

**Academia VacuVen**



**Procesamiento y análisis estadístico de  
datos para la investigación científica de  
las Ciencias de la Salud. (Curso Básico)**

**Alejandro Rísquez Parra  
Mariano Fernández Silano**

## Presentación



Dr. Alejandro Rísquez Parra

**Médico Pediatra, Magíster en Salud Pública: Epidemiología y Bioestadística; Servicios de Salud**  
**Profesor Asociado de la Catedra de Salud Pública de la Escuela de Medicina Luis Razetti, Facultad de Medicina. Universidad Central de Venezuela**  
**Jefe del Departamento de Medicina Preventiva y Social.**  
**Director del Centro de Vacunaciones de Venezuela. Vacuven**

**Médico Epidemiólogo, Doctor en Educación.**

**Profesor Asociado de la Catedra de Administración Sanitaria de la Escuela de Salud Pública, Facultad de Medicina. Universidad Central de Venezuela.**

**Coordinador de Investigación de la Facultad de Medicina**

**Jefe de la Unidad de Educación a Distancia de la Facultad de Medicina**



Dr. Mariano Fernández Silano

## BÁSICO de 8 horas académicas

**Duración:** 8 horas horarias distribuidas en dos sesiones de mañana y la tarde en sesiones de 4 horas cada sesión.

**Horario:** a convenir, con al menos un descanso intermedio de 1 hora.

**CERTIFICADO DE ASISTENCIA:** a los participantes que hayan completado el curso completo (mañana y tarde)

## Propósito

Brindar conocimientos básicos teórico-prácticos para el procesamiento y análisis de datos para investigación científica de las ciencias de la salud.

## Estrategia

Demostraciones teórico-prácticas presenciales con desarrollo de ejecución práctica de ejercicios reales de todo el contenido del programa.



## Objetivos

- ✓ Establecer el plan de proceso y análisis de datos.
- ✓ Comprender la relevancia del procesamiento y la depuración de los datos.
- ✓ Conocer y aprender a manejar archivos de datos básicos (Excel, Epi-info, Epidat, SPSS)
- ✓ Aplicar las pruebas estadísticas básicas descriptivas como las explicativas o inferenciales en paquetes estadísticos de datos tabulados y de bases de datos.
- ✓ Análisis e interpretación de los resultados de las pruebas estadísticas básicas descriptivas (medidas de tendencia central y de dispersión, correlación, regresión lineal) y las pruebas inferenciales (prueba Z, chi cuadrado, pruebas T, ANOVA) y las medidas de asociación e impacto (RR, OR, Hazzard, RA, NNT)



# Contenido:



## Primera sesión (4 horas)

- Introducción e importancia de las bioestadísticas
- Construcción del archivo de datos ( la base de datos)
- Variables
- Objetivos e hipótesis operativas y estadísticas
- Operacionalización de las variables
- Recolección de los datos y sus instrumentos (formularios, cuestionarios, encuestas, inspecciones, listas de cotejo, etc)
- Codificación y depuración de los datos.

## Segunda sesión (4 horas):

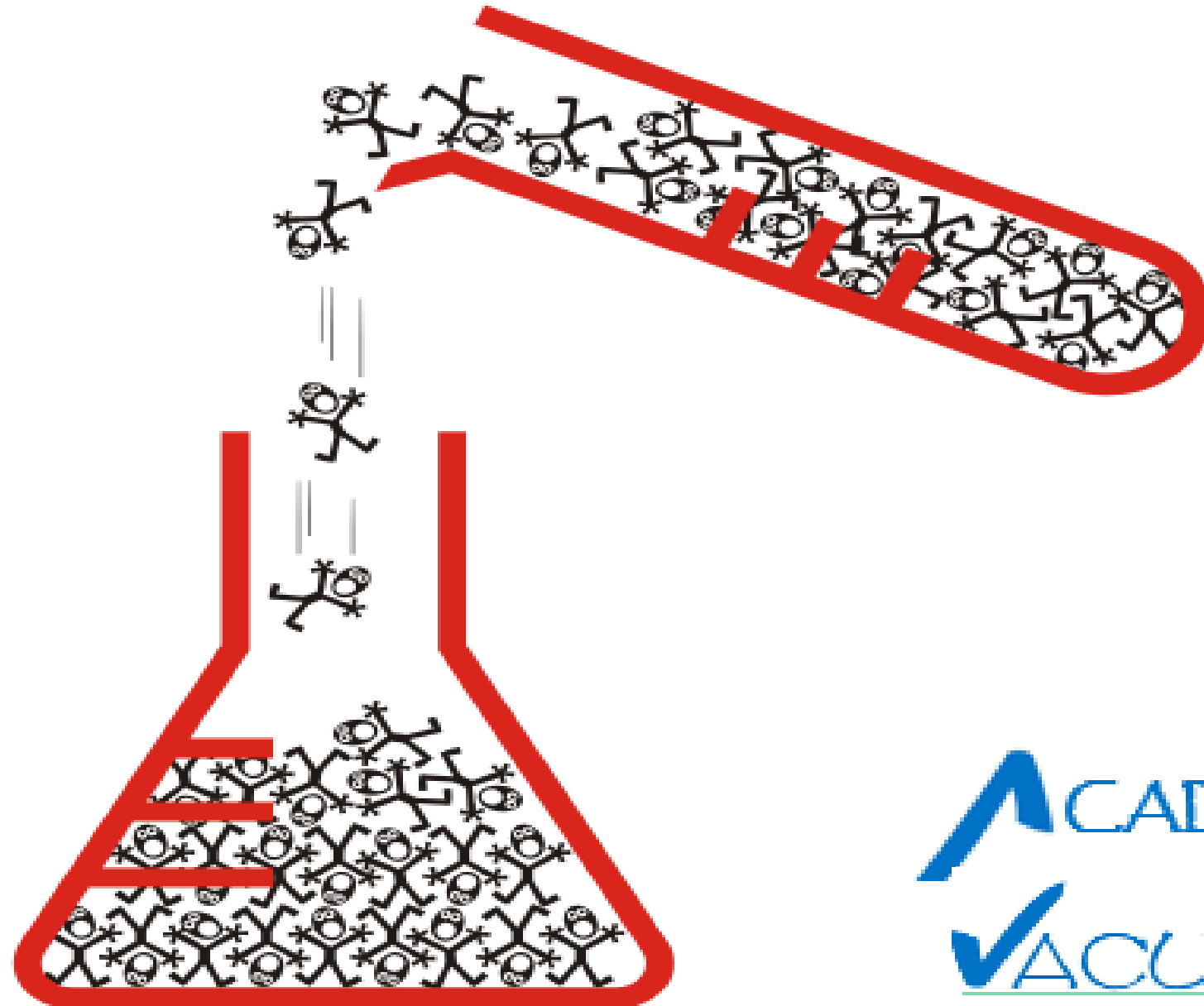
- Introducción a los paquetes estadísticos
- Excel, Epidat, Epiinfo, PSPP, SPSS
- Análisis exploratorio y preliminar.
- Análisis descriptivo:
  - Distribución de frecuencias (Binomial, Poisson y Curva Normal), medidas de tendencia central y de dispersión, cuadros, gráficos, y figuras. Correlación y regresión, pruebas de sensibilidad y especificidad, concordancia, análisis de supervivencia.
- Análisis univariado y bivariado.
- Análisis inferencial: Pruebas paramétricas y no paramétricas.
- Pruebas de comparación y de verificación de hipótesis.

## Sesión de cierre (20 minutos)

- Resumen del curso por facilitadores
- Intercambio de impresiones y preguntas y respuesta con participantes



Esperamos su  
mayor atención  
y que  
aprovechen al  
máximo su  
curso





## AGENDA INH

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

### Primera sesión :

1. Introducción e importancia de las bioestadísticas
2. Construcción del archivo de datos ( la base de datos)
3. Uso de bases de datos
4. Objetivos e hipótesis operativas y estadísticas
5. Operacionalización de las variables
6. Recolección de los datos y sus instrumentos (formularios, cuestionarios, encuestas, inspecciones, listas de cotejo, etc)

## AGENDA INH

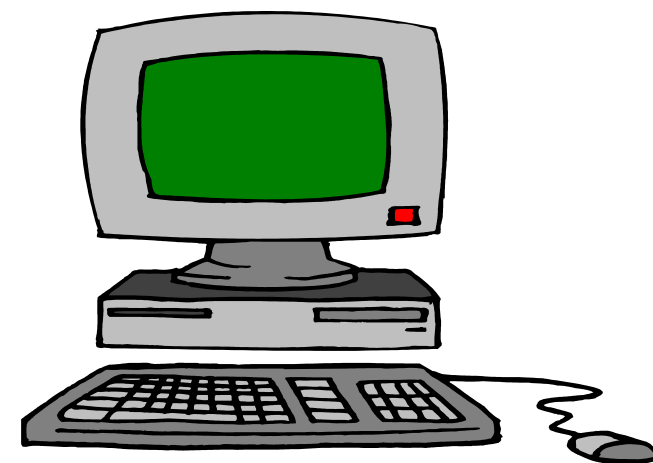
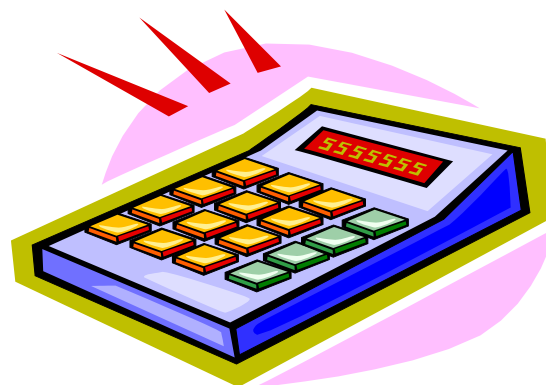
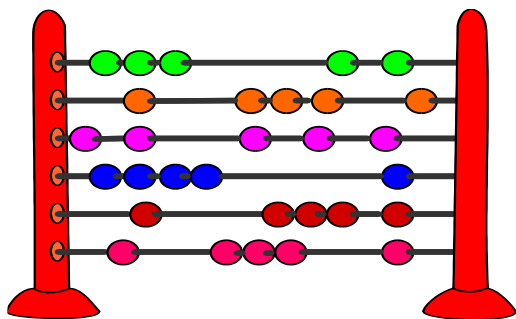
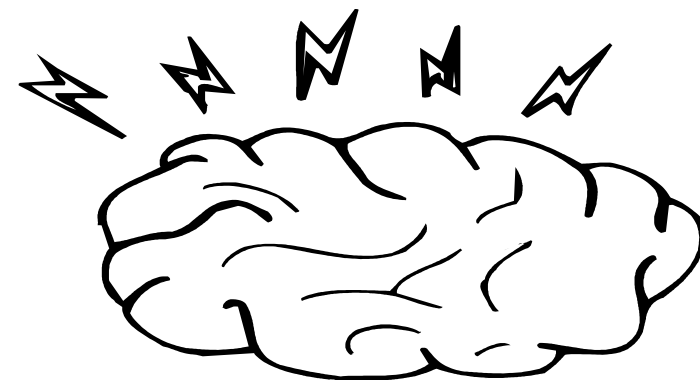
# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

## 1. Introducción e importancia de las bioestadísticas

Sin duda, es esencial que los investigadores en Ciencias de la Salud conozcan aspectos estadísticos fundamentales que les permitan llevar a cabo análisis básicos de una forma adecuada, interpretar correctamente los resultados y realizar una lectura crítica de los trabajos publicados por otros colegas.

# Estadística

...disciplina del estudio de los métodos científicos para la recolección, procesamiento, consolidación, reducción, presentación, análisis e interpretación de datos; y de hacer inferencias y sacar conclusiones de los datos numéricos.



# Ramas de la Estadística

## Estadística Descriptiva

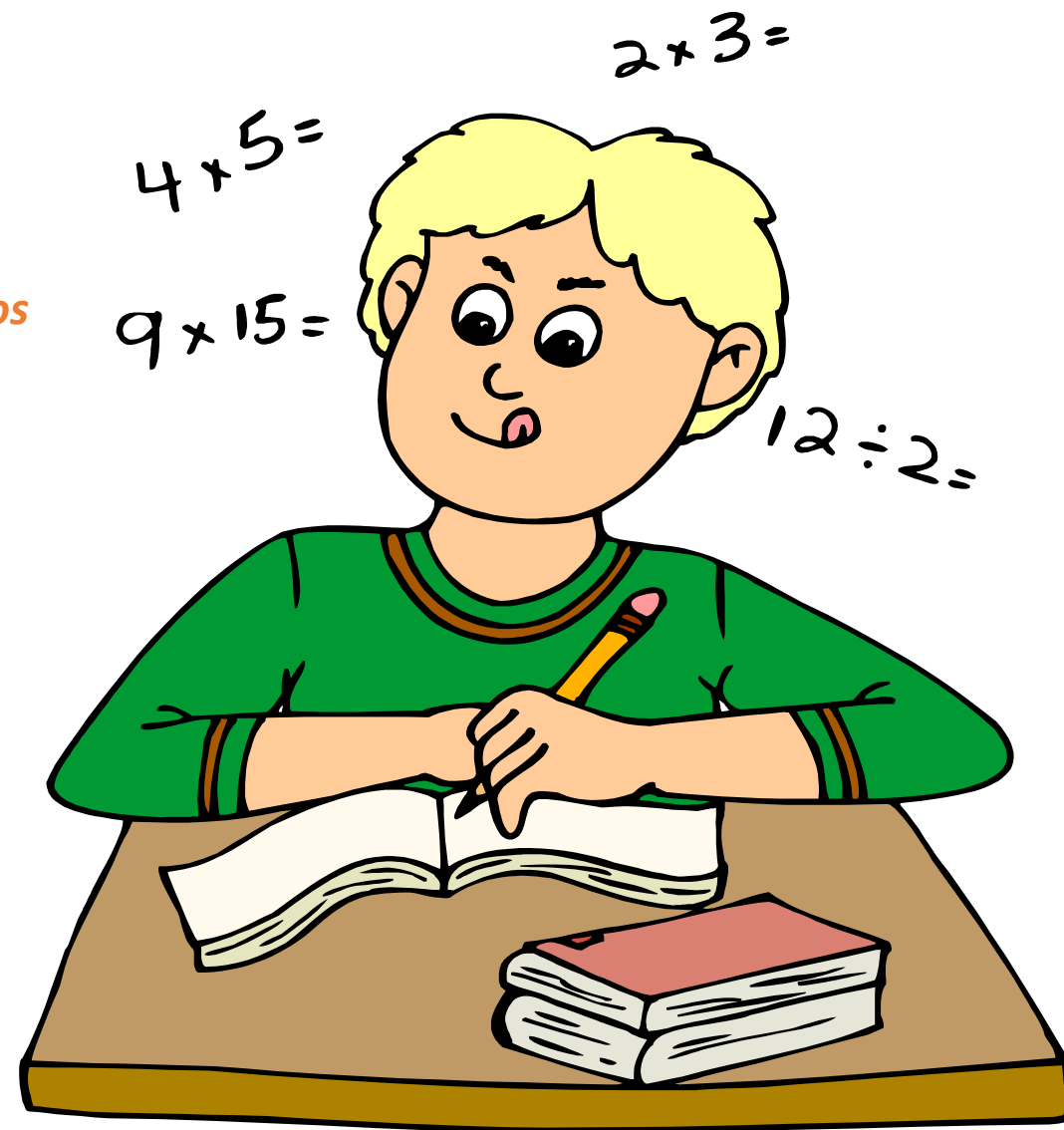
*Métodos para la organización, presentación y resumen de los datos*

## Estadística Vital

*Estadísticas derivadas del registro oficiales de demografía y la ocurrencia de las enfermedades notificables*

## Estadística Inductiva

*Métodos mediante el cual se generalizan los resultados de una muestra a la población de la cual fue obtenida*



# PRINCIPALES USOS DEL MÉTODO ESTADÍSTICO

## 1. Recolectar los datos de la mejor manera posible

- Diseñar formatos de recolección de datos*
- Organizar el procedimiento de recolección.*
- Diseñar y ejecutar la búsqueda.*
- Dirigir encuestas en la población.*

## 2. Describir las características de un grupo o situación

- Consolidación y tabulación de los datos.*
- Resumen de los datos.*
- Presentación de los datos.*

## 3. Analizar e interpretar los datos y sacar conclusiones

*Técnicas de análisis y el uso de conceptos probabilísticos.*



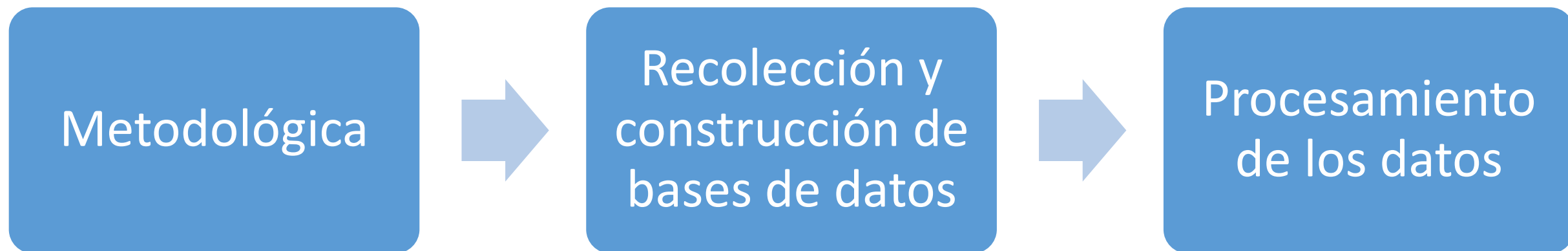
2. Construir una base de datos, para el análisis de los resultados de investigación científica.

Importancia.

Base de datos

Componentes

# El proceso de investigación



# Procesamiento de datos

- Luego de concluido la etapa metodológica y la de recolección de los datos
- Es el conjunto de métodos y procedimientos que permiten obtener información.



# Datos

- **Hecho mensurable (sujeto a una medición)**
- **Representa observaciones o hechos fuera de contexto y sin significado inmediato**
- **Representa la materia prima de la información**
- **Dentro de una organización, es el registro estructurado de una transacción**
- **Solamente indica qué es lo que ha pasado, sin aportar elementos para formarse un juicio de valor o para elaborar una interpretación o tomar una decisión**

# Datos

- **Un dato puede tener un valor escaso o nulo para un individuo en una situación concreta, pues por sí mismo no reduce la ignorancia o incertidumbre de quien tiene que tomar una decisión.**
- **La abundancia de datos no es necesariamente buena, porque puede dificultar la identificación de lo verdaderamente significativo**

# INFORMACION

- Es el significado que una persona asigna a un dato.
- Un dato se transforma en información cuando se lo procesa, es decir que es evaluado por un individuo concreto, que en un momento dado, trabaja sobre un problema para alcanzar un objetivo específico.-

# La base de datos o Archivo de datos

- Es una serie de ordenada de datos
- Productos de un instrumento de recolección: entrevista, cuestionario, lista de cotejo, registros oficiales, natalidad, etc..
- un archivo es un grupo de datos estructurados que son almacenados en algún medio y pueden ser usados por las aplicaciones.

# Estructura de la base de datos

Columnas, Variables, indicadores, cualidades, atributos.

Filas, casos, registros, sujetos muestrales

	A	B	C	D
1				
2				
3				
4				
5				
6				
7				
8				

## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

2. Utilización Instrumento de recolección de datos y construcción de las bases de datos. Se ejemplificara mediante los paquetes:

- google drive: formularios en línea
- Excel: hoja de cálculo extensamente utilizada
- SPSS: el paquete estadístico muy utilizado en ciencias sociales y médicas
- Epi info: programa epidemiológico libre, muy utilizado en los trabajos sobre riesgo

Epi Info  
Excel  
SPSS



# Los paquetes estadísticos

- conjunto de programas informáticos específicamente diseñados para el análisis estadístico de datos con el objetivo de resolver problemas de estadística descriptiva, inferencial o ambos.
- Existen muchos tipos y muy completos, adaptados a las múltiples necesidades.
- Lenguajes o entornos de programación
- Interfaz grafica
- Con manejadores de bases de datos o para datos tabulados
- Propietario o software libre



# Características

- Hoja de Cálculo
  - Gestor de Bases de Datos
  - Generador de Informes
  - Analizador de datos
  - Ejecutor de Minerías de Datos

# Lenguajes de programación

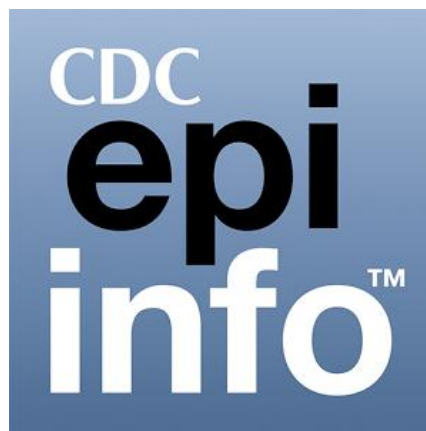


- Es un proyecto de software libre, resultado de la implementación GNU
- El ser un lenguaje de programación permite diseñar rutinas de análisis específicos diseñados por el analista.

# Interfaces graficas



# Epidemiológicos



# Ventajas

- En su mayoría son fáciles de aprender
- El manejo de datos es de fácil uso y facilita la gestión de los datos
- Los archivos de datos y los resultados pueden ser guardados en archivos y utilizados por diferentes programas
- Disponen de un amplio conjunto de métodos y procedimientos estadísticos.

- Base de prueba:

- Cedula
- Edad
- Sexo
- Profesión
- Peso

- Talla
- Edo civil
- Hijos
- Vacunas
- Color favorito
- Correo electrónico

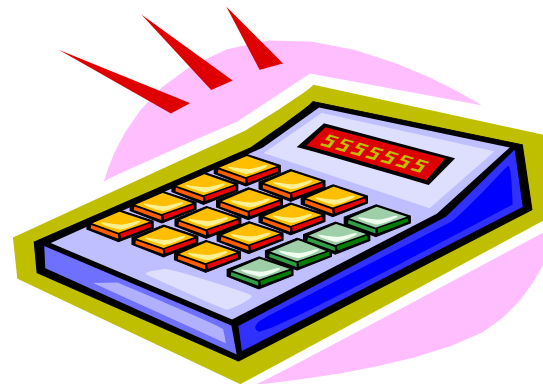
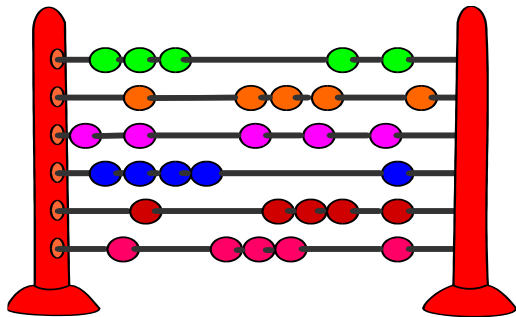
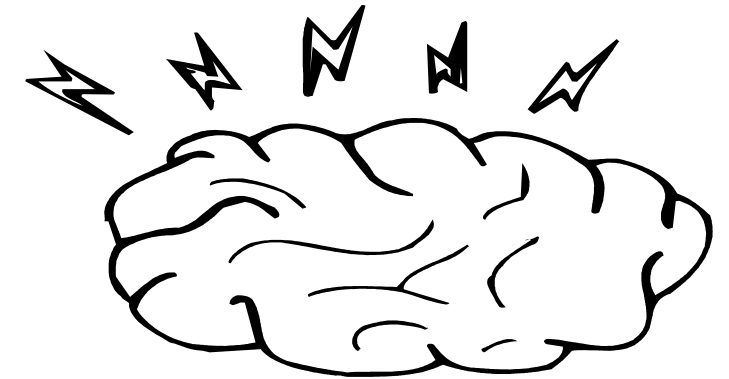
## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

3. Comprender los principales conceptos de la bioestadística:
- Diferentes tipos de estadística
  - Usos del método estadístico
  - Razonamiento epidemiológico

# Estadística

...disciplina del estudio de los métodos científicos para la recolección, procesamiento, consolidación, reducción, presentación, análisis e interpretación de datos; y de hacer inferencias y sacar conclusiones de los datos numéricos.





**Sin duda, es esencial que los investigadores en Ciencias de la Salud conozcan aspectos estadísticos fundamentales que les permitan llevar a cabo análisis básicos de una forma adecuada, interpretar correctamente los resultados y realizar una lectura crítica de los trabajos publicados por otros colegas.**

# El escándalo.

Ricardo Ocaña-Riola, Doctor en Ciencias Matemáticas, Profesor de Estadística - Escuela Andaluza de Salud Pública.

[http://www.divestadistica.es/es/2011\\_10/universo\\_estadistico\\_el\\_escandalo.html](http://www.divestadistica.es/es/2011_10/universo_estadistico_el_escandalo.html)

*¿Qué deberíamos pensar de un médico que usa el tratamiento equivocado, deliberadamente o por ignorancia, o que usa el tratamiento adecuado incorrectamente (como recetar la dosis errónea de un medicamento)?*

*La mayoría estaría de acuerdo en que este comportamiento sería poco profesional, carente de ética y por supuesto inaceptable.*

*Entonces, ¿qué deberíamos pensar de los investigadores que usan técnicas [estadísticas] equivocadas, deliberadamente o por ignorancia, usan las técnicas adecuadas incorrectamente, malinterpretan sus resultados, muestran solo parte de ellos, citan otras investigaciones selectivamente y dan conclusiones injustificadas?*

*Tendríamos que estar horrorizados. Aún hoy, numerosos estudios de la literatura médica, tanto de revistas generales como especializadas, han mostrado que todo lo descrito anteriormente es frecuente. Sin duda, un escándalo”*

# APLICACIONES DE LA ESTADÍSTICA

- ◆ Manejo adecuado de la variación
- ◆ Diagnóstico
- ◆ Pronóstico
- ◆ Selección del tratamiento
- ◆ Salud Pública y planificación
- ◆ Investigación médica

## Ramas de la Estadística

### Estadística Descriptiva

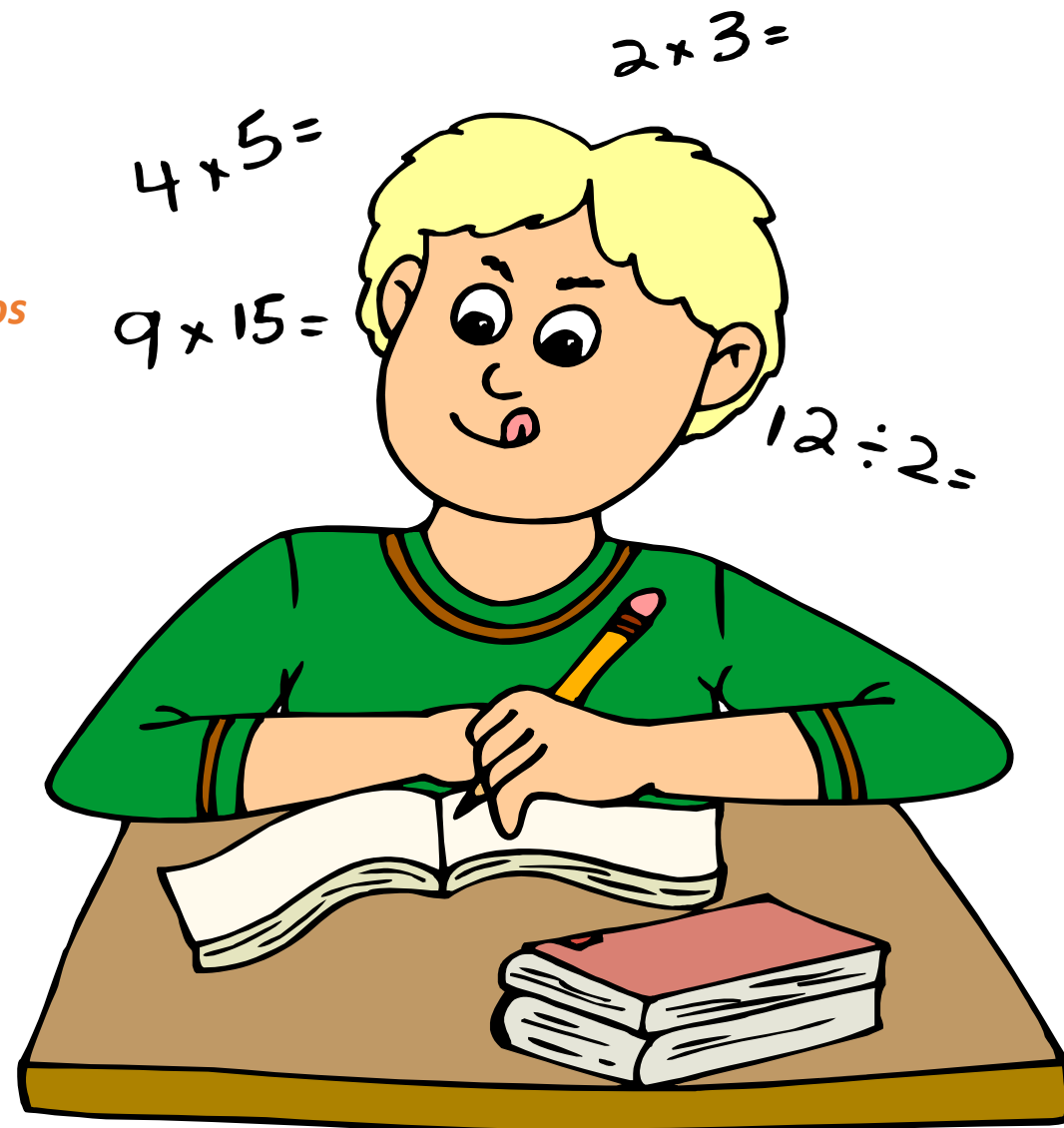
*Métodos para la organización, presentación y resumen de los datos*

### Estadística Vital

*Estadísticas derivadas del registro oficiales de demografía y la ocurrencia de las enfermedades notificables*

### Estadística Inductiva

*Métodos mediante el cual se generalizan los resultados de una muestra a la población de la cual fue obtenida*



## PRINCIPALES USOS DEL MÉTODO ESTADÍSTICO

### 1. Recolectar los datos de la mejor manera posible

- Diseñar formatos de recolección de datos*
- Organizar el procedimiento de recolección.*
- Diseñar y ejecutar la búsqueda.*
- Dirigir encuestas en la población.*

### 2. Describir las características de un grupo o situación

- Consolidación y tabulación de los datos.*
- Resumen de los datos.*
- Presentación de los datos.*

### 3. Analizar e interpretar los datos y sacar conclusiones

*Técnicas de análisis y el uso de conceptos probabilísticos.*





## **RAZONAMIENTO EPIDEMIOLÓGICO**

### **PROBLEMA**

Definición del área de investigación

El hecho o problema de salud debe ser medible y lo llaman daño.

### **HIPÓTESIS**

Se define como una suposición o conjetura que se plantea para explicar ciertos hechos o eventos y se emplea como base para llevar a cabo una investigación mediante la cual se busca demostrarla o refutarla.

### **GRUPO DE COMPARACIÓN**

Selección del diseño del estudio.

### **RECOLECCIÓN SISTEMÁTICA DE LA DATA**

## **ANÁLISIS ESTADÍSTICO**

**EVALUAR LA VALIDEZ DE UNA ASOCIACIÓN  
ESTADÍSTICA**

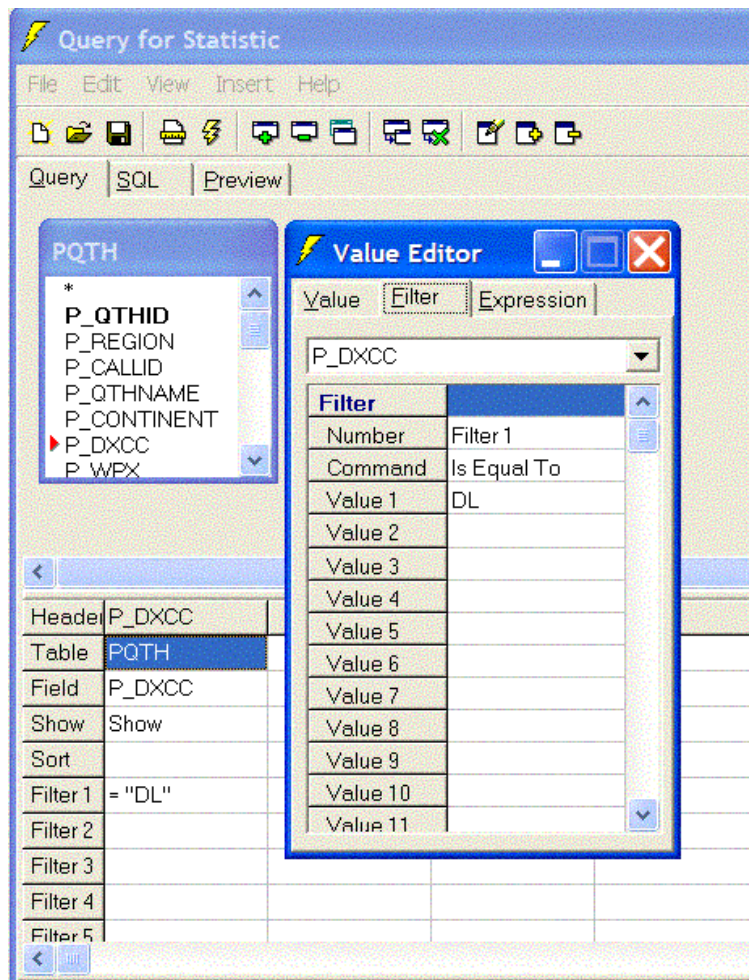
**EXCLUYENDO EL AZAR, SESGO Y LAS VARIABLES DE  
CONFUSIÓN**

**JUZGAR SI LA ASOCIACIÓN ES LÓGICA, SI TIENE LA  
MAGNITUD, LA CONSISTENCIA Y LA CREDIBILIDAD  
BIOLÓGICA.**

## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

4. Utiliza Excel para realizar cálculos generales.
  - Distingue observaciones, Variables e indicadores,
  - Utiliza cálculos y funciones en hoja de calculo



## TIPOS DE VARIABLES

- ◆ **NOMINALES**
- ◆ **dicotómicas**
- ◆ **multicotómicas**

- ◆ **ORDINALES**

- ◆ **NUMERICAS**

- ◆ **Discretas**
- ◆ **Intervalo de razón**



# VARIABLES NOMINALES

## Dicotómicas (binarias)

- ♦ Vivo o muerto
- ♦ Varón o hembra
- ♦ Presente o ausente

## Politómicas

- ♦ Soltero o casado o viudo o divorciado o unido
- ♦ Católico, budista, protestante, judío u otra

**Codificar como numéricas para facilitar su manejo**

# VARIABLES ORDINALES

**Leve, moderado o grave**

**Estadios I, II y III (cáncer)**

**Clases I, II, III y IV**

**(New York Heart Ass. Functional Classification)**

# **CLASIFICACIÓN FUNCIONAL DE LAS CARDIOPATÍAS (N.Y.H.A.)**

**CLASE I. Asintomático**

**CLASE II. Síntomas en actividad ordinaria**

**CLASE III. Síntomas en actividad < a la ordinaria**

**CLASE IV. Síntomas en reposo**

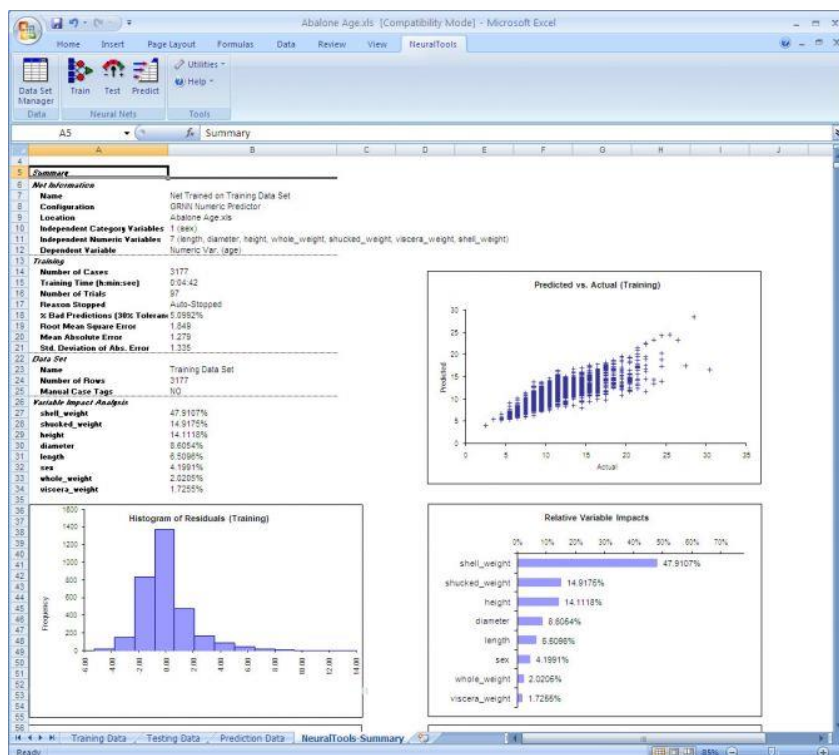
# VARIABLES NUMÉRICAS (dimensionales)

## DISCRETAS

- ◆ Número de hijos
- ◆ Número de admisiones

## CONTINUAS

- ◆ Edad
- ◆ Días de hospitalización
- ◆ Colesterol sérico
- ◆ Tensión arterial



# ÍNDICE

## Puntuación de Apgar

- ◆ Frecuencia cardíaca
- ◆ Frecuencia respiratoria
- ◆ Color
- ◆ Tono muscular
- ◆ Reflejo nasal al cateter

## Estado neurológico de Glasgow

## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

5. Describe las características de una población o muestra a través de las medidas estadísticas: Realización del análisis descriptivo de los datos:

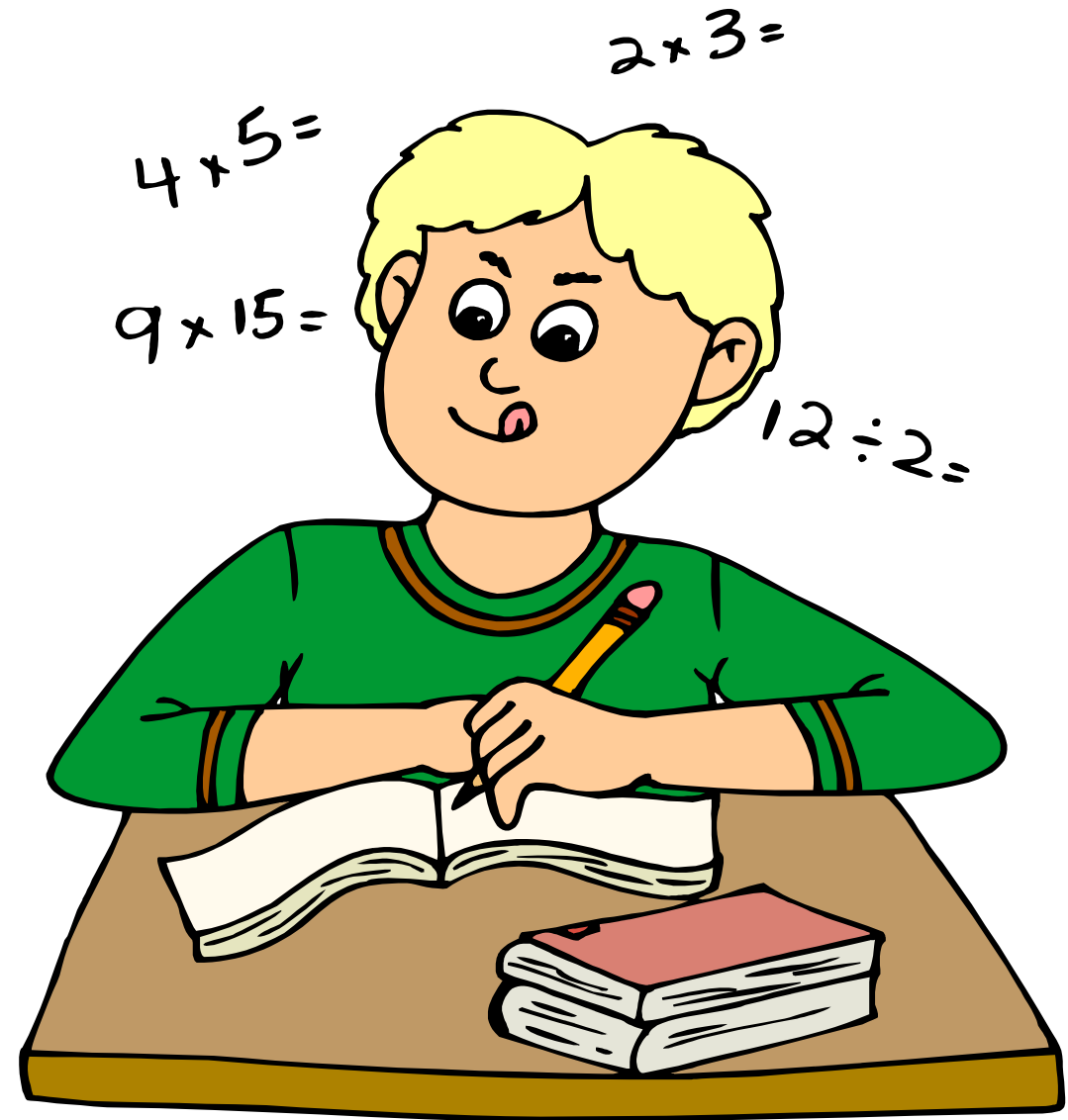
Técnicas de depuración de la información

Tabulación y graficación

Medidas de tendencia central y de dispersión.

**ANÁLISIS DESCRIPTIVO**

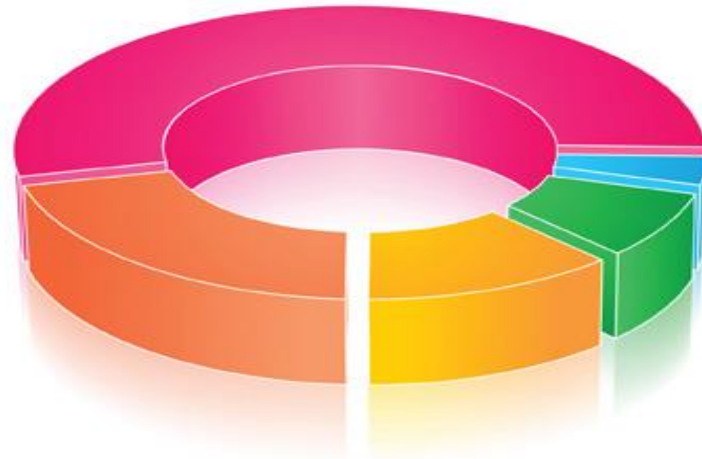
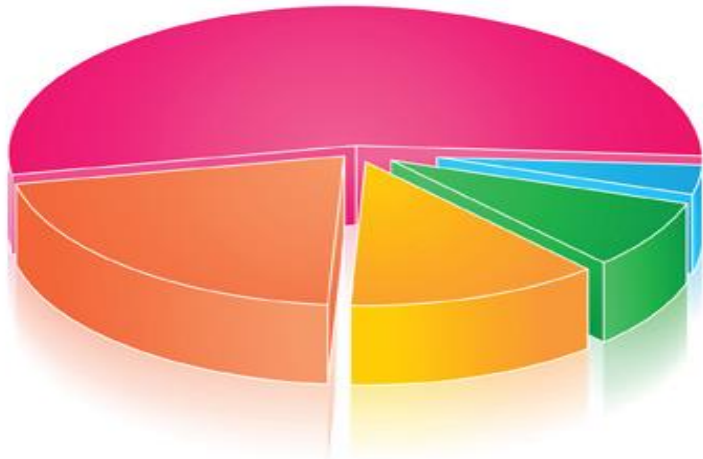
**ESTADÍSTICA DESCRIPTIVA**



# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)





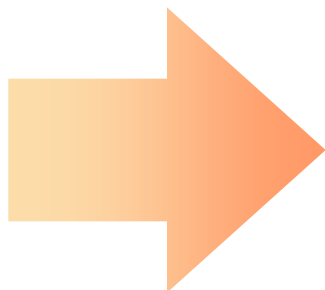


## **MEDIDAS DE TENDENCIA CENTRAL**

- ◆ Promedio
- ◆ Mediana
- ◆ Moda

## **MEDIDAS DE DISPERSIÓN**

- ◆ Desviación Estandar - varianza
- ◆ Percentiles - Cuartiles
- ◆ Recorrido intercuartil



**Variabilidad de los grupos**

## MEDIDAS DE TENDENCIA CENTRAL

- ◆ Promedio
- ◆ Mediana
- ◆ Moda

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

# MEDIDAS DE DISPERSIÓN

- ◆ Desviación Estandar - varianza
- ◆ Percentiles - Cuartiles
- ◆ Recorrido intercuartil

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

# CURVA DE NORMALIDAD

$$Z = \frac{X - \bar{X}}{\sigma}$$

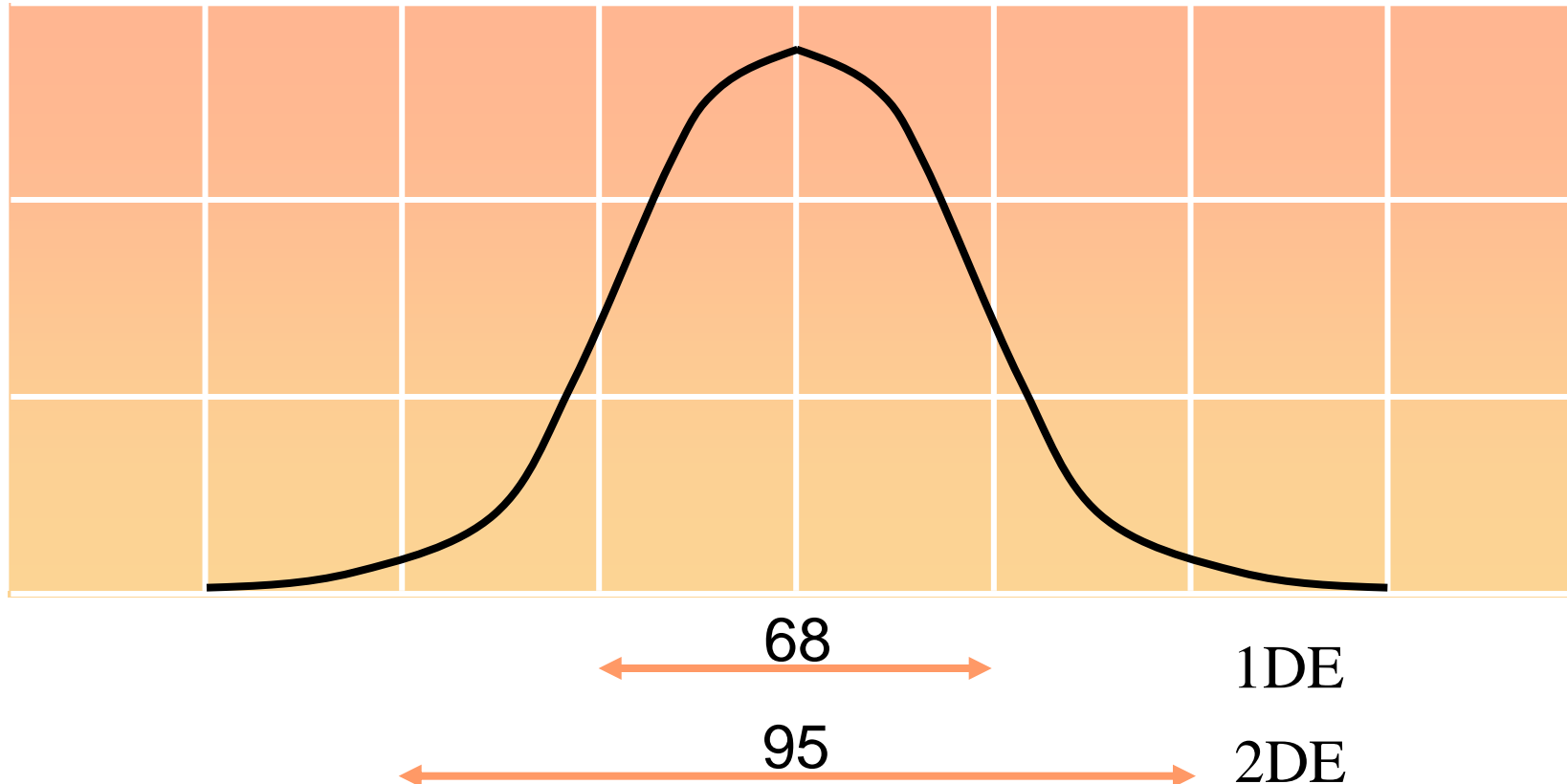
Donde:

Z = valor estadístico de la curva normal de frecuencias.

X = cualquier valor de una muestra estadística.

$\bar{X}$  = promedio o media aritmética obtenido de la muestra estadística, valor representativo.

$\sigma$  = desviación estándar.



## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

5. Aplica análisis descriptivo de correlación y regresión lineal simple.

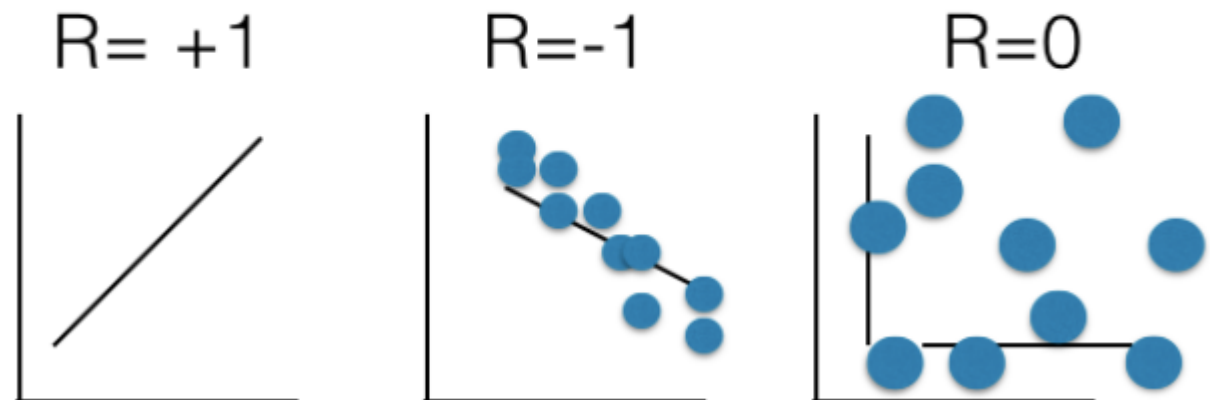
6. Utiliza análisis concordancia y sensibilidad y especificidad

( COEFICIENTE CORRELACIÓN DE PEARSON/ SPEARMAN)

## Características

1. Mide la fuerza de la asociación. Magnitud
2. Mide la dirección : sentido signo + o -
3. Sus valores oscilan entre de -1 hasta +1 ( -1,0,+1)
4. No tiene unidades , solo se expresa el valor del Coeficiente R

## Posibilidades de Correlación

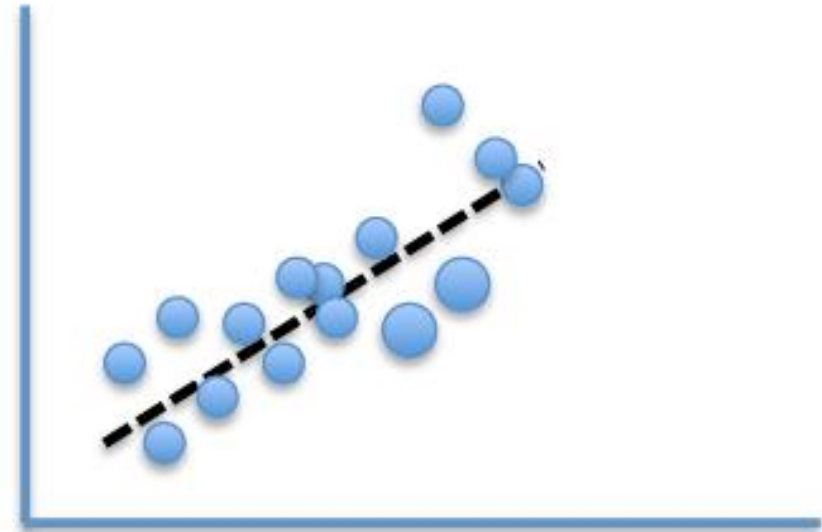


## Regresión (REGRESIÓN SIMPLE)

Meta: derivar una ecuación (la mejor ecuación posible) que me permita explicar los puntos que tengo, uno de las ideas principales es predecir valores que no tengo a priori .

La formula

$$Y=mX+b$$



La mayoría de Las veces el Calculo se hace Por la técnica de Los mínimos cuadrados.

Que valores obtengo de la regresión:

1.  $m$ =pendiente (inclinación varia entre -1 y +1)
2.  $X$ =variable independiente  $Y$  = dependiente
3.  $b$ = Corte de la recta al eje Vertical
4.  $p$ =significancia que se interpreta como la probabilidad de  $m=0$  y tiene una distribución (t) y  $df=n-2$



**Tabla 2. Datos hipotéticos de clasificación de una muestra de 100 radiografías por dos radiólogos**

Radiólogo 1	Radiólogo 2			Total
	Anormal	Dudosa	Normal	
Anormal	18	4	3	25
Dudosa	1	10	5	16
Normal	2	4	53	59
Total	21	18	61	100

**Valoración del Índice Kappa**

Valor de k	Fuerza de la concordancia
< 0.20	Pobre
0.21 – 0.40	Débil
0.41 – 0.60	Moderada
0.61 – 0.80	Buena
0.81 – 1.00	Muy buena

**Resultados de la exploración y biopsia prostática de una muestra de pacientes con sospecha de cáncer de próstata.**

Resultado del tacto rectal	Resultado de la biopsia prostática		
	Cáncer	Patología benigna	Total
Anormal	634	269	903
Normal	487	1251	1738
Total	1121	1520	2641

$$\text{Sensibilidad} = \frac{634}{634+487} = \frac{634}{1121} = 0,5656 \Rightarrow 56,56\%$$

$$\text{Especificidad} = \frac{1251}{269+1251} = \frac{1251}{1520} = 0,8230 \Rightarrow 82,30\%$$

$$\text{Valor predictivo positivo} = \frac{634}{634+269} = \frac{634}{903} = 0,7021 \Rightarrow 70,21\%$$

$$\text{Valor predictivo negativo} = \frac{1251}{487+1251} = \frac{1251}{1738} = 0,7198 \Rightarrow 71,98\%$$

$$\text{Razón de verosimilitud positiva} = \frac{\text{Sensibilidad}}{1-\text{Especificidad}} = \frac{0,5656}{1-0,8230} = 3,19$$

$$\text{Razón de verosimilitud negativa} = \frac{1-\text{Sensibilidad}}{\text{Especificidad}} = \frac{1-0,5656}{0,8230} = 0,53$$

## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

8. Describe las asociaciones entre las variables y aplica prueba de las hipótesis encontradas: Análisis estadístico inferencial.

Generalidades, universo y muestra

Prueba de hipótesis

Nociones para la selección de la prueba estadística

Pruebas más usadas:

Z curva normal,

pruebas T,

análisis de varianza (ANOVA),

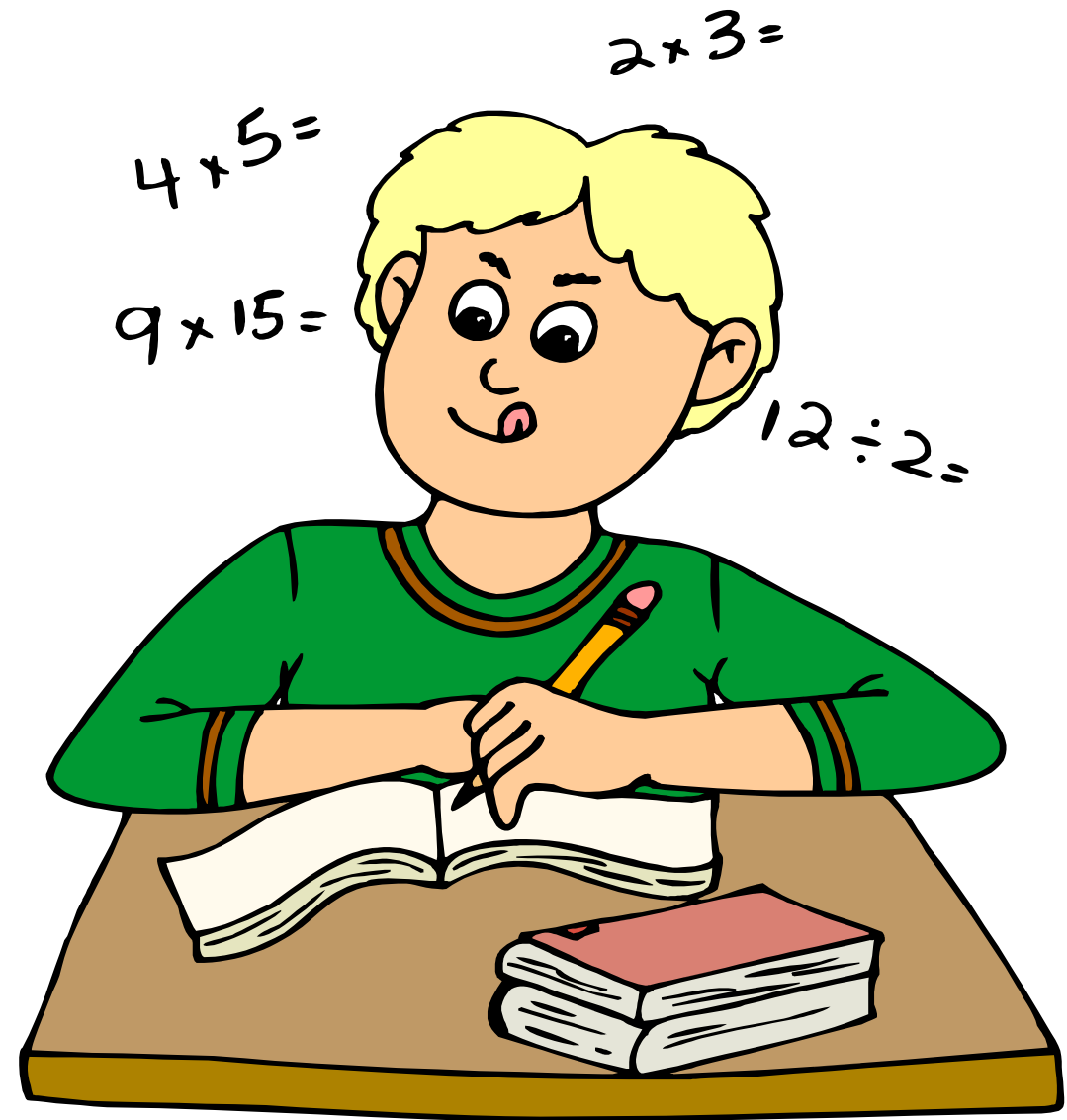
chi cuadrado,

medidas de asociación de riesgo: OR y RR



# ANÁLISIS INFERENCIAL

## ESTADÍSTICA INFERENCIAL



# INFERENCIA Y GENERALIZACIÓN

Universo y muestra

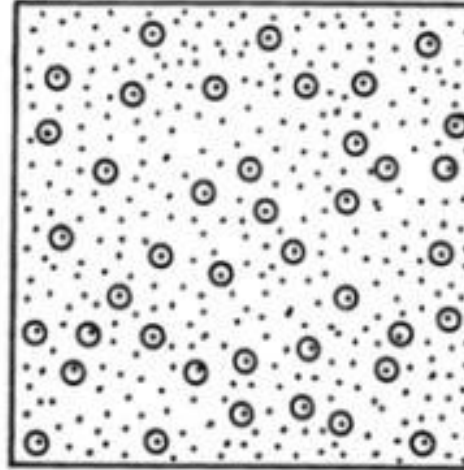
Población Objeto

Población a muestrear

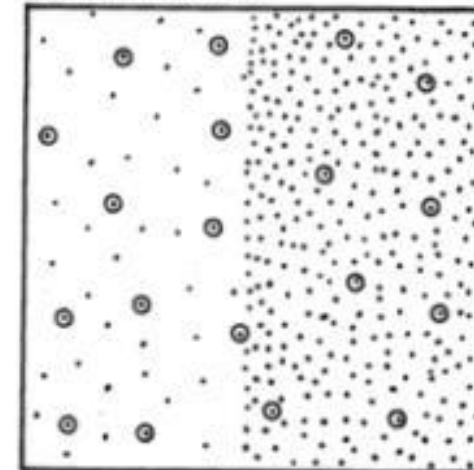
Muestra

# TÉCNICAS DE MUESTREO

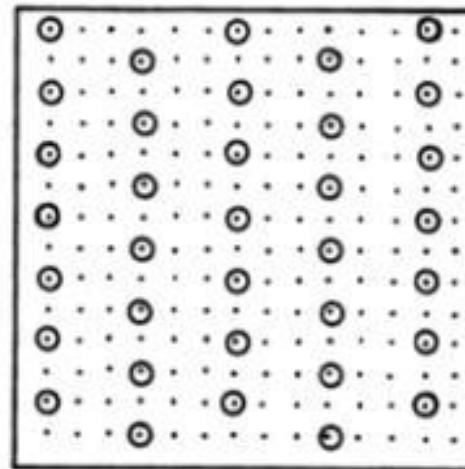
MUESTREO ALEATORIO SIMPLE



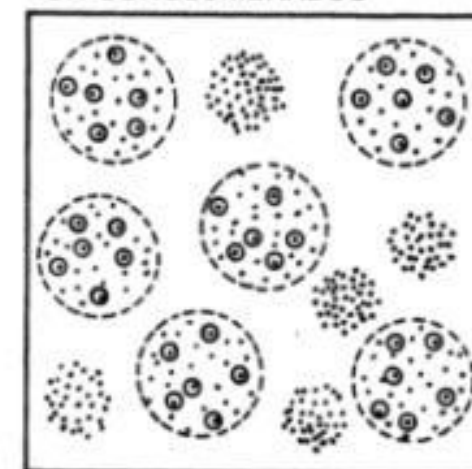
MUESTREO ALEATORIO ESTRATIFICADO

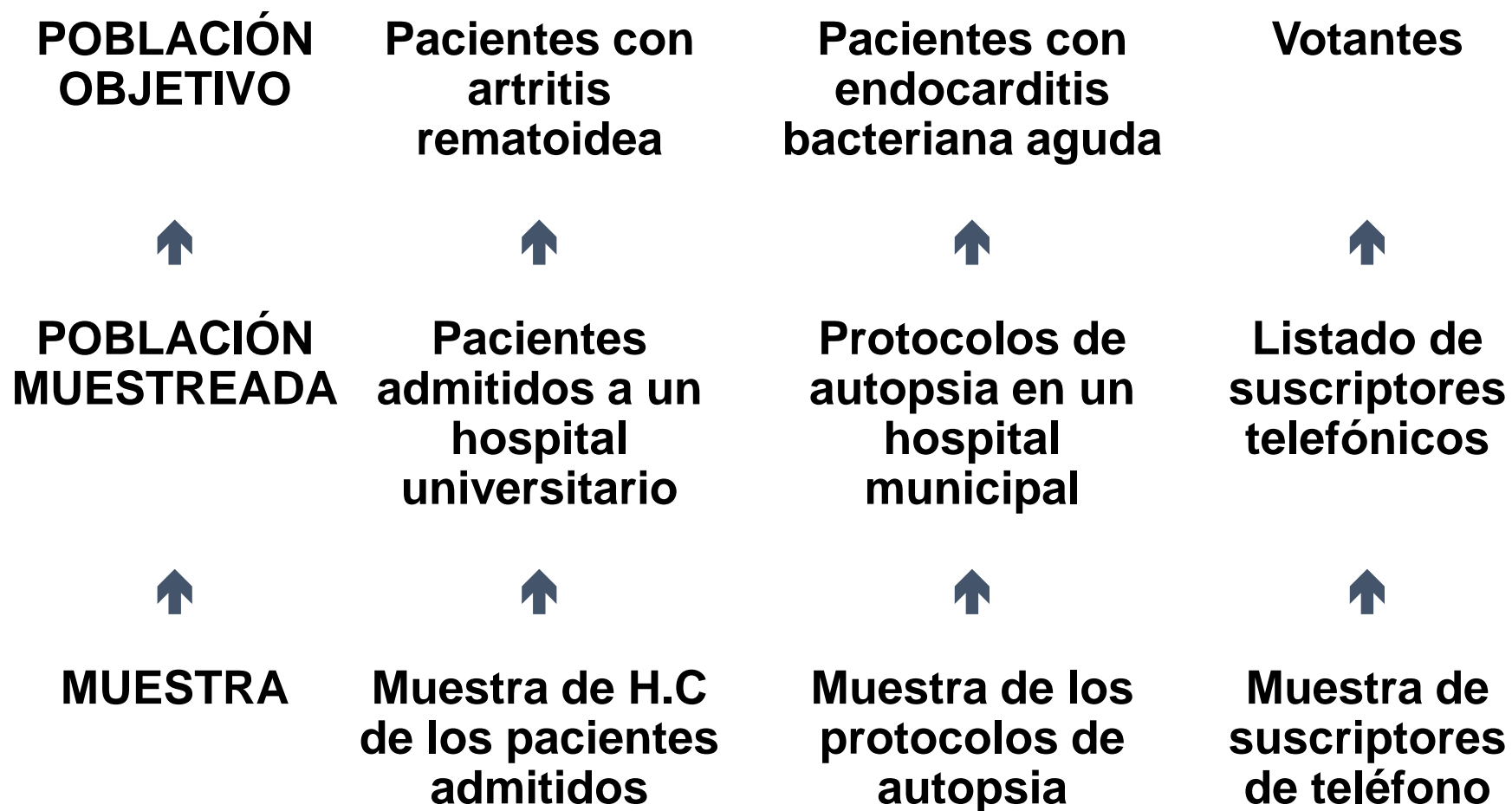


MUESTREO SISTEMÁTICO



MUESTREO POR CONGLOMERADOS





## **CARACTERÍSTICAS DE LOS PACIENTES CON ENDOCARDITIS BACTERIANA**

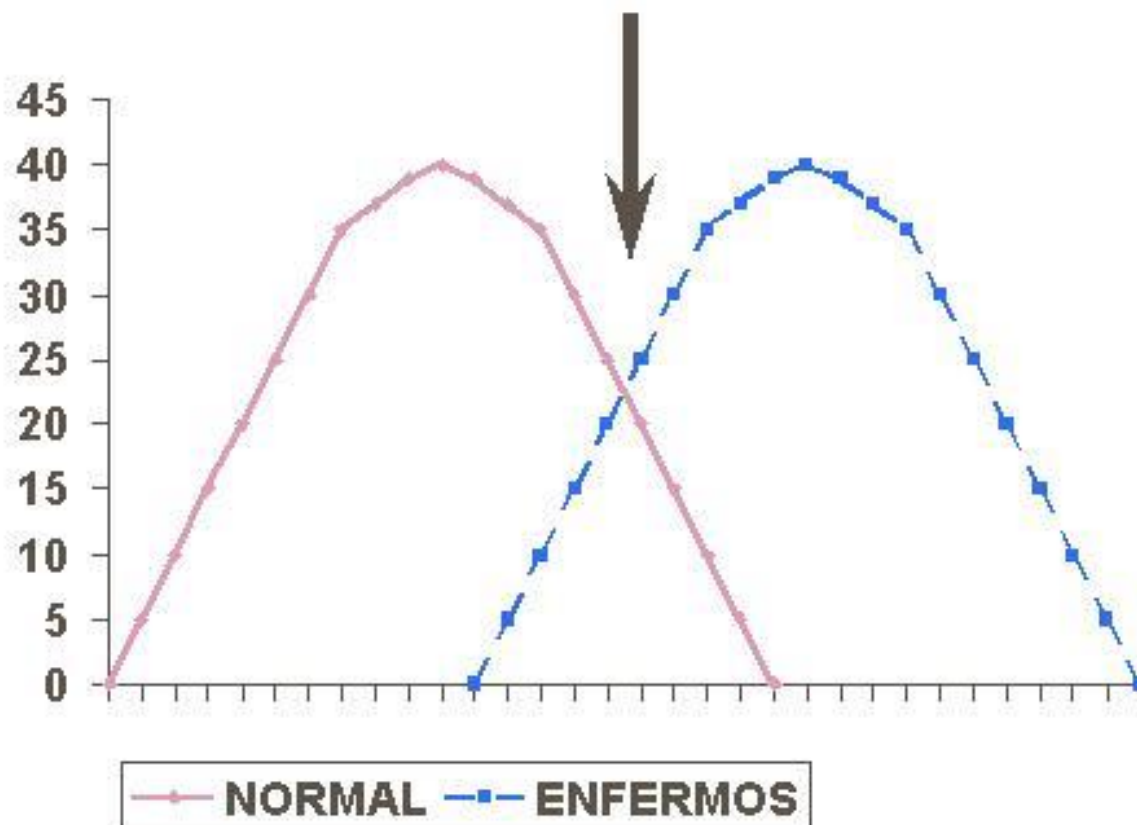
Característica	96	13	50
	Pacientes autopsiados	Muertes S/Autopsia	Pacientes egresados
Edad promedio (a)	56	51	36
De 65 y más años (%)	60	31	6
Enfermedad asociada (%)	60	51	21
Alteraciones mentales (%)	43	47	22
Esplenomegalia (%)	12	15	38
Murmullos (%)	66	100	100
Petequias (%)	35	38	54

*Datos de Cooper y otros*

*Tomado de: Colton Th. Statistics in Medicine. Little Brown and Company (Inc.) Boston, 1974; p 6.*



# NORMALIDAD PUNTOS DE CORTE



Un estadístico es una persona que teniendo su cabeza en el horno y sus pies en un balde de hielo, cuando se le pregunta como se siente, responde:  
**“En promedio, me siento bien”**

# **MEDIDAS DE ASOCIACIÓN Y DE IMPACTO**

- ◆ **RIESGO RELATIVO**
- ◆ **RAZÓN DE PRODUCTOS CRUZADOS**
- ◆ **RIESGO ATRIBUIBLE**

# EVALUANDO EL AZAR

- ◆ Verificación de hipótesis por pruebas estadísticas
  - ◆ Valor de “ $\rho$ ” probabilidad ( $< 0.05$ )
  - ◆ Intervalos de Confianza
- ◆ Tamaño de la muestra
- ◆ Poder del estudio ( $1 - \beta$ )

# ANÁLISIS ESTADÍSTICO

El AZAR en la jerga estadística es

*... una probabilidad muy pequeña para ser explicada por pura casualidad, suerte o chance*



**Tests estadísticos de significancia (“ $\rho$ ”)  
+ intervalo de confianza**

# ANÁLISIS ESTADÍSTICO

Las pruebas estadísticas SIRVEN para verificar o rechazar hipótesis:

**Hipótesis Nula (H0):** La droga no tiene efecto  
**Hipótesis Alternativa (H1):** La droga reduce la mortalidad

## Hipótesis Nula (H0)

No hay diferencia (“son  $\neq$ ”)

## Hipótesis Alternativa (H1) =

- ♦ de 2 extremos (“ $>$  o  $<$ ”)
- ♦ de 1 extremo (“ $>$  que” o “ $<$  que”)

# PROBABILIDADES DE LOS RESULTADOS PARA HIPÓTESIS

## POBLACIÓN MUESTREADA

	H <sub>0</sub> es falsa	H <sub>0</sub> es verdadera
Significante	$1 - \beta$ (Poder)	$\alpha$ (error tipo I)
No Significante	$\beta$ (error tipo II)	$1 - \alpha$ (confiabilidad)

# LOS ERRORES $\alpha$ Y $\beta$

Había una vez un Rey que era muy celoso con su Reina.

El tenía dos caballeros,  $\alpha$  quien era muy buenmozo y,  $\beta$  quien era muy feo. La reina estaba de amores con  $\beta$ .

Sin embargo, el rey sospechaba que la reina estaba traicionándolo con  $\alpha$  e hizo que lo decapitaran.

Así, el rey cometió dos tipos de errores; él sospechó de una relación (con  $\alpha$ ) que no existía y falló en detectar una relación (con  $\beta$ ) donde si la había en realidad.



Además de exigir a los laboratorios que las nuevas drogas no hagan daño, la FDA (Federal Drug Administration) está más interesada en disminuir la posibilidad de hacer errores  $\alpha$  con lo que evita las “drogas inútiles”.

Los laboratorios están interesados en disminuir el error  $\beta$  para poder así mercadear drogas potencialmente beneficiosas

# ANÁLISIS ESTADÍSTICO

Significancia estadística



Significancia clínica o práctica

Significancia estadística

NO

SI

Significancia Biológica

NO

Caso I

Caso II

SI

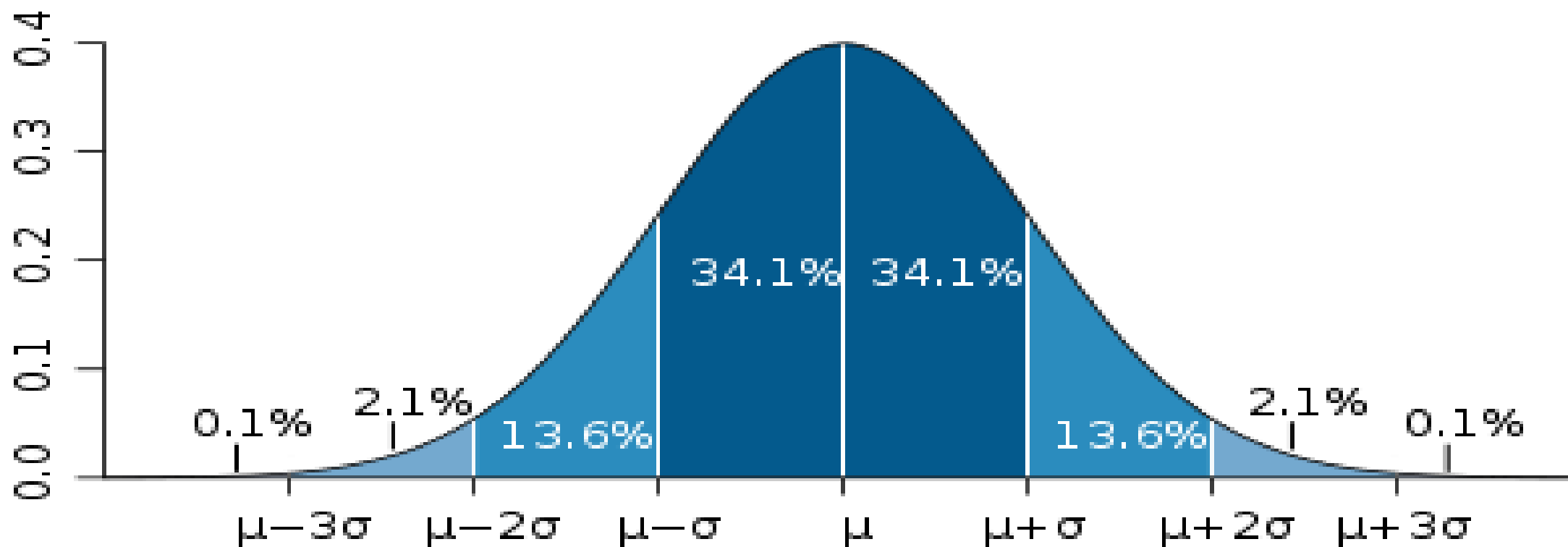
Caso III

Caso IV

*\* los casos II y III son problema.*

La **distribución normal estándar, o tipificada o reducida**, es aquella que tiene por media el valor cero,  $\mu = 0$ , y por desviación típica la unidad,  $\sigma = 1$ .

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



# PRUEBA Z

$$Z = \frac{X - \bar{X}}{\sigma}$$

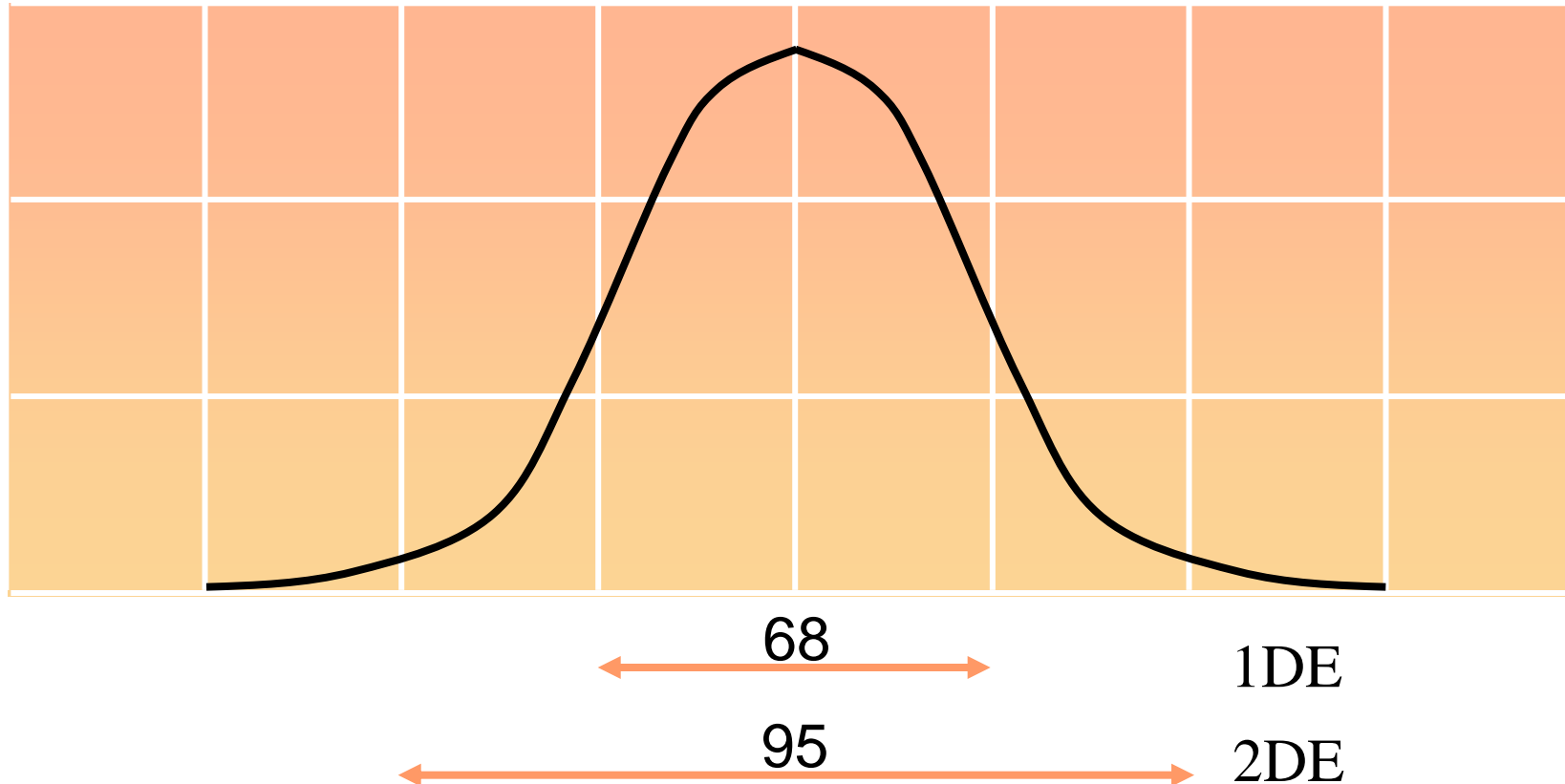
Donde:

Z = valor estadístico de la curva normal de frecuencias.

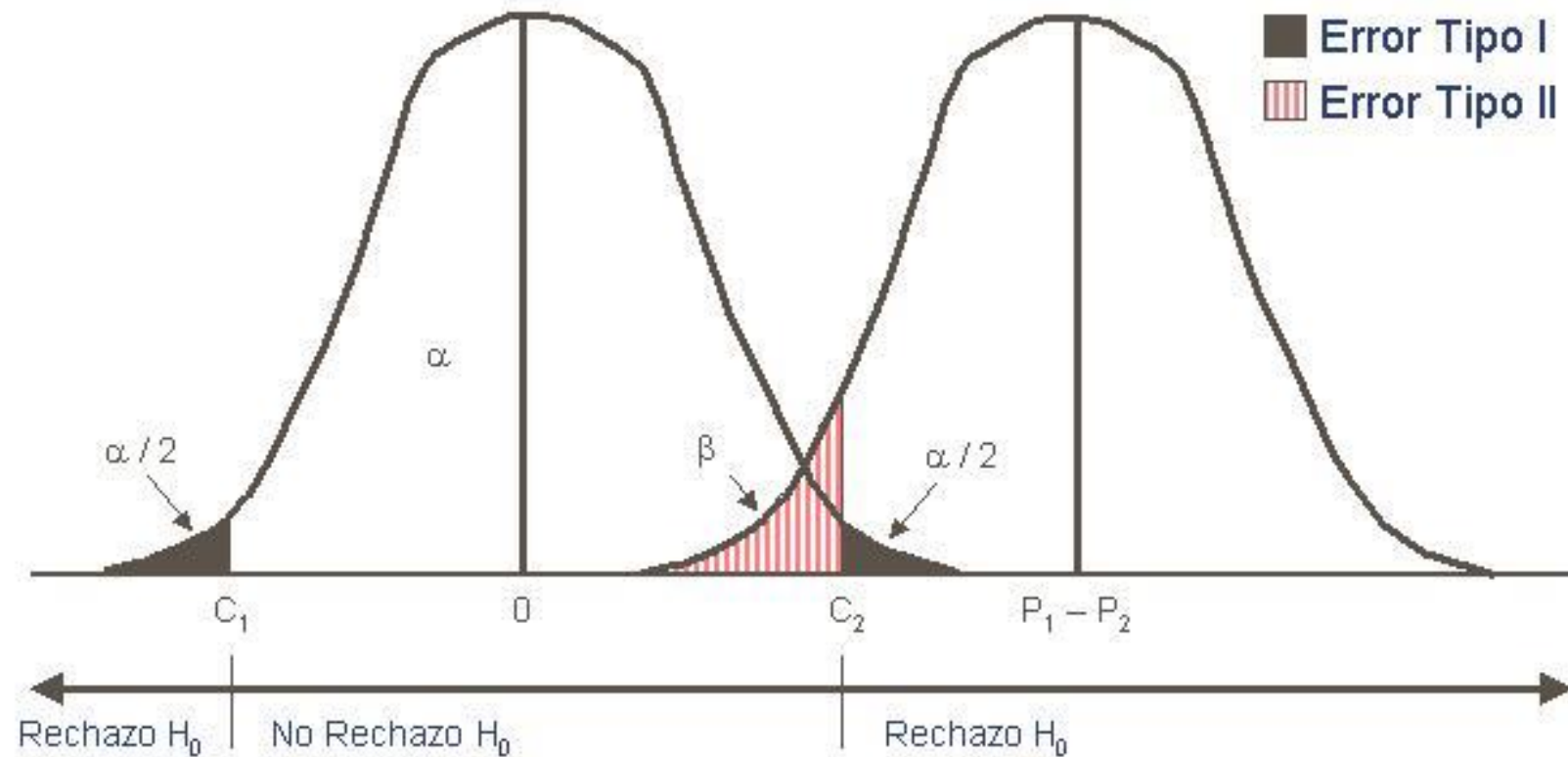
X = cualquier valor de una muestra estadística.

$\bar{X}$  = promedio o media aritmética obtenido de la muestra estadística, valor representativo.

$\sigma$  = desviación estándar.



## PROBABILIDADES DE LOS RESULTADOS POSIBLES EN ENSAYOS DE HIPÓTESIS, MOSTRANDO ERRORES $\alpha$ Y $\beta$



# Decisión

**Región de aceptación**

Intervalo en el que se acepta  $H_0$

**Región de rechazo**

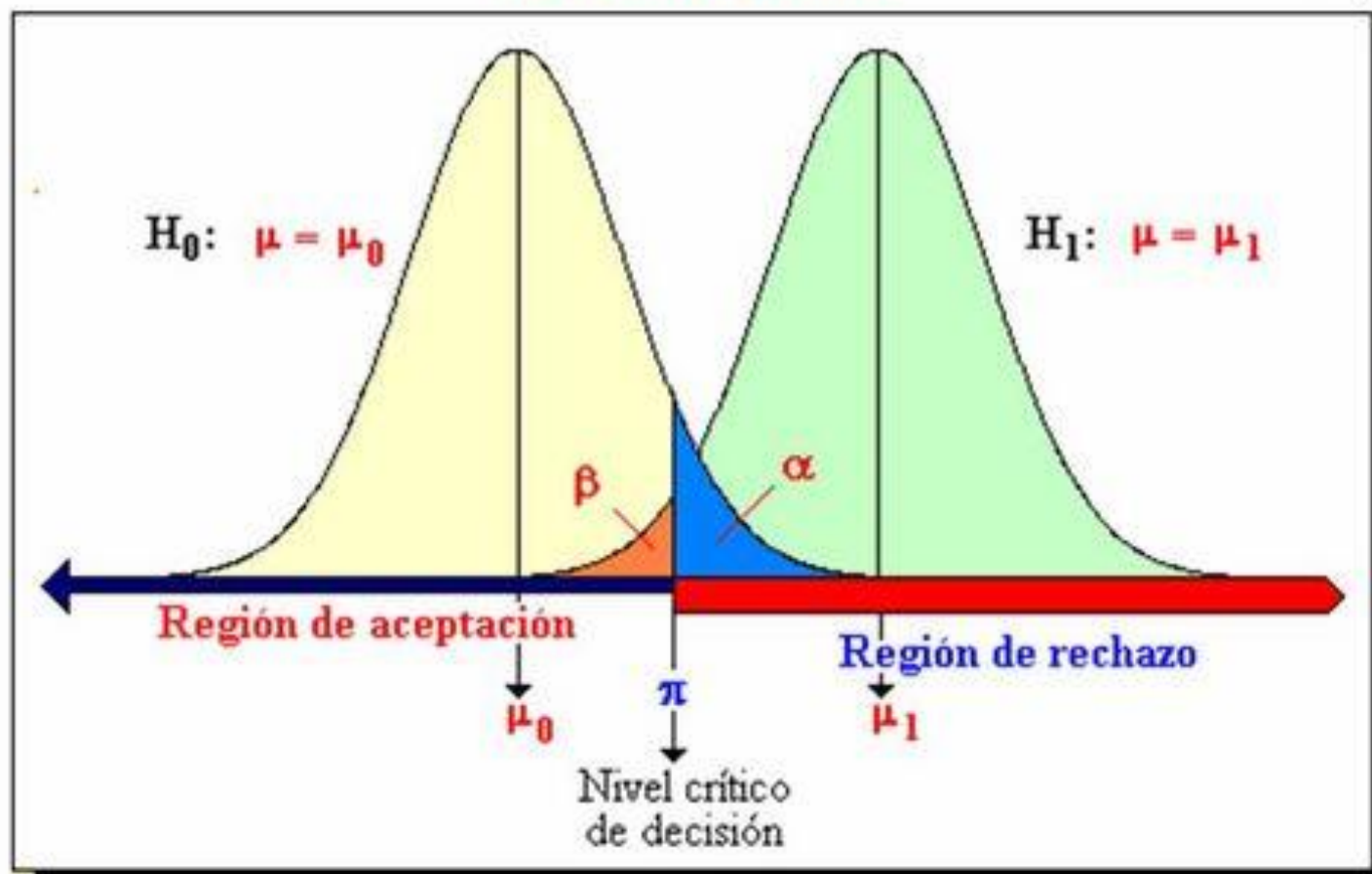
Intervalo en el que no se acepta  $H_0$

**Nivel(es) crítico(s) de decisión**

Extremos de la región de aceptación

**Función de decisión**

Estadístico que permite resolver el contraste



**Error de tipo I:**

Rechazar  $H_0$  siendo cierta  
(Nivel de significación)  
 $\alpha = \text{Prob}(\text{Error I})$

**Error de tipo II:**

Aceptar  $H_0$  siendo falsa  
 $\beta = \text{Prob}(\text{Error II})$

**Potencia del contraste:**

$$\eta = 1 - \beta$$

# Docimasia de hipótesis

- Planteamiento de la Hipótesis nula ( $H_0$ ) y alternativa ( $H_1$ )
- Seleccionar el nivel de significancia (95%)
- Elección del estadístico de prueba ( $Z, t, F...$ )
- Se formula la regla de decisión o se determina en el Software la prueba apropiada
  - Acepta  $H_0$  si  $p > 0,05$
  - Rechazas  $H_0$  si  $p < 0,05$
- Se toma una muestra , se realiza el calculo y se decide

## ANALISIS ESTADÍSTICO

El AZAR en la jerga estadística es

*... una probabilidad muy pequeña para ser explicada por pura casualidad, suerte o chance*

### Hipótesis Nula ( $H_0$ )

No hay diferencia (“son iguales”)

### Hipótesis Alternativa ( $H_1$ )

- de 2 extremos (“mayor o menor”)
- de 1 extremo (“mayor que” o “menor que”)

### Nivel de significancia 0,05

- si es igual o menor se considera probabilidad pequeña
- si es mayor se considera probabilidad grande





# ANALISIS ESTADÍSTICO

## Probabilidades de los resultados en ensayos de hipótesis

Población muestreada

H0 es falsa

H0 es verdadera

Significante

**1 - beta**  
**(Poder)**

**alfa**  
**(error tipo I)**

No  
Significante

**beta**  
**(error tipo II)**

**1 - alfa**  
**(confiabilidad)**

	H0 es falsa	H0 es verdadera
Significante	<b>1 - beta</b> <b>(Poder)</b>	<b>alfa</b> <b>(error tipo I)</b>
No Significante	<b>beta</b> <b>(error tipo II)</b>	<b>1 - alfa</b> <b>(confiabilidad)</b>

# ANALISIS ESTADÍSTICO

Significancia estadística  $\neq$  Significancia Clínica  
o práctica

Significancia estadística

No

Si

No

Caso I

Caso II

Significancia  
Biológica

Si

Caso III

Caso IV

	No	Si
No	Caso I	Caso II
Si	Caso III	Caso IV

Hipótesis nula:  $H_0: \mu_1 = \mu_2$  (las dos medias son iguales)

Hipótesis alternativa:  $H_1: \mu_1 \neq \mu_2$  (las dos medias no son iguales)

Atendiendo a los resultados del estadístico se decide:

		No rechazar	Rechazar
$H_0$	Verdadera	Decisión correcta	Error tipo I ( $\alpha$ )
	Falsa	Error tipo II ( $\beta$ )	Decisión correcta

*Tabla 2. Pruebas estadísticas de relación entre dos variables*

	Dicotómica	Nominal	Ordinal	Continua
<b>Dicotómica</b>				
Datos no apareados	$\chi^2$ ( $\pm$ Yates). Test de Fisher	$\chi^2$	U de Mann-Whitney	t de Student (datos independientes)
Datos apareados	Test de simetría de McNemar	$\chi^2$	Test de Wilcoxon	t de Student (datos apareados)
<b>Nominal</b>		$\chi^2$	Kruskal-Wallis (análisis de la varianza)	ANOVA (análisis de la varianza)
<b>Ordinal</b>			Rho Spearman	Rho Spearman
<b>Continua</b>				Correlación, regresión

# Análisis de Riesgo

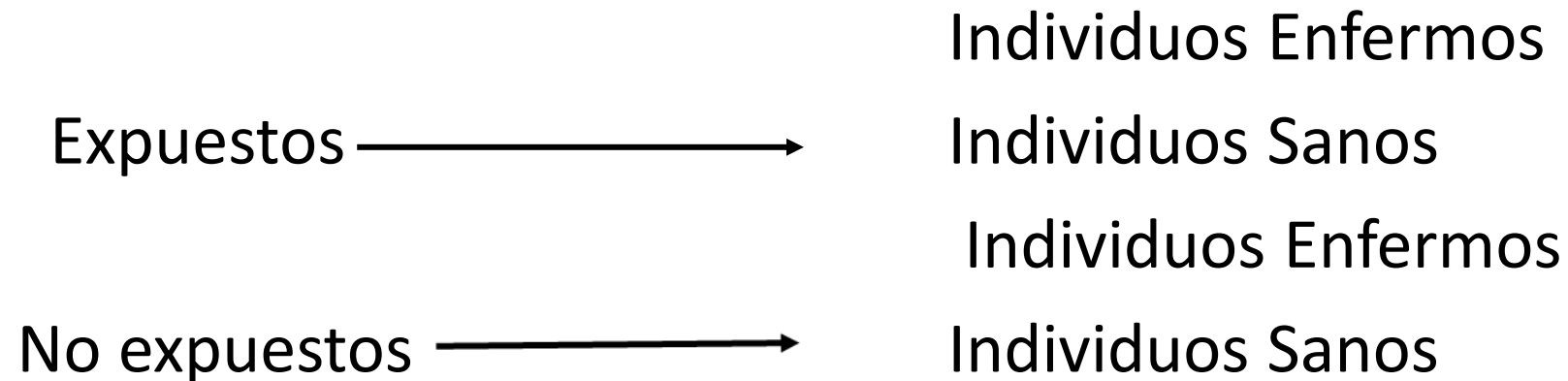
- **Asociación:** relación de dependencia estadística entre dos o más eventos, características u otras variables. Una asociación está presente si la probabilidad de ocurrencia de un evento depende de la ocurrencia de otro u otros.
- En epidemiología, la aplicación del término 'asociación' siempre implica la intención de establecer una **relación de causa a efecto** entre una exposición y una enfermedad o evento en salud.
- **Factor de riesgo:** característica o circunstancia detectable en individuos o grupos, asociada con una probabilidad incrementada de experimentar un daño o efecto adverso a la salud. En general, un factor de riesgo es un atributo o exposición que incrementa la probabilidad de ocurrencia de una enfermedad u otro daño a la salud.

# Pruebas de Significancia

- Las medidas de asociación estadística se basan en las llamadas **pruebas de significancia**. El propósito de estas pruebas es determinar si la presencia de un factor de riesgo evaluado está efectivamente relacionada con la frecuencia de la enfermedad. En dichas condiciones se espera que la prevalencia de exposición a dicho factor sea razonablemente más alta entre los que han enfermado o sufrido un daño a la salud que en aquellos aparentemente sanos.

# Diseños de investigación para estudiar asociaciones de riesgo

- **Cohorte:** Los estudios de cohorte consisten en el seguimiento de una o más cohortes de individuos sanos que presenta diferentes grados de exposición a un factor de riesgo en quienes se mide la aparición de la enfermedad o condición en estudio.



# Análisis

- **Riesgo Relativo:** razón entre el riesgo absoluto de enfermar o morir de aquellos con la exposición de interés y el riesgo absoluto de enfermar o morir de aquellos sin la exposición de interés.

	enfermo	no enfermo	
expuesto	a	b	a + b
no expuesto	c	d	c + d
	a + c	b + d	a + b + c + d

$$\text{riesgo relativo} = \frac{\text{incidencia en expuestos}}{\text{incidencia en no-expuestos}}$$

y, en la tabla 2x2, esto es:

$$\text{RR} = \frac{I_E}{I_{\bar{E}}} = \frac{a/a+b}{c/c+d}$$

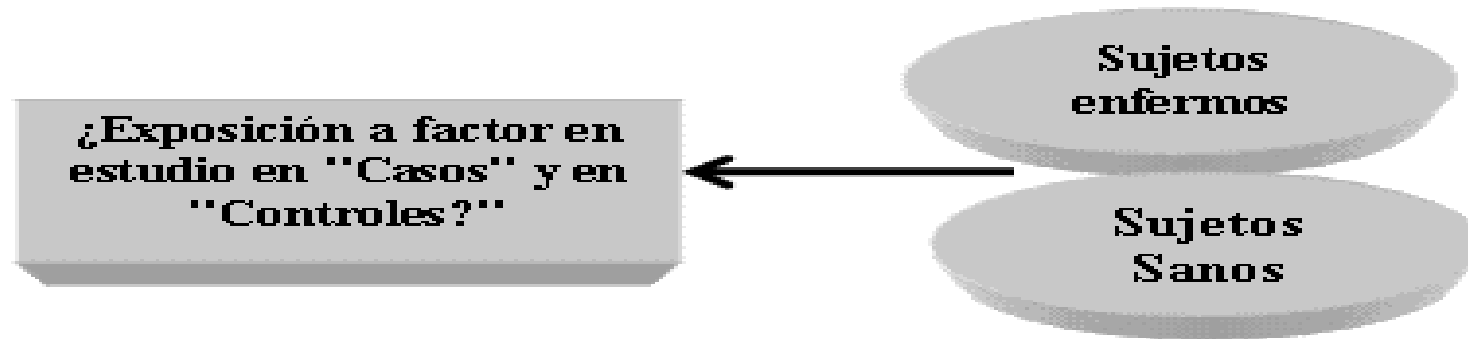


# Ejemplo

	Cáncer	Sanos	total
Expuestos	20	80	100
No expuestos	5	95	100
Total	25	175	200

$$RR = \frac{20/100}{5/100}$$

# Casos y controles



**Razón de la discrepancia (OR)** : Es el estimador del RR cuando no se conoce, la relación causa efecto.

$$OR = \frac{a \times d}{b \times c}$$

# Análisis inferencial

- la probabilidad de que una conclusión acerca de la población de referencia, basada en el análisis de datos de una porción de ella (muestra) o en la comparación con otra población, sea correcta o verosímil; ésto corresponde a la llamada **estadística inferencial**.

# Flujograma para toma de decisión de prueba estadística con promedios

*¿Es la muestra grande o pequeña?*

Grande  
mayor de 29  
en c/grupo

Pequeña  
menor de 30  
en c/grupo

*¿Cuántas muestras a comparar?*

*¿Cuántas muestras a comparar?*

Una

Dos  
pareadas

Dos  
Indep.

> dos  
Indep.

Una

Dos  
pareadas

Dos  
Indep.

> dos  
Indep.

*¿Difieren los promedio/varianzas?*

*¿Difieren los promedio/varianzas?*

Z Cuenta  
Distribución  
Normal

Anova  
(una o 2 vías)

Prueba de T  
muestra única

Prueba de T  
pareada

Prueba de T

Anova  
(una o 2 vías)

# Flujograma para toma de decisión de prueba estadística con proporciones

*¿Es la muestra grande o pequeña?*

Pequeña  
< 5 esperados  
por c/celda

Grande  
> 4 esperados  
por c/celda

*¿Cuántas muestras a comparar?*

*¿Cuántas muestras a comparar?*

Una

Dos  
pareadas

Dos  
Indep.

> dos  
Indep.

Una

Dos  
pareadas

Dos  
Indep.

> dos  
Indep.

*¿Difieren los promedio/varianzas?*

*¿Difieren los promedio/varianzas?*

Prob. exacta  
basado en  
distribución

Prueba de  
McNemar  
exacta

Prueba exacta  
de Fisher

Prueba Q  
de Cochran

Prueba de Chi2  
mejor ajuste

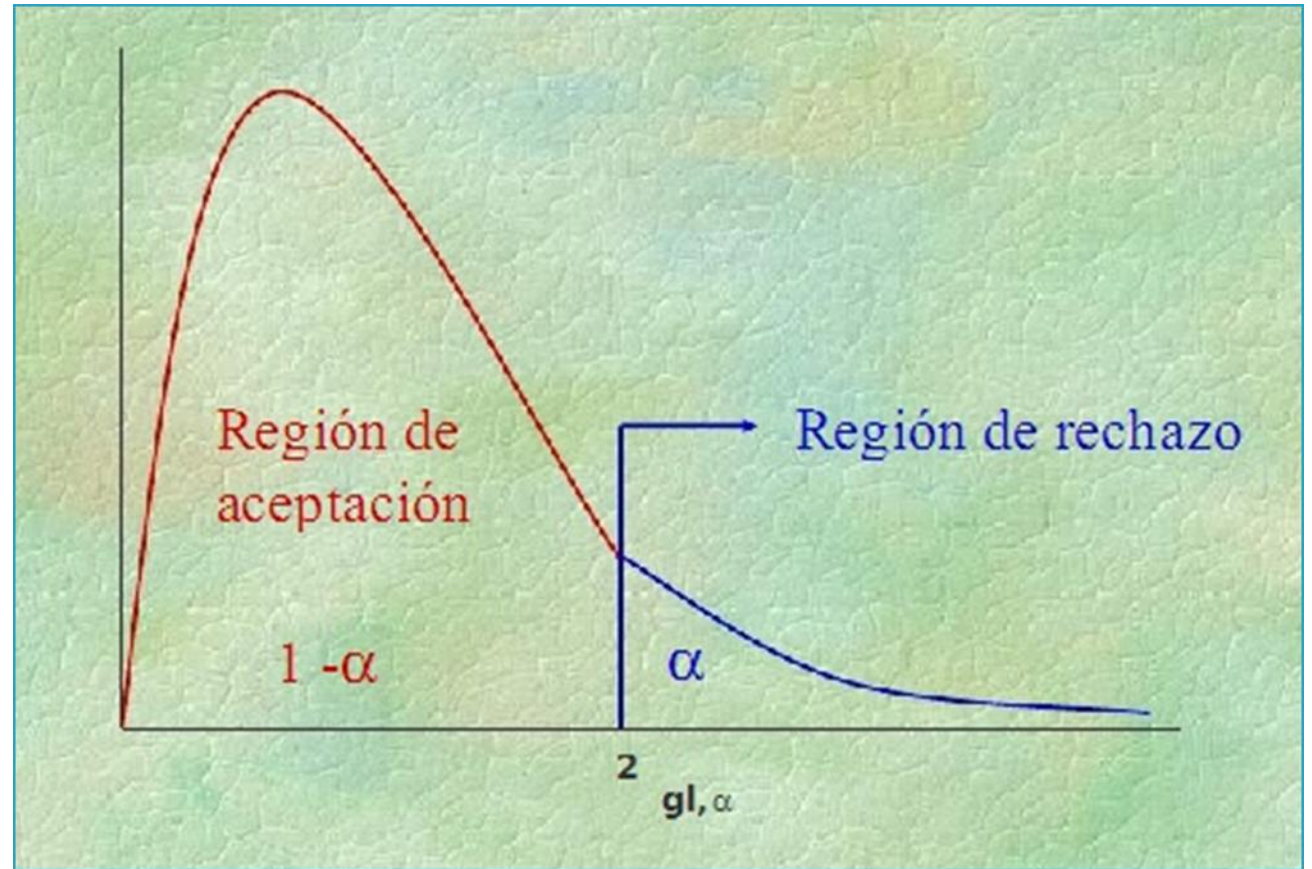
Prueba de Chi2  
pareada

Prueba de Chi2  
para asociación

Prueba de Chi2  
filas x columnas

# Prueba de $\chi^2$

- Es una distribución de probabilidad que compara una distribución observada con los posibles resultados esperados.
- Es considerada una prueba no paramétrica
- Luego de la distribución normal es una de las distribuciones mas ampliamente usadas



# Prueba de t de Student

- es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.
- Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.

## **PRUEBA ANOVA**

**Práctica con Excel, SPSS, EPIDAT con las bases de datos elaboradas en el curso.**





Calcula la probabilidad de que una familia que tiene cuatro hijos, tres de ellos sean niños.

Es una distribución binomial, los hijos sólo pueden ser niños o niñas.

Suceso A (éxito) tener un niño  $\Rightarrow p(A) = 0,5 \Rightarrow \mathbf{p = 0,5}$

Suceso  $\bar{A}$  tener una niña  $\Rightarrow p(\bar{A}) = 0,5 \Rightarrow \mathbf{q = 0,5}$

$\mathbf{n = 4}$  (hijos)  $\Rightarrow \mathbf{B(n, p)} \Rightarrow \mathbf{B(4; 0,5)}$

Probabilidad de tener tres niños  $\Rightarrow \mathbf{x = 3}$

$$P(x = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \Rightarrow \begin{matrix} k = 3 \\ n = 4 \\ p = 0,5 \\ q = 0,5 \end{matrix} \Rightarrow P(x = 3) = \binom{4}{3} \cdot (0,5)^3 \cdot (0,5)^{4-3}$$

$$\binom{n}{k} = \frac{n!}{k! (n-k)!} \Rightarrow \binom{4}{3} = \frac{4!}{3! (4-1)!} = \frac{4 \cdot \cancel{3!}}{\cancel{3!}} = 4$$

$$P(x = 3) = \binom{4}{3} \cdot (0,5)^3 \cdot (0,5)^{4-3} \Rightarrow P(x = 3) = 4 \cdot (0,5)^3 \cdot (0,5)^1 = \mathbf{0,25}$$

## Prueba de la Binomial

$$X \sim B(n, p)$$

Una prueba de laboratorio para descartar heroína en sangre tiene 92% de precisión.

Si se analizan 72 pruebas en un mes, calcula las siguientes probabilidades:

- a) 60 o menos están correctamente evaluadas CDF
- b) Menos de 60 están correctamente evaluadas PDF
- c) 60 están correctamente evaluadas

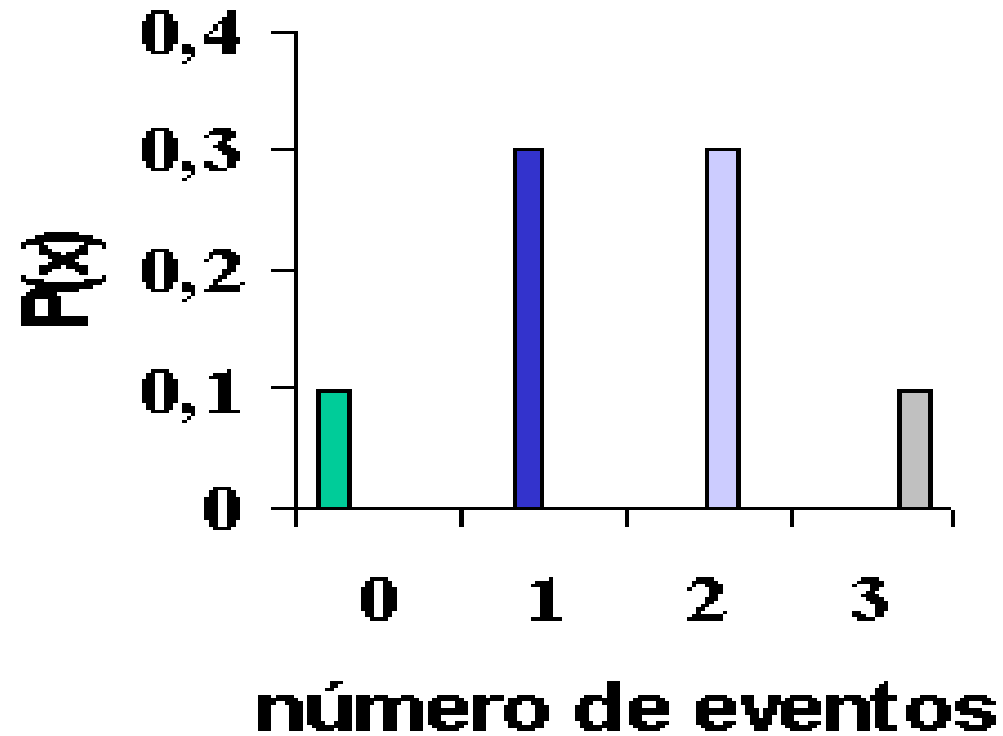
**72 muestras/mes = n**

**Probabilidad de 60 o menos. C= 60**

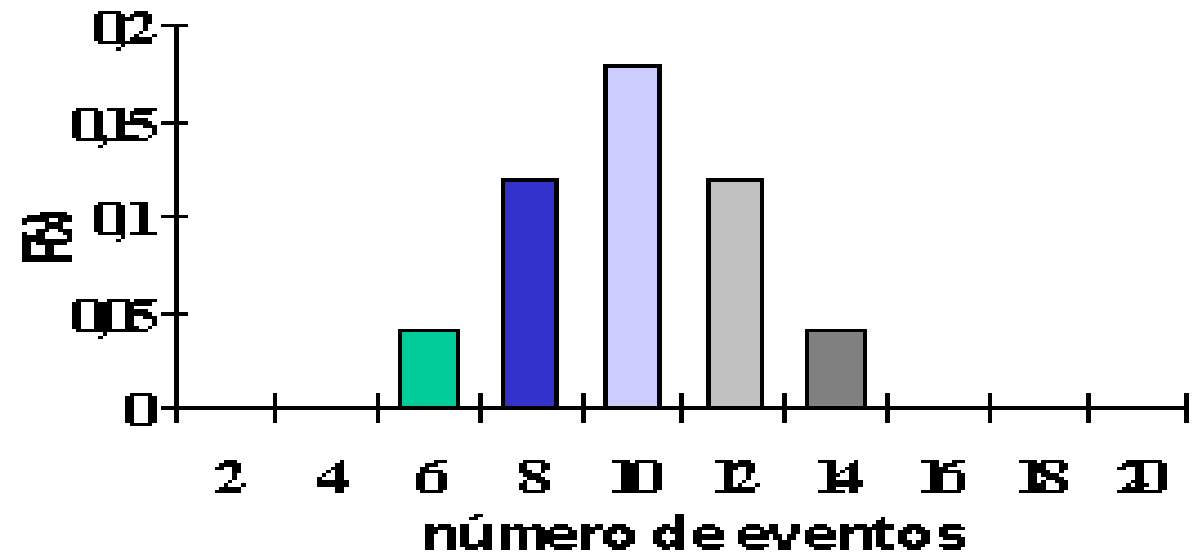
**Nivel de confianza, P= 0,92**

# Distribución binomial para $n$ igual a 3 y 20, donde $\pi = .50$

**$n=3$**



**$n=20$**



En teoría de probabilidad y estadística,

la **distribución de Poisson** es

una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo.

Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos "raros".

## PRUEBA DE POISSON

En una población la mortalidad de cáncer de pulmón es de 12. Si el número de muertes sigue una distribución de Poisson, calcula:

- a) Qué haya exactamente 10 muertes por cáncer en un año.
- b) 15 o más personas mueran de Ca de pulmón en un año
- c) 10 o menos personas mueran a causa de la enfermedad en 6 meses

$$f(k, \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

## AGENDA

# Procesamiento y análisis estadístico de datos para la investigación científica de las Ciencias de la Salud. (Básico)

9. Resumen y conclusiones del aprendizaje.

# Procesamiento y análisis de datos para la investigación científica de las Ciencias de la Salud. (Básico)

Mariano Fernández Silano ([mferna@gmail.com](mailto:mferna@gmail.com))

Alejandro Rísquez Parra ([risqueza@gmail.com](mailto:risqueza@gmail.com))

## REFERENCIAS

### I - Materiales para el uso de Epi Info:

Descarga programa para Windows: <http://wwwn.cdc.gov/epiinfo/7/index.htm>

Descarga programa para Tablets: <https://epiinfoandroid.codeplex.com/>

Selección de videos sobre Epi Info:

- Componentes Epi Info: <https://www.youtube.com/watch?v=TgECeZYC5BE>
- Como hacer un Cuestionario - Epi Info 7: <https://www.youtube.com/watch?v=CptueXCUJf4>
- Análisis de datos - Epi-info 7: <https://www.youtube.com/watch?v=hELfcjZzqtw>



# Procesamiento y análisis de datos para la investigación científica de las Ciencias de la Salud. (Básico)

Mariano Fernández Silano (mferna@gmail.com)

Alejandro Rísquez Parra ([risqueza@gmail.com](mailto:risqueza@gmail.com))

## REFERENCIAS

I

I - Recursos para el análisis estadístico en SPSS

A continuación encontrará una selección de videos que facilitaran el aprendizaje de la realización del análisis estadístico a través del paquete SPSS.

1.- La construcción de la Base de datos/ archivo de datos/ matriz de datos:

- <http://bioestadistico.com/principios-cientificos-de-una-matriz-de-datos>
- <http://bioestadistico.com/importacion-y-edicion-de-la-matriz-de-datos>
- <http://bioestadistico.com/la-matriz-de-datos-a-partir-de-las-fichas>
- <http://bioestadistico.com/la-matriz-de-datos-a-partir-de-las-fichas>

2.- Tabulación y graficación de los datos:

- <http://bioestadistico.com/tablas-de-frecuencias-absolutas-y-relativas>
- <http://bioestadistico.com/diagrama-de-sectores-y-grafico-de-barras>
- <http://bioestadistico.com/histograma-diagrama-de-caja-y-bigotes>
- [Tablas estadísticas y Tratamiento gráfico. Módulo 1](#)

3.- Medidas de tendencia central y dispersión

- <http://bioestadistico.com/los-percentiles-y-los-valores-de-normalidad>
- <http://bioestadistico.com/estimacion-puntual-e-intervalos-de-confianza>

4.- Algunas pruebas y análisis inferenciales

- [Prueba de Chi cuadrado, Chi cuadrado II](#)
- [Regresión lineal simple](#)
- [Análisis de tablas de contingencia](#)
- [Prueba de t de Student para muestras Independientes](#)
- [Prueba de t de Student para muestras dependientes](#)
- [Coeficiente de regresión](#)
- [Coeficiente de correlación](#)
- [ANOVA. Análisis de la varianza con un factor, ANOVA II](#)





# Procesamiento y análisis de datos para la investigación científica de las Ciencias de la Salud. (Básico)

Mariano Fernández Silano (mferna@gmail.com)

Alejandro Rísquez Parra ([risqueza@gmail.com](mailto:risqueza@gmail.com))

## REFERENCIAS

I

CURSO DE ESTADÍSTICA COMO MATERIAL DE SOPORTE

Dra. María Purificación Galindo Villardón

Dra. Purificación Vicente Galindo

Universidad de Salamanca

### MÓDULOS

Estadística Descriptiva: Tablas estadísticas y Tratamiento gráfico. Módulo 1

Universidad de Salamanca

<https://www.youtube.com/watch?v=BkSHNKjARYU>

**Estadística Descriptiva: Medidas de Síntesis (I). Módulo 2**

Universidad de Salamanca

<https://www.youtube.com/watch?v=E-Vpyi6hO9k>

**Estadística descriptiva: Medidas de síntesis (II). Módulo 2**

<https://www.youtube.com/watch?v=luX22-Epxzc>

**Análisis de relación entre dos variables cuantitativas Coeficiente de correlación de Pearson Módulo3**

<https://www.youtube.com/watch?v=1qkAU--IK8Y>

**Análisis de relación entre dos variables cuantitativas: Coeficiente de regresión. Módulo 3**

<https://www.youtube.com/watch?v=ZQb7wjzbfds>

**Análisis de la relación entre dos variables cualitativas. Chi cuadrado: significación Módulo 4**

<https://www.youtube.com/watch?v=qAHXnbp1IHY>

**Análisis de la relación entre dos variables, cualitativa y cuantitativa: T de Student Módulo 5**

<https://www.youtube.com/watch?v=ZQb7wjzbfds>

**Análisis de la relación entre dos variables cualitativa y cuantitativa: t de Student (II) Módulo 5**

<https://www.youtube.com/watch?v=zT-YWb3hIPw>

**Test no paramétricos: U de Mann-Whitney. Módulo 6**

<https://www.youtube.com/watch?v=dCG3VAfa11Y>

**Test no paramétricos: Test de Wilcoxon. Módulo 6**

[https://www.youtube.com/watch?v=B\\_7Wt49dTos](https://www.youtube.com/watch?v=B_7Wt49dTos)



## Procesamiento y análisis de datos para la investigación científica de las Ciencias de la Salud. (Básico)

Mariano Fernández Silano (mferna@gmail.com)

Alejandro Rísquez Parra ([risqueza@gmail.com](mailto:risqueza@gmail.com))

### REFERENCIAS

EJEMPLO DEL USO DEL EXCEL PARA CORRELACIÓN

Covarianza, Coeficiente de Correlación, Diagrama Dispersión

<https://www.youtube.com/watch?v=t6MCGQi-SGs>

Distribución normal y Teorema del límite.

Universidad a Distancia de Madrid. [www.udima.es](http://www.udima.es)

[https://www.youtube.com/watch?v=X2cGF9bC\\_Os](https://www.youtube.com/watch?v=X2cGF9bC_Os)

Calculo del chi cuadrado r x c

<https://www.youtube.com/watch?v=AgpWO1LiHQU>

TABLA 2 X2 PARA PRUEBAS DIAGNÓSTICAS

<https://www.youtube.com/watch?v=WF2RYFX-QRU>

Prueba binomial y de Poisson.

<http://es.slideshare.net/andreart4/seminario-viii-poisson-y-binomial-en-spss>

Introducción al paquete SPSS

Universidad de la Laguna

[http://es.slideshare.net/moibemo/introduccion-al-spss-1962856?next\\_slideshow=1](http://es.slideshare.net/moibemo/introduccion-al-spss-1962856?next_slideshow=1)

¿Qué demonios es un Odds Ratio OR? Medidas de efecto

Por: Adolfo Figueiras Guzmán

<https://www.youtube.com/watch?v=so4sV2KtQI0>

**Análisis de la relación entre dos variables cualitativas: Test Chi cuadrado. Módulo 4**

<https://www.youtube.com/watch?v=XvPEeQAjTW8>

**Cuestiones de supervivencia: de la curva de Kaplan-Meier al modelo de Cox**

Por: Guadalupe Gómez

<https://www.youtube.com/watch?v=Zcg2GAMB474>

**Diseñando un estudio de pruebas diagnósticas**

Por: María Xosé Rodríguez Álvarez

<https://www.youtube.com/watch?v=-w6dCgzH0ks>

**Excel: Kaplan-Meier survival curve with Excel**



# Procesamiento y análisis de datos para la investigación científica de las Ciencias de la Salud. (Básico)

Mariano Fernández Silano (mferna@gmail.com)

Alejandro Rísquez Parra ([risqueza@gmail.com](mailto:risqueza@gmail.com))

## REFERENCIAS

|

[Phil Chan](#)

<https://www.youtube.com/watch?v=4cXiECRvJwg>

How to Use SPSS-Kaplan-Meier Survival Curve

[TheRMUoHP Biostatistics Resource Channel](#)

<https://www.youtube.com/watch?v=f4X5csxtJkE>

Diseñando un estudio de etiología: estudios de casos y controles

Por: Alberto Ruano

<https://www.youtube.com/watch?v=5RdnegA5ggA>

ANOVA: Análise da Varianza

Por: Rosa M. Crujeiras Casais

<https://www.youtube.com/watch?v=zUXraaMQ7TY>

Descripción de un grupo y de prevalencias: Estudios transversales. El estudio más sencillo.

Por: Francisco Gude Sampedro

<https://www.youtube.com/watch?v=nyZyM186JoE>

Tablas de contingencia y test asociados

Por Mónica López Ratón

[https://www.youtube.com/watch?v=UAVAiMy\\_R88](https://www.youtube.com/watch?v=UAVAiMy_R88)

Diseñando un estudio de pronóstico: estudios de cohorte

Por: Salvador Pita, Sonia Pértega Díaz, M<sup>a</sup> Teresa Seoane Pillado

<https://www.youtube.com/watch?v=6JFzfitxGo0>

Epidat 4.0

Por: Gael Naveira Barbeito

<https://www.youtube.com/watch?v=EJOzJg1odFc>

