TEMA 5: Introducción al muestreo

Con el fin de extraer conclusiones a partir de una MUESTRA (subconjunto de la población), y extrapolarlos a toda la POBLACIÓN, se emplean técnicas de inferencia estadística.

Para que estas aporten resultados fiables, los individuos han de representar adecuadamente el comportamiento de toda la población en relación al carácter bajo estudio.

Por ello, se utiliza la Teoría de muestras o muestreo, la cual estudia procedimientos basados en el azar, para seleccionar una muestra representativa de la población.

Para que la muestra sea representativa, el proceso de selección tiene que ser aleatorio y cada elemento tiene que tener la misma oportunidad (probabilidad) de ser incluido en la muestra.

MUESTREO ALEATORIO:

4 tipos:

- Muestreo aleatorio SIMPLE
- Muestreo aleatorio ESTRATIFICADO
- Muestreo aleatorio CONGLOMERADO
- Muestreo aleatorio SISTEMÁTICO

→ Muestreo aleatorio SIMPLE: el método más sencillo.

1. Definir la población.

1.1. Asignar un número a cada individuo.

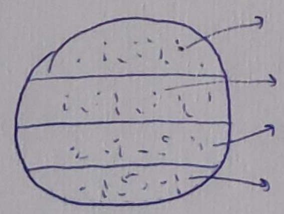
2. A través de algún medio, elegir aleatoriamente hasta completar la muestra.

Un individuo es elegible en más de una ocasión. La probabilidad es baja, pero es posible.

→ Muestreo aleatorio ESTRATIFICADO:

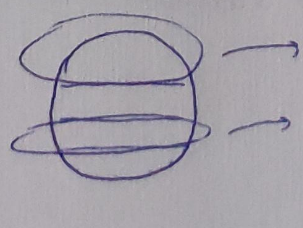
1. Organizar a la población en categorías típicas diferentes entre sí (ESTRATOS) que poseen gran homogeneidad respecto a alguna característica.

2. Se escogen individuos de cada estrato.



→ Muestreo aleatorio CONGLOMERADO:

Al igual que el ESTRATIFICADO, pero se seleccionarán los estratos enteros en lugar de individuos de cada estrato.



→ Muestreo aleatorio SISTEMÁTICO:

Al igual que el SIMPLE, pero se seleccionará aleatoriamente un solo elemento y es resto será sistemático:

$$i,\ i+k,\ i+2k,\ i+3k,\ \ldots,\ i+(n-1)k$$

Por ejemplo: $i=2$; → 2 ; 4 ; 8; 16 ...

Generalidades del muestreo aleatorio simple:

1. Cada elemento de la muestra se selecciona al azar. Por tanto, las repeticiones son independientes y en cada una de ellas los elementos de la población son equiprobables.

2. Una m.a.s (muestra aleatoria simple) de tamaño n, de una variable aleatoria X con distribución teórica F, es un vector formado por n variables aleatorias $(X_1, X_2, \ldots, X_n)$, independientes entre sí e igualmente distribuidas, con distribución común F.

$$X \sim N \longrightarrow X_1 \sim N\ ;\ X_2 \sim N\ .$$

Cada individuo / muestra seguirá la misma distribución que toda la población.

3. Consecuentemente, la función de distribución de una m.a.s. $(X_1, X_2, \ldots, X_n)$, correspondiente a una distribución de la población F, es:

$$F(X_1, X_2, \ldots, X_n) = F(X_1) \cdot F(X_2) \cdot \ldots \cdot F(X_n)$$

4. Los estadísticos, que sintetizan el comportamiento del conjunto de datos, son funciones de los valores de la muestra y por tanto, son a su vez variables aleatorias.

Principales distribuciones de los estadísticos

→ Distribución χ^2 de Pearson χ^2_n

Sean $X_1, X_2, \ldots, X_n$ n variables aleatorias normales con media 0 y varianza 1 independientes entre sí, entonces la variable

$$Y = X_1^2 + X_2^2 + \ldots + X_n^2$$

Sigue una distribución χ^2 (chi-cuadrado) de n grados de libertad.

$$Y \sim \chi^2_n$$

F. de DENSIDAD:

$$f(y) = \begin{cases} \dfrac{y^{(n/2)-1}}{2^{(n/2)}\,\Gamma\left(\frac{n}{2}\right)} e^{(-y/2)} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

F. de DISTRIBUCIÓN

donde $\Gamma(n)$ es la función gamma, definida para todo valor de $n \in \mathbb{R}^+$, por la integral impropia:

$$\Gamma(n) = \int_0^{+\infty} x^{n-1} e^{-x} \, dx \qquad n > 0$$

Propiedades de la función $\Gamma(n)$:

i. $\Gamma(1) = 1$

ii. Si $n \in \mathbb{Z}^+$, entonces $\Gamma(n) = (n-1)!$

iii. $\Gamma(n) = (n-1)\,\Gamma(n-1)$, $\forall n \in \mathbb{R}^+$ y $n > 1$

iv. En consecuencia, si se dispone de los valores $\Gamma(n)$ para $n \in [1, 2]$ se puede calcular $\Gamma(n)$ $\forall n > 1$

v. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

La distribución χ^2 únicamente toma valores positivos al ser suma de cuadrados y depende únicamente del parámetro n.

Propiedades de la distribución χ^2_n:

i. Definida en el intervalo $[0, +\infty)$

ii. Continua en todo su dominio.

iii. No es simétrica.

iv. Si $X_1, X_2, \dots, X_k$ son k variables aleatorias independientes con distribuciones $X_i \sim \chi^2_{n_i}$, entonces la v.a. $X = \sum_{i=1}^{k} X_i$ sigue una distribución χ^2_n siendo $n = \sum_{i=1}^{k} n_i$.

v. $\mu = n$; $\sigma^2 = 2n$.

La distribución χ^2_n de Pearson puede aproximarse por una distribución normal. Como se considera que $\chi^2_n = \sum_{i=1}^{n} Z_i^2$, siendo $Z_i \sim N(0,1)$, cada una de las variables aleatorias Z_i^2 seguirá una distribución χ^2_1 que tendrán $\mu = 1$; $\sigma^2 = 2$.

Por el teorema central del límite, la variable $\chi^2_n = \sum_{i=1}^{n} Z_i^2$ es asintóticamente $N(\mu = n, \sigma = \sqrt{2n})$.

Consecuentemente, para valores grandes de n, se puede aproximar la v.a. $X \sim \chi^2_n$ por una distribución normal $N(\mu = n, \sigma = \sqrt{2n})$.

En la práctica, cuando $\boxed{n \geq 30}$.

→ Distribución t de Student t_n

Sean $X, X_1, X_2, \dots, X_n$ n+1 variables aleatorias normales con media 0 y desviación típica σ independientes entre sí, entonces la variable

$$T = \frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}$$

Sigue una distribución t de Student de n grados de libertad.

$$T \sim t_n$$

Si se divide entre σ tanto el numerador como el denominador, entonces:

$$T = \frac{X/\sigma}{\sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\sigma}}} = \frac{Z}{\sqrt{X/n}}, \text{ donde } Z \sim N(0,1) \text{ y } X \sim \chi^2_n$$

Sumatorio de normales al cuadrado sigue una distribución de Pearson.

F. de DENSIDAD:

$$f(t) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$$

$\forall\, t \in \mathbb{R}$ y $n > 0$

Propiedades de la distribución t_n:

i. Definida en el intervalo $(-\infty, +\infty)$

ii. Continua en todo su dominio

iii. Simétrica respecto a la media

iv. $\mu = 0$; $\sigma^2 = \frac{n}{n-2}$ si $n > 2$

Las funciones de densidad de las distribuciones t de Student tienen forma acampanada como la distribución normal, pero son platicúrticas.

Como: $\lim_{n\to\infty} f(t) = \lim_{n\to\infty} \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$

que es la función de densidad de la $N(0,1)$, al aumentar el número de grados de libertad la distribución t_n tiende a la normal tipificada.

En la práctica, cuando $\boxed{n \geq 30}$.

→ Distribución F de Fisher - Snedecor $F_{n_1;n_2}$

Sean X_1 y X_2 dos variables aleatorias independientes entre sí y que siguen distribuciones χ^2 de Pearson de n_1 y n_2 grados de libertad respectivamente, es decir, $X_1 \sim \chi^2_{n_1}$ y $X_2 \sim \chi^2_{n_2}$ entonces la variable aleatoria

$$F = \frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}}$$

Sigue una distribución F de Fisher - Snedecor con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador.

$$F \sim F_{n_1;n_2}$$

F. de DENSIDAD

$$f(x) = \begin{cases} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \cdot \dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} x^{\frac{n_1}{2}-1} \cdot \left(1+\frac{n_1}{n_2}x\right)^{-\frac{(n_1+n_2)}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Propiedades de la distribución $F_{n_1;n_2}$:

i. Definida en el intervalo $[0, +\infty)$

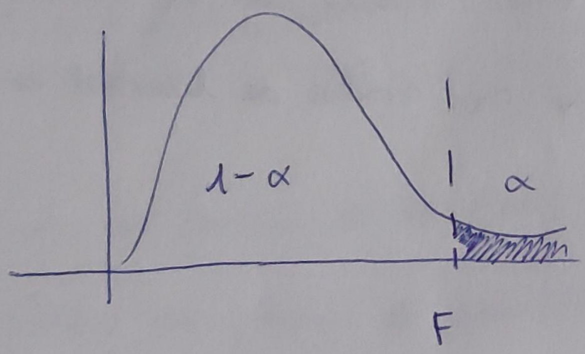
ii. Continua en todo su dominio

iii. No es simétrica.

iv. $\mu = \dfrac{n_2}{n_2-2} \quad (n_2 > 2)$

$$\sigma^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} \quad (n_2>4)$$

v. Se cumple que: $F_{1-\alpha;n_1,n_2} = \dfrac{1}{F_{\alpha;n_2,n_1}}$



Distribuciones en el muestreo de los estadísticos

→ Media muestral $\overline{X}$

POBLACIÓN:	MUESTRA:	DISTRIBUCIÓN del ESTADÍSTICO:
• Normal σ conocida		$\overline{X} \sim N(\mu, \sigma/\sqrt{n}) \Rightarrow Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
• Normal σ desconocida		$T = \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
• Cualquiera σ conocida	$n > 30$	$\overline{X} \sim N(\mu, \sigma/\sqrt{n}) \Rightarrow Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
• Cualquiera σ desconocida	$n > 100$	$\overline{X} \sim N(\mu, S/\sqrt{n}) \Rightarrow Z = \dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$

→ Diferencia de dos medias muestrales : $\overline{X_1} - \overline{X_2}$

POBLACIÓN :	DISTRIBUCIÓN del ESTADÍSTICO.
• Normales independientes σ_1, σ_2 conocidas	$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$
• Normales independientes σ_1, σ_2 desconocidas $\sigma_1 = \sigma_2$	$\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
• Normales independientes σ_1, σ_2 desconocidas $\sigma_1 \neq \sigma_2$	$\dfrac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim t_v$ donde $v = \dfrac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n+1} + \frac{\left(\frac{S_2^2}{m}\right)^2}{m+1}} - 2$
• Cualesquiera independientes σ_1, σ_2 conocidas $n > 15$ $m > 15$	$\bar{X}_1 - \bar{X}_2 \cong N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$
• Cualesquiera independientes σ_1, σ_2 desconocidos $n > 100$ $m > 100$	$\bar{X}_1 - \bar{X}_2 \cong N\left(\mu_1 - \mu_2, \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}\right)$

* Cuasivarianza.

→ Varianza muestral : s^2

POBLACIÓN:

DISTRIBUCIÓN del ESTADÍSTICO:

• Normal
μ conocida

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

• Normal
μ desconocida

$$\frac{ns^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

→ Cociente de dos cuasivarianzas muestrales : S_1^2/S_2^2

POBLACIÓN:

DISTRIBUCIÓN del ESTADÍSTICO:

• Normales

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,\, m-1}$$

→ Proporción muestral : $\hat{p}$

MUESTRA:

DISTRIBUCIÓN del ESTADÍSTICO:

$n > 100$

$$\hat{p} \cong N\left(p, \sqrt{\frac{pq}{n}}\right)$$

→ Diferencia de dos proporciones muestrales : $\hat{p}_1 - \hat{p}_2$

MUESTRA:

DISTRIBUCIÓN del ESTADÍSTICO:

$n > 100$
$m > 100$

$$\hat{p}_1 - \hat{p}_2 \cong N\left(p_1 - p_2 \,,\, \sqrt{\frac{p_1q_1}{n} + \frac{p_2q_2}{m}}\right)$$

ejemplo 1: el coeficiente que mide la dureza de unas láminas de acero sigue una distribución $N(1'6 ; 0'3)$. Se ha elegido al azar cinco láminas de acero.

a) Calcule la probabilidad de que la media de la dureza de las cinco láminas de acero elegidas al azar sea como mucho 1'5.

$$P(\bar{X} \leq 1'5) = P\left(Z \leq \frac{1'5 - 1'6}{0'3/\sqrt{5}}\right) = P(Z \leq -0'7454) =$$

$$\bar{X} \sim N\left(1'6, \frac{0'3}{\sqrt{5}}\right) \qquad = P(Z \geq 0'7454) = 1 - P(Z \leq 0'7454) =$$

$$= 1 - \text{pnorm}(0'7454, 0, 1) = 1 - 0'7720 =$$

$$= 0'228$$

b) Suponga ahora que la varianza de la distribución de la población es desconocida. Sabiendo que la varianza de la dureza de las cinco láminas al azar es 0,2; calcula la probabilidad de que la media muestral sea como mucho 2.

X := "Dureza de unas láminas de acero".

$$X \sim N(1'6, \sigma) \; ; \qquad T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \qquad \left|\begin{array}{l} \mu = 1'6 \\ n = 5 \\ s^2 = 0'2 \\ S^2 = 0'2 \cdot \frac{5}{4} = 0'25 \end{array}\right.$$

$$s^2 = 0'2 \rightarrow S^2 = 0'2 \cdot \frac{5}{4} = 0'25$$

$$P(\bar{X} \leq 2) = P\left(T \leq \frac{2 - \mu}{\sqrt{0'25}/5}\right) = P(T \leq 1'7889) = pt(1'7889, 4) =$$

$$= 0'9259$$

ejemplo 2: Con el fin de medir la calidad del aire de una ciudad, se analizan las concentraciones medias mensuales de SO_2 en dos estaciones. Se han escogido los datos de 10 meses al azar sobre las concentraciones medias de SO_2 en la estación A y los datos de 12 meses al azar en la estación B. La desviación típica de los 10 valores recogidos de la estación A es 2 mientras que la desviación típica de los 12 valores recogidos de la estación B es 1.
La concentración media mensual de SO_2 en la estación A sigue una distribución $N(15, \sigma_A)$ mientras que en la estación B sigue una distribución $N(13, \sigma_B)$.

Suponga que ambas variables son independientes y que las varianzas aun siendo desconocidas no son iguales. Calcule la probabilidad de que la media de la muestra A sobre la concentración media de SO_2 supere en más de una unidad a la media de la muestra B.

X_A := "Concentración media mensual de SO_2 en la estación A".

X_B := "Concentración media mensual de SO_2 en la estación B".

$n_A = 10$; $n_B = 12$ | $X_A \sim N(15, \sigma_A)$

$S_A = 2$; $S_B = 1$ | $X_B \sim N(13, \sigma_B)$

$$P(\overline{X_A} > \overline{X_B} + 1) = P(\overline{X_A} - \overline{X_B} > 1)$$

Diferencia de medias muestrales.

$P(\bar{X}_A - \bar{X}_B > 1)$

Diferencia de dos medias muestrales, cuyas poblaciones siguen distribuciones normales y son independientes, con $\sigma_1 \neq \sigma_2$.

* Cuasivarianza

$$* \quad \frac{S_1^2}{n} = \frac{s_1^2}{n-1}$$

Cuasivarianza   Varianza

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{m}}} \sim t_v$$

$$v = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n+1} + \frac{\left(\frac{S_2^2}{m}\right)^2}{m+1}} - 2$$

$$v = \frac{\left(\frac{2^2}{9} + \frac{1^2}{11}\right)^2}{\frac{\left(\frac{2^2}{9}\right)^2}{11} + \frac{\left(\frac{1^2}{11}\right)^2}{13}} - 2 = \frac{0'2866034078}{0'01859307922} - 2 =$$

$$= 15'41452089 - 2 = 13'41452085 \simeq 13$$

$$P\left(T > \frac{1 - (15-13)}{\sqrt{\frac{2^2}{9} + \frac{1^2}{11}}}\right) = P(T > -1'438) = P(T \leq 1'438) =$$

$$= pt(1'438, 13) = \boxed{0'9130}$$

ejemplo 3: atendiendo al último estudio elaborado por la agencia LV, el 90% de la población utiliza el teléfono móvil. Se ha elegido una muestra aleatoria simple compuesta por 200 personas.

a) Calcule la probabilidad de que de esas 200 personas elegidas al azar, la proporción de usuarios de teléfono móvil sea al menos 0'85.

X := "Número de personas que utiliza el teléfono móvil".

$X \sim B(n, p = 0'9)$

$n = 200$; $\hat{p} \sim N\left(p = 0'9, \sqrt{\frac{0'9 \cdot 0'1}{200}}\right) \equiv (0'9, 0'0212132034)$

$$P(\hat{p} \geq 0'85) = 1 - P(\hat{p} < 0'85) = 1 - P\left(Z \leq \frac{0'85 - 0'9}{\sqrt{\frac{0'9 \cdot 0'1}{200}}}\right) =$$

$$= 1 - P(Z \leq -2'3600) = 1 - P(Z \geq 2'36) = 1 - 1 + P(Z \leq 2'36) =$$

$$= \text{pnorm}(2'36, 0, 1) = \boxed{0'9909}$$

b) ¿Cuál es la probabilidad de que la diferencia entre la proporción muestral y la proporción poblacional sea como mucho del 2%?

$$P(|\hat{p} - p| \leq 0'02) = P(-0'02 \leq \hat{p} - p \leq 0'02) = \text{TIPIFICAR}$$

dividiendo entre σ.

$$= P\left(\frac{-0'02}{\sqrt{\frac{0'901}{200}}} < Z \leq \frac{0'02}{\sqrt{\frac{0'901}{200}}}\right) = P(-0'9400 < Z \leq 0'9400) =$$

$$= P(Z \leq 0'94) - P(Z \leq -0'94) = P(Z \leq 0'94) - (1 - P(Z \leq 0'94)) =$$

$$= 2P(Z \leq 0'94) - 1 = 2 \cdot 0'8264 - 1 = \boxed{0'6528}$$

ejemplo 4: el número de huéspedes mensuales de un hotel de lujo sigue una distribución normal cuya desviación típica es de 30 huéspedes. Se han escogido los datos sobre los huéspedes del hotel de 5 meses aleatorios.

a) ¿Cuál es la probabilidad de que la cuasivarianza sobre los huéspedes mensuales de una muestra de 5 meses sea como mínimo de 64 $(\text{huéspedes})^2$?

$X :=$ "Número de huéspedes mensuales de un hotel de lujo".

$X \sim N(\mu, 30)$

$n = 5$ meses

$P(S^2 > 64) = 1 - P(S^2 \leq 64)$

Varianza muestral, población normal, μ desconocida

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

$$\frac{(5-1)\cdot 64}{30^2} \Rightarrow 1 - P\left(C \leq \frac{64\cdot 4}{30^2}\right) \Rightarrow C \sim \chi_{n-1}$$

pchisq(0'2844, 4) = 0'0092.

$= 1 - P(C \leq 0'2844) = 1 - 0'0092 = 0'9908$

b) ¿Cuál es la probabilidad de que la cuasivarianza muestral de 5 meses sea al menos el doble que la varianza poblacional?

$$P(S^2 \geq 2\sigma^2) = 1 - P(S^2 \leq 2\sigma^2) = 1 - P\left(\frac{(n-1)S^2}{\sigma^2} \leq 2\cdot\frac{\sigma^2(n-1)}{\sigma^2}\right) =$$

$C \sim \chi^2_{n-1}$

$= 1 - P(C \leq 2\cdot 4) = 1 - P(C \leq 8)$

$= 1 -$ pChisq$(8, 4) = \boxed{0'0916}$

ejercicio 1:

$X :=$ "Nivel de sociabilidad de las personas".

$X \sim N(25, 5)$

a) $\cancel{\ }$

$P(22 < X < 30) = P(X<30) - P(X<22) = P\left(Z < \frac{30-25}{5}\right) - P\left(Z < \frac{22-25}{5}\right) =$

$= P(Z<1) - P(Z < -3/5) = pnorm(1, 0, 1) + pnorm(-3/5, 0, 1) - 1 =$

$= 0'8413 - 1 + 0'7257 = 0'567$.

b) $\overline{X} \sim N(25, 5/\sqrt{25})$ $\rightarrow Z = \frac{\overline{X} - \mu}{\sigma}$

$P(|\overline{x} - \mu| < 1) = P(-1 < \overline{X} - \mu < 1) =$

$= P\left(-\frac{1}{1_\sigma} < Z < \frac{1}{1_\sigma}\right) = P(-1 < Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) =$

$= P(Z \leq 1) - (1 - P(Z \leq 1)) = 2P(Z \leq 1) - 1 =$

$= 2 \cdot 0'8413 - 1 = \boxed{0'6826}$

ejercicio 2:

$X :=$ "Medidas proporcionadas ~~de~~ por un aparato".

$X \sim N(\mu, 0'1)$; muestra X_1 de 25 mediciones : $X_1 \sim N\left(\mu, \frac{0'1}{\sqrt{25}}\right)$

$P(S^2 > 0'014)$; población normal de μ desconocida

$\frac{(n-1)S^2}{\sigma^2} = \underbrace{\frac{nS^2}{\sigma^2}}_{C} \sim \chi^2_{n-1}$ $\longrightarrow$ $\frac{25 \cdot 0'014}{0'1^2} = 35$.

$P(C > 12'25) = 1 - P(C \leq 12'25) = 1 - pchisq(35, 24) =$

$= 1 - 0'9316 = \boxed{0'0684}$

ejercicio 3:

$X :=$ "Número de ordenadores recuperados".

$X \sim B(n, p=0'75)$; muestra: $n=150$

a) $P(0'65 \leq \hat{p} \leq 0'85) = P(\hat{p} \leq 0'85) - P(\hat{p} \leq 0'65) =$

$* \; \hat{p} \cong N\left(p, \sqrt{\frac{pq}{n}}\right) \equiv N\left(0'75, \sqrt{\frac{0'75 \cdot 0'25}{150}}\right) = N\left(0'75, \frac{\sqrt{2}}{40}\right)$

$= P\left(Z \leq \frac{0'85-0'75}{\sqrt{2}/40}\right) - P\left(Z \leq \frac{0'65-0'75}{\sqrt{2}/40}\right) =$

$= P(Z \leq 2\sqrt{2}) - P(Z \leq -2\sqrt{2}) =$

$=$ pnorm $(2\sqrt{2}, 0, 1)$ + pnorm $(2\sqrt{2}, 0, 1) - 1 =$

$= 2$ pnorm $(2\sqrt{2}, 0, 1) - 1 = 2 \cdot 0'9976 - 1 = \boxed{0'9952}$

b) $P(\hat{p} > 85\%) = 1 - P(\hat{p} \leq 0'84) =$

0'85

$=$

$\hat{p} \geq 85\%$

$*$ $\hat{p}$ sigue una normal.

$= 1 - P\left(Z \leq \frac{0'85-0'75}{\sqrt{2}/40}\right) = 1 - P(Z \leq 2\sqrt{2}) =$

$= 1 -$ pnorm $(2\sqrt{2}, 0, 1) = 1 - 0'9976 = \boxed{0'0024}$

ejercicio 4:

$X_A :=$ "Número de piezas que se deben repetir en la primera cadena".

$X_B :=$ "Número de piezas que se deben repetir en la segunda cadena".

$X_A \sim B(n, p = 0'07)$; $X_B \sim B(n, p = 0'02)$

muestra de 180 — muestra de 300

a) $P(\hat{p}_A > \hat{p}_B) = P(\hat{p}_A - \hat{p}_B > 0) =$

$$\left\{ \begin{array}{l} \hat{p}_A - \hat{p}_B \cong N\left(0'07 - 0'02, \sqrt{\frac{0'07 \cdot 0'93}{180} + \frac{0'02 \cdot 0'98}{300}}\right) \\ \cong N(0'05, 0'0206) \end{array} \right.$$

$$P = \left(Z > \frac{0 - 0'05}{0'0206}\right) = 1 - P(Z \leq -2'427184\ldots) =$$

$$= 1 - P(Z \geq 2'4272) = 1 - (1 - P(Z < 2'4272)) =$$

$$= P(Z < 2'4272) = \text{pnorm}(2'4272, 0, 1) = \boxed{0'9923921}$$

b) $P(\hat{p}_A - \hat{p}_B \geq 0'1) = 1 - P\left(Z < \frac{0'1 - 0'05}{0'0206}\right) =$

$$= 1 - P(Z < 2'427184\ldots) = 1 - \text{pnorm}(2'4272) =$$

$$= 1 - 0'9924 = \boxed{0'0076}$$

ejercicio 5:

$X :=$ "Calificación de los alumnos matriculados en informática".

$\mu_x = 5'7$

$\sigma_x^2 = 5'97$

muestra de 50 alumnos

$\sigma_x = \sqrt{5'97} = 2'4434$

a) población cualquiera, σ conocida, $n > 30$.

$$\bar{X} \sim N\left(5'7, \frac{2'4434}{\sqrt{50}}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$P(\bar{X} > 6) = 1 - P(\bar{x} \leq 6) = 1 - P\left(Z \leq \frac{6 - 5'7}{0'3455}\right) =$$

$$= 1 - P(Z \leq 0'8683) = 1 - \text{pnorm}(0'8683, 0, 1) =$$

$$= 1 - 0'8073849 = \boxed{0'1926151}$$

b) $$P(\bar{x} \geq 5) = 1 - P(\bar{x} < 5) = 1 - P\left(Z < \frac{5 - 5'7}{0'3455}\right) =$$

$$= 1 - P(Z < -2'0260) = 1 - (1 - P(Z < 2'0260)) =$$

$$= \text{pnorm}(2'0260, 0, 1) = \boxed{0'9786}$$

ejercicio 6:

$X :=$ "Diámetro de unas bolas de hierro"

$X \sim N(72, \sigma)$

$\boxed{\text{Muestra 1}}$ $\overline{X_1} = 65'0286$ $\quad s_1^2 = 99'8926061$ $\quad s_1 = 9'994629088$

$n = 7$; 78'9, 65'1, 55'2, 80'9, 57'4, 55'4, 62'3

$\boxed{\text{Muestra 2}}$ $\overline{X_2}$

$m = 7$

$P(\overline{X_2} > \overline{X_1}) \Rightarrow P(\bar{x}_2 - \bar{x}_1 > 0)$

Diferencia de medias muestrales, población normal, σ_1 y σ_2 desconocidos pero iguales.

Sabemos que $\bar{X}_1$ es 65'0286 ; por tanto:

$P(\overline{X_2} > 65'0286)$

Media muestral, población normal, σ desconocida:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \qquad T = \frac{65'0286 - 72}{10'7954/\sqrt{7}} \sim t_6$$

- $s_1^2 = 99'8926$

$$S_1^2 = 99'8926 \cdot \frac{7}{6} = 116'5413667$$

$$S_1 = \sqrt{116'5413667} = \boxed{10'79543268}$$

$P(\bar{X}_2 > 65'0286) = 1 - P(\bar{X}_2 \leq 65'0286) = 1 - P(T \leq -1'708560191) =$

$= 1 - (1 - P(T \leq 1'7085)) = pt(1'7085, 6) = \boxed{0'9308}$

ejercicio 7:

$n = 14$

$X \sim N(\mu, \sigma)$; Si $s^2 (n=14) > 40'5$, entonces $\sigma^2 \neq 20$

a) $\sigma^2 = 20$

$P(s^2 > 40'5)$ → varianza muestral, población normal μ desconocida.

$$C = \frac{n s^2}{\sigma^2} = \frac{(n-1) S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$C = \frac{14 \cdot 40'5}{20} = 28'35$$

$= 1 - P(s^2 \leq 40'5) = 1 - P(C \leq 28'35) = 1 - \text{pchisq}(28'35, 13)$

$= 1 - 0'991915 = 0'008084954 \simeq \boxed{0'0081}$

b) $\sigma^2 = 20$

$P(40'5 \leq s^2 \leq 42) = P(s^2 \leq 42) - P(s^2 \leq 40'5) =$

$$C_1 = \frac{n s^2}{\sigma^2} = \frac{14 \cdot 42}{20} = 29'4 \sim \chi^2_{n-1}$$

$$C_2 = \frac{n s^2}{\sigma^2} = \frac{14 \cdot 40'5}{20} = 28'35 \sim \chi^2_{n-1}$$

$= P(C \leq 29'4) - P(C \leq 28'35) =$

$\text{pchisq}(29'4, 13) - \text{pchisq}(28'35, 13) =$

$0'9942582 - 0'991915 = \boxed{0'0023432}$

ejercicio 8:

X_1 := "Número de automóviles que pasan a diario por la estación I".

X_2 := "Número de automóviles que pasan a diario por la estación II".

$X_1 \sim N(\mu, \sqrt{20})$; $X_2 \sim N(\mu, \sqrt{36})$

$\boxed{\text{MUESTRA I}}$ (X_1) $\boxed{\text{MUESTRA II}}$ (X_2)

$n=8$ $m=10$

a) $P\left(s_1^2/s_2^2 \leq 1'78\right)$

Cociente de dos cuasivarianzas muestrales, poblaciones normales.

$S_1^2 = s_1^2 \cdot \frac{n}{n-1}$

- $1'78 \cdot \frac{8}{7} = 2'0343$
- $1'78 \cdot \frac{10}{9} = 1'9777$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1;m-1} = F$$

$$P\left(\frac{S_1^2}{S_2^2} \leq \frac{1'78 \cdot \frac{8}{7}}{10/9}\right) = P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq 1'8309 \cdot \frac{36}{20}\right) =$$

$$= P\left(\underbrace{\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}}_{F_{7,9}} \leq 3'2955\right) = pf(3'2955, 7, 9) = \underline{0'9501132}$$

b) $P\left(\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq 2'44\right) = P\left(\dfrac{S_1^2}{S_2^2} \leq \dfrac{2'44 \cdot \sigma_1^2}{\sigma_2^2}\right) =$

$\left.\begin{array}{l}\sigma_1^2 = 20 \\ \sigma_2^2 = 26\end{array}\right\} = P\left(\dfrac{S_1^2}{S_2^2} \leq 1'8\overline{769230}\right)$

$$P\left(\dfrac{\dfrac{S_1^2 \cdot \frac{n}{n-1}}{\sigma_1^2}}{\dfrac{S_2^2 \cdot \frac{n}{n-1}}{\sigma_2^2}} \leq \dfrac{2'44 \cdot \frac{n}{n-1}}{\frac{m}{m-1}}\right) = P\left(\overbrace{\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}}^{F_{n-1;\, m-1}} \leq \dfrac{2'44 \cdot \frac{8}{7}}{\frac{10}{9}}\right) =$$

$= P(F \leq 2'5097) = pf(2'5097, 7, 9) = 0'9004$

ejercicio 9:

$X :=$ "Cantidad de nicotina de los cigarros de la marca XX".

$X \sim N(\mu, \sigma = 0'05)$

muestra: $n = 15$

$P(S \leq 0'065) = P(S^2 \leq 0'065^2) = P(S^2 \leq 0'004225) =$

$$C = \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1} \rightarrow \frac{15 \cdot 0'004225}{0'05^2} = 25'35$$

$= P(C \leq 25'35) = \text{pchisq}(25'35, 14) = \boxed{0'9687}$

ejercicio 10:

$X_1 :=$ "Producción diaria de las piezas de tipo A".

$X_2 :=$ "Producción diaria de las piezas de tipo B".

$X_1 \sim N(\mu, 3)$; $X_2 \sim N(\mu, 4'3)$ } Independientes entre sí.

muestra 1: $n = 5$

muestra 2: $m = 5$

a) $P(S_1^2 \geq 2S_2^2) = P\left(\frac{S_1^2}{S_2^2} \geq 2\right) =$ Cociente de dos cuasivarianzas muestrales.

$= \cancel{P(2S_2^2 \leq S_1^2)} = \cancel{P\left(\frac{S_2^2}{S_1^2} \leq \frac{1}{2}\right)}$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1;m-1}$$

$= 1 - P\left(\frac{S_1^2}{S_2^2} < 2\right) = 1 - P\left(\frac{S_1^2}{S_2^2} \leq 2\right) =$

$= 1 - P\Bigg(\underbrace{\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}}_{\sim F_{n-1;m-1}} \leq 2 \cdot \frac{\sigma_2^2}{\sigma_1^2}\Bigg) = 1 - P\left(F \leq 2 \cdot \left(\frac{3^2}{4'3^2}\right)^{-1}\right) =$

$= 1 - P(F \leq 4'1088) = 1 - \text{pf}(4'1088, 4, 4)$

$= 1 - 0'9000 = 0'1$

b) $P\left(\frac{S_1^2}{S_2^2} \leq a\right) = 0'95$

$P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq a \cdot \frac{\sigma_2^2}{\sigma_1^2}\right) = 0'95 \Rightarrow \boxed{pf(F \leq 2'05\bar{4}\, a) = 0'95}$

$\underbrace{}_{F_{n-1;m-1}}$ $\quad a \cdot \frac{4'3^2}{3^2} = 2'05\bar{4}\, a$

$qf(0'95, 4,4) = 6'388233 = 2'05\bar{4}\, a \rightarrow \boxed{a = 3'1095}$

ejercicio 11:

$$\left.\begin{array}{l} X_1 \sim N(150, \sigma) \\ X_2 \sim N(100, \sigma) \end{array}\right\} \text{Independientes.}$$

muestra 1: $n = 16$; $S_1^2 = 64$

muestra 2: $m = 16$; $S_2^2 = 36$

$m = 12$

a) Varianzas de las poblaciones son iguales: $\sigma_1^2 = \sigma_2^2$

$$\boxed{P(\bar{X}_1 - \bar{X}_2 \leq 45)}$$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

$$\frac{45 - (150 - 100)}{\sqrt{\frac{15 \cdot 64 + 15 \cdot 36}{16+16-2}}\sqrt{\frac{1}{16} + \frac{1}{16}}} = -2$$

($15 \to 11$, $16 \to 12$)

$-1'8130$

$P(T \leq -2) = pt(-2, 30) = 0'02731252$

$P(T \leq -1'8130) = pt(-1'8130, 26) = 0'0472$

b) Varianza de las poblaciones son diferentes: $\sigma_1^2 \neq \sigma_2^2$

$$\boxed{P(\bar{X}_1 - \bar{X}_2 \leq 45)}$$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim t_v \qquad v = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n+1} + \frac{\left(\frac{S_2^2}{m}\right)^2}{m+1}} - 2$$

$\downarrow$

$$\frac{45 - (150 - 200)}{\sqrt{\frac{64}{16} + \frac{36}{12}}} = -1'8898 \sim t_v$$

$$V = \frac{\left(\frac{64}{16} + \frac{36}{12}\right)^2}{\frac{\left(\frac{64}{16}\right)^2}{17} + \frac{\left(\frac{36}{12}\right)^2}{13}} - 2 = 27'99 \simeq 28$$

$$P(t \leq -1'8898) = pt(-1'8898, 28) = 0'03458952$$

ejercicio 12:

X:= "Duración de los focos de una determinada marca de automóvil".

$X \sim N(10, 3)$

muestra: n focos

$P(|\bar{X} - \mu| \leq 1) = 0'99$

$P(|\bar{X}| = 11) = 0'99$ Media muestral, población normal σ conocida.

$\bar{X} \sim N(10, \frac{3}{\sqrt{n}}) \rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$Z = \frac{11-10}{3/\sqrt{n}} = \frac{\sqrt{n}}{3}$

$P(Z = \frac{\sqrt{n}}{3}) = 0'99$

$P(-1 \leq \bar{X} - \mu \leq 1) = 0'99$

$P\left(\frac{-1}{\sigma/\sqrt{n}} \leq \overbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}^{Z} \leq \frac{1}{\sigma/\sqrt{n}}\right) = 0'99$

$P\left(\frac{-1}{\sigma/\sqrt{n}} \leq Z \leq \frac{1}{\sigma/\sqrt{n}}\right) = 0'99$

$P\left(Z \leq \frac{1}{\sigma/\sqrt{n}}\right) - P\left(Z \leq \frac{-1}{\sigma/\sqrt{n}}\right) = 0'99$

$P\left(Z \leq \frac{1}{\sigma/\sqrt{n}}\right) - \left(1 - P\left(Z \leq \frac{1}{\sigma/\sqrt{n}}\right)\right) = 0'99$

$2P\left(Z \leq \frac{1}{\sigma/\sqrt{n}}\right) = 1'99$

$P\left(Z \leq \frac{1}{\sigma/\sqrt{n}}\right) = 0'995$

qnorm$(0'995, 0, 1) = \underline{2'5758}$