

Ejercicios Resueltos de Estadística:
Tema 5: Inferencia: estimación y contrastes

1. Si $\bar{X} \sim N(40,10)$, calcular $\Pr(39 \leq \bar{X} \leq 41)$ para $n=10$. ¿En qué intervalo se obtendrán el 95% de los resultados?

SOLUCIÓN:

$$\Pr(39 \leq \bar{X} \leq 41) = \Pr\left(\frac{39-40}{\sqrt{10}} \leq \frac{\bar{X}-40}{\sqrt{10}} \leq \frac{41-40}{\sqrt{10}}\right) = \Pr(-0.31623 \leq \bar{X} \leq 0.31623)$$

$$Z = \frac{\bar{X}-40}{\sqrt{10}} \rightarrow N(0,1); \Pr(39 \leq \bar{X} \leq 41) = \Pr(Z \leq 0.31623) - \Pr(Z \leq -0.31623) =$$

$$= 2 \Pr(Z \leq 0.31623)$$

$$\text{Y por tanto, } \Pr(39 \leq \bar{X} \leq 41) = 2 * 0.6241 - 1 = .02482$$

$$\Pr(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon) = 0.95$$

$$\Pr(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon) = 2 * \Pr\left(Z \leq \frac{\varepsilon}{\sqrt{10}}\right) - 1$$

$$\Pr\left(Z \leq \frac{\varepsilon}{\sqrt{10}}\right) = \frac{1+0.95}{2} = 0.975 \rightarrow Z_{0.975} \rightarrow \varepsilon = 1.96\sqrt{10} = 6.1981$$

Por tanto, el intervalo es: (33.802, 46.198)

2. Si el contenido en gr. de un determinado medicamento X sigue una distribución $N(7.5, 0.3)$, calcular la probabilidad de que para una muestra de tamaño $n=5$, se obtenga medio menor que 7, $\Pr(\bar{X} \leq 7)$.

SOLUCIÓN:

A partir de una muestra de tamaño $n=5$ de una población normal $N(\mu=7.5, \sigma=0.3)$, tenemos que:

$$\Pr(\bar{X} \leq 7) = \Pr\left(\frac{\bar{X} - 7.5}{\frac{0.3}{\sqrt{5}}} \leq \frac{7 - 7.5}{\frac{0.3}{\sqrt{5}}}\right) = \Pr(Z \leq -3.7269)$$

Donde Z tiene una distribución normal estándar, y por tanto, $\Pr(\bar{X} \leq 7) = 0.0001$

3. Si la altura de un grupo de población sigue una distribución normal N(176,12), calcular la $\Pr(S \leq 10)$ para una muestra de tamaño 8.

SOLUCIÓN:

Considerando una muestra aleatoria de tamaño n de una población normal N(μ, σ), por el teorema de Fisher tenemos que:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

En particular, para una muestra de tamaño n=8 de una población normal N(176,12), el estadístico $\frac{7}{144}S^2$ sigue una distribución χ_7^2 , y por tanto

$$\Pr(S \leq 10) = \Pr(S^2 \leq 100) = \Pr\left(\frac{7}{144}S^2 \leq \frac{700}{144}\right) = \Pr(T \leq 4.8611)$$

Donde la variable T sigue una distribución χ_7^2 , es decir,

$$\Pr(S \leq 10) = 0.3232$$

4. Un ascensor limita el peso de sus cuatro ocupantes a 300Kg. Si el peso de un individuo sigue una distribución N(71,7), calcular la probabilidad de que el peso de 4 individuos supere los 300Kg.

SOLUCIÓN:

Teniendo en cuenta que el peso de cada individuo tiene una distribución normal $N(\mu = 71, \sigma = 7)$, si seleccionamos una muestra aleatoria de 4 individuos, tenemos que:

$$\Pr\left(\sum_{i=1}^4 X_i > 300\right) = \Pr\left(\frac{\sum_{i=1}^4 X_i}{4} > \frac{300}{4}\right) = \Pr(\bar{X} > 75) = \Pr\left(\frac{\bar{X} - 71}{\frac{7}{\sqrt{4}}} > \frac{75 - 71}{\frac{7}{\sqrt{4}}}\right) =$$

$$= \Pr(Z > 1.1429) = 1 - \Pr(Z \leq 1.1429)$$

donde Z tiene una distribución normal estándar, y por tanto,

$$\Pr\left(\sum_{i=1}^4 X_i > 300\right) = 1 - 0.8735 = 0.1265$$

5. Calcular la probabilidad de que la media μ se encuentre entre $\bar{X} \pm 3S$ para poblaciones normales y $n = 5$.

SOLUCIÓN:

A partir del teorema de Fisher, en el muestreo sobre poblaciones normales, tenemos que los estadísticos \bar{X} y S^2 son independientes, siendo la distribución del estadístico $T = \sqrt{n} * \frac{\bar{X} - \mu}{S}$ una t_{n-1} (t de Student de $n - 1$ grados de libertad). En particular, si consideramos una muestra aleatoria de tamaño $n = 5$, la probabilidad de que la media esté entre $\bar{X} \pm 3S$ viene dada por:

$$\Pr(\bar{X} - 3S < \mu < \bar{X} + 3S) = \Pr\left(-3 < \frac{\mu - \bar{X}}{S} < 3\right) = \Pr\left(-3\sqrt{5} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{5}}} < 3\sqrt{5}\right) = \Pr(-3\sqrt{5} < T < 3\sqrt{5})$$

donde T tiene una distribución t_4 , y por tanto:

$$\Pr(\bar{X} - 3S < \mu < \bar{X} + 3S) = \Pr(-3\sqrt{5} < T < 3\sqrt{5}) = 2 \Pr(T < 6.7082) - 1 = 2 * 0.9987 - 1 = 0.9974$$

6. Calcular un intervalo de confianza al nivel $\alpha = 0.05$ para la probabilidad de p de que un recién nacido sea niño si en una muestra de tamaño 123 se han obtenido 67 niños.

SOLUCIÓN:

Teniendo en cuenta que la proporción de varones recién nacidos puede modelizarse por una variable Bernoulli de parámetro p (probabilidad de que un recién nacido sea varón), el intervalo de confianza al nivel $\alpha = 0.05$ viene dado por:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Donde $n = 123$. $\hat{p} = \frac{67}{123}$ y $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, es decir,

$$(0.544715 - 0.0880096, 0.544715 + 0.0880096)$$

y por tanto, el intervalo $(0.0456706, 0.632725)$ contendrá a la proporción de varones nacidos con una probabilidad del 95%.

7. Calcular un intervalo de confianza al nivel $\alpha = 0.001$ para el peso exacto mediante los resultados obtenidos con 10 básculas:

7.20, 7.01, 7.36, 6.91, 7.22, 7.03, 7.11, 7.12, 7.03, 7.05

SOLUCIÓN:

Suponiendo que las medidas del peso de las básculas sigue una distribución normal $N(\mu, \sigma^2)$ con media el peso exacto, estamos interesados en encontrar un intervalo de confianza que contenga a la media de esta distribución, que a un nivel $\alpha = 0.001$ y desviación típica desconocida, esta determinado por:

$$\left(\bar{X} - t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

donde $n = 10$, $\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 7.1040$, $S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = 0.1286$, y utilizando la tabla de la distribución t de Student $t_{n-1; 1-\frac{\alpha}{2}} = t_{9; 0.9995} = 4.78091$. Por tanto, el intervalo de confianza al nivel 0.001 es: $(6.9096, 7.2984)$

Y representa que la media del peso estará en dicho intervalo con una probabilidad de acierto del 99.9%.

8. Calcular un intervalo de confianza al nivel $\alpha = 0.05$ para σ^2 mediante las desviaciones que se producen en un proceso de fabricación cuya distribución es $N(0, \sigma)$ a partir de la muestra

1.2, -2.2, -3.1, -0.2, 0.5, 0.6, -2.1, 2.2, 1.3

SOLUCIÓN:

Sabiendo que el proceso de fabricación sigue una distribución normal de media conocida $\mu = 0$, un intervalo de confianza para la varianza σ^2 al nivel $\alpha = 0.05$ es el siguiente:

$$\left(n \frac{T}{\chi^2_{n; 1-\frac{\alpha}{2}}}, n \frac{T}{\chi^2_{n; \frac{\alpha}{2}}} \right)$$

donde $n = 9$, $T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} = 3.05333$, y utilizando la tabla de la distribución χ^2 , se

tiene $\chi^2_{n; 1-\frac{\alpha}{2}} = \chi^2_{9; 0.975} = 2.7004$, es decir, (1.44458, 10.1763)

Es el intervalo que contendrá con un 95% de acierto las desviaciones que se producen en el proceso de fabricación.

9. Calcular qué tamaño muestral debemos tomar para obtener μ con una precisión de 0.001 a partir de una muestra de una población $N(\mu, 3)$.

SOLUCIÓN:

Teniendo en cuenta que el intervalo de confianza que contiene a la media μ de una población normal con varianza conocida es de la forma

$$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

es decir, el error que cometemos al estimar μ mediante un intervalo de confianza al nivel $\alpha = 0.05$, es

$$error = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{3}{\sqrt{n}}$$

Por tanto, si en esta estimación deseamos obtener la media con una precisión de 0.001, tenemos que calcular n tal que el error que se cometa esté acotado por esta cantidad, $error \leq 0.001$, es decir:

$$1.96 \frac{3}{\sqrt{n}} \leq 0.001 \Leftrightarrow \frac{5.88}{0.001} \leq \sqrt{n} \rightarrow n = 5880^2 = 34574400$$

10. Para estudiar la efectividad de un medicamento contra la diabetes se mide la cantidad de glucemia en sangre antes y después de la administración de dicho medicamento, obteniéndose los resultados siguientes:

Antes	7.2	7.3	6.5	4.2	3.1	5.3	5.6
Después	5.2	5.4	5.3	4.7	4.1	5.4	4.9

Estimar la reducción producida por el medicamento.

SOLUCIÓN:

Suponiendo que las cantidades de glucemia siguen una distribución normal, las muestras obtenidas mediante las medicaciones anteriores y posteriores a la administración del medicamento son apareadas, es decir, sobre cada individuo de la población se obtiene una observación de la variable y posteriormente se repite la observación una vez que el individuo ha tomado el medicamento siendo esta última observación dependiente de la primera.

Para estudiar el efecto de dicho medicamento, resulta de gran interés la variable diferencia entre ambas mediciones, puesto que nos permitirá estimar la reducción o incremento de la glucemia provocada por este medicamento.

En este caso, la diferencia de ambas observaciones sigue una distribución normal de media la diferencia de ambas y desviación típica desconocida, $N(\mu_A - \mu_D, \sigma^2)$, por lo que un intervalo de confianza al nivel $\alpha = 0.05$, viene dado por:

$$\left((\bar{X}_A - \bar{X}_D) - t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, (\bar{X}_A - \bar{X}_D) + t_{n-1; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right)$$

donde $n=7$, $\bar{X}_A - \bar{X}_D = \frac{\sum_{i=1}^n (\bar{X}_{Ai} - \bar{X}_{Di})}{n} = 0.60$ y

$$S = \sqrt{\frac{\sum ((\bar{X}_{Ai} - \bar{X}_{Di}) - (\bar{X}_A - \bar{X}_D))^2}{n-1}} = 1.17473$$

y utilizando la tabla de la distribución t de Student, $t_{n-1; 1-\frac{\alpha}{2}} = t_{6, 0.975} = 2.4469$, es decir, la reducción

de glucemia por el medicamento estará contenida a un nivel 0.05 en el intervalo $(-0.486448, 1.68645)$, siendo la reducción estimada 0.6.

11. Se ha hecho un estudio sobre la proporción de enfermos de cáncer de pulmón detectados en hospital que fuman, obteniéndose que de 123 enfermos 41 de ellos eran fumadores. Obtener un intervalo de confianza para dicha proporción. Estudiar si dicha proporción puede considerarse igual a la proporción de fumadores en la población si ésta es de un 29%.

SOLUCIÓN:

Como la proporción de fumadores en la población de los enfermos de cáncer de pulmón detectados en un hospital puede modelizarse por una variable Bernoulli de parámetro p, entonces:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

es un intervalo de confianza al nivel $\alpha = 0.05$, y a partir de las observaciones realizadas, $n = 123$, tenemos que $\hat{p} = \frac{41}{123}$ y $z_{1-\frac{\alpha}{2}} = 1.96$, es decir: $(0.250023, 0.416643)$

Por otra parte, si consideramos la proporción de fumadores en la población global, $p = 0.29$, observamos que esta proporción se encuentra dentro del intervalo de confianza obtenido para un nivel 0.05, es decir, como la proporción de fumadores en la población pertenece al intervalo que contiene a la proporción de fumadores entre los enfermos de cáncer con una probabilidad de

acierto del 955, podemos considerar que esta proporción de fumadores en los enfermos de cáncer se corresponde con la de los fumadores en la población global.

12. Sospechamos que nuestro cromatógrafo está estropeado, y queremos determinar si los resultados que nos proporciona son lo suficientemente precisos. Para ello, realizamos una serie de 8 mediciones del contenido de una solución de referencia que, sabemos, contiene 90% de un determinado compuesto. Los resultados que obtenemos son:

93.3, 86.8, 90.4, 90.1, 94.9, 91.6, 92.3, 96.5

Construir un intervalo de confianza al nivel de 95% para la varianza poblacional. ¿Que conclusiones podemos realizar?

SOLUCIÓN:

Sea la variable aleatoria X que representa el valor medio del contenido, si modelizamos el error incontrolable de medición por una variable aleatoria ε normal de media 0, y de varianza σ^2 desconocida, tenemos que $X \sim (\text{valor real}, \sigma^2)$, donde conocemos el valor real del contenido de la solución que es igual a 90. Por lo tanto, estamos en el caso en que la medida poblacional es conocida e igual a 90, y tenemos que construir un intervalo de confianza para la varianza poblacional, por lo que usamos el estadístico:

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2}$$

Cuya distribución en el muestreo es una χ_n^2 con n grados de libertad. Así, se deduce que, un intervalo de confianza del $100(1 - \alpha^2)\%$, para la varianza de una distribución normal con media conocida, está dado por:

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n-1; 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n-1; \frac{\alpha}{2}}^2}$$

En nuestro caso, para un nivel de confianza del 95%, observamos en la tabla de la distribución χ^2 que $\chi_{n-1; 1-\frac{\alpha}{2}}^2 = \chi_{8; 0.975}^2 \approx 17.53$ y $\chi_{n-1; \frac{\alpha}{2}}^2 = \chi_{8; 0.025}^2 \approx 2.18$, y calculamos el numerador anterior:

$$\sum_{i=1}^n (x_i - \mu)^2 = (93.3 - 90)^2 + \dots + (96.5 - 90)^2 = 95.41$$

por lo que obtenemos que el intervalo de confianza pedido es:

$$5.44 \leq \sigma^2 \leq 43.77$$

Esto representa que el intervalo para la dispersión es $2.33 \leq \sigma \leq 6.62$ (es una dispersión grande), de donde deducimos que la precisión del cromatógrafo es insuficiente.

13. En el departamento de control de calidad de una empresa, se quiere determinar si ha habido un descenso significativo de la calidad de su producto entre las producciones de dos semanas consecutivas a consecuencia de un incidente ocurrido durante el fin de semana. Deciden tomar una muestra de la producción de cada semana, si la calidad de cada artículo se mide en una escala de 100, obtienen los resultados siguientes:

Semana 1	93	86	90	90	94	91	92	96
Semana 2	93	87	97	90	88	87	84	93

Suponiendo que las varianzas de la puntuación en las dos producciones son iguales, construye un intervalo de confianza para la diferencia de medias al nivel de 95%. Interpreta los resultados obtenidos.

SOLUCIÓN:

En primer lugar, observamos que se disponen de dos poblaciones, la primera corresponde a la producción de la primera semana mientras que la segunda corresponde a la de la segunda semana. En este sentido, introducimos las dos variables X_1 que mide la puntuación de calidad de un artículo de la primera semana, y X_2 para la segunda.

Además, en el caso en el que las varianzas en las dos poblaciones son desconocidas pero iguales, X_1 y X_2 se asumen normales e independientes, utilizamos el estadístico:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde $S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, el cuál sigue una distribución $t_{n_1 + n_2 - 2}$ de Student de

$n_1 + n_2 - 2$ grados de libertad. Así, un intervalo de confianza al $100(1 - \alpha)\%$ para la diferencia entre medias de dos distribuciones normales, con varianzas desconocidas pero iguales es:

$$\bar{x}_1 - \bar{x}_2 - cS_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + cS_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde $c = t_{n_1+n_2-2; 1-\frac{\alpha}{2}}$. En particular, para los datos de este problema, se tiene

$\bar{x}_1 = 91.5, S_1^2 = 90.1, n_1 = 8, n_2 = 8, \bar{x}_2 = 89.9, S_2^2 = 17.8$, por lo que:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{7 * 9.1 + 7 * 17.8}{14} = 13.4$$

Y $c = t_{n_1+n_2-2; 1-\frac{\alpha}{2}} = t_{14; 0.975} \approx 2.145$, obteniendo el intervalo de confianza:

$$1.6 - 2.145\sqrt{13.4} \sqrt{\frac{1}{8} + \frac{1}{8}} \leq \mu_1 - \mu_2 \leq 1.6 + 2.145\sqrt{13.4} \sqrt{\frac{1}{8} + \frac{1}{8}}$$

es decir, el intervalo de confianza al nivel del 95% para la diferencia de medias es:

$$-2.31 \leq \mu_1 - \mu_2 \leq 5.56$$

Por último, podemos concluir que, con los datos de la muestra, es posible que la diferencia de las medias poblacionales sea igual o muy próximo a cero, en consecuencia no podemos afirmar que ha habido un descenso significativo de la calidad entre las dos semanas.

14. Eres el encargado de un departamento de producción en una fábrica y recibes un lote de 2000 piezas necesarias para la fabricación de un artículo. Tienes la responsabilidad de aceptar o rechazar el lote, si estimas que la calidad de éste no es suficiente. El fabricante te asegura que, en este lote, no hay más de 100 piezas defectuosas, pero decides tomar una muestra para estimar la proporción de las mismas.

a) ¿Cuántas piezas decides examinar para que, con un nivel de confianza del 95%, el error que cometes en la estimación de la proporción poblacional de defectuosas no sea mayor que 0.05?

b) Si decides tomar una muestra de 100 artículos escogidos al azar en el lote y realizas el recuento de piezas defectuosas en esta muestra, encontrado 4 artículos defectuosos. Construye para la proporción de defectuosos en el lote, un intervalo de confianza al nivel de 95% de confianza. ¿Se debe rechazar el lote?

SOLUCIÓN:

Observar que el tamaño de la población (el lote de piezas) es de 2000, por lo que podemos considerar que dicho tamaño es lo suficientemente grande como para que la podamos tratar como una población infinita.

a) Para determinar el tamaño muestral necesario para estimar la proporción poblacional si queremos que el error cometido sea menor que 0.05 con un nivel de confianza de $100(1 - \alpha)\%$, viene dado por:

$$n \geq \left(z_{1-\frac{\alpha}{2}} \right)^2 \frac{p(1-p)}{\text{error}^2}$$

Donde no conocemos el valor de la proporción exacta p , puesto esto es precisamente lo que queremos encontrar. Por lo tanto, debemos utilizar una aproximación de p , usando una muestra piloto o eligiendo el p más pesimista posible, es decir, el p que conlleva la elección del tamaño muestral más grande, $p=0.5$. así tenemos dos posibilidades:

1) el valor que, según el vendedor, es la proporción poblacional, $p = \frac{100}{2000} = 0.05$. En este caso, el error máximo admisible es $\text{error} = 0.05$ y el nivel de confianza 95%, $\alpha = 0.05$ y utilizando las tablas de la normal estándar $z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$. Por lo que se llega:

$$n \geq 1.96^2 * \frac{0.5 * 0.95}{0.0025} \approx 72.99$$

es decir, el valor de n que convendría usar es 73.

2) el valor más pesimista, $p=0.5$, y en este caso, de forma similar a la anterior, se tiene:

$$n \geq 1.96^2 * \frac{0.5 * 0.5}{0.0025} \approx 384.16$$

Por último, observar que tendremos que escoger entre las dos opciones anteriores según el tiempo, y/o el dinero, que nos llevará examinar cada artículo de la muestra.

b) en el caso de tomar una muestra de tamaño 100, en la cuál hemos encontrado 4 piezas defectuosas, tenemos que determinar un intervalo de confianza para la proporción poblacional.

Para ello, si consideramos la variable aleatoria \hat{P} que designa el estadístico proporción muestral, y p la proporción en la población, tenemos que la distribución el siguiente estadístico:

$$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Es aproximadamente una distribución normal estándar. A partir de esto, deducimos que un intervalo de confianza de nivel $100(1-\alpha)\%$ para la proporción p de la población, es de la forma:

$$\left(\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \hat{p} \leq \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Y aplicándolo a los datos del problema, $p = \frac{4}{100} = 0.04$, $\alpha = 0.05$ y $z_{1-\frac{\alpha}{2}} = z_{0.975} \approx 1.96$, por

lo que el intervalo de confianza al 95% para la proporción es:

$$0.04 - 1.96 \sqrt{\frac{0.04 * 0.96}{100}} \leq p \leq 0.04 + 1.96 \sqrt{\frac{0.04 * 0.96}{100}}$$

es decir; $0.002 \leq p \leq 0.078$

y teniendo en cuenta que el fabricante nos aseguraba que la proporción de defectuosos en el lote era de 0.05, concluimos que no está en contradicción con los resultados que hemos encontrado en nuestra muestra, por lo que decidimos confiar en el fabricante y aceptamos el lote.

15. Los tiempos de reacción, en mili segundos, de 17 sujetos frente a una matriz de 15 estímulos fueron los siguientes:

448, 460, 514, 488, 592, 490, 507, 513, 492, 534,523, 452, 464, 562, 584, 507, 461

Suponiendo que el tiempo de reacción se distribuye Normalmente, determine un intervalo de confianza para la media a un nivel de confianza del 95%.

SOLUCIÓN:

Mediante los cálculos básicos obtenemos que la media muestral vale 505,35 y la desviación típica 42,54.

Buscando en las tablas de la t de Student con 16 grados de libertad, obtenemos que el valor que deja por debajo una probabilidad de 0,975 es 2,12.

Sustituyendo estos valores en la expresión del intervalo de confianza de la media tenemos:

$$(505,35 - 2,12 * 42,54 / 4 \text{ ,, } 505,35 + 2,12 \cdot 42,54 / 4)$$

Operando (482,80 ,, 527,90)

16. Se considera una población representada por una variante ε , de suerte que la media poblacional es igual a 25, y la varianza poblacional es igual a 240. Supuesto extraídas muestras de tamaño 100, muestreo aleatorio simple, determinar la probabilidad de que el estadístico media muestral, A_x , este comprendido entre los valores 23,55 y 28,1.

SOLUCIÓN:

El estadístico media muestral :

$$A_x = \frac{\sum_{i=1}^n X_i}{n},$$

Habida cuenta el resultado obtenido en el problema 110, es tal que:

$$E[A_x] = 25$$

$$v[A_x] = \frac{240}{100} = 2,4$$

La probabilidad a obtener puede ser expresada en la forma:

$$P(23,55 \leq A_x \leq 28,1)$$

y sera conocida si determinamos la distribución de probabilidad de la variante A_x . Ahora bien, como quiera que A_x viene definida como suma de variantes X_i , independientes entre si, todas con igual distribución e igual a la distribución de probabilidad de la población, puesto que el muestro que se considera es el aleatorio simple, y habida cuenta que el tamaño muestral es suficientemente grande, $n=100$, procede efectuar la aproximación de la distribución de A_x mediante la distribución normal. De esta forma, pues, A_x sera uan variante con distribución :

$$N(25, \sqrt{2,4})$$

Y por ello:

$$P(23,55 \leq A_x \leq 28,1) = F(28,1) - F(23,55)$$

Al ser:

$$F(28,1) = P(A_x \leq 28,1) = P(\sqrt{2,4} * \varepsilon + 25 \leq 28,1) = P(\varepsilon \leq \frac{28,1 - 25}{\sqrt{2,4}})$$

$$F(23,55) = P(Ax \leq 23,55) = P(\sqrt{2,4} * \varepsilon + 25 \leq 23,55) = P(\varepsilon \leq \frac{23,55 - 25}{\sqrt{2,4}})$$

Donde ε representa una variante con distribución $N(0,1)$; leídos en tablas, los valores:

$$P\left(\varepsilon \leq \frac{28,1 - 25}{\sqrt{2,4}}\right) = P(\varepsilon \leq 2) = 0.97725$$

$$P\left(\varepsilon \leq \frac{23,55 - 25}{\sqrt{2,4}}\right) = P(\varepsilon \leq -0,93) = 0.17619$$

Habiéndose tomado $\sqrt{2,4} = 1,55$ es por tanto:

$$P(23,55 \leq \varepsilon \leq 28,1) = 0.97725 - 0.17619 = 0.80106$$

17. La duración aleatoria de las unidades producidas de un artículo, se distribuye según la ley normal, con desviación típica igual a seis minutos. Elegidas al azar cien unidades, resulto ser la duración media de 14,35 minutos. Elaborar el intervalo de confianza del 99% para la duración media de las unidades producidas.

SOLUCIÓN:

Si convenimos en representar por la variante ξ , la duración aleatoria, dicha variante representa la población de la cual sabemos posee una distribución:

$$N(0,6)$$

Para determinar el intervalo de confianza nos basaremos en la distribución del estadístico media muestral Ax que, sabemos, se distribuye como una variante:

$$N\left(0, \frac{6}{\sqrt{100}}\right)$$

Puesto que la distribución poblacional es normal con varianza conocida, resultando, entonces, dicho intervalo expresado en la forma:

$$Ix \equiv \left[Ax \pm \frac{\sigma}{\sqrt{n}} \right] * \lambda_{\alpha}$$

$$P\left[\frac{\sqrt{n} \cdot \|Ax - 0\|}{\sigma}\right] \leq \lambda_{\alpha=0,99}$$

Leído el valor en λ_{α} resulta ser:

$$\lambda_{\alpha} = 2,575$$

con lo que el intervalo de confianza para el que, con una probabilidad de 0,99, pertenece al parámetro θ , duración media, es:

$$(Ax - (0,6)2,575; Ax + (0,6)2,575)$$

Y habida cuenta la información de que disponemos, dada por la muestra, al ser:

$$Ax = 14,35 \text{ minutos}$$

Podemos esperar que en un 99% de los casos, la duración media este comprendida entre los valores:

$$(14,35 - (0,6)2,575; 14,35 + (0,6)2,575)$$

Esto es:

$$(12,805; 15,895)$$

Habiendo tomado:

$$(0,6)2,575 = 1,545$$

18. Se considera una población representada por una variante ξ , de suerte que la distribución de probabilidad poblacional viene definida por una función de densidad:

$$f(x) = \frac{1}{4} \text{ para } 0 \leq x \leq 4$$

$$f(x) = 0 \text{ para cualquier otro valor de } x$$

Determinar la distribución de probabilidad del estadístico media muestral, supuesto extraídas muestras de tamaño 3.600, muestreo aleatorio simple.

SOLUCIÓN:

Del comportamiento en probabilidad del estadístico α_x , sabemos que:

$$E[a_x] = \eta \quad , \quad V[a_x] = \frac{\sigma^2}{n}$$

Y en nuestro caso:

$$E[a_x] = 2 \quad V[a_x] = \frac{4}{10800}$$

Veamos entonces, si podemos determinar la distribución de probabilidad de a_x . Tal y como a_x . Es definido:

$$a_x = \frac{\sum_{i=1}^n x_i}{n}$$

a_x . Es una variante definida como suma de variantes x_i , y habida cuenta el muestreo que se considera, aleatorio simple, independientes en probabilidad, todas con igual distribución e igual a la distribución poblacional, por tanto con media y varianza finitas. De ahí que, usando del teorema central del límite, al ser el número de variantes x_i , que se componen, suficientemente grande, puesto que el tamaño muestral es 3600, resulta procedente aproximar la distribución normal. Por ello el estadístico a_x . Se comporta como una variante normal de parámetros:

$$N\left(2, \frac{2}{103,9}\right)$$

19. Se consideran dos poblaciones representadas por las variantes ξ_1 y ξ_2 , de suerte que las distribuciones de probabilidad de cada una de ellas son $N(\eta_1, \sigma_1)$ y $N(\eta_2, \sigma_2)$, respectivamente. Extraídas muestras, muestreo aleatorio simple, de tamaño n , de la primera población, y de tamaño m de la segunda, determinar la distribución del estadístico: $a_x - a_y$, donde:

$$a_x = \frac{\sum_{i=1}^n x_i}{n} \quad a_y = \frac{\sum_{j=1}^m y_j}{m}$$

Supuesto que las muestras obtenidas de cada población son independientes.

SOLUCIÓN:

Para la determinación de la distribución de probabilidad de la variante $a_x - a_y$, procederemos a obtener su función característica. Así al ser por definición:

$$\varphi_{a_x - a_y}(t) = E\left[e^{it(ax - ay)}\right]$$

De donde, al ser las variantes a_x y a_y independientes, puesto que se supone que las muestras extraídas de la primera población son independientes de las de la segunda, se tendrá:

$$\varphi_{a_x - a_y}(t) = E\left[e^{ita_x}\right] \cdot E\left[e^{-ita_y}\right] = \varphi_{a_x}(t) \cdot \varphi_{a_y}(-t)$$

Por lo que podemos decir que:

$$\varphi_{a_x - a_y}(t) = \left(e^{it\eta_1 - (1/2)t^2\left(\frac{\sigma^2}{n}\right)} \right) \left(e^{-it\eta_2 - (1/2)t^2(\sigma^2/m)} \right)$$

Función característica que expresa que la distribución de probabilidad de la variante $a_x - a_y$, es una distribución normal de parámetros:

$$E[ax - ay] = \eta_1 - \eta_2$$

$$V[ax - ay] = \frac{\sigma^2}{n} + \frac{\sigma^2}{m}$$

20. Se considera una población tal que sus elementos son susceptibles de poseer o no una característica c. Supuesto extraídas muestras de tamaño n, muestreo aleatorio simple, determinar: El valor probable del estadístico “proporción de elementos de la muestra que poseen la característica c”.

SOLUCIÓN:

Convengamos en representar por 1 a cada elemento de la población que posea la característica c, por 0 a los elementos que no posean y por Π a la proporción de elementos de la población que posean dicha característica.

Extraídas muestras de tamaño n, muestreo aleatorio simple, $(x_1, x_2, x_3, \dots, x_n)$, las componentes muestrales las componentes muestrales tomarán los valores 0 ó 1, según que el elemento extraído posea o no la característica c, la proporción de elementos de muestra que poseen tal característica vendrá expresada entonces:

$$p = \frac{\sum_{i=1}^n x_i}{n} = \frac{x}{n}$$

Donde x es por tanto, la suma de los componentes muestrales, igual a la suma de números 1 que hayan aparecido en la muestra. Al ser el muestreo aleatorio, x , representará el número aleatorio de elementos que aparecen en la muestra con la característica c y por ello será una variable aleatoria.

Definido, pues el estadístico p pasemos a determinar los parámetros fundamentales de su distribución.

1) Valor probable de p

Tal y como se ha establecido el estadístico p , se tendrá:

$$E[p] = E\left[\frac{x}{n}\right] = \frac{1}{n} E[x]$$

Y como quiera que la variante x viene expresada como suma de variantes x_i , que solo pueden tomar los valores 1 ó 0, independientes entre sí, puesto que el muestreo es aleatorio simple, su distribución de probabilidad será:

$$\beta(n, \pi)$$

Habiendo cuenta que el número máximo de veces que puede presentarse el elemento representado por 1 es n , tamaño muestral y ser π la probabilidad de extraer de la población un elemento representado por uno. Por ello pues:

$$E[p] = n\pi$$

De donde:

$$E[p] = 1/n E[x] = \pi$$

21. Se estudiaron 40 muestras de aceite crudo de determinado proveedor con el fin de detectar la presencia del níquel mediante una prueba que nunca da un resultado erróneo. Si en 5 de dichas muestras se observó la presencia de níquel, ¿podemos creer al proveedor cuando asegura que a lo sumo el 8% de las muestras contienen níquel?

SOLUCIÓN:

Llamemos p a la proporción de muestras que contienen níquel. Si la prueba nunca da un resultado erróneo la variable P_0 , que representa la proporción de pruebas positivas al analizar 40 muestras satisface.

$$\frac{P_0 - p}{\sqrt{\frac{p(1-p)}{40}}} = Z \approx N(0,1)$$

Contrastamos la hipótesis nula

$$H_0 : p = 0,08$$

Frente a la alternativa

$$H_a : p > 0,08$$

Al tratarse de un contraste unilateral, con la región crítica a la derecha, esta corresponde a valores de la distribución muestral superiores a $z_{0,95} = 1,645$, si consideramos un nivel de significación $\alpha = 0,05$

En nuestro caso $p_0 = \frac{5}{40} = 0,125$ de modo que

$$\frac{0,125 - 0,08}{\sqrt{\frac{0,08 \cdot 0,92}{40}}} = 1,049 < 1,645$$

Con lo que no podemos rechazar la hipótesis nula.

Llamaremos N al suceso que representa la presencia de níquel. Como la prueba es positiva en el 80% de los casos en que hay níquel pero también en su ausencia, con probabilidad igual a 0,01, la probabilidad de que una prueba resulte positiva es

$$\hat{p} = 0,79p + 0,01$$

Y ahora la variable P_0 , que representa la proporción muestral de pruebas positivas y no la de contenido real de níquel, satisface

$$\frac{P_0 - \hat{p}}{\sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{40}}} \approx N(0,1)$$

La hipótesis nula pasa a ser

$$H_0 : \hat{p} = 0,0732$$

Frente a la alternativa

$$H_a : p > 0,0732$$

Considerando el mismo nivel de significacion, tenemos que

$$\frac{0,125 - 0,0732}{\sqrt{\frac{0,0732 \cdot 0,9268}{40}}} = 1,2578 < 1,645$$

Llegando a la misma conclusion

22. La resistividad electrica de ciertas barras de aleacion de Cromo- molibdeno es una variable $N(12,5 ; 4,1)$. Un investigador acaba de calibrar un aparato que mide dicha resistividad y para comprobar que lo ha hecho bien utiliza el sistema consiguiente en medir cuatro barras y aceptar que el calibrado es bueno si encuentra al menos un valor inferior y otro superior a 12,5.

Determinar el nivel de significacion del contraste que esta llevando a cabo. ¿Es sensible el contraste a una mayor o menor dispersion de la variable resistividad?

SOLUCIÓN:

Planteamos la hipotesis nula:

$$H_0 : \mu = 12,5$$

Contra la alternativa

$$H_a : \mu \neq 12,5$$

Sea X el numero de valores mayores que 12,5 en una muestra de extension 4. Si la hipotesis nula es cierta, X es una variable $B(4 ; 0,5)$ independientemente del valor de la varianza.

Rechazar la hipotesis nula equivale

$$X=4 \text{ o bien } X=0$$

Por lo tanto, el nivel de significacion α es:

$$P(X=4) + P(X=0) = 0,125$$

El contraste no es sensible a la dispersion de la variable resistividad pues esta dispersion no afecta a la variable X, utilizada como estadigrafo, que es B(4,0,5) al margen del valor de la varianza de la variable resistividad.

23. Los rodamientos esfericos que fabrica una maquina deben de tener un diametro uniforme para ser aptos para su uso.El responsable de la maquina asegura que la varianza es $\sigma^2 = 0,025$.Medidos 50 rodamientos se obtuvo una varianza muestral $s^2 = 0,0272$. ¿Es compatible este resultado con la afirmacion previa?

SOLUCIÓN:

Planteamos la hipotesis nula:

$$H_0 = \sigma^2 = 0,025$$

Frente a la alternativa

$$H_a : \sigma^2 > 0,025$$

Pues una varianza σ^2 inferior a 0,025 no debe ser objeto de nuestra preocupacion, ya que supone menor dispersion y , por tanto mayor uniformidad de la variable diametro de rodamientos. Apoyandonos en la distribucion χ^2 con 49 crados de libertad dada por

$$\frac{50S^2}{\sigma^2}$$

Obtenemos para nuestros datos

$$\frac{50 \cdot 0,0272}{0,025} = 54,4$$

Teniendo en cuenta que, para 49 grados de libertad,

$$\chi_{49,0,95}^2 = 66,33$$

La desigualdad 54,4 menor que 66,33

Conduce a aceptar la hipotesis nula.

24. Se considera una poblacion representada por la variante ξ , de suerte que la distribucion poblacional es una distribucion $N(\eta, \sigma)$, con σ conocida. Supuesto extraidas muestras de tamaño n , muestreo aleatorio simple, determonar la distribucion de probabilidad del estadistica media muestral.

SOLUCIÓN:

Sabemos que $E[ax] = \eta$, $V[ax] = \frac{\sigma^2}{n}$

Veamos , entonces, si el conocimiento del modelo de distribucion de probabilidad poblacional permite determinar , no solo los parametros fundamentales de la distribucion de probabilidad del estadistico ax , que ya los conocemos como acabamos de ver, sino tambien su distribucion de probabilidad. Para ello, usaremos del concepto de funcion caracteristica y trass u obtencion, identificaremos el modelo que corresponde a la variante ax . Asi por definicion:

$$\varphi_{ax}(t) = E[e^{itx_1/n} \dots e^{itx_n/n}] = E[e^{itx_1/n}] \cdot \dots \cdot E[e^{itx_n/n}]$$

Por ser el muestreo que se emple aleatorio simple

De donde por, definicion da funcion caracteristica es:

$$E[e^{itx_n/n}] = \varphi_{xi}(t/n)$$

Con ello:

$$\varphi_{ax}(t) = \varphi_{x1}(t/n) \cdot \dots \cdot \varphi_{xn}(t/n)$$

Como quiera que todas las variantes poseen igual distribucion e igual a la distribucion poblacional, sera:

$$\varphi_{xi}(t/n) = \varphi_{\xi}(t/n)$$

Con lo que la funcion caracteristica resulta ser la correspondiente a una distribucion:

$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

Por tanto , el estadistico media muestral se comporta como una variante normal de parametros

$$\mu \text{ y } \frac{\sigma}{\sqrt{n}}$$

25. Se considera una poblacion representada por una variante ξ , de suerte que ϑ y σ^2 , representan los parametros media y varianza poblacional, respectivamente. Si estimamos la media poblacional, ϑ , a traves de la media muestral a_x comprobar que dicho estimador es consistente (Supuesto extraidas muestras de tamaño n , muestreo aleatorio simple).

SOLUCIÓN:

Por definicion, a_x sera estimador consistente de ϑ , si:

$$a_x \rightarrow \vartheta \Leftrightarrow P(|a_x - \vartheta| > \varepsilon) \rightarrow 0$$

Es decir, si la media muestral converge en probabilidad a la media poblacional.

Como del comportamiento en probabilidad de la media muestral sabemos que;

$$E[a_x] = \vartheta$$

$$V[a_x] = \sigma^2$$

se tendra que, habida cuenta la desigualdad de tchebueff de donde:

$$P(|a_x - \vartheta| > \varepsilon) \rightarrow 0$$

Por tanto, es estimador consistente de la media poblacional. Notese que este resultado es valido cualquiera que sea el tipo de distribucion poblacional, siempre que no carezca de valor medio y varianza, habida cuenta que, al aplicar la desigualdad de tchebycheff, solo hemos hecho uso de la informacion que proporciona el valor probable y la varianza.

26. Se considera una poblacion representada por una variante η , de suerte que la distribucion de probabilidad de la poblacion viene definida por la funcion de densidad:

$$f(x) = \frac{2(\theta - x)}{\theta^2} \text{ para } 0 \leq x \leq \theta$$

$$f(x) = 0 \text{ para cualquier otro valor}$$

Determinar el estimador, por el metodo de los momentos, del parametro poblacional. θ

SOLUCIÓN:

Teniendo en cuenta que solo es uno el parametro a estimar θ , el estimador por el metodo de los momentos , vendra dado como solucion de la ecuacion:

$$E[\xi] = \int_{-\infty}^{\infty} x dF(x/\theta) = a_1$$

Ecuacion obtenida al igualar le momento de orden 1 de la poblacion, al momento de orden 1 de la muestra. En nuestro caso, al ser la distribuciun poblacional de tipo continuo y teniendo en cuenta la definicion de la funcion de densidad ,sera:

$$\int_0^{\theta} x \frac{2(\theta - x)}{\theta^2} dx = a_x$$

Determinaremos la integral anterior.Se tendra:

$$\int_0^{\theta} x \frac{2(\theta - x)}{\theta^2} dx = \theta/3$$

Con lo que $\theta/3 = a_x$

Con ello: $\theta^* = 3a_x$

Sera el estimador, por el metodo de los momentos del parametro θ

27. Se considera una poblacion representada por una variable ξ , de suerte que la distribucion de probabilidad de la poblacion es normal, con desviacion tipica igual a 10.Determinar la probabilidad de que el estadistica media muestral supere el valor m edio poblacional en , por lo menos 0,2 , supuesto extraidas muestras de tamaño 100 , muestreo aleatorio simple.

SOLUCIÓN:

Si convenimos en representar por μ el valor medio poblacional , el suceso cuya probabilidad vamos a obtener puede ser expresado en la forma: $a_x - \eta \geq 0,2$. La diferencia : $a_x - \eta$ es aleatoria , por ser a_x una variable aleatoria, y su disatribucion de probabilidad podemos obtenerla a traves de la funcion caracteristica:

$$\varphi_{a_x - \eta}(t) = E[e^{it(ax-\eta)}] = e^{-it\mu} \cdot \varphi_{a_x}(t)$$

Donde:

$$\varphi_{ax}(t) = e^{it\mu - (1/2)t^2(\sigma^2/n)}$$

Así la función característica:

$$\varphi_{ax-\eta}(t) = e^{-1/2t^2\sigma^2/n}$$

Expresa que la variante $a_x - \eta$ se comporta con arreglo a la distribución de probabilidad:

$$N\left(0, \frac{\sigma}{\sqrt{n}}\right)$$

Conocida la distribución ($a_x - \eta$) en nuestro caso,

$$N\left(0, \frac{10}{\sqrt{100}}\right)$$

La probabilidad a obtener será determinada en la forma:

$$P(ax - \eta \geq 0,2) = P\left(\frac{10}{\sqrt{100}} \xi' \geq 0,2\right) = P(\xi' \geq 0,2)$$

donde ξ' , es una variante $N(0,1)$. Leída en tablas de probabilidad:

$$P(\xi' \geq 0,2) = 0,420$$

Resulta pues:

$$P(ax - \eta \geq 0,2) = 0,4207$$

28. En una muestra de 65 sujetos las puntuaciones en una escala de extroversión tienen una media de 32,7 puntos y una desviación típica de 12,64.

a) Calcule a partir de estos datos el correspondiente intervalo de confianza, a un nivel del 90%, para la media de la población.

b) Indique, con un nivel de confianza del 95%, cual sería el máximo error que podríamos cometer al tomar como media de la población el valor obtenido en la estimación puntual.

SOLUCIÓN:

a) Buscando en las tablas de la t de Student obtenemos que el valor que deja por debajo una probabilidad del 95% es 1,671 (aproximadamente). Sustituyendo los valores de esta muestra en la expresión del intervalo de confianza obtenemos:

$$(32,7 - 1,671 * 12,64 / 8, 32,7 + 1,671 * 12,64 / 8)$$

operando (30,06, 35,34)

b) En las tablas de la t de Student encontramos que el valor de la variable que deja por debajo una probabilidad de 0,975 es 2. En consecuencia a un nivel de confianza del 95% la media de la población puede valer

$$32,7 \pm 2 * 12,64 / 8$$

luego el máximo error que se puede cometer, a este nivel de confianza, es: 3,16

29. Se considera una población representada por una variante ξ , de suerte que la distribución de probabilidad de la población es $N(0,2)$. Si como estimador del parámetro poblacional θ se toma el estadístico media muestral, determinar el tamaño muestral n para que, con una probabilidad de 0,95, el error de la estimación producido no sea superior a 0,2.

SOLUCIÓN:

Si la estimación de la media poblacional θ se realiza a través del estadístico media muestral, A_x , el error de la estimación vendrá dado por:

$$\theta - A_x$$

Y habida cuenta el enunciado hemos de obtener el tamaño muestral para el que:

$$P[\theta - A_x \leq 0.2 = 0.95]$$

O bien :

$$P[A_x - 0.2 \leq \theta \leq 0.2 + A_x = 0.95]$$

Como quiera que el intervalo de confianza para la media poblacional obtenido en base a la distribución del estadístico media muestral es:

$$A_x - \frac{\sigma}{\sqrt{n}} * \lambda; A_x + \frac{\sigma}{\sqrt{n}} * \lambda$$

Puesto que suponemos la población normal con varianza conocida, se tendrá que la cota de error ,0,2 será igual a:

$$0,2 = \frac{\sigma}{\sqrt{n}} * \lambda$$

y por lo tanto:

$$n = \frac{\lambda^2 * \sigma^2}{0,2^2}$$

teniendo en cuenta que la λ es tal que:

$$P\left[\frac{\sqrt{n}}{\sigma} * \frac{Ax - \theta}{\sigma} \leq \lambda\right] = 0,95$$

Leyendo en las tablas el valor de $\lambda = 1,96$ con ello $n = 384,16$ con lo que al tener que ser el tamaño un número entero resultará 384.

Así pues tomando muestras de tamaño 384, podemos esperar que en el 95% de los casos, las estimaciones producidas por la media muestral produzcan errores no superiores a 0,2.

30. En una muestra aleatoria de 90 pacientes se mide el nivel de glucosa en sangre en ayunas. Se obtiene $\bar{X} = 132$ mg/dl y $s^2 = 109$. Construir el IC al 95%.

SOLUCIÓN:

Usando la fórmula general para cuando s^2 es desconocida:

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

podemos, o bien mirar a las tablas de la t , el valor de $t_{0,025}$ que para 89 grados de libertad es 1,99, o bien como $n > 30$ aproximar a la z y usar el valor 1,96.

$$IC\ 95\% : 132 \pm 1,99 \times \frac{\sqrt{109}}{\sqrt{90}} = 132 \pm 2,19$$

31. Para evaluar una vacuna para la gripe se selecciona un grupo de 200 individuos de riesgo. Se eligen 100 de ellos y se les suministra la vacuna; de ellos 10 pasan la gripe. Construir un IC al 95% para la probabilidad de pasar la gripe si se está vacunado. En los otros 100 pacientes sin vacunar la pasan 20. ¿Es eficaz la vacuna?

SOLUCIÓN:

La fórmula para calcular IC para proporciones es

$$\bar{p} \pm z_{\alpha/2} \sqrt{\frac{pq}{n}}$$

y aproximando p y q por sus estimaciones

$$0.1 \pm 1.96 \sqrt{\frac{0.1 * 0.9}{100}} = (0.04 - 0.16)$$

es decir, hay una probabilidad del 95% de que la probabilidad de pasar la gripe si se está vacunado esté comprendida entre el 4% y el 16%. Para los no vacunados

$$0.2 \pm 1.96 \sqrt{\frac{0.2 * 0.9}{100}} = (0.12 - 0.28)$$

32. El parámetro de una distribución de Poisson puede tomar uno de los cuatro valores siguientes: $a = 4; 4,5; 5,5; 6$. Decídase cual de ellos puede ser, considerando una muestra aleatoria simple de tamaño dos, tal que $x_1 = 3$ y $x_2 = 7$, y basándose en el principio de máxima verosimilitud.

SOLUCIÓN:

La muestra puede proceder de cuatro poblaciones distintas, según sea el valor que tome el parámetro a de la distribución de Poisson; de acuerdo con el principio de máxima verosimilitud, elegiremos el valor de a tal que la probabilidad de obtener la citada muestra sea máxima.

Calculamos la probabilidad para cada valor de a :

$$a = 4:$$

$$P(X_1 = 3; X_2 = 7) = 2 \cdot \frac{e^{-4} \cdot 4^3}{3!} \cdot \frac{e^{-4} \cdot 4^7}{7!} = 0,02326$$

$a = 4,5$:

$$P(X_1 = 3; X_2 = 7) = 2 \cdot \frac{e^{-4,5} \cdot 4,5^3}{3!} \cdot \frac{e^{-4,5} \cdot 4,5^7}{7!} = 0,02780$$

$a = 5,5$:

$$P(X_1 = 3; X_2 = 7) = 2 \cdot \frac{e^{-5,5} \cdot 5,5^3}{3!} \cdot \frac{e^{-5,5} \cdot 5,5^7}{7!} = 0,02798$$

$a = 6$:

$$P(X_1 = 3; X_2 = 7) = 2 \cdot \frac{e^{-6} \cdot 6^3}{3!} \cdot \frac{e^{-6} \cdot 6^7}{7!} = 0,02457$$

Cuando $a = 5,5$ obtenemos la mayor probabilidad de extraer la muestra (3; 7), por lo cual decidimos que el valor del parámetro a es 5,5.

33. La función de densidad $f(x; a) = ae^{-ax}$, $x \geq 0$, contiene el parámetro a , del cual se sabe que puede tomar uno de los tres valores siguientes: 0,5; 1; 1,5. Se toma una muestra aleatoria de tamaño uno, resultando un valor comprendido entre 1,7 y 2,6. Basándose en el principio de estimación de la máxima verosimilitud, determínese, a partir de la información suministrada por la muestra, por cual de los tres valores que puede tomar el parámetro debemos optar.

SOLUCIÓN:

Según el método de la máxima verosimilitud, elegiremos el valor de a que proporcione la mayor probabilidad de obtener la muestra.

$a = 0,5$:

$$P(1,7 \leq X \leq 2,6) = \int_{1,7}^{2,6} \frac{1}{2} e^{-x/2} dx = e^{-0,85} - e^{-1,3} = 0,1549$$

$a = 1$:

$$P(1,7 \leq X \leq 2,6) = \int_{1,7}^{2,6} e^{-x} dx = e^{-1,7} - e^{-2,6} = 0,1084$$

$a = 1,5$:

$$P(1,7 \leq X \leq 2,6) = \int_{1,7}^{2,6} \frac{3}{2} e^{-3x/2} dx = e^{-2,55} - e^{-3,90} = 0,0578$$

Al ser $a = 0,5$ el valor del parámetro que proporciona la mayor probabilidad, decidimos que $a = 0,5$, y la función de densidad es $f(x) = \frac{1}{2} e^{-x/2}$

34. Si una variable aleatoria tiene por función de densidad

$$f(x; a) = \frac{2a}{1-a} x^{\frac{3a-1}{1-a}}$$

$0 \leq x \leq 1$; $a > 0$ hállese el estimador del parámetro a por el método de máxima verosimilitud, en muestras aleatorias simples de tamaño n .

SOLUCIÓN:

La función de verosimilitud es:

$$L = L(x_1, \dots, x_n; a) = \frac{2a}{1-a} x_1^{\frac{3a-1}{1-a}} \dots \frac{2a}{1-a} x_n^{\frac{3a-1}{1-a}} = \frac{2^n a^n}{(1-a)^n} (x_1 \dots x_n)^{\frac{3a-1}{1-a}}$$

Tomamos logaritmos neperianos:

$$\ln L = n \ln 2 + n \ln(a) - n \ln(1-a) + \frac{3a-1}{1-a} \sum_{i=1}^n \ln x_i$$

Calculamos la derivada parcial de $\ln L$ respecto a a y la igualamos a cero, resolviendo la ecuación:

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} + \frac{n}{1-a} + \frac{2}{(1-a)^2} \sum_{i=1}^n \ln x_i = 0$$

$$\frac{2}{(1-a)^2} \sum_{i=1}^n \ln x_i = -\frac{n}{a(1-a)}; \quad \sum_{i=1}^n \ln x_i = \frac{na-n}{2a}$$

$$\hat{a} = \frac{n}{n - 2 \sum_{i=1}^n \ln x_i}$$

35. Calcúlese los estimadores máximo verosímiles de los parámetros a y b de las funciones de densidad

a) $f(x; y; a) = a^2 e^{-a(x+y)}, \quad x \geq 0; \quad y \geq 0$

b) $f(x; y; a; b) = abx^{a-1}y^{b-1}, \quad 0 \leq x \leq 1; \quad 0 \leq y \leq 1$

en muestras aleatorias simples de tamaño n .

SOLUCIÓN:

a)

$$L = L(x_1 y_1; \dots; x_n y_n; a) = a^2 e^{-a(x_1+y_1)} \dots a^2 e^{-a(x_n+y_n)} = a^{2n} e^{-a \sum_{i=1}^n (x_i+y_i)}$$

$$\ln L = 2n \ln a - a \sum_{i=1}^n (x_i + y_i)$$

$$\frac{\partial \ln L}{\partial a} = \frac{2n}{a} - \sum_{i=1}^n (x_i + y_i) = 0$$

$$\hat{a} = \frac{2n}{\sum_{i=1}^n (x_i + y_i)} = \frac{2}{\bar{x} + \bar{y}}$$

b)

$$L = L(x_1, \dots, x_n; y_1, \dots, y_n; a, b) = a^n b^n (x_1 \dots x_n)^{a-1} (y_1 \dots y_n)^{b-1}$$

$$\ln L = n \ln a + n \ln b + (a-1) \sum_{i=1}^n \ln x_i + (b-1) \sum_{i=1}^n \ln y_i$$

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} + \sum_{i=1}^n \ln x_i \Big| = 0$$

$$\frac{\partial \ln L}{\partial b} = \frac{n}{b} + \sum_{i=1}^n \ln y_i = 0$$

$$\hat{a} = -\frac{n}{\sum_{i=1}^n \ln x_i}; \quad \hat{b} = -\frac{n}{\sum_{i=1}^n \ln y_i}$$

36. Dada la variable aleatoria X , con función de densidad

$$f(x; a) = ae^{-ax}; \quad 0 \leq x; \quad 0 < a,$$

Calcúlese la distribución en el muestreo del estimador máximo verosímil del parámetro a , en muestras aleatorias simples de tamaño 2.

SOLUCIÓN:

La función de verosimilitud es

$$L = L(x_1; x_2; a) = ae^{-ax_1} ae^{-ax_2} = a^2 e^{-a(x_1+x_2)}$$

$$\ln L = 2 \ln a - a(x_1 + x_2);$$

$$\frac{\partial \ln L}{\partial a} = \frac{2}{a} - (x_1 + x_2) = 0;$$

$$\hat{a} = \frac{2}{x_1 + x_2} = \frac{1}{\bar{x}}$$

El estimador máximo verosímil es el inverso de la media muestral. Función de distribución del estimador:

$$\begin{aligned} G(y) &= P(\hat{a} \leq y) = P\left(\frac{1}{\bar{x}} \leq y\right) = P\left(\frac{1}{y} \leq \bar{x}\right) = 1 - P\left(\bar{x} \leq \frac{1}{y}\right) = 1 - P\left(X_1 + X_2 \leq \frac{2}{y}\right) = \\ &= 1 - \int_0^{\frac{2}{y}} \int_0^{\frac{2}{y}-x_1} L(x_1; x_2; a) dx_1 dx_2 = 1 - \int_0^{\frac{2}{y}} \int_0^{\frac{2}{y}-x_1} a^2 e^{-a(x_1+x_2)} dx_1 dx_2 = \left(1 + \frac{2a}{y}\right) e^{-2a/y} \end{aligned}$$

La función de densidad es

$$g(y) = G'(y) = \frac{4a^2}{y^3} e^{-2a/y}; \quad y \geq 0$$

37. En una población de tamaño 64 se estudia una característica X medida sobre sus individuos de la que se sabe que su media es 1012 y su desviación típica es 25. Hallar intervalos de confianza para el valor medio de la característica X con coeficientes de confianza del 90% y 95%.

SOLUCIÓN:

Estamos ante un caso de cálculo de intervalo de confianza para la media de una población normal con varianza conocida. En esta situación sabemos que:

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

lo que nos lleva al intervalo de confianza para la media siguiente:

$$\bar{x} \pm Z_{\alpha/2} \sigma/\sqrt{n}$$

Para un coeficiente de confianza del 90% se tiene que:

$$1 - \alpha = 0,9 \Rightarrow \alpha = 0,1 \Rightarrow \frac{\alpha}{2} = 0,05 \Rightarrow Z_{\alpha/2} = 1,645$$

El intervalo de confianza será entonces:

$$1012 \pm 1,645 \frac{25}{\sqrt{1012}} = [1006,86 \quad 1017,14]$$

Podemos concluir entonces que hay una probabilidad del 90% de que el valor medio de la característica esté entre 1006,86 y 1017,14.

Para un coeficiente de confianza del 95% se tiene que:

$$1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow Z_{\alpha/2} = 1,96$$

El intervalo de confianza será entonces:

$$1012 \pm 1,96 \frac{25}{\sqrt{1012}} = [1005,875 \quad 1018,125]$$

Podemos concluir entonces que hay una probabilidad del 95% de que el valor medio de la característica esté entre 1005,875 y 1018,125.

Se observa que al aumentar el coeficiente de confianza aumenta la amplitud del intervalo de confianza.

38. Se analizan 9 zumos de fruta y se ha obtenido un contenido medio de fruta de 22 mg por 100 cc de zumo. La varianza poblacional es desconocida, por lo que se ha calculado la cuasidesviación típica de la muestra que ha resultado ser 6,3 mg de fruta por cada 100 cc de zumo. Suponiendo que el contenido de fruta del zumo es normal, estimar el contenido medio de fruta de los zumos tanto puntualmente como por intervalos al 95% de confianza.

SOLUCIÓN:

Para la estimación puntual sabemos que en poblaciones normales un estimador lineal insesgado para la media poblacional es la media muestral, luego se puede estimar el contenido medio en fruta de los zumos en 22 mg por cada 100 cc de zumo.

Para la estimación por intervalos estamos ante un caso de cálculo de intervalo de confianza para la media de una población normal con varianza desconocida. En esta situación sabemos que:

$$\frac{\bar{x} - \mu}{S/\sqrt{n}} \rightarrow t_{n-1}$$

lo que nos lleva al intervalo de confianza para la media siguiente:

$$\bar{x} \pm t_{\alpha/2, n-1} S/\sqrt{n}$$

Para un coeficiente de confianza del 95% se tiene que:

$$1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow t_{0,05/2, 9-1} = t_{0,025, 8} = 2,0307$$

El intervalo de confianza será entonces:

$$22 \pm 2,307 \frac{6,3}{\sqrt{9}} = [17,15 \quad 26,84]$$

Podemos concluir entonces que hay una probabilidad del 95% de que el valor medio del contenido en fruta del zumo esté entre 17,15 y 26,84 mg por cada 100 cc de zumo.

39. Una firma comercial encuesta a 100 individuos para conocer sus opiniones sobre la elección de dos productos alternativos A y B recientemente fabricados. El resultado de la encuesta arroja que el producto A lo han elegido 55 individuos y el producto B 45. Hallar un intervalo de confianza al 95% para la proporción de individuos que eligen cada producto.

SOLUCIÓN:

Estamos ante el caso del cálculo de intervalos de confianza para proporciones, ya que este problema es ajustable al caso en que $\hat{p} = x/n$ es la proporción estimada del número de veces que aparece un suceso de Bernoulli (los encuestados eligen A o B exclusivamente) de entre n repeticiones de un experimento (x designa el número de veces que aparece el suceso, es decir el número de veces que los encuestados eligen A o B). En este caso el intervalo de confianza para la proporción se basa en el siguiente estadístico:

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-p)}{n}}} \rightarrow N(0,1)$$

lo que nos lleva al intervalo de confianza para p definido por:

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-p)}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-p)}{n}}$$

En nuestro caso, para el producto A tenemos el intervalo:

$$\left[\frac{55}{100} - 1,96 \sqrt{\frac{0,45 * 0,55}{100}} \quad \frac{55}{100} + 1,96 \sqrt{\frac{0,45 * 0,55}{100}} \right] = [0,45 \quad 0,65]$$

Para el producto B tenemos el intervalo:

$$\left[\frac{45}{100} - 1,96 \sqrt{\frac{0,55 * 0,45}{100}} \quad \frac{45}{100} + 1,96 \sqrt{\frac{0,55 * 0,45}{100}} \right] = [0,35 \quad 0,55]$$

Como conclusión podemos decir que hay una probabilidad del 95% entre 0,45 y 0,65 de que el producto elegido sea el A, y hay una probabilidad entre 0,35 y 0,55 de que el producto elegido sea el B.

40. En un proceso de fabricación de pilas alcalinas se sabe que su duración media es de 1100 horas y que dicha duración sigue una distribución normal. El nuevo proceso busca reducir la dispersión de la duración de las pilas por lo que se hace necesario construir

intervalos de confianza para la citada dispersión con coeficientes de confianza 90% y 98%. Construir dichos intervalos a partir de una muestra de tamaño 20 cuya dispersión es $\sqrt{2240}$ horas.

SOLUCIÓN:

Estamos ante el caso del cálculo de intervalos de confianza para la desviación típica de una distribución normal con media conocida. En este caso el intervalo de confianza para la desviación típica se basa en el siguiente estadístico:

$$\frac{n\hat{\sigma}^2}{\sigma^2} \rightarrow X_{n-1}^2$$

lo que nos lleva al intervalo de confianza para la desviación típica definido por:

$$\left(\sqrt{\frac{n\hat{\sigma}^2}{X_{\alpha/2,n}^2}}, \sqrt{\frac{n\hat{\sigma}^2}{X_{1-\alpha/2,n}^2}} \right)$$

En nuestro caso, para el nivel del 90% tenemos:

$$1 - \alpha = 0,9 \Rightarrow \alpha = 0,1 \Rightarrow \frac{\alpha}{2} = 0,05 \Rightarrow X_{0,05;20}^2 = 31,4, \quad X_{0,95;20}^2 = 10,9$$

El intervalo de confianza será entonces:

$$\left(\sqrt{\frac{20 * 2240}{31,4}}, \sqrt{\frac{20 * 2240}{10,9}} \right) = [119,42 \quad 203,2]$$

Para el nivel del 95% tenemos:

$$1 - \alpha = 0,98 \Rightarrow \alpha = 0,02 \Rightarrow \frac{\alpha}{2} = 0,01 \Rightarrow X_{0,01;20}^2 = 37,6, \quad X_{0,99;20}^2 = 8,26$$

El intervalo de confianza será entonces:

$$\left(\sqrt{\frac{20 * 2240}{37,6}}, \sqrt{\frac{20 * 2240}{8,26}} \right) = [109,2 \quad 232,88]$$

Podemos concluir entonces que hay una probabilidad del 90% de que la dispersión de la duración de las pilas en el nuevo proceso de fabricación esté entre 119,42 y 203,2 horas, y hay una probabilidad del 98% de que la dispersión de la duración de las pilas en el nuevo proceso de fabricación esté entre 109,2 y 232,88 horas.

41. Se sabe que la longitud de los diámetros de los tornillos fabricados por una máquina siguen una distribución normal y se busca un intervalo en el cual se encuentre la variabilidad de las longitudes de los tornillos fabricados por la máquina con una probabilidad del 80%. Construir dicho intervalo sabiendo que una muestra de 16 tornillos presenta una variabilidad cuantificada en 30.

SOLUCIÓN:

Estamos ante el caso del cálculo de intervalos de confianza para la varianza de una distribución normal con una media desconocida. En este caso el intervalo de confianza para la varianza se basa en el siguiente estadístico:

$$\frac{(n-1)S^2}{\sigma^2} \rightarrow X_{n-1}^2$$

lo que nos lleva al intervalo de confianza para la varianza definida por:

$$\left(\frac{(n-1)S^2}{X_{\alpha/2,n}^2}, \frac{(n-1)S^2}{X_{1-\alpha/2,n}^2} \right) = \left(\frac{n\sigma^2}{X_{\alpha/2,n}^2}, \frac{n\sigma^2}{X_{1-\alpha/2,n}^2} \right)$$

En nuestro caso, para el nivel 90% tenemos:

$$1 - \alpha = 0,8 \Rightarrow \alpha = 0,2 \Rightarrow \frac{\alpha}{2} = 0,1 \Rightarrow X_{0,1;15}^2 = 22,3, \quad X_{0,9;15}^2 = 8,55$$

El intervalo de confianza será entonces:

$$\left(\frac{16 * 30}{22,3}, \frac{16 * 30}{8,55} \right) = [21,43 \quad 56,14]$$

Luego el intervalos en el cual se encuentre la variabilidad de las longitudes de los tornillos fabricados por la máquina con una probabilidad del 80% es [21,43 56,14].

42. Un granjero dispone de dos explotaciones diferentes A y B con varias granjas cada una para la cría de pollos. Con el objetivo de estudiar la mortalidad de los pollos en las dos explotaciones observa el número de pollos muertos tomando una muestra de 4 granjas en la explotación A y otras 4 granjas en la explotación B obteniendo los siguientes resultados:

Nº de pollos muertos en las granjas de la explotación A: 16 14 13 17

Nº de pollos muertos en las granjas de la explotación B: 18 21 18 19

Suponiendo normalidad en las explotaciones, se trata de estudiar si la mortalidad de los pollos puede considerarse diferente en las dos explotaciones con un nivel de confianza del 95%. Resolver el problema bajo la hipótesis adicional de varianzas iguales en las explotaciones.

SOLUCIÓN:

Para resolver este problema hallaremos un intervalo de confianza para la diferencia del número medio de pollos muertos en ambas explotaciones al 95% y comprobaremos si dicho intervalo contiene el valor cero, en cuyo caso se puede aceptar la hipótesis de que la mortalidad media es similar en ambas granjas con una probabilidad del 95%. Estamos entonces ante el caso del cálculo de intervalos de confianza para la diferencia de medias de dos distribuciones normales con varianzas desconocidas y en general desiguales. En esta situación el intervalo de confianza para la diferencia de medias se basa en el siguiente estadístico:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \rightarrow t_f \text{ con } f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 + 1} + \frac{(S_2^2/n_2)^2}{n_2 + 1}} - 2$$

lo que nos lleva al intervalo de confianza para la varianza definido por:

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2, f} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Los valores de f y $t_{\alpha/2, f}$ para nuestro caso resultan ser:

$$f = \frac{\left(\frac{3,333}{4} + \frac{2}{4}\right)^2}{\frac{(3,333/4)^2}{5} + \frac{(2/4)^2}{5}} - 2 \cong 7 \Rightarrow t_{\alpha/2, f} = t_{0,025;6} = 2,365$$

El intervalo de confianza será entonces:

$$15 - 19 \pm 2,365 \sqrt{\frac{3,333}{4} + \frac{2}{4}} = [1,27 \quad 6,73]$$

Como este intervalo no contiene el valor cero, no se puede aceptar con una probabilidad del 95% que $\bar{x}_1 - \bar{x}_2 = 0$, es decir, $\bar{x}_1 = \bar{x}_2$. Por tanto, aceptamos la hipótesis de que las mortalidades en las dos granjas son significativamente distintas con un coeficiente de confianza del 95%. Además, como los dos extremos del intervalo son positivos, para todos sus valores $\bar{x}_1 - \bar{x}_2 > 0 \Rightarrow \bar{x}_1 > \bar{x}_2$, lo que indica que la mortalidad media es mayor en la granja A con una probabilidad del 95%.

Si suponemos ahora que las varianzas en las dos explotaciones son iguales, estamos ante el caso del cálculo de intervalos de confianza para la diferencia de medias de dos distribuciones normales con varianzas desconocidas e iguales. En esta situación el intervalo de confianza para la diferencia de medidas se basa en el siguiente estadístico:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightarrow t_{n_1+n_2-2} \text{ con } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

lo que nos lleva al intervalo de confianza para la varianza definido por:

$$\bar{x}_1 - \bar{x}_2 \pm (t_{\alpha/2, n_1+n_2-2}) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

El valor de S_p^2 para nuestro caso resulta ser:

$$S_p^2 = \frac{3,333(4-1) + 2(4-1)}{4+4-2} = 2,656$$

Además, $t_{\alpha/2, n_1+n_2-2} = t_{0,025;6} = 2,447$

El intervalo de confianza será entonces:

$$15 - 19 \pm 2,447 \sqrt{2,656} \sqrt{\frac{3,333}{4} + \frac{2}{4}} = [-6,82 \quad -1,18]$$

Como este intervalo no contiene el valor cero, no se puede aceptar con una probabilidad del 95% que $\bar{x}_1 - \bar{x}_2 = 0$, es decir, $\bar{x}_1 = \bar{x}_2$. Por tanto, aceptamos la hipótesis de que las mortalidades en las dos granjas son significativamente distintas con un coeficiente de confianza del 95%. Además, como los dos extremos del intervalo son negativos, para todos sus valores $\bar{x}_1 - \bar{x}_2 < 0 \Rightarrow \bar{x}_1 < \bar{x}_2$, lo que indica que la mortalidad media es mayor en la granja B con una probabilidad del 95%.

Es conveniente observar en este problema que la hipótesis de igualdad de varianzas en las dos poblaciones juega un papel muy importante, pues la ausencia de esta hipótesis lleva a un resultado contrario al obtenido con su presencia.

43. Al analizar 40 muestras de una aleación de bajo punto de fusión de tipo “wood” se ha detectado ausencia de cadmio en 12 de ellas. Determinar un intervalo de confianza para la proporción de muestras de dicha aleación que no contienen cadmio.

SOLUCIÓN:

La proporción observada es

$$p_0 = \frac{12}{40} = 0,3.$$

A un nivel de confianza del 95% el intervalo se obtiene resolviendo la inecuación

$$\frac{|0,3 - p|}{\sqrt{\frac{p(1-p)}{40}}} < 1,96$$

Lo que equivale a

$$(0,3 - p)^2 < (1,96)^2 \frac{p(1-p)}{40}.$$

Efectuando los cálculos obtenemos $0,1807 < p < 0,4543$.

Una alternativa es dar el intervalo

$$0,3 \pm 1,96 \sqrt{\frac{0,3 \cdot 0,7}{40}} = 0,3 \pm 0,1420, \text{ en el que se utiliza la proporción observada para estimar la varianza mediante la fracción } \frac{0,3 \cdot 0,7}{40}.$$

El resultado es $0,1580 < p < 0,4420$.

44. La cantidad de azufre encontrado en plantas secas de mostaza sigue una distribución normal X. se ha observado una muestra de extensión 9 con los siguientes resultados

0,7 0,8 0,6 0,95 0,65 1 0,9 0,2 0,55.

Si aceptamos como valor de σ el valor calculado de la cuasi-desviación típica muestral \hat{S} , ¿Cuál sería el tamaño mínimo de la muestra que habría de ser considerada para que el intervalo de confianza al 95% para el nivel medio de azufre tenga una longitud inferior a 0,1?

SOLUCIÓN:

Tenemos los siguientes datos muestrales

$$n = 9, \quad \sum x_i = 6,35, \quad \sum x_i^2 = 4,9675$$

de modo que

$$\bar{x} = \frac{6,35}{9} = 0,7055 \quad \text{y} \quad s^2 = \frac{4,9675}{9} - 0,7055^2 = 0,0541$$

$$\hat{S}^2 = \frac{9}{8}s^2 = 0,0609 \quad \text{y} \quad \hat{S} = 0,2468$$

Si aceptamos que $\sigma = \hat{S} = 0,2448$, la longitud del intervalo de confianza para una muestra de extensión n y un nivel de confianza del 95% es

$$2 \cdot 1,96 \frac{\sigma}{\sqrt{n}} = 2 \cdot 1,96 \frac{0,2464}{\sqrt{n}} = \frac{0,9674}{\sqrt{n}}.$$

Resolviendo la inecuación

$$\frac{0,9674}{\sqrt{n}} < 0,1$$

Obtenemos $n > 93,59$.

45. Consideremos un grupo de cuatro amigos jugando a los chinos. El primero en hablar ha pedido 8. el segundo estima, por máxima verosimilitud, cuántas lleva el primero. ¿Cuántas debe pedir si tiene en cuenta el número de chinos esperado para los dos restantes y el hecho cierto de que él no lleva ninguna?

SOLUCIÓN:

Llamemos θ al número de chinos que lleva el primer jugador y que debemos estimar por máxima verosimilitud. Si S es el número total de chinos, veamos qué valor de θ hace máxima la probabilidad de que $S=8$, que corresponde a la apuesta del primer jugador.

Representemos mediante una terna (x_1, x_2, x_3) las apuestas de los jugadores segundo, tercero y cuarto y llamemos T a la suma $x_1+x_2+x_3$.

- Si $\theta = 0$, el valor de $S=8$ equivale a $T=8$, lo que ocurre en los siguientes casos

$(3,3,2) \quad (3,2,3) \quad (2,3,3)$

Y, por lo tanto, con probabilidad igual a $\frac{3}{4^3} = \frac{3}{64}$, ya que cada caso tiene una probabilidad igual a $\left(\frac{1}{4}\right)^3$, dado que cada jugador tiene 4 elecciones diferentes.

- Si $\theta=1$, el valor $S=8$ equivale a $T=7$, lo que ocurre en los siguientes casos

(3,3,1) (3,1,3) (1,3,3) (3,2,2) (2,3,2) (2,2,3)

Y, por lo tanto, con probabilidad igual a $\frac{6}{4^3} = \frac{6}{64}$.

- Si $\theta=2$, el valor $S=8$ equivale a $T=6$, lo que ocurre en los siguientes casos

(3,3,0) (3,0,3) (0,3,3) (3,2,1) (3,1,2) (2,3,1) (2,1,3) (1,3,2) (1,2,3) (2,2,2)

Y, por lo tanto, con probabilidad igual a $\frac{10}{4^3} = \frac{10}{64}$.

- Si $\theta=3$, el valor $S=8$ equivale a $T=5$, lo que ocurre en los siguientes casos

(3,2,0) (3,0,2) (2,3,0) (2,0,3) (0,3,2) (0,2,3) (3,1,1) (1,3,1) (1,1,3) (2,2,1)
(2,1,2) (1,2,2)

Y, por lo tanto, con probabilidad igual a $\frac{12}{4^3} = \frac{12}{64}$.

Por lo tanto, la estimación de máxima verosimilitud es $\theta=3$.

La suma de chinos más probable para los dos últimos jugadores es 3, que coincide con el valor esperado, y se obtiene en los siguientes casos

(3,0) (2,1) (1,2) (0,3).

Consecuentemente, el segundo jugador debe pedir $3+0+3=6$.

46. La pérdida de peso de un determinado producto dietético en 16 individuos después de un mes fue (en kg):

3,2	2	2,5	3,3	5	4,3	2,9	4,1
3,6	2,7	3,5	4,2	2,8	4,4	3,3	3,1

Determinar un intervalo de confianza para la varianza con nivel de confianza del 99%, si la pérdida de peso es aproximadamente normal.

SOLUCIÓN:

Un intervalo de confianza para la varianza vendrá dado por

$$\frac{(n-1)S^2}{X^2_{1-\alpha/2}} \quad ; \quad \frac{(n-1)S^2}{X^2_{\alpha/2}}$$

donde los valores $X^2_{\alpha/2}$ y $X^2_{1-\alpha/2}$ corresponden al valor de la función de distribución Chi-cuadrado con n-1 grados de libertad.

En nuestro caso como $n = 16$, y $\alpha = 0,01$ tenemos que mirarlo en las tablas de percentiles de la distribución Chi-cuadrado con 15 grados de libertad $X^2_{0,005}(15) = 4,60087$ y $X^2_{0,995}(15) = 32,80149$

Si calculamos la varianza muestral, tenemos que $S^2 = 0,636958$; y por tanto sustituyendo los valores en los intervalos de confianza para la varianza tenemos

$$\frac{15 \cdot 0,636958}{32,80149} \quad ; \quad \frac{15 \cdot 0,636958}{4,60087} = \frac{9,55437}{32,80149} \quad ; \quad \frac{9,55437}{4,60087}$$

Y un intervalo de confianza del 99% para la varianza de la pérdida de peso es

$$\sigma^2 \in (0,291279 \quad ; \quad 2,07664).$$

47. Se consideran lo siguientes tiempos de reacción de un producto químico, en segundos:

1,4 1,2 1,2 1,3 1,5 1,3 2,2 1,4 1,1

Obtener un intervalo de confianza del 90% para el tiempo de reacción. Suponer la variable normal con desviación típica poblacional conocida $\sigma = 0,4$.

SOLUCIÓN:

Para calcular un intervalo de confianza para media de una población normal de la que conocemos la varianza poblacional, tenemos que usar el estadístico

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Por tanto, necesitamos calcular la media de la muestra. Que nuestro caso es:

$$\bar{X} = \frac{1,4 + 1,2 + 1,2 + 1,3 + 1,5 + 1,3 + 2,2 + 1,4 + 1,1}{9} = 1,4.$$

Como queremos obtener un intervalo con nivel de confianza del 90%, tenemos que $\alpha = 0,1$. Así un intervalo de confianza del 90% para la media obtendrá:

$$(\bar{X} - z_{1-\alpha/2} \cdot \sigma/\sqrt{n} ; \bar{X} + z_{1-\alpha/2} \cdot \sigma/\sqrt{n})$$

En nuestro caso, si miramos en la tabla de la función de distribución de la $N(0,1)$, tenemos que $z_{1-\alpha/2} = z_{0,95} = 1,64$ es el valor normal estándar que corresponde al percentil 95.

Un intervalo de confianza del 90% para la media μ será:

$$(1,4 - 1,64 \cdot 0,4/3 ; 1,4 + 1,64 \cdot 0,4/3) = (1,1813; 1,6186)$$

48. El tiempo, en minutos, que esperan los clientes de un determinado banco hasta que son atendidos sigue distribución normal de media desconocida y desviación típica igual a 3. Los tiempos que esperaron diez clientes elegidos al azar fueron los siguientes: 1,5- 2- 2,5- 3- 1- 5- 5,5- 4,5- 3- 3. Determinar un intervalo de confianza de coeficiente de confianza 0,95, para el tiempo medio de espera.

SOLUCIÓN:

Se pide determinar el intervalo de confianza para la media de una población normal de varianza conocida, σ ;

$$[\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} , \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}]$$

siendo \bar{X} la media muestral, n el tamaño y $z_{\alpha/2}$ el valor de la abscisa de una normal $N(0,1)$ que deja a la derecha un área de probabilidad $\alpha/2$.

De los datos del enunciado se desprende que es $\bar{X} = 3,1$, $n = 10$, $\sigma = 3$ y, a partir de la tabla 3 de la distribución normal, $z_{\alpha/2} = z_{0,05/2} = z_{0,025} = 1,96$. El intervalo pedido será, por tanto,

$$[3,1 - 1,96 \cdot 3/\sqrt{10}, 3,1 + 1,96 \cdot 3/\sqrt{10}] = [1,241, 4,959].$$

49. La cotización del dólar frente a la peseta sigue una distribución normal de media y varianza desconocidas. Elegidos 9 días al azar, la cotización del dólar en esos días fue: 145,3- 146,2- 145,8- 146- 146,1- 144,5- 145,2- 147- 144,2. Determinar un intervalo de confianza, de coeficiente de confianza 0,95, para la cotización media del dólar frente a la peseta.

SOLUCIÓN:

El intervalo de confianza solicitado es el de la media de una población normal de varianza desconocida y muestras pequeñas,

$$[\bar{X} - t_{n-1;\alpha/2} \cdot S/\sqrt{n}, \bar{X} + t_{n-1;\alpha/2} \cdot S/\sqrt{n}]$$

que para los datos del problema queda igual a

$$[145,59-2,306*0,88/3, 145,59+2,306*0,88/3] = [144,914, 146,266]$$

al ser $\bar{X} = 145,59$, $S = 0,88$ y obtenerse, a partir de la tabla 5 de la t de Student el valor $t_{n-1;\alpha/2} = t_{8;0,025} = 2,306$.

50. La duración en minutos de un determinado viaje es una variable aleatoria con distribución normal de media desconocida y desviación típica igual a 3. En una muestra tomada al azar de diez realizaciones del viaje en cuestión se obtuvieron los siguientes tiempos: 10.1, 6.5, 5.5, 7.9, 8.2, 6.5, 7.0, 8.1, 6.9 y 7.7. Se pide:

- a) Realizar la estimación de máxima verosimilitud de la duración media del viaje.
- b) Calcular la probabilidad de que, en valor absoluto, la diferencia entre media estimada la real sea menor que 1 minuto.

SOLUCIÓN:

Llamando X a la variable aleatoria duración en minutos del viaje en cuestión se tiene que es $X \rightarrow N(\mu, 3)$

a) Para una muestra de tamaño n de X, la función de verosimilitud será (CB-sección 5.2 o EII-sección 2.2)

$$L(\mu) = f_{\mu}(\chi_1, \dots, \chi_n) = \prod_{i=1}^n f_{\mu}(\chi_i) = \frac{1}{(18\pi)^{\frac{n}{2}}} \exp. \left\{ -\frac{1}{18} \sum_{i=1}^n (\chi_i - \mu)^2 \right\}$$

La cual tiene por logaritmo neperiano

$$\text{Log } L(\mu) = -\frac{n}{2} \log(18\pi) - \frac{1}{18} \sum_{i=1}^n 2(\chi_i - \mu)^2 = 0$$

La derivada de esta solución $\hat{\mu} = \bar{\chi}$; por tanto, el estimador de máxima verosimilitud de μ será la media muestral. La estimación de máxima verosimilitud será, por tanto, $\bar{\chi} = 7.44$.

b) La probabilidad pedida será

$$P \{ |\bar{\chi} - \mu| < 1 \}$$

De donde, como la distribución de la medida muestral para una muestra procedente de una población normal de varianza conocida es (CB-sección 5.4 o EII-sección 2.4) $\bar{\chi} \rightarrow N(\mu, 3/\sqrt{10/3})$, tipificando será

$$P \{ |\bar{\chi} - \mu| < 1 \} = P \{ |Z| < \sqrt{10/3} \} = P \{ |Z| < 1.054 \} = 0.70804$$

En donde $Z \rightarrow N(0, 1)$ y donde la última probabilidad se ha calculado con ayuda de interpolación: De la tabla 3 de la normal $N(0, 1)$ obtenemos que es $P \{ Z \geq 1.05 \} = 0.1469$ y $P \{ Z \geq 1.06 \} = 0.1446$, lo que nos dice que a la derecha de 1.05 y bajo la curva de densidad de la $N(0, 1)$ hay un área de probabilidad 0.1469 y a la derecha de 1.06 un área de probabilidad 0.1446. Es decir, que para un incremento de abscisa de 0.01 hay una disminución de probabilidad de $0.1469 - 0.1446 = 0.0023$. Para un incremento de abscisa de 0.004 – es decir, a la derecha de 1.054- habrá (mediante una regla de tres) una disminución de probabilidad de $0.0023 \times 0.004/0.01 = 0.00092$. Es decir,

$$P \{ Z \geq 1.054 \} = 0.1469 - 0.00092 = 0.14598$$

Por tanto,

$$\begin{aligned} P \{ |Z| < 1.054 \} &= 1 - P \{ |Z| \geq 1.054 \} \\ &= 1 - P \{ |Z| \geq 1.054 \} - P \{ |Z| \geq 1.054 \} \\ &= 1 - 2 \cdot P \{ |Z| \geq 1.054 \} \\ &= 1 - 2 \cdot 0.14598 = 0.70804 \end{aligned}$$

Por la simetría de la densidad $N(0, 1)$.

51. El tiempo de vida de una determinada especie animal sigue una distribución exponencial $\text{Exp}(\Theta)$ La cual tiene por función de densidad con $X > 0$ y siendo $\Theta > 0$ un parámetro desconocido. Con objeto de estimar el parámetro Θ y, en consecuencia, la ley de probabilidad que rige su tiempo de vida, se tomo al azar una muestra aleatoria de diez animales de la especie en estudio, obteniéndose los siguientes tiempos de vida en días: 1456, 900, 1450, 650, 666, 943, 790, 840, 790, 840 y 1500. Determinar la estimación de máxima verosimilitud del parámetro Θ .

SOLUCIÓN:

La variable aleatoria en estudio, X , es tiempo de vida de la especie. Para una muestra aleatoria simple de tamaño n de X la función de verosimilitud es (CB-sección 5.2 o EII-sección 2.2)

$$L(\Theta) = f_{\Theta}(\chi_1, \dots, \chi_n) = \prod_{i=1}^n f_{\Theta}(\chi_i) = \Theta^n \exp. \left\{ -\Theta \sum_{i=1}^n X_i \right\}$$

La cual tiene por logaritmo neperiano $\text{Log } L(\Theta) = n \log \Theta - \Theta \sum_{i=1}^n X_i$

Cuya derivada respecto a Θ igualada a cero es

$$\frac{\partial}{\partial \Theta} \log L(\Theta) = \frac{n}{\Theta} - \sum_{i=1}^n X_i = 0$$

Obteniéndose de la ultima igualdad el estimador de máxima verosimilitud

$$\hat{\Theta} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

De los datos se obtiene una media muestral $\bar{X} = 1006.1$, por lo que la estimación de máxima verosimilitud del parámetro Θ será

$$\hat{\Theta} = \frac{1}{\bar{X}} = 0.0009939.$$

52. Los saltamontes de la región africana de Asyut se caracterizan por tener una longitud media de 2cm pudiendo admitirse una distribución normal para la longitud de tales ortópteros. Elegida una muestra aleatoria de 20 de ellos, sus longitudes es cm fueron las siguientes:

1.90, 1.85, 2.01, 1.95, 2.05, 2.00, 1.97, 2.02, 1.89, 2.01

2.05, 1.95, 1.87, 2.05, 1.97, 1.85, 2.02, 1.95, 1.98, 2.05

Utilizando estos datos, se pide:

a) Determinar la estimación de máxima verosimilitud de la desviación típica poblacional σ .

b) Calcular la probabilidad de que el estimador de máxima verosimilitud de σ subestime el verdadero valor de dicho parámetro

SOLUCIÓN:

Si llamamos X a la longitud de los saltamontes es estudio, del enunciado se desprende que puede admitirse para X una distribución $N(2, \sigma)$, siendo la estimación de σ el objeto del problema

a) Para determinar el estimador de máxima verosimilitud de σ comenzaremos calculando la función de verosimilitud (CB-sección 5.2 o EII-sección 2.2)

$$L(\sigma) = f_{\sigma}(x_1, \dots, x_n) = \prod_{i=1}^n f_{\sigma}(x_i) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp. \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - 2)^2 \right\}.$$

Su derivada igualada a cero (ecuación de verosimilitud) será

$$\frac{\partial}{\partial \sigma} \text{Log } L(\sigma) = -\frac{n}{\sigma} + \frac{4\sigma \sum_{i=1}^n (x_i - 2)^2}{4\sigma^4} = 0$$

De donde se obtiene el estimador de máxima verosimilitud.

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - 2)^2}.$$

(Hacemos aquí la observación de que los estimadores de máxima verosimilitud respetan las transformaciones biyectivas; es decir, que si T es el estimador de máxima verosimilitud para Θ , entonces g(T) es el estimador de máxima verosimilitud para g(Θ) siempre que g sea una función biyectiva. En este sentido, podía haberse determinado el estimador de máxima verosimilitud para la varianza poblacional σ^2 - mas habitual y ya determinado en CB-sección 5.2 o EII-sección 2.2- y extrayendo la raíz cuadrada obtendríamos el da la desviación típica σ , al ser la función g(x) = \sqrt{x} una función biyectiva). A partir de los datos del enunciado obtenemos que la estimación de máxima verosimilitud será,

$$\hat{\sigma} = \sqrt{\frac{1}{20} \sum_{i=1}^{20} (X_i - 2)^2} = \sqrt{0.00546} = 0.0739$$

b) La probabilidad pedida será

$$P\{\hat{\sigma} < \sigma\} = P\left\{\frac{1}{20} \sum_{i=1}^{20} (X_i - 2)^2 < \sigma^2\right\} = P\left\{\frac{1}{\sigma^2} \sum_{i=1}^{20} (X_i - 2)^2 < 20\right\} = P\{\chi_{20}^2 < 20\} = 0.5292$$

Al tener $\sum_{i=1}^{20} (X_i - 2)^2 / \sigma^2$ una distribución χ_{20}^2 por estar ante un caso estimación de la varianza de una población normal de media conocida (CB-sección 5.6 o EII-sección 2.6), y en donde la ultima probabilidad se ha calculado por interpolación a partir de la tabla 4: De ella se obtiene que es $P\{\chi_{20}^2 < 16.27\} = 0.3$ y $P\{\chi_{20}^2 < 22.78\} = 0.7$, por lo que para un aumento de abscisas de $22.78 - 16.27 = 6.51$ se obtiene un aumento de probabilidad de 0.4; para un aumento de abscisa de $20 - 16.27 = 3.73$ se obtendrá un aumento de probabilidad de $0.4 \times 3.73/6.51 = 0.2292$. Por tanto, $P\{\chi_{20}^2 < 20\} = 0.3 + 0.2292 = 0.5292$.

53. El tiempo de vida en días X de los individuos de una población afectados de una nueva enfermedad es una variable aleatoria continua con función de densidad

$$f_{\phi}(x) = 2\phi^2 x^{-3} \quad \text{si } x > 0$$

Y $f_{\phi}(x) = 0$ si $x \leq \phi$ siendo $\phi > 0$ un parámetro desconocido.

Con objeto de estimar el parámetro ϕ , se extrajo una muestra aleatoria simple de dicha población, obteniéndose los siguientes tiempos de vida, en días, de los 10 individuos seleccionados, todos los cuales fallecieron por la enfermedad en estudio.

398, 356, 615, 265, 650, 325, 400, 223, 368, 680

Determinar la estimación de la máxima verosimilitud de ϕ .

SOLUCIÓN:

La función de densidad de la variable aleatoria en estudio X , tiempo de vida de los individuos de la población afectados por la enfermedad en estudio, nos indica que dichos individuos contraen la enfermedad de un momento desconocido, ϕ , (puesto que en este punto la función de distribución $F(x)$ empieza a crecer desde cero, o lo que es lo mismo la función de supervivencia $S(x) = 1 - F(x)$ vale 1, lo que quiere decir que todos los individuos están vivos), momento a partir del cual, y por la forma de dicha función de densidad, la probabilidad de sobrevivir va disminuyendo.

Es precisamente el inicio de las enfermedades objeto de la estimación. Para ello, utilizaremos el método de la máxima verosimilitud (CB-sección 5.2 o EII-sección 2.2). La función de verosimilitud de la muestra será:

$$L(\phi) = f_{\sigma}(\chi_1, \dots, \chi_n) = \prod_{i=1}^n f_{\sigma}(\chi_i) = 2^n \phi^{2n} \left(\prod_{i=1}^n x_i \right)$$

Si $\chi_1, \dots, \chi_n > \phi$

Como siempre, el método de la máxima verosimilitud se basa en asignar a ϕ el valor que maximice la función $L(\phi)$; el problema es que ahora ϕ aparece en el recorrido de la variable, es decir, que $L(\chi_i)$ toma un valor distinto de cero si $\phi < \chi_1, \dots, \chi_n$, y si algún χ_i , es tal que $\chi_i \leq \phi$ será $L(\phi) = 0$. En la estimación de ϕ habrá que tener también en cuenta, por tanto, el recorrido de $L(\phi)$

La función $L(\phi) = 2^n \phi^{2n} \prod_{i=1}^n x_i^{-3}$ crece al crecer ϕ , por lo que será tanto mayor cuanto mayor sea ϕ , y esto hasta que ϕ llegue al primer χ_i , a partir de donde, por los comentarios anteriores, $L(\phi)$ vale cero. Por tanto, el valor de ϕ que hace máxima $L(\phi)$ es el mínimo de los n valores $\{\chi_1, \dots, \chi_n\}$ el cual se suele denotar por $x_{(1)}$.

La estimación de máxima verosimilitud de ϕ será, a partir de los 10 datos de muestra, $x_{(1)}=223$.

Este problema es un ejemplo de lo importante que es la determinación del estimador de máxima verosimilitud es calcular el valor de ϕ que maximiza la función $L(\phi)$, máximo que en muchas ocasiones se podrá determinar derivando $L(\phi)$ respecto a ϕ e igualando a cero dicha derivada, pero que en otras ocasiones, como pasa en general al determinar el máximo de cualquier función, deberá utilizarse otras herramientas distintas de la derivada.

54. Una muestra de tamaño 10 de una población de mujeres presenta una altura media de 172 cm. y una muestra de 12 varones de otra población presenta una altura media de 176,7 cm. Sabiendo que ambas poblaciones son normales con varianzas 225 y 256 respectivamente, se trata de analizar si con una probabilidad del 95% se puede asegurar que los varones son más altos en media que las mujeres o viceversa.

SOLUCIÓN:

Para resolver este problema hallaremos un intervalo de confianza para la diferencia de medias al 95% y comprobaremos si dicho intervalo contiene el valor cero, en cuyo caso se puede aceptar la hipótesis de que las alturas medias son iguales con una probabilidad del 95%. Estamos entonces ante el caso del cálculo de intervalos de confianza para la diferencia de medias de dos distribuciones normales con varianzas conocidas. En esta situación el intervalo de confianza para la diferencia de medias se basa en el siguiente estadístico:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

lo que nos lleva al intervalo de confianza para la varianza definido por:

$$\bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

El intervalo de confianza será entonces:

$$170,2 - 176,7 \pm 1,96 \sqrt{\frac{225^2}{10} + \frac{256^2}{12}} = [-207,56 \quad 194,56]$$

Como este intervalo contiene el valor cero, se puede aceptar con una probabilidad del 95% que $\bar{x}_1 - \bar{x}_2 = 0$, es decir, $\bar{x}_1 = \bar{x}_2$. Por tanto, aceptaremos la hipótesis de que la estatura

de los varones y las mujeres de ambas poblaciones es similar con un coeficiente de confianza del 95%.

55. Los responsables municipales de la salud miden la radiactividad en el agua de una fuente natural en una zona abundante en granito. Realizadas 12 mediciones en diferentes fechas del año se observó una media de 3,6 picocurios con una desviación típica de 0,82. determinar, al 95% y al 99%, intervalos de confianza para la radiación media y para la varianza.

SOLUCIÓN:

Tenemos los siguientes datos muestrales

$$N=12, \quad \bar{x} = 3,6 \quad y \quad s = 0,82.$$

Para determinar el intervalo de confianza para σ^2 tenemos en cuenta que

$$\frac{12S^2}{\sigma^2}$$

Es un variable χ^2 con 11 grados de libertad. Los percentiles necesarios para la determinación de los intervalos al 95% y 99% de confianza son

$$\chi_{11;0,005}^2 = 2,603 \quad \chi_{11;0,025}^2 = 3,816 \quad \chi_{11;0,975}^2 = 21,92 \quad y \quad \chi_{11;0,995}^2 = 26,8.$$

Así el intervalo de confianza al 95% es

$$\left(\frac{12 \cdot 0,82^2}{21,92}, \frac{12 \cdot 0,82^2}{3,816} \right)$$

Y al 99%

$$\left(\frac{12 \cdot 0,82^2}{26,8}, \frac{12 \cdot 0,82^2}{2,603} \right)$$

Efectuando los cálculos se obtiene, respectivamente,

$$(0,3681; 2,1145) \quad y \quad (0,3011; 3,0998).$$

En lo relativo a la media, el intervalo se determina teniendo en cuenta que

$$\frac{\bar{x} - \mu}{S / \sqrt{11}}$$

Es una variable t de Student con 11 grados de libertad. El intervalo, para un nivel de confianza $p=1-\alpha$, viene dado por

$$\bar{x} \pm t_{11; 1-\alpha/2} \frac{s}{\sqrt{11}}$$

Siendo $t_{11; 1-\alpha/2}$ el valor del percentil $100 \cdot (1-\alpha/2)$ para 11 grados de libertad. Concretamente

$$t_{11; 0,975}=2,201 \quad \text{y} \quad t_{11; 0,995}=3,1058$$

y los intervalos de confianza son, respectivamente,

$$3,6 \pm 2,201 \frac{0,82}{\sqrt{11}} \quad \text{y} \quad 3,6 \pm 3,1058 \frac{0,82}{\sqrt{11}},$$

$$\text{Obteniéndose } 3,6 \pm 0,544 \quad \text{y} \quad 3,6 \pm 0,768.$$

56. Para m.a.s. de tamaño n, de una población con función de densidad

$$f(x|\theta) = \frac{5x^4}{\theta^5} I(0, \theta](x) \quad \text{con} \quad \theta > 0$$

a) Hállese el estimador de máxima verosimilitud T.

b) Determínese la región de confianza de grado 0,95 para θ , de la forma $(\lambda T, +\infty)$ para λ conveniente.

SOLUCIÓN:

a) La función de densidad de la muestra es

$$f(x_1, \dots, x_n | \theta) = 5^n \frac{\prod_{i=1}^n x_i^4}{\theta^{5n}} I_{[x_{(n)}, +\infty)}(\theta)$$

por lo que el EMV es $\hat{\theta} = X_{(n)}$.

a) La región de confianza, de la forma $(\lambda X_{(n)}, +\infty)$ estará de terminada, cuando se obtenga el valor de λ tal que

$$P\{\lambda X_{(n)} \leq \theta | \theta\} = 1 - \alpha \quad \text{para todo } \theta > 0$$

para lo que ha de cumplirse que

$$P\{X_{(n)} \leq y | \theta\} = P^n\{X \leq y | \theta\} = \left(\frac{y}{\theta}\right)^{5n}$$

De donde ha de ser $\lambda^{5n} = 1 - \alpha$; o lo que es lo mismo $\lambda = (1 - \alpha)^{\frac{1}{5n}}$ y

$$RC_{1-\alpha}(\theta) = \left((1 - \alpha)^{\frac{1}{5n}} X_{(n)}, +\infty \right).$$

57. Determínese, con m.a.s de tamaño n, el EMV para θ en los siguientes casos:

a) $f(x | \theta) = \frac{1}{\theta} I_{\{1,2,\dots,\theta\}}(x)$ con θ entero positivo

b) $f(x | \theta) = e^{-x+\theta} I_{[\theta,+\infty)}(x)$ con $\theta \in \mathbb{R}$

SOLUCIÓN:

a) La fusión de verosimilitud es

$$L(\theta | x_1, \dots, x_n) = \frac{1}{\theta^n} \prod_{i=1}^n I_{\{1,2,\dots,\theta\}}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^n I_{Z}(x_i) = I_{\{x_{(n)}, x_{(n)}+1, \dots\}}(\theta)$$

a medida que θ aumenta uno a uno desde $x_{(n)}$ esta función disminuye,

luego es $\hat{\theta} = X_{(n)}$

b) La fusión de verosimilitud es

$$L(\theta | x_1, \dots, x_n) = e^{-\sum_{i=1}^n x_i + n\theta} \prod_{i=1}^n I_{[\theta,+\infty)}(x_i) = e^{-\sum_{i=1}^n x_i + n\theta} I_{(-\infty, x_{(1)})}(\theta)$$

expresión que, como función de θ , es la máxima cuando θ es lo mayor posible, es decir es $\hat{\theta} = X_{(1)}$

58. En una muestra de 65 sujetos las puntuaciones en una escala de extroversión tienen una media de 32,7 puntos y una desviación típica de 12,64.

a) Calcule a partir de estos datos el correspondiente intervalo de confianza, a un nivel del 90%, para la media de la población.

b) Indique, con un nivel de confianza del 95%, cual sería el máximo error que podríamos cometer al tomar como media de la población el valor obtenido en la estimación puntual.

SOLUCIÓN:

a) Buscando en las tablas de la t de Student obtenemos que el valor que deja por debajo una probabilidad del 95% es 1,671 (aproximadamente). Sustituyendo los valores de esta muestra en la expresión del intervalo de confianza obtenemos:

$$(32,7 - 1,671 \cdot 12,64 / 8, 32,7 + 1,671 \cdot 12,64 / 8)$$

Operando

$$(30,06, 35,34)$$

b) En las tablas de la t de Student encontramos que el valor de la variable que deja por debajo una probabilidad de 0,975 es 2. En consecuencia a un nivel de confianza del 95% la media de la población puede valer

$$32,7 \pm 2 \cdot 12,64 / 8$$

luego el máximo error que se puede cometer, a este nivel de confianza, es: 3,16

59. Los tiempos de reacción, en mili segundos, de 17 sujetos frente a una matriz de 15 estímulos fueron los siguientes: 448, 460, 514, 488, 592, 490, 507, 513, 492, 534, 523, 452, 464, 562, 584, 507, 461.

Suponiendo que el tiempo de reacción se distribuye Normalmente, determine un intervalo de confianza para la media a un nivel de confianza del 95%.

SOLUCIÓN:

Mediante los cálculos básicos obtenemos que la media muestral vale 505,35 y la desviación típica 42,54.

Buscando en las tablas de la t de Student con 16 grados de libertad, obtenemos que el valor que deja por debajo una probabilidad de 0,975 es 2,12

Sustituyendo estos valores en la expresión del intervalo de confianza de la media tenemos:

$$(505,35 - 2,12 \cdot 42,54 / 4, 505,35 + 2,12 \cdot 42,54 / 4)$$

Operando

$$(482,80, 527,90)$$

60. Obtener el E.M.V. del parámetro λ de una distribución de Poisson.

SOLUCIÓN:

Para una muestra de tamaño n tendremos que la función de verosimilitud será:

$$L(x_1, x_2, \dots, x_n / \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \frac{e^{-\lambda} \lambda^{x_2}}{x_2!} \dots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!}$$

maximizar L será equivalente a minimizar el numerador de L , llamémosle L'

$$L' = e^{-n\lambda} \lambda^{\sum x_i}$$

maximizar L' es equivalente a maximizar su logaritmo $\ln L'$

$$\ln L' = -n\lambda + \sum x_i \ln \lambda$$

$$\frac{d \ln L'}{d\lambda} = -n + \sum x_i \frac{1}{\lambda} = 0$$

lo que se cumple para $\lambda = \frac{\sum x_i}{n} = \bar{x}$

Por lo tanto el estimador maximo-verosimil de λ es $\hat{\lambda} = \bar{x}$

61. De una población cuya distribución se desconoce se obtiene una muestra aleatoria de 2000 valores en que la media muestral resulta ser 225 y la desviación típica muestral 10. Suponiendo que la varianza muestral coincida con la de la población, estimar un intervalo para la media de la población con un nivel de confianza del 95%

SOLUCIÓN:

Según hemos visto

$$\mu \in \left[\bar{x} - \frac{\sigma}{\sqrt{n\sigma}}, \bar{x} + \frac{\sigma}{\sqrt{n\sigma}} \right] \text{ con una confianza de } 1 - \sigma$$

En este caso:

$$1 - \sigma = 0,95$$

$$\sigma = 0,05$$

$$\bar{x} = 225$$

$$s = 10$$

$$n = 2000$$

Sustituyendo $\mu \in [224, 226]$ con el 95% de confianza

62. La longitud, en centímetros, de las piezas fabricadas por una cierta máquina se distribuye según una $N(10, 0.25)$. Para muestras de tamaño 25, calcular:

a) $P(9.68 \leq \bar{X} \leq 10.1)$

b) $P(S^2 \leq 0.19)$

SOLUCIÓN:

a) La media muestral de muestras de tamaño n procedentes de una población $N(\mu, \sigma^2)$ se distribuye según una $N(\mu, \sigma^2/n)$. En nuestro caso, la media muestral lo hará según una ley $N(10, 0.01)$. Por tanto,

$$P(9.68 \leq \bar{X} \leq 10.1) = P(-3.2 \leq Z \leq 1) = 0.8406$$

Siendo una normal típica.

b) El teorema de Fisher afirma que

$$\frac{ns^2}{\sigma^2}$$

Se distribuye según una ley chi-cuadrado con $n-1$ grados de libertad. Por tanto, se tiene que

$$P(S^2 \leq \bar{X} \leq 0.19) = P\left(\frac{25 S^2}{0.25} \leq \frac{25 * 0.19}{0.25}\right) = P(U \leq 19) = 0.25$$

Donde U se distribuye χ^2_{24}

63. De una población normal de media y desviación típica desconocida se ha obtenido una muestra de 26 elementos que tiene como media aritmética 5 y desviación típica 1.2. ¿Cuál es la probabilidad de que la media poblacional sea superior a 5.3?

SOLUCIÓN:

Sabemos que la variable

$$Z = \frac{\bar{X} - \mu}{S} \sqrt{n}$$

Se distribuye según una ley t de Student con n-1 grados de libertad, y lo que se pide es

$$P\left(Z < \frac{5 - 5.3}{1.2} \sqrt{25}\right) = P(Z < -1.25) = 0.114$$

Donde Z sigue una t de Student con 25 grados de libertad.

64. De una población normal de varianza c^2 se extraen dos muestras de tamaño n. Hallar cuál debe ser el tamaño de ambas muestras, para que la probabilidad de que las medias muestrales difieran en más de $2c$, sea 0.05.

SOLUCIÓN:

La distribución de las dos medias muestrales es, suponiendo que μ es la media de la población, $N(\mu, c^2/n)$.

$$\bar{X} - \bar{Y} : N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

Como en este caso ambas muestras provienen de la misma población, la distribución de Z es:

$$Z = \bar{X} - \bar{Y} : N\left(0, \frac{2\sigma^2}{n}\right)$$

Y la probabilidad del suceso

$$|Z| = \left| \bar{X} - \bar{Y} \right| > 2c$$

Será

$$P(|Z| > 2c) = 1 - P(|Z| < 2c) = 1 - P(-2c < Z < 2c) = 0.05$$

O bien

$$P(-2c < Z < 2c) = 0.95 \rightarrow P(0 < Z < 2c) = 0.475 \rightarrow P(Z < 2c) = 0.975$$

Y pasando a una $N(0,1)$, Y , tendremos

$$P(Z < 2c) = P\left(Y < \frac{2c}{\sqrt{2/n}}\right) = P(Y < \sqrt{2n}) = 0.975$$

De donde la longitud, $\sqrt{2n} = 1.95$ esto es, $n=2$.

65. En una muestra de 100 personas de un barrio de Pamplona se ha observado una proporción de 0.18 personas que leen el periódico diariamente. ¿Puede ser que la verdadera proporción de personas que leen el periódico en ese barrio sea 0.20?

SOLUCIÓN:

$$\sigma_p = \sqrt{\frac{p_{pob} * q_{pob}}{N}} = \sqrt{\frac{0.20 * 0.8}{100}} = 0.04 ;$$

$$Z = \frac{p_{pob} * q_{pob}}{N} = \frac{0.18 - 0.20}{0.04} = -0.5 ;$$

$$Z = -0.5 > -1.645 \quad \text{ó} \quad |0.5| < |1.645|$$

Por consiguiente no puede rechazarse la hipótesis nula. Es decir, nada se opone a aceptar que la muestra procede de tal barrio donde un 20% de personas leen el periódico diariamente.

Con frecuencia p_{pob} es desconocida y, por lo tanto, también q_{pob} es desconocida. El error típico σ_p se calcula entonces, salvo excepciones, empleando la proporción p_m de la muestra y $q_m = 1 - p_m$;

$$\sigma_p = \sqrt{\frac{p_m * q_m}{N}}$$

Si en lugar de proporciones se toma la frecuencia absoluta el error típico de la frecuencia absoluta será: $\sigma_f = \sqrt{N * p_{pob} * q_{pob}}$ ya que $f = N * p$ y que $\sigma_f = N * \sigma_p$

66. Consideramos dos conjuntos de puntuaciones en un test de intuición sociológica, ambos al mismo grupo de 20 estudiantes. Llamamos \bar{X}_1 a la media de puntuaciones

obtenidas antes de que comenzase el curso y \bar{X}_2 a la media obtenida después de concluir el curso. Interesa determinar si \bar{X}_2 es significativamente diferente de \bar{X}_1 .

SOLUCIÓN:

Los datos son los siguientes:

Antes del curso	Después del curso	Diferencia
18	20	2
16	22	6
18	24	6
12	10	-2
20	25	5
17	19	2
18	20	2
20	21	1
22	23	1
20	20	0
10	10	0
8	12	4
20	22	2
12	14	2
16	12	-4
16	20	4
18	22	4
20	24	4
18	23	5
21	17	-4

$$\bar{X}_1 = 16.5 \quad \bar{X}_2 = 18.5 \quad \bar{D} = \frac{40}{20} = 2 \quad \sum (D - \bar{D})^2 = 164$$

Para resolver el problema partimos de la hipótesis nula: no hay diferencia significativa entre las dos medias; es decir, puede ser cero por azar.

1. Cálculo de la desviación típica de las diferencias:

$$S_D = \sqrt{\frac{\sum (D - \bar{D})^2}{N}} = \sqrt{\frac{164}{20}} = 2.863$$

2. Cálculo del error típico de la media de las diferencias:

$$\sigma_D = \frac{S_D}{\sqrt{N-1}} = \frac{2.863}{\sqrt{19}} = \frac{2.863}{4.3588} = 0.657$$

3. Hallamos la razón crítica t:

$$t = \frac{|\bar{D}|}{\sigma_D} = \frac{2}{0.657} = 3.044$$

4. Suponemos que estamos trabajando al nivel de confianza 1%. A ese nivel, buscamos en las tablas de distribución t de Student, la t correspondiente para $N - 1 = 20 - 1 = 19$ grados de libertad: $t = 2.54$ (se trata de prueba unilateral).

5. Como la t hallada $= 3.044 > 2.54$ rechazamos la hipótesis nula con 99% de probabilidades de acertar o lo que es lo mismo, a lo más con 1% de riesgo de equivocarnos; podemos decir que tal diferencia es significativa y, por lo tanto, que la media después de finalizar el curso es significativamente mayor que la media antes de comenzar el curso, al nivel de confianza 1%.

67. En una «reunión» una droga fue tomada por 14 personas, de las cuales 6 lo hacen por primera vez y 8 ya son habituales de ella. La droga produjo en el primer grupo sueños de duración 11,12,13,16,17 y 15 horas, mientras que en el segundo grupo 8,7,9,10,6,7,9 y 8 horas. Se pide:

- a) **Media y desviación típica de cada grupo**
- b) **Formar el estadístico que se distribuye según una t de Student de 12 grados de libertad, sabiendo que las poblaciones tienen la misma media y desviación típica.**

SOLUCIÓN:

- a) Grupo 1

$$\bar{x}_1 = \frac{11+12+13+16+17+15}{6} = 14$$

$$s_1^2 = \frac{(11-14)^2 + (12-14)^2 + (13-14)^2 + (16-14)^2 + (17-14)^2 + (15-14)^2}{6} = 4.67$$

Grupo 2

$$\bar{x}_2 = \frac{8+7+9+10+6+7+9+8}{8} = 8$$

$$s_2^2 = \frac{(8-8)^2 + (7-8)^2 + (9-8)^2 + (10-8)^2 + (6-8)^2 + (7-8)^2 + (9-8)^2 + (8-8)^2}{8} = 1.5$$

$$b) t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} * \frac{\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{14 - 8}{\sqrt{6 * 4.67 + 8 * 1.5}} * \frac{\sqrt{6 + 8 - 2}}{\sqrt{\frac{1}{6} + \frac{1}{8}}} = 6.084$$

68. Según una encuesta realizada sobre una muestra de 2500 personas elegidas al azar, el 80% está decidida a votar en las últimas elecciones.

a) ¿Puede ser cierto que llegue a votar el 85% de la población?

b) Con un 99% de nivel de confianza ente qué valores estará el porcentaje de los votantes de la población.

SOLUCIÓN:

a) Se trata de una comprobación de hipótesis.

$$\sigma_p = \sqrt{\frac{0.85 * 0.15}{2500}} = 0.0071 \left(\begin{array}{l} p_{pob} = 0.85 \\ q_{pob} = 1 - p_{pob} = 0.15 \end{array} \right)$$

Establecemos la hipótesis nula, la cual dice que puede ser cierto que vote el 85% de la población.

Hallamos la razón crítica

$$Z = \frac{0.85 * 8.15}{0.0071} = -7.04$$

Una $Z = |7.04| > 2.33$. Luego, de ser cierto que en la población votase el 85%, en una muestra de 2500 personas no podía haber salido un porcentaje de 80%. Por ello, rechazamos la hipótesis nula y afirmamos con 99% de probabilidades de acertar que no llegará a votar el 85% en circunstancias normales (= circunstancias similares al momento en que se hubiera realizado el muestreo).

Es decir, la probabilidad de que la suma de las rentas de 100 personas seleccionadas al azar supere los 725 millones de pesetas es tan sólo del 7,49%.

70. En una asignatura del colegio la probabilidad de que te saquen a la pizarra en cada clase es del 10%. A lo largo del año tienes 100 clases de esa asignatura. ¿Cuál es la probabilidad de tener que salir a la pizarra más de 15 veces?

SOLUCIÓN:

Salir a la pizarra es una variable independiente que sigue el modelo de distribución de Bernouilli:

"Salir a la pizarra", le damos el valor 1 y tiene una probabilidad del 0,10

"No salir a la pizarra", le damos el valor 0 y tiene una probabilidad del 0,9

La media y la varianza de cada variable independiente es:

$$\mu = 0,10$$

$$\sigma^2 = 0,10 * 0,90 = 0,09$$

Por tanto, la suma de las 100 variables se distribuye según una normal cuya media y varianza son:

$$\text{Media: } n * \mu = 100 * 0,10 = 10$$

$$\text{Varianza: } n * \sigma^2 = 100 * 0,09 = 9$$

Para calcular la probabilidad de salir a la pizarra más de 15 veces, calculamos el valor equivalente de la variable normal tipificada:

$$Y = \frac{15-10}{3} = 1,67$$

Luego:

$$P(X > 15) = P(Y > 1,67) = 1 - P(Y < 1,67) = 1 - 0,9525 = 0,0475$$

Es decir, la probabilidad de tener que salir más de 15 veces a la pizarra a lo largo del curso es tan sólo del 4,75%.

71. Suponga que de una población consistente en los valores 0, 2, 4, 6 y 8, se toman muestras de tamaño 2 con reemplazo.

X	Frecuencia	Frecuencia Relativa
0	1	1/5 = 0.2
2	1	1/5 = 0.2
4	1	1/5 = 0.2
6	1	1/5 = 0.2
8	1	1/5 = 0.2

Demostrar que es razonable aproximar la distribución muestral de \bar{x} por una distribución normal, una vez que se conoce la media y la desviación estándar de la distribución muestral.

SOLUCIÓN:

Se calcula la media poblacional, la varianza y desviación estándar poblacional

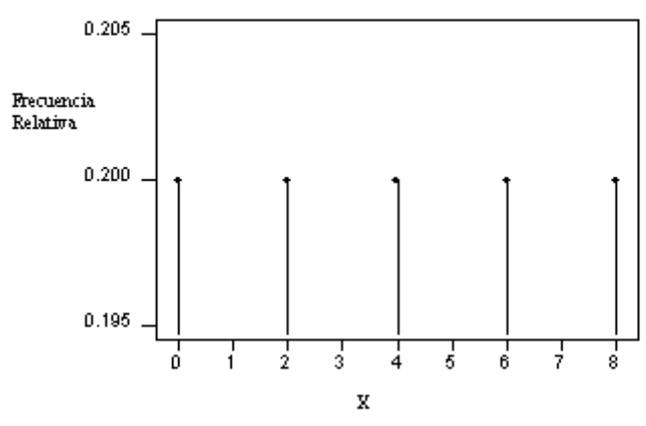
$$\mu = \frac{\sum_i^n X}{N} = \frac{0 + 2 + 4 + 6 + 8}{5} = \frac{20}{5} = 4$$

$$\sigma^2 = \sum_i^n \frac{(X - \mu)^2}{N} = \frac{(0 - 4)^2 + (2 - 4)^2 + (4 - 4)^2 + (6 - 4)^2 + (8 - 4)^2}{5} = \frac{40}{5} = 8$$

$$\sigma = \sqrt{8} = 2,83$$

A continuación se calcula la gráfica de la distribución de frecuencia para la población

Gráfica de la Población



Esta gráfica no puede considerarse acampanada o normal.

A posteriori se toman muestras de tamaño dos con reemplazo.

Muestra	\bar{X}	Muestra	\bar{X}	Muestra	\bar{X}
0, 0	0	4, 0	2	8, 0	4
0, 2	1	4, 2	3	8, 2	5
0, 4	2	4, 4	4	8, 4	6
0, 6	3	4, 6	5	8, 6	7
0, 8	4	4, 8	6	8, 8	8
2, 0	1	6, 0	3		
2, 2	2	6, 2	4		
2, 4	3	6, 4	5		
2, 6	4	6, 6	6		
2, 8	5	6, 8	7		

Se agrupa a las medias muestrales en la tabla de frecuencia siguiente:

X	0	1	2	3	4	5	6	7	8
F	1	2	3	4	5	4	3	2	1

Y se calcula la media poblacional de medias, la varianza de la medias y desviación estándar de las medias ó error estándar de las medias.

$$\mu_x = \sum_i \frac{f(X)}{N} = \frac{(1 * 0) + (2 * 1) + (3 * 2) + (4 * 3) + (5 * 4) + (4 * 5) + (3 * 6) + (2 * 7) + (1 * 8)}{25} =$$

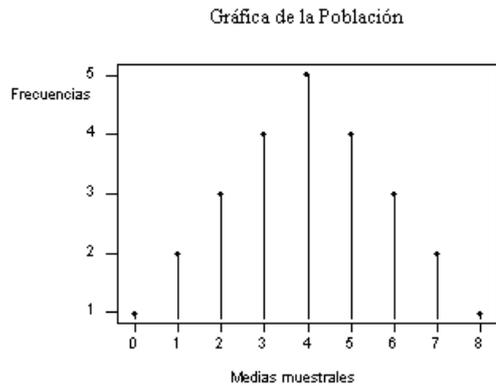
$$= \frac{100}{25} = 4$$

$$\sigma_x^2 = \frac{1(0 - 4)^2 + 2(1 - 4)^2 + 3(2 - 4)^2 + 4(3 - 4)^2 + 5(4 - 4)^2 + 4(5 - 4)^2 + 3(6 - 4)^2 + 2(7 - 4)^2 + 1(8 - 4)^2}{25} =$$

$$= \frac{100}{25} = 4$$

$$\sigma_x = \sqrt{4} = 2$$

Gráfica de la distribución de frecuencia para la población de medias maestras



De la apariencia acampanada de la distribución de las medias, concluimos que es razonable aproximar la distribución muestral de \bar{x} por una distribución normal, una vez que se conoce la media y la desviación estándar de la distribución muestral.

72. En un experimento de laboratorio se mide el tiempo de una reacción química. Se ha repetido el experimento 98 veces y se obtiene que la media de los 98 experimentos es de 5 segundos con una desviación de 0,05 segundos. ¿Cuál es la probabilidad de que la media poblacional μ difiera de la media muestral en menos de 0,01 segundos?

SOLUCIÓN:

Dado que la muestra es grande, por el teorema del límite central podemos suponer que la distribución de la media es una normal de media μ y desviación típica el error estándar. Por tanto, la probabilidad que nos preguntan, que es:

$$P(|\bar{X} - \mu| < 0,01) = P(-0,01 < \bar{X} - \mu < 0,01) = P\left\{ -\frac{0,01}{\frac{0,05}{\sqrt{98}}} < \frac{\bar{X} - \mu}{\frac{0,05}{\sqrt{98}}} < \frac{0,01}{\frac{0,05}{\sqrt{98}}} \right\} =$$

$$= P\left\{ -1,98 < \frac{\bar{X} - \mu}{\frac{0,05}{\sqrt{98}}} < 1,98 \right\}$$

Se puede aproximar por la probabilidad de una distribución normal estándar Z:

$$P(-1,98 < Z < 1,98) = 1 - 2 \cdot 0,0239 = 0,9522.$$

Por tanto, la probabilidad que nos piden es de 0,9522.

73. Se establece un control de calidad para un proceso de producción de balas. Se ha dispuesto que cuando el proceso está bajo control, el diámetro de las balas es de 1 cm., con una desviación típica de 0,003 cm. Cada hora se toman muestras de nueve balas y se miden sus diámetros. Los diámetros de media de diez muestras sucesivas, en centímetros, son:

1,0006 0,9997 0,9992 1,0012 1,0008
1,0012 1,0018 1,0016 1,0020 1,0022

Estableced cuáles son los límites de control y explicad qué podéis concluir sobre el proceso de producción en estos instantes.

SOLUCIÓN:

Observamos que la media $m = 1$ y que el error estándar es:

$$\frac{\sigma}{\sqrt{n}} = \frac{0,003}{\sqrt{10}} = 0,001$$

Por tanto, los límites de control serán 1,003 y 0,997. Observemos que absolutamente todas las medias que hemos obtenido de las sucesivas muestras están dentro del intervalo formado por los dos límites de control. Es decir, no hay ningún dato superior a 1,003 ni ningún dato inferior a 0,997. Por tanto, podemos concluir que el proceso de control ha sido correcto durante el tiempo que lo hemos analizado, y que no hemos detectado ninguna anomalía.

74. Un investigador quiere estimar la media de una población usando una muestra suficientemente grande para que la probabilidad de que la media muestral no difiera de la media poblacional en más del 25% de la desviación típica sea 0,95. Hallar el tamaño de muestra necesario.

SOLUCIÓN:

$$P(|\bar{X} - \mu| < 0,25\sigma) = 0,95 \longleftrightarrow P\left\{ \frac{-0,25\sigma}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{0,25\sigma}{\frac{\sigma}{\sqrt{n}}} \right\} = P(-0,25\sqrt{n} < Y_{N(0,1)} < 0,25\sqrt{n}) = 0,95$$

Realizando las operaciones resulta:

$$2F_{N(0,1)}(0,25\sqrt{n}) - 1 = 0,95 \Leftrightarrow F_{N(0,1)}(0,25\sqrt{n}) = 0,975 \Leftrightarrow 0,25\sqrt{n} = 1,96 = 62$$

75. Se considera una población representada por una variante ξ , de suerte que la media poblacional es igual a 25, y la varianza poblacional es igual a 240. Supuesto extraídas muestras de tamaño 100, muestreo aleatorio simple, determinar la probabilidad de que el estadístico media muestral a_x , este comprendido entre los valores 23,55 y 28,1

SOLUCIÓN:

El estadístico media muestral, a_x :

$$a_x = \frac{\sum_{i=1}^n x_i}{n}$$

Tenemos pues:

$$E[a_x] = 25$$

$$V[a_x] = \frac{240}{100} = 2,4$$

La probabilidad a obtener puede ser expresada de forma:

$$P(23,55 \leq a_x \leq 28,1)$$

Y será conocida si determinamos la distribución de probabilidad de la variante a_x . Ahora bien, como quiera que a_x viene definida como suma de variantes x_i , independientes entre sí, todas con igual distribución e igual a la distribución de probabilidad de la población, puesto que el muestreo que se considera es el aleatorio simple, y habida cuenta que el tamaño muestral es suficientemente grande, $n=100$, procede efectuar la aproximación de la distribución de a_x , mediante la distribución normal. De esta forma, pues, a_x será una variante con distribución:

$$N(25, \sqrt{2,4})$$

Y con ello:

$$P(23,55 \leq a_x \leq 28,1) = F(28,1) - F(23,55)$$

Al ser:

$$F(28,1) = P(a_x \leq 28,1) = P(\sqrt{2,4}\xi' + 25 \leq 28,1) = P(\xi' \leq \frac{28,1 - 25}{\sqrt{2,4}})$$

$$F(23,55) = P(a_x \leq 23,55) = P(\sqrt{2,4}\xi' + 25 \leq 23,55) = P(\xi' \leq \frac{23,55 - 25}{\sqrt{2,4}})$$

Donde ξ' representa una variante con distribución N (0,1); leídos en tablas, los valores:

$$P\left(\xi' \leq \frac{28,1 - 25}{\sqrt{2,4}}\right) = P(\xi' \leq 2) = 0,97725$$

$$P\left(\xi' \leq \frac{23,55 - 25}{\sqrt{2,4}}\right) = P(\xi' \leq -0,93) = 0,17619$$

Habiéndose tomado $\sqrt{2,4} = 1,55$ es por tanto:

$$P(23,55 \leq \xi \leq 28,1) = 0,97725 - 0,17619 = 0,80106$$

76. Se considera una población representada por una variante ξ , de suerte que la distribución de probabilidad de la población es normal, con desviación típica igual a 10. Determinar la probabilidad de que el estadístico media muestral supere al valor medio poblacional en, por lo menos 0,2, supuesto extraídas muestras de tamaño 100, muestreo aleatorio simple.

SOLUCIÓN:

Si convenimos en representar por μ el valor medio poblacional, el suceso cuya probabilidad vamos a obtener puede ser expresado en la forma:

$$a_x - \mu \geq 0,2$$

La diferencia $a_x - \mu$ es aleatoria, por ser a_x una variable aleatoria, y su distribución de probabilidad podemos obtenerla a través de la función característica:

$$\varphi_{a_x - \mu}(t) = E[e^{it(a_x - \mu)}] = E[e^{ita_x} * e^{-it\mu}] = e^{-it\mu} * E[e^{ita_x}] = e^{-it\mu} * \varphi_{a_x}(t)$$

Donde:

$$\varphi_{a_x}(t) = e^{-it\mu - (1/2)t^2(\sigma^2/n)}$$

La función característica sería:

$$\varphi_{a_x - \mu}(t) = e^{-it\mu} * e^{it\mu - 1/2t^2\sigma^2/n} = e^{-1/2t^2\sigma^2/n}$$

Esta expresa que la variante $(a_x - \mu)$ se comporta con arreglo a la distribución de probabilidad:

$$N\left(0; \frac{\sigma}{\sqrt{n}}\right)$$

Conocida la distribución de $(a_x - \mu)$, en nuestro caso,

$$N\left(0; \frac{10}{\sqrt{100}}\right)$$

La probabilidad a obtener será determinada en la forma:

$$P(a_x - \mu \geq 0,2) = P\left(\frac{10}{\sqrt{100}} \xi' \geq 0,2\right) = P(\xi' \geq 0,2)$$

Donde ξ' es una variante $N(0,1)$. Leída en tablas, la probabilidad:

$$P(\xi' \geq 0,2) = 1 - P(\xi' < 0,2) = 1 - 0,57926 = 0,42074$$

Resulta pues:

$$P(a_x - \mu \geq 0,2) = 0,42074$$

77. Para calcular $P\{\bar{X}_2 \leq \bar{X}_1\}$ es necesario determinar la distribución de $\bar{X}_1 - \bar{X}_2$ y para ello es preciso decidir si las varianzas poblacionales σ_1^2 y σ_2^2 son iguales o no. Habrá que razonar, por tanto, cual de las dos opciones es la más verosímil, a la vista de las observaciones.

SOLUCIÓN:

Se sabe que:

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \quad \text{tiene distribución } F_{n-1, m-1}$$

Como si F es una variable aleatoria con distribución $F_{(10,12)}$, es

$$P\{F > 2,7534\} = 0,05 \quad \text{y} \quad P\{F < 0,3433\} = 0,05$$

si las varianzas poblacionales fuesen iguales habría probabilidad 0,9 de obtener un cociente de cuasi varianzas muestrales comprendido entre 0,3433 y 2,7534. Puesto que razonar, por tanto, cual de las dos opciones es la más verosímil, a la vista de las observaciones. Se sabe que:

$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n + s_2^2/m}}$ tiene distribución t_{15} ya que se obtiene $f = 14,755$. Por tanto,

$$P\{\overline{X}_2 \leq \overline{X}_1\} = P\left\{t_{15} \geq \frac{-1,432}{\sqrt{2,734}}\right\} = P\{t_{15} \geq -0,866\} = 0,8$$

Como el tamaño de la muestra es suficientemente grande, puede utilizarse la aproximación

normal, para obtener $P\left\{\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n + s_2^2/m}} > \frac{-1,432}{\sqrt{0,523}}\right\} = P\{Z > -1,98\} = 0,97661$

Obsérvese que el tamaño de la muestra es lo suficientemente grande para poder utilizar las distribuciones poblacionales de los índices.

En cuestiones bursátiles la independencia es una suposición de dudosa veracidad. Por ello, este apartado la evita, aunque solamente la independencia entre ambos índices, ya que la independencia entre las observaciones de diferentes días resulta necesaria para obtener una muestra aleatoria simple. Ahora se trata de un caso de datos apareados, de forma que

$$P\{\overline{X}_1 - \overline{X}_2 \geq 0\} = P\left\{\frac{\overline{X}_1 - \overline{X}_2 - 1,432}{S^*/\sqrt{n}} \geq \frac{-1,432}{\sqrt{0,52}}\right\} = P\{t_{19} > -1,986\} = 0,969$$

78. Determinar las razones por las que es razonable utilizar el estadístico T_i / S_2^2 , siendo S_2^2 la cuasi-varianza de la muestra $Y_{i,y}$.

SOLUCIÓN:

$(n-1)S^2 / \sigma^2$ tiene distribución χ_{n-1}^2 ; luego

$$E[S^2] = \sigma^2 \quad \text{y} \quad V(S^2) = \frac{\sigma^4}{(n-1)^2} V((n-1)S^2 / \sigma^2) = \frac{2\sigma^2}{n}$$

Por otro lado, como $(X_i - \mu) / \sigma$ es $N(0,1)$, nT / σ^2 será χ_n^2 , por ser suma de n cuadrados de $N(0,1)$ independientes. Así pues,

$$E[T] = \sigma^2 \quad \text{y} \quad V(T) = \frac{\sigma^4}{n^2} V(nT / \sigma^2) = \frac{2\sigma^4}{n}$$

$$T_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_1)^2$$

Como $n_1 T_1 / \sigma_1^2$ tiene la distribución $\chi_{n_1}^2$ y $(n_2 - 1) S_2^2 / \sigma_2^2$ es $\chi_{n_2-1}^2$

$\frac{T_1 / \sigma_1^2}{S_2^2 / \sigma_2^2}$ tendrán distribución F_{n_1, n_2-1} . Por consiguiente, T_1 / σ_2^2 es el producto del parámetro

a estimar σ_1^2 / σ_2^2 por una variable con distribución F_{n_1, n_2-1}

79. La efectividad en días de un determinado antibiótico, sigue una distribución normal de media 14 días y desviación típica desconocida. Fue administrada a 16 enfermos, obteniéndose una cuasi desviación típica muestral de 1,4 días. Determinar la probabilidad de que la efectividad media en la muestra no supere los 3 días, que es el tiempo mínimo de efectividad requerido.

Preocupados por una posible subestimación de la varianza poblacional, que podría llevar a subestimar la probabilidad de que no se alcance la efectividad mínima, se desea determinar la probabilidad de que con una muestra de 16 enfermos se subestime la varianza en más de un 20%. Si la muestra es de 61 pacientes, esta probabilidad ¿aumenta o disminuye?

Determinar el tamaño de muestra necesario para que la probabilidad anterior sea 0,05.

SOLUCIÓN:

La probabilidad es $P\{X \leq 13\}$. Como la población es normal, pero no se conoce la varianza poblacional, habrá que utilizar la distribución t de Student.

$$P\{\bar{X} \leq 13\} = P\left\{\frac{\bar{X} - \mu}{S / \sqrt{n}} \leq \frac{13 - 14}{1,4 / \sqrt{16}}\right\} = P\{t_{15} \leq -2,857\} = 0,0063$$

Donde el valor se obtiene, por simetría, en la distribución t_{15} .

Como $(n-1) S^2 / \sigma^2$ tiene distribución χ_{15}^2 , la probabilidad de subestimación de σ^2 en más de un 20% es

$P\{s^2 < 0,8\sigma^2\} = P\{x_{15}^2 < 12\} = 0,3$, calculada a partir de las tablas por interpolación lineal. Para una muestra de 61 individuos, utilizando la aproximación de la distribución X_n^2 , es decir, la probabilidad disminuye al aumentar el tamaño de la muestra. Por último, para conseguir que:

$(0,05 = P\{z^2 < 0,8(n-1)\} \cong P\{z < \sqrt{1,6(n-1)} - \sqrt{2n-1}\})$, ha de ser $\sqrt{1,6(n-1)} - \sqrt{2n-1} = -1,645$. De donde $n = 117,61$; con tamaño muestral 118, se obtiene la precisión requerida.

80. Sea X_1, \dots, X_n , una m.a de una población $N(\mu, \sigma^2)$. Supongamos que X_{n+1} se distribuye $N(\mu, \sigma^2)$ y que X_1, \dots, X_{n+1} , son independientes. Obtener con todo detalle la distribución en el muestreo de $\frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$.

SOLUCIÓN:

En este caso se tiene que X_1, \dots, X_{n+1} , son $n+1$ v.a.i.i.d con distribución $N(\mu, \sigma^2)$. Por tanto, como se trata de obtener la distribución en el muestreo de $\frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$

Vamos a analizar la distribución de los elementos que intervienen en este estadístico. Se tiene que:

$$\left. \begin{array}{l} X_{n+1} : N(\mu, \sigma^2) \\ \bar{X} : N(\mu, \frac{\sigma^2}{n}) \end{array} \right\} \Rightarrow X_{n+1} - \bar{X} : N\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$$

Por otra parte, se verifica que

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

es tal que el estadístico del enunciado es

$$\frac{\frac{X_{n+1} - \bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}$$

Es decir, se trata de un cociente entre una distribución $N(0,1)$ y la raíz cuadrada del cociente de una distribución χ^2_{n-1} entre sus grados de libertad. Además son independientes ambas distribuciones. Por tanto, se trata de una distribución t_{n-1} .

81. La velocidad de una molécula de una gas de función de las coordenadas (x, y, z) de la misma respecto a un sistema cartesiano tridimensional, y es igual a la distancia de la molécula al origen del sistema.

Las coordenadas x, y, z son variables aleatorias independientes e igualmente distribuidas, según una ley $N(0, c^2)$

Si A y B son dos moléculas que se desplazan independientemente una de otra, calcular la probabilidad de que A se desplace a una velocidad igual o superior a cinco veces y media la de B.

SOLUCIÓN:

La velocidad v es $v = \sqrt{x^2 + y^2 + z^2}$ de donde $\frac{v^2}{c^2} = \left(\frac{x}{c}\right)^2 + \left(\frac{y}{c}\right)^2 + \left(\frac{z}{c}\right)^2$

Es una variable chi-cuadrado con 3 grados de libertad, ya que $\frac{x}{c}, \frac{y}{c}, \frac{z}{c}$ son variables $N(0,1)$ independientes.

Lo que se pide es la probabilidad siguiente.

$$P\left(\frac{v_A}{v_B} \geq 5,5\right) \text{ o lo que es igual } P\left(\frac{v_A}{v_B} \geq 5,5^2\right) = P\left(\frac{v_A^2/3c^2}{v_B^2/3c^2} \geq 30,25\right)$$

donde la variable $Z = \left(\frac{v_A^2/3c^2}{v_B^2/3c^2}\right)$

es una F de Snédecor con (3,3) grados de libertad, ya que es el cociente de dos variables chi-cuadrado, cada una de ellas dividida por sus grados de libertad.

Las tablas F de Snédecor indican que la probabilidad pedida es del orden del 1%

82. Sean X_1, X_2, X_3 y X_4 variables aleatorias independientes $N(0,1)$. Sean $a, b > 0$ y $R = a \cdot \sqrt{x_1^2 + x_2^2}$, $S = b \cdot \sqrt{x_3^2 + x_4^2}$. Calcular $P(R > S)$.

SOLUCIÓN:

$$P(R > S) = P(R^2 > S^2) = P\left(\frac{R^2}{S^2} > 1\right) = P\left(\frac{a^2}{b^2} \cdot \frac{x_1^2 + x_2^2}{x_3^2 + x_4^2} > 1\right) = P\left(\frac{x_1^2 + x_2^2}{x_3^2 + x_4^2} > \frac{b^2}{a^2}\right)$$

Pero $x_1^2 + x_2^2 + x_3^2 + x_4^2$ se distribuyen según leyes chi-cuadrado con dos grados de libertad, por lo que su cociente es una F de Snédecor con la pareja de grados de libertad (2,2).

Ahora bien como aparece en los elementos teóricos:

	Función de densidad f(x)
F de Snédecor	$\frac{k_n \cdot x^{(n/m)-1}}{c_n \cdot (m + nx)^{(n+2)/2}}$ $K_n = r\left(\frac{n+m}{2}\right) n^{\frac{n}{2}} \cdot m^{\frac{m}{2}}$ $c_n = r\left(\frac{n}{2}\right) \cdot r\left(\frac{m}{2}\right)$

La función de densidad de esta F de Snédecor es $f(x) = \frac{1}{(1+x^2)}$, $x > 0$

Y por tanto la probabilidad pedida es:

$$\int_{\frac{b^2}{a^2}}^{\infty} \frac{dx}{(1+x^2)} = \frac{1}{\left(1 + \frac{b^2}{a^2}\right)} = \frac{a^2}{a^2 + b^2}$$

83. De una población normal de media y desviación típica desconocida se ha obtenido una muestra de 26 elementos que tienen como media aritmética 5 y desviación típica 1,2. ¿Cuál es la probabilidad de que la media poblacional sea superior a 5.3?

SOLUCIÓN:

Sabemos que la variable $Z = \frac{\bar{X} - \mu}{S} \sqrt{n}$ se distribuye según una ley t de Student con n-1 grados de libertad, y lo que se pide es $P\left(Z < \frac{5 - 5,3}{1,2} \cdot \sqrt{25}\right) = P(Z < -1,25) = 0,1114$ donde Z sigue una t de Student con 25 grados de libertad.

84. Si X es una variable aleatoria con distribución $\gamma(n, \sigma)$ siendo n un número natural, determinar una función X que tenga distribución χ^2 .

SOLUCIÓN:

Si X tiene densidad $\frac{\theta^n}{\Gamma(n)} x^{n-1} e^{-\theta x}$, para $x > 0$, la variable aleatoria $Y = 2\theta X$ tendrá por densidad

$$\frac{\theta^n}{\Gamma(n)} \left(\frac{y}{2\theta}\right)^{n-1} e^{-y/2} \frac{1}{2\theta} = \frac{1}{2^n \Gamma(n)} y^{n-1} e^{-y/2}$$

para $y > 0$, que es la densidad de una χ^2 .

85. Se realiza un análisis de la duración de 40 pilas alcalinas obteniéndose los siguientes resultados:

Duraciones	Frecuencias absolutas
x_i	n_j

1,55-1,95	2
1,95-2,45	1
2,45-2,95	4
2,95-3,45	15
3,45-3,95	10
3,95-4,45	5
4,45-4,95	3

Ajustar las duraciones de las pilas alcalinas a una distribución normal con media 3,5 y desviación típica 0,7.

SOLUCIÓN:

Las probabilidades teóricas o esperadas p_i se calcularán mediante la distribución normal mediante:

$$P(1,45 \leq x \leq 1,95) = P\left(\frac{1,45 - 3,5}{0,7} \leq \frac{x - 3,5}{0,7} \leq \frac{1,95 - 3,5}{0,7}\right) = P(-2,93 \leq z \leq 2,21) = 0,012$$

Clases	n_i	p_i	np_i
1.45-1.95	2	0.012	0.5
1.95-2.45	1	0.0525	2.1
2.45-2.95	4	0.1475	5.9
2.95-3.45	15	0.2573	10.3
3.45-3.95	10	0.2573	10.7
3.95-4.45	5	0.175	7.0
4.45-4.95	3	0.0874	3.5

No obstante, suele ser habitual agrupar los intervalos consecutivos cuyas frecuencias teóricas son menores que 5 cm, en cuyo caso las letras se reducen a 4 y la tabla es la siguiente:

Clases	n_i	p_i	np_i
1.55-2.95	7	0.212	8.5

2.95-3.45	15	0.2573	10.3
3.45-3.95	10	0.2674	10.7
3.95-4.95	8	0.2624	10.5

Con estos datos ya podemos calcular el valor del estadístico del contraste de la chi-cuadrado. Los grados de libertad serán 4 ($n = 4$ y no se estima ningún parámetro).

$$\chi_4^2 = \sum_{i=1}^4 \frac{(n_i - np_i)^2}{np_i} = \frac{(7 - 8,5)^2}{8,5} + \frac{(15 - 10,3)^2}{10,3} + \frac{(10 - 10,7)^2}{10,7} + \frac{(8 - 10,5)^2}{10,5} = 3,05$$

Como tenemos que $\chi_{0,05;3}^2 = 7,815$ y $3,05 < 7,815$, resulta que el valor del estadístico chi-cuadrado cae fuera de la región crítica, por lo tanto se acepta la hipótesis nula de que el ajuste de los datos a una normal es correcto con un nivel de significación $\alpha = 0,5$.

86. Calcular los valores t_α correspondientes a una distribución t de Student en los siguientes casos:

- a) El área a la derecha de t_α es 0.20 y $n = 10$
- b) El área a la izquierda de t_α es 0.40 y $n = 8$
- c) El área a la derecha de t_α es 0.05 y $n = 50$

SOLUCIÓN:

a) Como $\frac{\alpha}{2} = 0.20 \rightarrow \alpha = 0.40$

$P(|t| > t_\alpha) = 0.40$ con $n = 5 \rightarrow t_\alpha = 0.0920$

b) Como $\frac{\alpha}{2} = 0.4 \rightarrow \alpha = 0.80$

$P(|t| > t'_\alpha) = 0.80$ con $n = 8 \rightarrow t'_\alpha = 0.262$ y $t_\alpha = -0.262$

c) Como para $n = 50$ no se encuentra el valor t_α en las tablas, ya que $n > 30$, se puede aproximar la distribución de t de Student por la $N(0,1)$

$P(|t| > t_\alpha) = 0.10$ ya que $\frac{\alpha}{2} = 0.05$

$$P(-t_{\alpha} \leq t \leq t_{\alpha}) = 1 - 0.10 = 0.90$$

$$P(-t_{\alpha} \leq t \leq t_{\alpha}) = 2P(t \leq t_{\alpha}) - 1 = 0.90$$

$$P(t \leq t_{\alpha}) = \frac{1.90}{2} = 0.95 \rightarrow t = 1.65$$

87. Calcular los valores de F_{α} correspondientes a una distribución F de Snédecor en los casos siguientes:

a) $\alpha = 0.01$ y (2,8) grados de libertad.

b) $\alpha = 0.05$ y (7,3) grados de libertad.

SOLUCIÓN:

a) $P(F > F_{\alpha}) = 0.01$ y $n_1 = 2$, $n_2 = 8$

$$F_{0.01,(2,8)} = 8.65$$

b) $P(F > F_{\alpha}) = 0.05$ y $n_1 = 7$, $n_2 = 3$

$$F_{0.05,(7,3)} = 27.67$$

88. Elegidas 26 personas al azar de una población y sometidas a un test de “modernismo” dan como media $\bar{X} = 40$ y $S = 6$. Se quiere saber si la verdadera media de la población puede ser tan alta como 44.

SOLUCIÓN:

-Se trata de una prueba unilateral.

-Trabajemos al nivel de confianza 1%.

-Calculamos el error típico de la media.

$$\sigma_{\bar{x}} = \frac{S}{\sqrt{N-1}} = \frac{6}{\sqrt{25}} = \frac{6}{5} = 1.20$$

Hallamos la razón crítica t:

$$t = \frac{\text{Estadístico} - \text{Parámetro}}{\text{Error}} = \frac{40 - 44}{1.20} = -3.33$$

Donde la razón crítica t tiene una distribución t de Student

Consultamos la tabla teniendo en cuenta:

-Los grados de libertad: $N-1 = 26 - 1 = 25$

-El nivel de confianza: 1%

Para 25 grados de libertad y al nivel de confianza del 1%, $t = -2.48$. Ahora bien, la t obtenida $= -0.33 < -2.48$ lo cual significa que de ser cierto que la media verdadera fuese 44, una media tan baja como 40 no podría resultar ni en 1 % de los casos. Luego con 99 % de probabilidades de acertar podemos decir que la media de la población no puede ser tan alta como 44 y, en consecuencia, rechazamos la hipótesis nula al nivel de confianza del 1%.

Si se pregunta por la fiabilidad de la media procederemos, en general, averiguando la diferencia máxima que estaremos dispuestos a admitir entre la media de una muestra y la media de la población a un determinado nivel de confianza. Según el ejemplo anterior, al nivel de confianza del 2 %.

$$\bar{X} - t_{0.02} \times \sigma_{\bar{X}} < \mu < \bar{X} + t_{0.02} \times \sigma_{\bar{X}}$$

$$\bar{X} - 2.48 \times 1.20 < \mu < \bar{X} + 2.48 \times 1.20$$

$$40 - 2.976 < \mu < 40 + t_{0.02} \times \sigma_{\bar{X}}$$

$$40 - 2.976 < \mu < 40 + 2.976$$

$$\mu \begin{cases} 42.976 \\ 37.024 \end{cases}$$

89. El fabricante de una dieta de adelgazamiento dice que su producto permite una reducción media de peso de 3,5kg. Con objetivo de investigar su eficacia, se seleccionaron al azar 40 personas, observando en ellas el peso antes de aplicar la dieta, X y el peso después de acabar el tratamiento, Y, lo que proporcionó una cuasi varianza para la diferencia de :

$$S_d^2 = 1/39 \sum_1^{40} [(X_i - Y_i) - (\bar{X} - \bar{Y})]^2 = 1,8$$

Si suponemos que tanto X como Y siguen distribuciones normales, determinar la probabilidad de que los individuos de la muestra haya una reducción media de masa de 3kg.

SOLUCIÓN:

Como las observaciones X_i e Y_i se realizan en los mismos individuos, antes y después del tratamiento, se trata de un problema de datos apareados. La diferencia de peso

$d = X - Y$ tiene media $\mu_d = \mu_x - \mu_y = 3,5$ y $S_d^2 = 1,8$.

Por tanto,

$$P\{\bar{d} > 3\} = P\left\{\frac{\bar{d} - \mu_d}{S_d \sqrt{n}} > \frac{3 - 3,5}{\sqrt{1,8/40}}\right\} = P\{t_{39} > -2,36\} = 0,988$$

después de interpolar en la tabla t de Student.

90. Si (x_1, \dots, x_n) es una muestra aleatoria simple de una $N(\mu_1, \sigma_1)$, con μ_1 conocida y σ_1 desconocida, comparar las distribuciones en el muestreo, la media y la varianza, de la cuasi varianza muestral S^2 y de

$$T = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

SOLUCIÓN:

$(n-1)S^2 / \sigma^2$ tiene distribución (x_{n-1}^2) ; luego

$$E[S^2] = \sigma^2 \quad \text{y} \quad V(S^2) = \frac{\sigma^4}{(n-1)^2} \quad V((n-1)S^2 / \sigma^2) = \frac{2\sigma^2}{n-1}$$

Por otro lado, como $(X_i - \mu) / \sigma$ es $N(0,1)$, $nT / (\sigma^2)$ será X_n^2 , por ser suma de n cuadrados de $N(0,1)$ independientes. Así pues,

$$E[T] = \sigma^2 \quad \text{y} \quad V(T) = \frac{\sigma^4}{n^2} V(nT / \sigma^2) = \frac{2\sigma^4}{n}$$

Ambos estadísticos se distribuyen alrededor de la media σ^2 , y la dispersión del segundo es inferior a la del primero. Si n es pequeño, T está considerablemente más concentrado alrededor de σ^2 que S^2 ; de manera que conviene hacer uso, mediante T del conocimiento de μ , en lugar de ignorarlo usando S^2 . Cuando n crece, la distribución de ambos estadísticos se concentra rápidamente alrededor de σ^2 y la diferencia entre ambos se hace inapreciable. (Naturalmente, si μ fuese desconocida, los valores de T no podrían determinarse).

91. La efectividad en días de un determinado antibiótico, sigue una distribución normal de media 14 días y desviación típica desconocida. Fue administrada a 16 enfermos, obteniéndose una cuasi desviación típica muestral de 1,4 días. Determinar la probabilidad de que la efectividad media en la muestra no supere los 3 días, que es el tiempo mínimo de efectividad requerido.

Preocupados por una posible subestimación de la varianza poblacional, que podría llevar a subestimar la probabilidad de que no se alcance la efectividad mínima, se desea determinar la probabilidad de que con una muestra de 16 enfermos se subestime la varianza en más de un 20%. Si la muestra es de 61 pacientes, esta probabilidad ¿aumenta o disminuye?

Determinar el tamaño de muestra necesario para que la probabilidad anterior sea 0,05.

SOLUCIÓN:

La probabilidad es $P\{X \leq 13\}$. Como la población es normal, pero no se conoce la varianza poblacional, habrá que utilizar la distribución t de Student.

$$P\{\bar{X} \leq 13\} = P\left\{\frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \frac{13 - 14}{1,4/\sqrt{16}}\right\} = P\{t_{15} \leq -2,857\} = 0,0063$$

Donde el valor se obtiene, por simetría, en la distribución t_{15} .

Como $(n-1) S^2 / \sigma^2$ tiene distribución χ^2_{n-1} , la probabilidad de subestimación de σ^2 en más de un 20% es

$$P\{s^2 < 0,8\sigma^2\} = P\{\chi^2_{15} < 12\} = 0,3,$$

calculada a partir de las tablas por interpolación lineal. Para una muestra de 61 individuos, utilizando la aproximación de la distribución χ^2_n

$$P\{s^2 < 0,8\sigma^2\} = P\{x_{60}^2 < 48\} \approx P\left\{\sqrt{2}\chi_{60}^2 - \sqrt{119} < \sqrt{96} - \sqrt{119}\right\} = \\ = P\{Z < -1.11\} = 0.1335$$

Es decir, la probabilidad disminuye al aumentar el tamaño de la muestra. Por último, para conseguir que:

$$0,05 = P\{\chi_{n-1}^2 < 0,8(n-1)\} \approx P\left\{Z < \sqrt{1,6(n-1)} - \sqrt{2n-1}\right\}$$

ha de ser $\sqrt{1,6(n-1)} - \sqrt{2n-1} = -1,645$ De donde $n = 117,61$; con tamaño muestral 118, se obtiene la precisión requerida.

92. Hallar el valor crítico de χ_α^2 de χ^2 en los siguientes casos:

a) $\alpha = 0.05$ y $n = 8$

b) $\alpha = 0.30$ y $n = 7$

c) $\alpha = 0.01$ y $n = 61$

SOLUCIÓN:

a) $P(\chi^2 > \chi_\alpha^2) = \alpha$ con $n = 8$

$$P(\chi^2 > \chi_{0.05}^2) = 0.05 \quad \rightarrow \quad \chi_{0.05}^2 = 15.507$$

b) $P(\chi^2 > \chi_{0.30}^2) = 0.30$ y $n = 7$ \rightarrow $\chi_{0.30}^2 = 8.383$

c) Para $n = 61$ no figura χ_α^2 en la tabla y se calcula mediante:

$$Z = \sqrt{\chi^2} - \sqrt{2n-1}$$

que se distribuye según una $N(0,1)$

$$P(Z > z_1) = 0.01 \quad \rightarrow \quad P(Z \leq z_1) = 1 - 0.01 = 0.99 \quad \rightarrow \quad z_1 = 2.33$$

$$Z = \sqrt{2\chi^2} - \sqrt{2 \times 61 - 1} > 2.33$$

$$\sqrt{2\chi^2} > 2.33 + 11 = 13.33$$

$$2\chi^2 > 117.689$$

$$\chi^2 > 88.844$$

$$P(\chi^2 > 88.844) = 0.01 \quad \rightarrow \quad \chi^2 = 88.844$$

93. Sea una v.a. X distribuida χ^2_{30} . Se trata de calcular:

- a) $P[X < 20.6]$
- b) $a / P[X > a] = 0.05$
- c) $c_{0.5}$ de X ,
- d) $E[X]$,

Por los tres procedimientos siguientes:

- i) utilizando las tablas de la chi-cuadrado,
- ii) utilizando la aproximación del teorema central del límite,
- iii) utilizando otra aproximación asintótica para la chi-cuadrado

Deducir de lo anterior que es mejor la aproximación iii) que la ii)

SOLUCIÓN:

a) i) Mirando en las tablas de la chi-cuadrado con 30 grados d libertad se tiene que $P[X > 20.599] = 0.9$, luego $P[X < 20.599] = 0.1$.

ii) por la aproximación del teorema central del límite

$$P[X < 20.6] = P\left[\frac{X-30}{\sqrt{60}} < \frac{20.6-30}{\sqrt{60}}\right] \cong P[Z < -1.2135] = 0.1124,$$

con $Z: N(0,1)$.

iii) La otra aproximación asintótica a la chi-cuadrado es aquel que si $X : \chi^2_{30}$, entonces $U = \sqrt{2X}$ se distribuye asintóticamente $N(\sqrt{59}, 1)$, luego

$$P[X < 20.6] = P[U < \sqrt{2 \times 20.6}] \cong P[Z < \sqrt{2 \times 20.6} - \sqrt{59}] = P[Z < -1.2624] = 0.1034$$

b) i) En este caso $a = 43.773$.

$$\text{ii) } P [X > a] = 0.05 \Leftrightarrow P \left[Z > \frac{a-30}{\sqrt{60}} \right] \cong 0.05, \text{ con } Z : N(0,1)$$

$$\text{De donde } a = 1.645 * \sqrt{60} + 30 = 42.742$$

$$\text{iii) } P [X > a] = P [U > \sqrt{2a}] \cong P [Z > \sqrt{2a} - \sqrt{59}]. \text{ Luego se busca un}$$

$$a / P [Z > \sqrt{2a} - \sqrt{59}] \cong 0.05 \text{ con } Z: N(0,1), \text{ y resulta que:}$$

$$\sqrt{2a} - \sqrt{59} = 1.645 \rightarrow a = 43.488$$

$$\text{c) i) En este caso } c_{0.5} = 29.336.$$

$$\text{ii) } P [X \leq c_{0.5}] = 0.5 \Leftrightarrow P \left[Z \leq \frac{c_{0.5}-30}{\sqrt{60}} \right] \cong 0.5, \text{ con } Z : N(0,1).$$

$$\text{De donde } c_{0.5} = 30$$

$$\text{iii) } P [X \leq c_{0.5}] = P [U < \sqrt{2c_{0.5}}] \cong P [Z < \sqrt{2c_{0.5}} - \sqrt{59}] \cong 0.5,$$

con $Z : N(0,1)$. De donde

$$\sqrt{2c_{0.5}} - \sqrt{59} = 0 \Rightarrow c_{0.5} \cong 29.5$$

$$\text{d) i) } E[X] = 30$$

ii) Como X se distribuye asintóticamente $N(30,60)$, se tiene que

$$E[X] \cong 30$$

iii) En este caso $E[U^2] = E[2X] \rightarrow E[X] = \frac{1}{2} E[U^2]$, donde U se distribuye asintóticamente $N(\sqrt{59}, 1)$, luego

$$E[U^2] = V[U] + E^2[U] \cong 59 + 1 = 60. \text{ Por tanto,}$$

$$E[X] \cong 30$$

Conclusión: De los apartados a), b), c) se observa que la aproximación asintótica de iii) es preferible a la del teorema central del límite, por darnos valores y probabilidades más cercanos a los verdaderos. Sin embargo, en d), las dos aproximaciones coinciden, y coinciden con el verdadero valor. Esto sucede por tratarse de la media, que suele mantenerse estable en el valor verdadero de la media, cuando se utilizan las aproximaciones asintóticas.

94. Extraemos una m.a.s. de 61 estudiantes universitarios, los cuales responden a una prueba de inteligencia espacial. Se obtiene una media muestral de 80 y una varianza

muestral de 100. ¿entre qué límites se hallará la verdadera puntuación media de la prueba, a un nivel de confianza del 99%?

SOLUCIÓN:

$$1 - \alpha = 0,99 ; \alpha = 0,01 ; 1 - \frac{\alpha}{2} = 0,995$$

Como la varianza poblacional es desconocida pero el tamaño muestral es mayor que 30, aunque la población no es normal, el intervalo correspondiente será:

$$\left[\bar{X}_n - t_{1-\frac{\alpha}{2}} \frac{\alpha}{2} (n-1) \frac{s_n}{\sqrt{n-1}} , \bar{X}_n + t_{1-\frac{\alpha}{2}} \frac{\alpha}{2} (n-1) \frac{s_n}{\sqrt{n-1}} \right]$$

Buscamos en las tablas de la distribución t de Student: $t_{0,995} (61 - 1) = 2,66$

Sabemos que $\bar{X}_{61} = 80$ y $s_{61}^2 = 100$. Sustituyendo en el intervalo y haciendo operaciones tenemos:

$$\left[80 - 2.66 \frac{10}{\sqrt{60}} , 80 + 2.66 \frac{10}{\sqrt{60}} \right] \Rightarrow \epsilon [76.57 , 83.43] \text{ con confianza del } 99\%$$

95. Uno de los líderes de un colectivo laboral desea plantear una cuestión a todos los miembros del grupo. Si más de la mitad respondieran NO entonces preferiría no plantearla. Para salir de dudas, elige al azar 40 trabajadores a los que hace la pregunta y sólo 12 responden NO. ¿ entre qué límites se hallará la verdadera proporción al nivel de confianza del 95%?

SOLUCIÓN:

vamos a calcular el intervalo de confianza con las dos expresiones.

$$1 - \alpha = 0.95 ; 1 - \frac{\alpha}{2} = 0.975 ; z_{1-\frac{\alpha}{2}} = 1.96 \quad \text{y} \quad P = \frac{12}{40}$$

Sustituyendo en los intervalos tendremos:

$$\text{a) } \left[0.3 - 1.96 \sqrt{\frac{0.3 \cdot 0.7}{40}} , 0.3 + 1.96 \sqrt{\frac{0.3 \cdot 0.7}{40}} \right] \rightarrow p \in [0.158, 0.442]$$

$$b) \left[0.3 - 1.96 \frac{0.5}{\sqrt{40}} + 0.3 - 1.96 \frac{0.5}{\sqrt{40}} \right] \rightarrow p \in [0.145, 0.455]$$

En ambos casos, el intervalo no llega a 0.5, por lo tanto podemos aconsejar al líder que formule la pregunta.

96. Con el fin de comparar el promedio de faltas de ortografía cometidas en una composición por dos clases similares de alumnos, se tomaron dos muestras de 7 y 8 alumnos, respectivamente y se observaron los siguientes errores:

Clase 1: 10,10, 12,12, 13, 13,14

Clase 2. 8 ,9 , 10, 10,10,10,12,12

Suponiendo que el numero de errores en ambas clases son normales, calcular es intervalo de confianza del 95% para la diferencia de medias:

- a) suponiendo que las varianzas poblacionales son iguales y valen $\sigma^2 = 1.44$**
- b) suponiendo que las varianzas son desconocidas pero iguales.**

SOLUCIÓN:

De los datos obtenemos:

- clase 1: $\bar{X}_7 = 12$ $s_7^* = 1.53$ $n = 7$ $s_7^2 = 2$ $s_7^{*2} = 2.33$
- clase 2: $\bar{Y}_8 = 10.125$ $s_8^* = 1.36$ $m = 8$ $s_8^2 = 1.61$ $s_8^{*2} = 1.84$

a) $1 - \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.975$; $z_{1-\frac{\alpha}{2}} = 1.96$

$$\mu_1 - \mu_2 \in \left[12 - 10.125 - 1.96 \sqrt{\frac{1.44}{7} + \frac{1.44}{8}}, 12 - 10.125 + 1.96 \sqrt{\frac{1.44}{7} + \frac{1.44}{8}} \right] =$$

$$= [0.658, 3.092]$$

b) $1 - \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.975 \rightarrow t_{1-\frac{\alpha}{2}}(n + m - 2) = t_{0.975}(13) = 2.16$

$$\mu_1 - \mu_2 \in \left[12 - 10.125 - 2.16 \sqrt{\frac{7 \cdot 2 + 8 \cdot 1.84}{7 + 8 - 2} \left(\frac{1}{7} + \frac{1}{8} \right)}, 12 - 10.125 + 2.16 \sqrt{\frac{7 \cdot 2 + 8 \cdot 1.84}{7 + 8 - 2} \left(\frac{1}{7} + \frac{1}{8} \right)} \right] = [0.215, 3.535]$$

Como el valor 0 no pertenece a ninguno de los dos intervalos, podemos concluir que las medias no serán iguales con una confianza del 95%.

97. Establecer el intervalo de confianza del 95% para la diferencia en las proporciones de alumnos de BUP que llega a la universidad procedentes de dos sistemas escolares distintos, si sabemos que:

Sistema escolar X: terminan BUP 600 alumnos; acceden universidad 90 alumnos.

Sistema escolar Y: terminan BUP 400 alumnos; acceden universidad 50 alumnos.

SOLUCIÓN:

$$1 - \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.975 \rightarrow z_{0.975} = 1.96$$

$$P_x = \frac{90}{600} = 0.15 \quad P_y = \frac{50}{400} = 0.125$$

Sustituyendo en el intervalo correspondiente:

$$P_x - P_y \in \left[0.15 - 0.125 - 1.96 \sqrt{\frac{0.15 \cdot 0.85}{600} + \frac{0.125 \cdot 0.875}{400}}, 0.15 - 0.125 + 1.96 \sqrt{4.8594 \cdot 10^{-4}} \right] = [-0.018, 0.068]$$

Como el valor 0 pertenece al intervalo, podemos concluir que la proporción de alumnos que acceden a la universidad es similar en los dos sistemas escolares.

98. En una muestra de 19 individuos se observa que un determinado trastorno emocional se produce a partir de una edad media de 50 años y una desviación típica de 6 años. Se supone que estamos ante un fenómeno que sigue la ley normal.

a) fijar los límites del intervalo de confianza para la varianza con un nivel de confianza del 99%

b) realizar lo mismo que en el apartado anterior pero suponiendo $n = 200$

SOLUCIÓN:

$$\text{a) } s_{19}^2 = 36, \quad 1 - \frac{\alpha}{2} = 0.995 \rightarrow X_{0.995}^2(18) = 37.2 \text{ y } X_{0.005}^2(18) = 6.26$$

$$\sigma^2 \in \left[\frac{19 \cdot 36}{37.2}, \frac{19 \cdot 36}{6.26} \right] = [18.39, 109.27]$$

b) como $n = 200 > 100$ utilizamos el intervalo de la nota: $z_{0.995} = 2.57$

$$\sigma^2 \in \left[\frac{200 \cdot 36}{199 + 2.57\sqrt{2} \cdot 200}, \frac{200 \cdot 36}{199 - 2.57\sqrt{2} \cdot 200} \right] = \left[\frac{7200}{250.4}, \frac{7200}{147.6} \right] = [28.75, 48.78]$$

Al utilizar un tamaño muestral mayor, se reduce a amplitud del intervalo.

99. En una prospección electoral se desea estimar el porcentaje de votos que obtendrá el partido del gobierno con una probabilidad del 90% y un error de estimación del $\pm 1\%$. Calcular el tamaño muestral necesario para conseguir una estimación de estas características.

SOLUCIÓN:

$$E = \pm 1\% \rightarrow E = 0.02$$

$$\text{El intervalo para } p \text{ es } \left[\bar{p} - \lambda_{\frac{\alpha}{2}} \frac{0.5}{\sqrt{n}}; \bar{p} + \lambda_{\frac{\alpha}{2}} \frac{0.5}{\sqrt{n}} \right]$$

$$E = 0.02 = 2 \lambda_{\frac{\alpha}{2}} \frac{0.5}{\sqrt{n}}$$

$$1 - \alpha = 0.99, \alpha = 0.01, \lambda_{\frac{\alpha}{2}} = 2.5733 \text{ (tras interpolar)}$$

Así que $\rightarrow 0.02 = 2.5733 \cdot \frac{0.5}{\sqrt{n}}$, despejando $\rightarrow n = 16554,68223$

Es decir habrá que mostrar a 16555 personas.

100. Los errores aleatorios que se cometen en las pasadas de una cierta balanza siguen una distribución normal de media 0 y desviación típica 1mg. Después de n=9 pesadas se han comprobado los errores (en mg): -0.07, 0.3, 1.8, -0.1, 2.0, 2.3, 0.62, 0.12, y 1.4.

Se pide contrastar la hipótesis de que la balanza funciona correctamente frente a la alternativa de que tiene un error sistemático positivo, con un nivel de significación de 0.1.

SOLUCIÓN:

Los errores aleatorios cometidos al pesar siguen una distribución normal $N(0, 1)$. Debemos contrastar H_0 (el error es aleatorio), frente H_1 (el error es sistemático positivo):

$$H_0: X \text{ es } N(0,1) \rightarrow f_0(x) = \frac{1}{\sqrt{2\pi \cdot 1^2}} \cdot e^{-\frac{1}{2} \frac{(x-0)^2}{1^2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}$$

$$H_1: X \text{ es } N(m,1) \rightarrow f_1(x) = \frac{1}{\sqrt{2\pi \cdot 1^2}} \cdot e^{-\frac{1}{2} \frac{(x-m)^2}{1^2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-m)^2}$$

Tenemos que plantear cual es la región crítica óptima para un m dado:

$$w = \left\{ \bar{x} / \frac{f_1(\bar{x})}{f_0(\bar{x})} \geq a \right\} = \left\{ x = (x_1, x_2, \dots, x_n) / \frac{f_1(x_1, x_2, \dots, x_n)}{f_0(x_1, x_2, \dots, x_n)} \geq a \right\} =$$

$$\left\{ \frac{\bar{x} / \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-m)^2}}{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}} \geq a \right\} = \left\{ \bar{x} / \frac{\left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x_1-m)^2} \right) \dots \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x_n-m)^2} \right)}{\left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x_1^2} \right) \dots \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x_n^2}} \geq a \right\} =$$

$$\begin{aligned}
&= \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n [x_i^2 - 2x_i m + m^2 - x_i^2]} \geq a \right\} = \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n [(x_i - m)^2 - x_i^2]} \geq a \right\} = \\
&= \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n (-2x_i m + m^2)} \geq a \right\} = \left\{ \bar{x} / -\frac{1}{2} \sum_{i=1}^n (-2mx_i + m^2) \geq \ln a \right\} = \\
&= \left\{ \bar{x} / m \sum_{i=1}^n x_i - \frac{1}{2} m^2 n \geq \ln a \right\} = \left\{ \bar{x} / \sum_{i=1}^n x_i \geq \frac{\ln a + \frac{1}{2} m^2 n}{m} \right\}
\end{aligned}$$

Anotación: $\left\{ \frac{\ln a + \frac{1}{2} m^2 n}{m} \right\} = b$

La región crítica óptima es: $w = \left\{ \bar{x} / \sum_{i=1}^n x_i \geq b \right\}$

Vamos a hallar la región crítica óptima para un nivel de significación $\alpha = 0,1$

Bajo la hipótesis H_0 (bajo la cual se construye la región crítica w) $\rightarrow X$ es $N(0,1)$

$$\sum_{i=1}^n x_i \text{ es } N(0, \sqrt{n})$$

Por tanto:

$$\frac{\sum_{i=1}^n x_i}{\sqrt{n}} = y \text{ es } N(0,1)$$

$$P(Y \geq Y_\alpha) = 0,1 \quad \rightarrow \quad P(|Y| \geq Y_\alpha) = 0,2$$

Mirando las tablas:

$$Y_\alpha = 1.28$$

Por tanto:

$$w = \left\{ \bar{x} / \frac{\sum_{i=1}^n x_i}{\sqrt{n}} \geq 1.28 \right\} \rightarrow w = \left\{ \bar{x} / \sum_{i=1}^n x_i \geq 1.28\sqrt{n} \right\} \rightarrow \text{Región crítica óptima}$$

En nuestro caso: $n=9$

$$\sum_{i=1}^9 x_i \geq 1.29 \cdot \sqrt{9} \rightarrow 8.37 \geq 3.84 \rightarrow \text{Rechazamos } H_0 \text{ y se acepta } H_1$$

101. Sea X v.a. N ($\theta,1$). Contrastar la hipótesis H_0 ($\theta = 0$) frente a la alternativa H_1 ($\theta = 4$) para una muestra de tamaño $n=100$ y un nivel de significación de 0.05.

SOLUCIÓN:

Sea X v.a. N ($\theta,1$)

Debemos contrastar las siguientes hipótesis:

$$H_0: \theta = 0$$

$$H_1: \theta = 4$$

Su función de densidad es: $\rightarrow f(x, \theta) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-\theta)^2}$

Y su función de densidad conjunta es: $\rightarrow f(\bar{x}, \theta) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2}$

Aplico Neyman Pearson:

$$w = \left\{ \bar{x} / \frac{f_1(\bar{x}, \theta = 4)}{f_0(\bar{x}, \theta = 0)} \geq a \right\} = \left\{ \bar{x} / \frac{\left(\frac{1}{\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n (x_i - 4)^2}}{\left(\frac{1}{\sqrt{2\pi}} \right)^n \cdot e^{-\frac{1}{2} \sum_{i=1}^n (x_i - 0)^2}} \geq a \right\} =$$

$$\begin{aligned}
&= \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n [(x_i-4)^2 - (x_i-0)^2]} \geq a \right\} = \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n (x_i^2 - 8x_i + 16 - x_i^2)} \geq a \right\} = \\
&= \left\{ \bar{x} / e^{-\frac{1}{2} \sum_{i=1}^n (16 - 8x_i)} \geq a \right\} = \left\{ \bar{x} / e^{-\frac{1}{2} \left[\sum_{i=1}^n 16 - \sum_{i=1}^n 8x_i \right]} \geq a \right\} = \left\{ \bar{x} / e^{-\frac{1}{2} (16n - 8 \sum_{i=1}^n x_i)} \geq a \right\} = \\
&= \left\{ \bar{x} / -8n + 4 \sum_{i=1}^n x_i \geq \ln a \right\} = \left\{ \bar{x} / \sum_{i=1}^n x_i \geq \frac{\ln a + 8n}{4} \right\} =
\end{aligned}$$

Si $n = 100$

$$= \left\{ \bar{x} / \sum_{i=1}^n x_i \geq \frac{\ln a}{4} + 200 \right\} = \sum_{i=1}^n x_i \geq \frac{\ln a}{4} + 200$$

Anotación $\rightarrow \frac{\ln a}{4} + 200 = b$

$$\bar{x}_{100} \geq b \rightarrow P(\bar{x}_{100} \geq b) = 0.05$$

X es $N(0,1) \rightarrow \bar{x}_{100}$ es $N\left(0, \frac{1}{\sqrt{n}}\right) =$ Supuesto cierto H_0

Al tipificar:

$$P\left(\frac{\bar{x}_{100} - 0}{\frac{1}{\sqrt{n}}} \geq \frac{b - 0}{\frac{1}{\sqrt{n}}}\right) = 0.05 \rightarrow P\left(\frac{\bar{x}_{100} - 0}{\frac{1}{10}} \geq \frac{b - 0}{\frac{1}{10}} = 10b\right) = 0.05$$

Anotación: $10b = c$

$$P(Y \geq c) = 0.05$$

$$P(|Y| \geq c) = 0.1$$

Mirando las tablas: $c = 1.64$, por tanto

$$10 \cdot b = 1.644 \rightarrow b = 0.1644 \rightarrow \bar{X}_{100} = 0.1644$$

Para $\bar{X}_{100} \geq 0.1644$ se rechaza H_0

Para $\bar{X}_{100} < 0.1644$ se acepta H_0

102. Sea X v.a $N(\theta, \sigma_0)$ donde (σ_0) es conocido. Contrastar la hipótesis $H_0 (\theta = \theta_0)$ frente a la alternativa $H_1 (\theta \neq \theta_0)$ por el test de razón de verosimilitud.

SOLUCIÓN:

Construimos la f^{on} de verosimilitud:

$$\theta = \theta_0$$

Una vez construida la f^{on} de verosimilitud, nos fijamos en los espacios.

El espacio paramétrico Ω queda definido por:

$$\Omega = \{-\infty < \theta < +\infty\}$$

El espacio paramétrico bajo la hipótesis nula es:

$$\omega = \{\theta_0\}$$

De aquí sacamos varias conclusiones inmediatas:

1) - $\dim \omega = 0 \rightarrow r$

2) - $\dim \Omega = 1 \rightarrow k$

El Máximo de L en

- ω es con $\theta = \theta_0$

- Ω es con $\theta = \bar{x}_n$

La razón de verosimilitud es:

$$\Lambda = \frac{L(\bar{x}, \theta_0)}{L(\bar{x}, \bar{x}_n)} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \cdot e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \theta_0)^2}}{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \cdot e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2}} = e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n [(x_i - \theta_0)^2 - (x_i - \bar{x}_n)^2]} =$$

$$= e^{-\frac{n}{2\sigma_0^2} (\bar{x}_n - \theta_0)^2}$$

Según el Teorema de Wilks:

$$U = -2 \ln \Lambda \text{ es } X_1^2 \text{ (el 1 sale de } k-r=1-0=1)$$

$$U = \frac{n}{\sigma_0^2} (\bar{x}_0 - \theta_0)^2 \approx X_1^2$$

Suponiendo un $E > 0$ dado, por ejemplo: $E = 0,05$

Tenemos que hallar:

$$X_E^2 / P(X_1^2 > X_E^2) = 0.05$$

Mirando las tablas:

$$X_E^2 = 3.841$$

Para una muestra \bar{x} se calcula U :

$$U = \frac{n}{\sigma_0^2} (\bar{x}_0 - \theta_0)^2$$

Y comparamos para el valor dado para X_E^2 :

- Si $U \leq 3.841$ se acepta H_0
- Si $U > 3.841$ se acepta H_1

103. Se hace un envío de latas de conserva, de las que se afirma que el peso medio es de 1000gr. Examinada una muestra de 5 latas se han obtenido pesos respectivos de 995, 992, 1005, 998 y 1000. ¿Puede mantenerse la hipótesis de que la media es 1000 con un nivel de significación $\alpha = 0.05$? Hallar un intervalo de confianza para la media, suponiendo normalidad en la población.

SOLUCIÓN:

Tenemos que contrastar la hipótesis:

$$H_0: \mu=1000 \quad (\text{hipótesis nula})$$

$$H_1: \mu \neq 1000 \quad (\text{hipótesis alternativa})$$

$$n=5 \rightarrow \mu=1000$$

Como X es v.a. N (μ , σ) con σ desconocida, usamos:

$$\frac{\bar{x}_n - \mu}{S_n} \cdot \sqrt{n-1} \text{ es } t_{n-1}$$

Por lo tanto necesitamos saber:

$$\bar{x}_n = \frac{995 + 992 + 1005 + 998 + 1000}{5} = 998$$

$$s_n^2 = \frac{1}{5} \left[(995 - 998)^2 + (992 - 998)^2 + (1005 - 998)^2 + (998 - 998)^2 + (1000 - 998)^2 \right] = \frac{98}{5} = 19.6$$

$$s_n = \sqrt{s_n^2} = \sqrt{19.6} = 4.43$$

Sustituyendo:

$$t_4 = \frac{998 - 1000}{4.43} \cdot \sqrt{4} = -0.903$$

$$P(|t_4| > t_\alpha) = 0.05$$

Mirando la tabla t de Student obtenemos:

$$t_\alpha = 2.776 \quad | -0.903 | > 2.776? \quad \text{NO, por lo tanto } H_0, \text{ es decir } \mu=1000$$

Para hallar el intervalo de confianza para μ :

$$P\left(-t_\alpha < \frac{\bar{x}_n - \mu}{s_n} \sqrt{n-1} < t_\alpha\right) = 1 - \alpha$$

$$P\left(-2.776 < \frac{998 - \mu}{4.43} \sqrt{4} < 2.776\right) = 0.95$$

El intervalo es:

$$I\left(998 - 2.776 \cdot \frac{4.43}{2} < \mu < 998 + 2.776 \cdot \frac{4.43}{2}\right)$$

$$I(991,9 < \mu < 1004.1)$$

También podemos decir que aceptamos H_0 porque $1000 \in I$

104. Se desean comparar dos nuevas líneas de trigo; para ello, se toman 10 fincas al azar, plantando en cada una de ellas, y en dos parcelas distintas, ambas líneas. La producción en las 10 fincas fue (en cierta unidad de medida) de 57, 49, 60, 55, 57, 48, 50, 61, 52 y 56 para la línea A y de 55, 48, 58, 56, 54, 48, 52, 56, 50 y 58 para la línea B. ¿Podemos aceptar que la producción es la misma, a un nivel $\alpha=0.05$ (y suponiendo la población normal bivalente)?

SOLUCIÓN:

$$n=10$$

$$\alpha=0.05$$

Finca	1	2	3	4	5	6	7	8	9	10
A	57	49	60	55	57	48	50	61	52	56
B	55	48	58	56	54	48	52	56	50	58

Suponemos la población normal bivalente.

Llamamos; $X=X_1, \dots, X_{10}$ a la producción en A

$Y=Y_1, \dots, Y_{10}$ a la producción en B

Debemos contrastar la hipótesis:

$$H_0: X_1=Y_1; X_2=Y_2; \dots, X_{10}=Y_{10} \rightarrow H_0: \mu_a=\mu_b \text{ (hipótesis nula)}$$

$$H_1: X_1 \neq Y_1; X_2 \neq Y_2; \dots, X_{10} \neq Y_{10} \rightarrow H_1: \mu_a \neq \mu_b \text{ (hipótesis alternativa)}$$

Pero en este caso las dos v.a. $X=(\text{producción de A})$, y $Y=(\text{producción en B})$ son dependientes, porque cada producción está tomada en la misma finca y depende de las características del terreno.

Ahora bien, si H_0 es cierta, se verifica: $E(X-Y) = E(X) - E(Y) = \mu_a - \mu_b = 0$ y la distribución de $X-Y$ será normal de media 0.

Quedando la hipótesis:

$H_0: E(X-Y)=0$ (hipótesis nula)

$H_1: E(X-Y) \neq 0$ (hipótesis alternativa)

Sea $D=(X-Y)$

$$D=2, 1, 2, -1, 3, 0, -2, 5, 2, -2$$

$$\bar{D} = \frac{2+1+2-1+3+0-2+5+2-2}{10} = 1$$

$$s_D^2 = \frac{1}{10} [(2-1)^2 + (1-1)^2 + (2-1)^2 + (-1-1)^2 + (3-1)^2 + (0-1)^2 + (-2-1)^2 + (5-1)^2 + (2-1)^2 + (-2-1)^2]$$

$$s_D^2 = 4.6 \rightarrow s_D = 2.14$$

Por lo tanto tenemos que:

$$\frac{\bar{D} - 0}{s_D} \cdot \sqrt{n-1} \text{ es } t_{n-1}$$

$$t_1 \approx \frac{1-0}{2.14} \cdot 3 \rightarrow t_9 \approx 1.4$$

Por otro lado tenemos:

$$P(t_9 > t_\alpha) = 0.05$$

Mirando en la tabla t de Student obtenemos:

$$t_\alpha = 2.262 \rightarrow ¿1.4 > 2.262?$$

No lo es, por lo tanto aceptamos H_0 , es decir no hay diferencia entre la producción de las dos líneas de trigo.

105. De 100 observaciones de una población normal se obtiene que $\bar{x} = 5$ y que $S = 2$. Contrastar con un nivel de significación del 5% la hipótesis de que la media de la población sea 7.

SOLUCIÓN:

$$H_0: \mu = 7 \quad = 1'96$$

$$H_1: \mu \neq 7$$

$$\alpha = 0'05 \quad \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{5 - 7}{2/10} = -10$$

$$\bar{x} = 5$$

$$S = 2$$

$$n = 100 \gg 30 \quad S \equiv \sigma = 2$$

Como $-10 \notin [-1'96, 1'96]$ rechazamos la hipótesis: Rechazamos la hipótesis de que $\mu = 7$.

106. Se escoge a 17 individuos al azar y se les mide. Resultando que su estatura media es 1,71m. Y que la desviación típica es 0'02 metros. Contrastar la hipótesis de que la estatura media nacional sea 1'75 con un nivel de significación del 5%.

SOLUCIÓN:

Distribución de la población $L(X) = N[\mu, \sigma]$

$$L\left[\frac{\bar{x} - \mu}{S} = (\sqrt{n-1})\right] = t_{n-1}$$

$$\bar{x} = 1'71 \quad t_{0,05}(16 \text{ g.l.}) = 2'12$$

$$S = 0'02$$

$$n = 17 < 30 \quad \frac{\bar{x} - \mu_0}{S} = (\sqrt{n-1}) = \frac{1,71 - 1,75}{0,02} \times 4 =$$

$$H_0: \mu = 1'75$$

$$H_1: \mu \neq 1'75 \quad = 8 \notin [-2,12; 2,12]$$

Por lo tanto Rechazamos la hipótesis de que la media sea 1'75

107. Un fabricante de refrescos sin burbujas desea sacar al mercado una variedad de su producto que tenga burbujas. Su director comercial opina que al menos el 50% de los consumidores verá con buenos ojos la innovación. Se realiza un sondeo de mercado y resulta que de 100 consumidores encuestados 40 son favorables a la innovación.

a) Contrastar la hipótesis del director comercial frente a la alternativa de que el % de aceptación es inferior, con un nivel de significación del 2,5%.

b) Si es aceptable la hipótesis de que el % de aceptación del nuevo producto es inferior o igual al 30% el fabricante decidirá no fabricarlo. Si es aceptable el criterio del director comercial entonces sí fabricarán el refresco con burbujas. Y si ninguna de las 2 hipótesis es aceptable procederán a hacer otro sondeo. Para tomar esta decisión trabajarán con un nivel de significación del 2,5% ¿ Por qué opción se decantarán?

SOLUCIÓN:

a)

$$H_0: p = 0,5$$

$$H_1: p < 0,5$$

$$\alpha = 0,025 \quad -\lambda_\alpha / \phi(-\lambda_\alpha) = 0,025 \quad -\lambda_\alpha \text{ es negativo}$$

y su opuesto es tal que $(+\lambda_\alpha) = 1 - 0,025 = 0,975$, así que $-\lambda_\alpha = -1,96$

$$\hat{p} = 0,4 \quad n = 100 \gg 30$$

$-2 < -1,96$ así es que rechazamos que $p = 0,5$

b)

$$H_0: p = 0,3$$

$$H_1: p > 0,3$$

$$\alpha = 0,025 \quad \lambda_\alpha = 1,96$$

$$\hat{p} = 0,4 \quad n = 100$$

$$\frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0,1}{0,5} = 2$$

$2 > 1,96$ así que rechazamos que $p = 0,3$

Como no se aceptan ninguna de las dos hipótesis habrá que hacer otro sondeo.

108. Contrastar la hipótesis de que dos poblaciones tienen la misma dispersión con un nivel de significación del 1% y sabiendo que la desviación típica de una muestra realizada sobre la primera población era 12 con un tamaño muestral de 25 y que en una muestra sobre la segunda de tamaño 30 la desviación típica resultó ser 7. Considérese que ambas poblaciones son normales.

SOLUCIÓN:

$$S^2x = 144$$

$$S^2y = 49 \quad \frac{S^2x}{S^2y} * \frac{n_x}{n_y} * \frac{(n_y - 1)}{(n_x - 1)} = \frac{144 * 25 * 29}{49 * 30 * 24} = 2.95918$$

$$F_{24,29} (\alpha = 0.01) = 2.49$$

Como $2.95918 > 2.49$, rechazamos la hipótesis de que las varianzas sean iguales.

109. Una empresa de publicidad desea comprobar si un determinado programa de televisión es visto por el 30% de la audiencia. Para ello escogen al azar 200 familias y resulta que de ellas 50 lo están viendo. Contrastar la hipótesis con un nivel de significación del 5%.

SOLUCIÓN:

$$H_0: p = 0.3$$

$$H_1: p \neq 0.3$$

$$\hat{p} = 50/200 = 0.25 \quad \frac{\hat{p} - p_0}{\sqrt{p_0 p_0 / n}} = -1.5430$$

$$n = 200$$

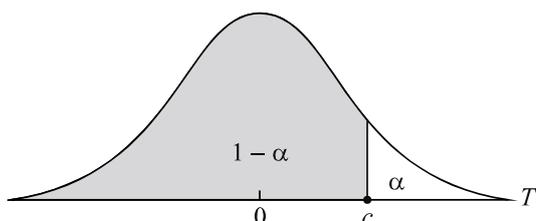
$$\alpha = 0.05 \quad \lambda_{\alpha/2} = 1.96$$

Como $-1.5430 \in [-1.96, 1.96]$ aceptamos la hipótesis de que sea visto por el 30% de la audiencia.

ANEXO

TABLA DE LA DISTRIBUCION t -Student

La tabla da áreas $1 - \alpha$ y valores $c = t_{1-\alpha, r}$, donde, $P[T \leq c] = 1 - \alpha$, y donde T tiene distribución t -Student con r grados de libertad...



$1 - \alpha$

r	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819

23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
□	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576