# Tema 9. Introducción a la Inferencia Estadística

## Presentación y Objetivos.

La inferencia utiliza el lenguaje de la probabilidad para sacar conclusiones de los datos y acompañar esas conclusiones por una declaración formal de la confianza que tenemos de que sean correctas. Así, comenzamos ubicando la Inferencia dentro del ciclo general de la Estadística. La Estadística Descriptiva y el Cálculo de Probabilidades, ya estudiados, nos servirán en nuestro objetivo de construir métodos que nos permitan realizar inferencias inductivas de la población partiendo de la muestra. Tales inferencias se formularán sujetas a un grado de confianza que podremos controlar. La primera etapa del Ciclo Estadístico es la selección de la muestra de la población de interés. El éxito del análisis final que se realice dependerá en gran medida del cuidado que se haya puesto en la selección de la muestra y en lo representativa que sea ésta de la población. La herramienta de inferencia que usaremos será la muestra aleatoria simple. Es esencial entender la distribución muestral para comprender los conceptos de inferencia. El estudio de las propiedades de la media muestral y su comportamiento asintótico nos lleva a la desigualdad de Tchebychey y al Teorema Central del Límite, resultado fundamental para el desarrollo de unidades posteriores. Por último se introducirán las distribuciones relacionadas con la distribución normal. Los Objetivos de esta Unidad Didáctica son:

- Entender cuáles son los objetivos y procedimientos de la Inferencia Estadística.
- Comprender la muestra aleatoria simple como variable aleatoria.
- Entender que el estadístico es una variable aleatoria y asimilar que surge de la transformación de la muestra aleatoria simple.
- Entender el concepto de distribución en el muestreo.
- Manejar la media muestral como variable aleatoria y asimilar la idea de aproximación hacia la media poblacional desde diferentes puntos de vista.

## Esquema Inicial

1. Introducción.
2. Muestreo.
3. Muestra aleatoria simple.
4. Media muestral. Propiedades.
5. Distribución asintótica de la media muestral.
6. Distribuciones asociadas a la Normal.

# Desarrollo del Tema

## 1. Introducción

La Figura 1 representa el Ciclo de la Estadística. Estaremos interesados en estudiar una característica determinada en todos los individuos de una Población. Ya que el estudio de todos y cada uno de sus elementos es inviable, seleccionamos una muestra de la misma. A través de los estadísticos descriptivos resumimos de manera concisa mucha de la información contenida en la muestra. Con esta información construimos un modelo matemático que refleje el comportamiento de la población. Este modelo, una vez validado, nos permite hacer suposiciones y predicciones sobre el conjunto de la población. Estas predicciones estarán sometidas a un error que el analista siempre podrá controlar. Por lo tanto, la **Inferencia Estadística** permite generalizar la información contenida en una muestra a la población de la que se extrajo, controlando el error que cometemos con tal generalización.

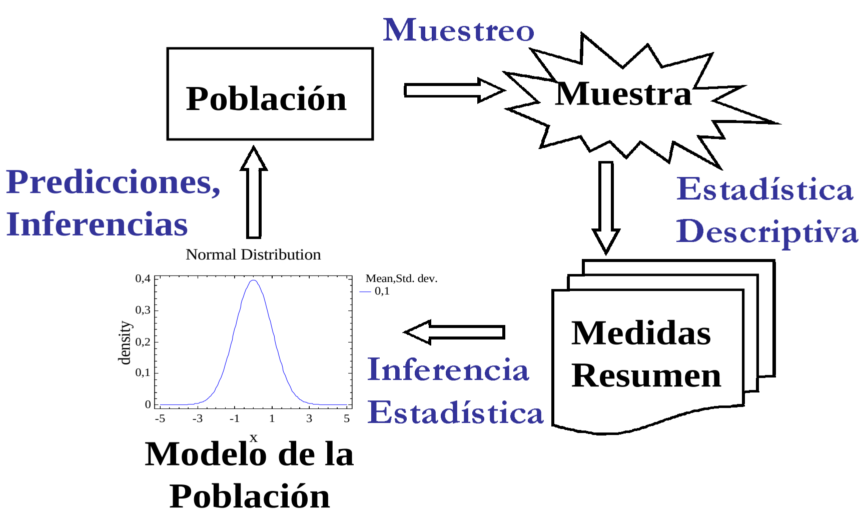


Figura 1: Ciclo de la Estadística

Los métodos de inferencia se clasifican atendiendo a diferentes criterios:

1. Según la información utilizada
   a) Métodos Clásicos
   b) Métodos Bayesianos
2. Según el grado de conocimiento del Modelo para la Población:
   a) Métodos paramétricos
   b) Métodos no paramétricos.

### 1.1. Métodos clásicos

Solamente utilizan la información contenida en la muestra (objetiva). Además, los parámetros son fijos (constantes) y desconocidos y la única información de ellos es la que proporcionan los datos (la muestra).

### 1.2. Métodos Bayesianos

Utilizan, además, fuentes de información subjetiva: conocimiento de especialistas, experimentos realizados anteriormente bajo las mismas o distintas condiciones, etc. Los parámetros se consideran variables aleatorias y esto permite introducir información de ellos a partir de una distribución a priori (información subjetiva).

### 1.3. Métodos paramétricos

Se supone que los datos provienen de un modelo para la población con distribución $P_X$ parcialmente conocida. Se sabe que es de una determinada forma pero sus parámetros o alguno de ellos son desconocidos y es lo que se intenta determinar. Posteriormente, el modelo elegido se somete a cierta crítica.

### 1.4. Métodos no paramétricos

Consideran condiciones muy generales respecto a la distribución $P_X$ y tratan de estimar su forma y contrastar su estructura. No hace hipótesis de qué distribución es. Pueden decir de ella que es simétrica, continua, discreta, nada,... Se utilizan para juzgar hipótesis hechas en los métodos paramétricos y ver así que no son contradictorias con la muestra.

## 2. Muestreo

Los conceptos básicos en este apartado son: población y muestra. El estudio de la población se realiza a través de muestras. El Muestreo es el procedimiento mediante el que se selecciona una muestra de una Población.

Se llama **POBLACIÓN** al conjunto de elementos de los que se va a estudiar una característica $X$. Normalmente no podremos utilizar toda la población, por ejemplo si:

- El estudio es destructivo, estudiar una característica implica la destrucción del objeto (vida media en bombillas, resistencias, etc.).
- Los elementos existen en concepto pero no en la realidad: poblaciones de piezas defectuosas que producirá una máquina.
- Es inviable económicamente el estudio de la población.

- La población se considera constituida por un número infinito de posibles resultados de la característica: por ejemplo, cuando la característica es una medición física, como el nivel de concentración de un contaminante, demanda de un producto, tiempo de espera en una unidad de servicio... Estudiar toda la población no solo llevaría mucho tiempo sino que incluso las propiedades de la población podrían haber cambiado con el mismo.

En estos casos seleccionaremos un conjunto representativo de elementos de la población al que llamaremos **MUESTRA**, en lugar de hacer un **CENSO**, que sería un estudio exhaustivo de todos sus elementos. La muestra debe reflejar la composición y características de la población de partida. Si la muestra está bien escogida será posible inferir características de la población a partir de los datos.

Es importante que la muestra escogida sea representativa de la población. Por ejemplo, sabemos que la altura media de los hombres es mayor que la de las mujeres. Por tanto si en una muestra de 500 estudiantes hay 400 hombres y 100 mujeres existirá un sesgo de selección. Para conseguir que la muestra garantice la representatividad de la población se pueden utilizar diversos procedimientos de muestreo. Detallaremos el muestreo aleatorio simple que es el que usaremos en el desarrollo de los próximos temas.

### Muestreo Aleatorio Simple

Este tipo de muestreo se utiliza cuando todos los elementos de la población son homogéneos respecto de la característica a estudiar, todos los elementos son indistinguibles desde el punto de vista de esta característica. Tiene las siguientes propiedades:

1. Cada elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra.

2. Las observaciones se realizan con reemplazamiento, de forma que la composición de la población es idéntica en todas las extracciones.

En adelante se considerará el **muestreo aleatorio simple en una población infinita**, por lo que se trabajará con una muestra aleatoria simple $X_1, \ldots, X_n$.

## 3. Muestra Aleatoria Simple

Se parte de una variable aleatoria $X$ que representará la característica que deseamos estudiar en una población. Por ejemplo, puede ser el tiempo de procesamiento, número de errores en compilación, tiempo de ejecución de un algoritmo, porcentaje de memoria utilizado, etc. Si $X$ es variable aleatoria discreta tendrá asociada una función de probabilidad $P(X = k)$ y si $X$ es variable aleatoria continua tendrá asociada una función de densidad $f(x)$.

Se considera una muestra aleatoria simple (m.a.s.) $X_1, \ldots, X_n$ de la variable aleatoria $X$, donde $X_i$ representa la v.a. $X$ en el sujeto o elemento $i$-ésimo de la muestra. La m.a.s.

$X_1, \ldots, X_n$ es la herramienta básica de la Inferencia Estadística y representa los distintos valores que pueden tomar todos los subconjuntos posibles de $n$ elementos de la población.

Formalmente, una **muestra aleatoria simple** de tamaño $n$ de una variable aleatoria $X$ de media $\mu$ y varianza $\sigma^2$, es una colección de variables aleatorias $X_1, \ldots, X_n$ de forma que:

- $X_1, \ldots, X_n$ son independientes.
- Cada $X_i$ tiene la misma distribución que la variable aleatoria $X$.

Por lo tanto, es un conjunto de $n$ variables aleatorias independientes e idénticamente distribuidas (de ahora en adelante i.i.d).

La **distribución conjunta** de esa m.a.s., dada la independencia de las variables será:

1. Si $X$ es una v.a. discreta entonces la función de probabilidad conjunta de la muestra es igual al producto de las funciones de probabilidad individuales:

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n) &= P(X_1 = x_1) \ldots P(X_n = x_n) = \\ &= \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} P(X = x_i) \end{aligned}$$

2. Si $X$ es v.a. continua, con función de densidad $f(x)$.

$$\begin{aligned} f(x_1, \ldots, x_n) &= f_1(x_1) f_2(x_2) \ldots f_n(x_n) = \\ &= \prod_{i=1}^{n} f(x_i) \end{aligned}$$

## Ejemplos

1. Calcular la distribución conjunta de una m.a.s. $X_1, \ldots, X_n$ de una variable aleatoria $X \sim P(\lambda)$.
2. Si $X \sim P(2)$, calcular la probabilidad de la muestra de tamaño 5, $(3, 1, 0, 2, 0)$.
3. Calcular la distribución conjunta de una m.a.s. $X_1, \ldots, X_n$ de una variable aleatoria $X \sim N(\mu, \sigma)$.

Un **Estadístico** es una función exclusivamente de la muestra, $T(X_1, \ldots, X_n)$. El valor de esta función cambiará muestra a muestra por lo que también será una variable aleatoria, con su correspondiente distribución, que llamaremos distribución en el muestreo del estadístico. Por lo tanto, **la distribución en el muestreo de un estadístico** $T$ es la distribución de probabilidad de $T$ que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño $n$, de la población de interés.

**Ejemplos:** Estadísticos más usuales.

# 4. Media Muestral. Propiedades

Supongamos que las variables aleatorias $X_1, \ldots, X_n$ constituyen una m.a.s. de una variable aleatoria $X$ con media $\mu$ y varianza $\sigma^2$. Se define la **MEDIA MUESTRAL** de $X_1, \ldots, X_n$ como la variable aleatoria (porque cambia de muestra a muestra), $\bar{X} = \dfrac{X_1 + \ldots + X_n}{n}$, i.e., es la media aritmética de los valores de la muestra. Su **ESPERANZA** y **VARIANZA** son:

$$\begin{aligned} E(\bar{X}) &= \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{n\mu}{n} = \mu \\ V(\bar{X}) &= \frac{\sum_{i=1}^{n} V(X_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

La varianza de la media disminuye a medida que $n$ crece[1].

<u>Observación:</u> La media de $\bar{X}$ es igual a la media de la distribución de la que se seleccionó la m.a.s., pero la varianza es $\frac{1}{n}$ la varianza de $X$. Así, la probabilidad de que $\bar{X}$ esté cerca de $\mu$ es mayor de que lo esté $X_i$. Precisemos esto más utilizando la desigualdad de Tchebychev:

$$\begin{aligned} P(|X - E(X)| \leq k\sigma) &\geq 1 - \frac{1}{k^2} \\ P(|X - E(X)| > k\sigma) &\leq \frac{1}{k^2} \end{aligned}$$

Si la aplicamos a $\bar{X}$, con $E(\bar{X}) = \mu$ y desviación típica $\frac{\sigma}{\sqrt{n}}$:

$$P(|\bar{X} - \mu| < k\frac{\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{k^2} \qquad \text{o bien,}$$

$$P(|\bar{X} - \mu| < k) \geq 1 - \frac{\sigma^2}{nk^2}$$

## Ejemplos

4. Supongamos que queremos seleccionar una muestra de una v.a. cuya media es desconocida y de la que sabemos que $\sigma = 2.0$. Queremos determinar el tamaño muestral para que la diferencia entre $\bar{X}$ y $\mu$ en valor absoluto sea menor que 1 con probabilidad de al menos 0.99.

$$P(|\bar{X} - \mu| \leq \underbrace{1}_{k}) \geq 1 - \frac{\sigma^2}{nk^2} = 1 - \frac{4}{n} \geq 0.99$$

$$\frac{4}{n} \leq 0.01 \Rightarrow n \geq 400$$

5. Seleccionamos una m.a.s. de tamaño $n = 25$ de una población con $\sigma = 2.4$. Calcular la probabilidad de que la diferencia entre la media muestral $\bar{X}$ y la media poblacional $\mu$ sea menor que 1.2.

[1] En general, $V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X,Y)$

# 5. Distribución asintótica de la media muestral. T.C.L.

Veremos que siempre que seleccionemos una m.a.s. de tamaño $n$ de cualquier distribución con media $\mu$ y varianza $\sigma^2$, la media muestral $\bar{X}_n$ tendrá una distribución aproximadamente Normal, $N(\mu, \frac{\sigma}{\sqrt{n}})$, cuando $n$ sea grande.

***Teorema Central del Límite (Lindeberg-Lévy)*** *Dadas $X_1, \ldots, X_n$ v.a.i.i.d con media $\mu$ y varianza $\sigma^2 < \infty$, entonces*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \stackrel{n\to\infty}{\longrightarrow} Z \sim N(0,1)$$

**Teorema de De Moivre:** Siendo $X_1, \ldots, X_n$ v.a.i.i.d. con distribución Bernoulli$(p)$, entonces,

$$\sum_{i=1}^{n} X_i \stackrel{n\to\infty}{\longrightarrow} Z \sim N(np, \sqrt{npq})$$

Así, aproximamos la binomial, que es suma de variables de Bernoulli, por una $N(np, \sqrt{npq})$, cuando $n$ sea grande.

Por lo tanto, la media muestral de un número suficientemente grande de datos es una variable aleatoria simétrica, concentrada alrededor de la media poblacional $\mu$, independientemente de la distribución de partida de $X$. En la práctica, realizaremos la aproximación descrita por el Teorema Central de Límite cuando $n \geq 30$.

## Ejemplos

6. Sabemos que la duración de un determinado componente eléctrico es una variable aleatoria con distribución no especificada, de la que lo único que sabemos es que $\sigma = 2$ horas. Calcular la probabilidad de que la media muestral se encuentre a no más de media hora del valor medio de la población, si tomamos una muestra de la duración de 35 componentes.

$$X = \text{duración del componente eléctrico}, \quad X_1, \ldots, X_{35} \text{ m.a.s.}$$

$$\text{Media poblacional } \mu \text{ desconocida, y} \quad \sigma = 2$$

Como $n = 35$, podemos utilizar la aproximación del Teorema Central del Límite.

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \equiv N(\mu, \frac{2}{\sqrt{35}}) \equiv N(\mu, 0.338)$$

Lo que nos piden es:

$$\begin{aligned}
P(|\bar{X}-\mu|<0.5) &= P(-0.5<\bar{X}-\mu<0.5) = \\
&= P\left(-\frac{0.5}{0.338}<\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<\frac{0.5}{0.338}\right) = \\
&= P(-1.479<Z<1.479) = \\
&= P(Z<1.479)-P(Z<-1.479) = \\
&= P(Z<1.479)-(1-P(Z<1.479)) = \\
&= 2P(Z<1.479)-1 = \\
&= 0.8584
\end{aligned}$$

Si utilizáramos Tchebychev:

$$P(|\bar{X}-\mu|<0.5)\geq 1-\frac{\sigma^2}{nk^2}=1-\frac{4}{35(0.5)^2}=0.5428$$

7. En un sistema con capacidad automática de recuperación de errores la probabilidad de una recuperación correcta es $p=0.4$. Hemos observado $n=200$ errores. Queremos saber, por ejemplo, cuál es la probabilidad de que el número de errores salvados correctamente sea menor que 100.

   Si hacemos $X_i=1$ si se solucionó el error, lo que sucede con probabilidad $c$, sabemos que, el número de errores solucionados converge, cuando $n$ es grande a una Normal. Del TCL sabemos que la distribución de la suma $\sum_{i=1}^{200} X_i$ será aproximadamente una distribución normal con media $np$ y y varianza $npq$, siendo $p$ la proporción de éxitos, que según la consideración de los expertos es 0.4. Así,

$$\begin{aligned}
P\left(\sum_{i=1}^{n} X_i<100\right) &= P\left(\sum_{i=1}^{n} X_i\leq 99\right)=P\left(\frac{\sum X_i-np}{\sqrt{npq}}\leq\frac{99.5-np}{\sqrt{npq}}\right)= \\
&= P\left(Z\leq\frac{99.5-80}{\sqrt{200\times 0.4\times 0.6}}\right)=P(Z\leq 2.81)=0.9975
\end{aligned}$$

8. Supongamos que el número de barriles de petróleo que produce un pozo diariamente es una v.a. con distribución no especificada. Si se observa la producción en 64 días, seleccionados de forma aleatoria, y si se sabe que la desviación típica del número de barriles producidos por día es $\sigma=16$, determinar la probabilidad de que la media muestral se encuentre a no más de 4 barriles del valor medio de la población.

# 6. Distribuciones asociadas a la Normal

## 6.1. Distribución $\chi^2$ de Pearson

<u>Definición:</u> Dadas $Z_1,\ldots,Z_n$ v.a.i.i.d $\sim N(0,1)$, definimos la variable aleatoria

$$X = Z_1^2 + \ldots + Z_n^2 = \sum_{i=1}^{n} Z_i^2$$

$X$ es una v.a. que depende de $n$ que es el número de sumandos y toma valores positivos. Se dice que esta v.a. sigue una distribución $\chi_n^2$ con $n$ grados de libertad.
Es un caso particular de la distribución gamma, $\gamma(\lambda = a = \frac{1}{2}, p = \frac{n}{2})$.
La figura 2 muestra la representación gráfica de esta distribución para distintos grados de libertad.

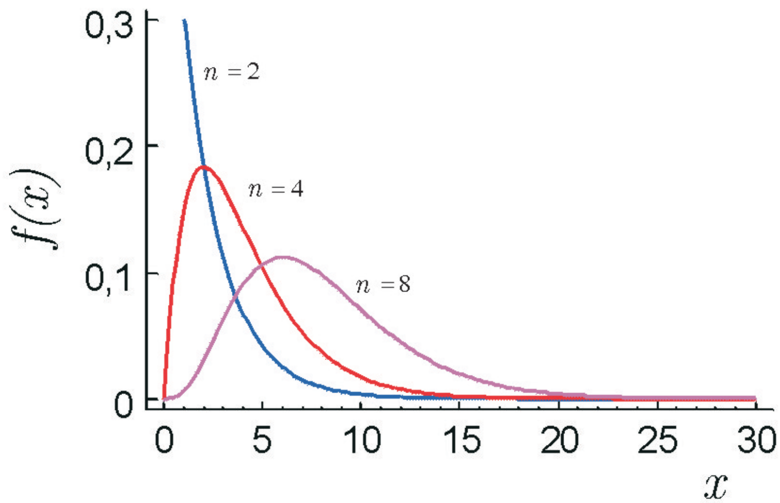

Figura 2: Distribución $\chi^2$ de Pearson

**Observación:**

La distribución $\chi_n^2$ es reproductiva respecto de $n$, es decir, dadas $X$, $Y$ variables aleatorias independientes con $X \sim \chi_{n_1}^2$, $Y \sim \chi_{n_2}^2$ entonces $X + Y \sim \chi_{n_1+n_2}^2$.

### 6.1.1. Medidas características

- Media
$$E(X) = n$$

- Varianza
$$V(X) = 2n$$

## 6.2. Distribución t de Student

Descubierta en 1908 por William Sealey Gosset, que la publicó bajo el pseudónimo de Student, cuando éste trabajaba en la Factoría Guiness.

Definición: Es la distribución de la siguiente v.a.:

$$T = \frac{Z}{\sqrt{\frac{X}{n}}}$$

donde $Z \sim N(0,1)$ y $X \sim \chi^2_n$, ambas INDEPENDIENTES. Se dice entonces que $T$ tiene una distribución t de Student con $n$ grados de libertad (que son los mismos que los de la $\chi^2$ que interviene en su definición).
La figura 3 muestra la representación gráfica de esta distribución para distintos valores del parámetro.

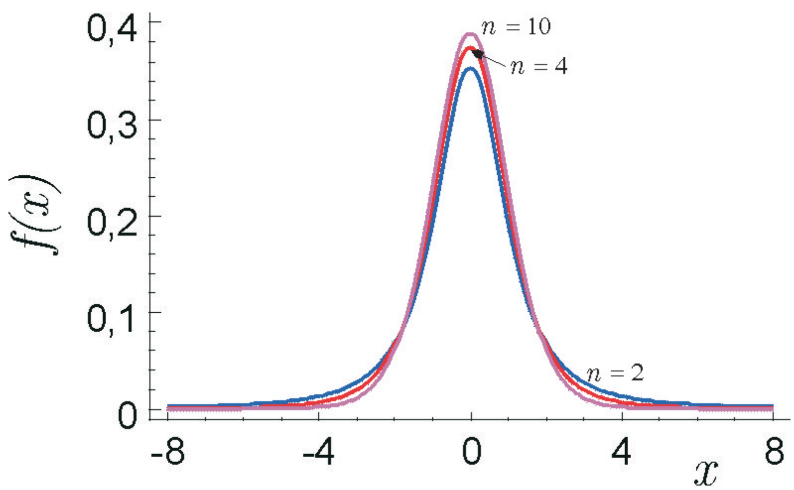


Figura 3: Distribución $t$ de Student

#### 6.2.1. Medidas características

- Media

$$E(T) = 0$$

- Varianza

$$V(T) = \frac{n}{n-2} \quad \text{si } n > 2$$

## 6.3. Distribución F de Fisher-Snedecor

Definición: Si $X$ e $Y$ son dos v.a. independientes, $X \sim \chi^2_n$ e $Y \sim \chi^2_m$, entonces la v.a.

$$F = \frac{\frac{X}{n}}{\frac{Y}{m}} = \frac{\frac{X_1^2+\ldots+X_n^2}{n}}{\frac{Y_1^2+\ldots+Y_m^2}{m}}$$

tiene una distribución F de Snedecor con $n$ y $m$ grados de libertad. Las $X_i \sim N(0,1)$ son independientes y las $Y_i \sim N(0,1)$ también son independientes.
La figura 4 muestra la representación gráfica de esta distribución para distintos valores de los parámetros.

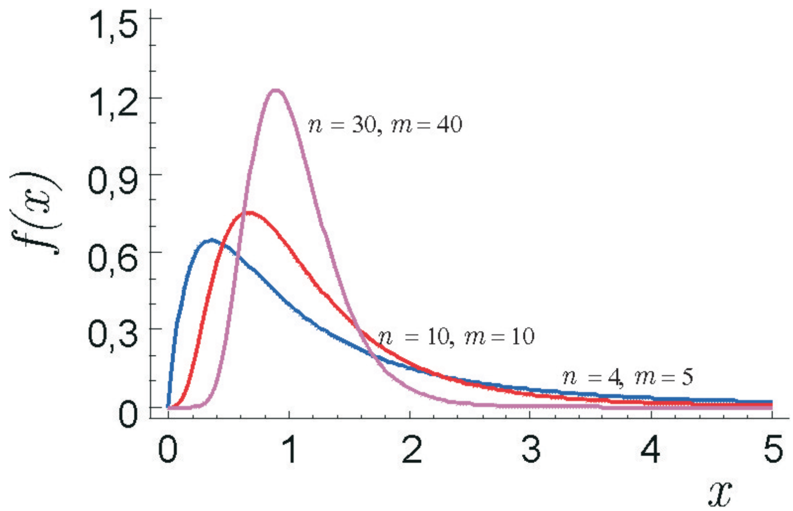


Figura 4: Distribución $F$ de Fisher-Snedecor

### 6.3.1. Medidas características

- Media
$$E(F) = \frac{m}{m-2} \qquad \text{si } m > 2$$

- Varianza
$$V(F) = \frac{m^2(2m+2n-4)}{n(m-2)^2(m-4)^2} \qquad \text{si } m > 4$$

**IMPORTANTE:** A partir de ahora los puntos $\mathbf{t_{n,\alpha}}; \chi^2_{\mathbf{n},\alpha}; \mathbf{F_{n,m,\alpha}}$ representarán, respectivamente, los valores de una distribución T de Student con $n$ grados de libertad, de una $\chi^2$ con $n$ grados de libertad y de una F con $n$ y $m$ grados de libertad, que dejan a la derecha un área o probabilidad de $\alpha$.

Por ejemplo, mirando en la Tabla de la T-Student, para 7 grados de libertad, el punto que deja a la derecha un área o probabilidad de 0.05 es 1.895, con lo que el punto $t_{7,0.05} = 1.895$. Para una distribución $\chi^2_{15}$, el punto que deja a la derecha un área o probabilidad de 0.9 es 8.547, con lo que el punto $\chi^2_{15,0.9} = 8.547$. Para una distribución $F$ con 10 y 7 grados de libertad, el punto que deja a la derecha un área o probabilidad de 0.05 es 3.637, con lo que $F_{10,7,0.05} = 3.637$.