UNIDAD DIDÁCTICA 5

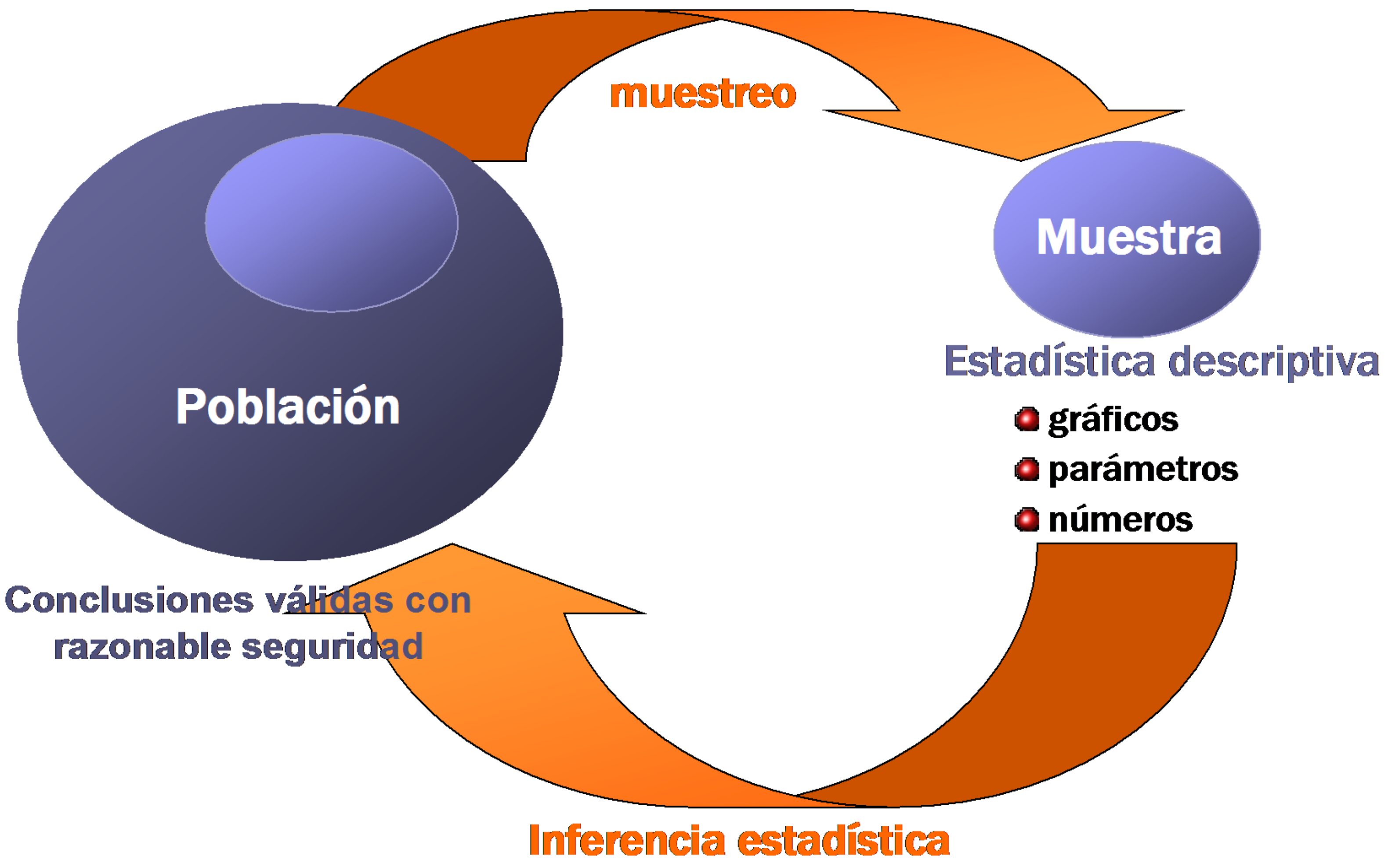
INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

(1ª parte: Distribuciones en el muestreo)

CONTENIDOS UD 5

1. Distribuciones en el muestreo
2. Inferencia básica en poblaciones normales
3. Análisis de la varianza
4. Introducción a la regresión lineal

Estadística Descriptiva e Inferencia Estadística



Contenidos distribuciones en el muestreo

1. Propiedades de la media y varianzas muestrales

2. Distribución Chi-cuadrado de Pearson

3. Distribución t de Student

4. Distribución F de Fisher

Propiedades de la media y varianza muestrales

Ejercicio 1:

Un programador ha desarrollado un sistema de depuración de programas, que en promedio tarda m=60 minutos. En una muestra de 10 programas seleccionados al azar se ha obtenido un tiempo medio de 62 minutos y una desviación típica de 2 minutos.

Una muestra como esta ¿es admisible cuando se muestrea una población de media m=60?

O por el contrario ¿la muestra está evidenciando que la media poblacional es mayor que 60 y hay, en consecuencia, que proceder a reajustar el sistema?

Propiedades de la media y varianza muestrales

- Para contestar a las preguntas anteriores es necesario tener una idea de hasta qué punto la media de una muestra puede diferir de la media m de la población.
- En este apartado se introducen algunos resultados respecto a las distribuciones de la media y desviación típicas muestrales.

Propiedades de la media y varianza muestrales

- Sea una población a cuyos individuos va asociada una variable aleatoria X.

- Para obtener conclusiones sobre la distribución de esta variable se toman muestras aleatorias constituidas por N individuos.

Propiedades de la media y varianza muestrales

- De una misma población es posible “a priori” obtener una gran cantidad de muestras diferentes de tamaño N.

- Existe por tanto una **población de posibles muestras.**

Propiedades de la media y varianza muestrales

- A cada individuo de esta nueva población se le pueden hacer corresponder diferentes características numéricas.

- Por ejemplo: la media muestral o la desviación típica muestral s, serán por tanto, nuevas variables aleatorias, cuyas características dependen de la distribución de X.

Propiedades de la media y varianza muestrales

Ejercicio 2:

¿Tienen sentido las siguientes expresiones?

La media de la media muestral.

La varianza de la media muestral.

La media de la varianza muestral.

La varianza de la varianza muestral.

¿Podrías intuitivamente avanzar cuáles crees que serán los valores de algunas de las características señaladas?

Propiedades de la media y varianza muestrales

- Estos valores dependen de la distribución de la variables aleatoria X.

- La base de la Inferencia Estadística es, precisamente, el conocimiento de las relaciones que ligan la distribución de las características muestrales con la distribución de la población.

Propiedades de la media y varianza muestrales

La media de las medias muestrales es la media de la población $E(\overline{X})=m$

La varianza de las medias muestrales es σ^2/N

La media de las varianzas muestrales es la varianza de la población $E(s^2)= \sigma^2$

Propiedades de la media y varianza muestrales

- Si N es grande las medias muestrales se distribuyen con el modelo normal $N(m, \sigma/\sqrt{N})$

Si la variable aleatoria sigue la distribución normal

- Las medias muestrales se distribuyen $N(m, \sigma/\sqrt{N})$ $\forall N$

- La distribución de las varianzas muestrales es asimétrica positiva para N pequeño

Propiedades de la media y varianza muestrales

Ejercicio 3:

El tiempo de transferencia de paquetes de un determinado tamaño a través de la red en milisegundos sigue una distribución N(m, σ=4).

a) Si se extrae una muestra aleatoria simple de tamaño N=25, calcular la probabilidad de que la diferencia entre la media poblacional y la media muestral sea mayor que 2 en valor absoluto.

b) Calcular cuánto debe valer como mínimo N si se desea una probabilidad inferior al 1% de que la diferencia en valor absoluto entre la media muestral y la media poblacional sea superior a 2.

Propiedades de la media y varianza muestrales

SOLUCIÓN:

a) Si X es N(m, σ=4) $\Rightarrow$ La media muestral $\overline{X}$ de una muestra aleatoria simple de tamaño 25 es N(m, σ=4/5) , y $\overline{X}$-m será N(0, σ=4/5).

Por tanto:

$P(|\overline{X}-m|>2) = P(|N(0,4/5)|>2) =$

$= P(N(0, 4/5)>2) + P(N(0, 4/5)<-2) =$

$= 2P(N(0,4/5)>2) = 2\,P(N(0,1)>10/4) =$

$= 2\,P(N(0,1)>2.5) =$

$= 2 \times 0.0062 =$ **0.0124**

Propiedades de la media y varianza muestrales

SOLUCIÓN:

b) Si la población es normal entonces $\bar{X}$ es $N(m, \frac{4}{\sqrt{N}})$.

$P(|\bar{X}-m|>2)<0{,}01 \Rightarrow 2\,P(\bar{X}-m>2)<0{,}01 \Rightarrow$

$P(\bar{X}-m>2)<0{,}005 \Rightarrow$

$P(N(0,1)>\frac{2}{4/\sqrt{N}})<0{,}005 \Rightarrow \quad \frac{\sqrt{N}}{2} \geq 2{,}58 \Rightarrow$

$N \geq (2x2{,}58)^2 = 26{,}62$

por lo que el valor de **N mínimo es 27**.

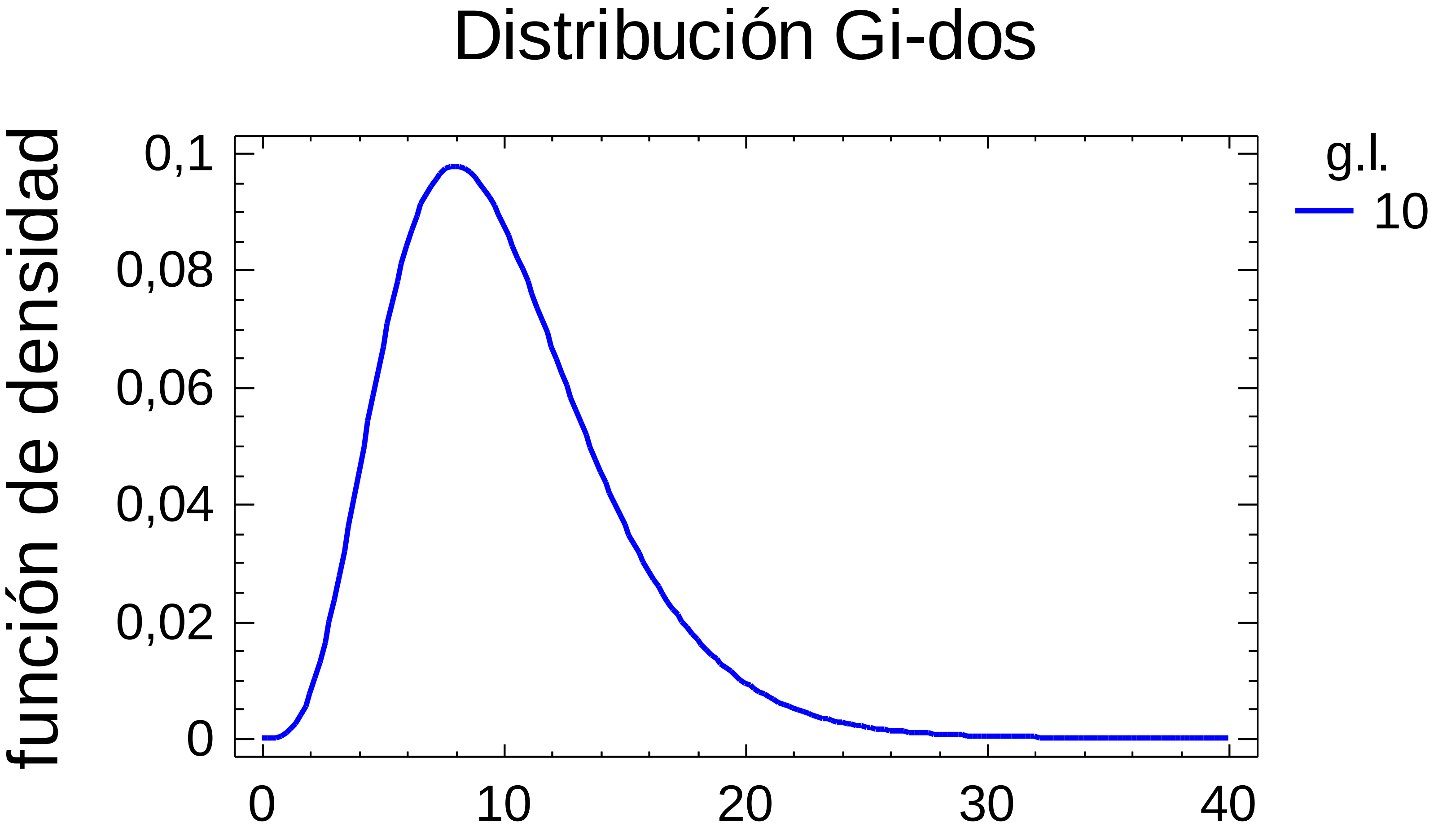
Distribución Chi-cuadrado de Pearson

- Se aplica en el estudio de la varianza muestral cuando los datos siguen la distribución normal.

- En el modelo teórico se define como la suma de ν términos al cuadrado independientes.

- La media es ν (grados de libertad) y la varianza es 2ν.

Distribución Chi-cuadrado de Pearson



Distribución Chi-cuadrado de Pearson

- Si s^2 es la varianza en una muestra de tamaño N extraída de una población normal de varianza σ^2 :

$$(N-1)\frac{s^2}{\sigma^2} \sim \chi^2 \text{ con } (N\text{-}1) \text{ g.l.}$$

Distribución Chi-cuadrado de Pearson

Ejercicio 4:

¿Cuál es la probabilidad de obtener una varianza muestral superior a 10 al sacar una muestra de tamaño 20 de una población normal de varianza igual a 5?

Distribución Chi-cuadrado de Pearson

SOLUCIÓN:

$P(s^2 > 10) = P\left((N-1)\frac{s^2}{\sigma^2} > (N-1)\frac{10}{\sigma^2}\right) = P\left(19\frac{s^2}{5} > 19\frac{10}{5}\right) =$

$= P(\chi_{19}^2 > 38)$

Buscando en la tabla se tiene

P(Chi-cuadrado con 19 g.l.>38,582)=0,005

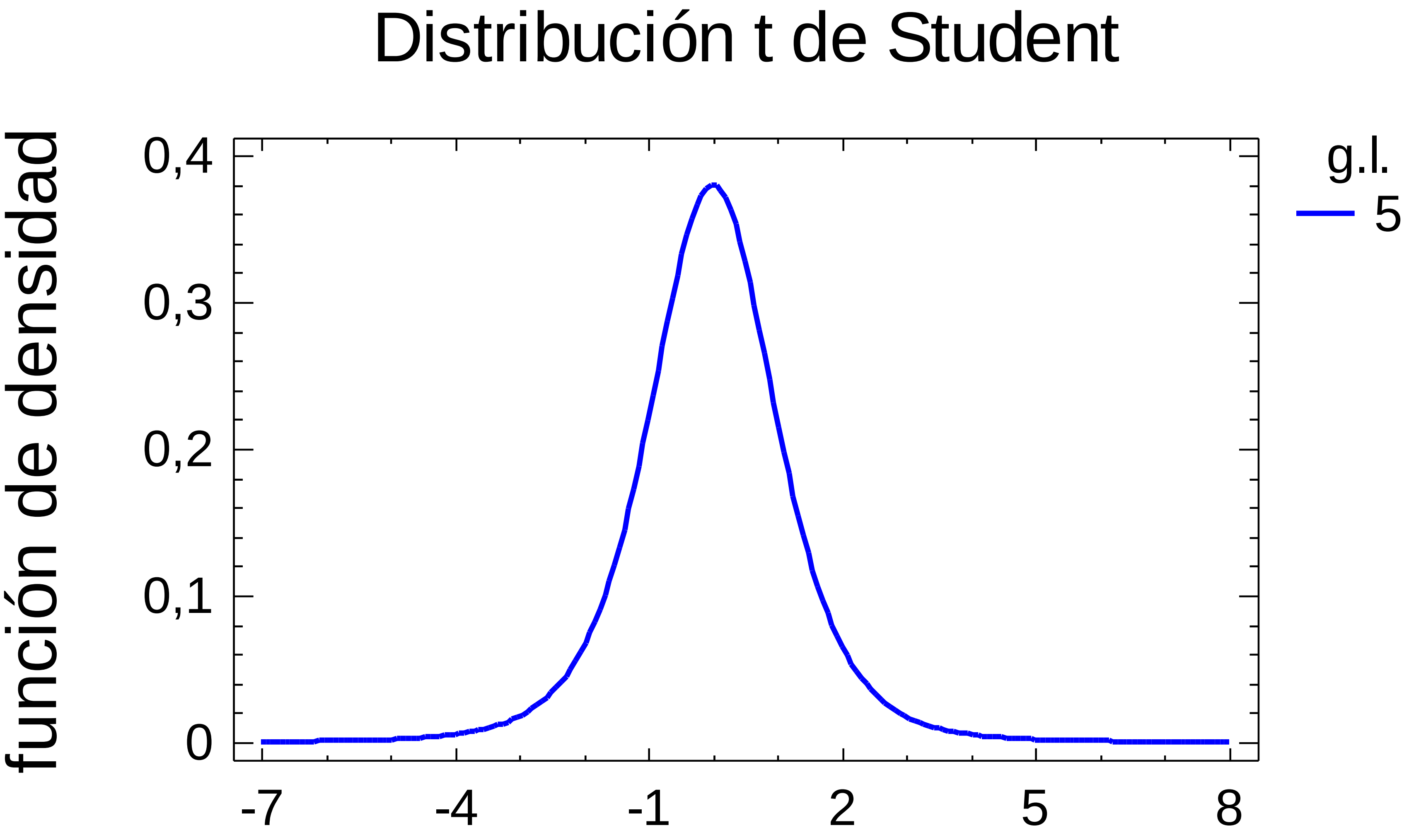
P(Chi-cuadrado con 19 g.l.>36,191)=0,01 $\Rightarrow$

0,005<P(Chi cuadrado con 19 g.l.>38)<0,01

Distribución t de Student

- Si $\overline{X}$ y s son la media y la desviación típica de muestras de tamaño N de una población normal de media m:

$$\frac{\overline{X}-m}{s/\sqrt{N}} \sim \text{t de Student con } (N-1) \text{ g.l.}$$


Distribución t de Student
función de densidad
0,4
0,3
0,2
0,1
0
-7
-4
-1
2
5
8
g.l.
5

Distribución t de Student

Ejercicio 5:

El voltaje de un circuito sigue una distribución normal. Se extraen muestras de tamaño N=10. La media muestral vale 90 y s=3.

Calcula un intervalo que contenga el valor de la media m de la población con un 80% de probabilidad.

Distribución t de Student

SOLUCIÓN:

$$\frac{\overline{X}-m}{s/\sqrt{N}} \sim \text{t de Student con } (N-1) \text{ g.l.}$$

$$\frac{90-m}{3/\sqrt{10}} \sim \text{t de Student con 9 g.l.}$$

$$P\left(-t < \frac{90-m}{3/\sqrt{10}} < t\right) = 0{,}8$$ *de la tabla t=1,383* $\Rightarrow$

[90-1,383x0,9487, 90+1,383x0,9487]$\Rightarrow$

[88,69, 91,31]

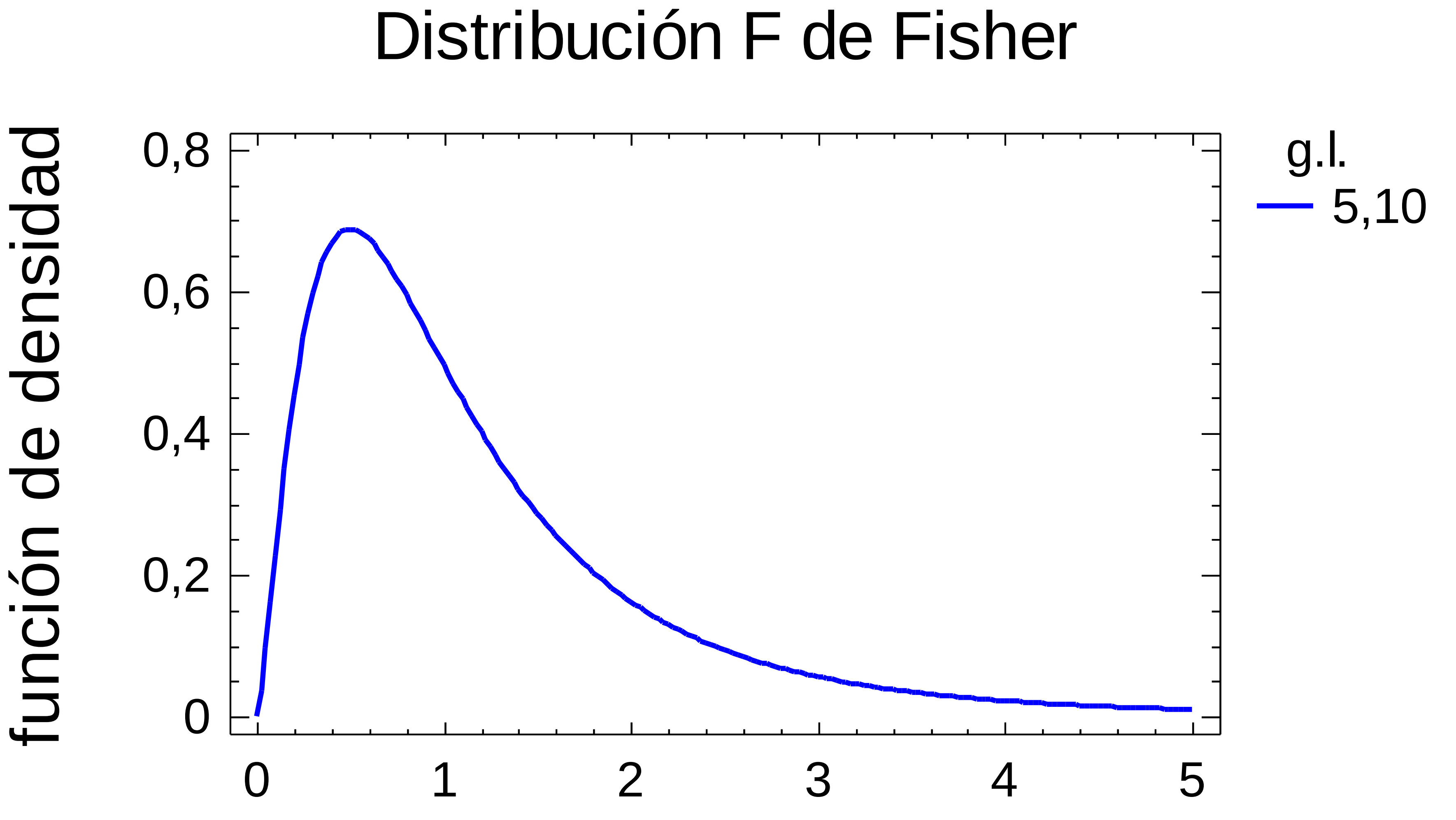
Distribución F de Fisher

- Desempeña un papel fundamental en los modelos de Regresión Lineal y de Análisis de la Varianza.

<u>Definición:</u>

- Siendo X_1 y X_2 dos variables Chi-cuadrado independientes con μ_1 y μ_2 grados de libertad, se dice que una variable F sigue una distribución F con μ_1 y μ_2 grados de libertad si:

$$F = \frac{X_1 / \mu_1}{X_2 / \mu_2}$$



Distribución F de Fisher
función de densidad
0,8
0,6
0,4
0,2
0
0
1
2
3
4
5
g.l.
5,10

Distribución F de Fisher

- La distribución F es marcadamente asimétrica. El coeficiente de asimetría decrece ligeramente cuando aumentan los grados de libertad del numerador y denominador.

- La media es cercana 1 ($E(F)=\mu_2/(\mu_2-2)$ con $\mu_2>2$)

Ejercicio:

Justificar intuitivamente el que la media de una distribución F sea cercana a 1.

Distribución F de Fisher

REPUESTA:

Como $E(X^2_\mu)=\mu$*, las dos expresiones en el numerador y en el denominador de la expresión que define la F tienen media igual a 1.*

Una F resulta, por tanto, igual al cociente de dos variables independientes, ambas con la misma media, por lo que es intuitivo que su valor medio sea cercano a 1. (decimos sólo “cercano” porque la media de un cociente no coincide exactamente con el cociente de las medias).

Distribución F de Fisher

- La tabla de la distribución F de 1 cola tiene tres entradas : una para μ_1 (grados de libertad del numerador, columnas), otra para μ_2 (grados de libertad del denominador filas) y otra para 2 valores distintos valores de la probabilidad de mayor que.

Ejercicio:

Con ayuda de la tabla, encuentra dos valores entre los que se encuentre la $F_{7,10}$ con probabilidad del 95% y del 99%.

Solución:

$P(\ 0{,}21 < F_{7,10} < 3{,}95\ ) = 0{,}95$ *ya que* $P(F_{7,10} > 0{,}21) = 0{,}975$ *y* $P(P(F_{7,10} > 3{,}95) = 0{,}025$

$P(\ 0{,}12 < F_{7,10} < 6{,}3\ ) = 0{,}99$ *ya que* $P(F_{7,10} > 0{,}12) = 0{,}995$ *y* $P(P(F_{7,10} > 6{,}3) = 0{,}005$

Distribución F de Fisher

Ejercicio:

¿Calcular un valor f tal que la probabilidad de que una variable F con 4 y 8 grados de libertad sea mayor que f sea igual a 5%?

SOLUCIÓN:

$P(F_{4,8} > f) = 0{,}05 \Rightarrow f = 3{,}84$

Distribución F de Fisher

- La distribución F se utiliza en la práctica para **comparar la variabilidad debida a diferentes fuentes**.
- Si s_1^2 es la varianza en una muestra de tamaño N_1 extraída de una población normal de varianza σ_1^2 y s_2^2 es la varianza de una muestra de tamaño N_2 extraída de una población normal de varianza σ_2^2, y ambas muestras son independientes, el cociente:

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \qquad (4)$$

Se distribuye como una variable F con (N_1-1) y (N_2-1) grados de libertad.

Distribución F de Fisher

Ejercicio:

Demostrar el resultado anterior, a partir de la expresión (2) de un apartado anterior.

SOLUCIÓN: *La expresión (2) dice que*

$$(N_1 - 1)\frac{s_1^2}{\sigma_1^2}\ es\ Gi-dos\ con\ (N_1 - 1)\ g.l.$$

$$(N_2 - 1)\frac{s_2^2}{\sigma_2^2}\ es\ Gi-dos\ con\ (N_2 - 1)\ g.l$$

Distribución F de Fisher

Por tanto:

$$\frac{(N_1-1)\dfrac{s_1^2}{\sigma_1^2}\Big/(N_1-1)}{(N_2-1)\dfrac{s_2^2}{\sigma_2^2}\Big/(N_2-1)}=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}\,\textit{será una}\,F_{N_1-1,N_2-1}$$

Distribución F de Fisher

- En particular si $\sigma_1^2=\sigma_2^2$ el cociente s_1^2/s_2^2 sigue una distribución F.

Ejercicio:

Calcula aproximadamente la probabilidad de que al extraer dos muestras de tamaño 25 de una misma población normal, la segunda varianza muestral resulte más del doble que la primera.

Distribución F de Fisher

***SOLUCIÓN*:**

$$P(s_2^2 > 2s_1^2) = P(\frac{s_2^2}{s_1^2} > 2) = P(F_{24,24} > 2) \approx 0{,}05$$

Ejercicios de evaluación

1- De una población normal de desviación típica 12 se extraen muestras de tamaño N. ¿Cuánto tiene que valer como mínimo N para que sea inferior al 3% la probabilidad de que la diferencia en valor absoluto entre la media muestral y la poblacional sea superior a 2?

Ejercicios de evaluación

2- En una población normal la desviación típica vale 10 y se extraen muestras de tamaño 25. ¿Qué porcentaje de las muestras tendrá una varianza menor que 151,7?

3- La temperatura que soporta un tipo determinado de equipo informático sigue una distribución normal. Si se extraen dos muestras de tamaños $N_1=9$ y $N_2=16$, calcula la probabilidad de que la primera varianza muestral resulte más de 4 veces el valor de la segunda varianza muestral.