Queremos contrastar la hipótesis H_0: "la duración media de los motores de una determinada marca es 180.000 km". Para ello se toma una muestra de 9 motores de automóviles, obteniéndose una media de 150.000 km y una varianza muestral $S^2=(30.000)^2$. Estudiar si debemos aceptar la hipótesis H_0 con un nivel de confianza del 95%.

***Solución*:**

Se trata de un contraste de hipótesis para la media de una población normal de varianza desconocida:

$$H_0 : \mu = 180.000$$
$$H_1 : \mu \neq 180.000$$

Datos: $n = 9; \bar{x} = 150.000$; $S = 30.000$; $\alpha = 0,05$

$$t_{n-1}=\left|\frac{\bar{X}-\mu}{S/\sqrt{n}}\right| = \left|\frac{150.000-180.000}{30.000/\sqrt{9}}\right|=3$$

Buscaremos un valor $t_{\alpha/2}$ tal que $P\left(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}\right) = 1-\alpha$.

DERIVE:

#1: NSOLVE(STUDENT(t, 8) = 0.975, t)

#2: t = 2.306004166

EXCEL: =DISTR.T.INV(0,05;8

2,305004166

SPSS: IDF. T(0.975,8) 2,3

Puesto que $|t| = 3>2,3$, **No podemos aceptar la hipótesis nula**, y en consecuencia la propaganda es falsa.

13.- Si el coeficiente medio de inteligencia de la población universitaria de la U.P.M. es $\mu = 95$ y σ=14, y se extrae una muestra de 49 estudiantes de esa población.
a) ¿Qué probabilidad hay de que resulte una media muestral igual o inferior a 92?
b) Hallar un intervalo de confianza al 95% para la media de la población universitaria, para la muestra de media 92.
c) ¿Podemos aceptar la hipótesis de ser la media $\mu = 95$.

***Solución*:**

La distribución de la media muestral es $\bar{\xi} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(95, \frac{14}{\sqrt{49}}\right) = N(95, 2)$

a) $F(92) = P\left(\bar{\xi} \le 92\right) \approx \boxed{0.06680720126}$

DERIVE: #1: NORMAL(92,95,2)
#2: 0.06680720126
EXCEL: =DISTR.NORM.(92;95;2;1) 0,0668072
SPSS: CDF. NORMAL(92,95,2) ,06680720

b) El intervalo de confianza para una población normal es:

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Para nuestros datos: $\bar{x} = 92; \sigma = 14; n = 49; \alpha = 0.05$

$$\text{Tenemos } Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \equiv N(0,1)$$

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1 - \alpha \Leftrightarrow P\left(Z < 1{,}96\right) = 0{,}975 \Rightarrow z_{0{,}025} = 1{,}96$$

DERIVE: #1: NSOLVE(NORMAL(z) = 0.975, z, Real)
#2: z = 1.959963962
EXCEL: =DISTR.NORM.INV(0,975;0;1) 1,9599628
O directamente
=INTERVALO.CONFIANZA(0,05;14;49) 3,91992797

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 92 \pm 3{,}919927969$$

SPSS: IDF. NORMAL(0.975, 0,1) 1,96

$$\Rightarrow I_{\alpha=0.05} = \left(92 - 1.96\frac{14}{\sqrt{49}}, 92 + 1.96\frac{14}{\sqrt{49}}\right) = \boxed{(88.08, 95.92)}$$

c) $\mu = 95 \in I_{\alpha=0.05} = (88.08, 95.92)$, SÍ SE ACEPTA

14.- Se prueban 12 piezas de un material A resultando un desgaste medio de 85 unidades con una desviación típica de 4. Por otra parte 10 piezas de un material B produce unos resultados de 81 unidades y una desviación típica de 5.
a) Comprobar que las varianzas poblacionales son iguales a un nivel de significación de 0,1.
b) ¿Podemos concluir con un nivel de significancia del 0,05 que el desgaste del material A excede el del material B en más de 2 unidades? Suponga que las poblaciones son normales.

***Solución*:**

$$n_A = 12; \bar{x}_A = 85; S_A = 4; n_B = 10; \bar{x}_B = 81; S_B = 5$$

a) Debemos en primer lugar contrastar la hipótesis de igualdad de varianzas

$$H_0 : \sigma_A^2 = \sigma_B^2$$

$$H_1 : \sigma_A^2 \neq \sigma_B^2$$

$$\frac{S_A^2}{S_B^2} \in \left(F_{1-\frac{\alpha}{2}, n_A-1, n_B-1}, F_{\frac{\alpha}{2}, n_A-1, n_B-1} \right)$$

$$\frac{S_A^2}{S_B^2} = \frac{4^2}{5^2} = 0.64 \in \left(F_{0.95,11,9}, F_{0.05,11,9} \right) = (0.345277309, 3,102485408)$$

EXCEL:=DISTR.F.INV(0,95;11;9) 0,34527731;
=DISTR.F.INV(0,025;11;9) 3,10248541
SPSS:IDF.F(0.05,1,9) ,34527731;IDF.F(0.95,11,9) 3,10248541

0,345 < 0,64 <3,1 y por tanto aceptamos la hipótesis de varianzas iguales.

b) Contrastamos ahora la diferencia de medias de dos poblaciones normales de varianzas desconocidas pero iguales y muestras pequeñas.

$$H_0 : \mu_A > \mu_B + 2 \Leftrightarrow \mu_A - \mu_B > 2$$

$$H_1 : \mu_A \leq \mu_B \Leftrightarrow \mu_A - \mu_B \leq 2$$

El estadístico de prueba es: $\dfrac{\left| \bar{x}_A - \bar{x}_B - (\mu_A - \mu_B) \right|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = t$ siendo

$$S^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)} = \frac{11 \cdot 4^2 + 9 \cdot 5^2}{11 + 9} = 20.05 \text{ con lo cual}$$

$$\frac{\left| \bar{x}_A - \bar{x}_B - 2 \right|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = \frac{85 - 81 - 2}{\sqrt{20.05}\sqrt{\dfrac{1}{12} + \dfrac{1}{10}}} = 1.043162796$$ y para α=0.05, $t_{0.025,20}$=2

EXCEL: =DISTR.T.INV(0,05;20) 2,01063472
SPSS:IDF.T(0.975,20) 2.01

Como 1.04 < 2 **ACEPTAMOS la hipótesis nula**.

15.- En medidas de ángulos con un cierto teodolito, un topógrafo asegura que la varianza que obtiene es igual o menor que 5. Se le pone a prueba y se le hacen 20 determinaciones, obteniéndose una varianza de 6. Si la variable medida del ángulo es normal. Se pide:
a) Intervalo de confianza del 99% para la varianza obtenida.
b) ¿Aceptamos su aseveración a un nivel de significación 0,01?

***Solución*:**

a) $P\left(\frac{(n-1).S^2}{k_2} < \sigma^2 < \frac{(n-1).S^2}{k_1}\right) = 1-\alpha$

Buscaremos los valores de k_1 y k_2 tales que: $\begin{matrix} P\left(\chi^2_{19} < k_1\right) = 0.005 \\ P\left(\chi^2_{19} < k_2\right) = 0.995 \end{matrix}$, obtenemos k_1=6,84397146 y k_2= 38,5822565

$$P\left(\frac{19\cdot 6}{38,5822565} < \sigma^2 < \frac{19\cdot 6}{6,84397146}\right) = 0,99 \Rightarrow \boxed{2,95 < \sigma^2 < 16,66}$$

b) H_0: $\sigma^2 \leq 5$.
H_1: $\sigma^2 > 5$

Aplicando el teorema de Fisher $(n-1)\frac{S^2}{\sigma^2} \equiv \chi^2_{n-1} \Rightarrow (20-1)\frac{6}{5} = 22.8$

$P\left(\chi^2_{n-1} > x_\alpha\right) = \alpha$ para n-1=19 y $\alpha = 0.01 \Rightarrow x_{\alpha=0.01} = 36.19$.

Por tanto 22.8<36.19 **se acepta la hipótesis nula**.

16.- Se tabulan los errores de cierre en nivelación obtenidos en 1000 polígonos. ¿Se puede admitir que el error de cierre se distribuye normalmente? Contrastar la bondad del ajuste mediante la χ^2_n con α=0,05.

error de cierre	n° de polígonos
0 -0,1	64
0,1-0,2	242
0,2-0,3	390
0,3-0,4	238
0,4-0,5	66

***Solución*:**

error	n_i	x_i	$x_i n_i$	$x_i^2 n_i$	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
<0	0				0,006209665	6,20966532	6,20966532
0 -0.1	**64**	0,05	3,2	0,16	0,060597536	60,5975359	0,19104344
0.1-0.2	**242**	0,15	36,3	5,445	0,241730337	241,730337	0,00030082
0.2-0.3	**390**	0,25	97,5	24,375	0,382924923	382,924923	0,13072203
0.3-0.4	**238**	0,35	83,3	29,155	0,241730337	241,730337	0,05756587
0.4-0.5	**66**	0,45	29,7	13,365	0,060597536	60,5975359	0,48164694
0.5<	0				0,006209665	6,20966532	6,20966532
	1000		**250**	**72,5**	**1**	**1000**	**13,280609**

Tenemos que calcular la media y la desviación típica de la distribución Normal. Para ello consideramos la muestra obtenida:

$$\bar{x} = \frac{\sum x_i n_i}{n} = 0{,}25;\ \sigma^2 = \frac{\sum x_i^2 n_i}{n} - \bar{x}^2 = 0{,}01$$

Para estimar la media y la desviación típica de la población utilizamos los resultados anteriores teniendo en cuenta que debemos considerar la cuasivarianza, es decir, $S^2 = \frac{n}{n-1}\sigma^2 = 0{,}01001001$ y por tanto s=0,1. Consideramos la población con distribución N(0.25,0.1).

La prueba de la bondad de ajuste de Pearson se basa en la distribución Chi-cuadrado con k-h-1 grados de libertad, en nuestro caso k=7 (nº de intervalos), h=2 (nº de parámetros) y para un nivel de significación $\alpha = 0.05$ resulta $P(\chi^2_{7-2-1} > 9{,}49) = 0{,}05$ y como D=$\frac{(n_i - np_i)^2}{np_i}$=13,28 > 9,49 **SE RECHAZA EL AJUSTE**.

17.- La siguiente tabla contiene los datos de dos muestras de tamaño 50. ¿Podemos admitir que tienen la misma distribución utilizando el estadístico de Pearson y calculando el p-valor?

1ª muestra	2ª muestra
3	1
2	4
4	9
11	5
11	10
9	9
5	4
3	4
2	4

Solución:

1ª muestra	2ª muestra	Frec. esperadas
3	1	2
2	4	3
4	9	6,5
11	5	8
11	10	10,5
9	9	9
5	4	4,5
3	4	3,5
2	4	3

Vemos que hay frecuencias esperadas menores que 5, luego agrupamos:

1ª muestra	2ª muestra	Frec. esperadas
5	5	5
4	9	6,5
11	5	8
11	10	10,5
9	9	9
10	12	11

$$D=\sum_{j=1}^{r}\sum_{i=1}^{k}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}=\frac{(5-5)^2}{5}+\frac{(4-6,5)^2}{6,5}+\frac{(11-8)^2}{8}+\frac{(11-10,5)^2}{10,5}+\frac{(9-9)^2}{9}+$$

$$+\frac{(10-11)^2}{11}+\frac{(5-5)^2}{5}+\frac{(9-6,5)^2}{6,5}+\frac{(5-8)^2}{8}+\frac{(10-10,5)^2}{10,5}+\frac{(9-9)^2}{9}+\frac{(12-11)^2}{11}$$

$$\approx 4,40251415$$

El estadístico D=4,4 con distribución de Pearson con (6-1)x(2-1)=5 grados de libertad resulta un $p=P(\chi_5^2\geq 4,4)\simeq 0,49$

No tenemos evidencia para rechazar la homogeneidad de las muestras.

Aceptamos la distribución común para todo nivel de significación menor que 0,49.

18.- A la vista de los datos que aparecen en la tabla adjunta, obtenidos después de una encuesta realizada para ver si la edad tiene influencia en los errores cometidos por los observadores. ¿Se puede admitir que el cometer errores es independiente de la edad?

Edad	Cometer error	No cometer error
<30	38	44
30-40	45	28
40-50	30	54
50-60	22	62
>60	20	57
	155	245

***Solución*:**

La probabilidad de cometer error independientemente de la edad será:

$P(\text{Error}) = \dfrac{155}{155+245} = 0.3875$. Calculamos la probabilidad de cometer error en cada intervalo correspondiente a la edad teniendo en cuenta el número de observaciones en cada intervalo.

Cometer error	No cometer error	Total	nP_i	$(n_i-nP_i)^2/nP_i$
38	44	82	31,775	1,21953186
45	28	73	28,2875	9,87388975
30	54	84	32,55	0,19976959
22	62	84	32,55	3,41943164
20	57	77	29,8375	3,24344889
155	245	400	0,3875	**17,9560717**

Utilizando el contraste de la bondad de ajuste de la distribución χ^2_{k-h-1} de Pearson con k-h-1 = 5-0-1 = 4 grados de libertad y observando el valor del estadístico $D = \sum_{i=1}^{5} \dfrac{(n_i - np_i)^2}{np_i} = 17{,}9560717$. $p = P(\chi^2_4 \geq 17{,}96) \simeq 0$

Para cualquier nivel de significación $\alpha > p$ se rechaza la hipótesis y consideramos que **el cometer errores no es independiente de la edad**.

19.- Se quieren comparar la vida útil de dos tipos de baterías para móviles, para ello obtenemos muestras de tamaño 9 de cada tipo, obteniendo los siguientes resultados $n_A = 9; \bar{x}_A = 5; S_A = 1,2; n_B = 9; \bar{x}_B = 3; S_B = 0,9$ **Contrastar la hipótesis de igualdad de medias al 95% suponiendo que la varianza es desconocida pero igual para las dos muestras.**

***Solución*:**

El estadístico de prueba es: $\dfrac{\left|\bar{x}_A - \bar{x}_B - (\mu_A - \mu_B)\right|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = t_{n_A + n_B - 2}$ siendo

$S^2 = \dfrac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)} = \dfrac{8 \cdot 1.2^2 + 8 \cdot 0.9^2}{8 + 8} = 1.125$ con lo cual:

$$\frac{\left|\bar{x}_A - \bar{x}_B\right|}{S\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = \frac{5 - 3}{\sqrt{1.125}\sqrt{\dfrac{1}{9} + \dfrac{1}{9}}} = 4$$

y para α=0.05, $t_{0.025,16}$=2,12

EXCEL: =DISTR.T.INV(0,05;16) 2,1199

SPSS:IDF.T(0.975,16) 2.12

Como 4 > 2 **RECHAZAMOS** la hipótesis de igualdad de medias.

20.- Al calcular cinco veces la distancia entre dos puntos, obtenemos los siguientes valores:

170,13m; 170,12m; 170,2m; 170,65m; 170,4

Se pide:

a) Intervalo de confianza del 80% para la media.

b) ¿Cuál será el número de mediciones necesaria para que el error sea inferior a 0,1 m con un nivel de significación de 0,2?

c) Intervalo de confianza del 90% para la varianza obtenida.

d) Si la variable medida del ángulo es normal ¿aceptamos una desviación típica inferior a 0,1 m con un nivel de significación 0,1?

Solución:

a) Datos: $n = 5; \overline{X} = 170,3;\; S^2 = 0,05095 \Rightarrow S = 0,22572107;\; \alpha = 0,2$

Tenemos una muestra de tamaño pequeño y varianza desconocida:

$$I_\alpha = \left(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

y en la distribución de Student

Buscaremos un valor $t_{\alpha/2}$ tal que $P\left(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}\right) = 1-\alpha \Rightarrow t_{0,2} = 1,53320627$.

$$\Rightarrow I_{\alpha=0,2} = \left(170,3 - 1,53320627\frac{0,22572107}{\sqrt{5}}, 170,3 + 1,53320627\frac{0,22572107}{\sqrt{5}}\right) =$$

$$\boxed{(170.145, 170.455)}$$

b)

$$\varepsilon = t_{\alpha/2}\frac{S}{\sqrt{n}} \Rightarrow n = \left(t_{\alpha/2}\frac{S}{\varepsilon}\right)^2 = 1,53^2\frac{0,05095}{0,1^2} \approx 11,92688549$$

Son necesarias $\boxed{12}$ mediciones

c)

$$P\left(\frac{(n-1).S^2}{k_2} < \sigma^2 < \frac{(n-1).S^2}{k_1}\right) = 1-\alpha$$

Buscaremos los valores de k_1 y k_2 tales que: $\begin{matrix} P\left(\chi_4^2 < k_1\right) = 0.05 \\ P\left(\chi_4^2 < k_2\right) = 0.95 \end{matrix}$, obtenemos

k_1= 0,7107230275 y k_2= 9,487729072

$$P\left(\frac{4 \cdot 0,05095}{9,487729072} < \sigma^2 < \frac{4 \cdot 0,05095}{0,7107230275}\right) = 0,9 \Rightarrow \boxed{0,0215 < \sigma^2 < 0,287}$$

d) H_0: $\sigma^2 \le 0,01$.

H_1: $\sigma^2 > 0,01$

Sabemos que $(n-1)\frac{S^2}{\sigma^2} \equiv \chi_{n-1}^2 \Rightarrow (5-1)\frac{0,05095}{0,1^2} = 20,38$

$P\left(\chi_{n-1}^2 > \chi_\alpha\right) = \alpha$ para n-1=4 y $\alpha = 0,1 \Rightarrow \chi_{\alpha=0,1} = 7,779440283$.

La región crítica es $\boxed{(7.779440283, +\infty)}$.

Por tanto 7.8<20,38 se rechaza la hipótesis nula, la desviación típica no es inferior a 0,1

21.- Se ha observado un ángulo cinco veces, obteniéndose los siguientes valores:
65º25'; 65º33'; 65º32'; 65º28'; 65º27'
Se pide:
a) Intervalo de confianza del 95% para la media del ángulo observado.
b) ¿Cuál será el número de observaciones necesaria para que el error sea inferior a 1 minuto con un nivel de significación de 0,05?
c) Intervalo de confianza del 99% para la varianza obtenida.
d) Si la variable medida del ángulo es normal ¿aceptamos una desviación típica inferior a 2 minutos a un nivel de significación 0.01? Escribir la región crítica.

Solución:

a)

Datos: $n = 5; \overline{X} = 65^{o}\,29'$; $S^2 = 11,5 \Rightarrow S = 3,39116499;\ \alpha = 0,05$

Tenemos una muestra de tamaño pequeño y varianza desconocida:

$$I_\alpha = \left(\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

y en la distribución de Student

Buscaremos un valor $t_{\alpha/2}$ tal que $P\left(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}\right) = 1-\alpha \Rightarrow t_{0,025} = 2,77644511$.

$$\Rightarrow I_{\alpha=0.05} = \left(65^{o}\,29' - 2,77644511\frac{3,39116499}{\sqrt{5}}, 65^{o}\,29' + 2,77644511\frac{3,39116499}{\sqrt{5}}\right) =$$

$$\boxed{\left(65^{o}24.79', 65^{o}\,33.21'\right)}$$

b)

$$\varepsilon = t_{\alpha/2}\frac{S}{\sqrt{n}} \Rightarrow n = \left(t_{\alpha/2}\frac{S}{\varepsilon}\right)^2 = 2,77^2\,\frac{11,5}{1^2} \approx 88,64944821$$

Son necesarias 89 observaciones

c)

$$P\left(\frac{(n-1).S^2}{k_2} < \sigma^2 < \frac{(n-1).S^2}{k_1}\right) = 1-\alpha$$

Buscaremos los valores de k_1 y k_2 tales que: $\begin{matrix} P\left(\chi_4^2 < k_1\right) = 0.005 \\ P\left(\chi_4^2 < k_2\right) = 0.995 \end{matrix}$, obtenemos k_1=0,20896909 y k_2= 14,860259

$$P\left(\frac{4\cdot 11,5}{14,860259} < \sigma^2 < \frac{4\cdot 11,5}{0,20896909}\right) = 0,99 \Rightarrow \boxed{3,10 < \sigma^2 < 222,23}$$

d) H_0: $\sigma^2 \le 2^2$.
H_1: $\sigma^2 > 2^2$

Sabemos que $(n-1)\frac{S^2}{\sigma^2} \equiv \chi^2_{n-1} \Rightarrow (5-1)\frac{11,5}{2^2} = 11,5$

$P\left(\chi^2_{n-1} > \chi_\alpha\right) = \alpha$ para n-1=4 y $\alpha = 0,01 \Rightarrow \chi_{\alpha=0,01} = 13,28$.

La región crítica es $(13.28, +\infty)$.

Por tanto 11,5<13,28 se acepta la hipótesis nula, la desviación típica es inferior a 2

22.- Se quiere comprobar si es cierto que un nuevo modelo de automóvil consume 5,7 litros por cada 100 km a 90 km/h tal y cómo afirma en su propaganda. Para ello se realizan 10 recorridos de 100 km a dicha velocidad, con los siguientes consumos:

5,8; 6,3; 6; 5,5; 6,1; 6,5; 5,5; 7; 6,8; 5,7

Se pide:

a) Intervalo de confianza para la media obtenida con un nivel de significación del 0,05.

b) ¿Se puede aceptar la afirmación del fabricante al 95%?

Solución:

a) $\overline{X}=\frac{\sum x_i}{n}=\frac{5,8+6,3+\ 6+\ 5,5+\ 6,1+\ 6,5+\ 5,5+\ 7+\ 6,8+\ 5,7}{10}=6,12$

$$S^2=\frac{\sum\left(x_i-\overline{X}\right)^2}{n-1}=0,27511\Rightarrow S=\sqrt{0,27511}\approx 0,52451035$$

Sabemos que el estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue una distribución t de Student con n-1 grados de libertad.

Así pues: $\frac{6,12-\mu}{0,52451035/\sqrt{9}}=t_{10-1}$,

Buscaremos el intervalo $I_\alpha=\left(\overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}},\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$, es decir,

$P\left(\overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right)=1-\alpha$.

En nuestro caso, queda:

$$I_{\alpha=0,05}=\left(6,12-2,26215726\frac{0,52451035}{\sqrt{10}},6,12+2,26215726\frac{0,52451035}{\sqrt{10}}\right)\approx$$

$$\boxed{(5.75,6.50)}$$

b)

$$H_0:\mu\le 5,7$$
$$H_1:\mu>5,7$$

$$t_{n-1}=t_{10-1}=\frac{6,12-5,7}{0,52451035/\sqrt{10}}=2,5$$

Buscaremos un valor t_α tal que $P\left(t_{n-1}<t_\alpha\right)=1-\alpha$. En nuestro caso t=1,8, por lo que se **RECHAZA** la hipótesis nula.

23.- Un jugador afirma que los dos dados que utiliza no están trucados. Para comprobar esta afirmación, se lanzan los dos dados 360 veces y se anota la suma de los resultados. Dichas sumas se muestran en la siguiente tabla, junto con su probabilidad correspondiente (para facilitar la comparación).

Suma	FRECUENCIA OBSERVADA	PROBABILIDAD
2	11	1/36
3	18	2/36
4	33	3/36
5	41	4/36
6	47	5/36
7	61	6/36
8	52	5/36
9	43	4/36
10	29	3/36
11	17	2/36
12	8	1/36

Para poder decidir si los dados están trucados o no, realizar un contraste χ^2. ¿Deberíamos jugar con estos dados?

Solución:

SUMA	n_i	P_i	nP_i	n_i-nP_i	$(n_i-nP_i)^2$	$(n_i-nP_i)^2/nP_i$
2	11	1/36	10	1	1	0.1
3	18	1/18	20	-2	4	0.2
4	33	1/12	30	3	9	0.3
5	41	1/9	40	1	1	0.025
6	47	5/36	50	-3	9	0.18
7	61	1/6	60	1	1	1/60
8	52	5/36	50	2	4	0.08
9	43	1/9	40	3	9	0.225
10	29	1/12	30	-1	1	1/30
11	17	1/18	20	-3	9	0.45
12	8	1/36	10	-2	4	0.4
	360	1	360			2.01

$D = \sum_{i=1}^{11} \frac{(n_i - nP_i)^2}{nP_i} = 2.01$. Cuando la distribución está determinada no hay parámetros que estimar y por tanto utilizamos $\chi^2_{11-0-1} = \chi^2_{10}$; siendo $p-\text{valor} = P(\chi^2_{10} \geq 2.01) = 0.9962629352$ y por tanto próximo a 1 y se acepta la hipótesis nula "**los dados no están trucados**". Podemos jugar con estos dados.

24.- Durante 50 días se ha observado la variable "número diario de cancelaciones de cuentas en una sucursal bancaria" anotándose la siguiente tabla de frecuencias:

Nº cancelaciones	Frecuencias observadas n_i
0	16
1	23
2	8
3	3
Sumas	50

Contrastar el hecho de que la distribución es de Poisson de parámetro λ=1.

Solución:

En el caso de la distribución de Poisson la media coincide con la varianza que es el valor del parámetro λ de la distribución. Para esta muestra utilizaremos $\lambda = 1$ y calculamos las probabilidades teniendo en cuenta las frecuencias esperadas $np_i \geq 5$.

Nº cancelaciones	Frecuencias observadas n_i	p_i	Frecuencias esperadas np_i	Frecuencias esperadas con $np_i \geq 5$	$\frac{(n - np_i)^2}{np_i}$
0	16	0,3678794	18,39397	18,39397	0,31157452
1	23	0,3678794	18,39397	18,39397	1,15339496
2	8	0,1839397	9,196985	13,21206	0,37035931
3	3	0,0803015	4,015075		
Sumas	**50**	**1**	**50**		**1,83532879**

Utilizando el contraste de la bondad de ajuste de la distribución χ^2_{k-h-1} de Pearson con k-h-1 = 3-1-1 = 1 grados de libertad y observando el valor del estadístico $D = \sum_{i=1}^{3} \frac{(n_i - np_i)^2}{np_i} = 1{,}8353287$.

Buscamos el p-valor $= P(\chi_1^2 \geq D) = P(\chi_1^2 \geq 1{,}8353287) = 0{,}175$

Para cualquier nivel de significación $\alpha < 0{,}175 = p$ **se acepta la hipótesis y consideramos que la distribución es de Poisson**.

25.- Para curar una enfermedad se sabe que existen cuatro tratamientos diferentes. Aplicados por separado a unos grupos de enfermos se han observado los resultados siguientes:

Tratamiento	Curados	No Curados
A	83	44
B	45	28
C	60	54
D	82	62
Total	270	188

¿Se puede asegurar que la eficacia de los cuatro tratamientos es la misma, con un nivel de confianza del 95%?

***Solución*:**

La probabilidad de curar independientemente del tratamiento será: $P(\text{Curar}) = \dfrac{270}{270+188} = 0,58951965$. Calculamos la probabilidad de curar en cada tratamiento teniendo en cuenta el número de pacientes con cada tratamiento.

Curados	No Curados	Total	np_i	$(n_i-np_i)^2/np_i$
83	44	127	74,8689956	0,88305221
45	28	73	43,0349345	0,08972902
60	54	114	67,2052402	0,77249164
82	62	144	84,8908297	0,09844286
270	**188**	**458**		**1,84371573**

Utilizando el contraste de la bondad de ajuste de la distribución χ^2_{k-h-1} de Pearson con k-h-1 = 4-0-1 = 3 grados de libertad y observando el valor del estadístico $D = \sum_{i=1}^{5} \dfrac{(n_i - np_i)^2}{np_i} = 1,84371573$.

Para $\alpha = 0,05$ $P(\chi^2_3 \geq \chi_{0,05}) = 0,05 \Rightarrow \chi_{0.05} = 7,8$ siendo D = 1,8 <7,8

se ACEPTA la hipótesis y consideramos que los tratamientos son igual de eficaces.

26.- Si el coeficiente medio de inteligencia de la población universitaria de la U.P.M. es $\mu = 95$ y σ=14, y se extrae una muestra de 49 estudiantes de esa población y el resultado de la varianza muestral S^2=100. ¿Podemos concluir con un nivel de significación del 0,05 que la varianza es menor de 14^2?

Solución:

Datos: $S^2 = 100; \sigma = 14; n = 49; \alpha = 0,05$

$$H_0:\ \sigma^2 < 14^2$$
$$H_1:\ \sigma^2 \geq 14^2$$

Aplicando el teorema de Fisher $(n-1)\dfrac{S^2}{\sigma^2} \equiv \chi^2_{n-1} \Rightarrow \chi = (49-1)\dfrac{100}{14^2} = 24{,}49$

$$P\left(\chi^2_{n\text{-}1} > x_\alpha\right) = \alpha \Rightarrow P\left(\chi^2_{48} > x_{\alpha=0,05}\right) = 0,05 \Rightarrow x_{\alpha=0,05} = 65{,}1707689$$

Por tanto $x_{\alpha=0,05} = 65{,}1707689 > 24{,}49 = \chi \Rightarrow$ **se ACEPTA la hipótesis nula**.

27.- Se lleva a cabo un experimento para comparar el desgaste por abrasivo de un material laminado. Se prueban 12 piezas del material mediante la exposición de cada pieza a una máquina para medir la profundidad del desgaste. La muestra da un desgaste medio de 53 unidades con una desviación estándar muestral de 4. Supongamos que la población es aproximadamente normal con varianza desconocida.
a) ¿Podemos concluir con un nivel de significación del 0,05 que el desgaste abrasivo medio del material es menor de 50 unidades?
b) Se justifica la suposición de varianza igual a 10 con un nivel de significación de 0,1.

Solución:

Datos: $n = 12; \overline{X} = 53;\ S = 4; \alpha = 0.05$

a) Sabemos que el estadístico $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue una distribución t de Student con n-1 grados de libertad.

$$H_0 : \mu < 50$$
$$H_1 : \mu \geq 50$$

Así pues:

$$t = \frac{53-50}{4/\sqrt{12}} = 2{,}598$$

Buscaremos un valor t_α tal que $P\left(t_{n-1} < t_\alpha\right) = 1-\alpha$. En nuestro caso aproximadamente $t_{0,05}=1{,}8<2{,}598=t$, por lo que se **RECHAZA** la hipótesis nula.

b) H_0: $\sigma^2=10$
H_1: $\sigma^2 \neq 10$

Aplicando el teorema de Fisher $(n-1)\dfrac{S^2}{\sigma^2} \equiv \chi^2_{n-1} \Rightarrow (12-1)\dfrac{16}{10} = 17{,}6$

Buscaremos los valores de k_1 y k_2 tales que:
$$P\left(\chi^2_{n-1} < k_1\right) = \frac{\alpha}{2} \Rightarrow P\left(\chi^2_{11} < k_1\right) = 0{,}05$$
$$P\left(\chi^2_{n-1} < k_2\right) = 1-\frac{\alpha}{2} \Rightarrow P\left(\chi^2_{11} < k_2\right) = 0{,}95$$
,

obtenemos k_1= 4,57481308 y k_2=19,6751376 para n-1=11 y $\alpha = 0{,}1$. -

Por tanto 4,6<17,6<19,7 **se ACEPTA la hipótesis nula**.

28.- Al realizar un muestreo aleatorio con 100 alumnos de la U.P.M para estudiar la estatura de los alumnos de esta Universidad, se obtuvieron los siguientes resultados agrupados en intervalos de amplitud 10 cm, así como los parámetros de la muestra $\overline{X} = 1.772$ y S = 0.1168.

	n_i
menos de 1.5	0
1.5 - 1.6	6
1.6 - 1.7	21
1.7 - 1.8	35
1.8 - 1.9	24
1.9 - 2.0	11
2.0 - 2.1	3
más de 2.1	0

Contrastar la hipótesis,
La muestra obtenida sigue una distribución normal **N($\overline{X} = 1.772$, S = 0.1168)**

a) Obtener el p-valor y dar su interpretación.
b) Para un nivel de significación $\alpha = 0.05$ calcular el valor crítico.
c) ¿Se acepta el ajuste para un nivel de significación $\alpha = 0.05$?

Solución:

Supuestamente hemos calculado la media y la desviación típica de la muestra obtenida:

$$\overline{X} = \frac{\sum x_i n_i}{n} = 1{,}772; \ \sigma = 0{,}1168$$

Consideramos la población con distribución N(1.772,0.1168).

n_i	$P_i = F(e_i) - F(e_{i-1})$	nP_i	$nP_i > 5$	$(n_i - nP_i)^2 / nP_i$
0	0,00993570	0,994		
6	0,06049339	6,049	7,043	0,15443331
21	0,19837353	19,837	19,837	0,06814156
35	0,32592605	32,593	32,593	0,17781797
24	0,26870796	26,871	26,871	0,30670732
11	0,11109752	11,110	13,656	0,00864832
3	0,02297503	2,298		
0	0,00249082	0,249		
100	1,000	100	100	**0,71574848**

$$D = \sum \frac{(n_i - np_i)^2}{np_i} \approx 0{,}71574848$$

a) La prueba de la bondad de ajuste de Pearson se basa en la distribución Chi-cuadrado con k-h-1 grados de libertad, en nuestro caso k=5 (nº de intervalos), h=2 (nº de parámetros)

p-valor= $P(\chi^2_{5-2-1} > 0{,}71574848) = \boxed{0{,}699160997}$ Se acepta para cualquier nivel de significación menor.

b) Para un nivel de significación $\alpha = 0.05$ resulta $P(\chi^2_{5-2-1} > x_{0,05}) = 0{,}05 \Rightarrow$

$x_{0,05} = \boxed{5{,}99}$

c) y como D=$\dfrac{(n_i - np_i)^2}{np_i}$ =0,716 $< 5{,}99$ **SE ACEPTA EL AJUSTE**.

29.- Para contrastar que la esperanza matemática μ de una distribución Normal es 10, se toma una muestra de tamaño 16 y se rechaza la hipótesis en el caso en que la media muestra sea mayor que 10, aceptándose en caso contrario. Sabiendo que la desviación típica de la población es σ=2 y habiendo obtenido una media muestral de 11. Calcular el p-valor y decidir si se acepta la hipótesis de que la media de la población es μ<10

Solución:

Se trata de un contraste de hipótesis para la media de una población normal de varianza conocida:

$$H_0 : \mu \leq 10$$
$$H_1 : \mu > 10$$

Sabemos que: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \equiv N(0,1)$

El valor del estadístico z bajo la hipótesis nula es:

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{11 - 10}{2 / \sqrt{16}} = 2$$

$$p = P\left(Z > 2\right) = 0{,}0227$$

Por ser muy próxima a cero RECHAZAMOS la hipótesis nula, y la media no puede ser menor que 10.

30.- Para los siguientes datos

x_i	n_i
0	10
1	30
2	50
3	10

Se pide:

a) Calcular las respectivas frecuencias teóricas correspodientes a una distribución Binomial de parámetros n=3 y p=0,53, es decir, $p_i=P(X=x_i)$

b) Hallar el valor del estadístico de contraste:

$$D=\sum_i \frac{(n_i-np_i)^2}{np_i}$$

c) Estudiar la Bondad del ajuste para un nivel de significiación α=0,05

Solución:

a) A partir de de la distribución Binomial B(3,0.53) calculamos las frecuencias teóricas:

$$P(X=k)=\binom{n}{k}p^k . q^{n-k}=\binom{3}{k}0,53^k(1-0,53)^{3-k}$$

$$P(X=0)=\binom{3}{0}0,53^0(1-0,53)^{3-0}\approx 0,1$$

$$P(X=1)=\binom{3}{1}0,53^1(1-0,53)^{3-1}\approx 0,35$$

$$P(X=2)=\binom{3}{2}0,53^2(1-0,53)^{3-2}\approx 0,39$$

$$P(X=3)=\binom{3}{3}0,53^3(1-0,53)^{3-3}\approx 0,16$$

b)

	n_i	$n\,p_i$	(n_i-np_i)	$(n_i-np_i)^2$	$\frac{(n_i-np_i)^2}{np_i}$
0	10	10	0	0	0
1	30	35	5	25	0,714
2	50	39	11	121	3,1
3	10	16	6	36	2,3
Σ	**100**	**100**			**6,114**

$$D=\sum_i \frac{(n_i-np_i)^2}{np_i}=6,114$$

c) Para α = 0,05 $\Rightarrow P(\chi^2_{4-2-1} \geq x_{0,05}) = 0,05 \Rightarrow x_{0,05} = 3,84$ siendo D = 6,114 menor que $\chi_{0,05}$ **aceptamos** la hipótesis de ser el ajuste bueno.

31.- Suponiendo una población normal con media 50 y desviación típica 8. Para una muestra de tamaño 9, ¿cuál es la probabilidad de que:
a) La media muestral sea superior a 55?
b) La desviación típica muestral sea superior a 10?

Solución:

Sabemos que la media muestral $\overline{X} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Para nuestros datos: $\overline{X} = 50; \sigma = 8; n = 9$

Tenemos $\overline{X} \equiv N\left(50, \frac{8}{\sqrt{9}}\right) = (50, 8/3)$

a) $P\left(\overline{X} \geq 55\right) = 1 - P\left(\overline{X} < 55\right) = 1 - F_{\overline{X}}(55) \approx \boxed{0{,}03039636175}$

b) Sabiendo que $\frac{(n-1).S^2}{\sigma^2} \equiv \chi^2_{n-1}$ si la población de partida es $N(\mu,\sigma)$

En nuestro caso $\frac{(9-1).S^2}{8^2} \equiv \chi^2_{9-1}$

$$P\left(S^2 > 10^2\right) = P\left(\frac{S^2}{8} > \frac{10^2}{8}\right) = P\left(\chi^2_8 > 12{,}5\right) = 1 - P\left(\chi^2_8 > 12{,}5\right) = 1 - F_{\chi^2_8}\left(12{,}5\right) \approx \boxed{0.13025}$$

32.- Según un estudio sobre los hábitos de consumo de las familias de la Comunidad de Madrid, el gasto mensual en productos de marcas blancas de las familias puede considerarse según una variable aleatoria *X* con ley Normal de desviación típica 50 euros. Los datos siguientes corresponden a una muestra aleatoria simple de tamaño *n* = 10 familias:

Familia	1	2	3	4	5	6	7	8	9	10
Gasto	325	193	203	305	251	240	233	238	354	311

a) ¿Qué estimación del gasto medio mensual se obtiene con un estimador insesgado?
b) Calcula un intervalo de confianza al 95% para el gasto medio mensual en productos de marcas blancas de las familias de la Comunidad de Madrid.
c) Si queremos trabajar con niveles de confianza del 90%, ¿obtendríamos un intervalo más o menos) amplio que el calculado anteriormente?

Solución:

a) La media muestral es un estimador insesgado para la media poblacional

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = 265,3$$

b) El intervalo de confianza para una población normal de varianza conocida es:

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Para nuestros datos:

$$\overline{X} = 265,3; \sigma = 50; n = 10; \alpha = 0.05$$

$$\text{Tenemos } Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \equiv N(0,1)$$

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha = 1-0.05 = 0.95 \Rightarrow F(z_{\alpha/2}) = P\left(Z < z_{\alpha/2}\right) = 0.975 \Rightarrow z_{\alpha/2} \approx 1.96$$

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 265,3 \pm 33,8463$$

$$\Rightarrow I_{\alpha=0.05} = \left(265.3 - 1.96\frac{50}{\sqrt{10}}, 265.3 + 1.96\frac{50}{\sqrt{10}}\right) = \boxed{(255.5, 275.1)}$$

c) Al disminuir el nivel de confianza disminuye la amplitud del intervalo

33.- Se sabe que la demanda diaria de un producto fabricado por cierta empresa sigue una distribución normal. Se toma una muestra aleatoria simple de la demanda registrada en 10 días distintos, obteniéndose una media muestral igual a 3,5 y una desviación típica muestral (o cuasidesviación típica) igual a 0,9. Se pide:
a) Un intervalo de confianza al 95% para la media de la demanda diaria de dicho producto.
b) El plan de producción de la empresa está basado en la suposición de que la demanda media diaria del citado producto es como mínimo igual a 4. Contrastar esta hipótesis, al nivel de significación de 0,05.

Solución:

a) Se trata de una población que sigue una distribución Normal de varianza desconocida, y muestras pequeñas, por lo que el intervalo de confianza es:

$$\overline{X} \pm t_{\alpha/2}\frac{S}{\sqrt{n}}$$

Datos: $n = 10; \overline{X} = 3,5$; $S = 0,9;$ $\alpha = 0.05$

Buscaremos un valor t_α tal que $P(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}) = 1-\alpha$.

$$P(|t_{n-1}| < t_{\alpha/2}) = 1-\alpha \Leftrightarrow P(|t_{n-1}| > t_{\alpha/2}) = \alpha \Rightarrow P(|t_9| > t_{\alpha/2=0,025}) = 0,05 \Rightarrow t_{0,025} = 2,262157127$$

$$\Rightarrow I_{\alpha=0.05} = \left(3,5-2,26\frac{0,9}{\sqrt{10}}, 3,5+2,26\frac{0,9}{\sqrt{10}}\right) = \boxed{(2.86, 4.14)}$$

b) Realizamos la hipótesis nula de que la demanda diaria es mayor que 4.

$$H_0 : \mu \geq 4$$
$$H_1 : \mu < 4$$

El estadístico es: $t=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} = \dfrac{3,5-4}{0,9/\sqrt{10}} \simeq -2,002775851$

Ahora se trata de un valor t_α tal que $P(t_{n-1} < t_\alpha) = \alpha$.

$$\Rightarrow P(t_9 < t_{\alpha=0,05}) = 0,05 \Rightarrow t_{\alpha=0,05} = -1,833112933$$

$t_{\alpha=0,05} = -1,833112933 > -2,002775851 = t_9$, luego se RECHAZA la hipótesis nula

34.- Se supone que el retraso de los trabajadores de una empresa es, por término medio, de 5 minutos; para contrastar dicha hipótesis se seleccionó al azar una muestra de 10 empleados y se midió su retraso: 0.1, 2, 7, 5.6, 7.4, 5.1, 6.1, 6, 4.5 y 9. Utilizar el p-valor.

Solución:

Se trata de una población que sigue una distribución Normal de varianza desconocida, y muestras pequeñas, por lo que la distribución a emplear es t_{n-1}

Datos: $n = 10; \overline{X} = 5,28; \ S = 2,6$

Realizamos la hipótesis nula de que la media diaria es 5.

$$H_0 : \mu = 5$$
$$H_1 : \mu \neq 5$$

$$|t| = \left| \frac{\overline{X} - \mu_0}{S / \sqrt{n}} \right| = \left| \frac{5,28 - 5}{2,6 / \sqrt{10}} \right| \simeq 0,3405529787$$

Calcularemos el p-valor

$p = 1 - P\left(-0,3405529787 < t_{n-1} < 0,3405529787\right) = 2P\left(0,3405529787 < t_{n-1}\right) = 0,74$.

luego se ACEPTA la hipótesis nula para cualquier nivel de significación menor que 0,74

35.- La siguiente tabla de contingencia contiene datos sobre el gasto mensual en llamadas telefónicas (en euros) y el sexo de 500 encuestados:

Gasto	Hombres	Mujeres
[0,6)	59	76
[6,12)	42	62
[12,30)	54	51
[30,60)	67	39
[60,120]	28	22

Justificar si es cierto que el gasto mensual medio de los hombres es superior al de las mujeres con un nivel de significación del 5%.

Solución: Debemos en primer lugar contrastar la hipótesis de igualdad de varianzas

$$H_0: \sigma_h^2 = \sigma_m^2$$
$$H_1: \sigma_h^2 \neq \sigma_m^2$$

$$\frac{S_h^2}{S_m^2} \in \left(F_{1-\frac{\alpha}{2}, n_h-1, n_m-1}, F_{\frac{\alpha}{2}, n_h-1, n_m-1} \right)$$

Datos:

$n_h = 250; \overline{X}_h = 28,896; \ S_h^2 = 728,824482; n_m = 250; \overline{X}_m = 22,368; \ S_m^2 = 643,735518$

$$\frac{S_h^2}{S_m^2} = \frac{728,824482}{643,735518} = 1,13218 \in \left(F_{0.975,249,249}, F_{0.025,249,249} \right) = (0.7795917, 1.28272281)$$

0,78 < 1,13 < 1,28 y por tanto **aceptamos la hipótesis de varianzas iguales**.

Contrastamos ahora la diferencia de medias de dos poblaciones normales de varianzas desconocidas pero iguales y muestras pequeñas.

$$H_0: \mu_h > \mu_m \Leftrightarrow \mu_h - \mu_m > 0$$
$$H_1: \mu_h \leq \mu_m \Leftrightarrow \mu_h - \mu_m \leq 0$$

El estadístico de contraste es: $\dfrac{\overline{x}_h - \overline{x}_m}{S\sqrt{\dfrac{1}{n_h} + \dfrac{1}{n_m}}} = t$ siendo

$$S^2 = \frac{(n_h - 1)S_h^2 + (n_m - 1)S_m^2}{(n_h - 1) + (n_m - 1)} = \frac{249 \cdot 728,824482 + 249 \cdot 643,735518}{249 + 249} = 686,28 \text{ con lo cual}$$

$$\frac{\overline{x}_h - \overline{x}_m}{S\sqrt{\dfrac{1}{n_h} + \dfrac{1}{n_m}}} = \frac{28,896 - 22,368}{\sqrt{686,28}\sqrt{\dfrac{1}{250} + \dfrac{1}{250}}} = 2,78602161 \text{ y para } \alpha\text{=0.05, resulta}$$

$t_{0.05,498} = -1.647919 < 2,78602161$ **ACEPTAMOS** la hipótesis de que el gasto medio de los hombres es superior al de las mujeres.

36.- ¿Se puede admitir la distribución uniforme de valores angulares en una triangulación de primer orden de un país en la que se ha tomado una muestra de tamaño 100 y se han obtenido los siguientes resultados? Utilizar el p-valor.

Medidas angulares	frecuencias observadas
<40	16
40-50	22
50-60	20
60-70	19
>70	23

Solución:

Consideramos la distribución uniforme de parámetros a=30 y b=80.

Clases	n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
< 40	16	20	-4	16	0,8
$[40,50)$	22	20	2	4	0,2
$[50,60)$	20	20	0	0	0
$[60,70)$	19	20	-1	1	0,05
$70 <$	23	20	3	9	0,45
sumas	100	100			1,5

Utilizando el contraste de la bondad de ajuste de la distribución χ^2_{k-h-1} de Pearson con k-h-1 = 5-0-1 = 4 grados de libertad y observando el valor de la distribución:

$$D=\sum_{i=1}^{5}\frac{(n_i - np_i)^2}{np_i}=1,5.$$

$$P-\text{valor} = P(\chi^2_4 \geq 1,5) = 0,82664147 \text{ próximo a } 1$$

Se **acepta** la hipótesis de uniformidad de la muestra.

37.- Un proceso industrial fabrica piezas cuya longitud en milímetros se distribuye según una Normal de media 190 y desviación típica 10.
Una muestra de 5 piezas proporciona los resultados siguientes: 185, 200, 186, 197 y 185. Utilizando un nivel de significación del 0,05, se pide:
a) Contrastar la hipótesis de que la media del proceso es 190.
b) Contrastar la hipótesis de la varianza del `proceso es 100.

Solución:

a) Se trata de un contraste de hipótesis para la media de una población normal de varianza desconocida:

$$H_0 : \mu = 190$$
$$H_1 : \mu \neq 190$$

Estimamos los parámetros de la distribución Normal a partir de la muestra:

X_i	$(Xi\text{-media})^2$
185	34225
200	40000
186	34596
197	38809
185	34225
953	**181855**
$\overline{X}=190{,}6$	$S^2=53{,}3$

Datos: $n = 5; \overline{X} = 190,6$; $S = 7,3;\ \alpha = 0,05$

$$t_{n-1}=\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| = \left|\frac{190,6-190}{7,3/\sqrt{5}}\right|=0{,}1837864091$$

Buscaremos un valor $t_{\alpha/2}$ tal que $P\left(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}\right) = 1-\alpha$.

Puesto que $|t| = 0{,}1837864091<2{,}776$, **Se acepta la hipótesis nula**

b) Para contrastar la hipótesis nula:

$$H_0 : \sigma^2 = 100$$
$$H_1 : \sigma^2 \neq 100$$

Buscaremos los valores de k_1 y k_2 tales que: $\begin{matrix} P\left(\chi_4^2 < k_1\right) = 0,025 \\ P\left(\chi_4^2 < k_2\right) = 0,975 \end{matrix}$, obtenemos $k_1=0{,}4844$ y $k_2= 11{,}14$

Aplicando el teorema de Fisher $(n-1)\dfrac{S^2}{\sigma^2} \equiv \chi_{n-1}^2 \Rightarrow \chi = (5-1)\dfrac{53,3}{100} = 2{,}132$

Por tanto $0,4844 < \chi_{\alpha=0,05} = 2{,}132<11{,}14$ **se ACEPTA la hipótesis nula**.

38.- Se quiere contrastar la hipótesis nula "una moneda de un euro está bien equilibrada", y que por tanto, cara y cruz son equiprobables; para ello se realizan 100 lanzamientos de una moneda de un euro y se observan 63 caras y 37 cruces. Contrastar la hipótesis utilizando el test de χ^2 con un nivel de significación α de 0,05.

Solución:

	Frecuencias observadas n_i	p_i	Frecuencias esperadas np_i	$\frac{(n_i - np_i)^2}{np_i}$
Cara	63	0,5	50	3,38
Cruz	37	0,5	50	3,38
Sumas	**100**	**1**	**100**	**6,76**

Utilizando el contraste de la bondad de ajuste de la distribución χ^2_{k-h-1} de Pearson con k-h-1 = 2-0-1 = 1 grados de libertad y observando el valor de la distribución para $\alpha = 0.05$, $P(\chi_1^2 > x) = 0{,}05$, se obtiene x=3,84 que es mayor que D= $\sum_{i=1}^{2} \frac{(n_i - np_i)^2}{np_i}$.=6,76. Se **rechaza** la hipótesis nula.

Contraste de hipótesis.

Contrastar una hipótesis estadísticamente es tomar una decisión sobre si cierta propiedad de una población es compatible con lo observado en una muestra de dicha población.

Llamamos ***hipótesis nula***, H_0, a la hipótesis que vamos a contrastar, H_0 representa la hipótesis que mantendremos mientras los datos no nos indiquen su falsedad.

Llamamos ***hipótesis alternativa***, H_1, a la hipótesis que se aceptará si H_0 se rechaza.

Las fases en un contraste de hipótesis son:

1) Definir la hipótesis a contrastar que llamaremos H_0.

2) Definir una medida de discrepancia D que mida la diferencia entre los valores observados y los esperados (de acuerdo con H_0).

3) Calcular D. Si la discrepancia D es muy grande, rechazaremos H_0; en caso contrario, aceptamos H_0.

Nivel de significación

Número α entre 0 y 1 que se elige subjetivamente en estadística para acotar superiormente el error de tipo I en un test, o bien para construir intervalos o regiones de confianza.

- En un intervalo de confianza del estimador de forma I_α tal que la $P(\theta \in I_\alpha) = 1 - \alpha$.
- En un contraste de hipótesis es el valor α es la probabilidad de cometer un error de tipo I, y determina un valor d_c de forma que: $\boxed{P(D > d_c) = \alpha}$

La figura siguiente muestra gráficamente este método. Si la discrepancia observada D cae dentro de la región de rechazo (probabilidad de rechazar y ser verdadera), rechazamos la hipótesis H_0, en caso contrario la aceptaremos.

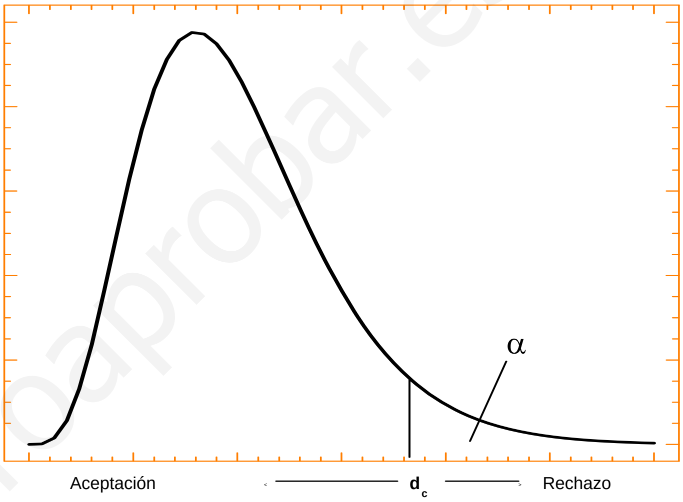
Definimos la ***región de rechazo*** por $D > d_c$ y la ***región de aceptación*** *de* H_0 será $D \le d_c$

Consideraciones acerca de α.

1) Aceptar o rechazar la hipótesis H_0 puede depender del valor α, siendo posible rechazar H_0 con $\alpha = 0.05$ y aceptar H_0 con $\alpha = 0.04$

2) Dar sólo el resultado del test no indica el grado de discrepancia. Se acostumbra a utilizar niveles de significación del 0.05 ó 0.01.

Si, por ejemplo se elige un nivel de significación del 0.05 entonces hay aproximadamente 5 ocasiones de cada 100 en que se rechazaría la hipótesis cuando debe ser aceptada.



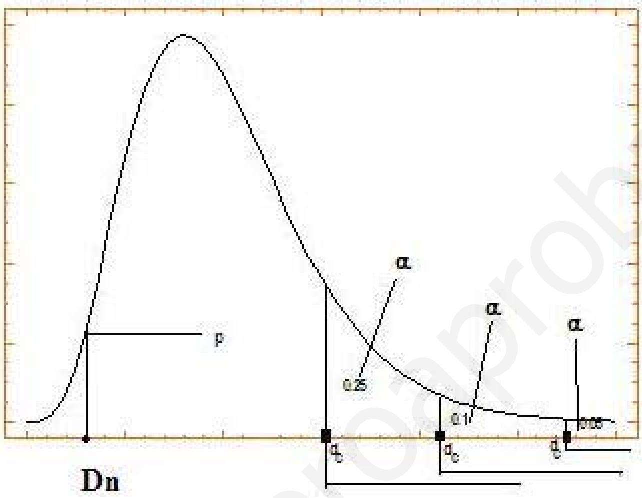
Nivel crítico

Se define el **nivel crítico** o **p valor** como el mínimo nivel de significación para el que, con los datos de una muestra concreta, se tendría que rechazar la hipótesis nula H_0 .

$$\boxed{p = P(D \geq D_n)}.$$

Es decir la probabilidad de obtener una discrepancia mayor o igual que la observada en la muestra. De esta forma, el valor de p no se fija a priori, sino que se determina en función de la muestra.

D es el estadístico y D_n es valor del estadístico obtenido con la muestra. Como se evidencia en la figura siguiente, cuanto menor sea el valor crítico, menor es la probabilidad de existir discrepancia como la observada, y menor es la certidumbre de H_0.



Esto es; cuanto más cercano a cero sea su valor con mayor confianza se rechazará H_0. Puesto que, $\boxed{p = P(D \geq D_n)}$ y D_n un valor fijo, si p es grande $\Rightarrow$ D_n es un valor pequeño, por tanto, para un valor fijo de $\alpha < p$ será $D_n < d_c$ y aceptamos la hipótesis H_0,

En general cuanto más próximo a 1 sea p con mayor evidencia se habrá de aceptar H_0 .

Si $p>0.25$ no existe suficiente evidencia para rechazar H_0.

Si $0.01<p<0.25$ existe incertidumbre entre rechazar o no rechazar H_0.

Si $p<0.01$ en general deberá ser rechazada la hipótesis H_0,

Si se ha fijado de antemano un nivel de significación α, se acepta H_0, si $p>\alpha$, y se rechaza H_0 si $p< \alpha$

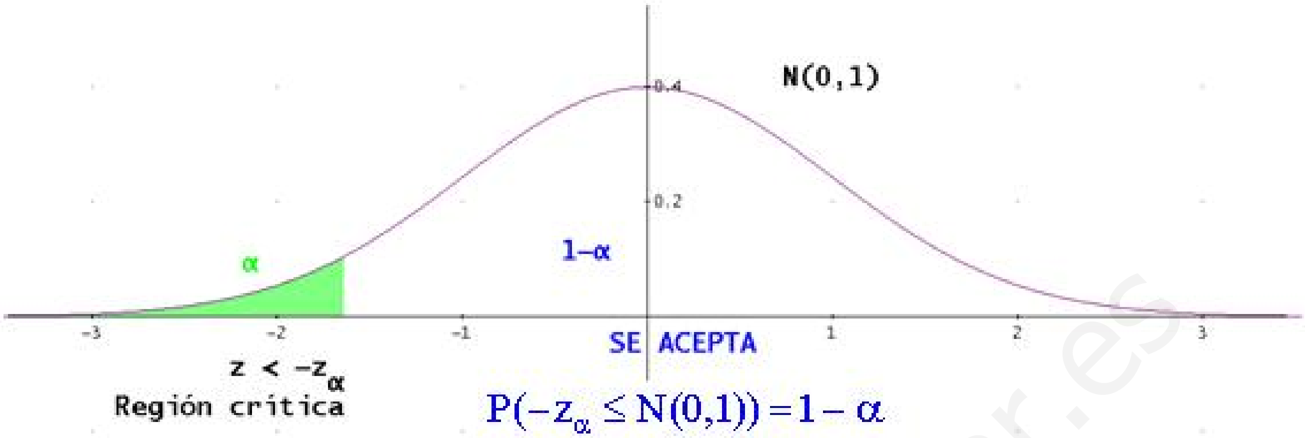
Pruebas relativas a la media

$H_0 : \mu = \mu_0$

Parámetros desconocidos	Estadístico	Distribución	H_1	Región crítica
μ	$z = \dfrac{\bar{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}$	N(0,1)	$\mu < \mu_0$ $\mu \neq \mu_0$ $\mu > \mu_0$	$z < -z_\alpha$ $\lvert z \rvert > z_{\alpha/2}$ $z > z_\alpha$
μ, σ^2	$t = \dfrac{\bar{X}-\mu_0}{\dfrac{S}{\sqrt{n}}}$	t_{n-1}	$\mu < \mu_0$ $\mu \neq \mu_0$ $\mu > \mu_0$	$t < -t_\alpha$ $\lvert t \rvert > t_{\alpha/2}$ $t > t_\alpha$

Contraste unilateral (cola de la izquierda)

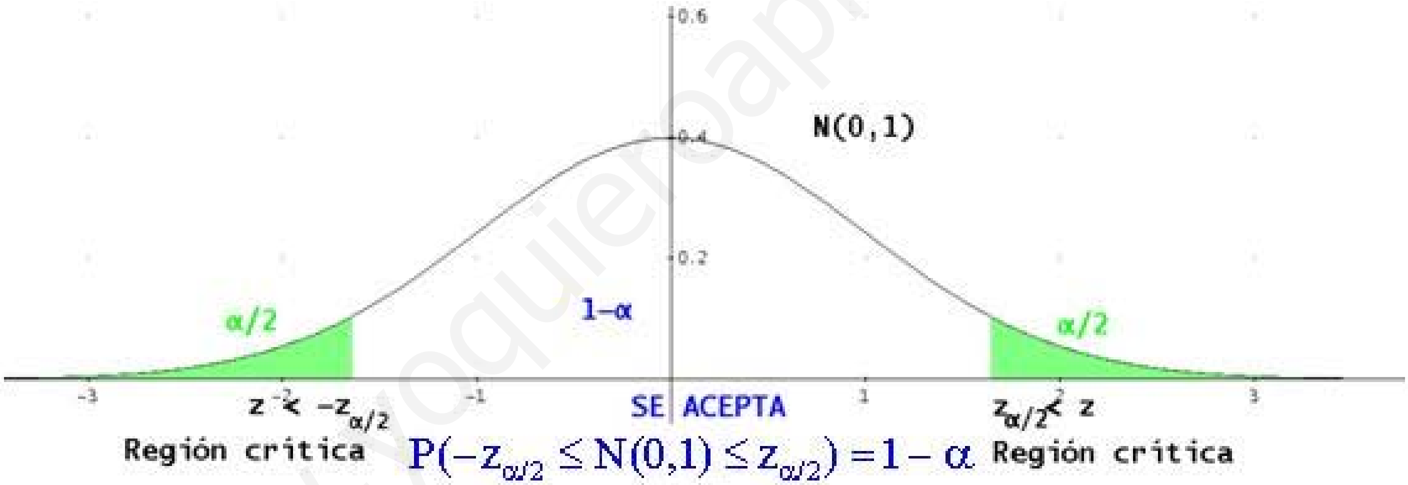
$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$



$$P(-z_\alpha \leq N(0,1)) = 1 - \alpha$$

Contraste bilateral (2 colas)

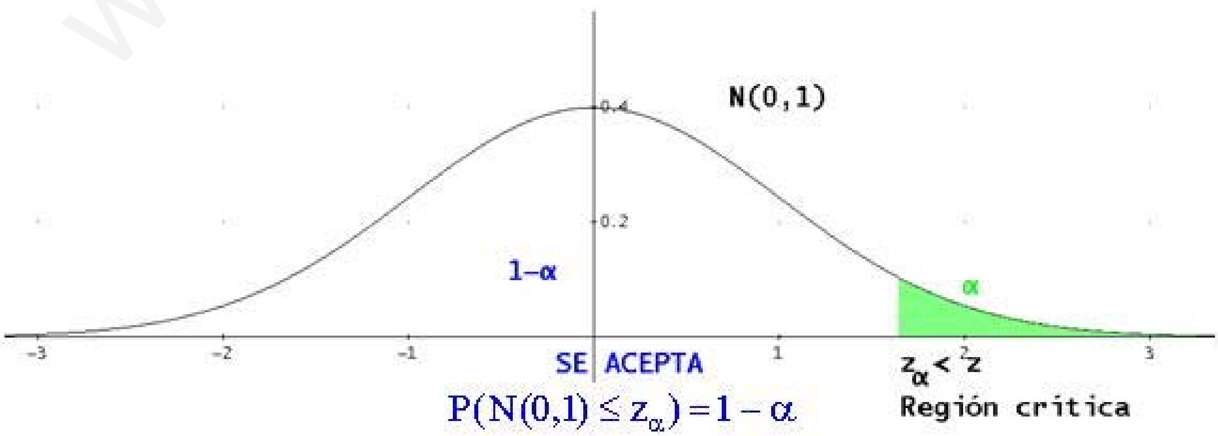
$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$



$$P(-z_{\alpha/2} \leq N(0,1) \leq z_{\alpha/2}) = 1 - \alpha$$

Contraste unilateral (cola de la derecha)

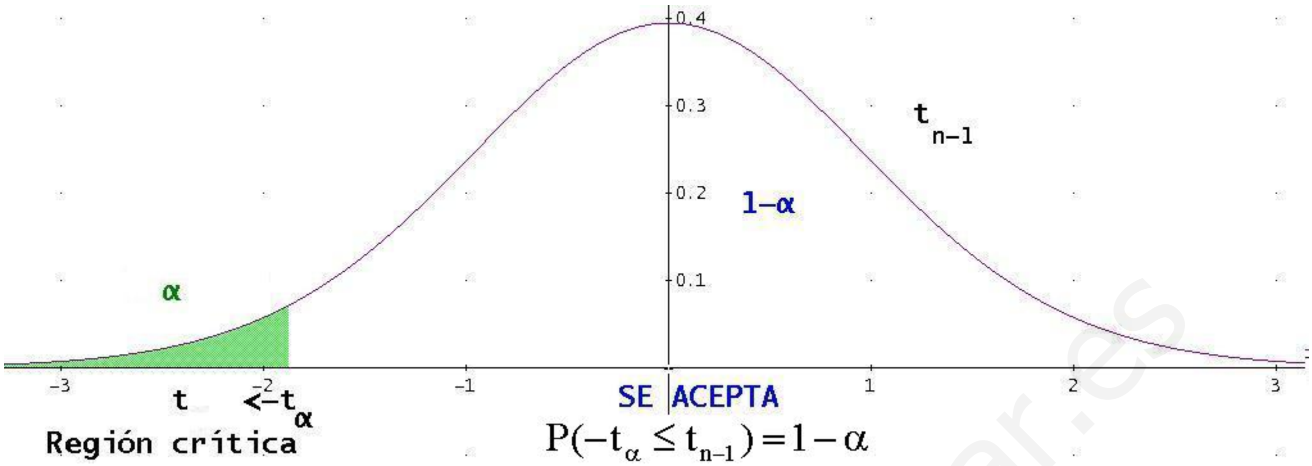
$$H_0 : \mu < \mu_0$$
$$H_1 : \mu \geq \mu_0$$



$$P(N(0,1) \leq z_\alpha) = 1 - \alpha$$

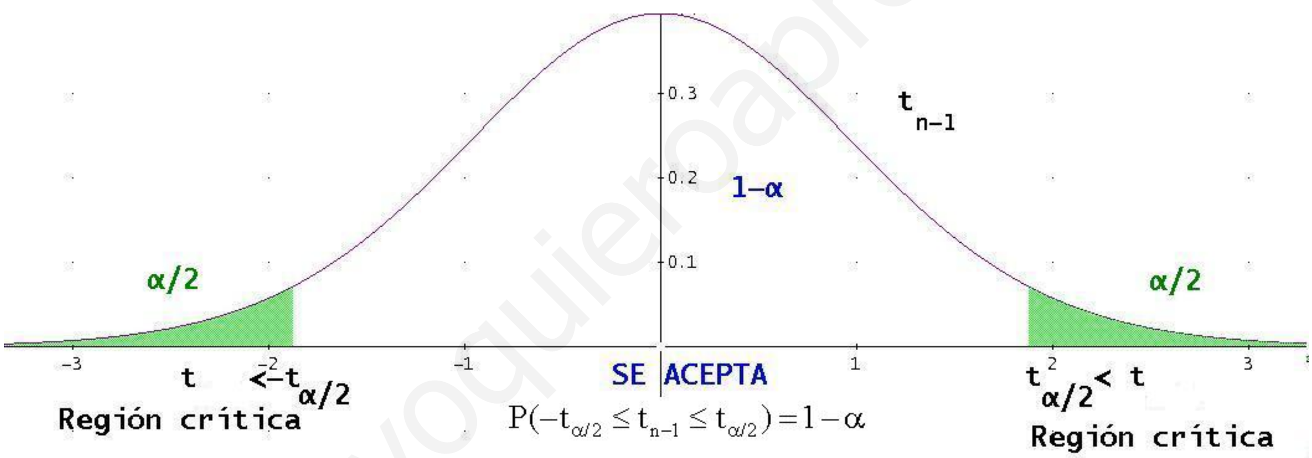
Contraste unilateral (cola de la izquierda)

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$



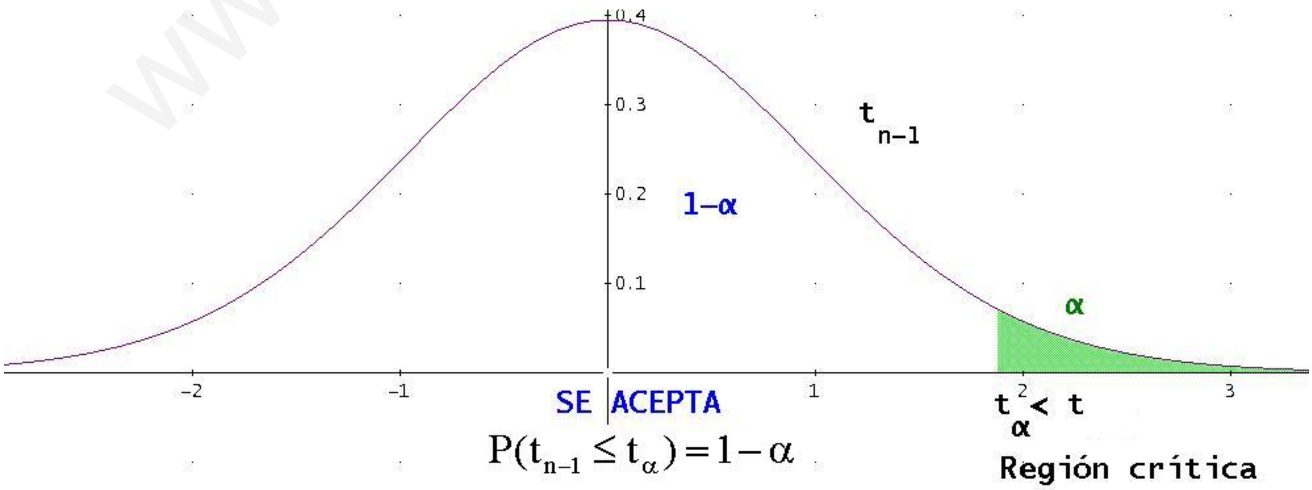
Contraste bilateral (2 colas)

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$



Contraste unilateral (cola de la derecha)

$$H_0 : \mu < \mu_0$$
$$H_1 : \mu \geq \mu_0$$



Pruebas relativas a la diferencias de medias $H_0: \mu_1 - \mu_2 = 0$

Parámetros desconocidos	Estadístico	Distribución	H_1	Región crítica
μ_1, μ_2	$\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	N(0,1)	$\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$ $\mu_1 - \mu_2 > 0$	$z < -z_\alpha$ $\lvert z \rvert > z_{\alpha/2}$ $z > z_\alpha$
$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ $\sigma_1^2 = \sigma_2^2$	$\dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}}$	$t_{n_1+n_2-2}$	$\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$ $\mu_1 - \mu_2 > 0$	$t < -t_\alpha$ $\lvert t \rvert > t_{\alpha/2}$ $t > t_\alpha$

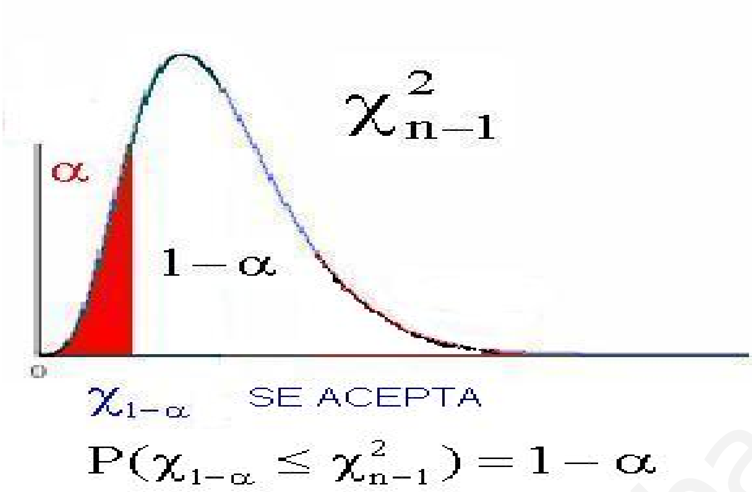
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Pruebas relativas a la varianza

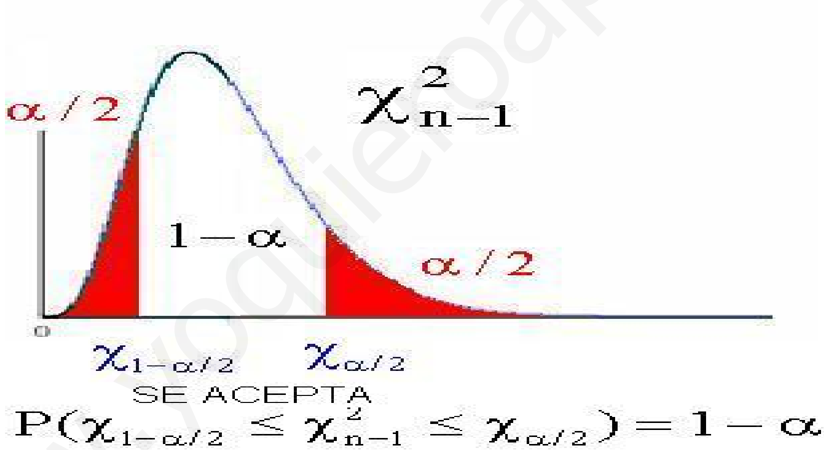
Parámetros desconocidos	Estadístico	Distribución	H_0	H_1	Región crítica
σ^2	$\chi = \frac{(n-1)S^2}{\sigma_0^2}$	χ^2_{n-1}	$\sigma = \sigma_0$	$\sigma < \sigma_0$	$\chi < \chi_{1-\alpha}$
				$\sigma \neq \sigma_0$	$\begin{cases} \chi < \chi_{1-\alpha/2} \\ \chi_{\alpha/2} < \chi \end{cases}$
				$\sigma > \sigma_0$	$\chi > \chi_{\alpha}$
σ_1^2, σ_2^2	$f = \frac{S_1^2}{S_2^2}$	F_{n_1-1,n_2-1}	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$f < F_{1-\alpha,n_1-1,n_2-1}$
				$\sigma_1^2 \neq \sigma_2^2$	$\begin{cases} f < F_{1-\alpha/2,n_1-1,n_2-1} \\ f > F_{\alpha/2,n_1-1,n_2-1} \end{cases}$
				$\sigma_1^2 > \sigma_2^2$	$f > F_{\alpha,n_1-1,n_2-1}$

Luis Sebastián

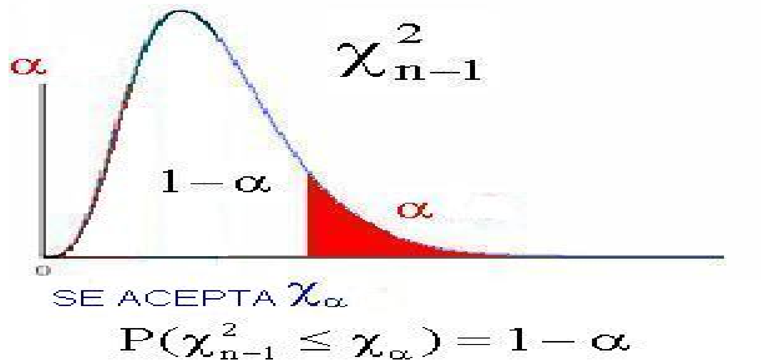
Contraste unilateral (cola de la izquierda) $H_0 : \sigma \geq \sigma_0$ $H_1 : \sigma < \sigma_0$



Contraste bilateral (2 colas) $H_0 : \sigma = \sigma_0$ $H_1 : \sigma \neq \sigma_0$



Contraste unilateral (cola de la derecha) $H_0 : \sigma < \sigma_0$ $H_1 : \sigma \geq \sigma_0$



Bondad de ajuste (test de la χ^2)

El problema que vamos a tratar es el de la conformidad de una distribución experimental y una distribución teórica; esto es, sustituir la distribución experimental (distribución de la muestra de la población), el histograma, o la distribución de frecuencias, por una distribución teórica conocida.

Sean:

n = tamaño de la muestra.
k = número de clases.
n_i = frecuencia absoluta de la clase i .
p_i = probabilidad de cada clase según la distribución teórica.
np_i = frecuencia absoluta de cada clase según la distribución teórica.
h = número de parámetros estimados a partir de la muestra.
λ = número de grados de libertad.

Las frecuencias observadas en la distribución de una muestra, se emplean para poner a prueba, la hipótesis de que la población de la cual se ha obtenido la muestra, no difiere en distribución, de la de alguna distribución conocida.

Supuesta conocida la distribución de Y. La hipótesis H_0 tiene la forma: *la población X de la cual se obtuvo la muestra tiene la misma distribución que la población Y*, formulamos la hipótesis alternativa H_1 *las poblaciones X e Y no tienen la misma distribución.*

Una medida de las discrepancias en este sentido, fue estudiada por Pearson construyendo el siguiente estadístico: $\boxed{D = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}}$, y demostró que, para $n \geq 30$ y $np_i \geq 5 \quad \Rightarrow \quad D \approx \chi^2_{k-h-1}$, esto es, la variable D sigue una distribución ji-cuadrado con $\lambda = k - h - 1$ grados de libertad.

Fijado un nivel de significación α, buscamos un valor χ_α tal que $P\left(\chi^2_\lambda \geq \chi_\alpha\right) = \alpha$

- Si $D < \chi_\alpha$ aceptamos la hipótesis H_0 de conformidad con el ajuste, siendo las diferencias $n_i - np_i$ debidas al azar.
- Si $D \geq \chi_\alpha$ rechazamos la hipótesis H_0, las diferencias $n_i - np_i$ son significativas y por tanto, las distribuciones son distintas.

Distribución binomial o de Bernouilli

Consideremos un experimento aleatorio que admite sólo dos resultados excluyentes:

Suceso A (éxito) con probabilidad P(A) = p.

Suceso $\overline{A}$ (fracaso) con probabilidad P($\overline{A}$) = 1-p = q.

A la variable aleatoria discreta ξ = “número de veces que ocurre el suceso A (éxito) en las n pruebas” se la denomina **variable binomial**.

Para calcular la distribución de probabilidad nos fijamos en el suceso favorable a A en k veces, $\overset{k)}{A \ldots . A}\overset{n-k)}{\overline{A} \ldots . \overline{A}}$ cuya probabilidad por ser sucesos independientes es el producto de las probabilidades $\overset{k)}{p \ldots . p}.\overset{n-k)}{q \ldots . q} = p^k . q^{n-k}$. Resultado que se puede repetir $\binom{n}{k}$ veces, luego P(ξ = k) $= \binom{n}{k} p^k . q^{n-k}$

La función de distribución correspondiente es: F(x) = $\sum_{k \le x} \binom{n}{k} p^k . q^{n-k}$

Una distribución binomial queda caracterizada cuando se conocen p y n y se escribe **B(n, p)**.

La esperanza matemática: $E[\xi]$ = n.p

La varianza: $V[\xi] = E[\xi^2] - (E[\xi])^2$ = n.p.q

Distribución de Poisson

Es una distribución que se presenta cuando tenemos una población n grande y la probabilidad de que ocurra un suceso determinado, tiene una probabilidad muy pequeña (ley de casos raros).
Una variable aleatoria ξ tiene **distribución de Poisson** de parámetro λ, si toma los valores enteros 0, 1, ..., n con probabilidad: $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ si $\lambda > 0$

Esperanza matemática: $E[\xi] = \lambda$

Varianza: $V[\xi] = E[\xi^2] - (E[\xi])^2 = \lambda$

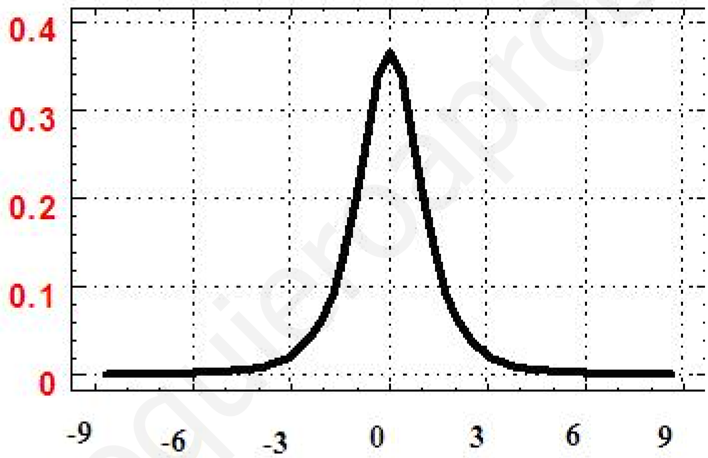
Distribución t de Student.

Sean $\xi_1,\xi_2,\ldots,\xi_n$ y ξ, n+1 variables aleatorias independientes entre sí con distribución $N(0,\sigma)$ cada una de ellas. Entonces la variable $t_n = \dfrac{\xi}{\sqrt{\dfrac{\xi_1^2+\ldots+\xi_n^2}{n}}}$ se denomina **t de Student** con n grados de libertad.

Función de densidad: $f(x) = \dfrac{1}{\sqrt{n\pi}} \dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)} \left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$

Su **media** $\mu = 0$ (n>1) y su **varianza** $\sigma^2 = \dfrac{n}{n-2}$ (n>2).

La gráfica de la función de densidad para n=3 grados de libertad:



Tiene su aplicación en la estimación de las características de una población con distribución normal mediante los llamados intervalos de confianza.

Student fue el seudónimo de William Sealy GOSSET (1.876,1.937), el estadístico y químico que descubrió la forma de la distribución t mediante una combinación de trabajo matemático y trabajo empírico con números aleatorios, una aplicación temprana de lo que ahora se llama el método de MonteCarlo.

Distribución χ^2 de Pearson

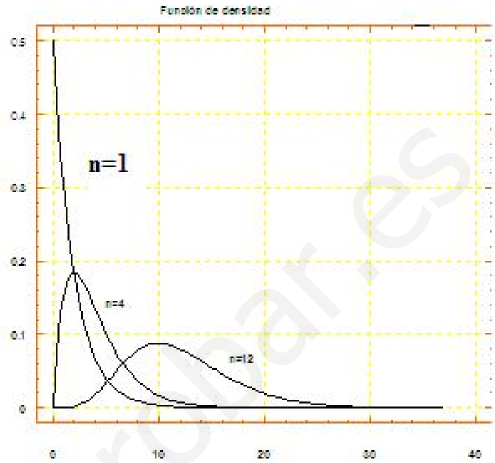
Sean $\xi_1, \xi_2, ..., \xi_n$, n variables aleatorias N(0,1) e independientes, entonces la expresión: $\chi_n^2 = \xi_1^2 + \xi_2^2 + ... + \xi_n^2$ es una variable aleatoria que recibe el nombre de **ji-dos chi-cuadrado de Pearson** (1.857,1936). El número de variables normales sumadas recibe el nombre de grados de libertad, y la representamos por χ_n^2.

La función de densidad de la variable aleatoria χ_n^2 es:

$$f(x) = \frac{x^{\left(\frac{n}{2}-1\right)}}{2^{n/2}\,\Gamma\left(n/2\right)} e^{-x/2} \quad \text{si } x \geq 0$$

La función de distribución de una variable aleatoria χ_n^2 viene dada por:

$$F(x) = P(\chi_n^2 \leq x_\alpha) = \int_0^{x_\alpha} f(x)dx$$

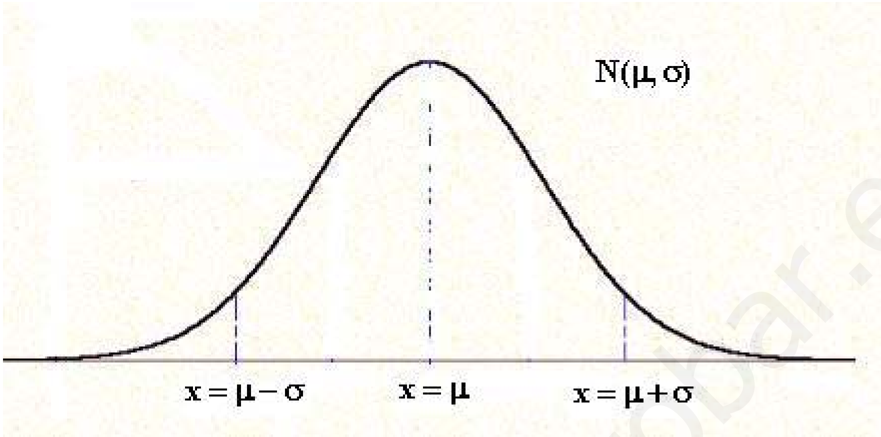


La **media**, **varianza** y **desviación típica** de una variable aleatoria ji-cuadrado con n grados de libertad son: $\mu = n$, $\sigma^2 = 2n$, $\sigma = \sqrt{2n}$ respectivamente.

Corresponde a una **distribución Gamma** de parámetros 1/2 y n/2.

Distribución Normal.

Una variable aleatoria continua X se dice que tiene una **distribución normal** o de **Laplace-Gauss** de media μ y desviación típica σ: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ es su función de densidad.



es la llamada **"campana de Gauss"**.

La función de distribución es: $F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$

La **esperanza matemática o media** es μ y la **varianza** es σ^2. Se denota $N(\mu, \sigma)$.

La Normal tipificada o estándar

$$\mathbf{Z = \frac{X-\mu}{\sigma} \equiv N(0,1)}$$

Intervalos de confianza para la media

a) Población normal con varianza conocida.

Sabemos que $\bar{\xi} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, luego $\eta = \frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}} \equiv N(0,1)$. Queremos calcular un intervalo I_α de forma que la $P(\mu \in I_\alpha) = 1 - \alpha$.

A $1-\alpha$ se le llama **nivel de confianza**

A α se le llama **nivel de significación** y es la probabilidad de que el parámetro no esté en el intervalo.

Buscaremos en la N(0,1) un valor λ de forma que $P(-\lambda \leq \eta \leq \lambda) = 1 - \alpha$ como λ, σ y n son conocidos, tenemos el intervalo $I_\alpha = \left(\bar{\xi} - \lambda\frac{\sigma}{\sqrt{n}}, \bar{\xi} + \lambda\frac{\sigma}{\sqrt{n}}\right)$.

El **intervalo de confianza** sería: $\bar{x} \pm \lambda\frac{\sigma}{\sqrt{n}}$

b) Población cualquiera de varianza finita y muestras grandes.

Sabemos que $\bar{\xi} \equiv N(\mu, \frac{\sigma}{\sqrt{n}})$. Razonando igual que antes, si la varianza es conocida el intervalo será $P\left(\bar{\xi} - \lambda\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{\xi} + \lambda\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ para n>30.

Si la varianza es desconocida la estimamos por la cuasivarianza, y queda:
$P\left(\bar{\xi} - \lambda\frac{S}{\sqrt{n}} \leq \mu \leq \bar{\xi} + \lambda\frac{S}{\sqrt{n}}\right) = 1 - \alpha$ para n>100 y el intervalo es $\bar{x} \pm \lambda\frac{S}{\sqrt{n}}$.

c) Población normal con varianza desconocida.

Buscaremos en un valor t_α tal que $P(-t_\alpha < t_{n-1} < t_\alpha) = 1 - \alpha$ y el correspondiente intervalo de confianza será: $P\left(\bar{\xi} - t_\alpha\frac{S}{\sqrt{n}} < \mu < \bar{\xi} + t_\alpha\frac{S}{\sqrt{n}}\right) = 1 - \alpha$

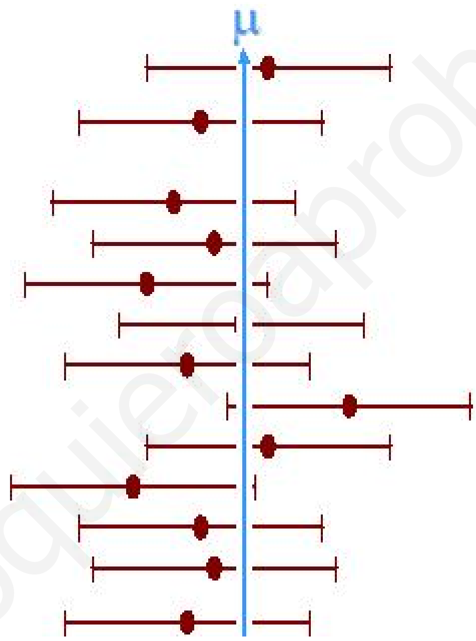
Para una muestra concreta: $\bar{x} \pm t_\alpha\frac{S}{\sqrt{n}}$ y si queremos determinar el tamaño muestral n, resulta $t_\alpha\frac{S}{\sqrt{n}} = \varepsilon$ de donde $n = \left(\frac{t_\alpha . S}{\varepsilon}\right)^2$.

Nivel o grados de confianza

Fijado un **nivel de confianza 1-α** buscaremos dos estadísticos $T_1=T(x_1,x_2,\ldots,x_n)$ y $T_2=T(x_1,x_2,\ldots,x_n)$ tales que el **intervalo de confianza** para θ con ese nivel de confianza sea:

$I_\alpha = \left(\hat{\theta}-h_1, \hat{\theta}+h_2\right) = (T_1, T_2)$ de forma que $P(\theta \in I_\alpha) = P(T_1 \leq \theta \leq T_2) = 1-\alpha$, siendo $0<\alpha<1$

Los **grados de confianza $1-\alpha$** (frecuentemente expresada en tanto por ciento equivalente) es la frecuencia relativa de veces que los intervalos de confianza contienen el parámetro de la población, entendiendo que el proceso de estimación se repite un número grande de veces.



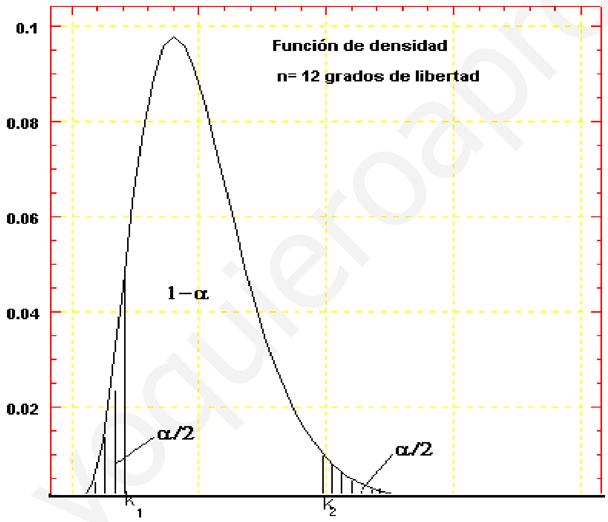
Intervalos de confianza para la varianza

Se sabe que $\frac{(n-1).S^2}{\sigma^2} \equiv \chi^2_{n-1}$ si la población de partida es $N(\mu,\sigma)$. Por tanto, para tomar el intervalo de confianza de nivel de significación α, buscamos los valores k_1 y k_2, tal que: $P\left(k_1 < \frac{(n-1).S^2}{\sigma^2} < k_2\right) = 1-\alpha$ y el correspondiente intervalo de confianza será:

$$P\left(\frac{(n-1).S^2}{k_2} < \sigma^2 < \frac{(n-1).S^2}{k_1}\right) = 1-\alpha$$

Se nos plantea el problema de que la distribución χ^2_{n-1} no es simétrica (como ocurría con la Normal y la t de Student) por lo que no es posible determinar con exactitud los valores k_1 y k_2 para que el intervalo esté centrado en S^2.

Una solución aproximada y generalmente buena es determinar k_1 y k_2 con las condiciones: $P\left(\chi^2_{n-1} < k_1\right) = \frac{\alpha}{2}$ y $P\left(\chi^2_{n-1} > k_2\right) = \frac{\alpha}{2}$



Estimador de la varianza de la población

La **varianza muestral** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2$ es un estimador insesgado de la varianza σ^2, ya que: $E\left[S^2\right] = \sigma^2$

Debe observarse que no hemos hecho ninguna hipótesis de cual sea la distribución de probabilidad de la variable ξ.

Por tanto, la **desviación típica muestral** será: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2}$

Esperanza matemática

La esperanza matemática es el valor medio teórico que resulta de sustituir las f_i (frecuencias relativas) por las P_i (probabilidades), y que no es sino una media aritmética ponderada. Se acostumbra a definirlo como **esperanza matemática** de ganancias, es decir, la ganancia que teóricamente esperaba obtener un jugador frente a unas determinadas reglas de juego.

Sea ξ una variable aleatoria, se define el operador $E[\xi]$ como:

$E[\xi] = \sum_{i=1}^{n} x_i P(X_i)$ para una variable discreta y finita.

$E[\xi] = \sum_{i=1}^{\infty} x_i P(X_i)$ para una variable discreta y no finita siempre que la serie sea convergente.

$E[\xi] = \int_{-\infty}^{\infty} t.f(t).dt$ cuando la variable ξ es continua con función de densidad f(x) y siempre que la integral sea absolutamente convergente.

Estimador Insesgado o centrado

Un estimador $\hat{\theta}$ se llama **insesgado** cuando $E\left[\hat{\theta}\right] = \theta$.

La **media muestral** $\overline{\xi} = \dfrac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$ es un estimador insesgado, ya que $E\left[\overline{\xi}\right] = \mu$, consistente, eficiente, puesto que $V\left[\overline{\xi}\right] = \dfrac{\sigma^2}{n}$, y suficiente.

Muestreo aleatorio simple

Se trata de extraer una muestra de una población.

Se caracteriza porque todos los elementos de la población tienen la misma probabilidad de ser elegidos. El procedimiento práctico de escoger la muestra, puede ser numerar los elementos de la población, apuntar los números en tarjetas, y sacarlas al azar. Si la muestra y la población son grandes, en vez de tarjetas se utilizan tablas de números aleatorios.

Este muestreo puede ser:

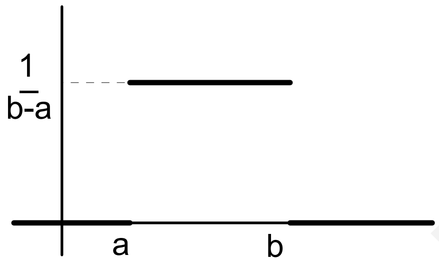
- ***Con reemplazamiento***: se elige un elemento de la población, se estudia y se reintegra a la población. Así sucesivamente, la probabilidad de obtener cualquier elemento se mantiene constante, interesa para que los elementos de la muestra sean independientes.
- ***Sin reemplazamiento***: los elementos elegidos en la muestra no se restituyen en la población. La probabilidad de obtener un elemento va aumentado al disminuir los elementos posibles.

Distribución Uniforme o Rectangular.

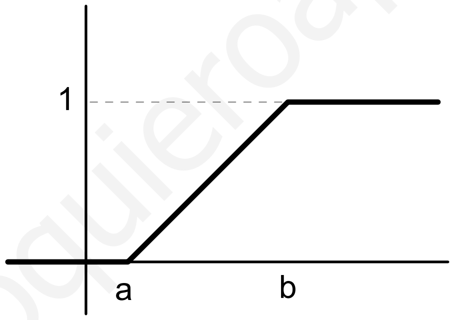
Se dice que una variable continua sigue una **distribución uniforme** en el intervalo $[a, b]$ cuando

su función de densidad es: $$f(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{1}{b\text{-}a} & \text{si } a \le x \le b \\ 0 & \text{si } b < x \end{cases}$$

Gráficamente:



La función de distribución, será: $$F(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x - a}{b - a} & \text{si } a \le x \le b \\ 1 & \text{si } b < x \end{cases}$$ cuya gráfica es:



Media: $E[\xi] = \int_{-\infty}^{\infty} x.f(x).dx = \dfrac{a+b}{2}$

Varianza: $V[\xi] = E[\xi^2] - (E[\xi])^2 = \dfrac{(a-b)^2}{12}$

Esta distribución depende de los parámetros a y b y se denota $U[a, b]$.

Para el caso **discreto** se caracteriza por tener todos su valores la misma probabilidad