Los resultados sobre inferencia en poblaciones normales son de enorme importancia debido a su gran aplicabilidad; de hecho, en virtud de los teoremas límite, estos resultados pueden aplicarse, al menos de forma aproximada, en gran cantidad de situaciones reales si los tamaños muestrales son suficientemente grandes.

Ya que las distribuciones normales están determinadas por su media y su varianza, la inferencia en tales poblaciones se realiza sobre dichos parámetros, y los estadísticos de interés son la media y la cuasivarianza muestral.

Los resultados de este tema se refieren a la distribución de diversas variables definidas en términos de la media y la cuasivarianza muestral de una y de dos poblaciones normales; estas distribuciones permiten obtener conclusiones probabilísticas relativas a dichos estadísticos.

Comenzamos definiendo las distribuciones teóricas que aparecen asociadas al muestreo de poblaciones normales.

2.1. χ^2 de Pearson, t de Student y F de Snedecor

Distribución χ^2 de Pearson

Es un caso particular de la distribución gamma:[1]

$$X \to \chi^2(n),\ n \in \mathbb{N} \Leftrightarrow X \to \Gamma(n/2, 1/2).$$

$n \to$ *grados de libertad.*

Función de densidad: $f(x) = \dfrac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0.$

Función generatriz de momentos: $M_X(t) = \dfrac{1}{(1-2t)^{n/2}}, \quad t < 1/2.$

Momentos: $E[X^k] = 2^k \dfrac{\Gamma(n/2+k)}{\Gamma(n/2)}, \quad k \in \mathbb{N} \to \begin{cases} \text{Media}: E[X] = n \\ \text{Varianza}: Var[X] = 2n. \end{cases}$

[1] $X \to \Gamma(p, a) \Leftrightarrow f(x) = \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, \ x > 0, \quad \left(\Gamma(p) = \int_0^{+\infty} x^{p-1} e^{-x} dx\right).$

Reproductividad:

$X_1,\dots,X_n$ independientes, $X_i \to \chi^2(k_i)$, $i=1,\dots,n \Rightarrow \sum_{i=1}^{n} X_i \to \chi^2\left(\sum_{i=1}^{n} k_i\right)$.

Demostración: Por la reproductividad de la gamma en el primer parámetro:

$X_1,\dots,X_n$ independientes, $X_i \to \Gamma(k_i/2, 1/2) \Rightarrow \sum_{i=1}^{n} X_i \to \Gamma\left(\sum_{i=1}^{n} k_i/2, 1/2\right)$. □

Relación con la distribución normal:

$i)$ $X \to \mathcal{N}(0,1) \;\Rightarrow\; X^2 \to \chi^2(1)$.

$ii)$ $X_1,\dots,X_n$ independientes, $X_i \to \mathcal{N}(0,1)$, $i=1,\dots,n \Rightarrow \sum_{i=1}^{n} X_i^2 \to \chi^2(n)$.

Demostración:

$$i)\; Y = X^2 \to X = \pm\sqrt{Y} \to f_Y(y) = f_X(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| + f_X(-\sqrt{y})\left|-\frac{1}{2\sqrt{y}}\right|$$
$$= \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2},\; y > 0.$$

Ya que $\Gamma(1/2) = \sqrt{\pi}$, la densidad de Y corresponde a una $\chi^2(1)$.

$ii)$ Se obtiene de $i)$, aplicando la reproductividad de la distribución χ^2.

Tema 2. Distribuciones en el muestreo de poblaciones normales A. Hermoso Carazo
└2.1. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor
└2.1.1. Distribución χ^2 de Pearson

Tablas y aproximaciones: Existen tablas para el cálculo de probabilidades de una $\chi^2(n)$ para distintos valores de n (usualmente, hasta $n = 50$). Para valores mayores de n, la expresión como suma de variables independientes e idénticamente distribuidas, con media y varianza finitas, da lugar, por el *teorema central del límite de Lèvy*, a la siguiente aproximación:

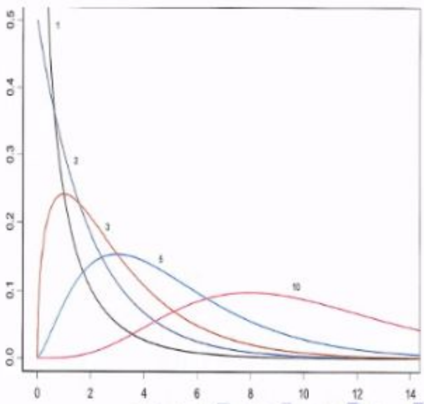
$$\chi^2(n) \approx \mathcal{N}(n, 2n).$$

Gráfica de la función de densidad de $\chi^2(n)$: Asimétrica a la derecha y unimodal.

$$n = 1 \to \lim_{x\to 0} f(x) = +\infty$$

$$n = 2 \to f(0) = 1/2$$

$$n \geq 3 \to f(0) = 0$$



Distribución t de Student

Corresponde al cociente entre una variable con distribución $\mathcal{N}(0,1)$ y la raíz cuadrada de una χ^2 dividida por sus grados de libertad, ambas independientes:

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \to \mathcal{N}(0,1) \\ Y \to \chi^2(n) \end{array}\right\} \Rightarrow T = \frac{X}{\sqrt{Y/n}} \to t(n) \quad (n \in \mathbb{N}).$$

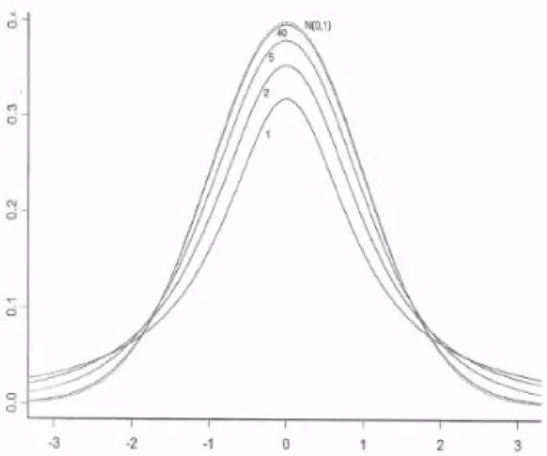
$n \to$ *grados de libertad.*

Función de densidad: $f(t) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\dfrac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}.$

Momentos: $\exists E[T^k] \Leftrightarrow k < n.$

- $n > 1 \Rightarrow \exists E[T] = 0.$
- $n > 2 \Rightarrow \exists Var[T] = \dfrac{n}{n-2}.$

Gráfica de la función de densidad de $t(n)$: Es similar a la de la $\mathcal{N}(0,1)$ (simétrica alrededor del cero y unimodal) y, de hecho, se aproxima a ella cuando $n \to +\infty$. La varianza es mayor que uno (más dispersión que en la normal); las colas son más gruesas y la gráfica es más aplastada (*platicúrtica*).



Tablas: Está tabulada (tablas hasta $n = 100$) y para n grande se aproxima por la $\mathcal{N}(0,1)$.

Distribución F de Snedecor

Es la distribución del cociente entre dos variables independientes con distribución χ^2, cada una dividida por sus grados de libertad:

$$\left.\begin{array}{l} X, Y \text{ independientes} \\ X \to \chi^2(m) \\ Y \to \chi^2(n) \end{array}\right\} \Rightarrow F = \frac{X/m}{Y/n} \to F(m,n) \quad (m, n \in \mathbb{N}).$$

$m, n \to$ *grados de libertad.*

Función de densidad:

$$g(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} f^{m/2-1} \left(1 + \frac{m}{n} f\right)^{-\frac{m+n}{2}}, \quad f > 0.$$

Momentos: $\exists E[F^k] \Leftrightarrow k < n/2$.

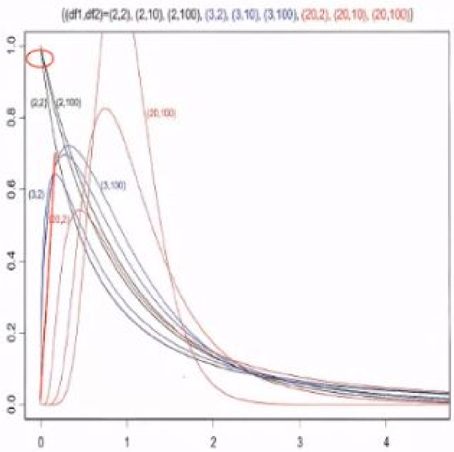
- $n > 2 \Rightarrow \exists E[F] = \dfrac{n}{n-2}$.
- $n > 4 \Rightarrow \exists Var[F] = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$.

Propiedades:

- $F \to F(m,n) \Leftrightarrow F^{-1} \to F(n,m)$.
- $T \to t(n) \Leftrightarrow T^2 \to F(1,n)$.

Gráfica de la función de densidad de $F(m,n)$: Asimétrica y unimodal:

- $F(1,n)$: $\lim\limits_{f \to 0} g(f) = +\infty$.
- $F(2,n)$: $\lim\limits_{f \to 0} g(f) = 1$.
- $F(m,n)$, $m > 2$: $\lim\limits_{f \to 0} g(f) = 0$.



Tablas: Como las anteriores, esta distribución está tabulada. Las tablas incluyen aproximaciones para valores grandes de m y n.

2.2. Muestreo en una población normal unidimensional

Sea $X \to \mathcal{N}(\mu, \sigma^2)$ y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. En esta sección veremos algunos resultados referentes a la distribución de los estadísticos media y cuasivarianza muestral:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}.$$

Herramientas de Cálculo de Probabilidades:

- *F. generatriz de momentos de* $Z \to \mathcal{N}(\mu, \sigma^2)$: $M_Z(t) = e^{t\mu + t^2\sigma^2/2}$, $t \in \mathbb{R}$.
- *T. Unicidad de la f.g.m: Si existe, la función generatriz de momentos determina la distribución de la variable.*
- *Funciones medibles de variables independientes son independientes.*
- *T. Multiplicación de Esperanzas:* X, Y *independientes,* $\exists E[X], E[Y]$, *entonces,* $\exists E[XY] = E[X]E[Y]$.
- *Caracterización de independencia por f.g.m.: Sean* $\boldsymbol{X}, \boldsymbol{Y}$ *vectores aleatorios con función generatriz de momentos* $M_{\boldsymbol{X}}$ *y* $M_{\boldsymbol{Y}}$:

  $\boldsymbol{X}, \boldsymbol{Y}$ *son independientes* $\Leftrightarrow \exists M_{(\mathbf{X},\mathbf{Y})}(t_1, t_2) = M_{\mathbf{X}}(t_1) M_{\mathbf{Y}}(t_2)$.

Teorema 2.2.1

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0, 1).$$

Demostración: En el Ejemplo 1.4.1. se probó, calculando su función generatriz de momentos, que $\overline{X} \to \mathcal{N}(\mu, \sigma^2/n)$. El resultado se deduce tipificando. □

Teorema 2.2.2

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

Demostración: Ya que $X_1, \ldots, X_n$ son independientes, las tipificadas también lo son y cada una tiene distribución $\mathcal{N}(0, 1)$. Por tanto, la variable de interés es suma de cuadrados de $\mathcal{N}(0, 1)$ independientes y su distribución es $\chi^2(n)$:

$$X_i \to \mathcal{N}(\mu, \sigma^2) \Rightarrow \frac{X_i - \mu}{\sigma} \to \mathcal{N}(0, 1) \Rightarrow \frac{(X_i - \mu)^2}{\sigma^2} \to \chi^2(1), \quad i = 1, \ldots, n$$

Teorema 2.2.3

$\overline{X}$ y $(X_1 - \overline{X}, \dots, X_n - \overline{X})$ son independientes.

Demostración: Calculamos la función generatriz de momentos conjunta para usar la caracterización de independencia correspondiente.

$$M_{(\overline{X}, X_1-\overline{X}, \dots, X_n-\overline{X})}(t, t_1, \dots, t_n) = E\left[e^{t\overline{X}+\sum\limits_{i=1}^n t_i(X_i-\overline{X})}\right]^{2}$$

$$= E\left[e^{\sum\limits_{i=1}^n X_i(t_i-\overline{t}+t/n)}\right] = E\left[\prod_{i=1}^n e^{X_i(t_i-\overline{t}+t/n)}\right] \overset{TME}{=} \prod_{i=1}^n E\left[e^{X_i(t_i-\overline{t}+t/n)}\right]$$

$$= \prod_{i=1}^n M_{X_i}(t_i - \overline{t} + t/n) = \prod_{i=1}^n e^{\mu(t_i-\overline{t}+t/n)+\frac{\sigma^2}{2}(t_i-\overline{t}+t/n)^2}$$

---

[2] $\overline{t} = \dfrac{\sum\limits_{i=1}^n t_i}{n} \to t\overline{X} + \sum\limits_{i=1}^n t_i(X_i - \overline{X}) = \sum\limits_{i=1}^n t_i X_i + \overline{X}(t - n\overline{t}) = \sum\limits_{i=1}^n t_i X_i + \sum\limits_{i=1}^n X_i\left(\dfrac{t}{n} - \overline{t}\right).$

$$M_{(\overline{X}, X_1-\overline{X}, \dots, X_n-\overline{X})}(t, t_1, \dots, t_n) = e^{\mu\sum\limits_{i=1}^n (t_i-\overline{t}+t/n)+\frac{\sigma^2}{2}\sum\limits_{i=1}^n (t_i-\overline{t}+t/n)^2}\ ^{3}$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2n} + \frac{\sigma^2}{2}\sum\limits_{i=1}^n (t_i-\overline{t})^2}, \quad t, t_1 \dots, t_n \in \mathbb{R}.$$

- $M_{\overline{X}}(t) = M_{(\overline{X}, X_1-\overline{X}, \dots, X_n-\overline{X})}(t, 0, \dots, 0) = e^{\mu t + \frac{\sigma^2 t^2}{2n}}, \quad t \in \mathbb{R}.$
- $M_{(X_1-\overline{X}, \dots, X_n-\overline{X})}(t_1, \dots, t_n) = M_{(\overline{X}, X_1-\overline{X}, \dots, X_n-\overline{X})}(0, t_1, \dots, t_n)$
  $= e^{\frac{\sigma^2}{2}\sum\limits_{i=1}^n (t_i-\overline{t})^2}, \quad t_1 \dots, t_n \in \mathbb{R}.$

Por tanto, la función generatriz de momentos conjunta es producto de las marginales, y los vectores son independientes. □

---

[3] $\sum\limits_{i=1}^n (t_i - \overline{t} + t/n) = \sum\limits_{i=1}^n (t_i - \overline{t}) + t = t.$
$\sum\limits_{i=1}^n (t_i - \overline{t} + t/n)^2 = \sum\limits_{i=1}^n (t_i - \overline{t})^2 + t^2/n + 2\sum\limits_{i=1}^n (t_i - \overline{t})t/n = \sum\limits_{i=1}^n (t_i - \overline{t})^2 + t^2/n.$

Lema de Fisher

$\overline{X}$ y S^2 son independientes.

Demostración: Consecuencia inmediata del Teorema 2.2.3, ya que S^2 sólo depende de $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$. □

Teorema 2.2.4

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1)$$

Demostración:

$$\begin{aligned}\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-\mu)^2\\ &= \underbrace{\sum_{i=1}^{n}(X_i-\overline{X})^2}_{(n-1)S^2} + n(\overline{X}-\mu)^2 + 2\underbrace{\sum_{i=1}^{n}(X_i-\overline{X})(\overline{X}-\mu)}_{0}.\end{aligned}$$

$$\Downarrow$$

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}}_{A} = \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{B} + \underbrace{\frac{n(\overline{X}-\mu)^2}{\sigma^2}}_{C}$$

- $A \to \chi^2(n)$ (Teorema 2.2.2).
- $C \to \chi^2(1)$ (es el cuadrado de una $\mathcal{N}(0,1) \to$ Teorema 2.2.1).
- B y C independientes (B es función de S^2 y C función de $\overline{X} \to$ Lema de Fisher).

Calculamos ahora la función generatriz de momentos de A, teniendo en cuenta que $A = B + C$ y B, C son independientes:

$$M_A(t) = E\left[e^{tA}\right] = E\left[e^{t(B+C)}\right] = E\left[e^{tB}e^{tC}\right] \overset{TME}{=} E\left[e^{tB}\right]E\left[e^{tC}\right] = M_B(t)M_C(t).$$

Así, como $A \to \chi^2(n)$ y $C \to \chi^2(1)$, tenemos:

$$\frac{1}{(1-2t)^{n/2}} = M_B(t)\frac{1}{(1-2t)^{1/2}}, \quad t < 1/2,$$

y despejando:

$$M_B(t) = \frac{1}{(1-2t)^{\frac{n-1}{2}}}, \quad t < 1/2.$$

Ya que esta es la función generatriz de momentos de la distribución $\chi^2(n-1)$, por el *teorema de unicidad de la función generatriz de momentos* se deduce que esa es la distribución de B.

Teorema 2.2.5 (Razón de Student)

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \to t(n-1).$$

Demostración: Basta aplicar la definición de la distribución t a las variables (independientes, por el *Lema de Fisher):*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1), \qquad \frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1).$$

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\sqrt{n}(\overline{X} - \mu)/\sigma}{S/\sigma} = \frac{\sqrt{n}(\overline{X} - \mu)}{S} \to t(n-1). \quad \square$$

ESQUEMA DE RESULTADOS

Variable	Distribución	Uso
$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$	Inferencia sobre μ cuando σ^2 es conocida
$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	$t(n-1)$	Inferencia sobre μ cuando σ^2 es desconocida
$\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$	$\chi^2(n)$	Inferencia sobre σ^2 cuando μ es conocida
$\dfrac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$	Inferencia sobre σ^2 cuando μ es desconocida

2.3. Muestreo en dos normales unidimensionales

En muchas ocasiones se plantea el problema de comparar dos distribuciones normales, para lo que deben compararse sus medias y sus varianzas.

Concretamente, los resultados que presentamos seguidamente son la base para hacer inferencia sobre la diferencia de medias y sobre el cociente de varianzas de ambas variables.

Consideramos dos variables normales, $X \rightarrow \mathcal{N}(\mu_1,\sigma_1^2)$, $Y \rightarrow \mathcal{N}(\mu_2,\sigma_2^2)$ y para realizar la inferencia, tomamos una muestra aleatoria simple de cada una de ellas, ambas independientes (lo que, en la práctica, no supone pérdida de generalidad). Trabajaremos con las medias y cuasivarianzas muestrales:

- $(X_1,\ldots,X_{n_1})$ m.a.s de $X \longrightarrow \overline{X}=\dfrac{\sum\limits_{i=1}^{n_1}X_i}{n_1}, \quad S_1^2=\dfrac{\sum\limits_{i=1}^{n_1}(X_i-\overline{X})^2}{n_1-1}.$
- $(Y_1,\ldots,Y_{n_2})$ m.a.s de $Y \longrightarrow \overline{Y}=\dfrac{\sum\limits_{j=1}^{n_2}Y_j}{n_2}, \quad S_2^2=\dfrac{\sum\limits_{j=1}^{n_2}(Y_j-\overline{Y})^2}{n_2-1}.$
- $(X_1,\ldots,X_{n_1})$ e $(Y_1,\ldots,Y_{n_2})$ independientes.

Extensión del Lema de Fisher

Los vectores $(\overline{X},\overline{Y})$ y (S_1^2,S_2^2) son independientes.

Demostración: Usamos la caracterización de independencia por funciones generatrices de momentos. Aplicando la independencia de las muestras (1) y el lema de Fisher (2) en cada una:

$$M_{(\overline{X},\overline{Y},S_1^2,S_2^2)}(s,t,u,v) \overset{(1)}{=} M_{(\overline{X},S_1^2)}(s,u)M_{(\overline{Y},S_2^2)}(t,v)$$
$$\overset{(2)}{=} M_{\overline{X}}(s)M_{S_1^2}(u)M_{\overline{Y}}(t)M_{S_2^2}(v) \overset{(1)}{=} M_{(\overline{X},\overline{Y})}(s,t)M_{(S_1^2,S_2^2)}(u,v). \qquad \square$$

Teorema 2.3.1

$$\overline{X}-\overline{Y} \to \mathcal{N}\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right).$$

Demostración: Por la independencia de $\overline{X}$ e $\overline{Y}$, $M_{\overline{X}-\overline{Y}}(t) = M_{\overline{X}}(t)M_{-\overline{Y}}(t)$ $= M_{\overline{X}}(t)M_{\overline{Y}}(-t)$. Por tanto:

$$M_{\overline{X}-\overline{Y}}(t) = e^{t\mu_1+\frac{t^2}{2}\frac{\sigma_1^2}{n_1}}e^{-t\mu_2+\frac{t^2}{2}\frac{\sigma_2^2}{n_2}} = e^{t(\mu_1-\mu_2)+\frac{t^2}{2}\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$$

Teorema 2.3.2

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2}}\sqrt{\frac{\sigma_1^2/n_1+\sigma_2^2/n_2}{n_1+n_2-2}}} \to t(n_1+n_2-2).$$

Demostración: Buscamos una $\mathcal{N}(0,1)$ y una $\chi^2(n_1+n_2-2)$, independientes, y aplicamos la definición de la t de Student:

- $A = \dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \to \mathcal{N}(0,1)$ (Teorema 2.3.1).
- $B = \dfrac{(n_1-1)S_1^2}{\sigma_1^2}+\dfrac{(n_2-1)S_2^2}{\sigma_1^2} \to \chi^2(n_1+n_2-2)$.
  Cada sumando es $\chi^2(n_i-1)$ (Teorema 2.2.4) y, al ser independientes, de la reproductividad de la χ^2 se tiene que la suma es $\chi^2(n_1-1+n_2-1)$.
- A y B son independientes (extensión del Lema de Fisher).

$$\frac{A}{\sqrt{B/n_1+n_2-2}} \to t(n_1+n_2-2).$$

Corolario

$$\sigma_1^2 = \sigma_2^2 \;\Rightarrow\; \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \to t(n_1+n_2-2)$$

Teorema 2.3.3

$$\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ n_1\sigma_1^2}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2 \Big/ n_2\sigma_2^2} \to F(n_1, n_2).$$

Demostración: Para aplicar la definición de la $F(n_1, n_2)$, buscamos una $\chi^2(n_1)$ y una $\chi^2(n_2)$ independientes:

- $A_1 = \sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2 \Big/ \sigma_1^2 \to \chi^2(n_1)$ (Teorema 2.2.2)
- $A_2 = \sum\limits_{j=1}^{n_2}(Y_j-\mu_2)^2 \Big/ \sigma_2^2 \to \chi^2(n_2)$ (Teorema 2.2.2)
- A_1 y A_2 son independientes (independencia de muestras).

$$\frac{A_1/n_1}{A_2/n_2} \to F(n_1, n_2).$$

Teorema 2.3.4

$$\frac{S_1^2 \big/ \sigma_1^2}{S_2^2 \big/ \sigma_2^2} \rightarrow F(n_1-1, n_2-1).$$

Demostración: Es análoga a la anterior, aplicando ahora el Teorema 2.2.4:

- $A_i = \dfrac{(n_i-1)S_i^2}{\sigma_i^2} \rightarrow \chi^2(n_i-1), \quad i=1,2$ (Teorema 2.2.4)
- A_1 y A_2 son independientes (independencia de muestras).

$$\frac{A_1/n_1-1}{A_2/n_2-1} \rightarrow F(n_1-1, n_2-1).$$

ESQUEMA DE RESULTADOS

Variable	Distribución	Uso
$\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2 \big/ n_1\sigma_1^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2 \big/ n_2\sigma_2^2}$	$F(n_1, n_2)$	Inferencia sobre $\dfrac{\sigma_2^2}{\sigma_1^2}$ (μ_1, μ_2 conocidas)
$\dfrac{S_1^2 \big/ \sigma_1^2}{S_2^2 \big/ \sigma_2^2}$	$F(n_1-1, n_2-1)$	Inferencia sobre $\dfrac{\sigma_2^2}{\sigma_1^2}$ (μ_1, μ_2 desconocidas)

ESQUEMA DE RESULTADOS

Variable	Distribución	Uso
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	$\mathcal{N}(0,1)$	Inferencia sobre $\mu_1-\mu_2$ (σ_1^2, σ_2^2 conocidas)
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}+\dfrac{(n_2-1)S_2^2}{\sigma_2^2}}\sqrt{\dfrac{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}{n_1+n_2-2}}}$	$t(n_1+n_2-2)$	
$\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	$t(n_1+n_2-2)$	Inferencia sobre $\mu_1-\mu_2$ ($\sigma_1^2=\sigma_2^2$ desconocidas)

Problema 1

Se toma una muestra aleatoria simple de tamaño 5 de una variable aleatoria con distribución $\mathcal{N}(2.5, 36)$. Calcular:

a) Probabilidad de que la cuasivarianza muestral esté comprendida entre 1.863 y 2.674.

b) Probabilidad de que la media muestral esté comprendida entre 1.3 y 3.5, supuesto que la cuasivarianza muestral está entre 30 y 40.

$a)$ $P\left(1.863 < S^2 < 2.674\right)$

$$\frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1) \quad \longrightarrow \quad Y = \frac{4S^2}{36} = \frac{S^2}{9} \to \chi^2(4).$$

$$P\left(1.863 < S^2 < 2.674\right) = P\left(\frac{1.863}{9} < Y < \frac{2.674}{9}\right) = P\left(0.207 < Y < 0.2971\right)$$
$$= P\left(Y > 0.207\right) - P\left(Y > 0.2971\right) = 0.995 - 0.99$$
$$= 0.005.$$

$$b)\ P\left(1.3 < \overline{X} < 3.5 \big/ 30 < S^2 < 40\right) \underset{\substack{\uparrow \\ \text{Lema de Fisher:}}}{=} P\left(1.3 < \overline{X} < 3.5\right).$$

$$\overline{X} \to \mathcal{N}(\mu, \sigma^2/n) \equiv \mathcal{N}(2.5, 36/5) \quad \Rightarrow \quad Z = \frac{\overline{X} - 2.5}{\sqrt{36/5}} \to \mathcal{N}(0,1).$$

$$\begin{aligned} P\left(1.3 < \overline{X} < 3.5\right) &= P\left(\frac{1.3 - 2.5}{\sqrt{36/5}} < Z < \frac{3.5 - 2.5}{\sqrt{36/5}}\right) \\ &= P\left(-0.44721 < Z < 0.37268\right) \\ &= P\left(Z < 0.37268\right) - P\left(Z < -0.44721\right) \\ &= 0.64531 - (1 - 0.67264) = 0.31795. \end{aligned}$$

Problema 2

La longitud craneal en una determinada población humana es una variable aleatoria que sigue una distribución normal con media 185.6 mm. y desviación típica 12.78 mm. ¿Cuál es la probabilidad de que una muestra aleatoria simple de tamaño 20 de esa población tenga media mayor que 190 mm.?

$$(X_1, \ldots, X_{20}) \text{ m.a.s de } X \to \mathcal{N}(185.6, 12.78^2) \quad \longrightarrow \quad P\left(\overline{X} > 190\right)$$

$$\overline{X} \to \mathcal{N}\left(185.6, \frac{12.78^2}{20}\right) \quad \Rightarrow \quad Z = \frac{\overline{X} - 185.6}{\sqrt{\frac{12.78^2}{20}}} = \frac{\overline{X} - 185.6}{\sqrt{8.16642}} \to \mathcal{N}(0,1).$$

$$P\left(\overline{X} > 190\right) = P\left(Z > \frac{190 - 185.6}{\sqrt{8.16642}}\right) = P\left(Z > 1.5397\right) = 0.06182.$$

Problema 3

¿De qué tamaño mínimo habría que seleccionar una muestra de una variable con distribución normal $\mathcal{N}(\mu, 4)$ para poder afirmar, con probabilidad mayor que 0.9, que la media muestral diferirá de la poblacional menos de 0.1?

$$(X_1, \ldots, X_n) \text{ m.a.s de } X \to \mathcal{N}(\mu, 4) \quad \longrightarrow \quad n \;/\; P\left(|\overline{X} - \mu| < 0.1\right) > 0.9.$$

$$\overline{X} \to \mathcal{N}\left(\mu, 4/n\right) \Rightarrow Z = \frac{\overline{X} - \mu}{2/\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \mu)}{2} \to \mathcal{N}(0, 1).$$

$$P\left(|\overline{X} - \mu| < 0.1\right) = P\left(|Z| < \frac{0.1\sqrt{n}}{2}\right) = P\left(|Z| < 0.05\sqrt{n}\right)$$

$$= 2P\left(Z < 0.05\sqrt{n}\right) - 1 > 0.9.$$

Por tanto:

$$P\left(Z < 0.05\sqrt{n}\right) > 0.95 \;\Rightarrow\; 0.05\sqrt{n} > 1.64486 \Rightarrow n > 1082.22 \Rightarrow n \geq 1083.$$

Problema 4

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable con distribución normal. Calcular la probabilidad de que la cuasivarianza muestral sea menor que un 50% de la varianza poblacional para $n = 16$ y para $n = 1000$.

$$(X_1, \ldots, X_n) \text{ m.a.s de } X \to \mathcal{N}(\mu, \sigma^2) \quad \longrightarrow \quad P\left(S^2 < 0.5\sigma^2\right).$$

$$Y = \frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1).$$

$$\boxed{n = 16} \quad Y = \frac{15S^2}{\sigma^2} \to \chi^2(15)$$

$$P\left(S^2 < 0.5\sigma^2\right) = P\left(Y < 7.5\right) = 0.059297.$$

$$\boxed{n = 1000} \quad Y = \frac{999S^2}{\sigma^2} \to \chi^2(999) \simeq \mathcal{N}(999, 1998) \Rightarrow \frac{Y - 999}{\sqrt{1998}} \approx Z \to \mathcal{N}(0, 1).$$

$$P\left(S^2 < 0.5\sigma^2\right) = P\left(Y < 499.5\right) \approx P\left(Z < -11.17\right) \approx 0.$$

Problema 5

Sean S_1^2 y S_2^2 las cuasivarianzas muestrales de dos muestras independientes de tamaños $n_1 = 5$ y $n_2 = 4$ de dos poblaciones normales con la misma varianza. Calcular la probabilidad de que S_1^2/S_2^2 sea menor que 5.34 o mayor que 9.12.

$(X_1, \ldots, X_5)$ m.a.s de $X \to \mathcal{N}(\mu_1, \sigma^2)$

$(Y_1, \ldots, Y_4)$ m.a.s de $Y \to \mathcal{N}(\mu_2, \sigma^2)$ $\quad \longrightarrow \quad P\left(\frac{S_1^2}{S_2^2} < 5.34 \text{ ó } \frac{S_1^2}{S_2^2} > 9.12\right)$

$(X_1, \ldots, X_5)$, $(Y_1, \ldots, Y_4)$ independientes

$$\frac{S_1^2 \big/ \sigma_1^2}{S_2^2 \big/ \sigma_2^2} = \frac{S_1^2}{S_2^2} \to F(4,3).$$

$$P\left(\frac{S_1^2}{S_2^2} < 5.34 \text{ ó } \frac{S_1^2}{S_2^2} > 9.12\right) = P\left(\frac{S_1^2}{S_2^2} < 5.34\right) + P\left(\frac{S_1^2}{S_2^2} > 9.12\right)$$

$$= 0.9 + 0.05 = 0.95.$$

Problema 6

Se consideran dos poblaciones de bombillas cuyas longitudes de vida siguen una ley normal con la misma media y desviaciones típicas 425 y 375 horas, respectivamente. Con objeto de realizar un estudio comparativo de ambas poblaciones, se considera una muestra aleatoria simple de 10 bombillas en la primera población y una de tamaño 6 en la segunda. ¿Cuál es la probabilidad de la diferencia entre las medias muestrales del primer y segundo grupo sea menor que la observada en dos realizaciones muestrales que dieron 1325 horas y 1215 horas, respectivamente?

$(X_1, \ldots, X_{10})$ m.a.s. de X: *longitud de vida de la primera población.*

$(Y_1, \ldots, Y_6)$ m.a.s. de Y: *longitud de vida de la segunda población.*

$(X_1, \ldots, X_{10})$, $(Y_1, \ldots, Y_6)$ independientes.

$$P\left(\overline{X} - \overline{Y} < 110\right)$$

$* \; X \to \mathcal{N}(\mu, 425^2) \Rightarrow \overline{X} \to \mathcal{N}\left(\mu, \dfrac{425^2}{10}\right).$

$* \; Y \to \mathcal{N}(\mu, 375^2) \Rightarrow \overline{Y} \to \mathcal{N}\left(\mu, \dfrac{375^2}{6}\right).$

$* \; \overline{X}, \; \overline{Y}$ independientes.

$$\overline{X} - \overline{Y} \to \mathcal{N}\left(0, \frac{425^2}{10} + \frac{375^2}{6}\right) \equiv \mathcal{N}\left(0, 41500\right) \Rightarrow Z = \frac{\overline{X} - \overline{Y}}{\sqrt{41500}} \to \mathcal{N}(0,1)$$

$$P\left(\overline{X} - \overline{Y} < 110\right) = P\left(Z < \frac{110}{\sqrt{41500}}\right) = P\left(Z < 0.5399\right) \approx 0.7054.$$