

**Pregunta 1:** ¿Qué diferencia existe entre la *Estadística descriptiva* y la *Inferencia estadística*?

**Pregunta 2:** Defina qué es la frecuencia absoluta y la frecuencia relativa

**Pregunta 3:** Si las variables son discretas ¿cuál es la representación gráfica de la distribución de frecuencias que se espera para una variable bidimensional (X, Y) ?

**Pregunta 4:** Discuta los valores que pueden tomar el *coeficiente de correlación* y la *covarianza*. ¿En qué unidades se mide cada uno? Analice que ventajas o desventajas tiene el usar uno u otro.

**Pregunta 5:** Dibuje esquemáticamente dos gráficos de dispersión para representar las variables (w;t). En el primero de ellos, suponga que existe una correlación lineal inversa entre ambas. En el segundo, suponga que existe una correlación exponencial.

**Pregunta 6:**

La tabla muestra para 6 países, la cantidad de habitantes de cada uno (en millones de personas) y la cantidad de nacimientos (en miles de personas) para el año 2020. Se pide:

- a) Construya un diagrama de dispersión para estos datos.
- b) ¿Usted entiende que puede darse una relación lineal entre la población de un país y la cantidad de nacimientos por año?
- c) Obtenga la recta de regresión.
- d) ¿Qué cantidad de nacimientos es esperable tener en un país que tenga una población de 30.000.000 de habitantes?

Países	Habitantes (Millones personas)	Nacimientos (Miles personas)
A	8.75	114.1
B	12.52	155.1
C	3.34	41.3
D	9.23	122.8
E	25.4	319.1
F	43.4	544.3

## RESPUESTAS:

1)

a) La estadística descriptiva, que se dedica a los métodos de recogida, descripción, visualización y resumen de datos originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numéricamente o gráficamente. Ejemplos básicos de parámetros estadísticos son: la media y la desviación estándar. Algunos ejemplos gráficos son: histograma, pirámide poblacional, clústeres, etc.

b) La inferencia estadística se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para modelar patrones en los datos y extraer inferencias sobre la población de estudio. Estas inferencias pueden tomar la forma de respuestas a preguntas sí/no (prueba de hipótesis), estimaciones de características numéricas (estimación), pronósticos de futuras observaciones, descripciones de asociación (correlación) o modelización de relaciones entre variables (análisis de regresión).

2)

Se llama frecuencia absoluta del valor  $X_i$  al número de veces que aparece repetida la observación en la recopilación de datos. Se representa por  $N_i$ .

Se llama frecuencia relativa del valor  $X_i$  al cociente entre la frecuencia absoluta de  $x_i$  y el número total de datos  $n$ . Se representa por  $f_i$  y, evidentemente, es la proporción en que se encuentra el valor  $X_i$  dentro del conjunto de datos en tanto por uno.

3) Un diagrama de dispersión.

4) La **covarianza** permite discernir si dos variables  $X$  y  $Y$  tienen una relación positiva, negativa o cero, pero no aporta información del grado de dependencia de una variable respecto a la otra (referencias bibliográficas 6, 10 y 17). Además, la covarianza depende de las unidades de medida empleadas para  $X$  y  $Y$ .

Es evidente que, por definición, el **coeficiente de correlación** lineal informa de las mismas cosas que lo hace la covarianza. Además, cumple una propiedad muy importante, está acotado por 1 y por  $-1$ .

• Ser adimensional y siempre estar comprendido entre  $-1$  y  $1$ .  
• Si hay relación lineal fuerte positiva,  $r_{xy} > 0$  y está cerca de  $1$ .  
• Si hay relación lineal negativa fuerte,  $r_{xy} < 0$  y está cerca de  $-1$ .  
• Si no hay relación lineal  $r_{xy}$  será  $0$ .