

Estadística Básica

Vicent Modesto i Alapont
UCIP. Hospital “La Fe”



Nivel conceptual: Diseño metodológico

- Protocolo de recogida de datos:
 - Definiciones operativas de las variables.
- Soporte informático: bases de datos relacionales
 - MS-Access
 - Se pueden usar hojas de cálculo (MS-Excel).

Tipos de Variables

- Variables independientes:
 - Causas (previas en el tiempo)
 - Aleatorizadas o bajo control experimental.
- Variables dependientes o resultado:
 - Consecuencias
 - Se miden al pasar el tiempo.

Tipos de Variables

- **Nominales:**
 - Categorías no ordenadas: Color del pelo
 - Dicotómicas o binarias.
- **Ordinales o de escala (tipo Likert):**
 - Categorías ordenadas: Mal-Regular-Bien
- **Intervalos o rangos:**
 - Semicuantitativas: < 2500 gr; 2500-3000 gr; > 3000gr
- **Cuantitativas: Numerales escalables**
 - Glicemia, HDL colesterol, ...

Aspecto de una matriz de datos

<i>Caso</i>	<i>Sexo</i>	<i>Historia</i>	<i>PaCO2_0</i>	<i>PaCO2_24</i>
1	0	2345	88	50
2	0	3216	49	48
3	1	5678	76	47
4	0	9807	74	77
5	1	1246	50	39
6	0	1367	47	43
7	1	1789	56	40

Nivel estadístico

- SPSS, SAS, STATA, BMDP, S-PLUS, Minitab (Programas buenos y caros)
- EpiInfo (Programa bueno y gratuito).
 - Recomendado por la OMS y CDC
- EpiDat (Datos tabulados)
 - Xunta de Galicia
 - Recomendado por OMS y OPanam Salud



	var	var	var	var	var	var	var	var	var
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									



	1	2	3	4	5	6	7	8	
	CASO	FECHA	SEXO	ALTURA	TABAQUIS	PESO	HTA	ALCOHOL	DISLI
1	1.00	01/02/55	M	170.00	25.00	80.00	2.00	50.00	
2	2.00	01/02/33	F	155.00	0.00	69.00	2.00	0.00	
3	3.00	05/08/30	F	160.00	10.00	58.00	1.00	0.00	
4	4.00	25/08/39	M	175.00	35.00	85.00	2.00	120.00	
5	5.00	21/09/42	M	165.00	20.00	80.00	2.00	80.00	
6	6.00	04/09/30	F	155.00	0.00	73.00	3.00	0.00	
7	7.00	05/04/36	F	170.00	15.00	69.00	0.00	40.00	
8	8.00	14/07/43	M	166.00	25.00	82.00	2.00	110.00	
9	9.00	25/01/35	M	180.00	30.00	95.00	2.00	150.00	
10	10.00	02/05/58	M	173.00	40.00	75.00	0.00	120.00	
11	11.00	23/06/43	M	185.00	25.00	96.00	2.00	150.00	
12	12.00	04/05/55	M	170.00	30.00	77.00	0.00	0.00	
13	13.00	05/02/33	F	157.00	0.00	80.00	3.00	0.00	
14	14.00	14/04/45	M	163.00	15.00	67.00	2.00	50.00	
15	15.00	15/03/52	F	168.00	23.00	69.00	2.00	100.00	
16	16.00	23/08/31	F	159.00	0.00	68.00	3.00	0.00	



Session

Retrieving worksheet from file: C:\MTBWIN\DATA\Choleste.mtw
Worksheet was saved on 6/5/1996

Descriptive Statistics

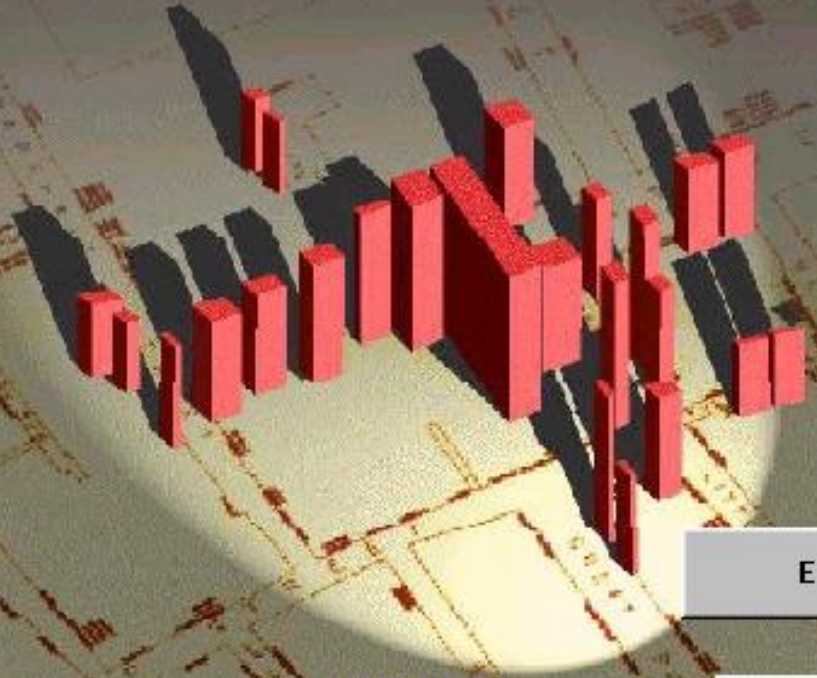
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
2-Day	28	253.93	268.00	254.15	47.71	9.02

Variable	Min	Max	Q1	Q3
2-Day	142.00	360.00	224.50	282.00

Data

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
↓	2-Day	4-Day	14-Day							
1	270	218	156							
2	236	234	*							
3	210	214	242							
4	142	116	*							
5	280	200	*							
6	272	276	256							
7	160	146	142							
8	220	182	216							

Epi Info 2002



Vistas

Grabar datos

Analizar Datos

Salir

Epi Map

Estadísticas nutricionales

Página WEB de EpiInfo

Revision 1

November 4, 2002

Salir



Anterior



Siguiete



Ultimo



Historial



Abrir



Bookmark



Imprimir



Maximizar

EpiInfo 2002

Ordenes

Datos

- Abrir (Read)/Importar
- Relacionar
- Escribir/Exportar (E)
- Unir

Variables

- Definir
- Anular definición
- Asignar
- Recodificar
- Documentar

Selección/Condiciones

- Seleccionar
- Cancelar selección
- Condiciones
- Ordenar
- Cancelar Ordenar

Estadísticas básicas

- Listar
- Frecuencias
- Tablas
- Emparejar/Aparear
- Medias
- Gráficos
- Mapas

Estadísticas avanzadas

Ayuda

Editor de órdenes - Nuevo programa

Archivo Edición Vista Fuentes Ejecutar Ayuda

Nuevo

Abrir

Guardar

Imprimir

Ejecutar

Ejecutar esta orden

Axioma de la estadística

Garbage in = Garbage out

A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.

Estadística Descriptiva

Estadística descriptiva

- El primer paso de todo análisis
- Objetivo:
 - Mediante calculos hechos en nuestra muestra, describir la población de la que se extrajo
 - ***Inferencia poblacional***: IC 95%
- Resume la matriz de datos.

Descripción de datos cualitativos

- Recuento y porcentajes
- IC95% de la proporción poblacional
- Utilización de tablas 2x2

Tabla 2x2

<i>Tratados con Propranolol</i>	<i>Supervivencia</i>		<i>Marginal</i>
	+	-	
+	38	7	45
-	29	17	46
<i>Marginal</i>	67	24	91

Ejercicio

- Existe la hipótesis que, entre los pacientes coronarios, uno de los principales determinantes de la mortalidad es el sexo femenino.
- Para contrastar dicha hipótesis se estudiaron 434 pacientes coronarios ingresados en la UCI del Hospital la Fe, observándose los siguientes datos:

Ejercicio

- **Mujeres ingresadas:109**
 - **Fallecieron:21; Supervivientes:88**
- **Varones ingresados:325**
 - **Fallecieron:25; Supervivientes:300**
- **Construye una tabla 2x2**

Solución al ejercicio

<i>Sexo</i>	<i>Fallecido</i>		<i>Total</i>
	<i>+</i>	<i>-</i>	
Mujer	21	88	109
Varón	25	300	325
Total	46	388	434

Descripción de variables cuantitativas

- Medidas de centralización.
 - Promedio o Media aritmética (m). Se calcula sumando los valores de las observaciones y dividiendo el resultado por el número de las mismas.
 - Mediana.
- Medidas de dispersión.
 - Desviación típica o estándar (DE).

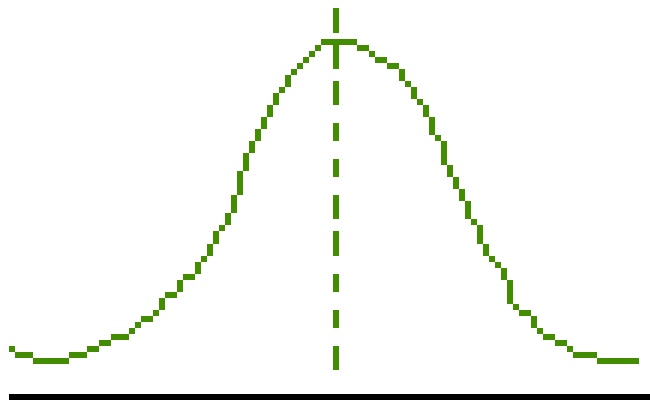
Parámetros de Centralización: Media Aritmética

$$m = \sum x / n$$

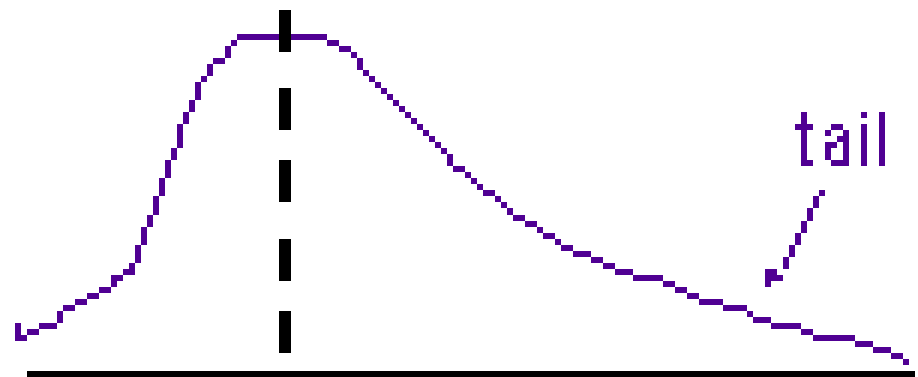
Varianza y Desviación Estándar

	x	$x-m$	$(x-m)^2$
	121	-15.86	251.45
	130	-6.86	47.02
	140	3.14	9.88
	121	-15.86	251.45
	169	32.14	1033.16
	107	-29.86	891.45
	170	33.14	1098.45
$\bar{m} = \Sigma x/n$	958	0	3583
	136.86		
$V = \Sigma (x-m)^2/n$			511.84
$DE = \sqrt{V}$			22.62

Normal and Skewed Distributions



- symmetrical
- mean, mode, median
- unimodal



- skewed to the left
- mode, median
- unimodal



Mean, Median, Mode

Respiratory rates

group 1 = (11, 12, 13, 13, 14, 15); $\bar{x} = 13$

group 2 = (11, 12, 13, 13, 14, 25); $\bar{x} = 17$

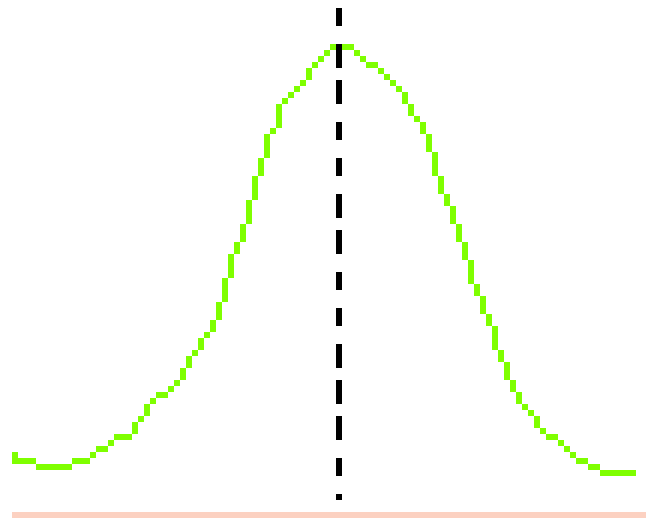
The mean is affected by extreme values

Median is 13 -- value that divides data in 50:50

Mode is 13 -- the most common value

Measures of Central Tendency

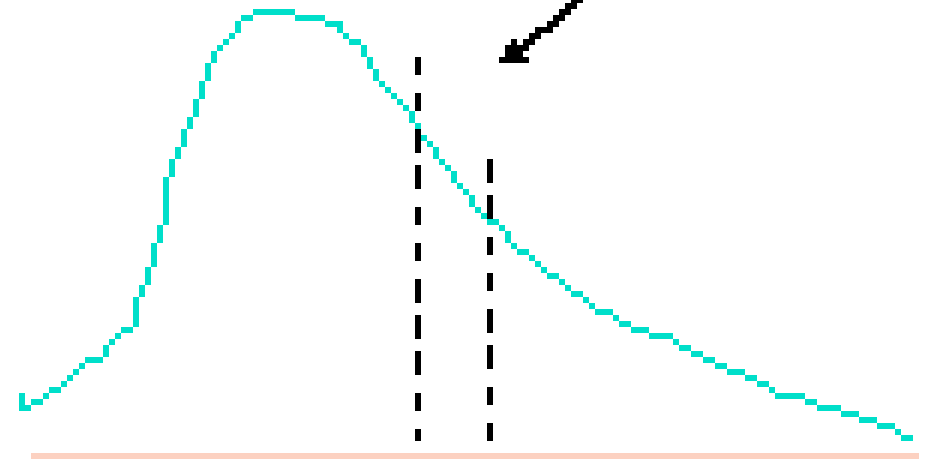
mean, median, mode



mode

median

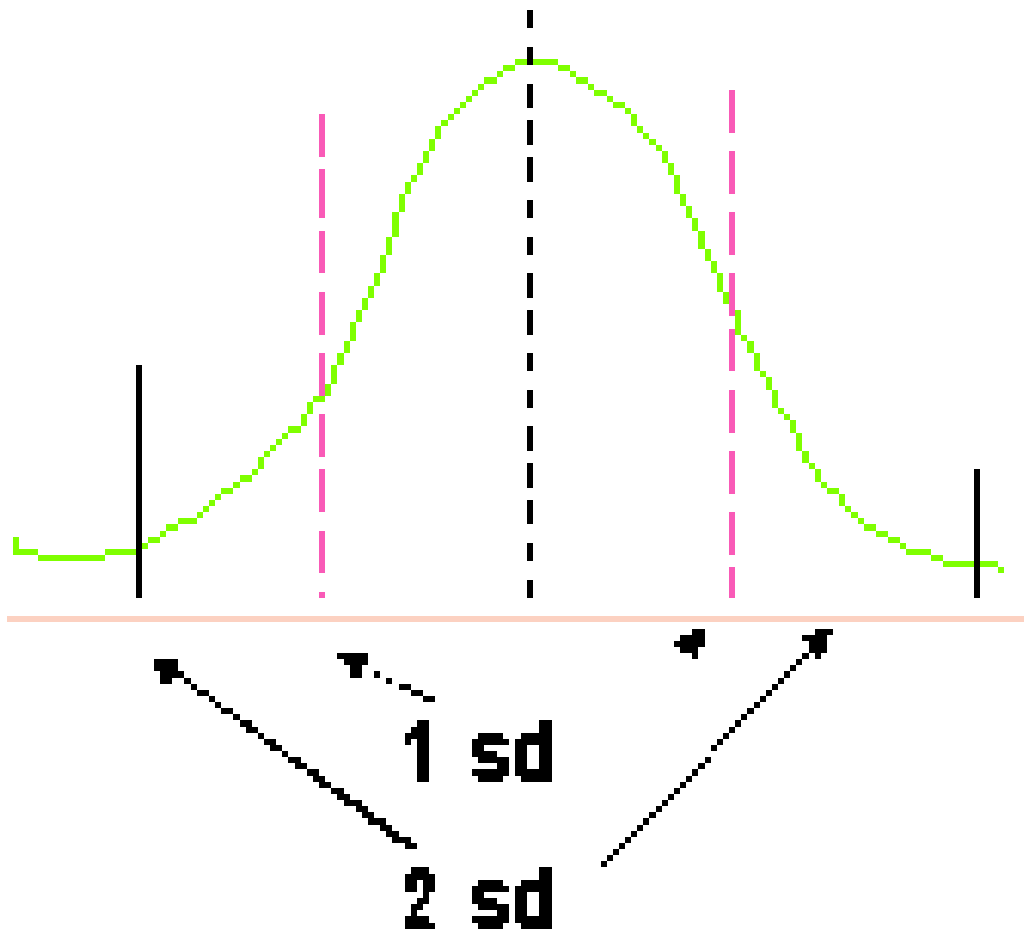
mean



Measures of Dispersion

- **Standard deviation:** 1 sd, 2 sd, 3 sd
- **Percentiles:** 25, 50, 75, 90, 95
- **Range:** largest value-smallest

Standard Deviation



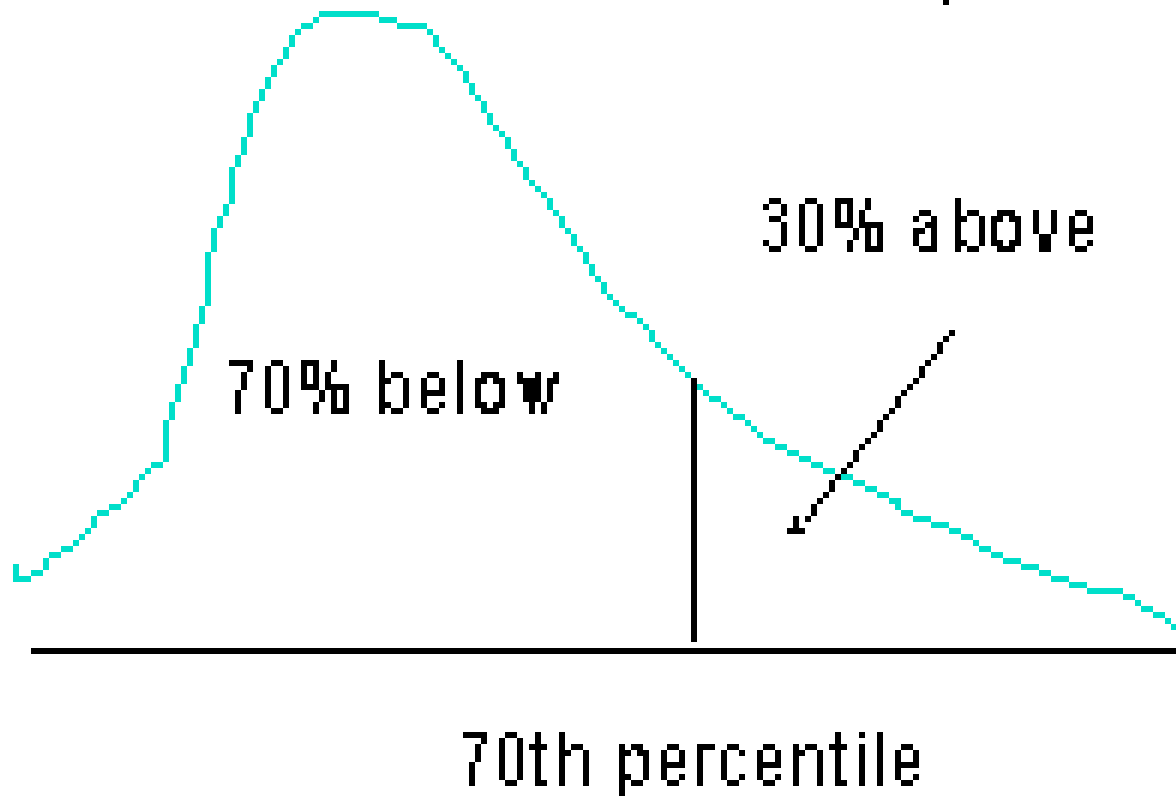
$sd^2 = \text{variance}$

1sd = 68%

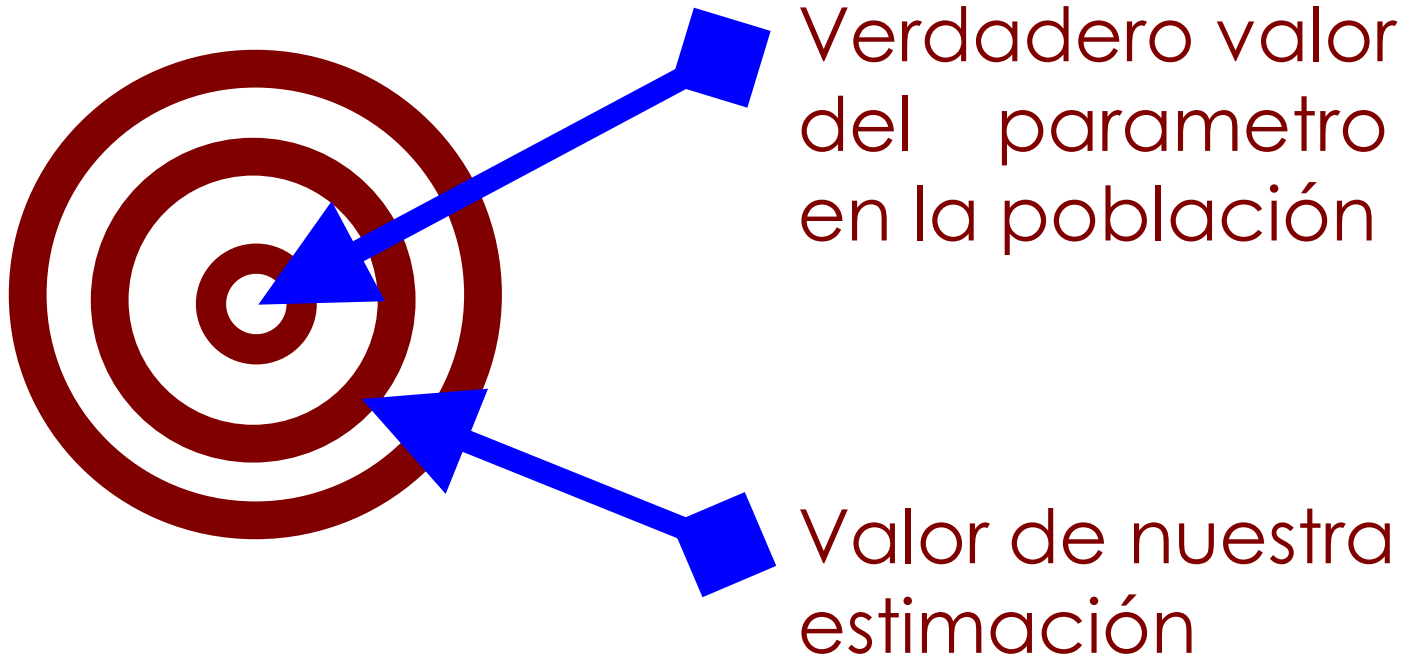
2sd = 95%

Percentiles

quartile 1 = 25%
quartile 2 = 50% (median)



Inferencia Poblacional



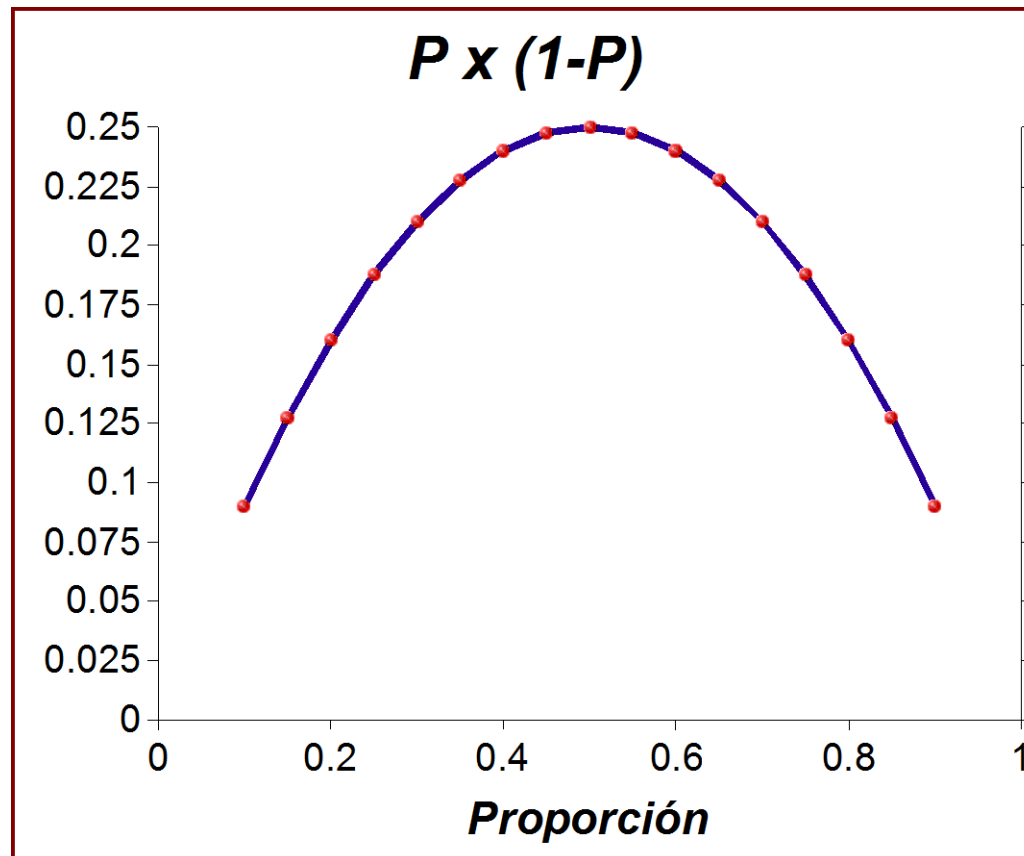
Error Estándar: Media Poblacional

$$EE = \frac{DE}{\sqrt{n}}$$

Error Estándar: Proporción Poblacional

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

Error estándar: Proporción Poblacional

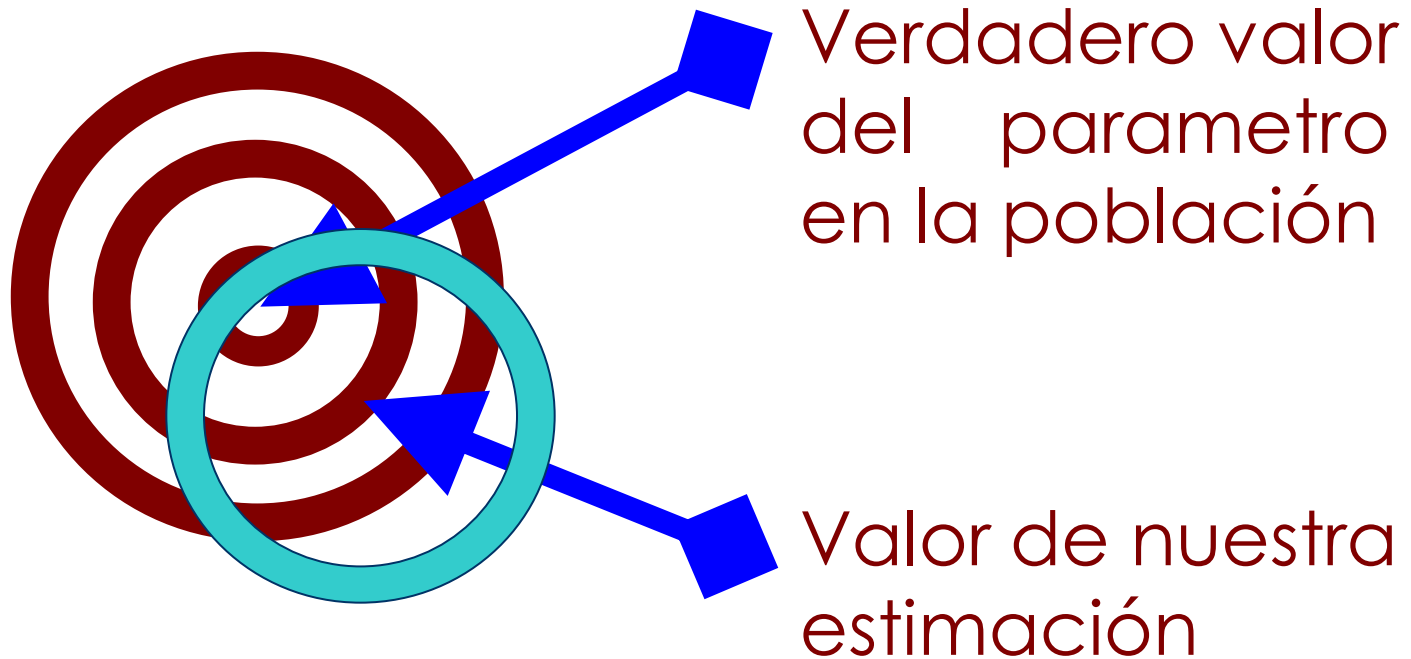


EE de una Media y una Proporción

$$EEM = \sqrt{\frac{\text{var}}{n}} \qquad EEP = \sqrt{\frac{p(1-p)}{n}}$$

- Numerador: *incertidumbre* muestral
- Denominador: tamaño muestral

Intervalo Confianza Poblacional



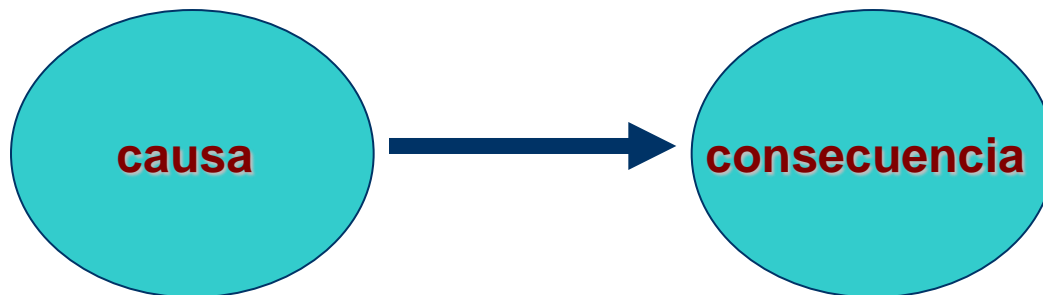
$$\text{IC (poblacional)} = \text{media} \pm \text{error estándar}$$

A decorative graphic on the left side of the slide, consisting of a light green L-shaped corner and a dark blue horizontal bar with rounded ends.

Estadística Analítica

Estadística “Analítica”

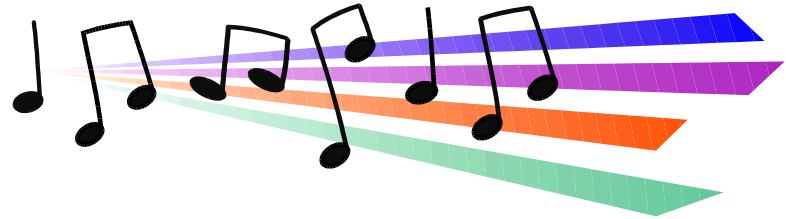
- **Descartar (o Confirmar) hipótesis**
- **Inferencia de relación causal**



...es calcular “la p”

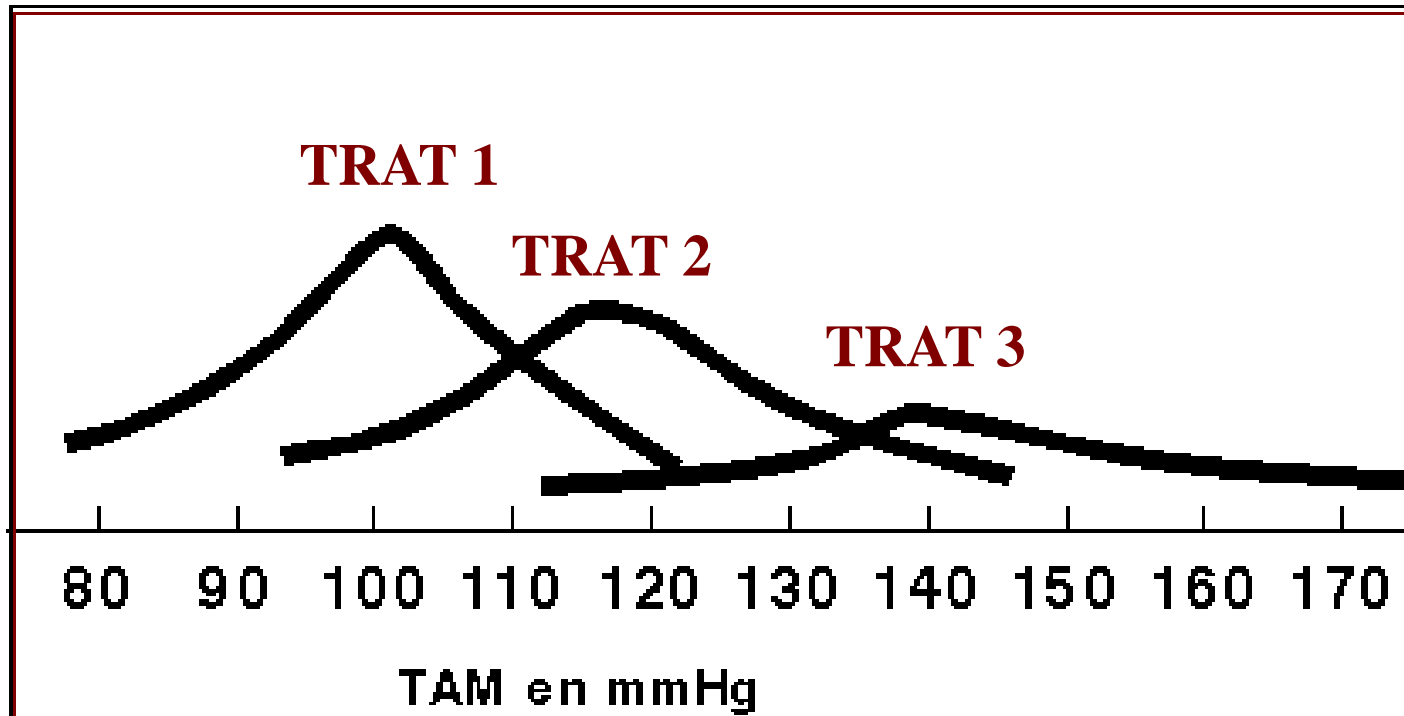
Fundamento de los tests estadísticos

- **Relación Señal/Ruido**
 - **La señal: diferencias entre medias o proporciones**
 - **El ruido: Error Estándar de la diferencia**



¿Es diferente el efecto entre los TRES tratamientos?

Test ANOVA



Test ANOVA

TRATAM	TA
1	100
1	87
1	125
1	80
1	89
1	120
1	123
1	120
1	114
1	180
1	150
1	127
1	130
1	98
1	100
2	89
2	100
2	109
2	90

TRAT 1	TRAT 2	TRAT 3
100	89	121
87	100	130
125	109	140
80	90	121
89	98	169
120	121	107
123	89	170
120	90	
114	98	
180	98	
150	99	
127	100	
130	108	
98	98	
100	99	
	110	
	89	

**Cálculo de estadístico F:
Bajo H0 se distribuye F Snedecor**

$$F = \frac{V \text{ Entre Grupos}}{V \text{ Intra Grupos}}$$

ANOVA como Razón Señal/Ruido

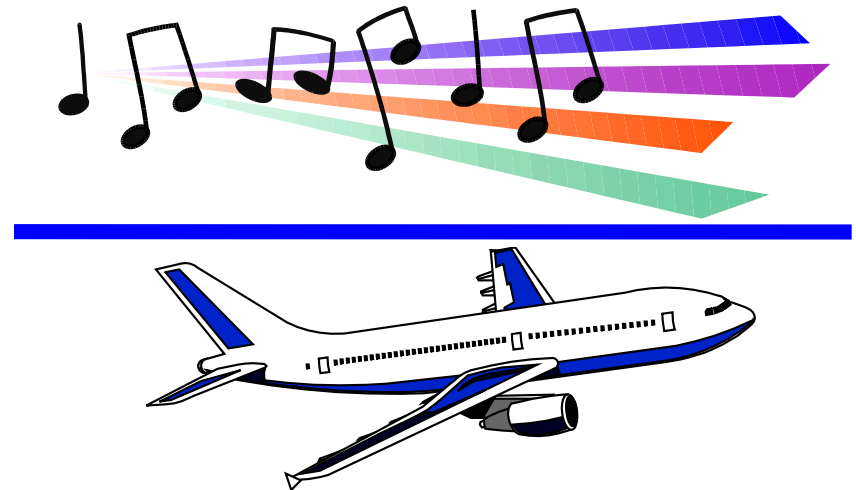
- **Señal:**

Var ENTRE grupos

$$F = \frac{V \text{ Entre Grupos}}{V \text{ Intra Grupos}}$$

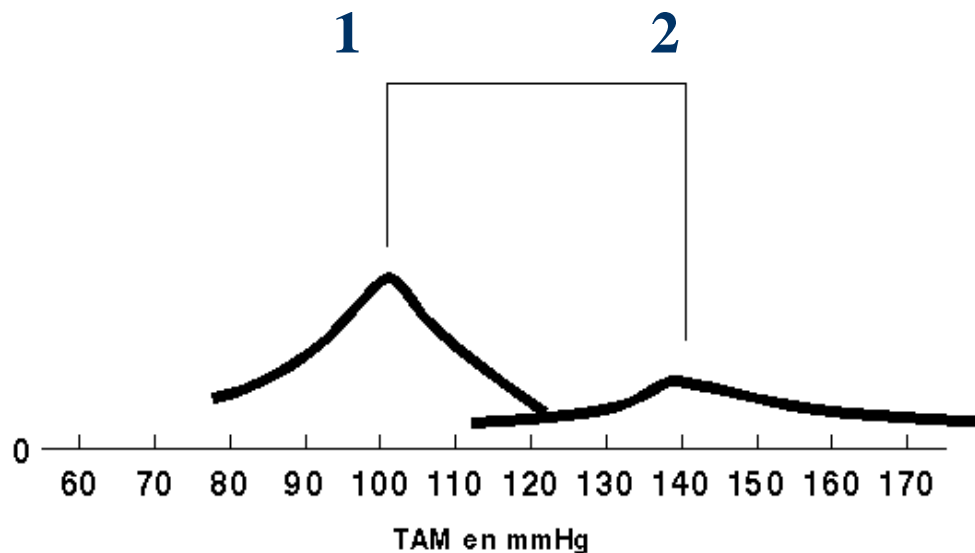
- **Ruido:**

Var INTRA grupo



¿Es diferente el efecto entre los DOS tratamientos?

Test “t” de Student Fisher



t-test

compare respiratory rates
between group of horses

group 1	group 2
11	11
12	12
13	13
14	13
15	14
	25

Horses in group 1 are
independent
from horses in group 2

H_0 : mean 1 = mean 2

H_a : mean 1 \neq mean 2

Cálculo de estadístico t:
Bajo H_0 se distribuye T Student-Fisher

t-test calculations

group 1 | group 2

11	11
12	12
13	13
14	13
15	14
	25

mean 13

14

s1 = 1.58

s2 = 5.16

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\text{var} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

var $\left(\frac{1}{n_1} + \frac{1}{n_2} \right)$

EEDM

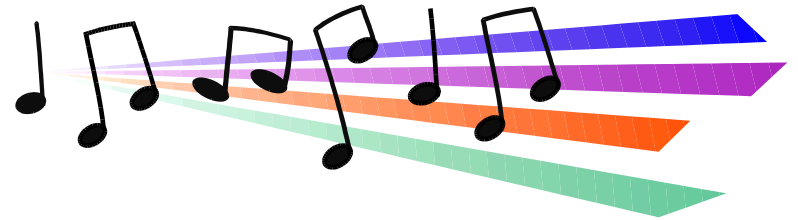
$$\text{var} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

La “t” de Student como Razón Señal/Ruido

- **Señal:**

DIFERENCIA medias

$$t = \frac{\text{Diferencia de Medias}}{\text{EE de la Diferencia}}$$



- **Ruido:**

VARIANZA diferencia medias



Diferencia de “Porcentajes” (Variables cualitativas)

Test Ji al cuadrado de Pearson

<i>Tratamiento</i>	<i>Remisión</i>		<i>Marginal</i>
	<i>+</i>	<i>-</i>	
<i>Estudio</i>	A	B	50
<i>Control</i>	C	D	50
<i>Marginal</i>	50	50	100

Diferencia en Recuento de Observados vs Esperados

<i>Tratamiento</i>	<i>Remisión</i>		<i>Marginal</i>
	+	-	
Estudio	37	13	50
Control	13	37	50
<i>Marginal</i>	50	50	100

The table displays observed counts (red) and expected counts (blue) for a 2x2 contingency table. The observed counts are 37 (Study, +) and 13 (Study, -) in the first row, and 13 (Control, +) and 37 (Control, -) in the second row. The expected counts are 25 for all four cells. Green double-headed arrows indicate the difference between observed and expected counts: 12 for the top-left cell (37-25), 12 for the top-right cell (13-25), 12 for the bottom-left cell (13-25), and 12 for the bottom-right cell (37-25).

Test de Ji al Cuadrado

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi_{obs}^2 = \frac{n(ad - bc)^2}{n_1 n_2 (a + c)(b + d)} =$$

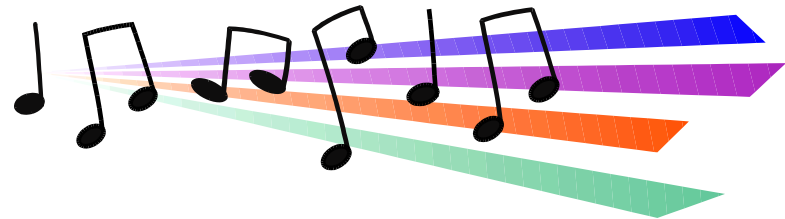
Cálculo de estadístico χ^2 :
Bajo H_0 se distribuye Ji al cuadrado

Ji al Cuadrado como Razón Señal/Ruido

- **Señal:**

Suma DISTANCIAS² obs-esp

$$= \sum \frac{(O - E)^2}{E}$$



- **Ruido:**

“Cuantos hay en todas las casillas”

mp

2

Fundamento de los tests estadísticos

- Señal
“pequeña”
vs ruido
grande
- $p > 0,05$
Test no
significativo



Fundamento de los tests estadísticos

- **Ruido**
pequeño vs
señal grande
- **Test**
significativo
 $p < 0,05$

