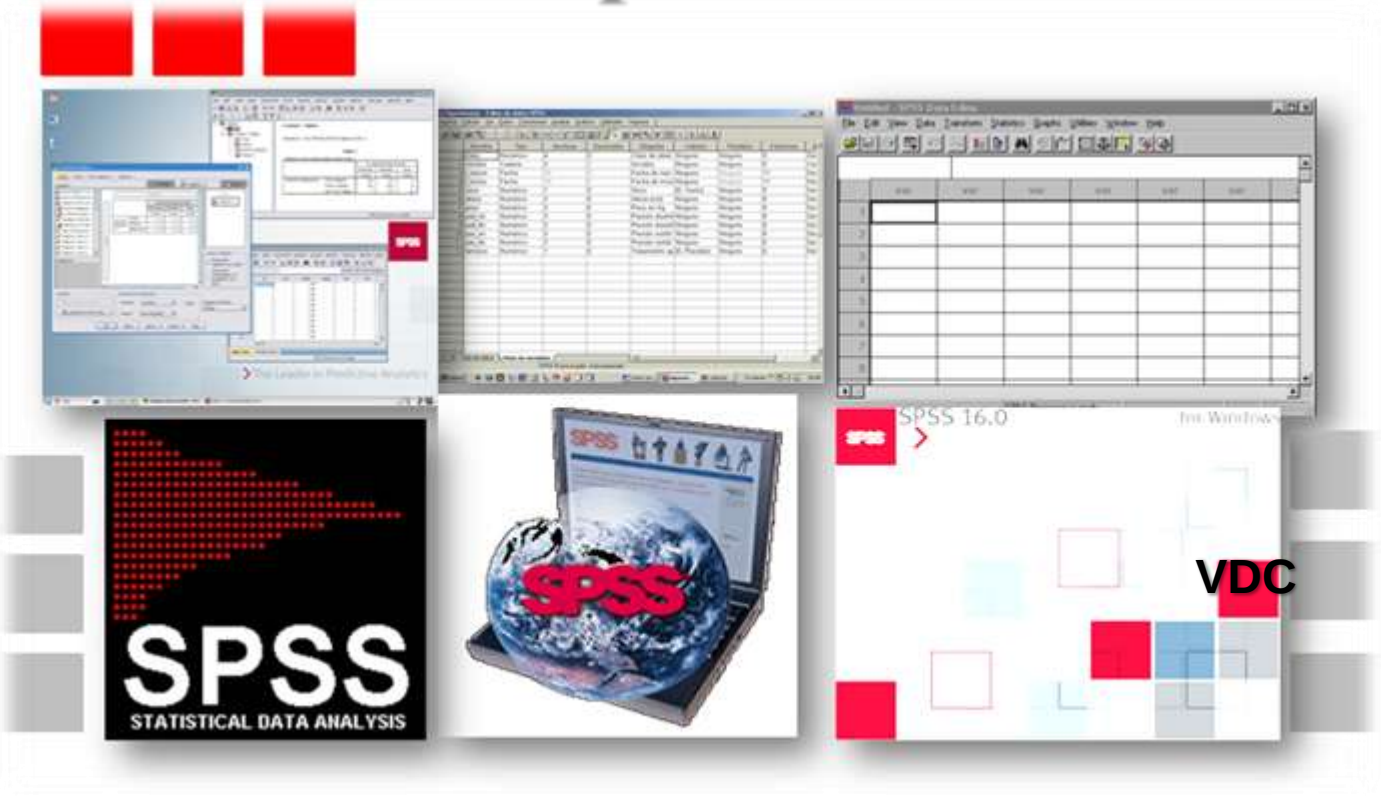
Estrategia de análisis estadístico de los datos. Inferencia Estadística y contraste de hipótesis



Prof. Mª JOSÉ PRIETO CASTELLÓ

MÉTODOS ESTADÍSTICOS. TÉCNICAS

ESTADÍSTICA DESCRIPTIVA

•Descripción de información contenida en un conjunto de datos.

•Resumen de la información contenida en los datos.

TEORÍA DE LA PROBABILIDAD

- Asignación de probabilidad a sucesos inciertos.
- Fundamenta y sirve de nexo entre la descriptiva y la inferencia.
- Ofrece resultados de interés: axiomas, teoremas, propiedades…

INFERENCIA ESTADÍSTICA

•Permite extraer conclusiones válidas sobre colectivos no observados. Cuantifica la variabilidad aleatoria.

•estimación y contraste de hipótesis

.

Trabajo Científico

Investigador

1º) Se plantea una duda sobre un tema.

2º) Descubre aspectos no explicados de un tema en los trabajos revisados.

Definir las Características del mismo

Una vez planteado el Trabajo

Se debe **ELABORAR UNA HIPÓTESIS**

Basada:

1.- Por nuestro conocimiento del Tema.

2.- Por los resultados obtenidos previamente por otros Autores.

Trabajo Científico

Para comprobar si la **HIPÓTESIS** es cierta

Debemos plantear un **EXPERIMENTO** o una **OBSERVACIÓN** sobre una **MUESTRA**

Recoger los **DATOS** que obtengamos

Trabajo Científico

Una vez obtenidos los **DATOS**

Llega el turno de la
ESTADÍSTICA DESCRIPTIVA

Su OBJETIVO es:

1.- **Identificar.**
2.- **Conocer.**
3.- **Describir.**
4.- **Resumir una MUESTRA a partir de los datos de ella.**

Trabajo Científico

Una vez recogidos los **DATOS**

Estos han de **ANALIZARSE**

De su análisis se produce la
aceptación o rechazo de la hipótesis
del estudio

la aceptación o rechazo de la hipótesis
a partir de los datos de la muestra

SOLO puede llevarse a cabo de manera fiable a través del uso de técnicas de
ESTADÍSTICA ANALÍTICA

Trabajo Científico

Si *después de utilizar las técnicas de la* ***ESTADÍSTICA ANALÍTICA*** *para analizar los datos obtenidos.*

Podemos ACEPTAR nuestra Hipótesis

SI podemos ACEPTAR nuestra Hipótesis

entonces

Habremos descubierto una ***nueva propiedad*** *de la población que estamos estudiando, a partir de una muestra*

Trabajo Científico

Si hemos descubierto una ***nueva propiedad***

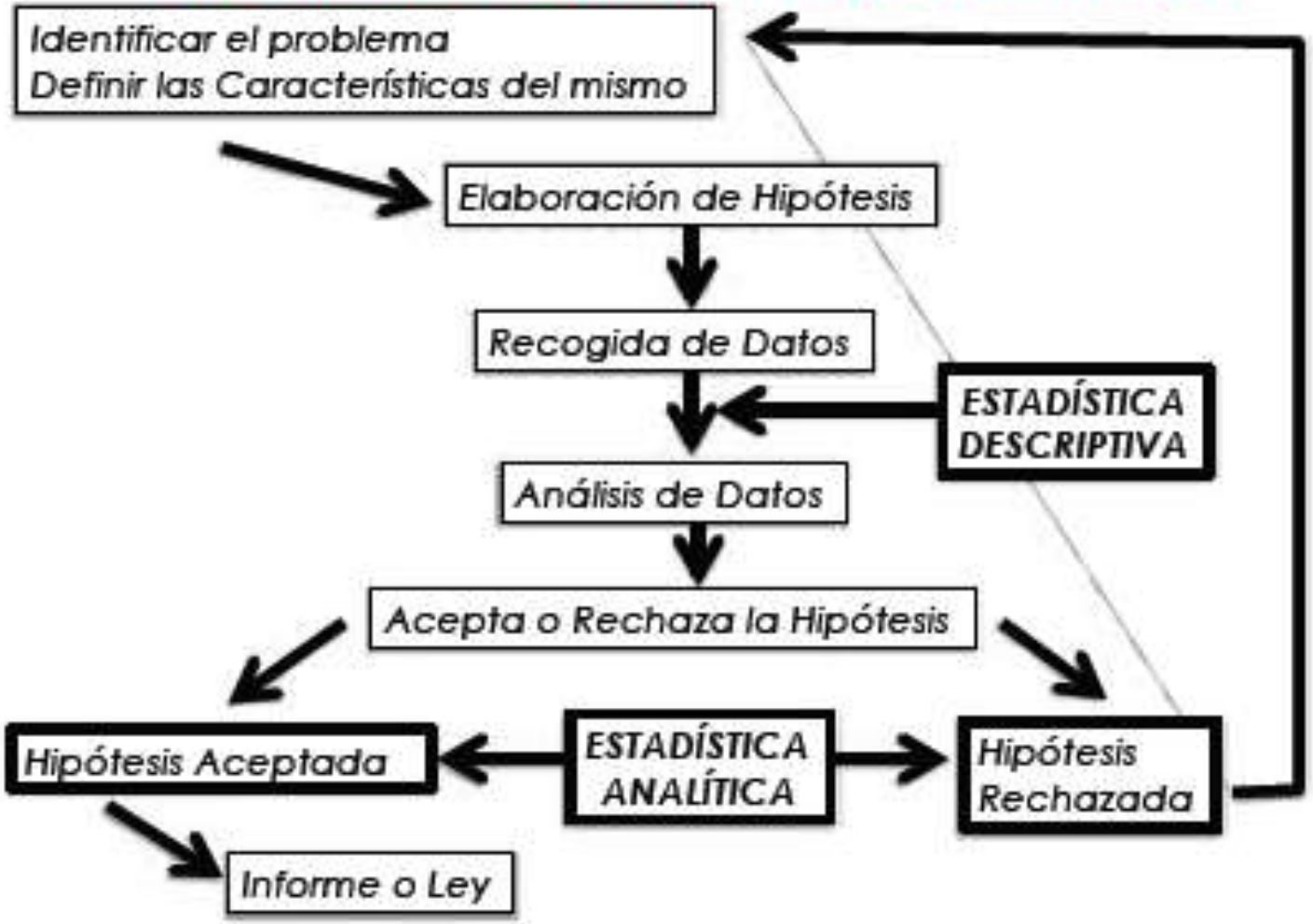
Entonces nos permitirá ***enunciar una ley o emitir un informe.***

SI NO podemos ACEPTAR nuestra Hipótesis

entonces

Volver al principio, ***replantear el problema inicial***

La estadística en el método científico



SPSS

ANÁLISIS ESTADÍSTICO DE DATOS

- **Estadística Descriptiva:**
  - **Cualitativas: frecuencias, porcentajes**
  - **Cuantitativas: descriptivos, medidas de centralizacíón, dispersión, posición..)**

- **Explorar**
- **Tablas de contingencia**

- **Estadística analítica (Inferencia estadística):**
  - **estimación de parámetros: puntual y por intervalos**
  - **contraste de hipótesis**

SPSS

Estadística analítica

DEDUCTIVA O INFERENCIAL. INFERENCIA ESTADÍSTICA.

Parte de la estadística que comprende los métodos y procedimientos para deducir propiedades (hacer inferencias) de una población, a partir de una pequeña parte de la misma (muestra).

La bondad de estas deducciones se mide en términos probabilísticos, es decir, toda inferencia se acompaña de su probabilidad de acierto.

SPSS

ESTIMACIÓN: proceso por el que intentamos aproximarnos al valor desconocido de un parámetro poblacional

Estimador: estadístico, Intervalo de confianza

CONTRASTE DE HIPÓTESIS: Problema de decisión entre dos hipótesis estadísticas (Ho y Ha)

afirmación sobre estadísticos de una muestra

-Paramétricos

-No paramétricos o de libre distribución

SPSS

ESTIMACIÓN: aproximarnos al valor desconocido de un parámetro poblacional

Estimador: estadístico, Intervalo de confianza

-Estimador (estadístico)

-Error estándar del estimador:

-Coeficiente (modelo de probabilidad)→Dist Normal

1,96 (95%)
2,58 (99%)

Linferior= estimador – coef x EE
Lsuperior=estimador+ coef x EE

IC Para una media	IC Para una proporción
95% prob: $\mu = x \pm 1{,}96 \times s/\sqrt{N}$	95% prob: $P = p \pm 1{,}96 \times \sqrt{pxq/N}$
99% prob: $\mu = x \pm 2{,}54 \times s/\sqrt{N}$	99% prob: $P = p \pm 2{,}54 \times \sqrt{pxq/N}$

Donde q=1-p

SPSS

ESTIMACIÓN: proceso por el que intentamos aproximarnos al valor desconocido de un parámetro poblacional

Estimador: estadístico, Intervalo de confianza

CONTRASTE DE HIPÓTESIS: Problema de decisión entre dos hipótesis estadísticas (Ho y Ha)

afirmación sobre estadísticos de una muestra

-Paramétricos

-No paramétricos o de libre distribución

SPSS

CONTRASTE DE HIPÓTESIS

OBJETIVO: analizar la validez de una teoría en base a datos muestrales.

PLANTEAMIENTO: dada una hipótesis, se construye la hipótesis contraria.

- Una de las hipótesis se denomina **H_0: HIPÓTESIS NULA**, de igualdad. Por ejemplo: un estadístico (media, varianza…) de una variable en una muestra es igual al de otra muestra, o igual a un valor determinado. O la forma de una distribución es igual a la normal.
- La contraria es **H_1: HIPÓTESIS ALTERNATIVA**, de diferencia.

En el contraste de hipótesis, se observa la/s muestra/s, se aplica un procedimiento de contraste y:

- O bien se concluye que no puede rechazarse H_0 (no existen diferencias significativas entre los elementos comparados)
- O bien se rechaza H_0 y se acepta H_1(sí existen diferencias significativas)

SPSS

CONTRASTE DE HIPÓTESIS

Significancia estadística y error

La realización de cualquier estudio clínico-epidemiológico pretende poner de manifiesto al final del mismo si existe o no asociación entre diferentes variables. Esta asociación puede ser resultado de que realmente exista la asociación indicada, pero esta asociación también puede ser producto del azar, de la presencia de sesgos o de la presencia de variables de confusión

En el proceso inferencial, siempre existe el riesgo de error o imprecisión ya sea por el azar o la variabilidad biológica del fenómeno a estudiar.

Cuanto más grande es el tamaño muestral, mayor es la precisión y la variabilidad explicada por el azar disminuye

SPSS

CONTRASTE DE HIPÓTESIS

Tipos de error

Error tipo I (a)

Riesgo de equivocarnos al rechazar la hipótesis nula.

La probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera (error de tipo I) se le denomina nivel de significación y es la "p".

Se le conoce también como error alfa.

El nivel de significación es el riesgo o la probabilidad que voluntariamente asume el investigador de equivocarse al rechazar la hipótesis nula, cuando en realidad es cierta. Este riesgo se establece normalmente en 0.05

Si $p < 0.05$ se considera significativo, en cuyo caso se rechaza la hipótesis nula y no significativo si $p > 0.05$ en cuyo caso no se rechaza.

Una "p" pequeña significa que la probabilidad de que los resultados obtenidos se deban al azar es pequeña.

SPSS

CONTRASTE DE HIPÓTESIS

Tipos de error

Error tipo II (b)

Consiste en aceptar la hipótesis nula cuando es falsa.

RECOMENDACIONES PARA EVITAR ERROR TIPO II:

- Incrementar el tamaño de la muestra.
- Utilizar test paramétricos (más potentes) en lugar de test no paramétricos.

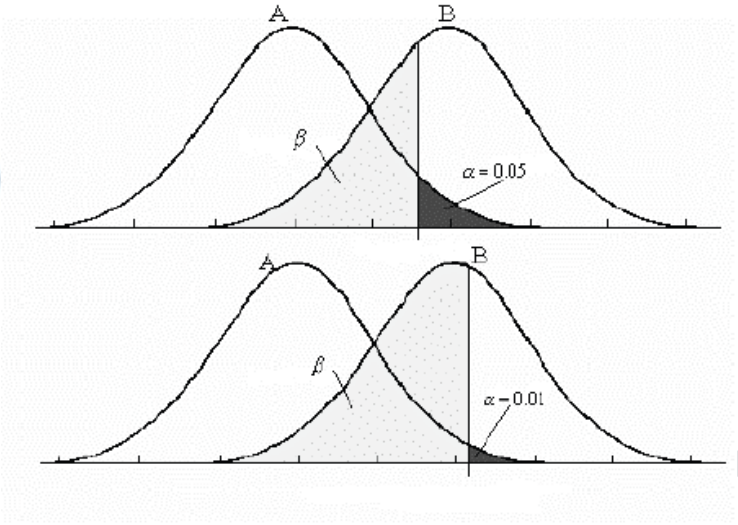
ETAPAS PARA RESOLVER UN CONTRASTE DE HIPÓTESIS

1. **Definir estadístico sobre el que se va a enunciar la hipótesis**
2. **Verificación de requerimientos previos**
3. **Formulación de hipótesis (Ho y Ha)**
4. **Regla de decisión (nivel de significación)**

	Ho cierto	**Ho falsa (Ha cierta)**
Acepto Ho	Correcto	Error tipo II (β)
Rechazo Ho	Error tipo I(α)	Correcto (probabilidad 1- α)

SPSS

CONTRASTE DE HIPÓTESIS



Errores:

RESULTADO DEL PROCEDIMIENTO DE CONTRASTE	H_0 (en realidad **no** hay diferencias)	H_1 (en realidad **sí** hay diferencias)
RECHAZO H_0 (el contraste dice que **sí** hay diferencias)	FALSO POSITIVO: **ERROR TIPO I** **(α)**	VERDADERO POSITIVO
NO PUEDO RECHAZAR H_0 (el contraste dice que **no** hay diferencias)	VERDADERO NEGATIVO	FALSO NEGATIVO: **ERROR TIPO II** **(β)**

SPSS

CONTRASTE DE HIPÓTESIS

Ejemplo práctico:

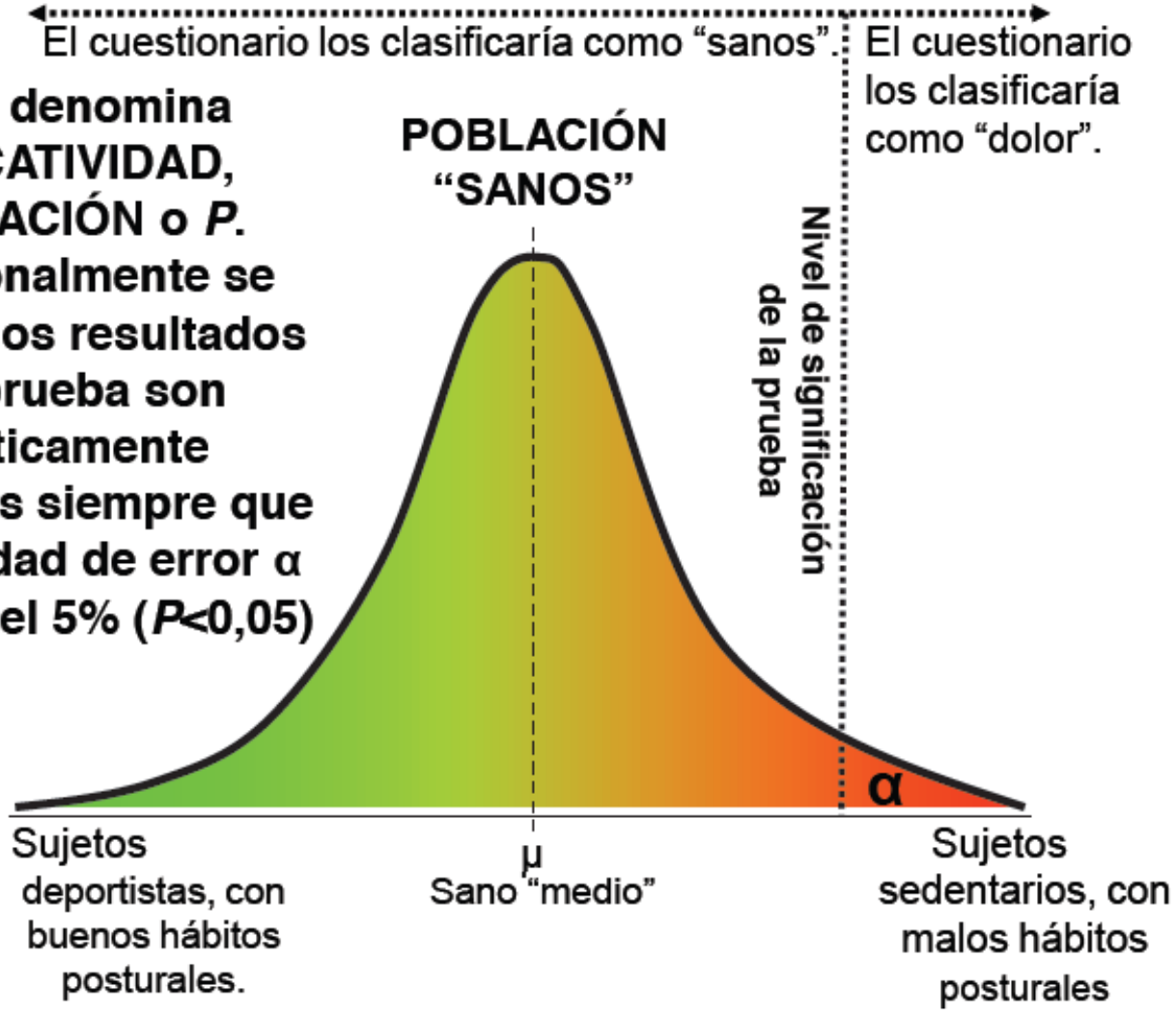
Pasamos un cuestionario sobre hábitos posturales a dos grupos de sujetos: unos sanos, sin antecedentes de dolor lumbar, y otro de enfermos con dolor lumbar.

- **H_0: según los datos de este cuestionario y las muestras que tenemos, NO podemos decir que las medias de la población "sanos" y la población "dolor" sean diferentes.**
- **H_1: según los datos de este cuestionario y las muestras que tenemos, las medias de la población "sanos" y la población "dolor" SÍ son diferentes.**

SPSS

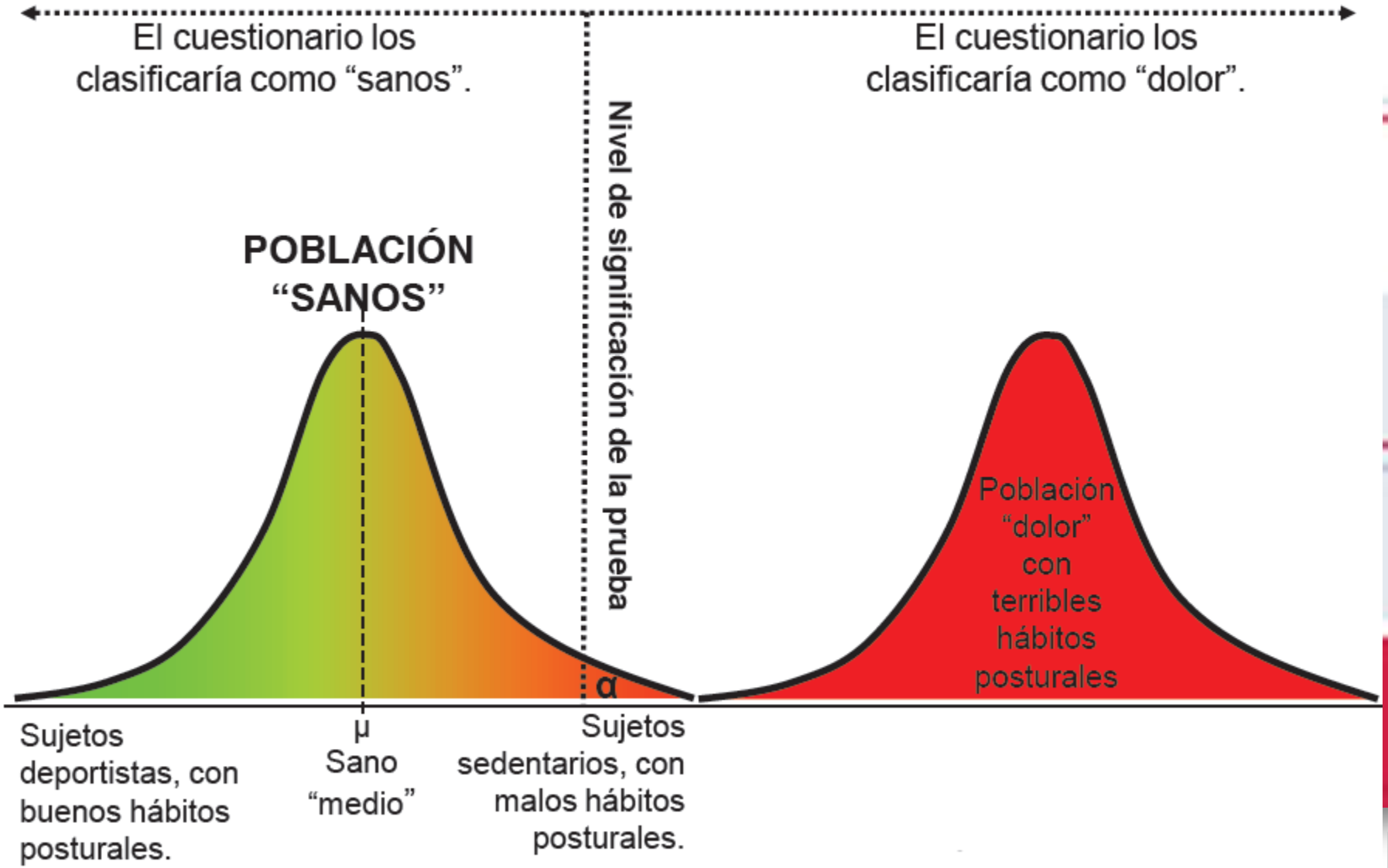
CONTRASTE DE HIPÓTESIS

A α se le denomina SIGNIFICATIVIDAD, SIGNIFICACIÓN o *P*. Convencionalmente se acepta que los resultados de una prueba son estadísticamente significativos siempre que la probabilidad de error α sea menor del 5% (*P*<0,05)



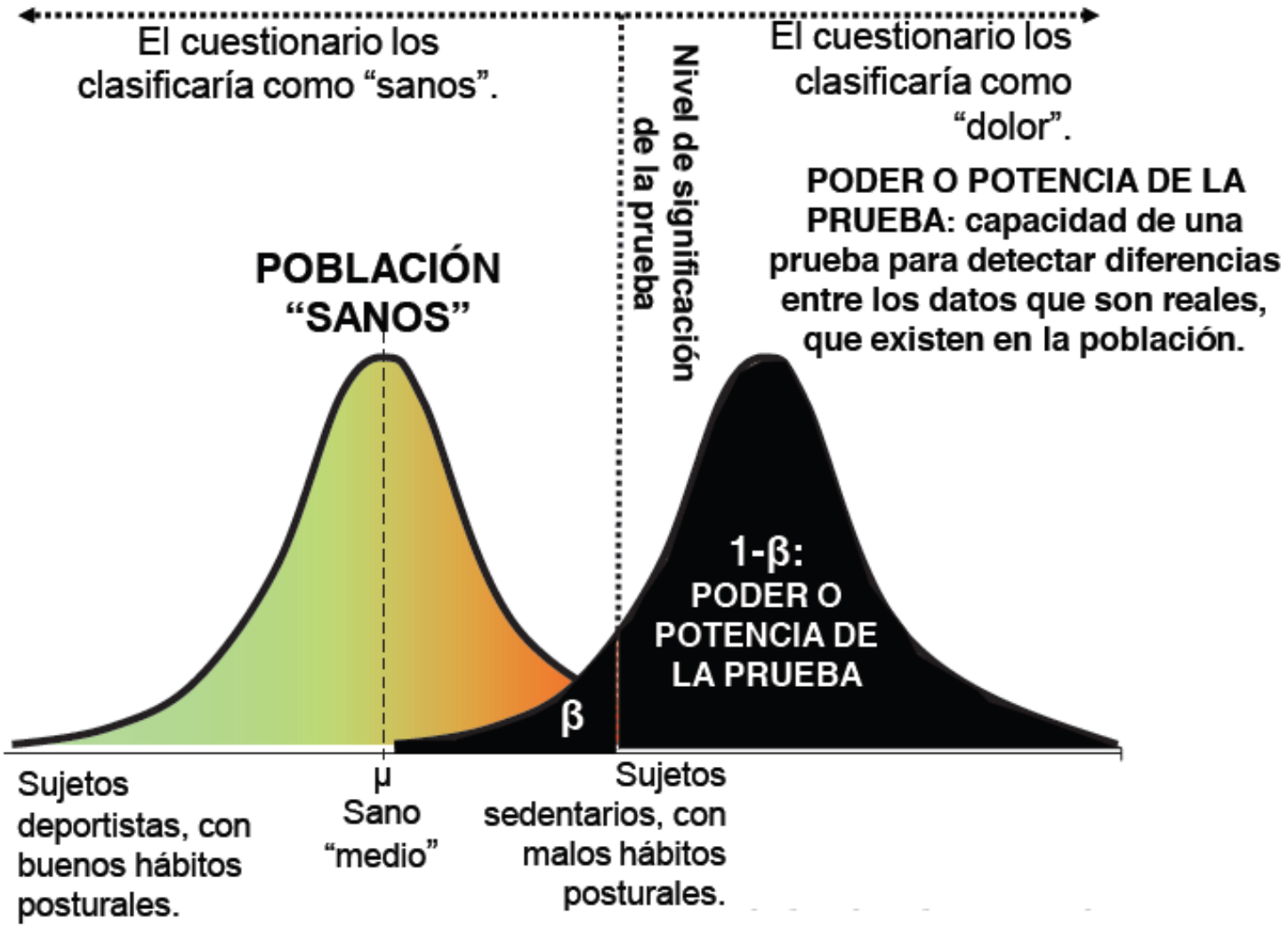
SPSS

CONTRASTE DE HIPÓTESIS



SPSS

CONTRASTE DE HIPÓTESIS

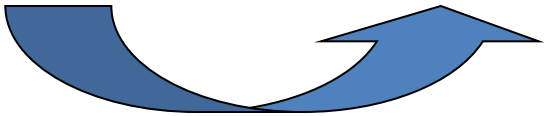


CONTRASTE DE HIPÓTESIS. TEST ESTADÍSTICOS

VARIABLE 1	VARIABLE 2	TEST
Cuantitativa	Cualitativa	T-test (k=2) ANOVA (K>2)
Cualitativa	Cualitativa	Test de la Chi cuadrado
Cuantitativa	Cuantitativa	Correlación de pearson Regresión

VARIABLE 1	VARIABLE 2	Test
Cuantitativa	Cualitativa	U de Mann Whitney. (K=2) Test de Kuskral-Wallis (k>2)
Cualitativa	Cualitativa	Test exacto de Fisher
Cuantitativa	Cuantitativa	Correlación de Spearman. Regresión no paramétrica

PARAMÉTRICOS **NO PARAMÉTRICOS**



Tamaño muestral < 30

Ausencia de normalidad

COMPARACIÓN DE MEDIAS

1 variable CUANTITATIVA – 1 variable CUALITATIVA)

Variable cualitativa 2 grupos

PARAMÉTRICO	NO PARAMÉTRICO
T-test para muestras independientes	U-Mann Whitney
T-test para muestras apareadas	Test de Wilconson

COMPARACIÓN DE MEDIAS

1 variable CUANTITATIVA – 1 variable CUALITATIVA)

Variable cualitativa K grupos

N< 30, NO normal

Varianzas no homogeneas

PARAMÉTRICO	NO PARAMÉTRICO
Análisis de varianza ANOVA de una vía	Test de Kruskall -Wallis
Análisis de varianza ANOVA de medidas repetidas	Test de Friedman

ASOCIACIÓN ENTRE 2 VARIABLES CUALITATIVAS

- **Tablas de contingencia**

-Chi cuadrado de pearson (2x2); P<0,05 si χ2>3,84; p<0,01 si χ2>6,63 (rxc: no válida N<20, más de 20% con valor esperado <5)

-Test exacto de fisher (en tablas 2x2) N<20 o χ2 no valida

-Corrección de Yates (2x2) cuando frecuencias observada menor de 5

-Chi cuadrado de Mantel y Haenszel (2x2) datos ordinales

-Test de Mac Nemar (apareado)

ASOCIACIÓN ENTRE 2 VARIABLES CUANTITATIVAS

- **CORRELACIÓN: Cuánta relación hay**
  r de Pearson → Rho de spearman

- **REGRESIÓN: qué forma tiene la relación**
  Expresión matemática v. respuesta = f (v. explicativa)

SPSS

DISTRIBUCIONES DE PROBABILIDAD

- Concepto.
- Teorema del límite central.
- Pruebas de normalidad:
  - Kolmogorov-Smirnov.
  - Shapiro-Wilk.
  - Gráficos Q-Q.

SPSS

DISTRIBUCIONES DE PROBABILIDAD

Se utilizan en estadística con el propósito de conocer como se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso.

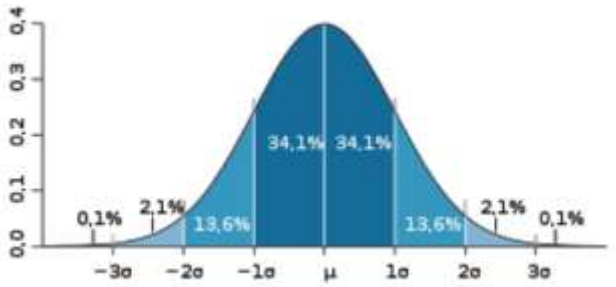
Se representan con una función o una curva, en la que el eje de abscisas muestra los diferentes valores numéricos que puede adoptar dicho resultado, y el eje de ordenadas la probabilidad de cada valor (su suma siempre será 1).

Hay varios tipos de distribuciones de probabilidad, que dependen del tipo de variable. Por ejemplo, la distribución uniforme discreta. Aparece cuando hay un número discreto de resultados igualmente probables.

SPSS

DISTRIBUCIONES DE PROBABILIDAD

DISTRIBUCIÓN NORMAL



Es la distribución de probabilidad que con más frecuencia aparece en estadística y teoría de probabilidades. *LAS PRUEBAS ESTADÍSTICAS MÁS USUALES REQUIEREN QUE LAS VARIABLES ANALIZADAS SE DISTRIBUYAN NORMALMENTE.*

- Es una distribución para variables contínuas.
- Es simétrica en torno a la media (coeficiente de asimetría = 0)
- Coinciden Media, Mediana y Moda.
- La probabilidad de un determinado resultado decrece al alejarse de la media poblacional. Es decir, más allá del valor de la media ±1,96 veces la DS sólo está el 5% de la población.

SPSS

DISTRIBUCIONES DE PROBABILIDAD

Teorema del límite central.

Cuando el nº de elementos de una muestra tiende al infinito, la distribución en el muestreo de la media muestral es aproximadamente normal, incluso aunque la distribución de la población de la cual se ha obtenido la muestra no sea normal.

Podemos afirmar que a partir de un número suficiente de individuos (n=30 habitualmente) una variable cuantitativa que resulte de la adición de un gran número de variables cuantitativas independientes e idénticamente distribuidas con varianzas finitas presentará una distribución aproximadamente normal.

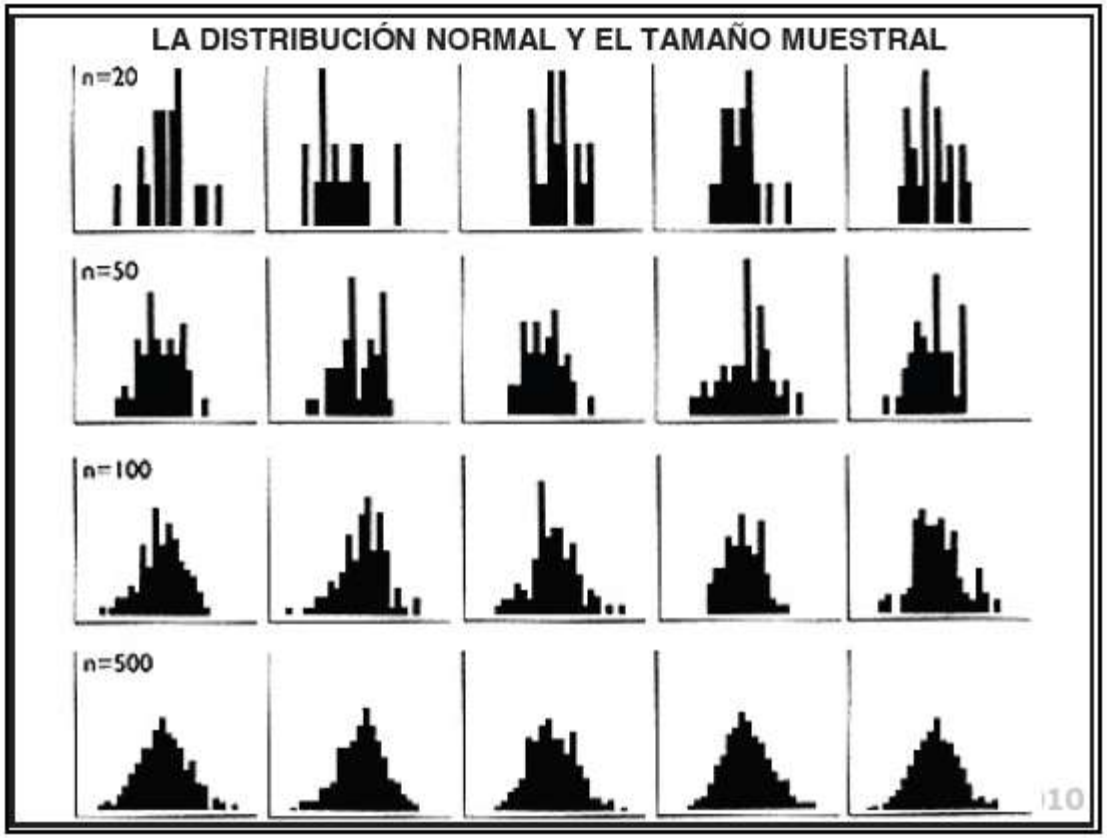
Dado que muchas poblaciones reales se distribuyen probabilísticamente con varianzas finitas, esto explica la prevalencia de la distribución normal.

De aquí la importancia de la distribución normal como modelo de fenómenos cuantitativos en ciencias naturales y sociales. Muchos fenómenos físicos y medidas psicológicas pueden ser bien aproximados mediante la distribución normal. Los mecanismos que subyacen a estos fenómenos son frecuentemente desconocidos, pero el uso de un modelo normal puede ser justificado asumiendo que muchos efectos pequeños e independientes están contribuyendo aditivamente a cada observación.

SPSS

DISTRIBUCIONES DE PROBABILIDAD

Teorema del límite central.



SPSS

DISTRIBUCIONES DE PROBABILIDAD

- Pruebas de normalidad:
  - Kolmogorov-Smirnov (K-S).

Se utiliza para determinar si la distribución de probabilidad de un número finito de observaciones difiere de una distribución hipotética.

Habitualmente se usa para comparar con la distribución normal, pero puede usarse para comprobar

DISTRIBUCIONES DE PROBABILIDAD

- Pruebas de normalidad:
  - Shapiro-Wilk (S-W):

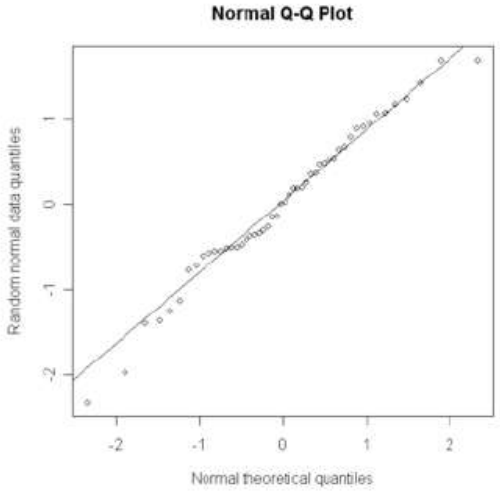
Comprueba si una muestra ha sido obtenida de una población distribuida normalmente.

Es más potente que el K-S para comprobar la normalidad, pero se usa menos.

Para N ≤ 50

SPSS

DISTRIBUCIONES DE PROBABILIDAD



- Pruebas de normalidad:
  - Gráficos Q-Q:

Si los datos representados se distribuyen normalmente, los puntos del diagrama deben ajustarse a la diagonal.

  - Gráficos Q-Q Normal sin tendencia:

Si los datos representados se distribuyen normalmente, los puntos del diagrama deben distribuirse aleatoriamente, sin mostrar ninguna pauta clara de variación.