Lección 8: Estimación para intervalos de confianza

Sea X una población y $(X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño n con función de masa P_θ (o función de densidad f_θ), $\theta \in \Theta$. En la lección anterior queriamos dar un estimación de θ, en esta vamos a localizar el verdadero parámetro θ en un intervalo, que llamaremos ïntervalo de confinanza". Para este fín necesitamos introducir las siguientes distrubuciones asociadas a la normal:

- **Distribución χ^2 de Pearson**. Sean $X_1, X_2, \cdots, X_n$ n variables aleatorias independiente, siguiendo cada una una distrubución $N(0,1)$. La distribución χ^2 de Pearson con n grados de libertad, abreviadamente χ^2_n, es la distribución de la variable aleatoria

$$\Sigma_{i=1}^n X_i^2.$$

Propiedades:

1. Es una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & x \geq 0 \end{cases}$$

2. Esperanza y varianza

$$E[\chi^2_n] = n, \qquad V(\chi^2_n) = 2n.$$

- **t de Student**. Sean $Y, X_1, X_2, \cdots, X_n$ $n+1$ variables aleatorias independiente, siguiendo cada una una distrubución $N(0,1)$. La distribución t de Student con n grados de libertad, abreviadamente t_n, es la distribución de la variable aleatoria

$$\frac{Y}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}}.$$

Propiedades

1. Es una variable aleatoria continua con función de densidad

$$f(x) = \frac{2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n^3}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad x \in \mathbb{R}.$$

f es una función par, y por lo tanto simétrica con respecto al eje de ordenadas.

2. Esperanza y varianza

$$E[t_n] = 0, \qquad V(t_n) = \frac{n}{n-2}.$$

- **F de Fisher-Snedecor**. Sean $X_1, X_2, \cdots, X_m$, $Y_1, Y_2, \cdots, Y_n$ $m+n$ variables aleatorias independiente, siguiendo cada una una distrubución $N(0,1)$. La distribución F de Fisher-Snedecor de m y n grados de libertad, abreviadamente $F_{m;n}$, es la distribución de la variable aleatoria
$$\frac{\frac{1}{m}\sum_{i=1}^{m} X_i^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}.$$
Propiedades:
  1. Es una variable aleatoria continua con función de densidad
$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{\Gamma\left(\frac{m+n}{2}\right) n^{\frac{n}{2}} m^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{\frac{n}{2}-1}}{(nx+m)^{\frac{m+n}{2}}} & x \geq 0 \end{cases}$$
  2. Esperanza y varianza
$$E[F_{m;n}] = \frac{m}{m-2}, \quad m > 2; \qquad V(F_{m;n}) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}, \quad m > 4.$$

1. Intervalos de confianza en las poblaciones normales

1.1. X una población normal siguiendo $N(\mu, \sigma)$ con σ conocido y $(X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño n. Intervalo de confianza para μ a nivel $1-\alpha$

Ejercicio 1 *Se consideran los siguientes tiempos de reacción de un producto químico, en segundos*
$$x_1 = 1,4; x_2 = 1,2; x_3 = 1,2; x_4 = 1,3; x_5 = 1,5$$
$$x_6 = 1,3; x_7 = 2,2; x_8 = 1,4; x_9 = 1,1.$$
Obténgase un intervalo de confianza del $90\,\%$ *para el tiempo de reacción. Supóngase la variable normal con desviación típica poblacional conocida* $\sigma = 0,4$.

Universo: población de cierto producto químico.

Experimento aleatorio: reacción del producto químico.

Característica que queremos estudiar de de la población: tiempo de reacción del producto en segundos. Sea X la variable aleatoria que cuantifica esta característica. X sigue una $N(\mu, 0,4)$, y por lo tanto su función de densidad va a tener un parámetro desconocido, μ, la media poblacional, que toma valores dentro de un espacio paramétrico Θ.

Para este propósito, temamos una muestra aleatoria simple de tamaño 9, $(X_1, X_2, \cdots, X_9)$ que adopta el valor

$$(x_1 = 1,4; x_2 = 1,2; \cdots, x_9 = 1,1).$$

Un estimador por intervalos de confianza de μ, al nivel $1 - \alpha = 1 - 0,10 = 0,90$, es una función que a cada posible valor $(x_1 = 1,4; x_2 = 1,2; \cdots, x_9 = 1,1)$ de una muestra aleatoria simple $(X_1, X_2, \cdots, X_9)$ le hace corresponder un intervalo $(T_1, T_2) = (T_1(x_1, x_2, \cdots, x_9), T_2(x_1, x_2, \cdots, x_9))$, tal que pata todo $p \in \Theta$

$$P_\mu \left(T_1(X_1, X_2, \cdots, X_9) < \mu < T_2(X_1, X_2, \cdots, X_9) \right) \tag{1}$$

$$= 1 - \alpha = 1 - 0,10 = 0,90.$$

(1) no lo tenemos que interpretar como que existe una probabilidad de $0,90$ de que el parámetro μ tome un valor en $(T_1(X_1, X_2, \cdots, X_9), T_2(X_1, X_2, \cdots, X_9))$, sino de que $0,90$ es la probabilidad de que este intervalo incluya el verdadero valor del parámetro desconocido μ antes de tomar la muestra aleatoria simple $(X_1, X_2, \cdots, X_9)$ y que esta adopta el valor

$$(x_1 = 1,4; x_2 = 1,2; \cdots, x_9 = 1,1).$$

Cuando esto ocurre, $(T_1(X_1, X_2, \cdots, X_9), T_2(X_1, X_2, \cdots, X_9))$ se transforma en el intervalo concreto

$$\left(T_1(1,4; 1,2; \cdots, 1,1), T_2(1,4; 1,2; \cdots, 1,1) \right),$$

y entonces la probabilidad de que este intervalo contenga a μ será 1 si lo contiene o 0 si no lo contiene.

El principal problema de los intervalos de confianza es su construcción.

El primer paso será la construcción de estadísticos $T(X_1, X_2, \cdots, X_9, \mu)$que satisfagan algunas propiedades. En el caso que nos ocupa, **la población sigue una** $N(\mu; 0,4)$**, desviación típica conocida**, el estadístico

$$\boxed{T(X_1, X_2, \cdots, X_9, \mu) = \frac{\overline{X} - \mu}{\frac{0,4}{\sqrt{9}}}} \tag{2}$$

con $\overline{X} = \frac{1}{9} \sum_{i=1}^{9} X_i$ satisface

- $T(X_1, X_2, \cdots, X_9, \mu) = \frac{\overline{X} - \mu}{\frac{0,4}{\sqrt{9}}} \sim N(0; 1)$.
- $T(X_1, X_2, \cdots, X_9, \mu) = \frac{\overline{X} - \mu}{\frac{0,4}{\sqrt{9}}}$, es una cantidad **pivotal** para μ, es decir, la función de distribución de $\frac{\overline{X} - \mu}{\frac{0,4}{\sqrt{9}}}$ no depende de μ.
- $T(x_1, x_2, \cdots, x_9, \mu) = \frac{\overline{X} - \mu}{\frac{0,4}{\sqrt{9}}}$ es una cantidad continua y monótona de μ.

Estas propiedades nos van a permitir construir un intervalo (T_1, T_2), T_1 y T_2 independientes de μ ,(segunda propiedad), tal que

$$P\left(T_1 < \frac{\overline{X}-\mu}{\frac{0,4}{\sqrt{9}}} < T_2\right) = 0,90, \tag{3}$$

y a partir de aquí obtener el intervalo de confianza para μ a un 90 %.

¿Cómo obtener el intervalo (T_1, T_2) de tal manera que se satisfaga (3)?. Hay infinitas maneras de hacer esto, pero parecería lógico de que el intervalo de confianza debe de tener la menor amplitud posible, ya que esto aumentaría la precisión de μ, y esto conduce a que el intervalo (T_1, T_2) también debe tener la menor longitud posible.

Ya que $\frac{\overline{X}-\mu}{\frac{0,4}{\sqrt{9}}} \sim N(0;1)$, puede demostrarse que lo acabado de indicar se consigue para $T_1 = -z_{\alpha/2}$ y $T_2 = z_{\alpha/2}$. Así, $z_{\alpha/2}$ debe de satisfacer

$$P\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{0,4}{\sqrt{9}}} < z_{\alpha/2}\right) = 0,90, \tag{4}$$

o lo que es lo mismo

$$1 - 2P\left(\frac{\overline{X}-\mu}{\frac{0,4}{\sqrt{9}}} > z_{\alpha/2}\right) = 0,90 \Longleftrightarrow P\left(\frac{\overline{X}-\mu}{\frac{0,4}{\sqrt{9}}} > z_{\alpha/2}\right) = 0,05 \Longleftrightarrow z_{\alpha/2} = 1,645.$$

De (4) y para el valor que adopta la media muestral

$$\bar{x} = \frac{1,4 + 1,2 + \cdots + 1,1}{9} = 1,4,$$

tenemos

$$-z_{\alpha/2} < \frac{\overline{x}-\mu}{\frac{0,4}{\sqrt{9}}} < z_{\alpha/2},$$

$$-1,645 < \frac{1,4-\mu}{0,133} < 1,645 \Longleftrightarrow -0,218 < 1,4-\mu < 0,218$$

$$\text{Intervalo de comfianza para } \mu \text{ al } 90\,\% = \boxed{(1,19; 1,62)}$$

$\square$.

Ejercicio 2 *¿Cuál debe ser el mínimo tamaño de una muestra aleatoria de una población normal $N(\mu; 5)$ para que el error al estimar μ no sea superior a $0,5$ con un nivel de confianza del 95 %?.*

Estamos trabajando con una población $X \sim N(\mu; 5)$, con media poblacional μ desconocida y desviación poblacional $\sigma = 5$ conocida. Sea $(X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño y $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ su media muestral.

Se puede demostrar que

- El estadístico $\frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}} \sim N(0,1)$.
- $\frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}}$ es una cantidad pivotal para μ, esto es, su función de distribución no depende de μ.
- $\frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}}$ es una función continua y monótona de μ.

Vamos a calcular el intervalo de confianza a un nivel del 95 %.

Tenemos que determinar $z_{\alpha/2}$ de tal manera que

$$P\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}} < z_{\alpha/2}\right) = 0,025$$

$$1-2P\left(\frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}} > z_{\alpha/2}\right) = 0{,}95 \Longleftrightarrow P\left(\frac{\overline{X}-\mu}{\frac{5}{\sqrt{n}}} > z_{\alpha/2}\right) = 0{,}025 \Longleftrightarrow z_{\alpha/2} = 1,96.$$

El intervalo de confianza para μ al 95 % es

$$\bar{x} - z_{\alpha/2}\frac{5}{\sqrt{n}} < \mu < \bar{x} - z_{\alpha/2}\frac{5}{\sqrt{n}}$$

Definimos el error del intervalo de confianza al 95 % para μ por

$$\boxed{z_{\alpha/2}\frac{5}{\sqrt{n}}}$$

Vamps a determinar n para que este error sea menor que $0,5$.

$$1,96\frac{5}{\sqrt{n}} \leq 0,5 \Longleftrightarrow n \geq \left(\frac{1,96\cdot 5}{0,5}\right)^2 = 384,16.$$

Necesitaríamos observar por lo menos 385 elementos para conseguir la estimación deseada para esta estimación por intervalos de confianza. Este es un resultado teórico. Lo normal es que pensásemos que tenemos que hacer como 400 observaciones.

□

1.2. X una población normal siguiendo $N(\mu,\sigma)$ con parámetro poblacionales μ y σ desconocidos y $(X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño n. Intervalo de confianza para μ a nivel $1-\alpha$

Ejercicio 3 *Se mide el tiempo de duración (en segundos) de un proceso químico realizado* 20 *veces en condiciones similares, obteniéndose los siguientes resultados*

$$x_1 = 93, x_2 = 90, x_2 = 97, x_4 = 90, x_5 = 93, x_6 = 91, x_7 = 96$$

$$x_8 = 94, x_9 = 91, x_{10} = 91, x_{11} = 88, x_{12} = 93, x_{13} = 95, x_{14} = 91$$

$$x_{15} = 89, x_{16} = 92, x_{17} = 87, x_{18} = 88, x_{19} = 90, x_{20} = 86.$$

Suponiendo que la duración sigue una normal $N(\mu, \sigma)$, hallar el intervalo de confianza al 90 % para la μ.

Sea $X =$"duración en segundos de un proceso químico", que sabemos que sigue una $N(\mu, \sigma)$, pero desconocemos los parámetros poblacionales, en este caso, $\theta = (\mu, \sigma)$. Queremos estimar el parámetro $g_1(\theta) = \mu$, mediante un intervalo de confianza al nivel del 90 %. Tomamos una muestra aleatoria de tamaño 20, $(X_1, X_2, \cdots, X_{20})$ que adapta el valor

$$(x_1 = 93, x_2 = 90, \cdots, x_{20} = 86).$$

Un estimador por intervalos de confianza para μ, $\theta = (\mu, \sigma) \in \Theta$, al nivel $1 - \alpha = 1 - 0,10 = 0,90$, es una función que a cada posible realización de la muestra aleatoria $(x_1 = 93, x_2 = 90, \cdots, x_{20} = 86)$, le hace corresponder un intervalo $(T_1, T_2) = (T_1(93, 90, \cdots, 86), T_2(93, 90, \cdots, 86))$, tal que, para todo $\theta = (\mu, \sigma) \in \Theta$ se tiene

$$P\left(T_1(X_1, X_2, \cdots, X_{20}) < \mu < T_2(X_1, X_2, \cdots, X_{20})\right) \tag{5}$$

$$= 1 - \alpha = 1 - 0,10 = 0,90.$$

Para construir el intervalo de confianza necesitamos un estimador que satisfaga ciertas propiedades.

Si $\overline{X} = \frac{1}{20}\sum_{i=1}^{20} X_i$ es la media muestral y S es la raíz cuadrada de la cuasivarianza, $S^2 = \frac{1}{n-1}\sum_{i=1}^{20}(X_i - \overline{X})^2$, se puede demostrar que el estadístico

$$\frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{20}}} \tag{6}$$

satisface

- $\frac{\overline{X}-\mu}{\frac{\hat{S}}{\sqrt{20}}} \sim t_{19}$.
- $\frac{\overline{X}-\mu}{\frac{\hat{S}}{\sqrt{20}}}$ es una cantidad pivotal, esto es que la función de distribución de la variable $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{20}}}$ no depende se $\theta = (\mu, \sigma)$.
- $\frac{\overline{x}-\mu}{\frac{\hat{s}}{\sqrt{20}}}$ es una función continua y monótona de μ.

Vamos a encontrar un número $z_{\alpha/2}$ tal que se satisfaga

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{20}}} < z_{\alpha/2}\right) = 0{,}9 \tag{7}$$

Por la segunda propiedad, $z_{\alpha/2}$ es independiente de $\theta = (\mu, \sigma)$ y por lo tanto de μ.

Por las propiedades de la t_{19}, (7) es equivalente a

$$1 - 2P\left(\frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{20}}} > z_{\alpha/2}\right) = 0{,}9 \Leftrightarrow P\left(\frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{20}}} > z_{\alpha/2}\right) = 0{,}05.$$

Si vamos a las tablas, obtenemos que

$$z_{\alpha/2} = 1{,}729,$$

entonces para este $z_{\alpha/2}$ podemos asegurar que

$$P\left(-1{,}729 < \frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{20}}} < 1{,}729\right) = 0{,}9,$$

y el intervalo de confianza al 90 % lo obtenemos despejando μ en la realización de la media y cuasivarianza muestral,

$$-1{,}729 < \frac{\overline{x} - \mu}{\frac{\hat{s}}{\sqrt{20}}} < 1{,}729,$$

$$-1{,}729 \cdot \frac{\hat{s}}{\sqrt{20}} < \overline{x} - \mu < 1{,}729 \cdot \frac{\hat{s}}{\sqrt{20}},$$

$$-\overline{x} - 1{,}729 \cdot \frac{\hat{\overline{s}}}{\sqrt{20}} < -\mu < -\overline{x} + 1{,}729 \cdot \frac{\hat{\overline{s}}}{\sqrt{20}},$$

$$\overline{x} - 1{,}729 \cdot \frac{\hat{s}}{\sqrt{20}} < \mu < \overline{x} + 1{,}729 \cdot \frac{\hat{s}}{\sqrt{20}}.$$

$$\bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = 91'25, \qquad \hat{s}^2 = \frac{1}{19}\sum_{i=1}^{20} x_i^2 - \frac{20}{19}\bar{x} = 8'6182.$$

El intervalo de confianza para μ al 90 % es $\boxed{(90; 11; 92, 39)}$.

$\square$

1.3. X una población normal $N(\mu, \sigma)$ con parámetros poblacionales μ y σ desconocidos y $(X_1, X_2, \cdots, X_n)$ una muestra aleatoria de tamaño n. Intervalo de confianza para σ a nivel $1 - \alpha$

Ejercicio 4 *Se mide el tiempo de duración (en segundos) de un proceso químico realizado* 20 *veces en condiciones similares, obteniéndose los siguientes resultados*

$$x_1 = 93, x_2 = 90, x_2 = 97, x_4 = 90, x_5 = 93, x_6 = 91, x_7 = 96$$

$$x_8 = 94, x_9 = 91, x_{10} = 91, x_{11} = 88, x_{12} = 93, x_{13} = 95, x_{14} = 91$$

$$x_{15} = 89, x_{16} = 92, x_{17} = 87, x_{18} = 88, x_{19} = 90, x_{20} = 86.$$

Suponiendo que la duración sigue una normal $N(\mu, \sigma)$, hallar el intervalo de confianza al 90 % para la σ.

Sea $X =$"duración en segundos de un proceso químico", que sabemos que sigue una $N(\mu, \sigma)$, pero desconocemos los parámetros. En este caso, $\theta = (\mu, \sigma)$.

Sea, $g_2(\theta) = \sigma$, que es el parámetro poblacional que queremos estimar por un intervalo de confianza. Para ello tomamos una muestra aleatoria simple de tamaño 20 $(X_1, X_2, \cdots; X_{20})$ y que adopta el valor

$$(x_1 = 93, x_2 = 90, \cdots, x_{20} = 86).$$

Un estimador por intervalos de confianza para σ, $\theta = (\mu, \sigma) \in \Theta$, al nivel $1 - \alpha = 1 - 0,10 = 0,90$, es una función que a cada posible realización de la muestra aleatoria $(x_1 = 93, x_2 = 90, \cdots, x_{20} = 86)$, le hace corresponder un intervalo $(T_1, T_2) = (T_1(93, 90, \cdots, 86), T_2(93, 90, \cdots, 86))$, tal que, para todo $\theta = (\mu, \sigma) \in \Theta$ se tiene

$$P\left(T_1(X_1, X_20, \cdots, X_{20}) < \sigma < T_2(X_1, X_2, \cdots, X_{20})\right) \tag{8}$$

$$= 1 - \alpha = 1 - 0,10 = 0,90.$$

Para construir el intervalo de confianza necesitamos un estimador que satisfaga ciertas propiedades.

Si $\overline{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$ es la media muestral y S es la cuasivarianza, $\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^{20} (X_i - \overline{X})^2$, se puede demostrar que el estadístico

$$\frac{20-1}{\sigma^2} \hat{S}^2 \tag{9}$$

satisface

- $\frac{20-1}{\sigma^2} \hat{S}^2 \sim \chi^2_{20-1}$.
- $\frac{20-1}{\sigma^2} \hat{S}^2$ es una cantidad pivotal, esto es que la función de distribución de la variable aleatoria $\frac{20-1}{\sigma^2} S^2$ no depende se $\theta = (\mu, \sigma)$.
- $\frac{20-1}{\sigma^2} \hat{s}^2$ es una función continua y monótona de σ.

Vamos a encontrar dos un números $z^1_{\alpha/2} < z^2_{\alpha/2}$ que satisfagan

$$P\left(z^1_{\alpha/2} < \frac{19}{\sigma^2} \hat{S}^2\right) = 0{,}95, \tag{10}$$

$$P\left(\frac{19}{\sigma^2} \hat{S}^2 < z^2_{\alpha/2}\right) = 0{,}95, \tag{11}$$

y por lo tanto

$$P\left(z^1_{\alpha/2} < \frac{19}{\sigma^2}\hat{S}^2 < z^2_{\alpha/2}\right) = 0{,}9 \tag{12}$$

De las propiedades de la χ^2_{19} se obtiene que $z^1_{\alpha/2} = 10,12$, que (11) es equivalente a

$$P\left(z^2_{\alpha/2} < \frac{19}{\sigma^2}\hat{S}^2\right) = 0{,}05,$$

y por lo tanto $z^2_{\alpha/2} = 30,1$.

De (12) y sabiendo que $\hat{s}^2 = 8,6182$

$$10,12 < \frac{19}{\sigma^2}\overline{s}^2 < 30,1 \Longleftrightarrow \frac{19}{30,1}\hat{s}^2 < \sigma^2 < \frac{19}{10,12}\hat{s}^2$$

$$5,44 < \sigma^2 < 16,18 \Longleftrightarrow 2,33 < \sigma < 4,02$$

Intervalo de confianza para σ al nivel de 90 % = $\boxed{(2,33; 4,02)}$

□

1.4. Dos poblaciones normales independientes: $X \sim N(\mu_1;\sigma)$ e $Y \sim N(\mu_2;\sigma)$ con la misma desviación típica poblacional. Intervalo de confianza al nivel $1-\alpha$ para $\mu_1 - \mu_2$

Ejercicio 5 *Se quieren comparar dos poblaciones X e Y para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos, (en calorías por gramo de masa), para pasar de $-0,72^oC$ a O^o usando reiterádamente los modelos característicos de cada población:*

- *Población X:*

$$x_1 = 79,98; x_2 = 80,04; x_3 = 80,02; x_4 = 80,04; x_5 = 80,03;$$

$$x_6 = 80,03; x_7 = 80,04; x_8 = 79,97; x_9 = 80,05; x_{10} = 80,03;$$

$$x_{11} = 80,02; x_{12} = 80,00; x_{13} = 80,02$$

- *Población Y:*

$$y_1 = 80,02; y_2 = 79,94; y_3 = 79,98; y_4 = 79,97;$$

$$y_5 = 79,97; y_6 = 80,03; y_7 = 79,95; y_8 = 79,97$$

Obtener un intervalo de confianza al 95 % para la diferencia de las medias poblacionales.

Vamos a suponer que ambas poblaciones son normales y que tienen la misma desviación típica poblacional.

La población $X \sim N(\mu_1, \sigma)$. Tomamos una muestra aleatoria de tamaño 13 $(X_1, X_2, \cdot\cdot\cdot, X_{13})$ con realización

$$(x_1 = 79,98, x_2 = 80,04, \cdot\cdot\cdot, x_{13} = 80,02).$$

Designamos por $\overline{X}$ y $\hat{S}_1^2$ su media y cuasivarianza muestral.

La población $Y \sim N(\mu_2, \sigma)$. Tomamos una muestra aleatoria de tamaño 8 $(Y_1, Y_2, \cdot\cdot\cdot, Y_8)$ con realización

$$(y_1 = 80,02, y_2 = 79,94, \cdot\cdot\cdot, y_8 = 79,97).$$

Designamos por $\overline{Y}$ y $\hat{S}_2^2$ su media y cuasivarianza muestral.

Queremos estimar el parámetro $\mu_1 - \mu_2$ por in intervalo de confianza.

Un estimador por intervalos de confianza para $\mu_1 - \mu_2$ al nivel $1 - \alpha = 1 - 0,05 = 0,95$, es una función que a cada par de posibles realizaciones de las muestras aleatorias

$$(x_1 = 79,98, x_2 = 80,04, \cdot\cdot\cdot, x_{13} = 80,02) \quad \text{y} \quad (y_1 = 80,02, y_2 = 79,94, \cdot\cdot\cdot, y_8 = 79,97),$$

le hace corresponder un intervalo (T_1, T_2)

$$P\left(T_1(X_1, \cdot\cdot, X_{13}, Y_1, \cdot\cdot, Y_8) < \mu_1 - \mu_2 < T_2(X_1, \cdot\cdot, X_{13}, Y_1, \cdot\cdot, Y_8)\right) \tag{13}$$

$$= 1 - \alpha = 1 - 0,05 = 0,95.$$

Para construir el intervalo de confianza necesitamos un estimador que satisfaga ciertas propiedades.

Se puede demostrar que el estadístico

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{13} + \frac{1}{8}}}, \tag{14}$$

donde

$$S_p^2 = \frac{(13-1)\hat{S}_1^2 + (8-1)\hat{S}_2^2}{13 + 8 - 2},$$

satisface

- $\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{13} + \frac{1}{8}}} \sim t_{13+8-2}$.
- $\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{13} + \frac{1}{8}}}$ es una cantidad pivotal, esto es que la función de distribución de la variable aleatoria $\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{13} + \frac{1}{8}}}$ no depende se μ_1, μ_2 y σ.

- $\frac{\overline{x}-\overline{y}-(\mu_1-\mu_2)}{\overline{s}_p\sqrt{\frac{1}{13}+\frac{1}{8}}}$ es una función continua y monótona de $\mu_1-\mu_2$.

Para obtener el intervalo de confianza, puede comprobarse que los valores muestrales son:

$$\overline{x}=80,02;\quad \overline{y}=79,97;\quad \hat{s}_1^2=0,13;\quad \hat{s}_2^2=1,6;\quad s_p=0,82.$$

Vamos a determinar $z_{\alpha/2}$ verificando

$$P\left(-z_{\alpha/2}<\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{13}+\frac{1}{8}}}<z_{\alpha/2}\right)=0{,}95. \tag{15}$$

Por las propiedades de la función de densidad de la t de Student, (15) es equivalente a

$$P\left(\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{13}+\frac{1}{8}}}>z_{\alpha/2}\right)=0{,}025,$$

y utilizando las tablas obtenemos $z_{\alpha/2}=2,093$.

De (15)

$$-z_{\alpha/2}<\frac{\overline{x}-\overline{y}-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{13}+\frac{1}{8}}}<z_{\alpha/2}$$

$$\Downarrow$$

$$\overline{x}-\overline{y}-z_{\alpha/2}s_p\sqrt{\frac{1}{13}+\frac{1}{8}}<\mu_1-\mu_2<\overline{x}-\overline{y}+z_{\alpha/2}s_p\sqrt{\frac{1}{13}+\frac{1}{8}},$$

sustituyendo, obtenemos que el intervalo de confianza al nivel $0,95$ para la diferencia de las medias poblacionales $\mu_1-\mu_2$ es $\boxed{(-0,72;0,82)}$.

□

1.5. Dos poblaciones normales independientes: $X\sim N(\mu_1;\sigma_1)$ e $Y\sim N(\mu_2;\sigma_2)$. Intervalo de confianza al nivel $1-\alpha$ para $\frac{\sigma_1^2}{\sigma_2^2}$

Ejercicio 6 *En un estudio sobre el tiempo de desarrollo de una especie de insectos en dos poblaciones aisladas X e Y, se obtuvieron los siguientes datos:*

Población $X\sim N(\mu_1,\sigma_1)$	$(X_1,X_2,\cdots,X_{13})$	$\overline{x}=4$	$\overline{s}_1=3$
Población $Y\sim N(\mu_2,\sigma_2)$	$(Y_1,Y_2,\cdots,Y_{11})$	$\overline{y}=5$	$\overline{s}_2=2,2$

1. *Hallar un intervalo de confianza para el cociente de las varianzas al nivel* $0,90$.

2. *Obtener un intervalo de confianza para $\mu_1 - \mu_2$, con un nivel de confianza de $0,95$, (suponiendo igualdad de varianzas).*

3. *¿Cuántos individuos habría que observar para estimar μ_1 con un error máximo de $\pm 0,2$ y un nivel de confianza de $0,95$?.*

1. La población $X \sim N(\mu_1, \sigma_1)$. Tomamos una muestra aleatoria de tamaño 13 $(X_1, X_2, \cdots, X_{13})$, con realización $(x_1, x_2, \cdots, x_{13})$ tal que $\overline{x} = 4$ y $\overline{s}_1 = 3$.

   La población $Y \sim N(\mu_2, \sigma_2)$. Tomamos una muestra aleatoria de tamaño 11 $(Y_1, Y_2, \cdots, Y_8)$, con realización $(y_1, y_2, \cdots, y_{11})$ tal que $\overline{y} = 5$ y $\overline{s}_2 = 2,2$. Queremos estimar el parámetro $\frac{\sigma_1^2}{\sigma_2^2}$ por in intervalo de confianza.

   Un estimador por intervalos de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$ al nivel $1 - \alpha = 1 - 0,10 = 0,90$, es una función que a cada par de posibles realizaciones de las muestras aleatorias

$$(x_1, x_2, \cdots, x_{13}) \quad \text{y} \quad (y_1, y_2, \cdots, y_{11}),$$

   le hace corresponder un intervalo (T_1, T_2) tal que

$$P\left(T_1(X_1, \cdot\cdot, X_{13}, Y_1, \cdot\cdot, Y_{11}) < \frac{\sigma_1^2}{\sigma_2^2} < T_2(X_1, \cdot\cdot, X_{13}, Y_1, \cdot\cdot, Y_{11})\right) \tag{16}$$

$$= 1 - \alpha = 1 - 0,10 = 0,90.$$

Para construir el intervalo de confianza necesitamos un estimador que satisfaga ciertas propiedades.

Se puede demostrar que el estadístico

$$\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2}, \tag{17}$$

satisface

- $\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\hat{S}_1^2}{\hat{S}_2^2} \sim F_{13-1,11-1}$.
- $\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\hat{S}_1^2}{\hat{S}_2^2}$ es una cantidad pivotal, esto es que la función de distribución de la variable aleatoria $\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\hat{S}_1^2}{\hat{S}_2^2}$ no depende se μ_1, μ_2, σ_1 y σ_2.
- $\frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\hat{s}_1^2}{\hat{s}_2^2}$ es una función continua y monótona de $\frac{\sigma_2^2}{\sigma_1^2}$.

Vamos a encontrar dos un números $z_\alpha^1 < z_\alpha^2$ que satisfagan

$$P\left(z_\alpha^1 < \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\hat{S}_1^2}{\hat{S}_2^2}\right) = 0{,}95, \tag{18}$$

$$P\left(\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\hat{S}_1^2}{\hat{S}_2^2} < z_\alpha^2\right) = 0{,}95, \tag{19}$$

y por lo tanto

$$P\left(z_\alpha^1 < \frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\hat{S}_1^2}{\hat{S}_2^2} < z_\alpha^2\right) = 0{,}9 \tag{20}$$

(19) es equivalente a

$$P\left(\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\hat{S}_1^2}{\hat{S}_2^2} > z_\alpha^2\right) = 0{,}05 \implies z_\alpha^2 = 2,91.$$

(18) es equivalente a

$$P\left(\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\hat{S}_1^2}{\hat{S}_2^2} < z_\alpha^1\right) = 0{,}05,$$

$$P\left(F_{12,10} < z_\alpha^1\right) < 0{,}05 \iff P\left(\frac{1}{z_\alpha^1} < \frac{1}{F_{12,10}}\right) \iff P\left(F_{10,12} > \frac{1}{z_\alpha^1}\right) = 0,05$$

$$\iff \frac{1}{z_\alpha^1} = 2,75 \iff z_\alpha^1 = 0{,}3636$$

De (20)

$$z_\alpha^1 < \frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\hat{s}_1^2}{\hat{s}_2^2} < z_\alpha^2,$$

$$\frac{\hat{s}_1^2}{\hat{s}_2^2}\cdot\frac{1}{z_\alpha^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{\hat{s}_1^2}{\hat{s}_2^2}\cdot\frac{1}{z_\alpha^1},$$

sustituyendo es esta desigualdad, obtenemos que el intervalo de confianza para estimar $\frac{\sigma_2^2}{\sigma_1^2}$ al nivel $0,90$ es $= \boxed{(0,639; 5,469)}$.

2. Del apartado anterior, el intervalo de confianza para $\frac{\sigma_1}{\sigma_2}$ al 90 % es $(0,8; 2,34)$. Ya que $1 \in (0,8; 2,34)$, parece razonable suponer en este apartado igualdad de desviaciones típicas para las dos poblaciones.

   Tenemos que dar un intervalo de confianza al nivel $0,95$ para la diferencia de las medias poblacionales $\mu_1 - \mu_2$.

   Utilizamos el estadístico

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{13}+\frac{1}{11}}} \sim t_{13+11-2},$$

   donde $S_p^2 = \frac{(13-1)\hat{S}_1^2 + (11-1)\hat{S}_2^2}{13+11-2}$. El $z_{\alpha/2}$ que satisface

$$P\left(-z_{\alpha/2} < t_{22} < z_{\alpha/2}\right) = 0,95, \tag{21}$$

es $z_{\alpha/2} = 2,074$.

Se puede comprobar que $s_p = 2,67$. De (21)

$$-z_{\alpha/2} < \frac{\overline{x} - \overline{y} - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{13} + \frac{1}{11}}} < z_{\alpha/2},$$

$$\overline{x} - \overline{y} - z_{\alpha/2}s_p\sqrt{\frac{1}{13} + \frac{1}{11}} < \mu_1 - \mu_2 < \overline{x} - \overline{y} + z_{\alpha/2}s_p\sqrt{\frac{1}{13} + \frac{1}{11}},$$

y el intervalo de confianza para $\mu_1 - \mu_2$ al 95 % para las realizaciones de las muestras aleatorias de las poblaciones dadas es $(-3,27; 1,27)$.

3. Supongamos que tomamos una muestra aleatoria de tamaño n. Vamos a suponer que este n es grande. Para estimar por intervalos de confianza la media poblacional μ_1 de la población X, vamos a utilizar el estadístico

$$\frac{\overline{X} - \mu_1}{\frac{\hat{S}_1}{\sqrt{n}}} \sim t_{n-1}.$$

que nos dá el intervalo de confianza al nivel del 95 %:

$$\overline{x} - z_{\alpha/2} \cdot \frac{\hat{s}_1}{\sqrt{n}} < \mu_1 < \overline{x} + z_{\alpha/2} \cdot \frac{\hat{s}_1}{\sqrt{n}},$$

y el error está viene dado por

$$\text{Error al 95 \% del intervalo de confianza} = \boxed{z_{\alpha/2} \cdot \frac{\hat{s}_1}{\sqrt{n}}}$$

En esta expresión $z_{\alpha/2}$ está definido por la condición

$$P(-z_{\alpha/2} < t_{n-1} < z_{\alpha/2}) = 0{,}95.$$

Tenemos que pensar en n grande, entonces para buscar $z_{\alpha/2}$ en las tablas, vamos a tomar como $n - 1 = \infty$. Si hacemos esto, obtenemos $z_{\alpha/2} = 1,96$.

En la expresión del error tampoco conocemos $\overline{s}_1$. Tomamos el $\overline{s}_1$ correspondiente a la muestra piloto.

$$\text{Error} = 1,96 \cdot \frac{3}{\sqrt{n}} \leq 0,2 \Longleftrightarrow n \geq \left(1,96 \cdot \frac{3}{0,2}\right)^2 = 864,36.$$

Por lo tanto, necesitamos del orden de 865 observaciones para estimar μ_1 con un error de $0,2$ y una confianza de $0,95$.

□

2. Intervalos de confianza en las poblaciones no normales

Cuando trabajamos con poblaciones no normales, por ejemplo Bernoulli o Poisson, para obtener intervalos de confianza por los métodos de la sección anterior, tenemos que trabajar con muestras grandes. Si el tamaño n tiende a infinito, la cantidad pivotal que se utiliza tiene una distribución límite independiente de los parámetros desconocidos. Desde un punto de vista práctico, traducimos que n es grande si $n \geq 30$.

2.1. Intervalo de confianza, al nivel $1-\alpha$, para la media de una población cualquiera con desviación típica conocida

Ejercicio 7 *Se toman 35 probetas de un cierto hormigón cuya resistencia a los 28 dias en N/mm^2 es la siguiente:*

29,2	*29*	*30,4*	*30,6*	*28,4*	*33,1*	*26,9*
26,7	*28*	*30,8*	*31,2*	*27*	*31,1*	*29,5*
29,2	*28,9*	*28,9*	*33,2*	*33,5*	*29,3*	*33,7*
28,4	*32,8*	*32,6*	*30,1*	*29,9*	*28,1*	*28,1*
33,9	*30,2*	*32*	*27,1*	*27,2*	*27,7*	*27,8*

Calcule, sabiendo que la desviación típica de la población es $4N/mm^2$, la media muestral y un lḿite superior de su valor al $5\,\%$ de significación;

la varianza muestral y un lḿite superior de su valor al $e\,\%$ de significación.

No sabemos con que tipo de población estamos trabajando. Conocemos la desviación típica poblacional y queremos estimar por intervalos de confianza al nivel $0,95$ la media poblacional. Tenemos una muestra aleatoria simple $(X_1, X_2, \cdots, X_{35})$ grande, ya que $n = 35 \geq 30$. Se comprueba que el estimador

$$\frac{\overline{X}-\mu}{\frac{4}{\sqrt{35}}}$$

tiene una función de distribución que se aproxima bien por una $N(0;1)$.

El número $z_{\alpha/2} = 1,96$ satisface que

$$P(-1,96 < N(0,1) < 1,96) = 0{,}95,$$

entonces podemos pensar que

$$P\left(-1,96 < \frac{\overline{X}-\mu}{\frac{4}{\sqrt{35}}} < 1,96\right)$$

estará bastante próximo a $0,95$. A efectos prácticos, tomamos

$$P\left(-1,96 < \frac{\overline{X}-\mu}{\frac{4}{\sqrt{35}}} < 1,96\right) = 0,95. \tag{22}$$

Un cálculo nos da $\overline{x} = 29,841$. Si utilizamos (22)

$$-1,96 < \frac{\overline{x}-\mu}{\frac{4}{\sqrt{35}}} < 1,96,$$

$$\overline{x} - 1,96 \cdot \frac{4}{\sqrt{35}} < \mu < \overline{x} - 1,96 \cdot \frac{4}{\sqrt{35}} \iff 28,51 < \mu < 31,16$$

y el intervalo de confianza al nivel $0,95$ para la media poblacional es= $\boxed{(28,51; 31,16)}$.

$\square$

Ejercicio 8 *Se ha tomado una muestra aleatoria de 100 individuos a los que se ha medido el nivel de glucosa en sangre, obteniéndose una media muestral de 110 mg/cc. Se sabe que la desviación típica de la población es de 20 mg/cc.*

1. *Obtener un intervalo de confianza, al 90 %, para el nivel medio de glucosa en sangre en la población.*
2. *¿Qué error se comete con la estimación anterior?.*

1. La Población X no sabemos qé distribución de probabilidad sigue. Sabemos que la desviación típica poblacional es de 20 mg/cc y que la muestra es de tamaño grande, $100 \geq 30$, entonces la distribución del estadístico $\frac{\overline{X}-\mu}{\frac{20}{\sqrt{100}}}$ se apróxima bien por la de una normal $N(0,1)$.

   Sea $z_{\alpha/2} = 1,64$ tal que

   $$P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 0,90$$

   y pensamos que

   $$P(-z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{20}{\sqrt{100}}} < z_{\alpha/2}) = 0,90.$$

   De esto

   $$-z_{\alpha/2} < \frac{\overline{x}-\mu}{\frac{20}{\sqrt{100}}} < z_{\alpha/2},$$

   y el intervalo de confianza al 90 % el nivel medio de glucosa de la población es $(106,72; 113,28)$.

2. El error en el intervalo de confianza al 90 % el nivel medio de glucosa de la población es

$$\text{Error} = z_{\alpha/2}\frac{20}{\sqrt{100}} = 3,24.$$

Comentarios: Hemos obtenido al 90 % un intervalo de confianza $(106,72; 113,28)$. Esto significa que tenemos una probabilidad del 90 % de que dicho intervalo contenga la media real y desconocida μ de glucosa en sangre de la población. Dicho de otra forma, si contruimos de esta forma 100 intervalos, en promedio, 95 de ellos contendrán el valor medio μ real y desconocido de glucosa en sangre en la población.

El intervalo de confianza para la media μ es

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

El error

$$\text{Error} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

¿Cómo podemos hacer más pequeño el error? Será más pequeño, conforme sea mas grande el tamaño de la muestra aleatoria simple.

2.2. Intervalo de confianza para una distribución de Bernoulli

Ejercicio 9 *En una población se desea conocer la probabilidad de que un individuo sea alérgico al polen de las acacias. En 100 individuos tomados al azar se observaron 10 alérgicos. Hallar el intervalo de confianza al* 95 % *para la probabilidad pedida. ¿Cuántos individuos se deberían observar para que con probabilidad de* $0,95$*, el error máximo en la estimación de alérgicos sea del* $0,01$*?*

El universo Ω son los individuos de una cierta población.

Experimento aleatoria: Tomar un individuo de esa población y ver si es alérgico al polen.

Definimos la variable aleatoria

$$X = \begin{cases} 1 & \text{el individuo es alérgico al polen} \\ 0 & \text{en caso contrario} \end{cases}$$

La variable aleatoria es discreta, su rango es $\{0,1\}$ con función de masa

$$P_p(X = n) = p^n(1-p)^{1-n}, \qquad n = 0,1.$$

P_p depende de un parámetro p, desconocido, (que toma valores dentro de un espacio paramétrico Θ), que es el que queremos estimar. En este caso, la población X es de Bernoulli.

Para este objetivo,tomamos una muestra aleatoria de tamaño 100, $(X_1, X_2, \cdots, X_{100})$.

- Las X_i, $i = 1, 2, \cdots, 100$, son independientes
- Para $i = 1, 2, \cdots, 100$,

$$P_p(X_i = n) = p^n(1-p)^{1-n}, \qquad n = 0, 1.$$

la realización de la muestra aleatoria nos ha dado $(x_1, x_2, \cdots, x_{100})$, donde 10 componentes de este vector de $\mathbb{R}^{100}$ son iguales a 1 y las 90 restantes iguales a 0. $n = 100$ es bastante grande desde el punto de vista práctico, $(100 > 30)$.

El estadístico

$$\frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X}}{100}}}$$

tiene un distribución límite que se apróxima bien por una $N(0,1)$. El número $z_{\alpha/2} = 1,96$ satisface que

$$P(-1,96 < N(0,1) < 1,96) = 0,95,$$

entonces pensaremos que

$$P\left(-1,96 < \frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X}}{100}}} < 1,96\right) = 0,95 \tag{23}$$

es cierto.

$\bar{x} = \frac{10}{100} = \frac{1}{10}$. De (23)

$$-1,96 < \frac{\overline{x} - p}{\sqrt{\frac{\overline{x}(1-\overline{x})}{100}}} < 1,96$$

$$\overline{x} - 1,96\sqrt{\frac{\overline{x}(1-\overline{x}}{100}} < p < \overline{x} + 1,96\sqrt{\frac{\overline{x}(1-\overline{x})}{100}} \iff 0,04 < p < 0,16 \tag{24}$$

y el intervalo de confianza para p al nivel $0,95$ es= $\boxed{(0,04; 0,16)}$.

Si sustituimos en (24) 100 por n obtenemos

$$\overline{x} - 1,96\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} < p < \overline{x} + 1,96\sqrt{\frac{\overline{x}(1-\overline{x})}{n}},$$

definimos

$$\text{Error en el intervalo de confianza al } 95\,\% \text{ para } p = \boxed{1,96\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}}$$

En el error desconocemos $\overline{x}$, pero vamos a tomar el que obtuvimos de la prueba piloto.

$$\text{Error} = 1,96\sqrt{\frac{0,1(1-0,1)}{n}} \leq 0,01 \iff n \geq 3458.$$

Luego para garantizar que el error máimo en la estimación de la proporción de alérgicos sea del 0,01, tendremos que tomar del orden de 3460 individuos.

$\square$

2.3. Intervalo de confianza para una distribución de Poisson

Ejercicio 10 *Se supone que el número de erratas por página en un libro sigue una distribución de Poisson. Elegidas al azar 95 páginas de un libro de 1266 páginas, se obtuvieron los siguientes resultados*

Número de erratas	*0*	*1*	*2*	*3*	*4*	*5*
Número de páginas	*40*	*30*	*15*	*7*	*2*	*1*

Hallar el intervalo de confianza al 90 % *para el número medio de erratas por página en todo el libro.*

El universo: 1266 páginas del libro.

Experimento aleatorio: tomar una página y ver el número de páginas.

Sea X la variable aleatoria que da el número de erratas en una página. X es una variable aleatoria discreta que sigue una distribución de Poisson $P(\lambda)$ de parámetro λ desconocido y con función de masa

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}, \qquad k=0,1,2,cdot\cdots.$$

Ya que $\lambda = E[X]$, el enunciado nos dice que estimemos este parámetro por un intervalos de confianza al 90 %.

Para este objetivo, tomamos una muestra aleatoria de tamaño 95 $(X_1, X_2, \cdots, X_{95})$, con realización $(x_1, x_2, \cdots, x_{95})$. De este vector de $\mathbb{R}^{95}$, 40 componentes son 0, 30 son 1, 15 son 2, 7 son 3, 2 son 4 y una componente es igual a 5.

El tamaño de la muestra aleatoria es grande. Se puede comprobar que la función de distribución del estadístico

$$\frac{\overline{X}-\lambda}{\frac{\overline{X}}{95}}$$

se aproxima bien por la de una $N(0,1)$.

Se puede comprobar que

$$P(-1,645 < N(0,1) < 1,645) = 0,90$$

pensamos entonces que

$$P\left(-1,645 < \frac{\overline{X}-\lambda}{\frac{\overline{X}}{95}} < 1,645\right) = 0,90 \tag{25}$$

es cierto.

Puede comprobarse que $\overline{x} = 0,989$. De (25)

$$-1,645 < \frac{\overline{x}-\lambda}{\frac{\overline{x}}{95}} < 1,645$$

$$\overline{x} - 1,645\frac{\overline{x}}{95} < p < \overline{x} + 1,645\frac{\overline{x}}{95} \iff 0,82 < p < 1,16$$

y el intervalo de confianza al 90 % para λ es= $\boxed{(0,82; 1,16)}$.

$\square$

3. Ejercicios

Ejercicio 11 *Calcular un intervalo de confianza al nivel* $1 - \alpha = 1 - 0,001 = 0,999$ *para el peso exacto mediante los resultados obtenidos con 10 básculas:*

$$7,2;\ 7,01;\ 7,36;\ 6,91;\ 7,22;\ 7,03;\ 7,11;\ 7,12;\ 7,03;\ 7,05.$$

<u>Indicación:</u> Supóngase que la población es normal.

<u>Solución:</u> $(6,9096; 7,2984)$.

Ejercicio 12 *Calcule qué tamaño muestral debemos tomar para obtener* μ *con una precisión de 0,001 a partir de una muestra de una población* $N(\mu, 3)$.

<u>Solución:</u> Del orden de 34574400.

Ejercicio 13 *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 al* 95 %. *Estudiar si dicha proporción puede considerarse igual a la proporción de fumadores en la población si ésta es de un* 29 %.

El intervalo de confianza es $(0,250023; 0,416643)$.

Si consideramos la proporción de fumadores en la población, $q = 0,29$, observamos que esta proporción se encuentra dentro del intervalo de confianza obtenido para un nivel de 0,95; 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 95 %, podemos considerar que esta proporción de fumadores en los enfermos de cáncer se corresponde con la de los fumadores en la población.

$\square$

Ejercicio 14 *Una noticia en el periódico dice que , de 1000 personas encuestadas sobre una cuestión, 556 se muestran a favor y 444 en contra, y concluye afirmando que el* 55,6 % *de la población se muestra a favor, con un margen de error de* ±3 %. *¿Cuál es el nivel de confianza de esta afirmación?*

Universo: personas de una cierta provincia, por ejemplo.

Experimento aleatorio: tomar una persona de esa provincia y preguntarle si está a favor o en contra de cierta cuestión.

Consideramos la variable aleatoria

$$X = \begin{cases} 1 & \text{si la persona está a favor de la cuestión} \\ 0 & \text{en caso contrario} \end{cases}$$

La variable aleatoria es discreta, con rango $\{0,1\}$ y función de masa

$$P(X = n) = p^n(1-p)^{1-n} \qquad n = 0, 1.$$

Desconocemos es parámetro p, y el periódico nos está hablando sobre la estimación del parámetro p con un intervalo de confianza, y queremos saber a qué nivel ha calculado el intervalo de confianza. Estamos ante una población X de Bernoulli.

El periódico ha tomado una muestra aleatoria simple de tamaño 1000 $(X_1, X_2, \cdots, X_{1000})$ con realización $(x_1, x_2, \cdots, x_{1000})$ y de las 1000 componentes de este vector, 556 son unos y 444 son ceros. El tamaño de la muestra es grande. Si $\overline{X} = \frac{1}{1000}\sum_{i=1}^{1000} X_1$ es la media muestral, el estadístico

$$\frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{1000}}}$$

tiene una función de distribución próxima a la de una normal $N(0,1)$. Para $\alpha \in (0,1)$, sea $z_{\alpha/2}$ tal que

$$P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 1 - \alpha \tag{26}$$

Pensamos que este $z_{\alpha/2}$ satisface

$$P(-z_{\alpha/2} < \frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{1000}}} < z_{\alpha/2}) = 1 - \alpha \tag{27}$$

De los datos que suministra el periódico, $\overline{x} = \frac{556}{1000}$. De (27) el intervalo de confianza para p al nivel $1 - \alpha$ es

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{1000}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{1000}}\right),$$

y el error de este intervalo de confianza para p al nivel $1-\alpha$ es

$$\text{Error} = z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{1000}} = z_{\alpha/2}\sqrt{\frac{\frac{556}{1000}(1-\frac{556}{1000})}{1000}},$$

y el periódico nos dice que este es como máximo de $0,03$.

$$z_{\alpha/2}\sqrt{\frac{\frac{556}{1000}(1-\frac{556}{1000})}{1000}} = 0,03 \Longrightarrow z_{\alpha/2} = 1,91.$$

(26) es equivalente a

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2} \Longleftrightarrow P\left(N(0,1) > 1,91\right) = \frac{\alpha}{2} \Longrightarrow \frac{\alpha}{2} = 0,0281 \Longrightarrow \alpha \sim 0,06.$$

El nivel de confianza de la afirmación será de $1-\alpha \sim 1-0,6 = 0,94$.

$\square$

Ejercicio 15 *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 cuasivarianza típica de la muestra que ha resultado ser de 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.*

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 coon varianza desconocida. En este caso utilizamos el estadístico

$$\frac{\overline{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \to t_{n-1},$$

que nos dá el intervalo al $95\,\%$ de confianza $(17,15; 26,84)$.

Podemos concluir 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.

$\square$.

Ejercicio 16 *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 siguientes resultados:*

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

Suponiendo que las varianzas en las dos producciones son iguales, construya un intervalo de confianza para la diferencia de las medias al nivel de 95 %*. Interprete los resultados.*

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.

La variable aleatoria X_1 mide mide la puntiación de calidad de un artículo de la primera semana. Suponemos que sigue una normal $N(\mu_1; \sigma)$.

La variable aleatoria X_2 mide mide la puntiación de calidad de un artículo de la primera semana. Suponemos que sigue una normal $N(\mu_2; \sigma)$.

En ambas poblaciones tomamos una muestra aleatoria de tamaño 8. Si designamos por $\overline{X}_1$ y $\overline{X}_2$ las medias muestrales de las poblaciones, sabemos que

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{8} + \frac{1}{8}}} \sim t_{8+8-2} \text{ y es una cantidad pivotal para } \mu_1 - \mu_1,$$

donde $S_p^2 = \frac{(8-1)\hat{S}_1^2 + (8-1)\hat{S}_2^2}{8+8-2}$. Vamos a determinas $z_{\alpha/2}$ tal que

$$P\left(-z_{\alpha/2} < \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{8} + \frac{1}{8}}} < z_{\alpha/2}\right) = 0{,}95$$

$$\Longleftrightarrow P\left(\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{8} + \frac{1}{8}}} > z_{\alpha/2}\right) = 0,025 \Longleftrightarrow z_{\alpha 2} = 2,145$$

Calculamos las medias y cuasivarianzas muestrales

$$\overline{x}_1 = 91,5;\ \hat{s}_1^2 = 9,1;\ \overline{x}_2 = 89,9;\ \hat{s}_2^2 = 17,8 \text{ y } s_p^2 = 13,4.$$

$$-z_{\alpha/2} < \frac{\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{8} + \frac{1}{8}}} < z_{\alpha/2}$$

$$\overline{x}_1 - \overline{x}_2 - z_{\alpha/2} s_p\sqrt{\frac{1}{8} + \frac{1}{8}} < \mu_1 - \mu_2 < \overline{x}_1 - \overline{x}_2 + z_{\alpha/2} s_p\sqrt{\frac{1}{8} + \frac{1}{8}}$$
$$-2,31 < \mu_1 - \mu_2 < 5,56.$$

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, (el intervalo de confianza incluye al cero), en consecuencia no podemos afirmar que ha habido un descenso significativo de la calidad entre las dos semanas.

□

Ejercicio 17 *Se quiere estudiar la proporción p de declaraciones de renta que presenta algún defecto. En una muestra preliminar, (muestra piloto), de tamaño 50 se han observado 22 declaraciones defectuosas. ¿Cuál es el tamaño muestral necesario para estimar p cometiendo un error máximo de 0,01 con una probabilidad del 0,99?.*

Universo: declaraciones de renta, por ejemplo, en la capital de Madrid.

Experimento aleatorio: tomar una declaración de la renta y ver si es defectuosa o no. Cuantificamos el experimento aleatorio con la variable aleatoria

$$X = \begin{cases} 1 & \text{la declaración de remta no es defectuosa} \\ 0 & \text{en caso contrario} \end{cases}$$

X es una población de Bernoulli, con función de distribución

$$P(X = n) = p^n(1-p)^{1-n} \qquad n = 0, 1.$$

Desconocemos es parámetro p, para estimarlo vamos a tomar una muestra aleatoria $(X_1, X_2, \cdots, X_n)$ de tamaño n para dar un intervalo de confianza al nivel $1 - 0,01 = 0,99$. Nos preguntamos sobre el tamaño de la muestra.

. Si $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es la media muestral, el estadístico

$$\frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}}$$

tiene una función de distribución próxima a la de una normal $N(0,1)$. $z_{\alpha/2} = 2,58$ satisface que

$$P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 1 - \alpha = 0,99 \tag{28}$$

Pensamos que este $z_{\alpha/2}$ satisface también

$$P(-z_{\alpha/2} < \frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}} < z_{\alpha/2}) = 1 - \alpha = 0,99 \tag{29}$$

El intervalo de confianza para p al nivel $1 - \alpha = 0,99$ es

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right),$$

y el error de este intervalo de confianza para p al nivel $0,99$ es

$$\text{Error} = z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}.$$

Queremos calcular n de tal manera que

$$z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \leq 0,01. \tag{30}$$

En la desigualdad anterior no conocemos $\overline{x}$. Vamos a tomar una muestra aleatoria piloto $(X_1, X_2, \cdots, X_{50})$ de tamaño 50 para estimar este dato. Según la realización de esta muestra aleatoria $\overline{x} = \frac{28}{50}$.

Sustituyendo en (30)

$$2,58\sqrt{\frac{\frac{28}{50}\left(1-\frac{28}{50}\right)}{n}} \leq 0,01 \implies n \geq 16402.$$

Necesitaríamos estudiar si son defectuosas o no del orden de 16400 declaraciones.

$\square$

Ejercicio 18 *El contenido en nicotina de los cigarrillos de una marca determinada sigue una distribución $N(\mu, \sigma)$. Se toma una muestra de 5 cigarrillos obteniéndose en esa muestra un contenido medio de 21,2 mg. con cuasivarianza muestral $\overline{s} = 2,05$.*

1. *Intervalo de confianza para μ, con un nivel de confianza del $90\,\%$.*
2. *Intervalo de confianza para σ^2, con un nivel de confianza del $95\,\%$.*

Solución: 1. $(19,25; 23,15)$; 2. $(1,51; 34,73)$.

Ejercicio 19 *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 número de errores en ambas clases son normales, calcular el intervalo de confianza del $95\,\%$ para la diferencia de las medias:

1. *suponiendo que las varianza poblacionales son iguales y valen $\sigma^2 = 1,44$;*
2. *suponiendo que las varianzas poblacionales son iguales pero desconocidas*

- Clase 1.

  Universo o población: alumnos de la clase 1.

Experimento aleatorio: redactar una composivión.

La característica que que queremos estudiar de esta población, y la cuantificamos con la variable aleatoria X, es el número de faltas de ortografía en la composición.

X es una población normal $N(\mu_1, \sigma)$. Tomamos una muestra aleatoria de tamaño 7, $(X_1, X_2, \cdots; X_7)$ cuya realización es $(x_1 = 10, x_2 = 10, \cdots; x_7 = 14)$. Si $\overline{X} = \frac{1}{7}\sum_{i=1}^{7} X_i$ y $\hat{S}_1^2 = \frac{1}{6}\sum_{i=1}^{7}(X_i - \overline{X})^2$ designan la media y cuasivarianza muestral respectivamente, se tiene que

$$\overline{x} = 12, \qquad \hat{s}_1^2 = 2.$$

- Clase 2.

  Universo o población: alumnos de la clase 2.

  Experimento aleatorio: redactar una composivión.

  La característica que que queremos estudiar de esta población, y la cuantificamos con la variable aleatoria Y, es el número de faltas de ortografía en la composición.

  Y es una población normal $N(\mu_2, \sigma)$. Tomamos una muestra aleatoria de tamaño 8, $(Y_1, Y_2, \cdots; Y_8)$ cuya realización es $(y_1 = 8, y_2 = 9, \cdots; y_8 = 12)$. Si $\overline{Y} = \frac{1}{8}\sum_{i=1}^{8} Y_i$ y $\hat{S}_2^2 = \frac{1}{7}\sum_{i=1}^{8}(Y_i - \overline{Y})^2$ designan la media y cuasivarianza muestral respectivamente, se tiene que

$$\overline{y} = 10,125; \qquad \hat{s}_2^2 = 1,61.$$

1. Suponemos que $\sigma^2 = 1,44$. Sabemos que

$$\overline{X} \to N\left(\mu_1, \frac{1,2}{\sqrt{7}}\right), \qquad \overline{Y} \to N\left(\mu_2, \frac{1,2}{\sqrt{2}}\right),$$

$$\overline{X} - \overline{Y} \to N\left(\mu_1 - \mu_2, \sqrt{\frac{1,44}{7} + \frac{1,44}{8}}\right), \qquad \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1,44}{7} + \frac{1,44}{8}}} \to N(0,1).$$

   Vamos a utilizar el estadístico $\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1,44}{7} + \frac{1,44}{8}}}$ para determinar el intervalo de confianza al 95 % para la diferencia de las medias.

   El $z_{\alpha/2} = 1,96$ satisface que

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1,44}{7} + \frac{1,44}{8}}} < z_{\alpha/2}\right) = 0,95 \tag{31}$$

   de donde deducimos que

$$-z_{\alpha/2} < \frac{\overline{x} - \overline{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{1,44}{7} + \frac{1,44}{8}}} < z_{\alpha/2},$$

$$\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\frac{1,44}{7} + \frac{1,44}{8}} < \mu_1 - \mu_2 < \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\frac{1,44}{7} + \frac{1,44}{8}},$$
$$0,658 < \mu_1 - \mu_2 < 3,092.$$

2. Utilizamos el estadístico
$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{7} + \frac{1}{8}}} \to t_{7+8-2}.$$
El número $z_{\alpha/2} = 2,16$ satisface
$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{7} + \frac{1}{8}}} < z_{\alpha/2}\right) = 0,95,$$
entonces
$$-z_{\alpha/2} < \frac{\overline{x} - \overline{y} - (\mu_1 - \mu_2)}{\overline{s}_p\sqrt{\frac{1}{7} + \frac{1}{8}}} < z_{\alpha/2},$$
$$\overline{x} - \overline{y} - z_{\alpha/2}s_p\sqrt{\frac{1}{7} + \frac{1}{8}} < \mu_1 - \mu_2 < \overline{x} - \overline{y} + z_{\alpha/2}s_p\sqrt{\frac{1}{7} + \frac{1}{8}},$$
y utilizando que $s_p = \sqrt{\frac{(7-1)\hat{s}_1^2 + (8-1)\hat{s}_2^2}{7+8-2}}$ obtenemos que
$$0,215 < \mu_1 - \mu_2 < 3,535.$$

Como el valor cero no pertenece a ninguno de los dos intervalos, podemos concluir que las medias poblacionales no serán igual con una confianza del 95 %.

□

Ejercicio 20 *En una muestra de 19 individuos se observó 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 una ley normal.*

1. *Fijar los límites del intervalo de confianza para la varianza con un nivel de confianza del* 99 %.
2. *realizar lo mismo que en el apartado anterior pero suponiendo que el tamaño de la muestra es* 101.
3. *realizar lo mismo que en el apartado anterior pero suponiendo que el tamaño de la muestra es* 200.

El objetivo de este problema es comparar los intervalos de confianza de muestras "pequeñasz "grandes". El enunciado nos dá la media y desviación muestral. En ambos apartados suponemos que es la misma y que solo cambia el tamaño de la muestra.

1. Vamos a utilizar el estadístico

$$\frac{19-1}{\sigma^2}\hat{S}^2 \to \chi^2_{19-1}.$$

Sea $z^1_{\alpha/2} = 6,26$ y $z^2_{\alpha/2} = 37,2$ verificando

$$P\left(\frac{18}{\sigma^2}\hat{S}^2 > z^1_{\alpha/2}\right) = 0,995, \tag{32}$$

$$P\left(\frac{18}{\sigma^2}\hat{S}^2 < z^2_{\alpha/2}\right) = 0,995 \Longleftrightarrow P\left(\frac{18}{\sigma^2}\hat{S}^2 > z^2_{\alpha/2}\right) = 0,005 \tag{33}$$

De (32) y (33) tenemos

$$\frac{18}{\sigma^2}\hat{s}^2 > z^1_{\alpha/2} \Longleftrightarrow \frac{18}{z^1_{\alpha/2}}\hat{s}^2 > \sigma^2,$$

$$\frac{18}{\sigma^2}\hat{s}^2 < z^2_{\alpha/2} \Longleftrightarrow \frac{18}{z^2_{\alpha/2}}\overline{s}^2 z < \sigma^2,$$

$$\frac{18}{z^2_{\alpha/2}}\hat{s}^2 z < \sigma^2 < \frac{18}{z^1_{\alpha/2}}\hat{s}^2.$$

Como $\hat{s}^2 = \frac{19}{18}\cdot 36$ tenemos

$$\frac{18\cdot 36}{37,2} < \sigma^2 < \frac{19\cdot 36}{6,26},$$

y el intervalo de confianza al nivel 99 % para la varianza poblacional es $(18,39; 109,27)$.

2. No vamos a hacerlo con tanto detalle como el anterior.

El estadístico es $\frac{101-1}{\sigma^2}\hat{S}^2 \sim \chi^2_{101-1}$. Sea $z^1_{\alpha/2} = 67,3$ y $z^2_{\alpha/2} = 140,2$ verificando

$$P\left(\frac{100}{\sigma^2}\hat{S}^2 > z^1_{\alpha/2}\right) = 0,995, \tag{34}$$

$$P\left(\frac{100}{\sigma^2}\hat{S}^2 < z^2_{\alpha/2}\right) = 0,995 \Longleftrightarrow P\left(\frac{18}{\sigma^2}\hat{S}^2 > z^2_{\alpha/2}\right) = 0,005 \tag{35}$$

De (34) y (35) tenemos

$$\frac{100}{\sigma^2}\hat{s}^2 > z^1_{\alpha/2} \Longleftrightarrow \frac{100}{z^1_{\alpha/2}}\overline{s}^2 > \sigma^2,$$

$$\frac{100}{\sigma^2}\hat{s}^2 < z^2_{\alpha/2} \Longleftrightarrow \frac{100}{z^2_{\alpha/2}}\hat{s}^2 z < \sigma^2,$$

$$\frac{100}{z^2_{\alpha/2}}\hat{s}^2 z < \sigma^2 < \frac{100}{z^1_{\alpha/2}}\hat{s}^2.$$

Como $\hat{s}^2 = \frac{101}{100} \cdot 36$ tenemos

$$\frac{101 \cdot 36}{140,2} < \sigma^2 < \frac{101 \cdot 36}{67,3},$$

y el intervalo de confianza al nivel 99 % para la varianza poblacional es $(25,93; 54,02)$.

3. Este apartado sería análo al anterior. Solamente observar que tendríamos que trabajar con χ^2_{199} y en las tablas no encontramos la χ^2 con 199 grados de libertad. Cuando $n > 100$, nos fijamos en la última fila de distribución de χ^2, donde aparece la Z_α. Definimos un $z_{\alpha/2}$ verificando que

$$P\left(Z\alpha > z_{\alpha/2}\right) = \frac{\alpha}{2} = \frac{0,001}{2} = 0,005 \Longrightarrow z_{\alpha/2} = 2,58$$

y

$$z^2_{\alpha/2} = (n-1) + z_{\alpha/2}\sqrt{2 \cdot n} = 199 + 2,58\sqrt{2 \cdot 200},$$
$$z^1_{\alpha/2} = (n-1) - z_{\alpha/2}\sqrt{2 \cdot n} = 199 - 2,58\sqrt{2 \cdot 200}.$$

Si seguimos el apartado anterior, el intervalo de confianza es

$$\left(\frac{200 \cdot 36}{z^2_{\alpha/2}}, \frac{200 \cdot 36}{z^2_{\alpha/2}}\right) = (28,75; 48,78).$$

Como se observa, al utilizar una muestra mayor, se reduce la amplitud del intervalo.

□

Ejercicio 21 *Se supone que los caudales medios diarios de un río se distribuyen según una ley normal de esperanza 1 l/s y desviación típica 0,2 l/s.*

1. *Calcular la probabilidad de que un día dado el caudal medio esté entre 0,7 l/s y 1,2 l/s.*

2. *A partir de una muestra de n datos,* $n > 30$*, se determina el intervalo de confianza (con nivel de significación del* 5 %*) de la esperanza, resultando este* $(1,14; 1,26)$*. ¿Cuál es el valor medio de los caudales muestreados?. ¿Cuál fue el tamaño medio de la muestra?.*

1. Sea X la variable aleatoria que mide el caudal medio diario del rio. Sabemos que sigue una $N(1; 0,2)$.

$$P(0,7 < X < 1,2) = P\left(\frac{0,7-1}{0,2} < \frac{X-1}{0,2} < \frac{1,2-1}{0,2}\right)$$

$$= P\left(-1,5 < \frac{X-1}{0,2} < 1\right) = 1 - P\left(\frac{X-1}{0,2} < 1\right) - P\left(\frac{X-1}{0,2} < 1,2\right)$$
$$= 1 - 0,1587 - 0,0668 = 0,7745.$$

2. Suponemos que la desviación poblacional la conocemos, y utilizamos la muestra para dar un intervalo de confianza de la media a un nivel de significación del 5 %, $\alpha = 0,5$.

   Sea $(X_1, X_2, \cdots, X_n)$ una muestra de tamaño n. Utilizamos el estadístico

   $$\frac{X-\mu}{\frac{0,2}{\sqrt{n}}} \to N(0;1).$$

   Sea $z_{\alpha/2} = 1,96$ tal que

   $$P\left(-1,96 < \frac{X-\mu}{\frac{0,2}{\sqrt{n}}} < 1,96\right) = 0,95.$$

   Utilizando esto, el intervalo de confianza al 95 % para μ es

   $$(\overline{x} - 1,96 \cdot \frac{0,2}{\sqrt{n}}; \overline{x} + 1,96 \cdot \frac{0,2}{\sqrt{n}}.$$

   Según el enunciado

   $$\left\{ \begin{array}{l} (\overline{x} - 1,96 \cdot \frac{0,2}{\sqrt{n}} = 1,14 \\ \overline{x} + 1,96 \cdot \frac{0,2}{\sqrt{n}} = 1,26 \end{array} \right. \Longrightarrow \left\{ \begin{array}{l} \overline{x} = 1,2 \\ n \sim 42,68 \end{array} \right.$$

   El tamaño de la muestra debe ser del orden de 43.

□

Ejercicio 22 *Si el porcentaje de individuos daltónicos de una cierta población en una muestra aleatoria de tamaño 36 es igual al* 30 %*. Estimar la proporción p de individuos daltónicos de una población con un nivel de confianza de* 0,95*.*

Solución: $(0,1503; 0,4497)$.

Ejercicio 23 *En un determinado barrio se seleccionó al azar una muestra de 100 personas cuya media de ingresos mensuales resultaba igual a 1060 con una desviación típica de 200 . Si se toma un nivel de confianza del* 95 %*:*

1. *¿cuál es el intervalo de confianza para la media de los ingresos mensuales de toda la población?;*

2. *si se toma un nivel de significación igual a 0,01, ¿cuál es el tamaño muestral necesario para estimar la media de ingresos mensuales con un error menor que 30 ?*

INTERVALOS DE CONFIANZA MÁS FRECUENTES

1. $X \sim N(\mu, \sigma)$, $(X_1, X_2, \cdots, X_n)$ muestra aleatoria de tamaño n con realización $(x_1, x_2, \cdots, x_n)$ y

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2.$$

Intervalos de confianza al nivel $1-\alpha$ ‘para μ:

a) σ conocida

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

donde $z_{\alpha/2} > 0$ y $P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}$.

b) σ desconocida

$$\left(\overline{x} - z_{\alpha/2}\frac{\hat{s}}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\hat{s}}{\sqrt{n}}\right),$$

donde $z_{\alpha/2} > 0$ y $P\left(t_{n-1} > z_{\alpha/2}\right) = \frac{1}{2} - \frac{\alpha}{2}$.

Intervalo de confianza al nivel $1-\alpha$ ‘para σ^2:

$$\left(\frac{n-1}{z_{\alpha 2}^2}\hat{s}^2, \frac{n-1}{z_{\alpha 2}^1}\hat{s}^2\right),$$

donde $0 < z_{\alpha 2}^1 < z_{\alpha 2}^2$ y

$$P\left(\chi_{n-1}^2 > z_{\alpha 2}^1\right) = 1 - \frac{\alpha}{2}, \qquad P\left(\chi_{n-1}^2 > z_{\alpha 2}^2\right) = \frac{\alpha}{2}.$$

2. Dos poblaciones normales independientes

- Sea X una pobleción normal $N(\mu_1; \sigma_1)$, $(X_1, X_2, \cdots, X_m)$ una muestra aleatoria simple de tamaño m con realización $(x_1, x_2, \cdots, x_m)$ y $\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ y $\hat{s}_1^2 = \frac{1}{m-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{m-1}\sum_{i=1}^{n} x_i^2 - \frac{m}{m-1}\overline{x}$.
- Sea Y una pobleción normal $N(\mu_2 1; \sigma_2)$, $(Y_1, Y_2, \cdots, Y_n)$ una muestra aleatoria simple de tamaño n con realización $(y_1, y_2, \cdots, y_n)$ y $\overline{y} = \frac{1}{m}\sum_{i=1}^{n} y_i$ y $\hat{s}_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n}{n-1}\overline{y}$.
- $\overline{s}_p^2 = \frac{(m-1)\overline{s}_1^2 + (n-1)\overline{s}_2^2}{m+n-2}$.

Intervalos de confianza de $1-\alpha$ para $\mu_1 - \mu_2$:

a) Son conocidas la desviaciones típicas poblacionales σ_1 y σ_2:

$$\left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right)$$

donde $z_{\alpha/2}$ satisface

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

b) Desconocidas las desviaciones típicas poblacionales pero iguales $\sigma_1 = \sigma_2$.

$$\left(\overline{x} - \overline{y} - z_{\alpha/2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}, \overline{x} - \overline{y} - z_{\alpha/2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(t_{m+n-2} > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

Intervalo de confianza de $1-\alpha$ para $\frac{\sigma_1^2}{\sigma_2^2}$:

$$\left(\frac{\hat{s}_1^2}{\hat{s}_2^2} \cdot \frac{1}{z_\alpha^2}, \frac{\hat{s}_1^2}{\hat{s}_2^2} \cdot \frac{1}{z_\alpha^1}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(F_{m-1,n-1} > z_\alpha^2\right) = \frac{\alpha}{2}, \quad P\left(F_{n-1,m-1} > \frac{1}{z_\alpha^1}\right).$$

3. Poblaciones no normales. Tamaño de la muestra n grande, ($n \geq 30$).

a) Intervalo de confianza de $1-\alpha$ para μ, X población cualquiera con desviación típica poblacional σ conocida.

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

b) Intervalo de confianza de $1-\alpha$ para μ, X población cualquiera con desviación típica poblacional σ desconocida.

$$\left(\overline{x} - z_{\alpha/2}\frac{\overline{s}}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\overline{s}}{\sqrt{n}}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

c) $X \sim B(1,p)$, Bernoulli. Intervalo de confianza $1-\alpha$ para p:

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$

d) $X \sim P(\lambda)$, Poisson. Intervalo de confianza $1-\alpha$ para λ:

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}}{n}}\right),$$

donde $z_{\alpha/2}$ satisface

$$P\left(N(0,1) > z_{\alpha/2}\right) = \frac{\alpha}{2}.$$