Análisis y Procesamiento de Datos

F. Hidrobo, K. Tucci, M. Uzcátegui

Universidad de Los Andes
Facultad de Ciencias
SUMA

{hidrobo,kay,maye}@ula.ve

Objetivos del Curso

Objetivos

- Proveer al estudiante de una visión general, y de nociones operativas, de las técnicas comúnmente utilizadas en el análisis y procesamiento de datos.
- Proveer al estudiante de la capacidad para extraer información relevante de un conjunto de datos.
- Proveer al estudiante de la capacidad para representar información relevante extraída de los datos, mediante el uso de técnicas de visualización científica.

Esquema de Evaluación

Dos tipos

- **Proyectos**: Análisis y procesamiento de datos
- **Artículo**: Cada proyecto, o tarea, deberá ser presentado en forma de artículo

¿Qué vamos a usar en el curso?

- Lenguajes de programación (C, Fortran, Java, etc.)
- Matlab, Octave, Maple, etc.
- Paquetes estadísticos.
- Visualización (Paquetes, Entornos, Bibliotecas, etc.)

Definición

Estadística

Conjunto de métodos para la recolección, organización, análisis e interpretación de datos.

Algunos la consideran una ciencia, otros una rama de las matemáticas.

Clasificación

Estadística Descriptiva

Describe, de forma cuantitativa, las características principales de un conjunto de datos. Resume dichas características a través del cálculo de un conjunto de métricas.

Estadística Inferencial

Permite deducir (hacer inferencias) características o propiedades de un conjunto de datos (**TODO**) a partir de información tomada para un subconjunto (**PARTE**) de éste.

Conceptos Básicos

- **Población**: Conjunto de “objetos” agrupados según alguna(s) característica(s) o propiedad(as).
- **Muestra**: Subconjunto de “objetos” pertenecientes a una población.
- **Muestra aleatoria**: Subconjunto de “objetos” seleccionados “aleatoriamente” de una población.
- **Muestreo**: Procedimiento y/o técnica que se utiliza para obtener muestras de una población.

Conceptos Básicos

Datos y Variables Estadísticas

- **Dato estadístico**: Un valor que se obtiene a través de la observación o medición de algún concepto (atributo).
    - Valor Cualitativo: No numérico.
      Género: Masculino, Femenino
      Color: Verde, Amarillo, Rojo
    - Valor Cuantitativo: Numérico (Tiene asociado una unidad de medida).
      Volumen: 22.5 mts^3
      Edad: 12 años

Conceptos Básicos

Datos y Variables Estadísticas

- **Parámetro**: Una característica descriptiva GLOBAL de una población.
- **Estadígrafo**: Característica descriptiva GLOBAL de una muestra que se usa para estimar o inferir un parámetro.
- **Estadístico**: Es cualquier cantidad cuyo valor se pueda calcular a partir de datos muestrales. Un estadístico es una variable aleatoria y estará denotada por una letra mayúscula; una minúscula se emplea para representar el valor calculado u observado del estadístico.

Medidas de Tendencia

Medidas de Tendencia Central

Valor numérico que se sitúa hacia el "centro" de la distribución

Las más usadas

- **Promedio (Media aritmética)**. La suma de cada valor del dato estadístico dividido por el número de muestras.

$$\bar{x} = \frac{\sum_i^n x_i}{n}$$

$$\mu = \frac{\sum X}{N}$$

Otros Promedios

- **Promedio Ponderado**.

$$\bar{x}_w = \frac{\sum_i^n (w_i * x_i)}{\sum_i^n w_i}$$

- **Media Armónica**.

$$H = \frac{n}{\sum_i^n \frac{1}{x_i}}$$

- **Media Geométrica**.

$$GM = \sqrt[n]{\prod_i^n x_i}$$

Medidas de Tendencia

Las más usadas

- **Mediana**. Valor central del grupos de datos ordenados.

$$
\begin{array}{ll}
x_{(\frac{n}{2}+1)} & si \;\; n \; es \; impar \\
\dfrac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & si \;\; n \; es \; par
\end{array}
$$

- **Moda**. Valor que más se repite en la muestra.
- **Media Cuadrática**

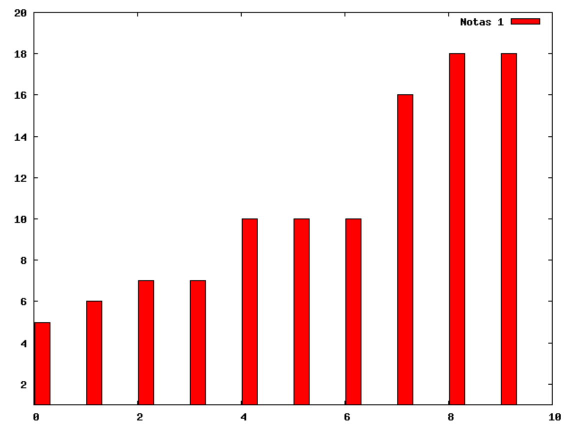
$$\bar{x^2} = \frac{\sum_i^n {x_i}^2}{n}$$

Notas del curso 1

Datos	18 16 5 6 7 10 18 10 10 7
Ordenados	5 6 7 7 10 10 10 16 18 18
Promedio	10,7
Moda	10
Mediana	10

Representación Gráfica

Diagrama de Barras



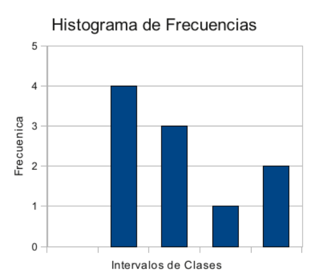
Representación Gráfica

Distribución de Frecuencias

Agrupamiento de datos en categorías mutuamente excluyentes, que indican el número de observaciones en cada categoría.

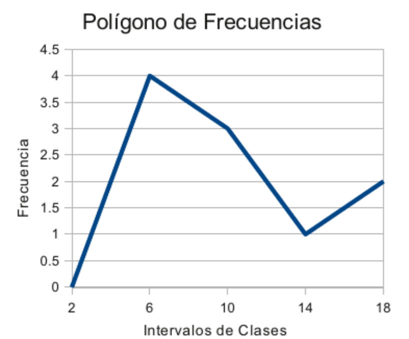
Clases: Notas de 1 a 4, de 5 a 8, de 9 a 12, de 13 a 16 y de 17 a 20.

Intervalo de clase: 4



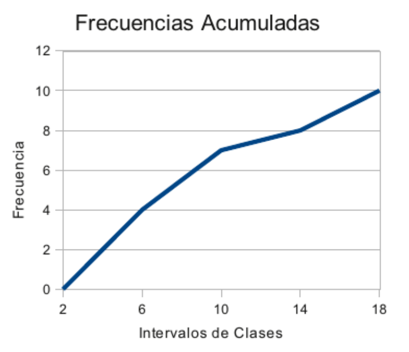
Representación Gráfica

Polígono de Frecuencias



Representación Gráfica

Distribución de Frecuencias Acumuladas



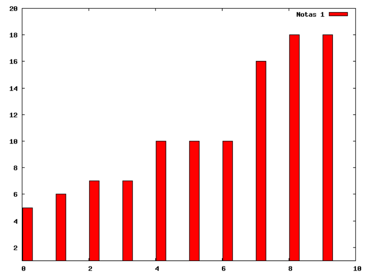
Ejemplo

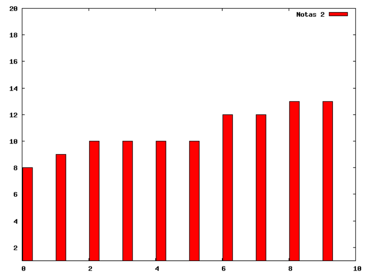
Notas del curso 2

Datos	10 13 13 12 8 9 10 12 10 10
Ordenados	8 9 10 10 10 10 12 12 13 13
Promedio	10,7
Moda	10
Mediana	10

Comparación de Datos

Gráficamente (Estadística Descriptiva)





Moda, Mediana y Promedio IGUALES.
¿Conclusiones?

Medidas de Tendencia

Dispersión Estadística

Conocida también como variabilidad o variación estadística, es la variabilidad o extensión en una variable o una distribución de probabilidad.

Una medida de dispersión estadística es un número real no negativo, que vale cero si todos los datos son idénticos, y aumenta a medida que los datos están más dispersos.

Medidas de Dispersión

Medidad comunes de dispersión

- **Rango**:
  Es la diferencia entre el valor máximo y el mínimo en la muestra
  $$x_{max} - x_{min}$$
  Observación: Solamente toma en cuenta los valores extremos.
  En el ejemplo anterior:
  Rango "Notas 1" es 13 (18 - 5)
  Rango "Notas 2" es 5 (13 - 8)

Medidas de Dispersión

Medidad comunes de dispersión

- **Varianza**: mide la dispersión de los valores respecto a un valor central (media), específicamente el cuadrado de las desviaciones respecto a dicho valor.

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

$y_i = x_i + k \Rightarrow S_y^2 = S_x^2$

$y_i = x_i * k \Rightarrow S_y^2 = k^2 * S_x^2$

En el ejemplo:

Varianza "Notas 1" es 24,23

Varianza "Notas 2" es 3,57

Medidas de Dispersión

Medidad comunes de dispersión

- **Desviación Típica (Desviación Estándar)**: Es la raíz cuadrada (positiva) de la varianza, se utiliza para eliminar el uso de unidades cuadráticas.

$$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}}$$

  En el ejemplo:
  Desviación estándar "Notas 1" es 4,92
  Desviación estándar "Notas 2" es 1,89

Medidas de Dispersión

Medidas comunes de dispersión

- **Coeficiente de Variación**: cociente entre la desviación típica y el promedio. Con esta medida se elmina la influencia de la escala escogida en las mediciones efectuadas.

$$d = \frac{S_x}{\overline{x}}$$

En el ejemplo:
Coeficiente de variación "Notas 1" es 0,46
Coeficiente de variación "Notas 2" es 0,18

Medidas de Dispersión

Medidas comunes de dispersión

- **Desigualdad de Chebyshev:** Para una colección de datos, con promedio $\bar{x}$ y desviación estándar S_x, se cumple que la proporción de datos contenidos en en el intervalo $(\bar{x} - k * S_x; \bar{x} + k * S_x)$ es igual o mayor que $(1 - \frac{1}{k^2})$, para cualquier valor real de k mayor que 1.

Correlación estadística

Covarianza

Es una medida de que tanto dos VARIABLES cambian juntas (Varianza es un caso especial de covarianza cuando las dos variables son idénticas):

$$S_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})}{n-1}$$

Correlación estadística

Coeficiente de correlación

Mide el grado de asociación lineal entre las variables:

$$r = \frac{S_{xy}}{S_x * S_y}$$

Propiedades:

- $r \in [-1, 1]$
- Si $|r| = 1$ existe una correlación lineal perfecta entre X y Y.
- Si $r \approx 0$ la asociación lineal es débil.
- r no tiene dimensión, si cambia la escala o el origen se mantiene constante.

Teoría de Probabilidades

Conceptos básicos

- **Experimento**: Proceso mediante el cual se obtiene alguna observación. Ejemplo "Lanzar un dado"
- **Resultado**: es un suceso particular proveniente de un experimento.
- **Evento**: es un conjunto de uno o más resultados de un experimento
  - **Simple**: Resultado básico que no puede ser dividido. Ejemplo "Obtener un valor UNO al lanzar un dado"
  - **Compuesto**: Conjunto de eventos simples. Ejemplo "Obtener un valor IMPAR al lanzar un dado"

Conceptos básicos de Probabilidades

Espacio Muestral

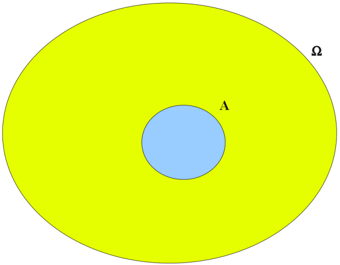
- **(Ω)**: TODOS los eventos SIMPLES de un experimento. Ejemplo. En el caso de lanzar un dado el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$

Ejemplos

- Lanzar dos modedas.
$$\Omega = \{CC, SS, CS, SC\}$$
- Lanzar una moneda y un dado.
$$\Omega = \{C1, S1, C2, S2, C3, S3, C4, S4, C5, S5, C6, S6\}$$
- Lanzar dos dados.
$$\Omega = \{11, 12, 13, 14, 15, 16, 21, \cdots, 64, 65, 66\}$$

Teoría de Conjuntos para eventos

Diagrama de Venn

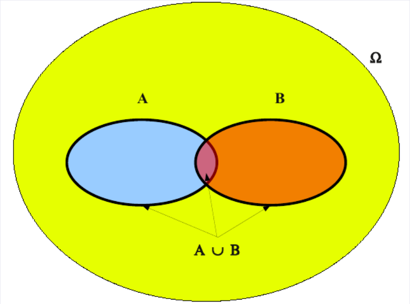


Representación del evento A perteneciente al espacio muestral Ω. Ω es el espacio muestral obtenido al "lanzar dos dados", A puede ser, por ejemplo, que la suma de los valores sea SIETE (7).

Operaciones sobre conjuntos

UNION

El evento **Unión**, denotado $A \cup B$, consiste de todos los eventos simples que estan contenidos en A o B y en ambos.
$A \cup B$ puede ser descrito como que ocurre por lo menos uno de los dos eventos A o B.



Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:

Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

Operaciones sobre conjuntos

Ejemplo de UNION

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

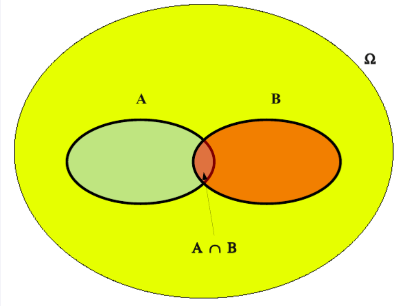
$A \cup B = \{16,25,34,43,52,11,21,22,31,33,41,42,44,51,55,62,63,66, 61\}$

Operaciones sobre conjuntos

INTERSECCIÓN

El evento **Intersección**, denotado $A \cap B$, consiste de todos los eventos simples comunes de A y B.

$A \cap B$ puede ser descrito como que ocurre ambos eventos A y B.



Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 7:

$$A = \{16, 25, 34, 43, 52, 61\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

$$A \cap B = \{61\}$$

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 8:

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 8:

$$A = \{26, 35, 44, 53, 62\}$$

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 8:

$$A = \{26, 35, 44, 53, 62\}$$

B el evento que el primer número sea divisible entre el segundo

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 8:

$$A = \{26, 35, 44, 53, 62\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

Operaciones sobre conjuntos

Ejemplo de INTERSECCIÓN modificado

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras de 8:

$$A = \{26, 35, 44, 53, 62\}$$

B el evento que el primer número sea divisible entre el segundo

$$B = \{11, 21, 22, 31, 33, 41, 42, 44, 51, 55, 61, 62, 63, 66\}$$

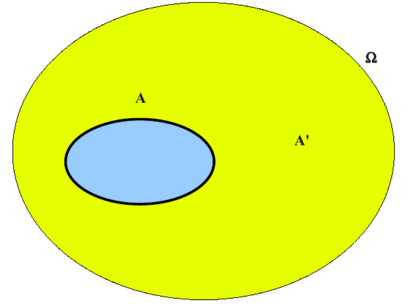
$$A \cap B = \{44, 62\}$$

Operaciones sobre conjuntos

COMPLEMENTO

El evento **Complemento**, denotado A', consiste de todos los eventos simples que no están en A.

A' puede ser descrito como que NO ocurre A.



$$A \cup A' = \Omega$$

Operaciones sobre conjuntos

Ejemplo de COMPLEMENTO

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sea:

Operaciones sobre conjuntos

Ejemplo de COMPLEMENTO

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sea:
A el evento que la suma de las caras sea PAR:

Operaciones sobre conjuntos

Ejemplo de COMPLEMENTO

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sea:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

Operaciones sobre conjuntos

Ejemplo de COMPLEMENTO

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sea:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

El evento A' será que la suma de las caras NO SEA PAR (sea IMPAR)

Operaciones sobre conjuntos

Ejemplo de COMPLEMENTO

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sea:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

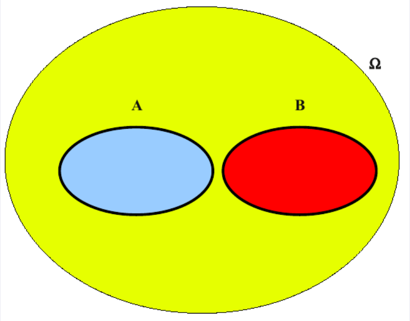
El evento A' será que la suma de las caras NO SEA PAR (sea IMPAR)

$$A' = \{12, 14, 16, 21, 23, 25, \cdots, 61, 63, 65\}$$

Relación entre Eventos

Eventos Disjuntos

Dos eventos son DISJUNTOS, mutuamente excluyentes, si no pueden ocurrir al mismo tiempo, en otras palabras, ellos no tienen eventos en común. A y B son disjuntos si $A \cap B = \emptyset$



Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:

Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras sea PAR:

Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

B el evento que la suma de las caras de 7:

Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

B el evento que la suma de las caras de 7:

$$B = \{16, 25, 34, 43, 52, 61\}$$

Operaciones sobre conjuntos

Ejemplo de Eventos Disjuntos

Sea Ω el espacio muestral resultante de "Lanzar DOS dados":
$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, \cdots, 64, 65, 66\}$
Sean:
A el evento que la suma de las caras sea PAR:

$$A = \{11, 13, 15, 22, 24, 26, \cdots, 62, 64, 66\}$$

B el evento que la suma de las caras de 7:

$$B = \{16, 25, 34, 43, 52, 61\}$$

A y B son disjuntos,

$$A \cap B = \emptyset$$

Relación entre eventos

Otras relaciones

- Eventos independientes: la ocurrencia de un evento no afecta la ocurrencia del otro.
  Ejemplo: Obtener UNO al lanzar un dado, y DOS al lanzar el otro dado.
- Conjunto colectivamente exhaustivo: por lo menos uno de los eventos debe ocurrir cuando se realiza un experimento.

Definición de Probabilidad

Si $S \in \Omega$, entonces $P(S)$: es una medida de la posibilidad relativa de que el evento S ocurra.
La probabilidad puede definirse como una función $P : \Omega \rightarrow \mathbb{R}$

Definiciones de Probabilidad

Definición de Probabilidad

Si $S \in \Omega$, entonces $P(S)$: es una medida de la posibilidad relativa de que el evento S ocurra.
La probabilidad puede definirse como una función $P : \Omega \rightarrow \mathbb{R}$

Definiciones de Probabilidad

- **Clásica**: cuando hay n resultados posibles, se asigna la probabilidad a un suceso antes de que éste ocurra.

Definición de Probabilidad

Si $S \in \Omega$, entonces $P(S)$: es una medida de la posibilidad relativa de que el evento S ocurra.
La probabilidad puede definirse como una función $P : \Omega \rightarrow \mathbb{R}$

Definiciones de Probabilidad

- **Clásica**: cuando hay n resultados posibles, se asigna la probabilidad a un suceso antes de que éste ocurra.
- **Empírica**: para asignar la probabilidad, el número de veces que ocurre un evento se divide por el número total de observaciones.

Definición de Probabilidad

Si $S \in \Omega$, entonces $P(S)$: es una medida de la posibilidad relativa de que el evento S ocurra.
La probabilidad puede definirse como una función $P : \Omega \to \mathbb{R}$

Definiciones de Probabilidad

- **Clásica**: cuando hay n resultados posibles, se asigna la probabilidad a un suceso antes de que éste ocurra.
- **Empírica**: para asignar la probabilidad, el número de veces que ocurre un evento se divide por el número total de observaciones.
- **Subjetiva**: se basa en cualquier información disponible.

Reglas de probabilidad

Básicas (Axiomas)

1. $0 \leq P(S) \leq 1$
2. $P(\Omega) = 1$
3. $P(\emptyset) = 0$

Otras Reglas

1. Dado un evento A, $P(A') = 1 - P(A)$
2. Si $M = \{\bigcup_i^n A_i\}$, y A_i es disjunto de A_j, $\forall i \neq j$, entonces $P(M) = \sum_i^n P(A_i)$
   En general, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. Si A y B son independientes, $P(A \cap B) = P(A) * P(B)$

Probabilidad condicional

- Es la probabilidad de que ocurra un evento determinado, dado que otro evento ya haya ocurrido.
- La probabilidad de que ocurra el evento A dado que el evento B ha ocurrido se denota por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad si \;\; P(B) > 0$$

Probabilidad condicional

Ejemplo

Se tiran dos dados y se sabe que el primero no tiene el número 5. ¿Cuál es la probabilidad de que la suma de los dados sea 7?
Para resolver, llamemos
B el evento: "el primer dado no es 5".
A el evento: "la suma de los dados es 7".
$P(B) = 30/36$. 36 parejas posibles, 6 tienen 5 en el primer dado.
$P(A \cap B) = 5/36$. Porque sólo se obtiene 7, con las parejas (1,6), (2,5), (3,4), (4,3) y (6,1). La pareja (5,2) suma SIETE pero tiene un 5 en el primer dado. Así:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{5/36}{30/36} = \frac{5}{30}$$

Probabilidad condicional

Regla General de la multiplicación

Usando la probabilidad condicional, la probabilidad conjunta está definida por

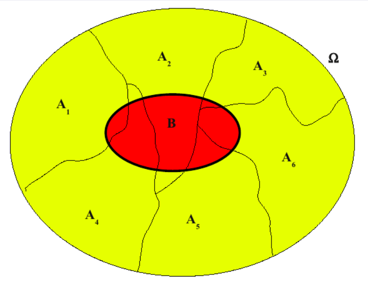
$$P(A \cap B) = P(A/B) * P(B)$$

$$P(A \cap B) = P(B/A) * P(A)$$

Teoremas

Probabilidad Total

Dada una partición de Ω en n eventos A_i, tal que $\Omega = \{\cup A_i, \forall i(1 \leq i \leq n)\}$, y A_i es disjunto de A_j, $\forall i \neq j$, y dado un evento B.

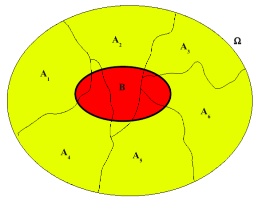


$$P(B) = \sum_{i}^{n} \{P(B/A_i) * P(A_i)\}$$

Teoremas

Teorema de Bayes

$$P(A_i/B) = \frac{P(B/A_i) * P(A_i)}{\sum_j^n \{P(B/A_j) * P(A_j)\}}$$



Principios de Conteo

Fórmula de multiplicación

Dado un espacio muestral Ω_1 con m eventos, y un espacio muestral Ω_2 con n eventos, ¿Cuántas parejas de eventos se pueden formar tomando un evento de Ω_1 y un evento de Ω_2?.

Principios de Conteo

Fórmula de multiplicación

Dado un espacio muestral Ω_1 con m eventos, y un espacio muestral Ω_2 con n eventos, ¿Cuántas parejas de eventos se pueden formar tomando un evento de Ω_1 y un evento de Ω_2?.

$$T_{12} = n * m$$

Principios de Conteo

Fórmula de multiplicación

Dado un espacio muestral Ω_1 con m eventos, y un espacio muestral Ω_2 con n eventos, ¿Cuántas parejas de eventos se pueden formar tomando un evento de Ω_1 y un evento de Ω_2?.

$$T_{12} = n * m$$

Si hay m formas de hacer una cosa, y n formas de hacer otra, existirán $m * n$ formas de hacer ambas.

Principios de Conteo

Ejemplo de multiplicación

Dado el espacio muestral Ω_1 "lanzar un dado", y un espacio muestral Ω_2 "lanzar un moneda", ¿Cuántas parejas de eventos se pueden formar tomando un evento de Ω_1 y un evento de Ω_2?.

Principios de Conteo

Ejemplo de multiplicación

Dado el espacio muestral Ω_1 "lanzar un dado", y un espacio muestral Ω_2 "lanzar un moneda", ¿Cuántas parejas de eventos se pueden formar tomando un evento de Ω_1 y un evento de Ω_2?.

$$T_{12} = 6 * 2 = 12$$

Principios de Conteo

Variaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento o en el orden de colocación de ellos.

Principios de Conteo

Variaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento o en el orden de colocación de ellos.

$$V_n^m = n*(n-1)*(n-2)*(n-3)....(n-m+1) = \frac{n!}{(n-m)!}$$

Principios de Conteo

Variaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento o en el orden de colocación de ellos.

$$V_n^m = n * (n-1) * (n-2) * (n-3) \ldots (n-m+1) = \frac{n!}{(n-m)!}$$

Si los elementos se pueden repetir se llaman variaciones con repetición.

Principios de Conteo

Variaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento o en el orden de colocación de ellos.

$$V_n^m = n*(n-1)*(n-2)*(n-3)....(n-m+1) = \frac{n!}{(n-m)!}$$

Si los elementos se pueden repetir se llaman variaciones con repetición.

$$VR_n^m = n*(n)*(n)*(n)....(n) = n^m$$

Principios de Conteo

Ejemplo de Variaciones 1 (Clave de Cajero)

Clave de Cajero. Se toman 4 números de un total de 10, no puede tener números repetidos
¿Cuántas CLAVES diferentes pueden usarse?:

Principios de Conteo

Ejemplo de Variaciones 1 (Clave de Cajero)

Clave de Cajero. Se toman 4 números de un total de 10, no puede tener números repetidos
¿Cuántas CLAVES diferentes pueden usarse?:

$$V_{10}^{4} = 10 * 9 * 8 * 7 = \frac{10!}{(10-4)!} = 5040$$

Principios de Conteo

Ejemplo de Variaciones 1 (Clave de Cajero)

Clave de Cajero. Se toman 4 números de un total de 10, no puede tener números repetidos
¿Cuántas CLAVES diferentes pueden usarse?:

$$V_{10}^{4} = 10 * 9 * 8 * 7 = \frac{10!}{(10-4)!} = 5040$$

¿Cuál es la probabilidad de que si pierde la tarjeta alguién ADIVINE su clave?
Si le permiten 3 intentos, será

$$3/5040 = 5{,}952380952380953 * 10^{-4}$$

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

Clave de Cajero. Se toman 4 números de un total de 10, puede tener números repetidos
¿Cuántas CLAVES diferentes pueden usarse?:

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

Clave de Cajero. Se toman 4 números de un total de 10, puede tener números repetidos
¿Cuántas CLAVES diferentes pueden usarse?:

$$VR_{10}^{4} = 10 * 10 * 10 * 10 = 10000$$

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

¿Cuál es la probabilidad de que si pierde la tarjeta alguién ADIVINE su clave?

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

¿Cuál es la probabilidad de que si pierde la tarjeta alguién ADIVINE su clave?

- Si le permiten 3 intentos, será: $P(Clave) = 3/10000 = 3{,}0 * 10^{-4}$

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

¿Cuál es la probabilidad de que si pierde la tarjeta alguién ADIVINE su clave?

- Si le permiten 3 intentos, será: $P(Clave) = 3/10000 = 3{,}0 * 10^{-4}$
- Adivinar la pregunta de la cédula: $P(CI) = 10^{-4}$

Principios de Conteo

Ejemplo de Variaciones 2 (Clave de Cajero)

¿Cuál es la probabilidad de que si pierde la tarjeta alguién ADIVINE su clave?

- Si le permiten 3 intentos, será: $P(Clave) = 3/10000 = 3{,}0 * 10^{-4}$
- Adivinar la pregunta de la cédula: $P(CI) = 10^{-4}$
- Si suponemos INDEPENDENCIA:

$$P(Clave \cap CI) = P(Clave) * P(CI) = 3 * 10^{-8}$$

Principios de Conteo

Combinaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento (Sin importar el orden)

Principios de Conteo

Combinaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento (Sin importar el orden)

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

Principios de Conteo

Combinaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento (Sin importar el orden)

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

Si los elementos se pueden repetir se llaman combinaciones con repetición.

Principios de Conteo

Combinaciones. Grupos de m elementos escogidos de los n elementos de un conjunto, teniendo en cuenta que dos grupos son distintos si difieren en algún elemento (Sin importar el orden)

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

Si los elementos se pueden repetir se llaman combinaciones con repetición.

$$CR_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}$$

Principios de Conteo

Ejemplo de Combinaciones (KINO)

KINO. Se toman 15 números de un total de 25, sin importar el orden.

¿Cuántos CARTONES diferentes pueden vender?:

Principios de Conteo

Ejemplo de Combinaciones (KINO)

KINO. Se toman 15 números de un total de 25, sin importar el orden.
¿Cuántos CARTONES diferentes pueden vender?:

$$C_{25}^{15} = \frac{25!}{15!(25-15)!}$$

Principios de Conteo

Ejemplo de Combinaciones (KINO)

KINO. Se toman 15 números de un total de 25, sin importar el orden.

¿Cuántos CARTONES diferentes pueden vender?:

$$C_{25}^{15} = \frac{25!}{15!(25-15)!}$$

TOTAL=3268760

Principios de Conteo

Ejemplo de Combinaciones (KINO)

¿Cuántos CARTONES pueden ganar algo?:

Principios de Conteo

Ejemplo de Combinaciones (KINO)

¿Cuántos CARTONES pueden ganar algo?:

Con 15 $\rightarrow C_{15}^{15} * C_{10}^{0} = 1$

Con 14 $\rightarrow C_{15}^{14} * C_{10}^{1} = 150$

Con 13 $\rightarrow C_{15}^{13} * C_{10}^{2} = 4725$

Con 12 $\rightarrow C_{15}^{12} * C_{10}^{3} = 54600$

Principios de Conteo

Ejemplo de Combinaciones (KINO)

¿Cuántos CARTONES pueden ganar algo?:

Con 15 $\rightarrow C_{15}^{15} * C_{10}^{0} = 1$

Con 14 $\rightarrow C_{15}^{14} * C_{10}^{1} = 150$

Con 13 $\rightarrow C_{15}^{13} * C_{10}^{2} = 4725$

Con 12 $\rightarrow C_{15}^{12} * C_{10}^{3} = 54600$

Total $= 59476$

Principios de Conteo

Ejemplo de Combinaciones (KINO)

¿Cuántos CARTONES pueden ganar algo?:

Con 15 $\rightarrow C_{15}^{15} * C_{10}^{0} = 1$

Con 14 $\rightarrow C_{15}^{14} * C_{10}^{1} = 150$

Con 13 $\rightarrow C_{15}^{13} * C_{10}^{2} = 4725$

Con 12 $\rightarrow C_{15}^{12} * C_{10}^{3} = 54600$

Total $= 59476$

Probabilidad de GANAR algo:

$$\frac{59476}{3268760} = 0{,}01819$$

Menos del 1.82 % GANAN ALGO

Principios de Conteo

Permutaciones. Variaciones cuando $m = n$

Principios de Conteo

Permutaciones. Variaciones cuando $m = n$

$$P_n = V_n^n = n!$$

Principios de Conteo

Permutaciones. Variaciones cuando $m = n$

$$P_n = V_n^n = n!$$

Si los elementos se pueden repetir se llaman permutaciones con repetición, se debe tomar en cuenta que un elemento se puede repetir a veces, otro b veces, etc., siendo $a + b + c \cdots = n$

Principios de Conteo

Permutaciones. Variaciones cuando $m = n$

$$P_n = V_n^n = n!$$

Si los elementos se pueden repetir se llaman permutaciones con repetición, se debe tomar en cuenta que un elemento se puede repetir a veces, otro b veces, etc., siendo $a + b + c \cdots = n$

$$P_n^{a,b,c\cdots k} = \frac{n!}{a!b!\cdots k!}$$

Principios de Conteo

Ejemplo de Permutaciones

El ejemplo "clásico" de permutaciones.
¿Cuántas maneras hay de ordenar los elementos de un conjunto?.

Principios de Conteo

Ejemplo de Permutaciones

El ejemplo “clásico” de permutaciones.
¿Cuántas maneras hay de ordenar los elementos de un conjunto?.

Suponga que tenemos 3 elementos: A, B y C.
Podemos tener los siguientes ordenamientos:

$$\{A,B,C\},\{A,C,B\},\{B,A,C\},\{B,C,A\},\{C,A,B\},\{C,B,A\}$$

Principios de Conteo

Ejemplo de Permutaciones

El ejemplo "clásico" de permutaciones.
¿Cuántas maneras hay de ordenar los elementos de un conjunto?.

Suponga que tenemos 3 elementos: A, B y C.
Podemos tener los siguientes ordenamientos:

$$\{A,B,C\},\{A,C,B\},\{B,A,C\},\{B,C,A\},\{C,A,B\},\{C,B,A\}$$

Probabilidad de que un orden comience por A es:

$$\frac{2}{6} = 1/3$$

Variables Aleatorias

Definición

Es una función que asocia un número real con cada elemento del espacio muestral

Variables Aleatorias

Definición

Es una función que asocia un número real con cada elemento del espacio muestral

Ejemplo

Se lanzan tres monedas, $\Omega = \{CCC, CCS, CSC, SCC, SSC, SSS\}$. Si nos interesa el número de caras (C) que se obtienen, a cada punto de Ω se le asignará 0, 1, 2 o 3. Estos valores son cantidades aleatorias determinadas por el resultado del experimento; éstos son valores que toma la variable aleatoria X.

Tipos de variables aleatorias

Variable aleatoria discreta

Cuando está definida sobre un espacio muestral que tiene un conjunto finito o infinito numerable de posibilidades.

Ejemplo: El presentado anteriormente.

Tipos de variables aleatorias

Variable aleatoria discreta

Cuando está definida sobre un espacio muestral que tiene un conjunto finito o infinito numerable de posibilidades.
Ejemplo: El presentado anteriormente.

Variable aleatoria continua

Cuando está definida sobre un espacio muestral que tiene un conjunto infinito NO numerable de posibilidades.
Ejemplos: Temperaturas, Presiones, Pesos, Posiciones (Variables correspondientes a datos medidos). En muchos casos, los valores posibles de la variable aleatoria continua son los mismos valores que contiene el espacio muestral continuo.

Distribuciones de Probabilidad

Se lanzan TRES monedas ($\Omega = \{CCC, CCS, CSC, CSS, SSS, SSC, SCS, SCC\}$), se define la variable aleatoria X como el número de caras (C) obtenidas:

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Distribuciones de Probabilidad

Se lanzan TRES monedas ($\Omega = \{CCC, CCS, CSC, CSS, SSS, SSC, SCS, SCC\}$), se define la variable aleatoria X como el número de caras (C) obtenidas:

x	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Si es posible representar las probabilidades de una variable aleatoria X mediante alguna fórmula, entonces esa fórmula sería una función de los valores numéricos de x, por ejemplo $f(x)$.

$$f(x) = P(X = x)$$

A $(x, f(x))$ se le llama **función de probabilidad** o **distribución de probabilidad** de la variable aleatoria X.

Distribuciones Discretas de Probabilidad

Definición

El conjunto de pares ordenados $((x, f(x))$ es una función de probabilidad de la variable discreta X si, para cada resultado posible x, se cumple:

1. $f(x) \geq 0$
2. $\sum_{x} f(x) = 1$
3. $P(X = x) = f(x)$

Ejemplos

Distribución Acumulada

Definición

La distribución acumulada $F(x)$ de una variable aleatoria discreta X con distribución de probabilidad $f(x)$ es:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

Distribución Acumulada

Definición

La distribución acumulada $F(x)$ de una variable aleatoria discreta X con distribución de probabilidad $f(x)$ es:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

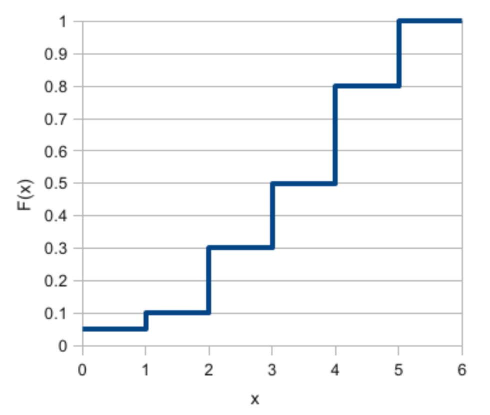
Ejemplo

Si se lanzan TRES monedas
¿Cuál es la probabilidad de obtener a la sumo 1 cara?

$$P(X \leq 1) = f(0) + f(1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$$

Distribución Acumulada Discreta

Gráfica



Distribución uniforme discreta

La variable aleatoria X toma los valores $x_1, x_2, \cdots, x_k$ con idénticas probabilidades.

$$f(x, k) = \frac{1}{k} \quad x = x_1, x_2, \cdots, x_k$$

Ejemplo

Al lanzar un dado cada elemento del espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$ ocurre con probabilidad $1/6$. Por la tanto, se tiene una distribución uniforme:

$$f(x; 6) = \frac{1}{6} \quad x = 1, 2, 3, 4, 5, 6$$

Tendencia en Distribuciones Discretas de Probabilidad

Promedio

- Registra la ubicación central de los datos.
- Es el valor promedio a largo plazo de la variable aleatoria.
- También se le conoce como su valor esperado, E(x), en una distribución de probabilidad.
- Es un promedio ponderado.

Tendencia en Distribuciones Discretas de Probabilidad

Promedio

- Registra la ubicación central de los datos.
- Es el valor promedio a largo plazo de la variable aleatoria.
- También se le conoce como su valor esperado, E(x), en una distribución de probabilidad.
- Es un promedio ponderado.

Distribución Uniforme Discreta

$$\mu = \sum_{i=1}^{n} x_i * f(x_i)$$

Para un dado: $\mu = (1 + 2 + 3 + 4 + 5 + 6)/6 = 21/6 = 3{,}5$

Dispersión en Distribuciones Discretas de Probabilidad

Varianza

- Mide el tamaño de la dispersión de una distribución.
- La varianza de una distribución discreta es representada por σ^2
- La desviación estándar es $\sigma = \sqrt{\sigma^2}$

Dispersión en Distribuciones Discretas de Probabilidad

Varianza

- Mide el tamaño de la dispersión de una distribución.
- La varianza de una distribución discreta es representada por σ^2
- La desviación estándar es $\sigma = \sqrt{\sigma^2}$

Distribución Uniforme Discreta

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 * f(x_i)$$

Para un dado:

$\sigma^2 = \frac{1}{6}((-2{,}5)^2 + (-1{,}5)^2 + (-0{,}5)^2 + 0{,}5^2 + 1{,}5^2 + 2{,}5^2) = \frac{17{,}5}{6} \approx 2{,}92$

Proceso de Bernoulli

Características

- El experimento consiste de n pruebas que se repiten
- Cada prueba produce un resultado que se puede clasificar como **éxito** o **fracaso**
- La probabilidad de un éxito, p, permanece constante en cada prueba
- Las pruebas que se repiten son independientes

Distribución Binomial (Proceso de Bernoulli)

Si se tiene un experimento que tiene como resultado un éxito con probabilidad p y un fracaso con probabilidad $q = 1 - p$ (**Experimento de Bernoulli**), la distribución de probabilidad, el número de éxitos en n pruebas independientes, de la variable aleatoria X es una Distribución Binomial:

$$b(x; n; p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \cdots, n$$

Ejemplo de Distribución Binomial

Un jugador de baloncesto tiene un porcentaje de acierto del 75 % desde la línea de "tiros libres", si en un partido ejecuta 15 lanzanmientos de este tipo,

- ¿Cuál es la probabilidad de que acierte 10?
- ¿Cuál es la probabilidad que falle TODOS?
- ¿Cuál es la probabilidad que acierte TODOS?
- ¿Cuál es la probabilidad de que al menos acierte 10?

Medidas en una Distribuición de Binomial

Promedio

$$\mu = n * p$$

En el ejemplo: $\mu = 15 * 0{,}75 = 11{,}25$

Varianza

$$\sigma^2 = n * p * (1 - p)$$

En el ejemplo: $\sigma^2 = 15 * 0{,}75 * 0{,}25 = 2{,}8125$

Distribución Multinomial

Si en un experimento se pueden producir k resultados diferentes $R_1, R_2, \cdots, R_k$ con probabilidades $p_1, p_2, \cdots, p_k$, entonces la distribución de probabilidad de las variables aleatorias $X_1, X_2, \cdots, X_k$, que representan el número de ocurrencias para $R_1, R_2, \cdots, R_k$ en n pruebas independientes es:

$$f(x_1, x_2, \cdots, x_k; p_1, p_2, \cdots, p_k; n) = \binom{n}{x_1, x_2, \cdots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

con

$$\sum_{i=1}^{k} x_i = n \quad \text{y} \quad \sum_{i=1}^{k} p_i = 1$$

Distribución multinomial

Ejemplo

Lanzar un par de dados 6 veces, calcular la probabilidad que:

- La suma de DOS (2) veces 7 u 11 (R_1). $p_1 = 8/36 = 2/9$
- Se obtiene UN (1) par igual (R_2). $p_2 = 6/36 = 1/6$
- Cualquier otro resultado TRES (3) veces (R_3).$p_3 = 22/36 = 11/18$

Distribución multinomial

Ejemplo

Lanzar un par de dados 6 veces, calcular la probabilidad que:

- La suma de DOS (2) veces 7 u 11 (R_1). $p_1 = 8/36 = 2/9$
- Se obtiene UN (1) par igual (R_2). $p_2 = 6/36 = 1/6$
- Cualquier otro resultado TRES (3) veces (R_3).$p_3 = 22/36 = 11/18$

Fórmula

$$P = \binom{6}{2,1,3}\left(\frac{2}{9}\right)^2\left(\frac{1}{6}\right)\left(\frac{11}{18}\right)^3 = 0{,}1127$$

Distribución Hipergeométrica

- Hay dos resultados posibles
- La probabilidad de un éxito no es la misma en cada ensayo.
- Para obtener la probabilifdad se cuenta el número de éxitos en un número fijo de ensayos.

Distribución Hipergeométrica

- Hay dos resultados posibles
- La probabilidad de un éxito no es la misma en cada ensayo.
- Para obtener la probabilifdad se cuenta el número de éxitos en un número fijo de ensayos.

$$P(x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$$

N: Tamaño de la Población
S: Número de éxitos en la Población
n: Tamaño de la muestra
x: Número de éxitos en la muestra

Distribución Hipergeométrica

Ejemplo

En un edificio de 40 aptos. el 80 % son propietarios, si se elige una junta de condominio de 5 miembros; ¿Cuál es la probabilidad de que hayan 4 propietarios en la junta?.

Distribución Hipergeométrica

Ejemplo

En un edificio de 40 aptos. el 80 % son propietarios, si se elige una junta de condominio de 5 miembros; ¿Cuál es la probabilidad de que hayan 4 propietarios en la junta?.

N=40, S=32, n=5 y x=4

$$P(x=4)=\frac{\binom{32}{4}\binom{8}{1}}{\binom{40}{5}}=0{,}4372$$

Distribución de Poisson

- Describe la cantidad de veces que ocurre un evento en un intervalo determinado.
- Es una forma límite de la distribución binomial, cuando la probabilidad de éxito es muy pequeña y n es grande.

Distribución de Poisson

- Describe la cantidad de veces que ocurre un evento en un intervalo determinado.
- Es una forma límite de la distribución binomial, cuando la probabilidad de éxito es muy pequeña y n es grande.

Fórmula

$$P(x; \lambda t) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

λ es el número promedio de resultados por unidad de tiempo.

Distribución de Poisson

Ejemplo

La probabilidad de que un "paracaidas" falle es de 0.001, si se prueban 1000 paracaidas,
¿Cuál es la probabilidad que fallen 2?

Distribución de Poisson

Ejemplo

La probabilidad de que un "paracaidas" falle es de 0.001, si se prueban 1000 paracaidas,
¿Cuál es la probabilidad que fallen 2?
$\lambda t \approx \mu = n * p = 0{,}001 * 1000 = 1$
$P(2) = \frac{1^2 e^{-1}}{2!} = 0{,}1839$

Distribuciones de Probabilidad Conjunta (Discretas)

La función $f(x,y)$ es una **distribución de probabilidad conjunta** de las variables aleatorias **discretas** X y Y si:

1. $f(x,y) \geq 0 \quad \forall (x,y)$
2. $\displaystyle\sum_x \sum_y f(x,y) = 1$
3. $P(X = x, Y = y) = f(x,y)$

Dada una región A en el plano xy:

$$P[(X,Y) \in A] = \sum_x \sum_y f(x,y) \quad \forall (x,y) \in A$$

Distribución de Probabilidad Conjunta

Ejemplo

Para iniciar un juego de Baloncesto, un entrenador dispone de 5 jugadores buenos, 4 regulares y 3 malos; Si X es el número de jugadores buenos en el quinteto abridor y Y el número de jugadores regulares.

¿Cuál es la probabilidad de tener 4 buenos y 1 regular?

Distribución de Probabilidad Conjunta

Ejemplo

Para iniciar un juego de Baloncesto, un entrenador dispone de 5 jugadores buenos, 4 regulares y 3 malos; Si X es el número de jugadores buenos en el quinteto abridor y Y el número de jugadores regulares.

¿Cuál es la probabilidad de tener 4 buenos y 1 regular?

$$f(4,1) = \frac{\binom{5}{4}\binom{4}{1}\binom{3}{0}}{\binom{12}{5}} = 0{,}0253$$

Distribuciones Continuas de Probabilidad

- Como X es continua, $P(X = x) = 0$
- Está definido en los casos: $P(a \leq X \leq b)$ y $P(X < b)$
- $f(x)$ en el caso continuo no representa $P(X = x)$. En este caso se llama **función de densidad de probabilidad**
- Como X está definida en un espacio continuo, $f(x)$ podría tener discontinuidades; aunque en general se trabaja con funciones continuas.

Función de densidad de probabilidad

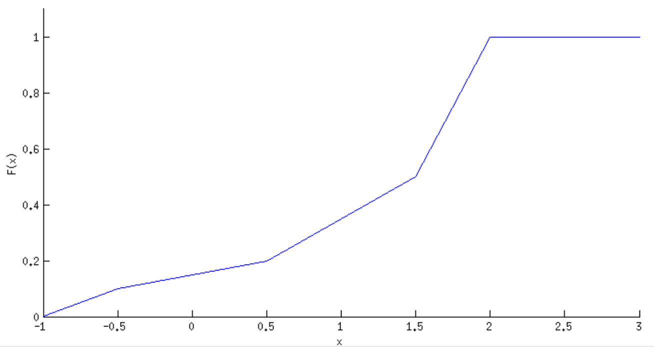
Definición

$f(x)$ es una función de densidad para la variable aleatoria continua X, definida en el conjunto de los números reales, si:

1. $$f(x) \geq 0 \quad \forall x \in R$$
2. $$\int_{-\infty}^{\infty} f(x)dx = 1$$
3. $$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Distribución Acumulada Continua

Gráfica



Distribución Uniforme Continua

La función de densidad de la variable uniforme continua en X en el intervalo $[A, B]$ es:

$$\begin{aligned} f(x; A, B) &= \frac{1}{B - A}, \quad A \leq x \leq B \\ &= 0 \qquad\qquad \text{en cualquier otro caso} \end{aligned}$$

La media y la varianza de la distribución uniforme son:

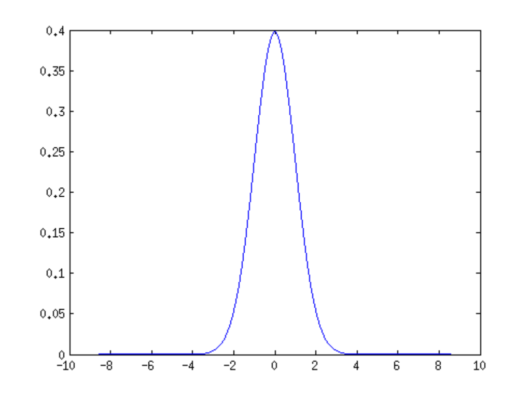
$$\mu = \frac{A + B}{2} \quad \text{y} \quad \sigma^2 = \frac{(B - A)^2}{12}$$

Distribución Normal

- La distribución normal es una curva acampanada y presenta sólo un máximo en el centro.
- La media aritmética, la mediana y la moda de la distribución son iguales y están localizadas en el valor máximo (pico).
- Es simétrica con respecto a su media.
- Decrece en ambas direcciones a partir del valor central (Asintótica).

Distribución Normal

Gráfica



Distribución Normal

La función de densidad de la variable aleatoria normal X, con media μ y varianza σ^2, es:

$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(1/2)[(x-\mu)/\sigma]^2}, \quad -\infty < x < \infty$$

La distribución de una variable aleatoria normal con media cero (0) y varianza uno (1) se llama **distribución normal estándar**

Uso de la distribución normal estándar

Ejemplo

Una empresa fabrica "focos" que tienen una duración que se distribuye normalmente con una media de 800 horas y desviación estándar de 40 horas. Encuentre la probabilidad de un foco dure entre 778 y 834 horas.

Distribución Gamma

La función gamma se define como:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \;\; \forall \alpha > 0$$

La variable aleatoria continua X tiene una distribución gamma, con parámetros α y β, si su función de densidad está dada por:

$$\begin{aligned} f(x) &= \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad && x > 0 \\ &= 0 && \text{en cualquier otro caso} \end{aligned}$$

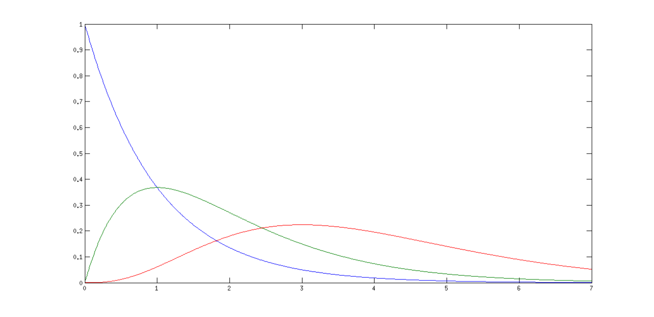
cuando $\alpha > 0$ y $\beta > 0$.

Distribución Gamma

Media y Varianza

$$\mu = \alpha\beta \quad \sigma^2 = \alpha\beta^2$$

Gráfica



Distribución Exponencial

Caso especial de un Gamma cuando $\alpha = 1$

La variable aleatoria continua X tiene una **distribución exponencial**, con parámetro β, si su función de densidad está dada por:

$$\begin{aligned} f(x) &= \frac{1}{\beta} e^{-x/\beta}, \quad x > 0 \\ &= 0 \qquad\qquad \text{en cualquier otro caso} \end{aligned}$$

donde $\beta > 0$.

Media y Varianza

$$\mu = \beta \quad \sigma^2 = \beta^2$$

Distribución Exponencial

Ejemplo

Un componente electrónico tiene un tiempo de vida T. El tiempo promedio de vida del componente es 5 años. Si se instalan cinco componentes de este tipo. ¿Cuál es la probabilidad de que al menos DOS (2) funcionen 8 o más años?

Distribución Exponencial

Ejemplo

Un componente electrónico tiene un tiempo de vida T. El tiempo promedio de vida del componente es 5 años. Si se instalan cinco componentes de este tipo. ¿Cuál es la probabilidad de que al menos DOS (2) funcionen 8 o más años?

Tomando un distribución exponencial con $\beta = 5$:

$$P(T > 8) = \frac{1}{5}\int_8^{\infty} e{-}t/5dt = e^{-8/5} = 0{,}2019$$

Ahora X es el número de componentes que funcionan después de 8 años. Entonces se usa una distribución binomial:

$$P(X \geq 2) = \sum_{x=2}^{5} b(x;5,p) = 1 - \sum_{x=0}^{1} b(x;5,p) = 1 - 0{,}7334 = 0{,}2666$$

Distribución χ^2

Caso especial de gamma cuando $\alpha = \nu/2$ y $\beta = 2$

La variable aleatoria continua X tiene una distribución χ^2, con ν grados de libertad, si su función de densidad está dada por:

$$
\begin{aligned}
f(x) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad && x > 0 \\
&= 0 && \text{en cualquier otro caso}
\end{aligned}
$$

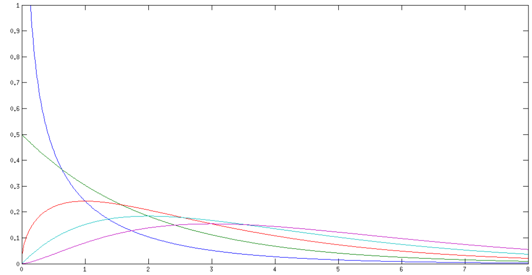
donde ν es un entero positivo.

Media y Varianza

$$\mu = \nu \quad \sigma^2 = 2\nu$$

Distribución Gamma

Gráfica



Distribución Logarítmica Normal

La variable aleatoria continua x tiene una distribución logarotmica normal si la variable aleatoria $Y = ln(X)$ tiene una distribución normal con media μ y desviación *sigma*. La función de densidad de X es:

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma x} e^{-[ln(x)-\mu]^2/(2\sigma^2)}, & x \geq 0 \\ &= 0 & x < 0 \end{aligned}$$

Media y Varianza

$$E(X) = e^{\mu+\sigma^2/2} \quad Var(X) = e^{2\mu+\sigma^2}.(e^{\sigma^2} - 1)$$

Distribución Logarítmica Normal

Ejemplo

Se supone que la concentración de un cierta sustancia en un bebida comercial, en partes por millón, tiene una distribución logarítmica normal con parámetros $\mu = 3{,}2$ y $\sigma = 1$. ¿Cual es la probabilidad de que en una bebida la sustancia excede las 8 partes por millón?

Distribución Logarítmica Normal

Ejemplo

Se supone que la concentración de un cierta sustancia en un bebida comercial, en partes por millón, tiene una distribución logarítmica normal con parámetros $\mu = 3{,}2$ y $\sigma = 1$. ¿Cual es la probabilidad de que en una bebida la sustancia excede las 8 partes por millón?

$$P(X > 8) = 1 - P(X \leq 8)$$

Como $In(X)$ se distribuye según una normal se calcula $\dfrac{In(8) - 3{,}2}{1} = -1{,}12$. Con este valor se busca en la tabla de distribución normal acumulada.

Distribución de Weibull

Propuesta por el físico sueco Waloddi Weibull (1939). La variable aleatoria continua X tiene una **distribución de Weibull**, con parámetros α y β si su función de densidad está dada por:

$$f(x) = \alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}, \quad x > 0$$

$$= 0 \qquad \text{en cualquier otro caso}$$

donde $\alpha > 0$ y $\beta > 0$.

Media y Varianza

$$\mu = \alpha^{-1/\beta}\Gamma\left(1+\frac{1}{\beta}\right), \quad \sigma^2 = \alpha^{-2/\beta}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma\left(1+\frac{1}{\beta}\right)\right\}$$

Introducción a la inferencia estadística

Métodos por lo que se realizan inferencias o generalizaciones acerca de una población.

Introducción a la inferencia estadística

Métodos por lo que se realizan inferencias o generalizaciones acerca de una población.

- **Método Clásico**. Se basan en la información que se obtiene de un muestra aleatoria.

Introducción a la inferencia estadística

Métodos por lo que se realizan inferencias o generalizaciones acerca de una población.

- **Método Clásico**. Se basan en la información que se obtiene de un muestra aleatoria.
- **Método Bayesiano**. Además de los datos obtenidos de la muestra, usa conocimiento subjetivo previo sobre la distribución de probabilidad.

Introducción a la inferencia estadística

Areas

Introducción a la inferencia estadística

Areas

- **Estimación**. Sobre la base de datos tomados en una muestra, se estima el valor de algún parámetro de la(s) población (Ejemplo: Encuesta)

Introducción a la inferencia estadística

Areas

- **Estimación**. Sobre la base de datos tomados en una muestra, se estima el valor de algún parámetro de la(s) población (Ejemplo: Encuesta)
- **Pruebas de hipótesis**. Se plantea una hipótesis sobre algún(os) parámetro(s) de la población(os) y se hace una prueba tomando una (algunas) muestras y se acepta o rechaza la hipótesis. (Ejemplo: medición de dos productos)

Estimación Puntual

Una **estimación puntual** de algún parámetro de la población θ es un solo valor $\hat{\theta}$ de un estadístico $\hat{\Theta}$.

Estimación Puntual

Una **estimación puntual** de algún parámetro de la población θ es un solo valor $\hat{\theta}$ de un estadístico $\hat{\Theta}$.

Ejemplo

El valor $\bar{x}$ del estadístico $\bar{X}$, que se calcula con una muestra de tamaño n, es una estimación puntual del parámetro poblacional μ.

Estimación Puntual

Una **estimación puntual** de algún parámetro de la población θ es un solo valor $\hat{\theta}$ de un estadístico $\hat{\Theta}$.

Ejemplo

El valor $\bar{x}$ del estadístico $\bar{X}$, que se calcula con una muestra de tamaño n, es una estimación puntual del parámetro poblacional μ.

Observación

En general, el estimador permite realizar la estimación del parámetro poblacional con un error diferente de cero (0). En algunas casos se pueden usar como para estimar un parámetro poblacional diferentes valores calculados de la muestra (Ejemplo: Mediana y Media para estimar μ)

Estimador Puntual insesgado

Un estimador es **insesgado** si la distribución muestral tiene una media igual al parámetro estimado. Así, el estadístico $\hat{\Theta}$ es un estimador **insesgado** del parámetro θ si:

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta$$

Ejemplo

S^2 es un estimador insesgado de la varianza (σ^2).

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\right]$$

Sin embargo, S es un estimador sesgado de desviación estándar (σ)

Varianza de un estimador puntual

Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores insesgados del mismo parámetro θ, se usa el estimador que tiene la menor varianza. Si $\sigma^2_{\hat{\Theta}_1} < \sigma^2_{\hat{\Theta}_2}$, se dice que $\hat{\Theta}_1$ es un estimador más **eficiente** de θ que $\hat{\Theta}_2$.

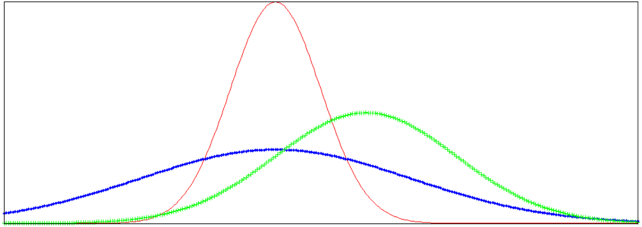
Varianza de un estimador puntual

Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores insesgados del mismo parámetro θ, se usa el estimador que tiene la menor varianza. Si $\sigma^2_{\hat{\Theta}_1} < \sigma^2_{\hat{\Theta}_2}$, se dice que $\hat{\Theta}_1$ es un estimador más **eficiente** de θ que $\hat{\Theta}_2$.

Definición

Si se consideran "todos" los posibles estimadores insesgados de algún parámetro θ, el de menor varianza se denomina **estimador más eficiente** de θ.

Distribuciones Muestrales de estimadores de θ



Estimación por intervalo

La estimación por intervalo es un intervalo con un límite inferior y un límite superior, que dependen del valor del estadístico $\hat{\Theta}$, donde se estima que esté el parámetro poblacional θ.

$$\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}$$

Estimación por intervalo

La estimación por intervalo es un intervalo con un límite inferior y un límite superior, que dependen del valor del estadístico $\hat{\Theta}$, donde se estima que esté el parámetro poblacional θ.

$$\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}$$

Dadas las variables aleatorias $\hat{\Theta}_{inf}$ y $\hat{\Theta}_{sup}$, si encontramos $\hat{\theta}_{inf}$ y $\hat{\theta}_{sup}$ tal que: $P(\hat{\Theta}_{inf} < \theta < \hat{\Theta}_{sup}) = 1 - \alpha$
Para $0 < \alpha < 1$, se tiene una probabilidad de $1 - \alpha$ de seleccionar una variable aleatoria que produzca un intervalo que contenga a θ.

Estimación por intervalo

La estimación por intervalo es un intervalo con un límite inferior y un límite superior, que dependen del valor del estadístico $\hat{\Theta}$, donde se estima que esté el parámetro poblacional θ.

$$\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}$$

Dadas las variables aleatorias $\hat{\Theta}_{inf}$ y $\hat{\Theta}_{sup}$, si encontramos $\hat{\theta}_{inf}$ y $\hat{\theta}_{sup}$ tal que: $P(\hat{\Theta}_{inf} < \theta < \hat{\Theta}_{sup}) = 1 - \alpha$
Para $0 < \alpha < 1$, se tiene una probabilidad de $1 - \alpha$ de seleccionar una variable aleatoria que produzca un intervalo que contenga a θ.

$\hat{\theta}_{inf} < \theta < \hat{\theta}_{sup}$ es el **intervalo de confianza** de $(1 - \alpha)100\,\%$.
$(1 - \alpha)$ es el **coeficiente de confianza**, y los extremos del

Estimación de la media

Como $\sigma^2_{\bar{X}} = \sigma^2/n$, la varianza de $\bar{X}$ disminuye cuando aumenta n. Entonces, $\bar{x}$ es una **estimación precisa** de μ cuando n es grande.

Estimación de la media

Como $\sigma^2_{\bar{X}} = \sigma^2/n$, la varianza de $\bar{X}$ disminuye cuando aumenta n. Entonces, $\bar{x}$ es una **estimación precisa** de μ cuando n es grande.

Por el teorema del límite central, la distribución muestral de $\bar{X}$ está distribuida de forma aproximadamente normal con $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Estimación de la media

Como $\sigma^2_{\bar{X}} = \sigma^2/n$, la varianza de $\bar{X}$ disminuye cuando aumenta n. Entonces, $\bar{x}$ es una **estimación precisa** de μ cuando n es grande.

Por el teorema del límite central, la distribución muestral de $\bar{X}$ está distribuida de forma aproximadamente normal con $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Si $\bar{x}$ es la media de una muestra aleatoria de tamaño n de una población con varianza σ^2, conocida, un intervalo de confianza de $(1-\alpha)100\,\%$ para μ está dado por:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Error de Estimación para la media

Si se utiliza como estimación para μ a $\bar{x}$, se puede tener una cofianza de $(1-\alpha)100\,\%$ de que el error no será mayor a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Error de Estimación para la media

Si se utiliza como estimación para μ a $\bar{x}$, se puede tener una cofianza de $(1-\alpha)100\,\%$ de que el error no será mayor a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Tamaño de n para limitar el error

Si se utiliza como estimación para μ a $\bar{x}$, se puede tener $(1-\alpha)100\,\%$ de confianza de que el error no será mayor que una cantidad específica e cuando el tamaño de la muestra es:

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

Estimación μ cuando σ^2 es desconocida

Dos casos

Estimación μ cuando σ^2 es desconocida

Dos casos

1. **Muestras pequeñas ($n < 30$)**. Se define una variable aleatoria $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$. Esta variable tiene una distribución t de Student con $n-1$ grados de libertad. Entonces, el intervalo de confianza para μ es:
$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$
donde $t_{\alpha/2}$ es el valor de t con $\nu = n-1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación μ cuando σ^2 es desconocida

Dos casos

1. **Muestras pequeñas ($n < 30$)**. Se define una variable aleatoria $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$. Esta variable tiene una distribución t de Student con $n - 1$ grados de libertad. Entonces, el intervalo de confianza para μ es:
$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$
donde $t_{\alpha/2}$ es el valor de t con $\nu = n - 1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.
2. **Muestras grandes ($n > 30$)**. Cuando n es grande T se distribuye aproximadamente como una normal estándar, es decir, como Z, y se puede usar el intervalo de confianza para z sustituyendo σ por s.

Estimación de un proporción

Si $\hat{p}$ es la proporción de éxitos en una muestra aleatoria de tamaño n, y $\hat{q} = 1 - \hat{p}$, un intervalo de confianza aproximado de $(1-\alpha)100\,\%$ para el parámetro binomial p está dado por:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

donde $z_{\alpha/2}$ es el valor de z que deja un área de $\alpha/2$ a la derecha.

Estimación de la diferencia entre dos medias

Varianzas conocidas

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones con varianzas conocidas σ_1^2 y σ_2^2, respectivamente, un intervalo de confianza de $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ está dado por:

$$(\bar{x}_1 \text{ - } \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 \text{ - } \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

donde $z_{\alpha/2}$ es el valor z que deja un área de $\alpha/2$ a la derecha.

Estimación de la diferencia entre dos medias

Varianzas desconocidas iguales

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de muestras aleatorias independientes con tamaño n_1 y n_2, respectivamente, de poblaciones aproximadamente normales con varianzas iguales pero desconocidas, un intervalo de confianza de $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ está dado por:

$$(\bar{x}_1 \text{ - } \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 \text{ - } \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde s_p es la estimación de unión de la desviación estándar poblacional:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

y $t_{\alpha/2}$ es el valor de t con $\nu = n_1 + n_2 - 2$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la diferencia entre dos medias

Varianzas desconocidas y diferentes

Si $\bar{x}_1$ y s_1^2, y $\bar{x}_2$ y s_2^2 son las medias y varianzas de muestras aleatorias pequeñas independientes con tamaño n_1 y n_2, respectivamente, de distribuciones aproximadamente normales con varianzas desconocidas y diferentes, un intervalo de confianza de $(1-\alpha)100\,\%$ para $\mu_1 - \mu_2$ está dado por:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

donde $t_{\alpha/2}$ es el valor de t con:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)\right]}$$

grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la diferencia entre dos medias

Muestras Pareadas (No independientes)

Si $\bar{d}$ y s_d son la media y la desviación estándar de las diferencias distribuidas normalmente de n pares aleatorios de mediciones, un intervalo de confianza de $(1-\alpha)100\,\%$ para $\mu_D = \mu_1 - \mu_2$ está dado por:

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

donde $t_{\alpha/2}$ es el valor con $\nu = n-1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha.

Estimación de la Varianza

Una muestra

Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza de $(1-\alpha)100\,\%$ para σ^2 es:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores de χ^2 con $\nu = n-1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1-\alpha/2$, respectivamente, a la derecha.

Estimación de la razón de dos varianzas

Si s_1^2 y s_2^2 son las varianzas de muestras independientes de tamaño n_1 y n_2, respectivamente, de poblaciones normales, entonces un intervalo de confianza de $(1-\alpha)100\,\%$ para σ_1^2/σ_2^2 es:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(\nu_1,\nu_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(\nu_2,\nu_1)$$

donde $f_{\alpha/2}(\nu_1,\nu_2)$ es un valor de f con $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$ grados de libertad, que deja un área de $\alpha/2$ a la derecha de $f_{\alpha/2}(\nu_1,\nu_2)$ y $1-\alpha/2$, respectivamente, a la derecha; $f_{\alpha/2}(\nu_2,\nu_1)$ es un valor similar con $\nu_2 = n_2 - 1$ y $\nu_1 = n_1 - 1$ grados de libertad.

Prueba de Hipótesis

Hipótesis Estadística

Es una declaración o conjetura con respecto a una o varias poblaciones que involucra el valor de algún parámetro de éstas.

Ejemplos

- El 20 % de los rayos lanzados en el experimento "pegan" fuera del valor umbral.
- El número de particulas en el sistema es 25000.

Prueba de Hipótesis

Definición

La **prueba de hipótesis** es un procedimiento basado en la evidencia de la muestra y la teoría de las probabilidades, que se usan para determinar si la hipótesis es una declaración razonable y no debe ser rechazada, o es irrazonable y debe ser rechazada.

Prueba de Hipótesis

Definición

La **prueba de hipótesis** es un procedimiento basado en la evidencia de la muestra y la teoría de las probabilidades, que se usan para determinar si la hipótesis es una declaración razonable y no debe ser rechazada, o es irrazonable y debe ser rechazada.

La aceptación de un hipótesis simplemente implica que la información no es SUFICIENTE para rechazarla. Mientras, el rechazo implica que hay una probabilidad muy pequeña de obtener la información muestral observada cuando, de hecho, la hipótesis es verdadera.

Prueba de Hipótesis

Definiciones

- **Hipótesis nula** H_0: Una declaración sobre el valor de un parámetro de la población.
- **Hipótesis alternativa** H_1: Una declaración que se acepta si los datos de la muestra proporcionan evidencia de que la hipótesis nula es falsa.
- **Nivel de significancia**: La probabilidad de rechazar la hipótesis nula cuando es verdadera.

Prueba de Hipótesis

Definiciones

- **Error tipo I**: Rechazar de la hipótesis nula cuando es verdadera.
- **Error tipo II**: Aceptar de la hipótesis nula cuando es falsa.
- **Estadístico de prueba**: Un valor obtenido a partir de la información muestral, que se usa para determinar si se rechaza la hipótesis nula.
- **Valor crítico**: Punto de división entre la región en la que se rechaza la hipótesis nula y la región en la que se acepta la hipótesis nula.

Prueba de hipótesis

Esquema de funcionamiento

1. Se plantean las hipótesis nula y la alternativa.
2. Se selecciona el nivel de significancia.
3. Se identifica el estadístico de prueba.
4. Se formula la regla de decisión.
5. Se toma una muestra y se decide: se acepta o se rechaza H_0.

Prueba de hipótesis de una cola

Definición

Una prueba es de una cola cuando la hipótesis alternativa, H_1 indica una dirección:

$$H_0 : \theta = \theta_0 \qquad\qquad H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0 \quad \text{o quizá} \quad H_1 : \theta < \theta_0$$

Prueba de hipótesis de una cola

Definición

Una prueba es de una cola cuando la hipótesis alternativa, H_1 indica una dirección:

$$H_0 : \theta = \theta_0 \qquad\qquad H_0 : \theta = \theta_0$$

$$H_1 : \theta > \theta_0 \quad \text{o quizá} \quad H_1 : \theta < \theta_0$$

Ejemplos

- H_1: El número de rayos que pegan fuera del umbral es menor a 100 ($\mu < 100$)
- H_1: El número de usuarios del comedor universitario es mayor a 4000 ($\mu > 4000$)

Prueba de hipótesis de dos colas

Definición

Una prueba es con dos colas cuando no se especifica ninguna dirección en la hipótesis alternativa H_1:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

Prueba de hipótesis de dos colas

Definición

Una prueba es con dos colas cuando no se especifica ninguna dirección en la hipótesis alternativa H_1:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0 \end{aligned}$$

Ejemplos

- H_1: El número de rayos que pegan fuera del umbral es distinto de 100 ($\mu \pm 100$)
- H_1: Los alumnos por sección en la facultad de ciencias no es igual a 7 ($\mu \pm 7$)

Prueba de hipótesis para medias

Desviación estándar conocida

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Prueba de hipótesis para medias

Desviación estándar conocida

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Desviación estándar desconocida, *n* grande

$$z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Prueba de hipótesis para medias

Desviación estándar conocida

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Desviación estándar desconocida, n grande

$$z = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Desviación estándar desconocida, n pequeño

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Prueba para la proporción de la población

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Donde:

π: Es la proporción de la población

p: Es la proporción en la muestra

Ejemplo para una muestra

En fabricante de un producto indica que el mismo contiene 16 onzas. La desviación estándar del proceso es 0.5 onza. Una muestra de 36 botellas de reveló un peso promedio de 16.12 onzas por botella. ¿En un nivel de significancia del .05 el proceso está fuera de control? ¿Es decir, podemos concluir que la cantidad por botella es diferente a 16 onzas?

$$\mu = 16$$

$$\sigma = 0{,}5$$

$$\bar{x} = 16{,}12$$

$$n = 36$$

Prueba de Hipótesis para el ejemplo

1. $H_0 : \mu = 16 \quad H_1 : \mu \neq 16$
2. Nivel de significancia: 0.05
3. Estadístico de prueba: z (Desviación estándar conocida)
4. Regla de decisión: Rechazar H_0 si $z > 1{,}96$ ó $z < -1{,}96$
5. Cálculo de z:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{16{,}12 - 16}{0{,}5/\sqrt{36}} = 1{,}44$$

Conclusión: No se rechaza la hipótesis nula (No se puede decir que la media es diferente a 16)

El valor P

Un valor P es el nivel (de significancia) más bajo en el que el valor observado del estadístico de prueba es significativo

Ejemplo

$H_0 : \mu = 10$, $H_1 : \mu \neq 10$.
Si se elige un $\alpha = 0{,}05$, para una distribución normal estándar, obtenemos una región crítica $z > 1{,}96$, $z < -1{,}96$,
Con el valor de z se puede hacer el planteamiento: "El valor del estadístico de prueba es significativo", para el ejemplo planteado: "La media difiere significativamente del valor 10".
Si se obtiene un valor del estadístico tal que $z = 1{,}87$, el valor no es significativo. Pero el riesgo del error tipo I si se rechaza H_0 no es severo.

Elección del tamaño de la muestra

Existen 3 factores que determinan el tamaño de la muestra, ninguno de los cuales tiene relación con el tamaño de la población. Éstos son:

- El nivel de confianza deseado.
- El máximo error permisible.
- La variación en la población.

Elección del tamaño de la muestra

Existen 3 factores que determinan el tamaño de la muestra, ninguno de los cuales tiene relación con el tamaño de la población. Éstos son:

- El nivel de confianza deseado.
- El máximo error permisible.
- La variación en la población.

$$n = \left(\frac{z * s}{E}\right)^2$$

Elección del tamaño de la muestra

Caso de proporciones

$$n = p(1-p)\left(\frac{z}{E}\right)^2$$

Donde:

- p: es la proporción estimada, por experiencia o tmada de una prueba piloto.
- z: el valor z asociado al grado de confianza seleccionado.
- E: el error máximo permisible

Pruebas de hipótesis para dos muestras

Poblaciones independientes y varianzas conocidas

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Pruebas de hipótesis para dos muestras

Poblaciones independientes y varianzas conocidas

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Poblaciones independientes y varianzas desconocidas (iguales)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Donde:

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Pruebas de hipótesis para dos muestras

Poblaciones independientes y varianzas desconocidas (diferentes)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

donde t tiene una distribución *student* con grados de libertad aproximados:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)\right]}$$

Pruebas de hipótesis para dos muestras

Muestras Pareadas

$$t = \frac{\bar{d} - d_0}{s_d / \sqrt{n}}$$

Donde:

$\bar{d}$: es la media de las diferencias en la muestra

d_0: es la media de las diferencias en la población

s_d: es la desviación estándar de las diferencias en la muestra

n: es el tamaño de la muestra (número de pares)

Prueba de hipótesis sobre varianza

Una muestra

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

donde:

n: Tamaño de la muestra

s^2: Varianza muestral

σ_0^2: Valor de σ^2 dado para la hipótesis nula.

Si H_0 es verdadera, χ^2 en un valor de la distribución ji cuadrada con $n-1$ grados de libertad. Para una prueba de dos colas en el nivel de significancia α, la región crítica es $\chi^2 < \chi^2_{1-\alpha/2}$ y $\chi^2 > \chi^2_{\alpha/2}$. Si $\sigma^2 < \sigma_0^2$, la región crítica es $\chi^2 < \chi^2_{1-\alpha}$, y para $\sigma^2 > \sigma_0^2$, la región crítica es $\chi^2 > \chi^2_{\alpha}$.

Prueba de hipótesis sobre varianza

Dos muestras

$$f = \frac{s_1^2}{s_2^2}$$

s_1^2 y s_2^2 son las varianzas calculadas de las dos muestras. Si las dos poblaciones son aproximadamente normales y la hipótesis nula es verdadera, la razón $f = s_1^2/s_2^2$ es un valor de la distribución F con $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$ grados de libertad. Por tanto, las regiones críticas de tamaño α que corresponden a las alternativas unilaterales $\sigma_1^2 < \sigma_2^2$ y $\sigma_1^2 > \sigma_2^2$ son, respectivamente, $f < f_{1-\alpha}(\nu_1, \nu_2)$ y $f > f_\alpha(\nu_1, \nu_2)$. Para $\sigma_1^2 \neq \sigma_2^2$, la región crítica es $f < f_{1-\alpha/2}(\nu_1, \nu_2)$ y $f > f_{\alpha/2}(\nu_1, \nu_2)$.

Regresión

Elementos de regresión

- Gráfica de datos (Diagrama de dispersión): Representación gráfica de la relación entre dos variables.
- Análisis de correlación: Técnicas estadísticas para medir el nivel de asociación entre dos variables.
- En cualquier estudio, se puede suponer que:
$$\exists f | f(x) = y$$
Donde:
  - $x \in \Re^n$: variables de regresión (variables independientes). Estas pueden ser medidas, en cuyo caso no son aleatorias, o pueden tener alguna distribución aleatoria.
  - $y \in \Re$: variable dependiente (respuesta)

Regresión

Caso General

Se puede definir una ecuación o una función que aproxime a f, $\hat{f}(x) \approx f$ que se denomina **ecuación de regresión** o **función de regresión**.

Regresión Lineal Simple

$x \in \Re$; $y \in \Re$.
$\{(x_i, y_i); i = 1, 2, \cdots, n\}$ es una muestra aleatoria de tamaño n.
Se define $Y|x$ como la variable Y que corresponde a un valor fijo de x con media $\mu_{Y|x}$ y varianza $\sigma^2_{Y|x}$.
$\mu_{Y|x} = \alpha + \beta x$, es la **ecuación de regresión de población**, donde los **coeficientes de regresión** α y β son los parámetros a estimar a partir de los datos muestrales.

Regresión Lineal simple

$\mu_{Y|x}$ puede ser estimada con $\hat{y}$ a partir de la regresión de la muestra:

$$\hat{y} = a + bx$$

Donde a y b son las estimaciones para la intersección y la pendiente de la recta.

Cada par de observaciones satisface la ecuación:

$$y_i = a + bx_i + e_i$$

Donde $e_i = y_i - \hat{y}_i$ se denomina el residuo y describe el error de ajuste del modelo en el *i-ésimo* punto de los datos.

Método de mínimos cuadrados

Se define la suma de los cuadrados de los errores alrededor de la línea de regresión, denotada por SSE. El método de los mínimos cuadrados consiste en encontrar los valores de a y b, tal que SSE sea mínima.

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Método de mínimos cuadrados

Dada la muestra $\{(x_i, y_i);\ i = 1, 2, \cdots, n\}$, las estimaciones por mínimos cuadrados de a y b de los coeficientes de regresión α y β se calculan a partir de las fórmulas:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \bar{y} - b\bar{x}$$

Inferencias sobre los coeficientes de regresión

Estimador de σ^2

Un estimador insesgado de σ^2 es:

$$s^2 = \frac{SSE}{n-2} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-2} = \frac{S_{yy} - bS_{xy}}{n-2}$$

Donde:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2;\ S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2;\ S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Inferencias sobre los coeficientes de regresión

Intervalo de confianza para β

Un intervalo de confianza de $(1-\alpha')100\,\%$ para el parámetro β en la línea de regresión $\mu_{Y|x} = \alpha + \beta x$ es:

$$b - \frac{t_{\alpha'/2}s}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha'/2}s}{\sqrt{S_{xx}}}$$

donde $t_{\alpha'/2}$ es un valor de la distribución t con $n-2$ grados de libertad.

Prueba de Hipótesis $H_0 : \beta = \beta_0$

$$t = \frac{b - \beta_0}{s/\sqrt{S_{xx}}}$$

Inferencias sobre los coeficientes de regresión

Intervalo de confianza para α

Un intervalo de confianza de $(1-\alpha')100\,\%$ para el parámetro α en la línea de regresión $\mu_{Y|x} = \alpha + \beta x$ es:

$$a - \frac{t_{\alpha'/2}\, s \sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n S_{xx}}} < \alpha < a + \frac{t_{\alpha'/2}\, s \sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n S_{xx}}}$$

$t_{\alpha'/2}$ es un valor de la distribución t con $n-2$ grados de libertad.

Prueba de Hipótesis $H_0 : \alpha = \alpha_0$

$$t = \frac{a - \alpha_0}{s\sqrt{\sum_{i=1}^{n} x_i^2 / n S_{xx}}}$$

Predicción de Valores por regresión

Un intervalo de confianza de $(1-\alpha')100\,\%$ para la respuesta media $\mu_{Y|x_0}$ es:

$$\hat{y}_0 - t_{\alpha'/2} s\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha'/2} s\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}$$

$t_{\alpha'/2}$ es un valor de la distribución t con $n-2$ grados de libertad.

Un intervalo de predicción de $(1-\alpha')100\,\%$ para la respuesta media $\mu_{Y|x_0}$ es:

$$\hat{y}_0 - t_{\frac{\alpha'}{2}} s\sqrt{1+\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}} < y_0 < \hat{y}_0 + t_{\frac{\alpha'}{2}} s\sqrt{1+\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}}$$

Correlación

Si X es una variable aleatoria, entonces se tiene una distribución normal bivariada:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$
$$\times\ exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

para $-\infty < x < \infty$ y $-\infty < y < \sigma$, donde

$$\rho^2 = \beta^2\frac{\sigma_X^2}{\sigma_Y^2}$$

La constante ρ (rho) se llama **coeficiente de correlación poblacional**.

Estimación del Coeficiente de Correlación

La medición ρ de la asociasión lineal entre dos variables X y Y se estima mediante el coeficiente de correlación muestral r:

$$r = b\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$-1 \leq r \leq 1$.

- Se debe tener cuidado con las conclusiones basadas en este valor.
- r^2: La variación de Y que puede explicarse por la relación lineal con los valores de X.

Prueba de Hipótesis sobre ρ

$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$

Se toma el valor:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

t es un vlaor de la t-student con $n-2$ grados de libertad.

$H_0 : \rho = \rho_0 (\rho_0 \neq 0) \quad H_1 : \rho \neq \rho_0$

Se toma el valor z de una normal estándar:

$$z = \frac{\sqrt{n-3}}{2} ln \left[\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0} \right]$$

Regresión Lineal Múltiple

Para el caso de k variables independientes $x_1, x_2, \cdots, x_k$, la media de $Y|x_1, x_2, \cdots, x_k$ está dada por el modelo de regresión lineal múltiple:

$$\mu_{Y|x_1,x_2,\cdots,x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

la respuesta estimada se obtiene de la ecuación de regresión de la muestra:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

Donde cada coeficiente de regresión β_i se estima por b_i de los datos de la muestra, usando mínimos cuadrados.

Regresión Polinomial

Para el caso del ajuste de la ecuación polinomial:

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r$$

a los pares de observaciones $\{(x_i, y_i);\ i = 1, 2, \cdots, n\}$, con $n \geq r + 1$. Cada observación, y_i, satisface la ecuación:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_r x_i^r + \varepsilon_i$$

o

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + \cdots + b_r x_i^r + e_i$$

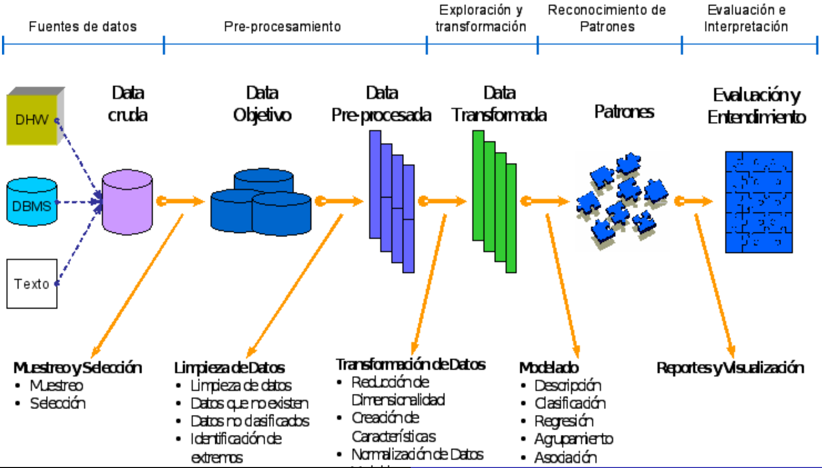
Donde:

r: es el grado del polinomio

ε_i: es error aleatorio

Introducción a la Minería de datos

Mapa Conceptual - Minería de Datos



Niveles de Entendimiento

Es el nivel previo de la conciencia. Se relaciona con el futuro.

Es humano. Es una apreciación del **Por qué**. Se puede generar nuevo conocimiento usando lo que se sabe y nueva información.

Es información útil, contextual, tácita. Experiencia (datos + información), responde a **Cómo**. Se puede aprender.

Data con un significado por vía de una relación. Puede o no ser útil. Responde a **Quién**, **Qué**, **Dónde**, **Cuándo**.

Existe o no, pero no tiene significado por si misma. Es codificable y explícita, es

Niveles de Entendimiento

Data:

- Requiere un medio de almacenamiento.
- Se debe capturar (registrar) y codificar.
- Es muy abundante.

Información:

- Se crea mediante la relación de datos.
- Es abundante y económica.
- Se puede distribuir

Conocimiento:

- Es la información útil.
- Se forma desde los patrones de comportamiento.
- Requiere de la vivencia del humano para ser aprendido.
- Es el ¿Cómo?
- No es fácilmente codificable, por lo tanto no es fácilmente almacenable ni recuperable.
- Es costoso (Entrenamiento, Aprendizaje, Experiencia

Tipos de Conocimiento

- **Explícito**. Es el que sabemos que tenemos y somos concientes cuando lo ejecutamos, se encuentra estructurado y esquematizado para su difusión. Se pueden transmitir y vender. (Ej. Procedimiento de Trabajo)

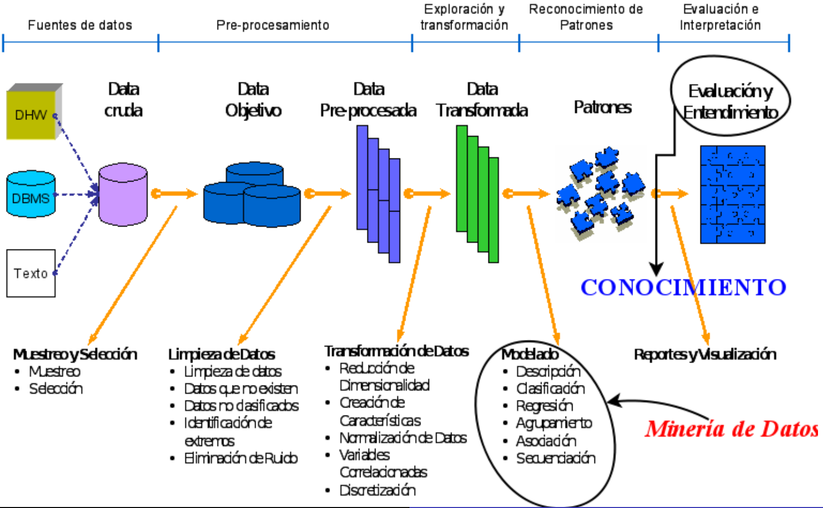
Tipos de Conocimiento

- **Explícito**. Es el que sabemos que tenemos y somos concientes cuando lo ejecutamos, se encuentra estructurado y esquematizado para su difusión. Se pueden transmitir y vender. (Ej. Procedimiento de Trabajo)
- **Implícito**. Sabemos que tenemos el conocimiento, pero no nos damos cuenta que lo estamos utilizando, simplemente lo ejecutamos y ponemos en práctica de una manera habitual. Se puede explicar "Por qué". Se usa de manera habitual, pero no mecánicamente. (Ej. Hablar un idioma).

Tipos de Conocimiento

- **Explícito**. Es el que sabemos que tenemos y somos concientes cuando lo ejecutamos, se encuentra estructurado y esquematizado para su difusión. Se pueden transmitir y vender. (Ej. Procedimiento de Trabajo)
- **Implícito**. Sabemos que tenemos el conocimiento, pero no nos damos cuenta que lo estamos utilizando, simplemente lo ejecutamos y ponemos en práctica de una manera habitual. Se puede explicar "Por qué". Se usa de manera habitual, pero no mecánicamente. (Ej. Hablar un idioma).
- **Tácito**. Permanece en un nivel inconsciente e intuitivo, se encuentra desarticulado, lo implementamos y ejecutamos de una manera mecánica sin darnos cuenta. Se transmite mediante la observación y la imitación. Es difícil de extraer, pero es muy valioso. (Ej. La forma de escribir o hablar.)

Minería de datos



Minería de Datos

Definición

- Proceso de descubrir CONOCIMIENTO a partir de los datos.
- Proceso de extraer CONOCIMIENTO útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos fórmatos.
- El conocimiento se puede manifestar como: patrones, reglas, restricciones, tendencias, etc.

En la minería de datos la tarea fundamental es encontrar modelos inteligibles a partir de los datos. El proceso debería ser automático o semi-automático (asistido) y el conocimiento generado debe ayudar en la toma de desiciones.

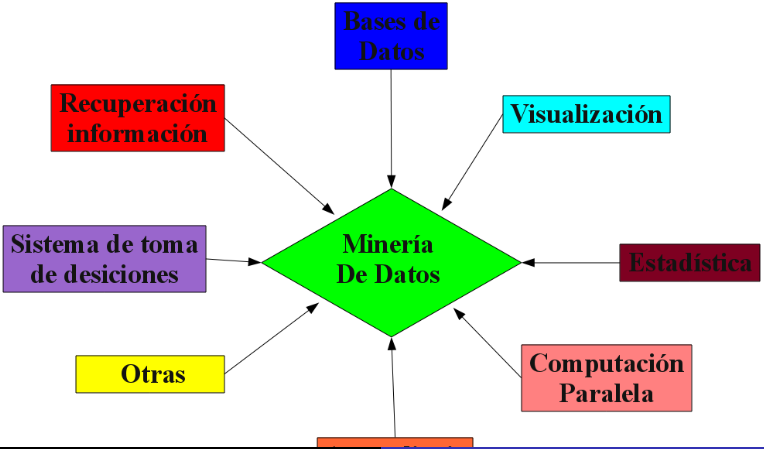
Propiedades del Conocimiento extraído

- **Válido**. Los patrones deben seguir siendo preciso para datos nuevos (con algún grado de certidumbre), y no solo para los que se usaron en su obtención.
- **Novedoso**. Debe aportar algo desconocido para el sistema y, preferiblemente, para el usuario.
- **Potencialmente útil**. La información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- **Comprensible**. Los patrones no comprensibles imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una informanción incompresible no proporciona conocimiento.

Tipos de Datos

- Bases de Datos Relacionales. Colección de Tablas (Cada columna un atributo y cada fila un registro o tupla)
- Otras Bases de Datos
    - Bases de Datos Espaciales (Ej. SIG, Redes de Transporte, etc.)
    - Bases de Datos Temporales (Ej. Tendencias Económicas, Electrocardiogramas, etc.)
    - Bases de Datos Documentales (Ej. Bibliotecas digitales, Indices, Fichas, etc.)
    - Bases de Datos Multimedia (Combinación de imágenes, audio y vídeo)
- La Word Wide Web

Relación de la Minería de Datos con otras disciplinas



Fases del proceso de extracción de Conocimiento

1. Integración y Recopilación
2. Selección
3. Preprocesamiento
4. Minería de Datos
5. Evaluación e interpretación
6. Difusión y uso

Integración y Recopilación

Desde diversas fuentes: Base de Datos, texto, imágenes, video, sonido, etc. Uso de almacenes de datos multidimensionales organizados y estructurados (*data warehousing*).

OLTP $\Rightarrow$ OLAP

SQL $\Rightarrow$ Análisis Multidimensional

Selección

Selección de atributos relevantes. Selección de muestras (Muestreo).

Generación de la “Vista Mineable”. Se puede reducir el tiempo de procesamiento y/o reducir las requisitois computacionales.

Preprocesamiento

Limpieza

Mejora de la calidad de los datos, eliminación de atributos irrelevantes o eliminación de datos extremos, tratamiento de datos faltantes.

Transformación

Discretizar atributos continuos, cuantificar atributos no numéricos, etc. Se puede aplicar técnicas de "Lógica difusa".

Minería de Datos

Es la más relevante, incluso se puede utilizar el termino para nombrar a todo el proceso.

Generación de modelos desde lo datos recopilados y seleccionados. Uso de varios modelos.

- Selección del Tipo de tarea de minería más apropiado
- Selección del Tipo de modelo
- Algoritmo para obtener el tipo de modelo con la tarea seleccionada.

Evaluación e interpretación

Evaluación y uso de los resultados obtenidos, reformulación del modelo.

La evaluación es un proceso complejo, y muchas veces subjetivo. Se pueden establecer tres criterios para evaluar patrones: Presición, Comprensibilidad e Interés. Además, dependiendo del tipo de modelo se debe aplicar la medida de evaluación más apropiada.

Difusión y uso

Se puede aplicar el modelo para:

- Recomendar acciones sobre el modelo
- Aplicar el modelo en conjunto de datos diferentes

Los resultados pueden integrarse a un sistema de TOMA DE DESICIONES.

Se debe aplicar algún esquema de monitorización que permita la adaptación y adecuación a nuevos datos (Evolución del Modelo).

Tareas de la Minería de Datos

Tipos de Aprendizaje

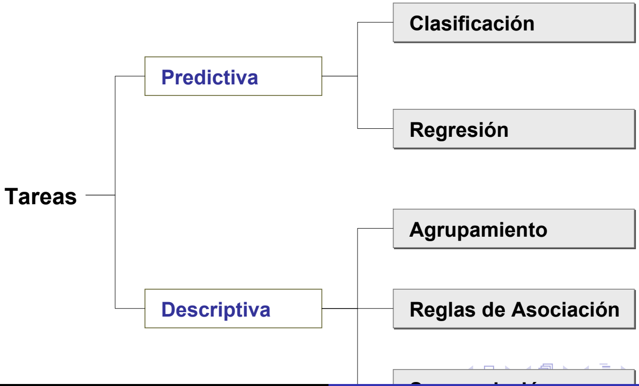
Supervisado

Existe una "guía" que sugiere una categoría para cada elemento del conjunto de entrenamiento. El objetivo es reducir el error de entrenamiento.

No Supervisado

No existe ninguna guía, el sistema realiza agrupamientos en forma natural sobre los patrones de entrada, para determinar la clase a la que pertenece.

Tareas de la Minería de Datos



Clasificación

- Intenta clasificar algunos objetos en un número finito de clases, en función a sus propiedades (características)
- El objetivo es encontrar un función de mapeo que permita separar la clase 1 de la clase 2, y éstas de la clase 3 $\cdots$.
- Las variables (atributos) pueden ser numéricas o categóricas (no numéricas).
- El modelo se construye con datos completos, cada registro tiene una clase predefinida.
- Busca formas de separar la data en clases pre-definidas.

Técnicas

Arboles de desición, Redes Neuronales, Clasificador Bayesiano,

Regresión

- Intenta determinar la función que mapea un conjunto de variables de entrada X (independiente), en una (o más) variables de salida Y (dependiente).
- Es básicamente numérica.
- Está basada en supuestos estadísticos.

Técnicas

Arboles de desición, Redes Neuronales, Regresión logística, SVM, etc.

Agrupamiento (*Clustering*)

Clasificación No Supervisada

- Intenta agrupar una serie de objetos en grupos.
- Cada objeto es representado por un vector $n-$dimensional.
- Los objetos que forman cada grupo deben ser disimilares.
- La similaridad es medida del grado de proximidad.
- Luego cada grupo es etiquetado.

Técnicas

- K-means (agrupamiento exclusivo)
- Fuzzy C-means (agrupamiento con traslape)
- Angulo de distribución mínima
- Método de autoorganización (SOM)

Reglas de Asociación

- Analiza los datos para descubrir reglas que identifiquen patrones o comportamientos.
- Reglas de la forma $A \Rightarrow B$.
- Usa algoritmos intensivos en procesamiento.
- Análisis de la cesta de la compra (market basket analysis).

Característica de las reglas

- Cobertura (*support*): Número de instancias que la regla predice correctamente.
- Confianza (*confidence*): Porcentaje de veces que la regla se cumple cuando se puede aplicar.

Secuenciación

- Buscar secuencias que son usualmente probables.
- Requiere entrenamiento, lista de eventos, conocimiento de eventos interesantes.
- Debe ser robusto en la fase de adicionar eventos con ruido.
- Usado en el análisis de fallas y predicción.

Técnicas

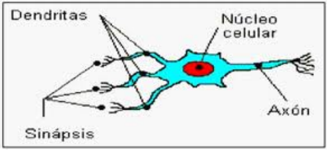
- Modelos de Markov (Cadenas de Markov)
- Agrupamiento MDD (Maximal Dependence Decomposition Clustering)

Redes Neuronales Artificiales

- Son una rama de la Inteligencia Artifial
- Intentan emular el funcionamiento de las Redes Neuronales Biológicas
- Son usadas en: Predicción, Clasificación y control (Modelado, simulación, control de procesos, manejo de fallas, dianóstico médico, etc.)
- Su éxito se debe principalmente a dos factores:
  - **Potencia**: Son técnicas no lineales sofisticadas capaces de modelar complejas de alta dimensionalidad.
  - **Fáciles de usar**: Aprenden con el ejemplo. Se recopilan datos representativos, y éstos mediante un algoritmo de entrenamiento para aprender. Se debe conocer el comportamiento desea de la Red, pero no como lograrlo.

Redes Neuronales Artificiales

Neurona Biológica (Natural)



- Es una célula especializada que puede propagar una señal electroquímica.
- Los axones de una neurona se conectan con las dendritas de otras neuronas a través de la sinapsis.
- Una neurona se activa cuando la señal total recibida en el núcleo celular excede cierto umbral.
- Si se activa emite una señal electroquímica a lo largo de su axón.

Redes Nueronales

Concepto Biológico

- Una Red Neuronal es una red de muchos procesadores simples ("Neuronas"), cada uno de los cuales tiene una cierta capacidad de almacenamiento o memoria.
- Las unidades están conectadas por canales de comunicación que transmiten datos codificados.
- Las unidades trabajan con datos locales, las entradas se reciben por unas conexiones ("dendritas") y se transmiten por otras ("axones").

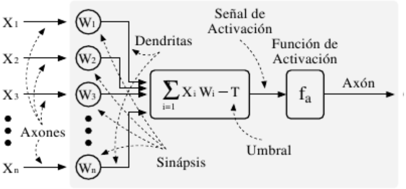
Redes Nueronales

El cerebro, una red neuronal natural

- Compuesto principalmente por un enorme número de neuronas (aprox. 10^{11}), y un mayor número de conexiones (aprox. 10^{14}).
- La sinapsis es una separación entre el axón de una neurona y la dendrita de otra. Este espacio está lleno de una sustancia neurotransmisora que permite transmitir la señal de una neurona a otra. La intensidad de la señal recibida depende de la eficacia de la sinapsis.
- Donal Hebb (1949) postuló que el aprendizaje consistía principalmente de alterar la "fuerza" o eficacia de las conexiones sinápticas.

La neurona artificial

En 1943 McCulloch y Pitts introdujeron un modelo de neurona que podía implantarse con circuitos. Una "neurona articial":



- Recibe varias entradas poderadas (pesos) por sus dendritas. Los pesos corresponden a la eficacia sináptica biológica.
- La suma ponderada de las entradas se resta al umbral; si el resultado es positivo (señal de activación) la neurona se dispara.

Usos de Redes Neuronales

Las Redes Neuronales son especialmente útiles en los problemas de clasificación, aproximación y transformación; en los cuales pueda aceptarse un cierto grado de imprecisión y que se disponga de datos para el entrenamiento.
Minería de Datos: No puede aplicarse alguna técnica clásica.

- Es importante entender que no hay métodos para entrenar a las Redes que pueda crear mágicamente la información que no esté en los datos del entrenamiento.
- Actualmente, simular el conocimiento y la emoción humana es ciencia ficción (con cualquier técnica).

Clasificación y Regresión

Problemas de predicción

- **Clasificación**: Determinar a cuáles, de un número de clases discretas, pertenece un caso dado de la entrada.
- **Regresión**: Predecir el valor de la variable generalmente continua.

En la mayoría de los casos, las redes neuronales tienen una sola variable de salida que puede estar compuesta por múltiples unidades de salida.

Clasificación

Algoritmo de Aprendizaje

- **Supervisadas**: El comportamiento E/S deseado es conocido y la red se entrena, ajustando los *pesos* para que su salida coincida con los valores deseados. Una vez entrenada la red "sabrá" como comportarse.
  - **Auto-asociativo**: El valor de salida es igual que el de la entrada.
  - **Hetero-asociativo**: El valor de la salida es diferente que el de la entrada.
- **No supervisadas**: A la red no se le proporciona ningún conjunto E/S (solo entrada).

Clasificación según la topología

- **Feedforward**: Las conexiones entre las *neuronas* no forman ciclos. Generalmente responden rápidamente y la mayoría de ellas se pueden entrenar usando métodos convencionales
- **Feedback**: Hay ciclos en las conexiones. Cada vez que se le da una entrada la NN debe iterar por un tiempo potencialmente largo antes de que produzca una respuesta. Generalmente es más difícil de entrenar que las *feedforward*.

Las Feedforward se utilizan cuando el conjunto de posibles valores de entrada y salida están acotados, en cambio las Feedback se utilizan cuando estos conjuntos no están acotados, por ejemplo en procesamiento de señales y series temporales.

Clasificación según los tipos de datos

- **Cualitativo**: Las variables solamente toman un número finito de valores. Generalmente, muchos de ellos pertenecen a una misma categoría. Este tipo de Redes se utilizan en problemas de *clasificación*.
- **Cuantitativo**: Las variables son numéricas y representan medidas de una cualidad. Las variables deben tomar valores tales que conserven alguna relación aritmética con las cualidades que están representando.

Las primeras decisiones que se deben tomar son: qué variables se van a utilizar; cuántos y cuáles casos se tendrán en cuenta.

E/S Cualitativas

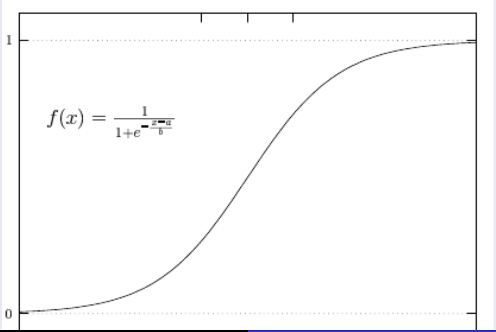
Las Redes Neuronales con variables cualitativas pueden ser de dos estados (`Encendido, Apagado`) o de múltiples estados (`Rojo, Verde, Azul`). La numeración puede ser:

- Ordinal { `Rojo=1, Verde=2, Azul=3`}
  Puede confundir la aritmética, `Verde` no es el promedio de `Rojo` y `Azul`.
- *one of N* { `Rojo=(1,0,0), Verde=(0,1,0), Azul=(0,0,1)`}
  Aumenta en forma prohibitiva el tamaño de la red haciendo más difícil su entrenamiento.

Hay que buscar en cada caso la manera en que mejor se adapte a la información que se maneja.

E/S Cuantitativas

Generalmente para el manejo de las entradas se elige una función que pueda aceptar cualquier valor y que produzca una salida en un rango limitado para evitar saturación. Una función muy utilizada es la logística. Es muy importante tener en cuenta el escalamiento.



Aprendizaje en Redes Neuronales

Consiste en la modificación de los *pesos* y se pueden considerar todas la reglas de aprendizaje como variantes de la regla propuesta por Hebb (*Hebbian learning rule* en 1949).

Básicamente la regla consiste en fortalecer aquellas sinapsis que unen a dos neuronas (i y j), si éstas se activan simultáneamente. Por ejemplo, la versión no supervisada puede ser

$$\Delta W_{ij} = \gamma X_i X_j$$

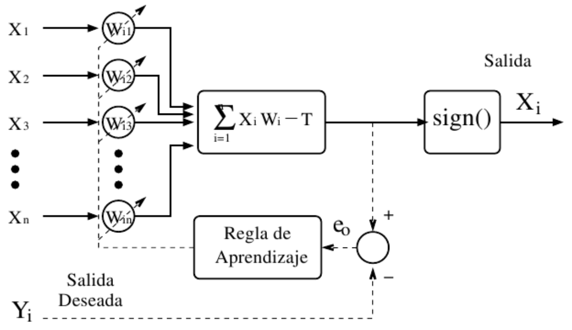
y la supervisada

$$\Delta W_{ij} = \gamma (X_i - Y_i) X_j$$

donde, $\gamma > 0$ representa la tasa de aprendizaje y Y_i la salida deseada suministrada por el supervisor.

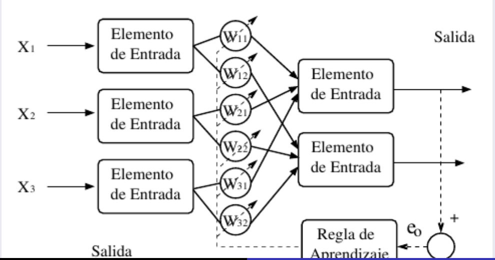
Adaline *Adaptative Linear Element*

Es el sistema más siple y consiste en un sumador lineal de las entradas $X_j \in \{-1, 1\}$ ponderadas por los *pesos* W_{ij}. A la salida del sumador hay una función signo que retorna -1 o $+1$ dependiendo del signo de las suma.



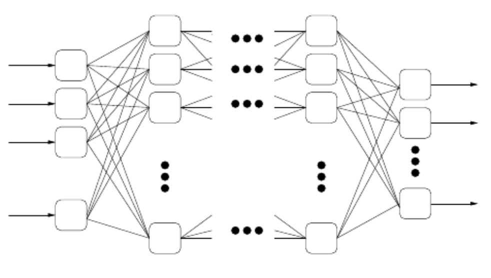
El Perceptrón

El Perceptrón al igual que Adaline básicamente es un sumador lineal de entradas que cambia la entrada bipolar $X_j \in \{-1, 1\}$ por una entrada binaria $X_j \in \{0, 1\}$ ponderadas. Consiste en una capa de entrada de N elementos que se conectan con los M elementos de salida. Entre las capas de entrada y la de salida se encuentran los *pesos*



Madaline y Perceptron Multicapa (MLP)

En 1969 Minsky y Papert demostraron que con el **Perceptron** de una capa de entrada y otra de salida *no es posible* entrenar la red para comportarse como una compuerta XOR. Para resolver este tipo de problemas se propuso una topología Multicapa.



Con una capa de entrada, una o varias ocultas y una de salida.

Aprendizaje en redes multicapa

Aplicando alguna de las generalizaciones de las reglas Hebbianas, el aprendizaje comprende un proceso iterativo de dos pasos:

1. Se presenta una entrada a la red y se calcula la salida que se compara con el valor deseado, obteniendose el *error*
2. El error es enviado hacia atrás *back-propagation* para calcular el cambio de cada uno de los pesos.

Hay que hacerlo de este modo, debido q que se desconoce la salida deseada de los elementos que se encuentran en las capas ocultas *hidden*

Técnicas y Herramientas