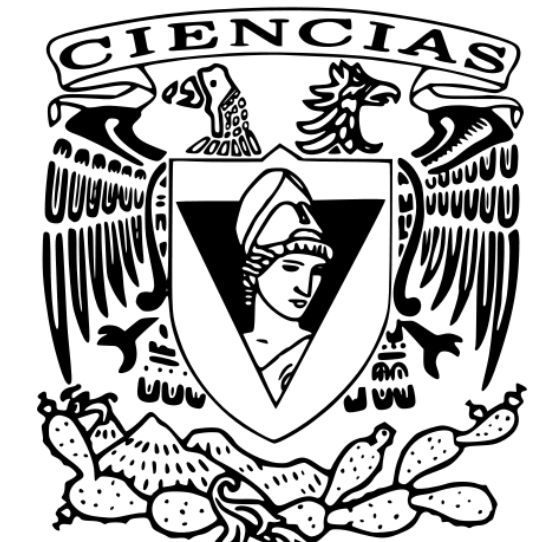


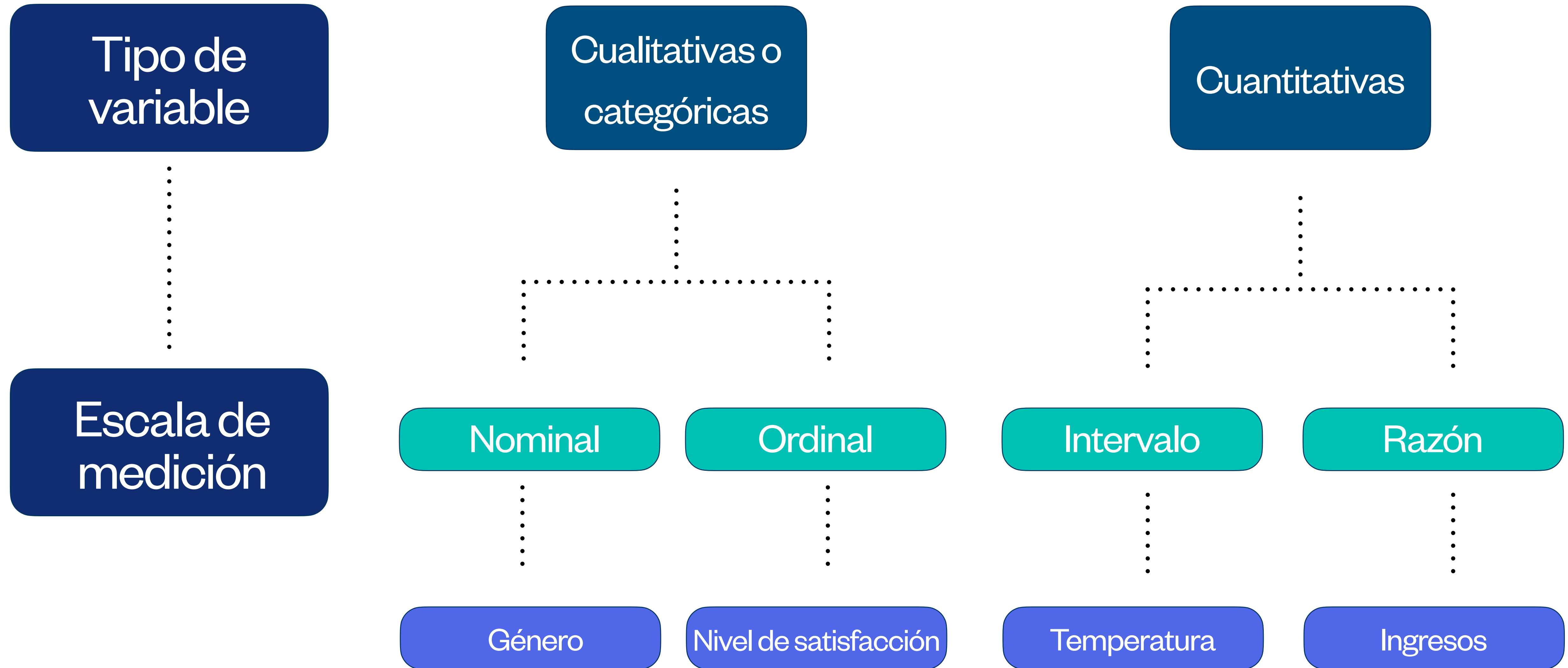
Estadística descriptiva



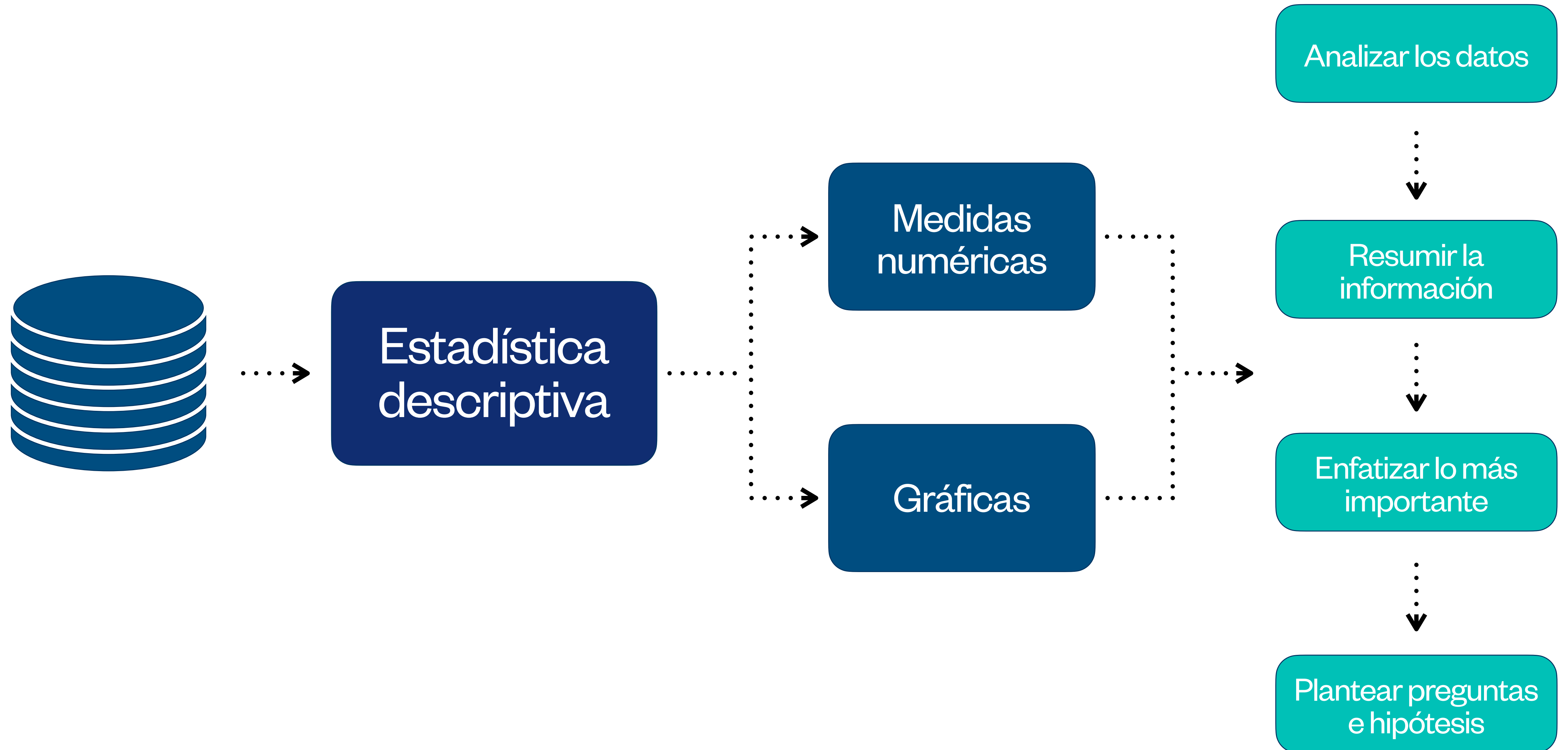
José Antonio Perusquía Cortés
Inferencia Estadística Semestre 2026-I



Datos y variables



¿Qué es la estadística descriptiva?



Medidas numéricas

Tendencia
central

Localización

Dispersión

Forma

Correlación

Medidas de tendencia central

- Indican el valor donde se centran los datos e.g.:

- **Media:** El promedio de las observaciones

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

- **Mediana:** El valor que separa las observaciones en dos partes iguales

$$\text{med}(\mathbf{x}) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ es impar} \\ \frac{1}{2} \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right) & \text{si } n \text{ es par} \end{cases}$$

- **Moda:** El valor más repetido

Medidas de localización

- Indican los valores que dividen la muestra ordenada e.g. :
 - **Cuartiles** : Son los tres valores que dividen las observaciones en cuatro partes iguales y usualmente denotados por $q_{.25}$, $q_{.50}$ (mediana) y $q_{.75}$
 - **Deciles** : Son los nueve valores que dividen la muestra en 10 partes iguales
 - **Percentiles** : Son los 99 valores que dividen la muestra en 100 partes iguales

Medidas de dispersión

- Indican la variabilidad de los datos e.g.:

- **Desviación estándar**

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

- **Rango**: Definido como $r = x_{(n)} - x_{(1)}$

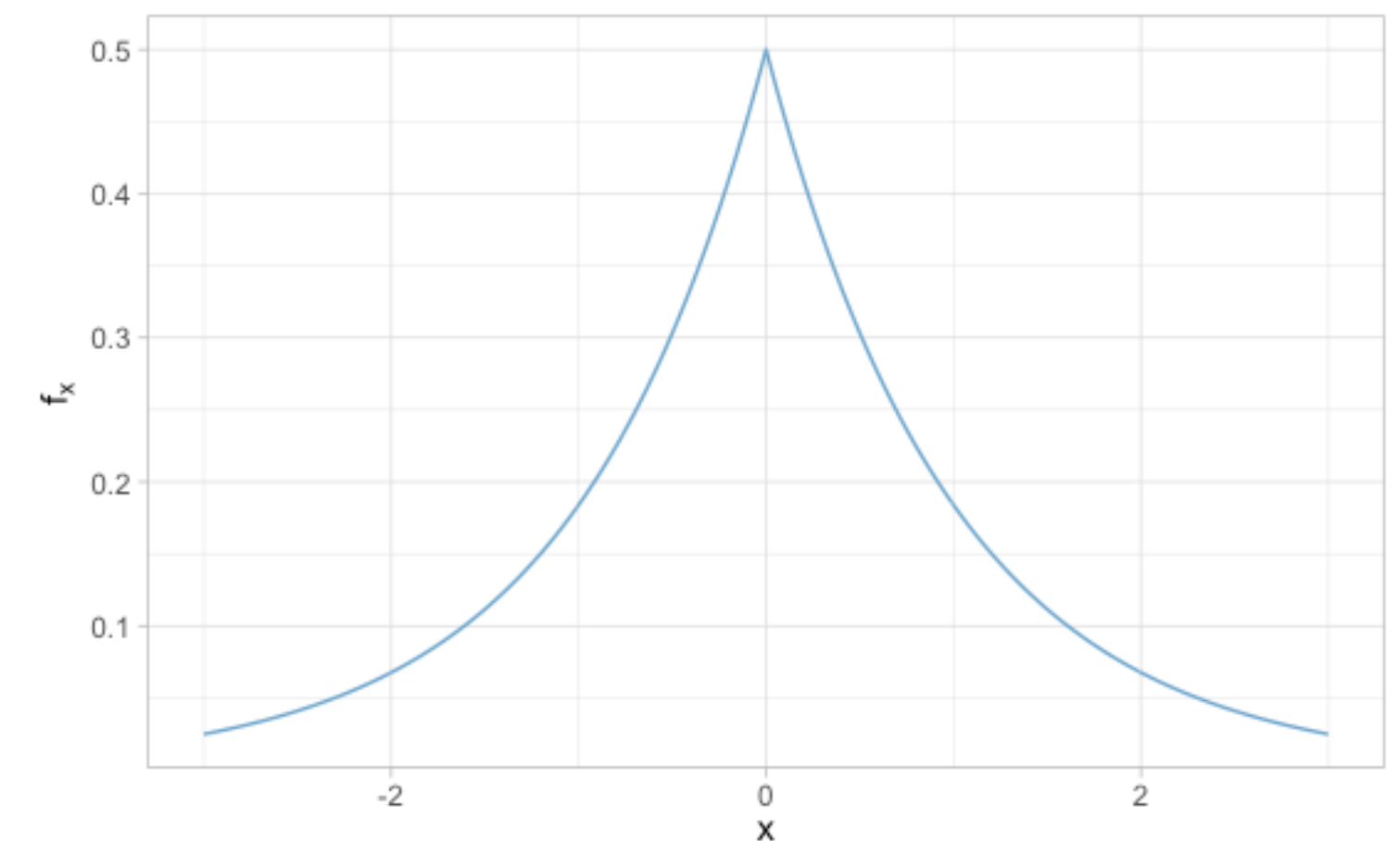
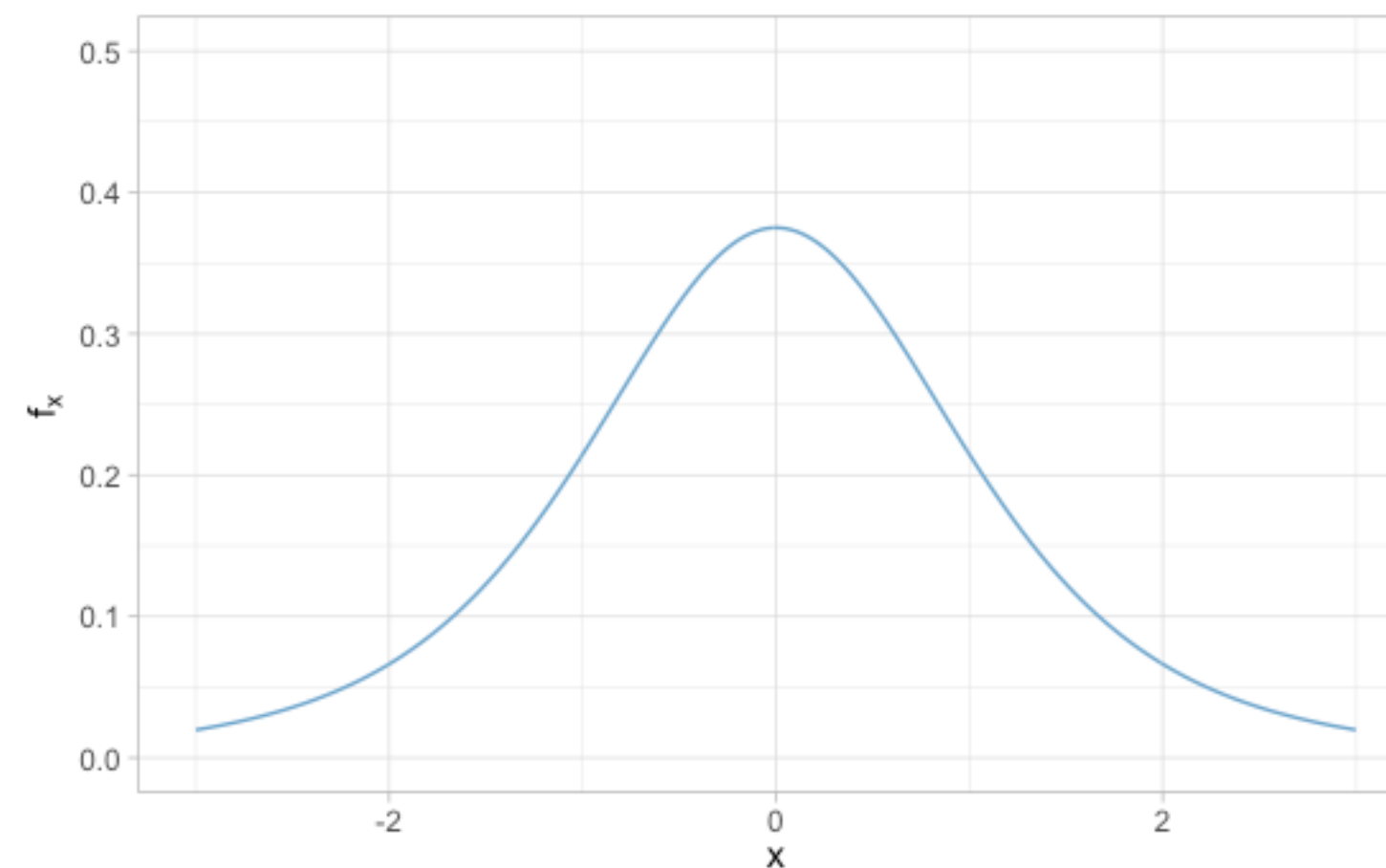
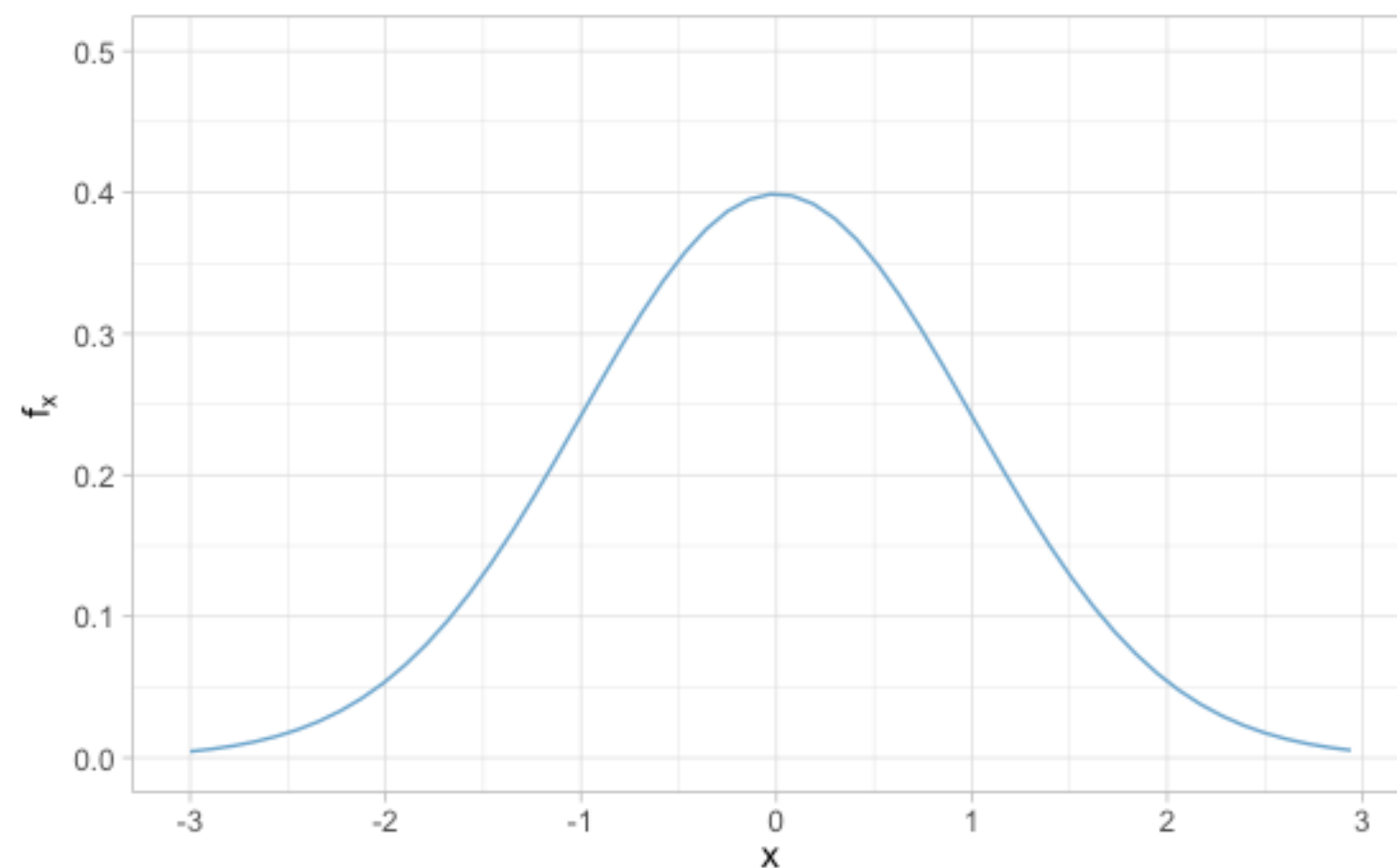
- **Rango intercuartílico**: Utilizado para identificar valores atípicos y definido como

$$\text{IQR} = q_{.75} - q_{.25}$$

- **Coeficiente de variación**: Utilizado para comparar las distribuciones y definido como σ/\bar{x}

Medidas de forma

- Nos indican la forma de la distribución:
 - **Curtosis**: Mide que tan achatada es una distribución en relación a una distribución gaussiana cuya curtosis es 3
 - **Mesocúrtica**: Si la curtosis es igual a 3
 - **Platicúrtica**: Si la curtosis es menor a 3
 - **Leptocúrtica**: Si la curtosis es mayor a 3

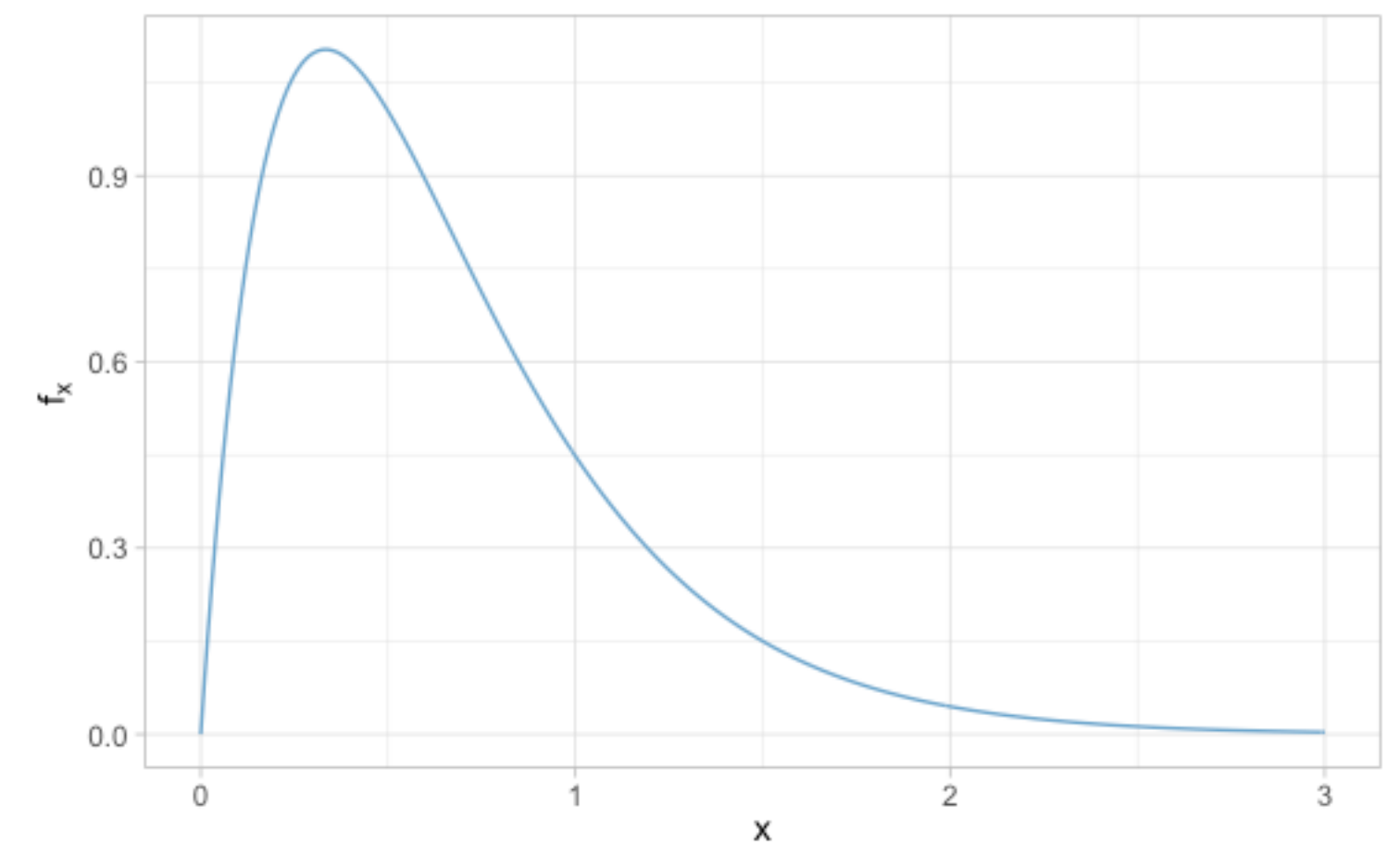
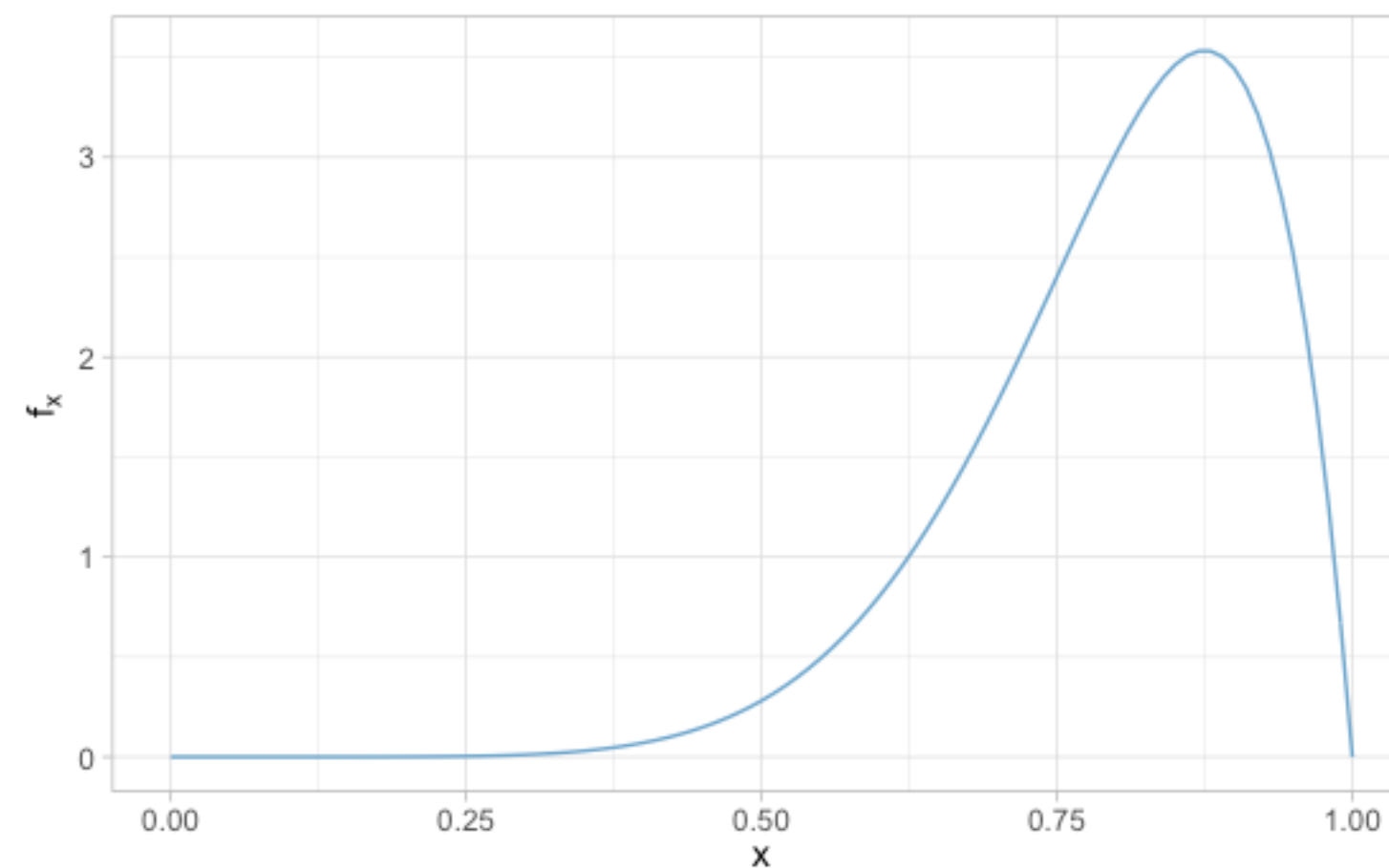
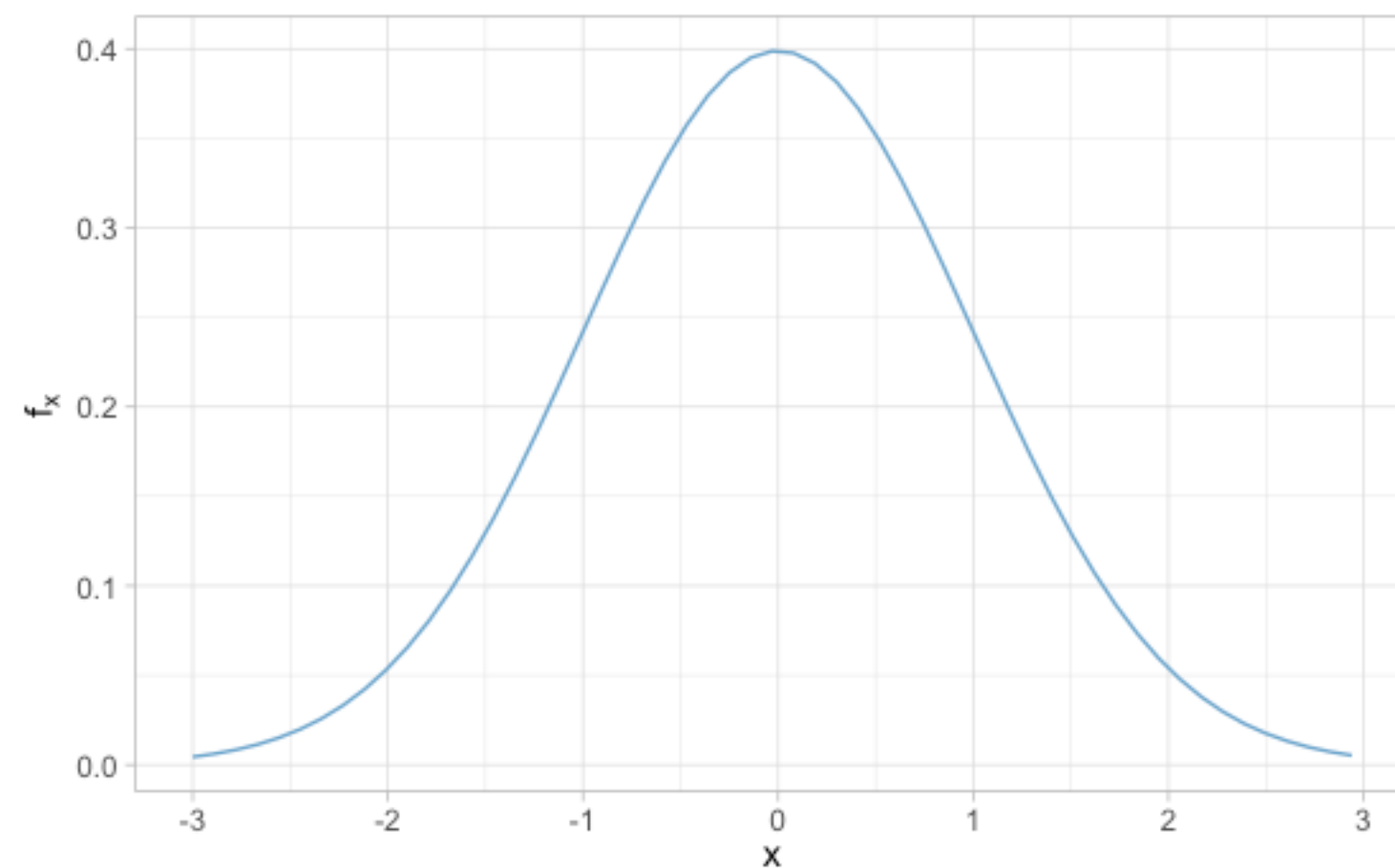


Medidas de forma

- Nos indican la forma de la distribución :

- Coeficiente de asimetría (o sesgo)

- **Simétrica** : si el coeficiente es cero.
- **Asimétrica negativa** (o a la izquierda) : si el coeficiente es menor a cero.
- **Asimétrica positiva** (o a la derecha) : si el coeficiente es mayor a cero.



Medidas de correlación

- Miden el grado de efecto de una variable en otra :
 - **Coeficiente de Pearson.**
 - **Coeficiente ρ de Spearman.**
 - **Coeficiente τ de Kendall.**
- Se estudian a fondo en el curso de Modelos No Paramétricos y de Regresión

Representaciones gráficas

Tabla de
frecuencias

Gráfica de
barras

Histograma

Box plot

Pie/dona

Diagrama de
dispersión

Diagrama de
correlación

Diagrama de
tallo y hojas

¡Muchas más!

Tabla de frecuencias

- Útiles para resumir la información de variables categóricas
- Para una muestra x_1, \dots, x_n
 - Se encuentran los valores únicos x_1^*, \dots, x_k^*
 - Se cuenta el número de veces que estos valores únicos ocurren (**frecuencias**) y se denotan por f_1, \dots, f_k de tal forma que $f_1 + \dots + f_k = n$
 - Se calculan las **frecuencias relativas**, $f_i^* = f_i/n$, de tal forma que $f_1^* + \dots + f_k^* = 1$
 - Se obtienen las **frecuencias acumuladas** como $F_i = f_1^* + \dots + f_i^*$
- Las frecuencias (relativas) se utilizan para construir las gráficas de barras, de pie/dona

Gráficas de barras y pie/dona

