Summary statistic: números que se usan para describir la data

Tipos de variables

* Numérica: toma un gran rango de valores.
* Discreta: solo toma un conjunto de valores positivos con saltos (0,1,2,…)
* Categórica: representa una categoría a través de un numero

Relación

* Dependiente: se explica por la otra
* Independiente: no se explica por la otra

Muestra: porción de la población

Evidencia anecdótica: no se sabe si representan a la población

Muestreo al azar: reduce la parcialidad, aumenta la posibilidad de no respuesta.

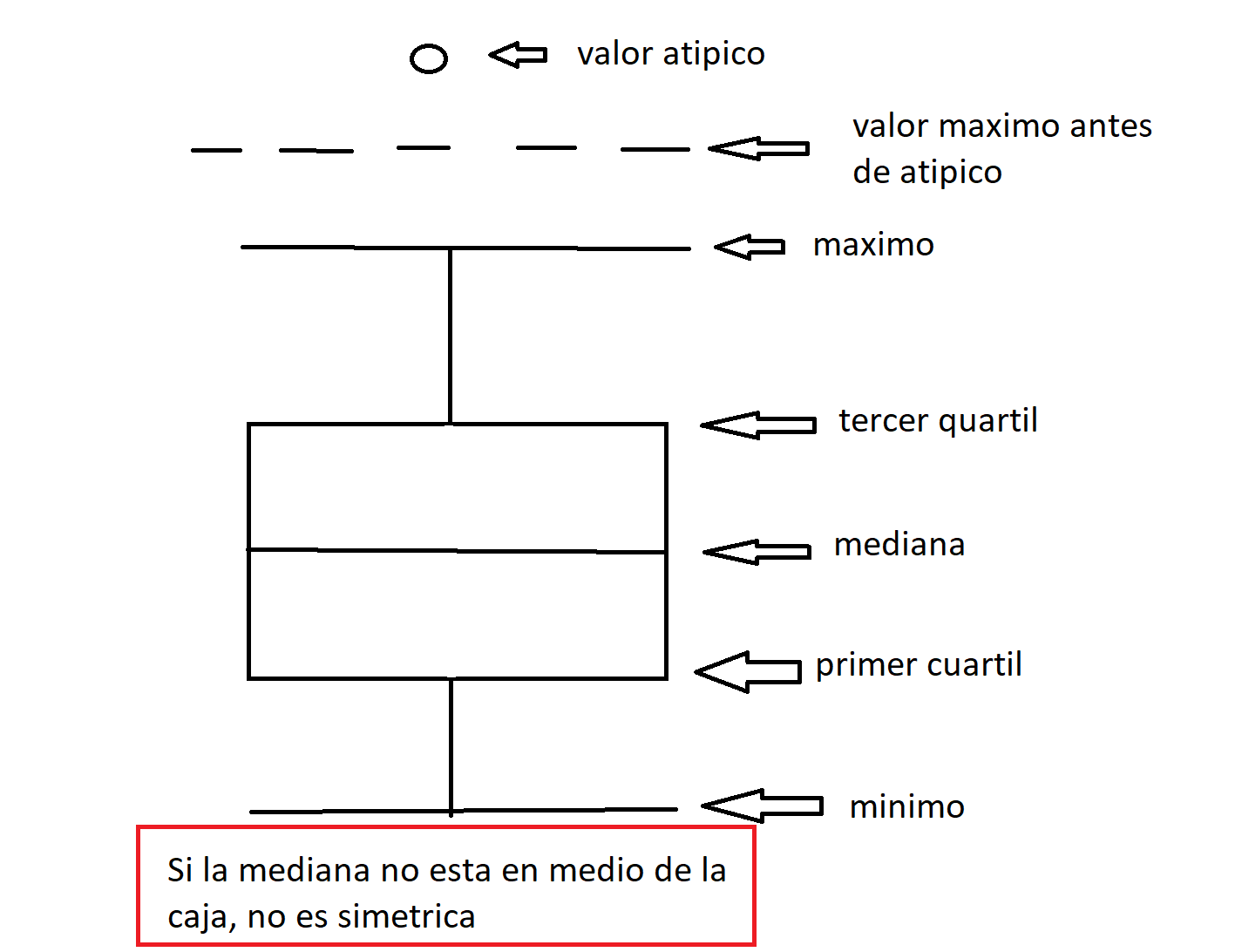
Variable explicativa: explica a la otra variable de respuesta.

Tipos de estudio

* Observación: no interfiere directamente
* Experimentos: interfiere directamente
* Experimento al azar: los grupos se dividen al azar. Reduce parcialidad.

Gráficos

* Dispersión/scatterplots (pareada): muestran la relación entre 2 variables (lineal/no lineal)
* Histogramas: dan la información sobre la densidad de la data. Entre más grande más común. Cuentan con cola a la izquierda, a la derecha y con simetría.
* Box plot: muestra los datos según los cuartiles.



Unimodal: concentración en un lado

Bimodal: concentración en 2 lados

Multimodal: concentración en 3 o más lados.

Desviación estándar: que tan cerca esta la data de su media.

Valores atípicos: sirven para ver que tan sesgada esta la distribución, identificar errores y ver propiedades de la data.

Distribución normal

Aceptación:

* Independiente
* N>30
* sesgado

Cambiar la media mueve la curva.

Cambiar la desviación estándar la comprime o la expande.

Z= ;

Error estándar (SE)=

Intervalo de confianza:

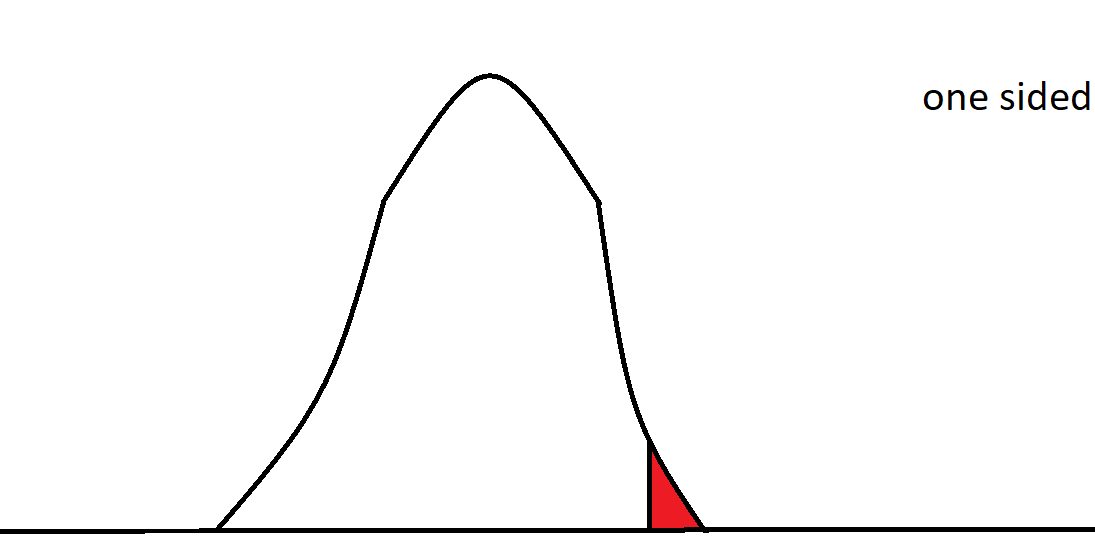
Es cuan confiable es que este dentro del intervalo. **NO** es probabilidad, es confiabilidad.

Error tipo 1: H0 es verdadera pero esta se rechaza.

Error tipo 2: no rechazar H0 cuando se debe rechazar.

Al reducir un tipo de error el otro aumenta.

P valor: probabilidad de que la data sea favorable para la HA.



Teorema del limite central: la distribución de la media de la muestra es aproximadamente normal si viene de una distribución normal.

No pareados: si x esta en el grupo 1, no es necesario que este en el 2.

Distribución t

Para muestras pequeñas y centradas en 0.

Condiciones de aceptación:

* Independencia dentro de la muestra.
* Muestra que venga de una población aproximadamente normal.

Grados de libertad:

Intervalo de confianza:

Para diferencia de muestras:

Desviación estándar agrupada: 2 grupos con desviaciones estándar similares se pueden tratar como idénticos.

Error estándar de la media: desviación estándar de las muestras.