



# Que es la estadística?

Ciencia, método, herramienta.



# para qué sirve la estadística?

Análisis de fenómenos económicos, sociales, políticos, educacional, biológico, científico.

Es necesario registrar y ordenar continuamente, para seleccionar lo sucedido y lo que puede suceder.

Se han creado métodos, reglas, principios, que permiten la observación, ordenamiento, cuantificación y análisis llamado estadística.



# Que es la estadística?

Sistema o método usado en la recolección, organización, análisis y descripción numérica de la información.

Sistema de estudio de comportamiento de hechos o fenómenos en grupo.



# Fases de la estadística

- **Estadística Descriptiva o Deductiva:** Aspectos característicos de grupo (promedios, variabilidad de datos, etc.) que permiten hacer comparaciones específicas.
- **Estadística Inductiva o Inferencial:** Busca dar explicaciones y causas del comportamiento de un conjunto de observaciones, probar la significación o validez de los resultados que se pueden extender mas allá de la muestra.





# Términos comunes

- Población: Conjunto de elementos o características del cual se extrae la muestra.
- Muestra: Subconjunto de la población, que puede ser utilizada para representarla.
- Elementos: Son las entidades acerca de las que se reúnen datos.
- Parámetros: Números que resumen datos obtenidos de características.
- Variable o característica: Representación a través de números u otros símbolos de una variable.



# Características

- Cuantitativas o variables
- Cualitativas o atributos
- Edad, peso, altura, nombre, color de piel, sexo, intereses, ubicación, fecha de nacimiento



# Variables

- Variables continuas
- Variables discretas

Edad, peso, altura, número de hijos, fecha de nacimiento, caudal, volúmen, población de un municipio.





# Investigaciones estadísticas

- Conocer la realidad en una observación o fenómeno.
- Determinar lo típico o normal de esa observación
- Determinar los cambios que presenta el fenómeno.
- Relacionar dos o más fenómenos
- Determinar las causas que originan el fenómeno.
- Hacer estimativos sobre el comportamiento futuro del fenómeno
- Obtener conclusiones de un grupo menor (muestra), para hacerlas extensivas a un grupo mayor (población).
- Determinar el grado de validez y confiabilidad de una muestra.



# Tabla de distribución de frecuencias.

Agrupadas

No agrupadas



# Gráficos estadísticos

Histogramas de frecuencia, ojiva, diagramas de paretto, otras tablas y diagramas.



# Medidas de tendencia central

Media, mediana, moda, cuartiles, deciles, percentiles. (interpolación)



# Media Aritmética

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Dónde:

$x_i$  = el valor de cada dato

$n$  = el número total de datos.



# Media aritmética ponderada

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n n_i * y_i$$

Dónde:

$n_i$  = Frecuencia absoluta

$y_i$  = marca de clase

$n$  = el número total de datos.



# Media aritmética ( $M_x$ )

## Ventajas

- ✓ Es una medida fácil de entender y muy usada.
- ✓ Es estable en el muestreo.
- ✓ Es sensible a cualquier cambio en los datos de distribución.
- ✓ Es adaptable cuando se trata de hacer cálculos matemáticos posteriores con él.

## Desventajas

- ❖ Es sensible a valores extremos.
- ❖ Cuando la distribución es asimétrica puede no ser un valor único representativo de la serie.

# Media aritmética de varias muestras

$$\bar{x} = \sum_{i=1}^m \frac{n\bar{x}_1 + n\bar{x}_2 + \cdots + n\bar{x}_m}{n_1 + n_2 + \cdots + n_m}$$

Dónde:

$n_i$  = Frecuencia absoluta

$\bar{x}_i$  = media de la submuestra

$n$  = el número total de datos.



# Media geométrica MG

$$M_G = \sqrt[n]{x_1 * x_2 * \dots * x_n}$$

$$M_G = \sqrt[n]{\prod_{i=1}^n y_i^{n_i}}, \text{ para datos agrupados}$$

*Donde:  $x_i$  = Observaciones,  $y_i$  = Marca de Clase,  
 $n_i$  = Frecuencia Absoluta*

menos sensible a datos extremos y no pueden haber valores = 0

# Media Armónica H

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}$$

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{n_i}{y_i}} \text{ para datos agrupados}$$

Recomendada para medir velocidades, tiempos ,no puede usarse con valores iguales a cero,



# Media cuadrática

RMS (Root Mean Square)

$$M_2 = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$$

$$M_2 = \sqrt{\frac{\sum_{i=1}^n y_i^2 n_i}{n}} \text{ para datos agrupados}$$

# Mediana Me

Cuando el total de datos es **impar**, la mediana se determina por la siguiente fórmula:

$$i = X_{\frac{n+1}{2}}$$

Mientras que si el total de datos es **par**, la mediana se determina así:

$$i = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

Índice del valor  
ordenado de la media



# Mediana Me

$$M_e = l_i + \left[ \frac{\frac{n}{2} - N_{i-1}}{n_i} \right] * c$$

*Donde:*

$l_i$ : Límite inferior del intervalo que contiene  $n/2$

$n$ : número de datos de la muestra

$N_{i-1}$ : Frecuencia absoluta acumulada anterior al intervalo que contiene  $n/2$

$n_i$ : Frecuencia absoluta del intervalo que contiene  $n/2$

$c$ : Amplitud del intervalo que contiene  $n/2$

Para datos agrupados

# Moda Mo(Md)

Observación con la mayor frecuencia, puede no existir o existir más de una, bimodal o multimodal

$$M_o = l_i + \left[ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right] * c$$

*Donde:*

*$l_i$ : límite inferior del intervalo con mayor frecuencia absoluta*

*$\Delta_1$ : Diferencia entre la mayor frecuencia absoluta y la anterior*

*$\Delta_2$ : Diferencia entre la mayor frecuencia absoluta y la siguiente*

*$c$ : Amplitud del intervalo con mayor frecuencia absoluta*



# Medidas de posición no central

# Cuartiles $Q_k$

$$i = \left(\frac{K}{4}\right) n,$$

Índice del cuartil e datos ordenados

$$Q_k = l_i + \left[ \frac{k \left(\frac{n}{4}\right) - N_{i-1}}{n_i} \right] * C$$

$k$ : número del cuartil,  $k = 1, 2, 3$ .

$n$ : número total de datos.

$l_i$ : límite inferior del intervalo que contiene a  $k \left(\frac{n}{4}\right)$ .

$N_{i-1}$ : Frecuencia absoluta acumulada anterior al intervalo que contiene a  $k \left(\frac{n}{4}\right)$ .

$n_i$ : Frecuencia absoluta del intervalo que contiene a  $k \left(\frac{n}{4}\right)$ .

$c$ : amplitud del intervalo.



# Deciles $D_k$

$$i = \left(\frac{K}{10}\right) n$$

Índice del decil de los datos ordenados

$$D_k = l_i + \left[ \frac{k \left(\frac{n}{10}\right) - N_{i-1}}{n_i} \right] * C$$

$k$ : número del decil,  $k = 1, 2, 3, 4, 5, 6, 7, 8, 9$ .

$n$ : número total de datos.

$l_i$ : límite inferior del intervalo que contiene a  $k \left(\frac{n}{10}\right)$ .

$N_{i-1}$ : Frecuencia absoluta acumulada anterior al intervalo que contiene a  $k \left(\frac{n}{10}\right)$ .

$n_i$ : Frecuencia absoluta del intervalo que contiene a  $k \left(\frac{n}{10}\right)$ .

$c$ : amplitud del intervalo.

# Percentiles $P_k$

$$i = \left(\frac{K}{100}\right)n$$

Índice del percentil de los datos ordenados

$$P_k = l_i + \left[ \frac{k\left(\frac{n}{100}\right) - N_{i-1}}{n_i} \right] * C, \text{ para datos agrupados}$$



# Medidas de dispersión

desviación típica, varianza, medidas de  
asimetría y apuntamiento.



# **Amplitud o rango**



# Rango intercuartil

$$RIC = Q_3 - Q_1$$

# Varianza

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}, \text{ para una población}$$

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}, \text{ para una muestra}$$

para datos agrupados

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i^2 * n_i)}{N} - \mu^2, \text{ para una población}$$

$$S^2 = \frac{\sum_{i=1}^n (y_i^2 * n_i)}{n - 1} - \bar{X}^2, \text{ para una muestra}$$



# Desviación estándar

$$S = \sqrt{S^2}, \text{ para la muestra}$$

$$\sigma = \sqrt{\sigma^2}, \text{ para la población}$$

# Coeficiente de variación

$$CV = \frac{S}{\bar{X}} * 100$$

según Dane

- ✓ CV menor o igual al 7%, las estimaciones se consideran precisas.
- ✓ CV entre el 8% y el 14%, las estimaciones tienen precisión aceptable.
- ✓ CV entre el 15% y 20%, la precisión es regular.
- ✓ CV mayor del 20% indica que la estimación es poco precisa.





# Ejemplo Coeficiente de variación

Pensemos en una población de elefantes y otra de ratones. La población de elefantes tiene un peso medio de 5.000 kilogramos y una desviación típica de 400 kilogramos. La población de ratones tiene un peso medio de 15 gramos y una desviación típica de 5 gramos

# Puntaje típico o estandarizado

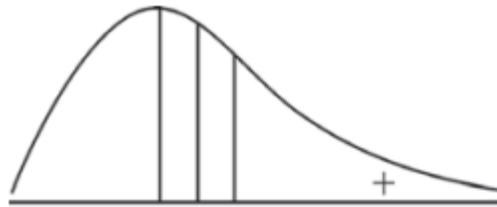
$$Z = \frac{x - \bar{x}}{s}$$

usado para distribuciones normales  
Comparar dos datos individuales

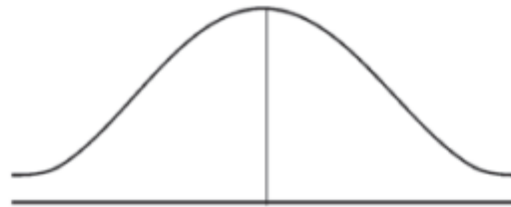
Una persona presenta un examen de admisión para ingresar a la facultad de Ingeniería, en dos Universidades en Colombia. En la primer Universidad logró una calificación de 85 y en la segunda obtuvo una calificación de 60, además se sabe que el promedio de calificación de la primer Universidad es 90 con una desviación estándar de 10, y en la Segunda Universidad la media de la calificación es 55 con una desviación de 12. En qué Universidad el aspirante logró mejor desempeño (Puntuación relativa), Martínez (2000).

# Medidas de asimetría

Distribución de frecuencias.



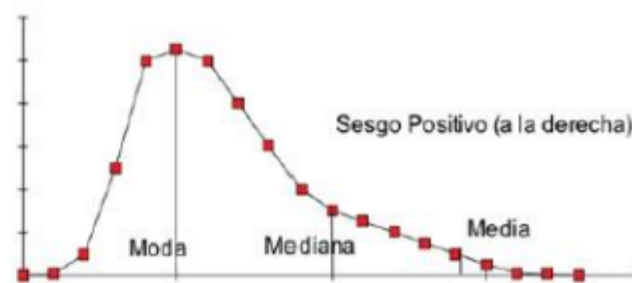
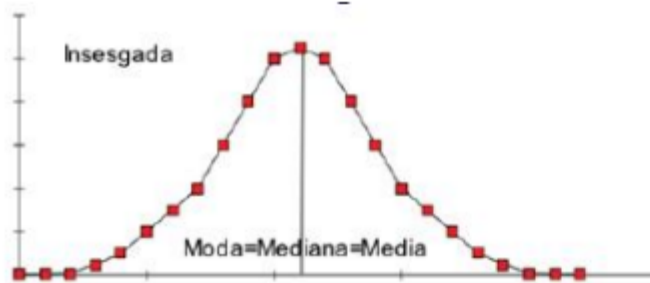
$M_d \ M_e \ M_x$   
 $M_d < M_e < M_x$   
*Asimétrica positiva*



$M_l = M_e = M_d$   
*Simétrica*



$M_x \ M_e \ M_d$   
 $M_l < M_e < M_d$   
*Asimétrica negativa*



# Medidas de asimetría

$$a) As = \frac{M_1 - M_o}{S}$$

$$b) As = \frac{3(M_1 - M_e)}{S}$$

$$c) As = \frac{m^3}{S^3} ; m^3 = \frac{(\sum (y_i - \bar{y}))^3 n_i}{n}$$

Donde:

$M_1$ : Media;  $M_o$ : Moda;  $M_e$ : Mediana;  $S$ : Desviación Estandar

$m^3$ : Tercer momento

Si  $A_s = 0$  Simétrico    Si  $A_s > 0$  Asimétrico positivo    Si  $A_s < 0$  Asimétrico negativo

# Medidas de asimetría

$$A_s = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_1}$$

Si  $A_s = 0$  Simétrico

Si  $A_s > 0$  Asimétrico positivo

Si  $A_s < 0$  Asimétrico negativo

# Asimetría de Fisher

$$g_1 = \frac{\sum (x_i - \bar{x})^3}{n * s^3}; \text{para datos sin agrupar}$$

$$g_1 = \frac{\sum n_i (y_i - \bar{x})^3}{n * s^3}; \text{para datos agrupados}$$

Dónde:

Sesgo = 0, simetría en la distribución de los datos

Sesgo > 0 Sesgo positivo o a la derecha

Sesgo < 0 Sesgo negativo o a la izquierda

Si  $A_s = 0$  Simétrico

Si  $A_s > 0$  Asimétrico positivo

Si  $A_s < 0$  Asimétrico negativo

# Medidas de apuntamiento Ap

Curtosis: Grado de agudeza

$$g_2 = \frac{\sum (x_i - \bar{x})^4}{n * s^4}; \text{ para datos sin agrupar}$$

$$g_2 = \frac{\sum n_i (y_i - \bar{x})^4}{n * s^4}; \text{ para datos agrupados}$$

$g_2 = 3$ , la distribución es normal o mesocúrtica

$g_2 > 3$ , la curva es leptocúrtica o aguda

$g_2 < 3$ , la curva es platicúrtica o achatada





# **Estadística Descriptiva**

Resumen de estadística descriptiva Excel



# UT2

Generalidades, probabilidad elemental, espacio muestral, reglas básicas de probabilidad, permutaciones, combinaciones, variaciones.



# Probabilidad

Cómo medimos la incertidumbre?



# Probabilidad

- Se dirá que un **hecho es cierto**, cuando son favorables todos los casos posibles. Un ejemplo podría ser, el de comprar todos los billetes de un sorteo, por lo tanto será un hecho cierto que ganará el sorteo.
- Llamaremos un **hecho verosímil** a un suceso susceptible de realizarse, pero su probabilidad favorable es menor que la unidad y mayor que 0,5.
- Si la probabilidad es igual a 0,5 será un **hecho dudoso**, ya que las probabilidades ventajosas y desventajosas son iguales. Tal es el caso del lanzamiento de una moneda, en la aparición de cara o sello.
- **Hecho inverosímil**, se presenta cuando la probabilidad es menor que 0,5 y mayor que cero.
- **Hecho imposible**, es cuando no existe posibilidad alguna de salir favorecido; por ejemplo el individuo que no compra lotería, la probabilidad que tiene de ganar será cero.



# Probabilidad empírica

$$P = \frac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$$



# Probabilidad apriori

Moneda, cuál es la probabilidad de que salga cara?



# Probabilidad objetiva y subjetiva

Objetiva = empírica, clásica

Subjetiva = Criterio personal

# Probabilidad con base en las frecuencias relativas

$$\text{Probabilidad de ocurrencia} = \frac{\text{Número de veces de ocurrencia del evento en el pasado}}{\text{Número total de observaciones}}$$





# Conceptos básicos

Experimento = Mismas condiciones en la realización de pruebas.

Prueba = Realización de un acto

Punto muestral = Resultado de una prueba o suceso

Evento = conjunto de uno o varios puntos muestrales (subconjunto de espacio muestral)

Espacio muestral = Conjunto de todos los resultados posibles



# Espacio muestral

elegir un alumno del curso de estadística en la facultad.

lanzamiento de una moneda teórica

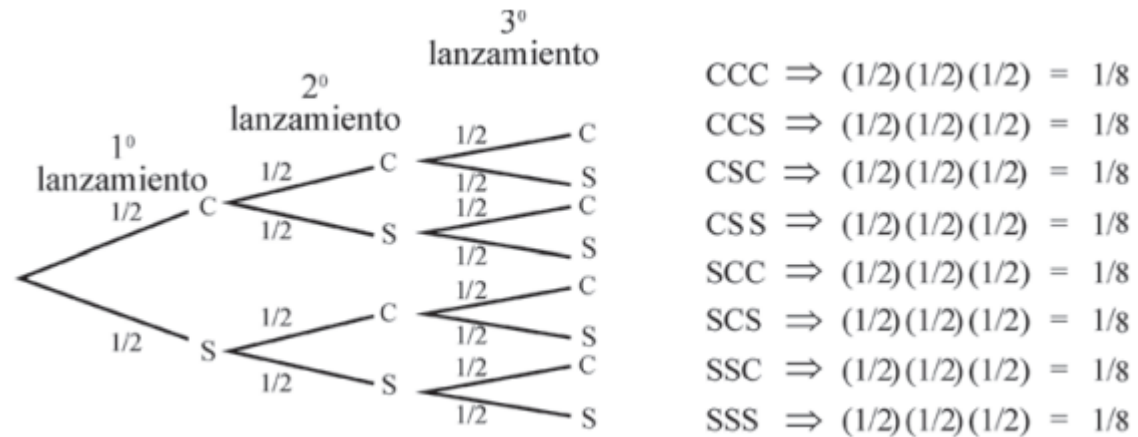
lanzamiento de dos monedas, asignando la probabilidad a cada suceso.

lanzamiento de tres monedas

Lanzamiento de 3 dados.

# Diagrama de árbol

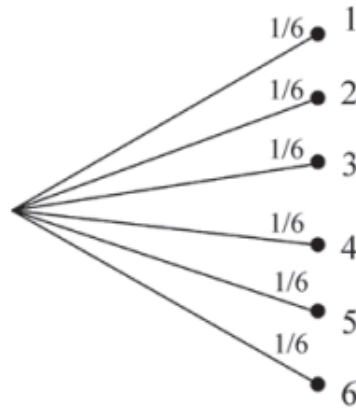
Lanzamiento de 3 monedas



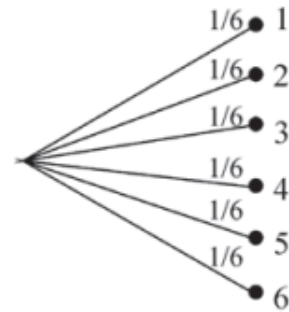
# Diagrama de árbol

Lanzamiento de 3 dados

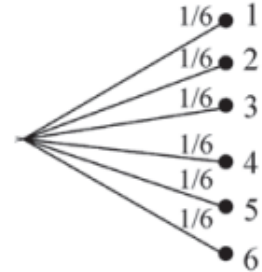
1º Lanzamiento



2º Lanzamiento



3º Lanzamiento



La probabilidad de obtener las caras ( 3 4 2 )  $\rightarrow (1/6)(1/6)(1/6) = \frac{1}{216}$



# Esperanza matemática

Si  $p$  es la probabilidad de éxito de un suceso en un solo ensayo, el número de sucesos o la esperanza de ese suceso en  $n$  ensayos, estará  $E = np$  dado por el producto de  $n$  y la probabilidad de éxito  $p$ .



# Diagrama de Vent

A decorative vertical strip on the left side of the slide, composed of a repeating pattern of triangles in various shades of gray and white, creating a 3D effect.

# **REGLA DE MULTIPLICACIÓN, PERMUTACIONES Y COMBINACIONES**

Cuando el diagrama de árbol es muy extenso.



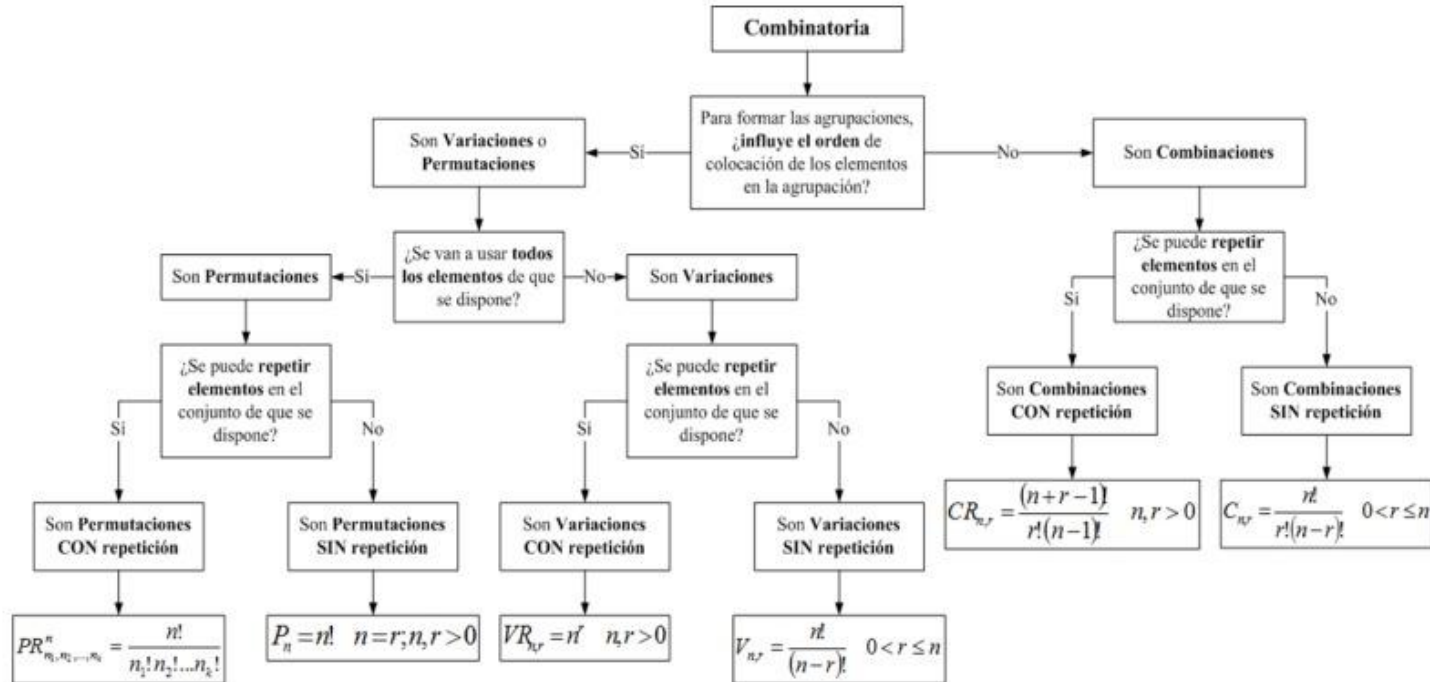
## REGLA DE MULTIPLICACIÓN Y SUMA

Cuando un evento se puede realizar de  $M$  formas diferentes y otro evento se puede realizar de  $N$  formas diferentes, el número de combinaciones posibles será de  $M \times N$

Cuando además, si ocurre uno no puede ocurrir el otro, entonces, el evento A o el evento B, se realizarán de  $m+n$  formas



# Combinatoria






# Probabilidad

Probabilidad condicionada, teorema de Bayes.

Distribuciones de probabilidad discretas,  
Bernoulli, Binomial



# Probabilidad




Distribuciones de probabilidad continuas,  
Distribución normal. Otras distribuciones Gumbel



# UT3

Diseño de muestreo, tipos de muestreo, muestreo aleatorio, distribución muestral, estimación, intervalo de confianza, límite de confianza.




Tamaño óptimo de la muestra en poblaciones finitas e infinitas, Técnicas de muestreo (MIL STD 105D).




# UT4

Hipótesis estadística, hipótesis nula, hipótesis alterna, tipos de error, prueba unilateral y bilateral, nivel de significación y puntos críticos. Procedimiento para pruebas de hipótesis.



Pruebas con  $\sigma$  conocido. Teoría de las muestra  
pequeñas, distribución t de student, grados de  
libertad.






## Análisis de varianza. Tabla ANOVA. Utilización software




# UT5

Diagrama de dispersión, tipos de regresión, regresión lineal simple. Función matemática.



Análisis de correlación, Coeficiente de correlación, Utilización de software.



Análisis de regresión aplicado, proyecto  
integrador de fin de curso.