Universidad de Santander
UDES

VIGILADA MINEDUCACIÓN | SNIES 2832

Estadística

Estadísticos y distribuciones muestrales

Daniel Martínez Bello

Universidad de Santander
Maestría en Biotecnología

Octubre de 2025

Muestra al azar

A partir de un número de resultados

$$x_1, \ldots, x_n$$

que se obtienen de repetir un experimento **independientemente** un número de veces. Estos resultados pueden considerarse como valores de variables aleatorias

$$X_1, \ldots, X_n$$

los cuales son independientes y tienen la misma distribución que una variable aleatoria X, o sea

$$X_1, \ldots, X_n \xrightarrow{i.i.d} X$$

y expresamos, "$X_1, \ldots, X_n$" son independientes e identicamente distribuidas como X, o que "$X_1, \ldots, X_n$" son una muestra al azar de X.

Estadístico

La información contenida en los resultados observados $x_1, \ldots, x_n$ se debe condensar en alguna forma de función conocida, que no dependa de los parámetros, como puede ser

$$t(x_1, \ldots, x_n)$$

Si la función t es tal que $t(X_1, \ldots, X_n)$ es una variable aleatoria, entonces

$$T_n = t(X_1, \ldots, X_n)$$

se conoce como un estadístico, donde un estadístico k-dimensional es un vector

$$\underset{\sim}{T}_n = T(T_{n1}, \ldots, T_{nj})$$

donde, para $j = 1, \ldots, k$

$$T_{nj} = t_j(X_1, \ldots, X_n)$$

corresponde a un estadístico de una dimensión.

Muestra aleatoria de una variable aleatoria

Se dice que las variables aleatorias $X_1, X_2, \cdots, X_n$ forman una muestra aleatoria (simple) de tamaño n de una variable aleatoria X si

1. Las X_i son variables aleatorias independientes.
2. Cada X_i tiene la misma distribución de probabilidad.

Distribución Normal

$$X \sim \text{Normal}(\mu, \sigma^2)$$

$$f(x) = \text{Normal}(X|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{1}{2\sigma^2}(X-\mu)^2\right)$$

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

$$moda(X) = \mu$$

Distribución Normal

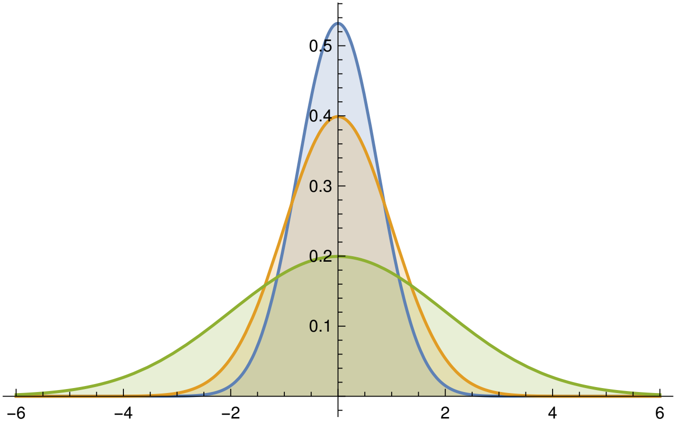
- La situación especial, con $\mu = 0$ y $\sigma = 1$, se le da el nombre de distribución Normal estándar (típica), y este proceso de azar se denota como Z, por tanto $Z \sim \text{Normal}(0; 1)$
- Cualquier variable aleatoria normal $X \sim \text{Normal}(\mu, \sigma^2)$ puede volverse una distribución Normal estándar a través de

$$\frac{X - E(X)}{DE(X)} \sim \text{Normal}(0, 1)$$

- Probabilidades de la forma $P(X \leq x)$ [también denotado como $\Phi(x)$] se encuentran tabuladas. $\Phi(x)$ es la distribución acumulada de una variable aleatoria Normal estándar.

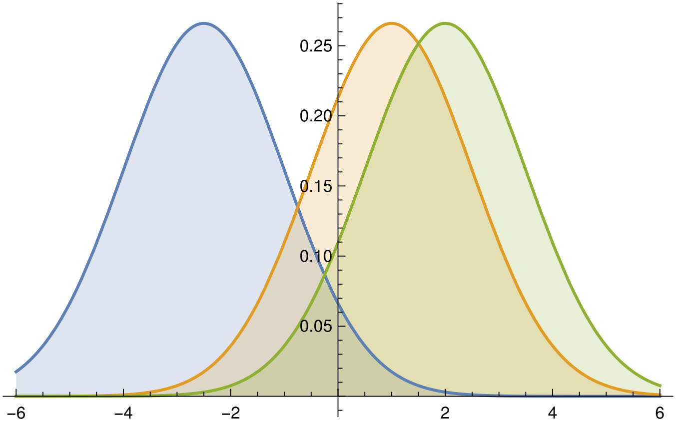
Distribución Normal

Normal[0, 0,75], Normal[0, 1] y Normal[0, 2]



Distribución Normal

Normal$[-2{,}5, 1{,}5]$, Normal$[1, 1{,}5]$ y Normal$[2, 1{,}5]$



Teorema del límite central

Sea $X_1, X_2, \cdots$ una secuencia de variables aleatorias independientes con la misma distribución de una variable aleatoria X.
Suponiendo que $E(X^2) < \infty$, sea $E(X) = \mu$, $Var(X) = \sigma^2$ y sea $S_n = X_1 + \cdots + X_n$. Si $\sigma > 0$ entonces

$$Z_n = \frac{S_n - n\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim \text{Normal}(0, 1)$$

Aproximación Normal a la Distribución Binomial, si X se distribuye Binomial con parámetros n y π

$$\frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$$

Esta aproximación se da cuando $n\pi(1-\pi) \geq 10$

Teorema del límite central

Sea $X_1, X_2, \cdots$ una secuencia de variables aleatorias independientes con la misma distribución de una variable aleatoria X.
Suponiendo que $E(X^2) < \infty$, sea $E(X) = \mu$, $Var(X) = \sigma^2$ y sea $S_n = X_1 + \cdots + X_n$. Si $\sigma > 0$ entonces

$$Z_n = \frac{S_n - n\mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim \text{Normal}(0, 1)$$

Aproximación Normal a la Distribución Binomial, si X se distribuye Binomial con parámetros n y π

$$\frac{X - E(X)}{\sqrt{Var(X)}} = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$$

Esta aproximación se da cuando $n\pi(1-\pi) \geq 10$

La suma de la muestra

Si se tiene como modelo de la población X con $E(X) = \mu$ y $\operatorname{Var}(X) = \sigma^2$,y se extrae una muestra al azar de variables aleatorias $X_1, \ldots, X_n$,independientes y con la misma distribución de X, y se calcula el estadístico $S_n = \sum_{i=1}^{n} X_i$, con las siguientes propiedades

$$E(S_n) = E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\mu$$

$$Var(S_n) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = n\sigma^2$$

$$\text{desviación estándar} = \sqrt{n}\sigma = \text{ error estándar (EE)}$$

En virtud del teorema del límite central

$$\frac{S_n - E(S_n)}{EE(S_n)} \xrightarrow{d} Z \sim \text{ Normal}(0; 1)$$

La media de la muestra

La media muestral como un modelo de azar se define como,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ya que $\bar{X} = \frac{1}{n} S_n$

$$\begin{aligned} E(X) &= \mu \\ \text{Error estándar}(\bar{X}) &= \frac{\sigma}{\sqrt{n}} \\ \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} &\xrightarrow{d} Z \sim \text{Normal}(0;1) \end{aligned}$$

Una realización particular produce el valor $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

La proporción de la muestra

La proporción de la muestra es de hecho una media muestral, cuando la muestra se obtiene de una población 0-1 con $P(X = 1) = \pi$

$$P = \frac{1}{n} \sum_{i=1}^{n} X_i$$

ya que $\bar{X} = \frac{1}{n} S_n$

$$E(P) = \pi$$

$$\text{Error estándar}(P) = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$$

$$\frac{P - \pi}{\frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}} \xrightarrow{d} Z \sim \text{ Normal}(0; 1)$$

Una realización particular produce el valor $p = \frac{1}{n} \sum_{i=1}^{n} x_i$

Resultados importantes asociados a transformación de variables aleatorias distribuidas Normal

- Para $i = 1, \cdots, n$: $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$ y si $\underset{\sim}{X} = (X_1, \cdots, X_n)$ son independientes, entonces

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)$$

- Si $X_1 \sim \text{Normal}(0, 1)$, $X_2 \sim \chi^2(n))$, X_1 y X_2 son independientes, entonces

$$T = \frac{X_1}{\sqrt{\frac{x_2}{n}}} \sim t(n)$$

- Si $X_1 \sim \chi^2(n_1))$, $X_2 \sim \chi^2(n_2))$, X_1 y X_2 son independientes, entonces

$$F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1; n_2)$$

Distribución Chi cuadrado

$$X \sim \chi^2_\nu$$

$$\nu > 0$$

$$f(x) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}$$

$$E(X) = \nu$$

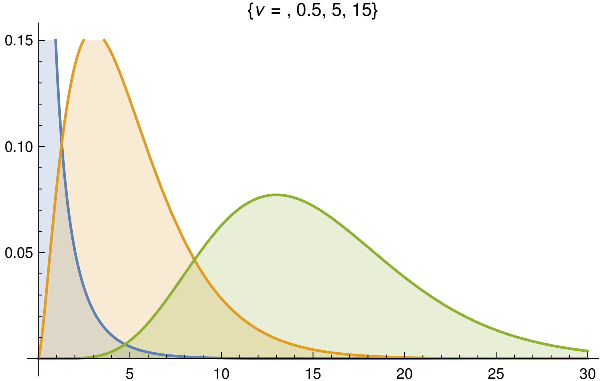
$$Var(X) = 2\nu$$

$$moda(X) = \nu - 2$$

$$Gamma(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$$

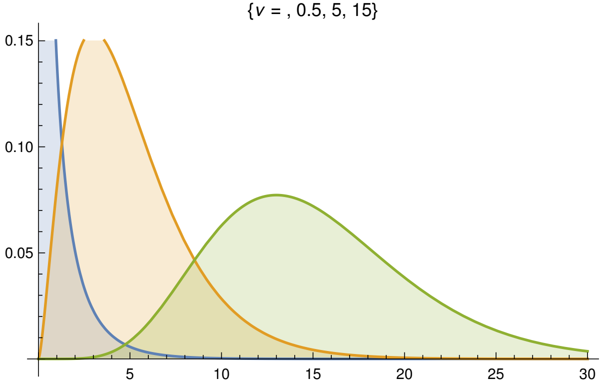
Distribución Chi cuadrado

Función de densidad



Distribución Chi cuadrado

Función de distribución acumulada



Distribución Chi cuadrado

Si para $i = 1, \ldots, k$, $X_i \sim \text{Normal}(\mu_i, \sigma^2)$ y si $X_1, \ldots, X_k$ son variables aleatorias independientes, entonces

$$\sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \sim \chi^2(n)$$

Distribución t-Student

$$X \sim t_\nu(\mu, \sigma^2)$$

$$G.F. \nu > 0, \;\text{posición } \mu, \;\text{escala } \sigma^2$$

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left(1 + \frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}$$

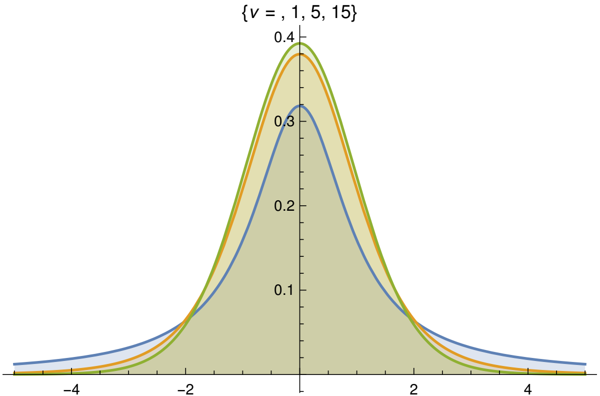
$$E(X) = \mu, \;\text{para } \nu > 1$$

$$Var(X) = \frac{\nu}{\nu - 2}\sigma^2, \;\text{para } \nu > 2$$

$$Moda(X) = \mu$$

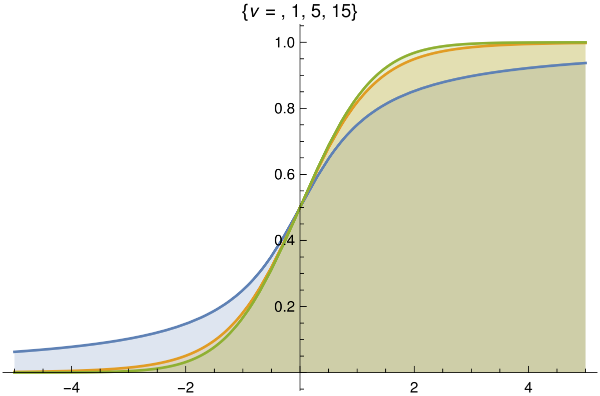
Distribución t de Student

Función de densidad



Distribución t de Student

Función de distribución acumulada



Distribución t-Student

Si $X_1 \sim \text{Normal}(0,1)$, $X_2 \sim \chi^2(n)$, y X_1 y X_2 son independientes, entonces

$$\frac{X_1}{\sqrt{\frac{X_2}{n}}} \sim t(n)$$

Distribución de Probabilidad F

La distribución F modela el comportamiento de una variable aleatoria obtenida como la razón de la varianza de una muestra de una población Normal 1 (s_1^2) con respecto a la varianza de una muestra de una población Normal 2 (s_2^2), ambas poblaciones teniendo la misma varianza σ^2.

$$\frac{s_1^2}{s_2^2} \sim F(\nu_1, \nu_2)$$

La forma de la distribución de densidad F se determina de acuerdo a los grados de libertad de la muestra 1 ν_1 y los grados de libertad de la muestra 2 ν_2.

Distribución de Probabilidad F

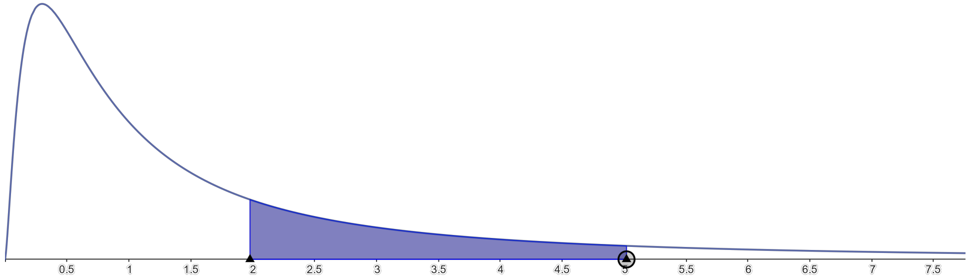
$X \sim F(n_1, n_2)$ denota que la variable aleatoria X tiene distribución F con parámetros n_1 y n_2, los cuales son enteros positivos conocidos como grados de libertad para el número y el denominador. La F distribución se conoce como la *distribución de razon de varianzas* o la *distribución Fisher–Snedecor*. Una variable aleatoria X se distribuye F con n_1 y n_2 grados de libertad con densidad de probabilidad

$$f(x) = \frac{\Gamma((n_1+n_2)/2)(n_1/n_2)^{n_1/2} x^{n_1/2-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[(n_1/n_2)x+1]^{(n_1+n_2)/2}} \qquad x > 0,$$

para $n_1 = 1, 2, \ldots$ y $n_2 = 1, 2, \ldots$. La distribución F distribución se utiliza para inferencia estadística de razones de varianzas de dos poblaciones normales.

Distribución F

$n_1 = 7$ y $n_2 = 2$



Distribución F

$n_1 = 14$ y $n_2 = 10$

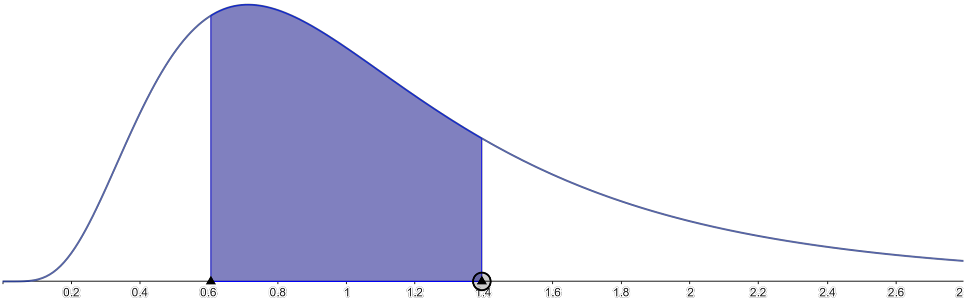
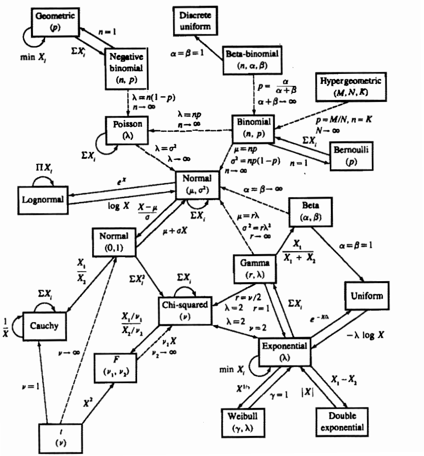


TABLE OF COMMON DISTRIBUTIONS 627



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

Tomado sin permiso, con propósitos académicos de Casella y Berger, Statistical Inference, second edition, 2002