

Poblaciones y muestras

- En ciencia se buscan los efectos de manipular una cosa sobre otra.
 - E.g., datura, floripondio
- Nos interesan las medidas de poblaciones, no de individuos.
- E.g., intención de voto
- Aleatoriedad
 - Encuesta en el centro vs lista y dado
- Demostración de TCL
 - Distribución normal y desviación estándar
 - Mousekherramienta misteriosa

De las muestras obtenemos datos al medir variables.

Variables

Todo lo que es medible y toma más de un valor

- Cualitativa vs cuantitativa
- Continua o discreta
- Dependiente o independiente
- Niveles de medición
 - Nominal
 - Sexo, raza, religión
 - Número no implica valor (2 católicos = 1 musulmán)
 - Solo frecuencias
 - Ordinal: cantidad relativa
 - Posición en un carrera, escalas de acuerdo/desacuerdo
 - Intervalar
 - Temperatura, inteligencia?, personalidad?
 - No tiene sentido hablar de razones sin 0 absoluto
 - De razón: se mide cantidad real, objetiva de la propiedad
 - Temperatura en grados Kelvin
 - Ejercicios:
 - Número de teléfono (nominal)
 - Ranking cuadro de honor (ordinal)
 - Peso (razón)
 - Distancia viajada (razón)

Ejercicios de variables

- Nivel socioeconómico
- Longitud del cabello
- Sexo
- Edad
- Horas que pasa viendo televisión
- Comida favorita
- Postura política
- Inteligencia
- Número de teléfono
- Peso
- Países visitados

Estadística descriptiva

Medidas de tendencia central

- Resumen rápido de los datos. ¿En dónde se agrupan? ¿Cuál es la puntuación típica?
- Al elegir, importa la escala y la forma de la distribución
 - Moda: con escalas **nominales**
 - Puede haber empates
 - Puede no ser representativa
 - Mediana: de **ordinal en adelante**
 - Solo dice la posición
 - $\frac{N+1}{2}$
 - Media: de **intervalo o razón**
 - Sensible a valores extremos
 - Mejor predictor en ausencia de otra información (minimiza errores)
 - $\bar{X} = \frac{\sum X}{N}$

Medidas de dispersión

- Indican diferencias o consistencia entre puntajes

- Rango: puntaje mayor – menor
- Varianza y desviación estándar: cuantifican cuan distintos son entre sí los puntajes
- Desviación media sería:

$$\frac{\sum(X - \bar{X})}{N}$$

Pero el numerador equivaldría a 0 porque la suma de las desviaciones siempre es 0.

- Varianza muestral:

$$S_x^2 = \frac{\sum(X - \bar{X})^2}{N}$$

- Pero está al cuadrado. Es inflada y no tiene sentido. No es el promedio de las desviaciones, pero sí indica variabilidad.

Edad	X - Media	(X - Media) ²
2	-3	9
3	-2	4
4	-1	1
5	0	0
6	1	1
7	2	4
8	3	9
Media = 5	Total = 0	Total = 28

$$\frac{28}{7} = 4 \text{ Pero nadie tiene una desviación tan grande.}$$

Un mejor estimador es la desviación estándar:

$$\sqrt{S_x^2} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

- Sí está en unidades interpretables
- Indica diferencia media con el promedio de la muestras
- Señalar relación con la distribución normal

S_X^2 = Varianza de la muestra

S_X = Desviación estándar de la muestra

σ_X^2 = Varianza observada de la población

σ_X = Desviación estándar observada en la población

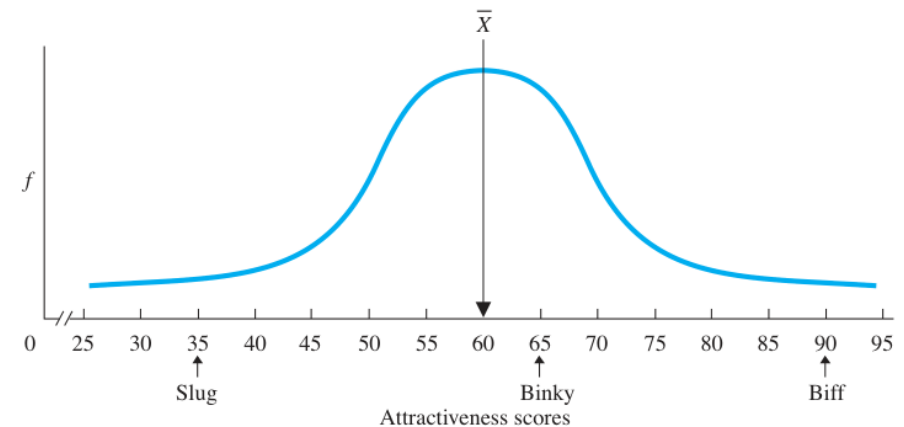
s_X^2 = Varianza estimada de la población

s_X = Desviación estándar estimada en la población

Puntajes z

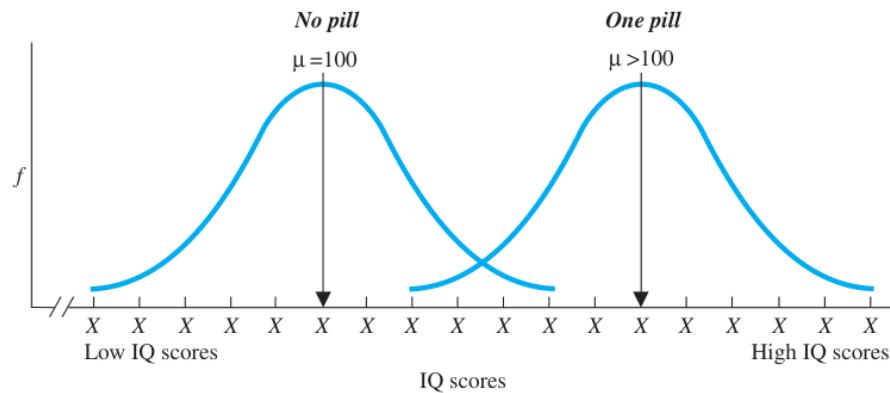
$$z = \frac{X - \bar{X}}{S_X}$$

Describen un puntaje en términos de qué tan lejos se encuentran de la media en desviaciones estándar.



Pruebas de hipótesis

- Establecer si una relación que ocurre en una muestra también ocurre en la población en general
- Probar hipótesis estadísticas H_0 vs H_1 .



$H_0: \mu = 100$

$H_1: \mu \neq 100$

Si la media de la muestra es de 105 parecería que la píldora funcionó. ¿O no?

No necesariamente. Existe el error de muestreo.

No se puede demostrar que la píldora funciona (a menos que se pruebe toda la población), pero sí se puede tener cierta seguridad de que los datos vienen de una distribución distinta de $\mu = 100$.

Prueba z: busca demostrar que una media muestral está lo bastante lejos de la media poblacional para no ser parte de la misma población.

- Se tiene una muestra aleatoria
- La variable dependiente tiene distribución aproximadamente normal
- Se conocen la media y la desviación estándar de la población

