MODULO ANÁLISIS DE DATOS

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
POSTGRADO AUTOFINANCIADO EN MATEMÁTICA
Carrera de Matemáticas
Prof. M.sc. Iván Aliaga C.

11 de noviembre de 2025

Objetivo del curso

Desarrollar en el estudiante competencias avanzadas en el análisis exploratorio, descriptivo e inferencial de datos, que sirvan de fundamento para la construcción, estimación y validación de modelos matemáticos aplicados a problemas científicos, tecnológicos y socioeconómicos.

Referencias

Navidi, W. (2015). Estadística para ingenieros y científicos. McGraw-Hill. Triola, M. F. (2018). Estadística. Pearson, 12ª edición.

Pardo M. Ruiz D., C. (2019). Análisis de Datos con SPSS 13 Base, McGraw Hill.

McKinney, W. (2017). Python for Data Analysis. O'Reilly.

Dennis D. Wackerly, Dennis D Wackerly, William Mendenhall, III, Richard L. Scheaffer (2009) Estadística Matemática Con Aplicaciones, México Editoriales y graficos S.A. 2002Edition: 6a edición

Dorin, Federico; Goldszier, Patricia; Perrotti, Daniel E. (2018) Los números índices y su relación con la economía, CEPAL.

Metodología de aprendizaje

- Clases expositivas.
  - Google meet.
  - Pausas despues de 55 minutos (o cada 1h, o contínuo).
- Evaluaciones
  - Evaluación Continua (35 puntos, 7pts por semana, total 5 semanas). Estas pueden ser elaboradas como:
    - Representación gráfica de conceptos: Mapas mentales.
    - Presentaciones de ejercicios a través de videoconferencias: sincrónico o asincrónico.
    - Evaluaciones y cuestionarios en línea (sencillos).
    - Debates a través de foros (existe una opción en Moodle): esto sirve para enterarse lo necesario a reforzar la siguiente semana.
  - Evaluación parcial:
    - 1er Parcial (15 %).
    - 2do Parcial (15 %).
  - Evaluación final:
    - Trabajo final del módulo (investigación bibliográfica, exposiciones personales o grupales) (15 puntos).
    - Asistencia (20 puntos)

Contenido

1. Análisis de distribuciones
    - Descripción numérica de las distribuciones.
    - Descripciones numéricas: mediana, media, recorrido, cuartiles, desviación típica.
    - Descripción gráfica: simétricas y asimétricas
    - Gráficos de series temporales. - Diagramas de caja.
    - Transformaciones de los datos. - Distribuciones normales.
    - Curva de densidad de probabilidad. Media y mediana en una curva de densidad.
    - Distribución normal, distribución normal estándar.
    - Propiedades de la distribución normal - Cálculos con distribuciones normales.
2. Análisis de relaciones
    - Conjunto de dos variables (dos variables numéricas).
    - Diagramas de dispersión.
    - Correlación lineal de Pearson. - Recta de regresión.
    - Conjunto de dos variables (dos variables categóricas).
    - Tablas de contingencia. - Distribuciones marginales.
    - Diagramas de barras. - Distribuciones condicionales.
3. Estadística económica
    - Números índice, simples y complejos.
    - Índice de precios al consumo. - Medidas de desigualdad y concentración
    - Medidas de desigualdad y curvas de Lorenzo: Cálculo e interpretación.
    - Series temporales: representación gráfica.
    - Componentes de las series temporales, tendencias, fluctuaciones a medio y corto plazo.

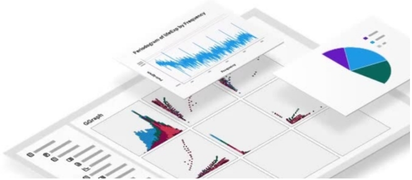
Programas

- SPSS https://www.ibm.com/es-es/products/spss-statistics

Inicio / Productos / SPSS Statistics

IBM® SPSS Statistics

Analice datos y prevea tendencias más rápido ahora con una IA que interpreta resultados.

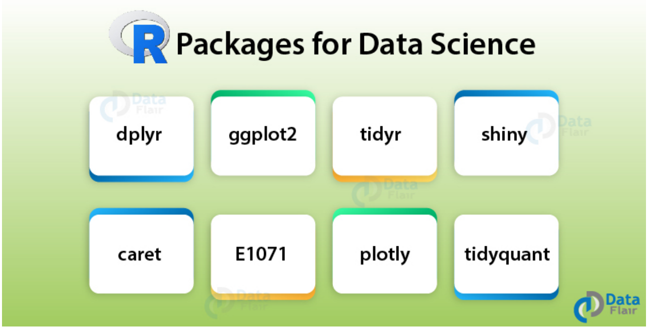


Programas

- Python (versión)
  https://repo.anaconda.com/archive/Anaconda3-2025.06-1-Windows-x86_64.exe

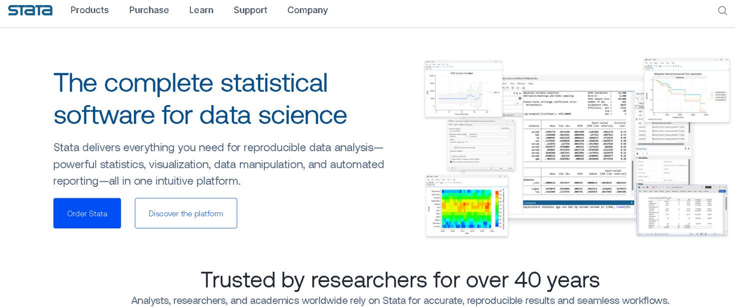
Programas

- R (x86_64-w64-mingw32, x86_64, mingw32, ucrt, x86_64, mingw32, , 4, 4.3, 2025, 02, 28, 87843, R, R version 4.4.3 (2025-02-28 ucrt), Trophy Case), https://cran.r-project.org/bin/windows/base/



Programas

- Stata (Version 15) https://www.stata.com/



Programas

- EViews (Version 14) https://www.eviews.com/home.html

Programas

- Matlab (versión 2019) https://yadi.sk/d/mlDQ8uMFVWVevw
- Matlab (versión 2016) https://bit.ly/3oMwOwv

Aspecto fundamental de análisis de datos

Los datos pueden ser representados mediante una matriz de $n \times d$ con n filas y d columnas.

$$\mathbf{D} = \left(\begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_d \\ \hline \mathbf{x}_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ \mathbf{x}_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array}\right)$$

- las filas son d-tuplas

$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})$$

- las columnas son n-tuplas

$$X_j = (x_{1j}, x_{2j}, \ldots, x_{nj})$$

Atributos

Los Atributos se clasifican en dos principales tipos

- **Atributos numéricos o cuantitativos** : valores en el dominio reales o enteros.
  - *Escala de intervalo*: Solo las diferencias son significativas
    Ejemplo, `temperatura, nota de clases, etc..`
  - *Escala de razón*: Las diferencias de razones son significativas
    Ejemplo `Edad, Salario, Gasto, etc.`
- **Atributos Categóricos o cualitativos**: el dominio es un conjunto compuesto por un conjunto de símbolos o categorias.
  - *Nominales*: Solo la igualdad es significativa
    Ejemplo, dominio(`Sexo`) = { `M, F`}
  - *Ordinales*: Tanto la igualdad como la desigualdad con orden son significativos
    Ejemplo, dominio(`Educación`) = { `primaria`, `secundaria`, `Universidad`, `Magister`}, ... etc,.

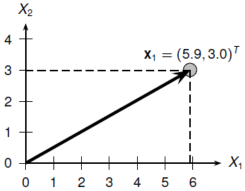
Representación geométrica

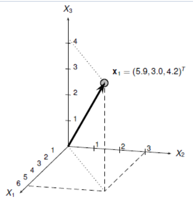
Para una matriz numérica $\boldsymbol{D}$, cada fila es un vector columna $d-$ dimensional.

$$\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix} = \begin{pmatrix} x_{i1} & x_{i2} & \cdots & x_{id} \end{pmatrix}^t \in \mathbb{R}^d$$

Mientras que cada atributo es un vector columna $n-$ dimensional.

$$X_j = \begin{pmatrix} x_{1j} & x_{2j} & \cdots & x_{nj} \end{pmatrix}^t \in \mathbb{R}^n$$





Descripción de Datos

Una vez que el investigador (ingeniero, científico de datos, etc.) está a frente a los datos surge una pregunta que es fundamental:

Qué técnica de análisis debe emplearse?

- ¿Se quiere responder a preguntas descriptivas o inferencia?.
- ¿Número de variables o columnas a utilizar?.
- ¿Número de instancias o filas a utilizar?
- ¿Qué tipo de variables son las que están presentes en la investigación?.

Tablas de Frecuencia

Las tablas de frecuencia son tablas arregladas para un conjunto de datos de una variable cuantitativa. Datos.

Ejemplo

Los siguientes datos representan el número de días de baja por enfermedad de 50 trabajadores de una compañía de los últimas 6 semanas.

2,2,0,0,5,8,3,4,1,0,0,7,1,7,1,5,4,0,4,0,1,8,9,7,0,
1,7,2,5,5,4,3,3,0,0,2,5,1,3,0,1,0,2,4,5,0,5,7,5,1

X: *Número de días de baja por enfermedad*.

Es una variable Cuantitativa discreta medida en una escala de razón.

Tabla de frecuencia

Representación de datos

Valor	Frecuencia Abs.
0	12
1	8
2	5
3	4
4	5
5	8
6	0
7	5
8	2
9	1
total	50

Cuadro 1: Tabla de frecuencia de Días de baja por enfermedad

Gráfico de lineas

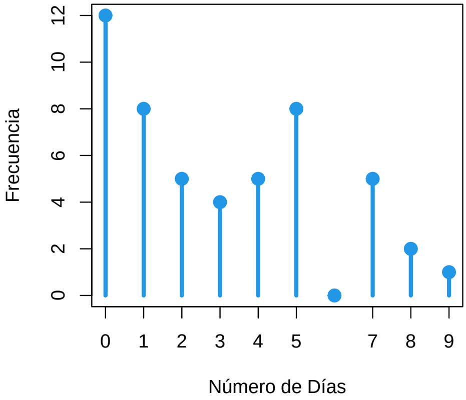


Figura 1: Gráfica de lineas

Gráfico de barras

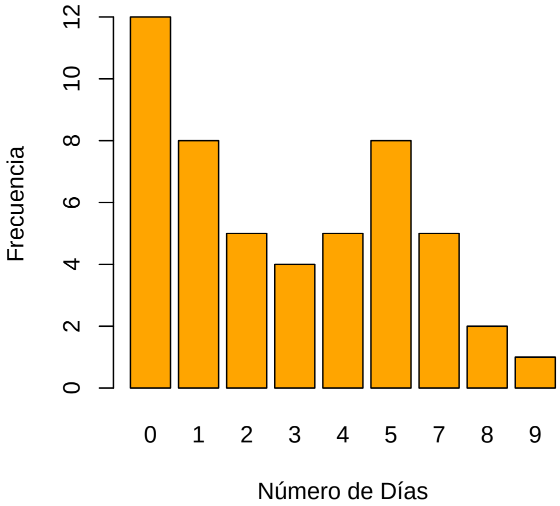


Figura 2: Gráfica de barras

Polígono de frecuencia

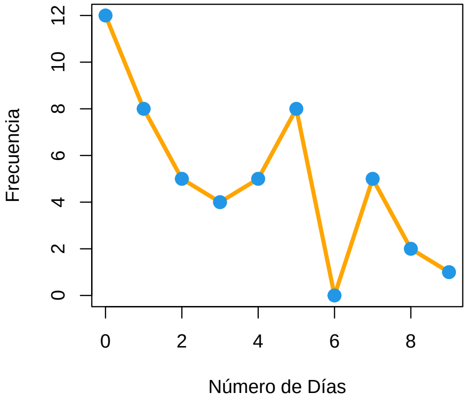


Figura 3: Polígono de frecuencia

Tablas de frecuencia

Construcción de tabla de frecuencia relativa

Valor X	Frecuencia Abs. (n_i)	Frecuencia Rel. (f_i)	F. Acum(N_i)
0	12	$\frac{12}{50} = 0{,}24$	12
1	8	$\frac{8}{50} = 0{,}16$	20
2	5	$\frac{5}{50} = 0{,}10$	25
3	4	$\frac{4}{50} = 0{,}08$	29
4	5	$\frac{5}{50} = 0{,}10$	34
5	8	$\frac{8}{50} = 0{,}16$	42
6	0	$\frac{0}{50} = 0{,}00$	42
7	5	$\frac{5}{50} = 0{,}10$	47
8	2	$\frac{2}{50} = 0{,}04$	49
9	1	$\frac{1}{50} = 0{,}02$	50
total	50	1.00	

Cuadro 2: Tabla de frecuencia de Días de baja por enfermedad

Polígono de frecuencia Relativa

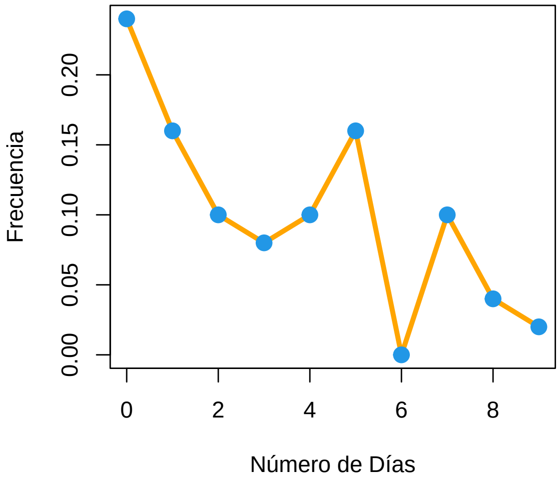


Figura 4: Poligono de frecuencia

Pie Chart

Valor X	f_i	$360 * f_i$ (A^o)
0	0,24	86,4
1	0,16	57,6
2	0,10	36,0
3	0,08	28,8
4	0,10	36
5	0,16	57,6
6	0,00	0
7	0,10	36
8	0,04	14,4
9	0,02	7,2
total	50	1.00

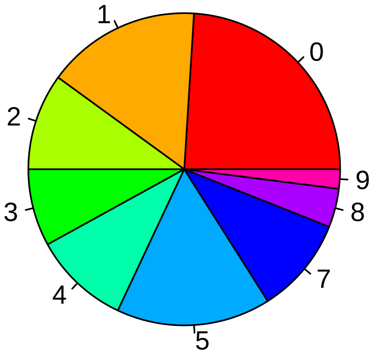


Figura 5: Diagrama Circular

Datos agrupados

Para variables de tipo cuantitativos contínuos y el tamaño de la muestra es mayor a 15, se debe agrupar la información. **Ejemplo** A continuación se presenta los niveles de colesterol (mg) en la sangre de 40 estudiantes de los primeros años

213	174	193	196	220	183	194	200
192	200	200	199	178	183	188	193
187	181	193	205	196	211	202	213
216	206	195	191	171	194	184	191
221	212	221	204	204	191	183	227

Datos ordenados

171	174	178	181	183	183	183	184	187	188
191	191	191	192	193	193	193	194	194	195
196	196	199	200	200	200	202	204	204	205
206	211	212	213	213	216	220	221	221	227

Cálculos estadísticos de la muestra

$$\begin{aligned} Max(X) &= & Max(x_1, x_2, \ldots, x_n) = 227 \\ Min(X) &= & Min(x_1, x_2, \ldots, x_n) = 171 \\ Rango(X) &= & Max(X) - Min(X) = 227 - 171 = 56 \\ n &= & length(X) = Tamaño(X) = 40 \end{aligned}$$

Cuantas Categorías o clases

Redondear al entero superior en cada caso.

1. (Criterio Estandar $n < 100$) $k = \sqrt{n}$, $k = \sqrt{40} = 6{,}3245553 \approx 7$.
2. (Criterio Scott)

$$k = \frac{Rango}{1 + 3{,}322\ log\ n} = \frac{56}{1 + 3{,}322\ log\ 40} = 4{,}224994 \approx 5$$

3. (Criterio Sturges,Herbert A. Sturges 1926)

$$k = log_2(n) + 1 = \frac{log\ n}{log\ 2} + 1 = 6{,}321928 \approx 7$$

Construcción de los intervalos de clase

Dado que se conoce la amplitud, a.

$$a = \frac{Rango}{k}$$

La construcción de los intervalos podría ser de la forma: Construcción:

$$\begin{aligned}
l_0 &= & X_{min} \\
l_1 &= & l_0 + a \\
l_2 &= & l_1 + a = l_0 + 2a \\
l_3 &= & l_2 + a = l_0 + 3a \\
\vdots \\
l_k &= & l_0 + (k)a
\end{aligned}$$

Tabla de frecuencia de los niveles de colesterol en la sangre

Calculando la amplitud

$$a = \frac{Rango}{k} = \frac{56}{7} = 8$$

	Clases	n_i	N_i	f_i
1	[171,179]	3	3	0.07
2	(179,187]	6	9	0.15
3	(187,195]	11	20	0.28
4	(195,203]	7	27	0.17
5	(203,211]	5	32	0.12
6	(211,219]	4	36	0.10
7	(219,227]	4	40	0.10

Tabla de frecuencia de los niveles de colesterol en la sangre

Calculando la amplitud con $k = 6$

$$a = \frac{Rango}{k} = \frac{56}{6} = 9{,}33310 \approx 10$$

	Clases	n_i	N_i	f_i
1	[171,181]	3	3	0.07
2	(181,191]	7	10	0.17
3	(191,201]	16	26	0.40
4	(201,211]	5	31	0.12
5	(211,221]	6	37	0.15
6	(221,231]	3	40	0.07

Histograma

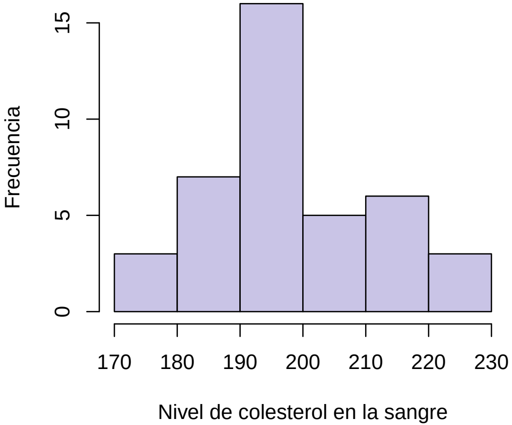


Figura 6: Histograma

Pasos para construir un histograma

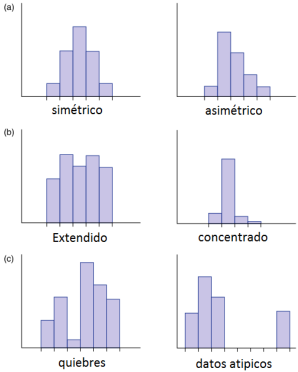
1. Ordenar los datos en orden creciente
2. Elige intervalos de clase de manera que se cubren todos los puntos de datos.
3. Construir una tabla de frecuencia.
4. Dibujar barras adyacentes que tienen alturas determinadas por las frecuencias en el paso anterior.

Histograma

La importancia de un histograma resulta que nos permite organizar y presentar datos gráficamente con el fin de llamar la atención sobre ciertas características importantes, tales como.

1. La simetría de los datos.
2. La asimetría (sesgo) de los datos.
3. En que intervalo (clase) esta concentrado la información.
4. Si algunos datos (intervalo) de los valores están muy separados de los demás

Histogramas de ejemplo



Conjuntos normales, a) Simetría, b) Concentración de datos c) Quiebres y valores atípicos

Diagrama de Tallos y hojas

Una manera de mostrar un conjunto de datos de tamaño pequeño o moderado es utilizar un diagrama de tallo y hojas. que se obtiene dividiendo cada valor de datos en dos partes la primera en los tallos y la segunda en las hojas, el tallo puede contener los primeros dígitos y las hojas contienen los últimos dígitos.

- Los datos que se muestran a continuación corresponde al consumo eléctrico de 42 hogares de cierto estrato socioeconómico.

  Construir el diagrama de Tallos y Hojas correspondiente.

223	213	245	275	213	245
202	239	265	275	230	202
213	213	215	220	199	213
218	245	265	233	213	218
223	245	265	225	210	202
205	250	218	255	230	223
255	230	235	230	255	227

Diagrama de tallos y hojas

Tallo	Hojas
19	9
20	2 2 2 5
21	0 3 3 3 3 3 3 5 8 8 8
22	0 3 3 3 5 7
23	0 0 0 0 3 5 9
24	5 5 5 5
25	0 5 5 5
26	5 5 5
27	5

Estadígrafos y Parámetros

- Estadígrafo Son las medidas descriptivas inherentes a una muestra o cálculos numéricos con respecto a la muestra.
- Parámetro Son las medidas o características descriptivas inherentes a las poblaciones o cálculos numéricos con respecto a la población.

Estadígrafo Media Muestral

Suponer que se tiene n valores o datos designado como:

$$x_1, x_2, \ldots, x_n$$

El estadígrafo media muestral indica el centro promedio de los datos y se designa como:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Que es la formula para **datos no agrupados**.

Media muestral para datos agrupados en k clases

la media muestral para datos agrupados es la siguiente.

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i m_i}{n} = \frac{n_1 m_1 + n_2 m_2 + \ldots + n_k m_k}{n}$$

Donde: $m_i = \frac{l_{i-1} + l_i}{2}$ Representa la marca de clase (m_i)

Propiedades de la media

1. Todo conjunto de datos derivado de una variable cuantitativa tiene una media.
2. $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$
3. Si un conjunto de datos $x_1, x_2, \ldots, x_n$ posee media $\bar{x}$ entonces $x_1 + b, x_2 + b, \ldots, x_n + b$ tiene media $\bar{x} + b$.
4. Si un conjunto de datos $x_1, x_2, \ldots, x_n$ posee media $\bar{x}$ entonces $a * x_1, a * x_2, \ldots, a * x_n$ tiene media $a * \bar{x}$.

Desventajas de la media

1. Es muy sensible a valores extremos de la variable ya que todas las observaciones intervienen en el calculo de la media, la aparición de una observación extrema hará que la media se desplace en esa dirección.
2. No es recomendable usar la media como medida en distribuciones asimétricas
3. Depende de la división en intervalos en el caso de variables continuas.

Varianza y Desviación Muestral

Estos estadígrafos miden miden su dispersión o variabilidad de un conjunto de datos. Por ejemplo, se tienen los siguientes datos de dos muestras donde se tienen el mismo promedio.

$$A : 1, 2, 5, 6, 6, \qquad B : -40, 0, 5, 20, 35$$

$$\bar{A} = 4, \quad \bar{B} = 4$$

Como las dos muestras tienen el mismo promedio no se puede afirmar que ambas muestras tienen la misma representación (promedio). Ambas muestras difieren en su variabilidad

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad \text{,Varianza Muestral}$$

$$s = \sqrt{s^2} \quad \text{Desviación Muestral}$$

Varianza y Desviación Muestral

Así de esta forma se tiene lo siguiente:

$$A : 1, 2, 5, 6, 6, \qquad B : -40, 0, 5, 20, 35$$

$$\bar{x}_A = 4, \quad \bar{x}_B = 4$$

$$s_A^2 = 5{,}5, \quad s_B^2 = 792{,}5$$

$$s_A = \sqrt{5{,}5} = 2{,}35, \quad s_B = \sqrt{792{,}5} = 28{,}15$$

Se afirma que en la muestra B se tiene más variabilidad(dispersión) con respecto al promedio que la muestra A.

Varianza Muestral y Desviación Muestral

Para datos agrupados la varianza se define como:

$$s^2 = \frac{\sum_{i=1}^{k} n_i(x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^{k} n_i(m_i - \bar{x})^2}{n-1}$$

Propiedades de la Varianza

1. Todo conjunto de datos derivado de una variable cuantitativa tiene una media y una varianza.
2. $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$
3. Si un conjunto de datos $x_1, x_2, \ldots, x_n$ posee media $\bar{x}$ y varianza muestral s^2 entonces $x_1 + b, x_2 + b, \ldots, x_n + b$ donde b es una constante cualquiera, entonces el nuevo conjunto de datos tiene varianza s^2.
4. Si un conjunto de datos $x_1, x_2, \ldots, x_n$ posee media $\bar{x}$ y varianza s^2 entonces $a * x_1, a * x_2, \ldots, a * x_n$ donde a es una constante cualquiera, entonces el nuevo conjunto de datos tiene varianza $a^2 * s^2$.

Conjuntos de datos normales

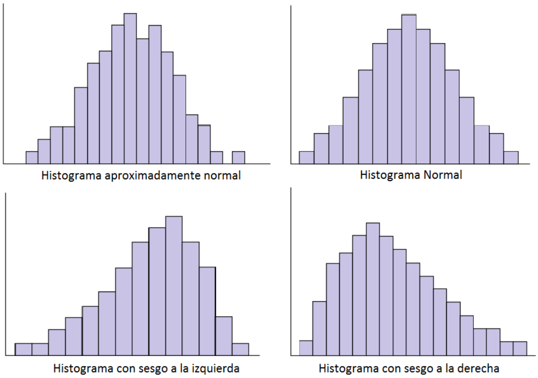
Muchos de los grandes conjuntos de datos que uno encuentra en la práctica tienen los histogramas que son a menudo simétrica alrededor de su punto más alto en forma de campana. Tales conjuntos de datos se dice que son normales, y sus histogramas son llamados histogramas normales.

Definición

Un conjunto de datos se dice que es normal si un histograma que describe que tiene la siguiente propiedades:

- El intervalo medio es el más alto en frecuencia
- Al Pasar del intervalo medio en cualquier dirección, la altura disminuye de tal manera que todo el histograma tiene la forma de campana.
- El histograma es simétrico alrededor de su intervalo medio.

Histogramas normales



Regla Empírica

Regla empírica

Si un conjunto de datos es aproximadamente normal con media muestral $\bar{x}$ y desviación estándar muestral s, a continuación, los siguientes puntos son verdaderas.

- Aproximadamente el 68 por ciento de las observaciones se encuentran dentro de:

$$\bar{x} \pm s$$

- Aproximadamente el 95 por ciento de las observaciones se encuentran dentro de:

$$\bar{x} \pm 2s$$

- Aproximadamente el $99,7$ por ciento de las observaciones se encuentran dentro de:

$$\bar{x} \pm 3s$$

Variable aleatoria

Definición

Dado un experimento aleatorio ϵ y $\boldsymbol{\Omega}$ asociados. Una función que asigna a cada elemento (suceso) $w \in \boldsymbol{\Omega}$, uno y solamente un número real $x = \boldsymbol{X}(w)$, se llama **variable aleatoria**.

$$\boldsymbol{X} : \boldsymbol{\Omega} \; \rightarrow \; \mathbb{R}$$

El dominio de la variable aleatoria $\boldsymbol{X}$ es $\boldsymbol{\Omega}$ y el rango (recorrido o soporte) es el subconjunto de $\mathbb{R}$ y se denota como: $R_{\boldsymbol{X}}$.

$$R_{\boldsymbol{X}} = \{x \in \mathbb{R} / \; \boldsymbol{X}(w) = x, \;\; w \in \boldsymbol{\Omega}\}$$

Notación: las variables aleatorias se denotan con letras mayúsculas: $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X_1}, \boldsymbol{X_2}, \ldots$ y los valores particulares que toma la variable aleatoria se denotan con letras minusculas $x, y, z, x_1, x_2, \ldots$

Variables Aleatorias Discretas

Definición (Variable Aleatoria Discreta)

Si el rango (recorrido o soporte) de la variable aleatoria $\boldsymbol{X}$, es un conjunto finito o infinito numerable, se llama **Variable Aleatoria Discreta**, En este caso.

$$R_{\boldsymbol{X}} = \{x_1, x_2, x_3, \ldots\}$$

Definición (Función de cuantía o ley de probabilidad)

Sea $\boldsymbol{X}$ una variable aleatoria discreta con rango $R_{\boldsymbol{X}}$, la función de cuantía está definida por:

$$p(x) = \mathbb{P}(\boldsymbol{X} = x) = \sum_{\{w \in \boldsymbol{\Omega} / \boldsymbol{X}(w) = x\}} \mathbb{P}(\{w\})$$

Que satisface las siguientes condiciones

1. $p(x) > 0, \ \forall x \in R_{\boldsymbol{X}}$
2. $\displaystyle\sum_{x \in R_{\boldsymbol{X}}} p(x) = \sum_{x \in R_{\boldsymbol{X}}} \mathbb{P}(\boldsymbol{X} = x) = 1$

Ejemplo de variable aleatoria

Se lanza una moneda tres veces, sea $\boldsymbol{X}$ la variable aleatoria que está definida por:

$$\boldsymbol{X}(w) = n_C - n_S$$

Donde: n_C y n_S representa el número de caras y el número de sellos obtenidos respectivamente.

EL dominio de $\boldsymbol{X}$ es el espacio muestral $\boldsymbol{\Omega}$

$$\begin{aligned}
\boldsymbol{\Omega} = \{CCC,\ CCS,\ CSC,\ SCC,\ CSS,\ SCS,\ SSC,\ SSS\} \\
\boldsymbol{X}(CCC) = 3 - 0 = 3 \\
\boldsymbol{X}(CCS) = \boldsymbol{X}(CSC) = \boldsymbol{X}(SCC) = 2 - 1 = 1 \\
\boldsymbol{X}(SSC) = \boldsymbol{X}(SCS) = \boldsymbol{X}(CSS) = 1 - 2 = -1 \\
\boldsymbol{X}(SSS) = 0 - 3 = -3
\end{aligned}$$

Luego:

$$R_{\boldsymbol{X}} = \{-3, -1, 1, 3\}$$

Ejemplo de variable aleatoria discreta

En el ejemplo anterior, la variable aleatoria $\boldsymbol{X}$ está definida por:

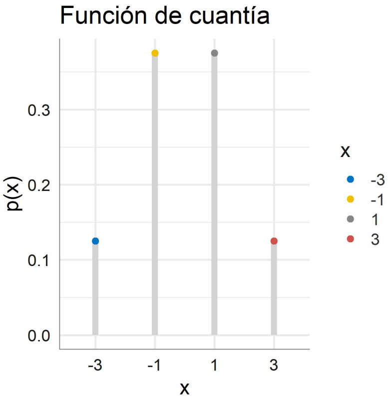
$$\boldsymbol{X}(w) = n_C - n_S$$

Cuya rango o espacio soporte es: $R_{\boldsymbol{X}} = \{-3, -1, 1, 3\}$ La función de cuantía se puede obtener así:

$$\begin{aligned} p(3) &= \mathbb{P}(\boldsymbol{X} = 3) = \mathbb{P}(CCC) = \frac{1}{8} \\ p(1) &= \mathbb{P}(\boldsymbol{X} = 1) = \mathbb{P}(CCS) + \mathbb{P}(CSC) + \mathbb{P}(SCC) = \frac{3}{8} \\ p(-1) &= \mathbb{P}(\boldsymbol{X} = -1) = \mathbb{P}(SSC) + \mathbb{P}(SCS) + \mathbb{P}(CSS) = \frac{3}{8} \\ p(-3) &= \mathbb{P}(\boldsymbol{X} = -3) = \mathbb{P}(SSS) = \frac{1}{8} \end{aligned}$$

Representación gráfica de la función de cuantía

x	-3	-1	1	3
p(x)	1/8	3/8	3/8	1/8



Función de Distribución Acumulada

Definición (Función de distribución)

Sea $\boldsymbol{X}$ una variable aleatoria discreta con rango $R_{\boldsymbol{X}}$ y función de probabilidad, $p(x_i) = \mathbb{P}(\boldsymbol{X} = x_i)$, sea $x \in \mathbb{R}$ un número real cualquiera, la **función de Distribución** de $\boldsymbol{X}$ se denota por $F(x)$ se define como:

$$F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \sum_{x_i \leq x} p(x_i) = \sum_{x_i \leq x} \mathbb{P}(\boldsymbol{X} = x_i)$$

Ejemplo función de distribución

En el ejemplo anterior se tiene:

$$p(-3) = \frac{1}{8}, \quad p(-1) = \frac{3}{8}, \quad p(1) = \frac{3}{8}, \quad p(3) = \frac{1}{8}$$

1. Si $x < -3$, es claro que $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = 0$.
2. Si $x \leq -3$, entonces $F(-3) = \mathbb{P}(\boldsymbol{X} \leq -3) = \sum_{x_i \leq -3} p(x_i) = p(-3) = \frac{1}{8}$.
3. Si $x \in [-3, -1)$, entonces $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \sum_{x_i \leq x} p(x_i) = p(-3) = \frac{1}{8}$.
4. Si $x = -1$, entonces
   $F(-1) = \mathbb{P}(\boldsymbol{X} \leq -1) = \sum_{x_i \leq -1} p(x_i) = p(-3) + p(-1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$.
5. Si $x \in [-1, 1)$, entonces $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \sum_{x_i \leq x} p(x_i) = p(-3) + p(-1) = \frac{4}{8}$.

Ejemplo función de distribución

6 Si $x = 1$, entonces
$$F(1) = \mathbb{P}(\boldsymbol{X} \leq 1) = \sum_{x_i \leq 1} p(x_i) = p(-3) + p(-1) + p(1) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}.$$

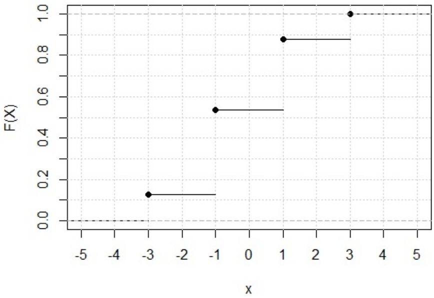
7 Si $x \in [1, 3)$, entonces $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \sum_{x_i \leq x} p(x_i) = p(-3) + p(-1) + p(1) = \frac{7}{8}$.

8 Si $x = 3$, entonces
$$F(3) = \mathbb{P}(\boldsymbol{X} \leq 3) = \sum_{x_i \leq 3} p(x_i) = p(-3) + p(-1) + p(1) + p(3) = 1.$$

Representación Gráfica de la Función de Distribución Acumulada

x	$p(x) = \mathbb{P}(\boldsymbol{X} = x)$	F(x)
-3	1/8	1/8
-1	3/8	4/8
1	3/8	7/8
3	1/8	1

$$F(x) = \begin{cases} 0 & , x < -3 \\ 1/8 & , -3 \leq x < -1 \\ 4/8 & , -1 \leq x < 1 \\ 7/8 & , 1 \leq x < 3 \\ 1 & , x \geq 3 \end{cases}$$



Propiedades de la función de distribución acumulada

Notación

$$F(\infty) = \lim_{x \to \infty} F(x); \quad F(-\infty) = \lim_{x \to -\infty} F(x)$$

Propiedad 1

Si $0 \leq F(x) \leq 1, \quad \forall x \in \mathbb{R}$, entonces $F(x)$ es una probabilidad para cualquier x real y las probabilidades están limitadas por 0 y 1.

Propiedad 2

$F(x)$ es una función no decreciente.
Sean $a, b \in \mathbb{R}$ tales que $a \leq b$ entonces:

$$\{x/\boldsymbol{X} \leq a\} \subset \{x/\boldsymbol{X} \leq b\}$$
$$F(a) = \mathbb{P}(\boldsymbol{X} \leq a) \leq \mathbb{P}(\boldsymbol{X} \leq b) = F(b)$$
$$F(a) \leq F(b)$$

Propiedades de la función de distribución acumulada

Propiedad 3

a) $F(\infty) = \mathbb{P}(\{x/\boldsymbol{X} < \infty\}) = \mathbb{P}(\boldsymbol{X} < \infty) = 1.$

b) $F(-\infty) = \mathbb{P}(\{x/\boldsymbol{X} < -\infty\}) = \mathbb{P}(\boldsymbol{X} < -\infty) = 0.$

Propiedad 4

Sea $a, b \in R_{\boldsymbol{X}}$, si x es tal que $a \leq x < b$, entonces $F(x) = F(a)$, es decir, la función $F(x)$ es constante e igual a $F(a)$ para todo $x \in [a, b)$, $F(x)$ es una función **escalonada**.

Propiedades de la función de distribución acumulada

Nota

a)

$$\mathbb{P}(\boldsymbol{X} \geq a) = 1 - \mathbb{P}(\boldsymbol{X} < a) = \begin{cases} 1 - F(a-1) & \text{Si } a \text{ es entero} \\ 1 - F([\![a]\!]) & \text{Si } a \text{ no es entero} \end{cases}$$

b)

$$\mathbb{P}(a < \boldsymbol{X} \leq b) = F(b) - F(a)$$

c)

$$\mathbb{P}(a \leq \boldsymbol{X} \leq b) = F(b) - F(a) + \mathbb{P}(\boldsymbol{X} = a)$$

d)

$$\mathbb{P}(a < \boldsymbol{X} < b) = F(b) - F(a) - \mathbb{P}(\boldsymbol{X} = b)$$

Ejemplo

La variable aleatoria $\boldsymbol{X}$ tiene la siguiente función de distribución

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 1/2 & 1 \leq x < 2 \\ 5/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Calcular:

a) La distribución de probabilidad de $\boldsymbol{X}$

b) $\mathbb{P}(\boldsymbol{X} < 0{,}5)$

c) $\mathbb{P}(\boldsymbol{X} \geq 2)$

Ejemplo solución

a) La distribución de probabilidad de $\boldsymbol{X}$

x	$p(x) = \mathbb{P}(\boldsymbol{X} = x)$
0	$1/8$
1	$3/8$
2	$1/8$
3	$3/8$

b)

$$\begin{aligned}\mathbb{P}(\boldsymbol{X} < 0{,}5) &= F(0{,}5)\\ &= F(0) = 1/8\end{aligned}$$

c)

$$\begin{aligned}\mathbb{P}(\boldsymbol{X} \geq 2) &= 1 - \mathbb{P}(\boldsymbol{X} < 2)\\ &= 1 - \mathbb{P}(\boldsymbol{X} \leq 1)\\ &= 1 - F(1) = 1 - 1/2 = 1/2\end{aligned}$$

VARIABLES ALEATORIAS CONTÍNUAS - Introducción

Definición

Si el rango $R_{\boldsymbol{X}}$, de una variable aleatoria $\boldsymbol{X}$ es un intervalo sobre la recta de los números reales, se llama **variable aleatoria contínua**

Definición (Función de Densidad de Probabilidad)

Sea $\boldsymbol{X}$ una variable aleatoria contínua con rango $R_{\boldsymbol{X}}$, **La función de densidad de probabilidad** asociado a la variable aleatoria contínua, es una función $f(x)$ integrable que satisface las siguientes condiciones:

1. $f(x) \geq 0$ para todo $x \in \mathbb{R}$ ó $f(x) > 0, x \in R_{\boldsymbol{X}}$.
2. $$\int_{R_{\boldsymbol{X}}} f(x)dx = 1; \text{ó} \quad \int_{-\infty}^{+\infty} f(x)dx = 1$$

Observaciones

1. $f(x)$ no representa una probabilidad.
2. Para cualquier valor $a \in \mathbb{R}$, $\quad \mathbb{P}(\boldsymbol{X} = a) = 0$

$$\mathbb{P}(\boldsymbol{X} = a) = \mathbb{P}(a < \boldsymbol{X} \leq a) = \int_a^a f(x)dx = 0$$

   Este resultado admite que $\boldsymbol{X}$ asume todos los valores de algún intervalo, $\mathbb{P}(\boldsymbol{X} = a) = 0$ no es equivalente a que el evento $\boldsymbol{X} = a$ es imposible.

3. Una consecuencia de 2 es el siguiente resultado.

$$\mathbb{P}(a \leq \boldsymbol{X} \leq b) = \mathbb{P}(a < \boldsymbol{X} \leq b) = \mathbb{P}(a < \boldsymbol{X} < b) = \mathbb{P}(a \leq \boldsymbol{X} < b)$$

Ejemplo

Sea $\boldsymbol{X}$ una variable aleatoria continua con función de densidad:

$$f(x) = \begin{cases} a(\ 3x\ \ -x^2\ ) & 0 \leq x \leq 3 \\ 0 & e.o.c \end{cases}$$

a) hallar el coeficiente a

b) Construir la gráfica de la funcion de distribución

c) Calcular $\mathbb{P}(1 \leq \boldsymbol{X} \leq 2)$.

d) Calcular $\mathbb{P}(-1 < \boldsymbol{X} < \frac{3}{2})$

Ejemplo

a) hallar el coeficiente a

$$\int_{R_{\boldsymbol{X}}} f(x)dx = 1$$

$$\int_0^3 a(\ 3x\ \ - x^2\ )\ dx = 1$$

$$a\left[\frac{3}{2}x^2 - \frac{1}{3}x^3\right]_0^3 = 1$$

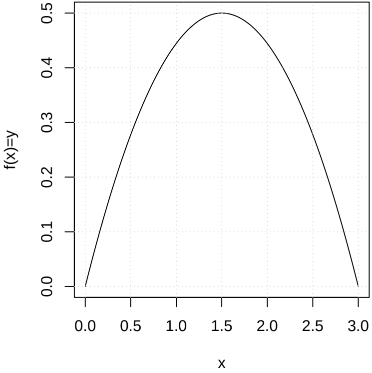
$$a\left(\frac{27}{2} - 9\right) = 1$$

$$a = \frac{2}{9}$$

$$f(x) = \begin{cases} \frac{2}{9}(\ 3x\ \ - x^2\ ) & 0 \leq x \leq 3 \\ 0 & e.o.c \end{cases}$$

Ejemplo

b) La gráfica de la función $f(x)$ en el intervalo $[0,3]$ es una parabola. $y = \frac{2}{3}x - \frac{2}{9}x^2$.



Ejemplo

c)

$$\mathbb{P}(1 \leq \boldsymbol{X} \leq 2) = \int_1^2 \frac{2}{3}x - \frac{2}{9}x^2 dx = \frac{13}{27}$$

d)

$$\mathbb{P}(-1 < \boldsymbol{X} < \frac{3}{2}) = \int_{-1}^0 0 \; dx + \int_0^{3/2} \frac{2}{3}x - \frac{2}{9}x^2 dx = \frac{1}{2}$$

Funciones de Distribución

Definición

Sea $\boldsymbol{X}$ una variable aleatoria contínua y función de densidad $f(x)$. **La función de distribución o función de distribución acumulada** de la variable aleatoria contínua $\boldsymbol{X}$, es una función $F(x)$ que se define por:

$$F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \forall x \in \mathbb{R}$$

Ejemplo

Del anterior ejemplo se tiene la siguiente función de densidad, se requiere hallar la función de distribución.

$$f(x) = \begin{cases} \frac{2}{9}(\ 3x \ \ - x^2\ ) & 0 \leq x \leq 3 \\ 0 & e.o.c \end{cases}$$

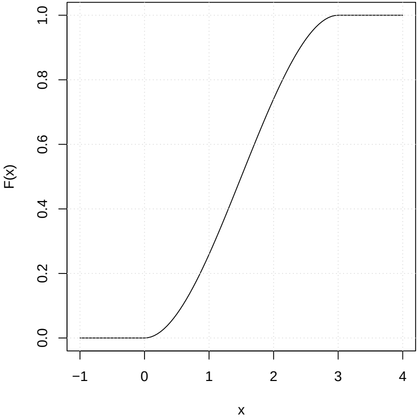
Sol.

1. Si $x < 0$, $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \displaystyle\int_{-\infty}^{x} f(t)\ dt = 0.$
2. Si $0 \leq x < 3$,
$F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \displaystyle\int_{-\infty}^{0} f(t)\ dt + \int_{0}^{x} f(t)\ dt = 0 + \int_{0}^{x} \frac{2}{9}\left(3t - t^2\right)\ dt = \frac{x^2}{9} - \frac{2x^3}{27}.$
3. Si $x \geq 3$, $F(x) = \mathbb{P}(\boldsymbol{X} \leq x) = \displaystyle\int_{-\infty}^{x} f(t)\ dt =$
$\displaystyle\int_{-\infty}^{0} f(t)\ dt + \int_{0}^{3} f(t)\ dt + \int_{3}^{x} f(t)\ dt = 0 + \int_{0}^{3} \frac{2}{9}\left(3t - t^2\right)\ dt + 0 = 1.$

Representación gráfica de la función de Distribución

En Resumen

$$F(x) = \begin{cases} 0 & , x < 0 \\ \dfrac{x^2}{3} - \dfrac{2x^3}{27} & , 0 \leq x < 3 \\ 1 & , x \geq 3 \end{cases}$$



Propiedades de la función de Distribución

1. $\mathbb{P}(\boldsymbol{X} \leq a) = \mathbb{P}(\boldsymbol{X} < a) = F(a) = \int_{-\infty}^{a} f(t)\ dt$
2. $\mathbb{P}(\boldsymbol{X} > a) = 1 - \mathbb{P}(\boldsymbol{X} \leq a) = 1 - F(a) = 1 - \int_{-\infty}^{a} f(t)\ dt$
3. $$\begin{aligned} \mathbb{P}(a \leq \boldsymbol{X} \leq b) = \mathbb{P}(a < \boldsymbol{X} \leq b) = \mathbb{P}(a \leq \boldsymbol{X} < b) = \mathbb{P}(a < \boldsymbol{X} < b) \\ = \mathbb{P}(\boldsymbol{X} \leq b) - \mathbb{P}(\boldsymbol{X} \leq a) = F(b) - F(a) \end{aligned}$$
4. $0 \leq F(x) \leq 1,\ \forall x \in \mathbb{R}.$
5. $$F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(t)\ dt = 0$$
$$F(\infty) = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \int_{-\infty}^{x} f(t)\ dt = 1$$

Propiedades de la función de Distribución

6. La función de distribución es no decreciente, esto es si $a \leq b$, entonces

$$F(a) \leq F(b)$$

7. $F(x)$ es continua por la derecha en todos los puntos.

$$\lim_{h \to 0} F(x+h) = F(x), \quad \forall \in \mathbb{R}$$

con $h > 0$.

8. Del segundo teorema fundamental del cálculo se tiene que si $F(x)$ es un función derivable, entonces.

$$f(x) = \frac{d}{dx} F(x)$$

Distribuciones de probabilidad comunes

Se dice que la variable aleatoria $\boldsymbol{X}$ tiene **distribución discreta uniforme** si la función de cuantía esta dada por:

x	p(x)
x_1	$\frac{1}{n}$
x_2	$\frac{1}{n}$
$\vdots$	$\vdots$
x_n	$\frac{1}{n}$

Esta variable tiene un número finito de puntos x_i que toma con probabilidades iguales a $p(x_i)$.

Distribución Discreta Uniforme - Ejemplo

Si la v.a. $\boldsymbol{X}$ representa el número de la cara que queda hacia arriba al lanzar una vez un dado legal, halle la función de cuantia (distribución de probabilidad).

Sol.

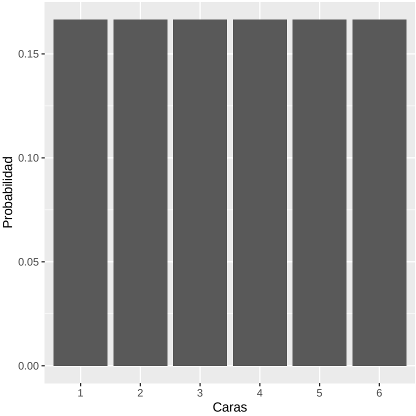
x	$p(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Cuya esperanza y varianza es:

$$\mathbb{E}(X) = \frac{7}{2} = 3{,}5 \ , \qquad \mathbb{V}(X) = \frac{35}{12} = 2{,}9166667$$

Distribución Discreta Uniforme - Ejemplo

Representación gráfica.



El Proceso Bernoulli

1. La observación consiste en clasificarlos en 2 categorias (por ejemplo éxito y fracaso, asiste y no asiste, etc.)
2. La proporción de elementos éxito en la población es constante y no se modifica cualquiera sea la cantidad observada.
   Sea p la probabilidad de observar un éxito y $1 - p$ a la probabilidad de observar un fracaso.
3. Las observaciones son independientes, es decir, la probabilidad de éxito es siempre la misma.

Se define la v.a. $\boldsymbol{X}$ de la siguiente forma:

$$\boldsymbol{X}(\text{éxito}) = 1 \quad y \quad \boldsymbol{X}(\text{fracaso}) = 0$$

El proceso Bernoulli

De esta forma:

$$\mathbb{P}(\boldsymbol{X}=1)=p \quad y \quad \mathbb{P}(\boldsymbol{X}=0)=1-p$$

Así la función de cuantia para la v.a. se define como:

$$\mathbb{P}(\boldsymbol{X}=x)=\begin{cases} p^{x}\ (1-p)^{1-x} & x=0,1 \\ 0 & e.o.c \end{cases}$$

Notación Una v.a. que tiene distribución Bernoulli con parámetro p se denota como:

$$\boldsymbol{X} \sim b(p)$$

Observación Si $\boldsymbol{X} \sim b(p)$ entonces:

$$\mathbb{E}(\boldsymbol{X})=\mu=p$$
$$\mathbb{V}(\boldsymbol{X})=\sigma^{2}=p(1-p)$$

Distribución Binomial

Suponga que se realiza n repeticiones de un *Proceso Bernoulli* bajo las siguientes condiciones:

1. La probabilidad p de que ocurra un éxito permanece constante de un ensayo a otro.
2. Las repeticiones del ensayo son independientes.

Se define la v.a. $\boldsymbol{X}$ como el número de exitos en las n repeticiones, así:

$$R_{\boldsymbol{X}} = \{0, 1, 2, \ldots, n\}$$

La función de cuantia de la v.a. $\boldsymbol{X}$ es:

$$\mathbb{P}(\boldsymbol{X} = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x \, (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & e.o.c \end{cases}$$

Donde:

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$$

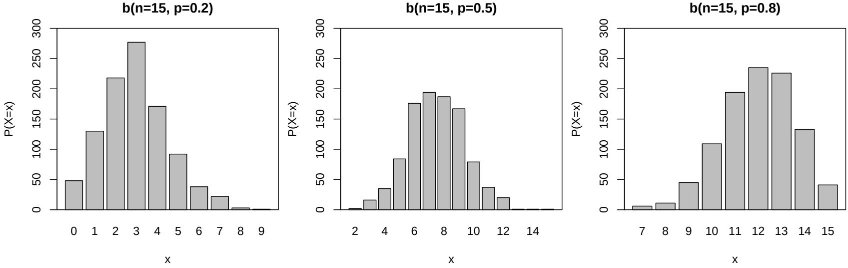
Distribución Binomial

Notación Una v.a. que tiene distribución Binomial con parámetro p y n se denota como:

$$\boldsymbol{X} \sim b(n, p)$$

Observación Si $\boldsymbol{X} \sim b(n, p)$ entonces:

$$\mathbb{E}(\boldsymbol{X}) = \mu = np$$
$$\mathbb{V}(\boldsymbol{X}) = \sigma^2 = np(1-p)$$



Ejemplo v.a. Binomial.

En la tienda de alquiler de automóviles, cada vez que un cliente arrienda un automóvil debe pagar como mínimo 14bs. Si arrienda un auto tipo Abc debe pagar 30bs. más, y si arrienda un auto tipo no Abc debe pagar 50bs. más. Se sabe que la probabilidad de que un cliente arriende un auto tipo Abc es de 0,7 de cinco clientes que arriendan autos en esa tienda.

a) Determine la distribución de probabilidades de los clientes que arriendan automóviles tipo Abc.

b) Determine la utilidad esperada que producen a la tienda los cinco clientes que arriendan automóviles.

Sol. (a).

Sea $\boldsymbol{X}$ el número de clientes que arriendan automóviles tipo Abc, entonces:

$$n = 5; \quad \mathbb{P}(Exito) = 0{,}7; \qquad R_X = \{0, 1, 2, 3, 4, 5\}$$

$$\boldsymbol{X} \sim b(n = 5, p = 0{,}7)$$

Ejemplo v.a. Binomial.

La función de cuantía es:

$$\mathbb{P}(\boldsymbol{X}=x)=\begin{cases}\begin{pmatrix}5\\x\end{pmatrix}(0{,}7)^x\ (0{,}3)^{5-x} & x=0,1,2,3,4,5\\ 0 & e.o.c\end{cases}$$

$$\mathbb{E}(\boldsymbol{X})=np=5*0{,}7=3{,}5$$

(b)

Precio (Bs.)	Nro. de Clientes
14	5
30	$\boldsymbol{X}$
50	$5-\boldsymbol{X}$

Ecuación de la utilidad:

$$\begin{aligned}\boldsymbol{U}&=70+30\boldsymbol{X}+50(5-\boldsymbol{X})\\&=320-20\boldsymbol{X}\end{aligned}$$

Utilidad esperada

$$\begin{aligned}\mathbb{E}(\boldsymbol{U})&=\mathbb{E}(320-20\boldsymbol{X})\\&=320-20\mathbb{E}(\boldsymbol{X})\\&=320-20(3{,}5)=250\end{aligned}$$

Distribución contínua Uniforme

Suponga que ocurra en evento en que una v.a. **toma valores en un intervalo finito, de manera que estos se encuentran distribuidos uniformemente sobre el intervalo**. Entonces, la probabilidad es la misma para todos los valores de la v.a.

Se dice que una v.a. $\boldsymbol{X}$ está distribuida uniformemente sobre el intervalo $[a, b]$ si su función de densidad está dado por:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & e.o.c \end{cases}$$

Y la función de distribución acumulada de la distribución uniforme está dada por:

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Distribución Uniforme

Para indicar que la v.a. $\boldsymbol{X}$ tiene distribución uniforme en el intervalo $[a, b]$, se usa la notación:

$$\boldsymbol{X} \sim U(a, b)$$

Observación

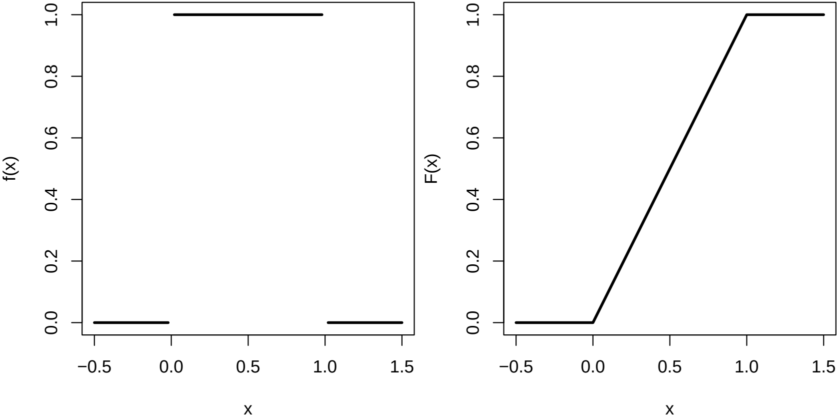
Si $\boldsymbol{X} \sim U(a, b)$, entonces

$$\mathbb{E}(\boldsymbol{X}) = \mu = \frac{a+b}{2}$$

$$\mathbb{V}(\boldsymbol{X}) = \sigma^2 = \frac{(b-a)^2}{12}$$

Gráfica

Si $\boldsymbol{X} \sim U(0,1)$ entonces:



Ejemplo de la Distribución Uniforme

Dos gerentes A y B deben encontrarse en cierto lugar entre las 7p.m. y 8p.m. para firmar un contrato. Cada uno espera al otro a lo más 10 minutos, ¿Cuál es la probabilidad de que no se encuentren sabiendo que A llega a las 7.30 p.m.?.

Sol. Sea $\boldsymbol{X}$ el tiempo de llegada del gerente B que puede hacerlo en cualquier instante aleatorio entre las 7 p.m. y las 8 p.m. o entre 0 y 60 minutos. Luego, $\boldsymbol{X} \sim U(0, 60)$ y su función de densidad es:

$$f(x) = \begin{cases} \dfrac{1}{60}, & 0 \leq x \leq 60 \\ 0, & e.o.c \end{cases}$$

Puesto que A llega a las 7.30 p.m. o a los 30 minutos después de las 7 p.m. y espera a lo más 10 minutos, B no se encontrará con A si B llega de 7 p.m. a menos de 7.20 p.m. o si llega despues de las 7.40 p.m. la probabilidad de que A y B no se encuentre es:

$$\mathbb{P}(0 \leq X < 20 \ \cup \ 40 < X \leq 60) = \int_0^{20} \frac{1}{60} dx + \int_{40}^{60} \frac{1}{60} dx = \frac{2}{3}$$

Distribución Normal

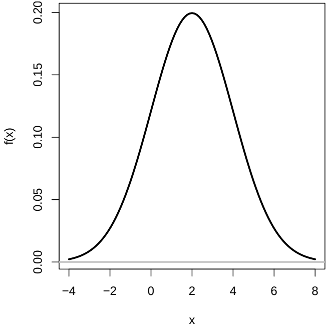
Definición

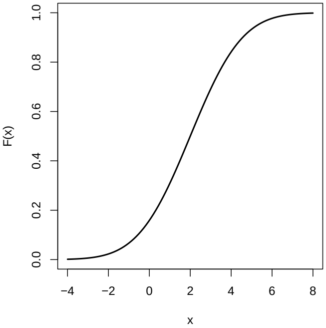
Se dice que la variable aleatoria continua, $\boldsymbol{X}$ que toma valores reales $\mathbb{R}$, se distribuye **Normalmente** o **Normal** con parámetros μ y σ y se denota por $\boldsymbol{X} \sim N(\mu, \sigma^2)$, si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < +\infty$$

Distribución Normal

La gráfica para $\mu = 2; \sigma = 2$:





Propiedades

- f es simétrica con respecto al eje vertical $\boldsymbol{X} = \mu$.

$$f(\mu + x) = f(\mu - x)$$

- f tiene un valor máximo en $x = \mu$ (su moda). Este valor máximo es:

$$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

- f tiene al eje $\boldsymbol{X}$ una asíntota horizontal, ya que :

$$\underset{x \to -\infty}{Lim} f(x) = 0 \;\text{ y }\; \underset{x \to +\infty}{Lim} f(x) = 0$$

- Tiene puntos de inflexión en $x = \mu - \sigma$ y $x = \mu + \sigma$, por tanto, es cóncava hacia abajo en el intervalo $\mu - \sigma < x < \mu + \sigma$, y cóncava hacia arriba en cualquier otra parte.

Propiedades

Teorema

Si la variable aleatoria $\boldsymbol{X}$ tiene distribución normal $N(\mu, \sigma^2)$, entonces:

$$\mathbb{E}(\boldsymbol{X}) = \mu \quad \mathbb{V}(\boldsymbol{X}) = \sigma^2$$

Distribución Normal estándar y uso de la tabla normal

Si la variable aleatoria $\boldsymbol{X} \sim N(\mu, \sigma^2)$, entonces la variable aleatoria:

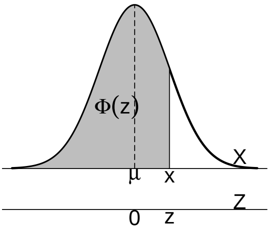
$$\boldsymbol{Z} = \frac{\boldsymbol{X} - \mu}{\sigma}$$

Tiene Distribución normal $N(0,1)$ y tiene nombre propio se denomina **Distribución Normal Estándar**, esto es:

$$\mathbb{E}(\boldsymbol{Z}) = 0 \ \text{ y } \ \mathbb{V}(\boldsymbol{Z}) = 1$$

La función de densidad de la variable aleatoria $\boldsymbol{Z}$ y su función de Distribución acumulativa es:

$$f(z) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{z^2}{2}\right\}; \quad F(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{t^2}{2}\right\} dt$$



Tabulación de la distribución normal

La probabilidad:

$$\mathbb{P}(a \leq \boldsymbol{Z} \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{z^2}{2}\right\} dz$$

No puede calcularse por métodos de integración directa. Sin embargo usando métodos de integración numérica se han tabulado los valores de la función de distribución acumulada de la normal estándar.

$$\Phi(Z) = \mathbb{P}(\boldsymbol{Z} \leq z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{t^2}{2}\right\} dt$$

Propiedades.-

1. $\mathbb{P}(a \leq \boldsymbol{Z} \leq b) = \Phi(b) - \Phi(a)$.
2. $\Phi(-z) = 1 - \Phi(z)$.
3. $\mathbb{P}(-a \leq \boldsymbol{Z} \leq a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1$.
4. Si la variable aleatoria $\boldsymbol{X}$ tiene distribución $N(\mu, \sigma^2)$, entonces, la variable aleatoria $\boldsymbol{Z} = (\boldsymbol{X} - \mu)/\sigma$ tiene distribución $N(0, 1)$.

$$\mathbb{P}(a \leq \boldsymbol{X} \leq b) = \mathbb{P}\left[\frac{a-\mu}{\sigma} \leq \boldsymbol{Z} \leq \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Tabulación de la distribución normal

En la tabla de probabilidades de la normal estándar se puede encontrar la probablidad $1-\alpha$, o el valor $z_{1-\alpha}$ de $\boldsymbol{Z}$ mediante la relación:

$$\mathbb{P}(\boldsymbol{Z} \leq z_{1-\alpha}) = 1-\alpha \quad \text{ó} \quad \Phi(z_{1-\alpha}) = 1-\alpha$$

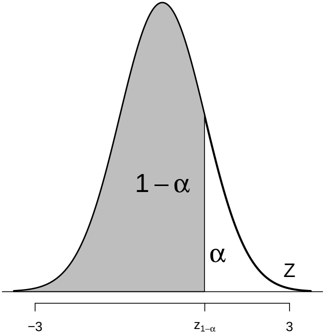
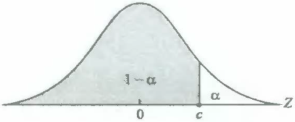


tabla Normal Estandar

TABLA DE LA DISTRIBUCION NORMAL ESTANDAR

La tabla da áreas $1-\alpha$ y valores $c = Z_{1-\alpha}$, donde, $P[Z \leq c] = 1-\alpha$, y donde Z tiene distribución normal $N(0,1)$.



Z	Segundo decimal de Z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

tabla Normal Estandar cont.

1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	1.000	1.000	1.000
4.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Ejemplos

Utilizando la tabla de probabilidades normal hallar:

1. $\mathbb{P}(\boldsymbol{Z} \leq 1{,}2)$, SOL 0,8849
2. $\mathbb{P}(0{,}81 \leq \boldsymbol{Z} \leq 1{,}94)$, SOL 0,1828
3. $\mathbb{P}(\boldsymbol{Z} \leq -1{,}28)$, SOL $\Phi(-1{,}28) = 1 - \Phi(1{,}28) = 1 - 0,8997 = 0{,}1003$
4. $\mathbb{P}(-0{,}46 \leq \boldsymbol{Z} \leq 2{,}21)$, SOL $\Phi(2{,}21) + \Phi(0{,}46) - 1 = 0{,}9864 + 0{,}6772 - 1 = 0{,}6636$
5. $\mathbb{P}(\boldsymbol{Z} \geq -0{,}68)$, SOL $\Phi(0{,}68) = 0{,}7517$

Ejemplos

```
from scipy import stats
Z = stats.norm()
print("P(Z < 1.2) = ")
```

```
P(Z < 1.2) =
```

```
print(Z.cdf(1.2))
```

```
0.8849303297782918
```

Ejemplos

Hallar el valor de z tal que:

- $\mathbb{P}(\boldsymbol{Z} \leq z) = 0{,}8621$, Sol $z = 1{,}09$
- $\mathbb{P}(\boldsymbol{Z} \leq z) = 0{,}2236$, Sol $z = -0{,}76$
- $\mathbb{P}(-z \leq \boldsymbol{Z} \leq z) = 0{,}95$, Sol $z = \pm 1{,}96$
- $\mathbb{P}(-z \leq \boldsymbol{Z} \leq z) = 0{,}9900$, Sol $2\Phi(z) - 1 = 0{,}9900$, implica $\Phi(z) = 0{,}9950$

Ejemplos

```
from scipy import stats
Z = stats.norm()
# P(Z < z) = 0.8621
print(Z.ppf(0.8621))
```

```
1.0898028460795839
```

Ejemplos

Suponga que el ingreso familiar mensual en una comunidad tiene distribución nomal con media 600Bs. y Desviación estándar 100Bs.

a) Calcular la probabilidad de que el ingreso de una familia escogida al azar sea menor que 400Bs.

b) Si el 5 % de las familias con mayores ingresos deben pagar un impuesto, ¿A partir de que ingreso familiar se debe pagar el impuesto?.

Sol. Sea $\boldsymbol{X}$ la variable aque representa los ingresos familiares mensuale, la distribución de $\boldsymbol{X}$ es $N(600, 100^2)$.

a) $\mathbb{P}(\boldsymbol{X} < 400) = \mathbb{P}\left(\boldsymbol{Z} < \dfrac{400-600}{100}\right) = \mathbb{P}(\boldsymbol{Z} < -2) = 0{,}0228.$

b) Se debe hallar k tal que $\mathbb{P}(\boldsymbol{X} \geq k) = 0{,}05$ o $\mathbb{P}(\boldsymbol{X} < k) = 0{,}95$, entonces $0{,}95 = \mathbb{P}(\boldsymbol{X} < k) = \mathbb{P}\left(\boldsymbol{Z} < \dfrac{k-600}{100}\right)$, de donde resulta $\dfrac{k-600}{100} = 1{,}645$, $k = 764{,}5$.