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.

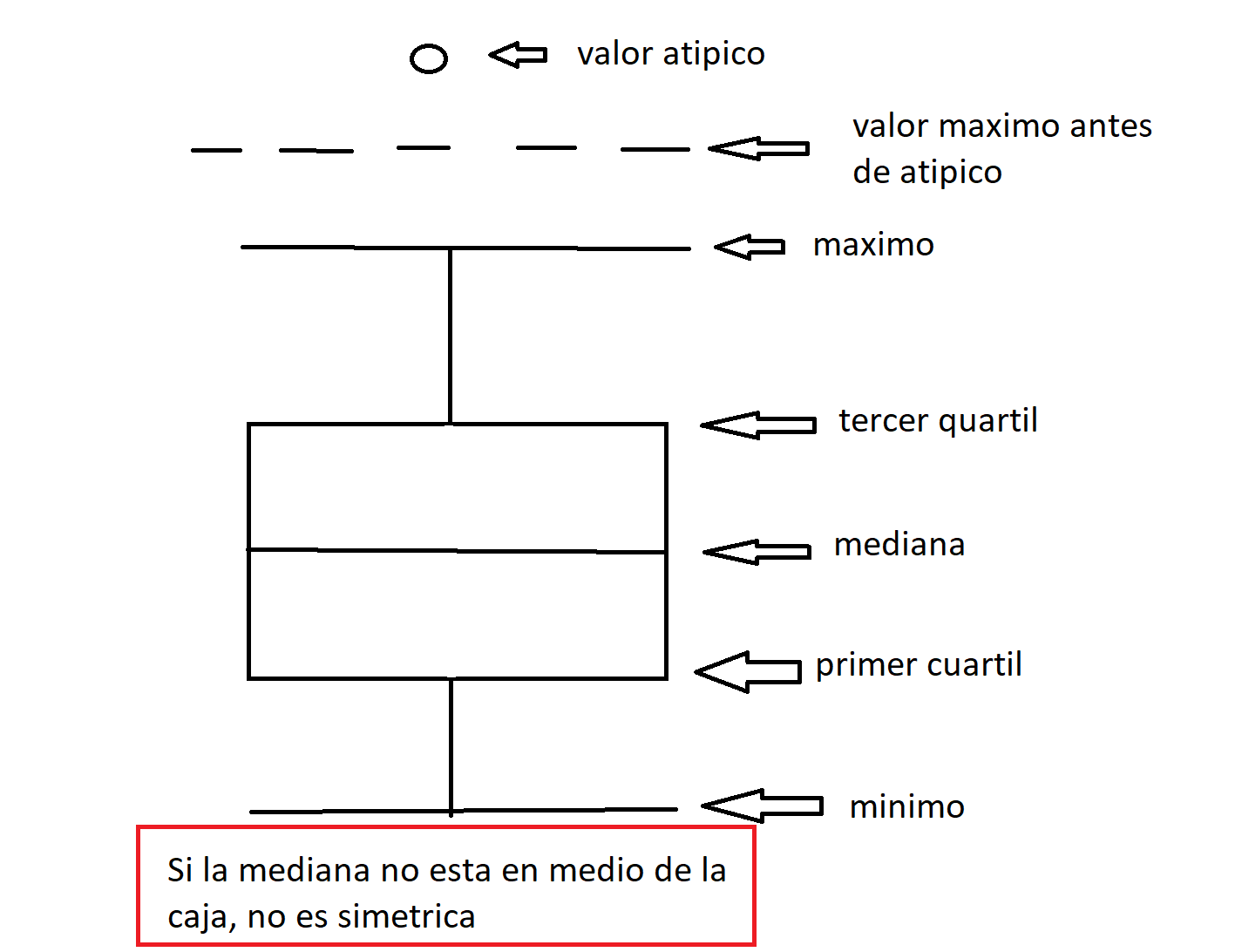
Variable explicativa: explica a la otra variable de respuesta.

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:

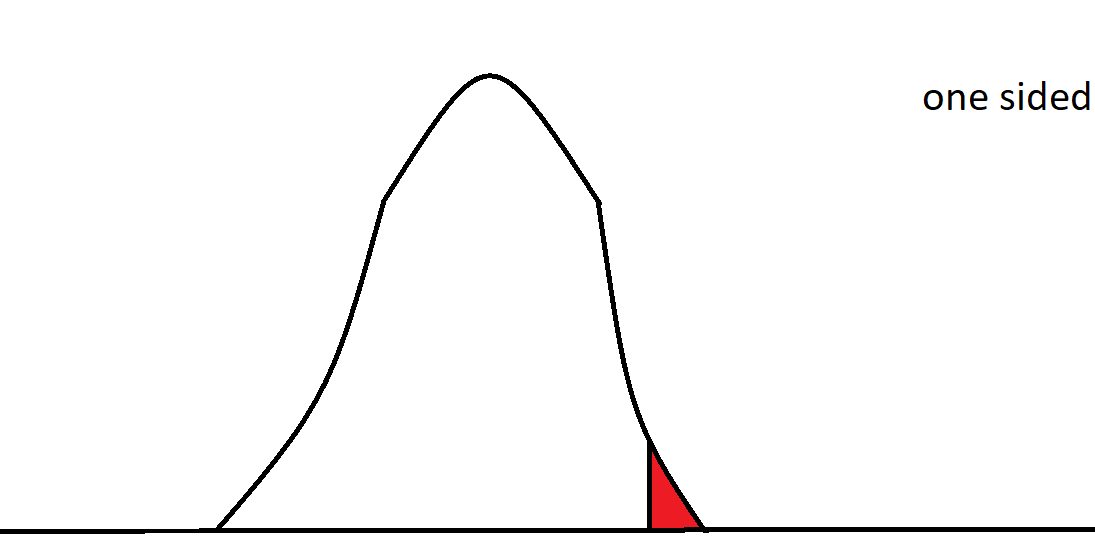
Es cuan confiable es que este dentro del intervalo. **NO** es probabilidad, es confiabilidad.

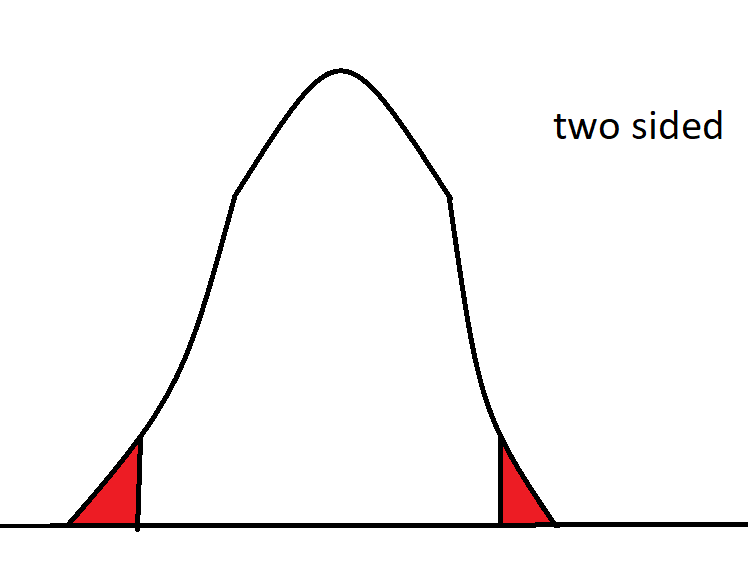
Error tipo 1: H0 es verdadera pero esta se rechaza.

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

Comparar medias.

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.

Poder estadístico: que tan posible es el detectar un efecto que nos interese (rechazar H0)

Poder=z\*100%

Inverso: (poder de x%)

Nivel de significación: probabilidad de realizar un error de tipo 1

Tamaño del efecto: muestras pequeñas provocan error del tipo 2

Proporciones

La distribución es casi moral cuando la distribución de 0´s y 1´s no esta muy cargada hacia un lado.

Condiciones:

* Muestra independiente
* Al menos 10 éxitos (1) y 10 fallos (0), esto se conoce como condición de éxito-fallo

Si se cumple, la muestra tiene media p y error estándar:

Intervalo de confianza:

Margen de error

Diferencia de 2 proporciones

Condiciones:

* Distribución aproximadamente normal
* Las 2 muestras deben ser independientes entre si

Punto estimado:

Fisher

asociación entre 2 variables asociativas

Condiciones:

* Muestras independientes
* Muestras aleatorias
* Muestras < 10% de la población
* Cada observación debe estar solo en una muestra.

Mc Nemar

Comparar el cambio en la distribución de proporciones entre dos mediciones de una variable dicotómica.

Condiciones:

* Al azar
* Datos pareados
* Se estudian 2 variables
* Suma de negativo a positivo y de positivo a negativo sean mayores a 25

Test chi cuadrado

Se usa para:

* Dada una muestra que puede ser muchos grupos, se revisa si es representativa
* Se ve si pertenece a alguna distribución

Condiciones:

* Independencia

Ejemplo de hipotesis:

H0: son elegidos al azar

HA: no son elegidos al azar

Los grados de libertad afectan la forma, el centro y el ancho de la distribución.

Entre más grados de libertad:

* Mas centro hay
* Aumenta la variabilidad
* Mayor simetría

Grados de libertad= categorías -1

Calcular valor esperado en tabla de 2 vias