

Ejemplo Clase 2

Solución problema 2. a)

Para calcular la edad promedio de jóvenes con ingreso entre 20 y 25 mil pesos, aislaremos esa columna, creando una tabla univariada:

Edad	Frec. Absoluta
[18; 20)	1
[20; 22)	2
[22; 24)	4
[24; 26)	1
[26; 28]	4

Por lo que podemos calcular el promedio como ya sabemos. Así, $\bar{x}_{[20;25)} = 23,83$.

Solución problema 2. b)

Sabemos que con "el sueldo más frecuente" se pide la moda. Por lo que aislamos la fila de estudiantes entre 24 y 26 años y construimos una tabla univariada. Con esto, calculamos la moda dándonos cuenta que el intervalo modal es $[15; 20)$. Entonces, $Mo = 16,67$.

Solución problema 2. c)

Para conocer la cantidad de estudiantes, podemos calcular el porcentaje usando el percentil, y de ahí obtener lo pedido. Esta vez aislaremos la fila del total para construir una tabla univariada.

Ingreso	Frec. Absoluta
[10; 15)	16
[15; 20)	18
[20; 25)	12
[25; 30)	12

Ahora, usamos la fórmula del percentil para obtener el porcentaje. Con esto, 23.000 es el percentil 71,04, por lo que un 71,04 % de los 58 estudiantes cumplen nuestro requisito. Por lo tanto, son aproximadamente 41 los estudiantes que ganan a lo más 23 mil pesos a la semana.

Solución problema 2. d)

Para saber qué estrato es más homogéneo, necesitamos calcular el coeficiente de variación; éste requiere de la media y la desviación estándar de los estratos.

Para, el intervalo $[24; 26)$, la media es 17,5 y la desviación estándar es 4,47; por lo que el coeficiente de variación es 0,26.

Para el siguiente intervalo, la media es 18,75 y la desviación estándar es 5,99; entonces, el coeficiente de variación es 0,32.

Por lo tanto, sí podemos afirmar que el ingreso de los estudiantes de entre 24 y 26 años es más homogéneo que el de los estudiantes de entre 18 y 20 años.

Solución problema 2. e)

Para saber si las variables están relacionadas, calcularemos el coeficiente de correlación. Para esto, necesitamos la covarianza de X e Y , la varianza de cada variable y la media de cada variable.

$$\bar{X} = 23,9 ; \bar{Y} = 19,22 ; S_X^2 = 8,16 ; S_Y^2 = 29,79$$

$$\text{Cov}(X, Y) = 1,2128 \Rightarrow r \approx 0,08$$

Por lo tanto, como r es muy pequeño, concluimos que las variables están muy débilmente relacionadas directamente.

Solución problema 2. f)

Para realizar un modelo lineal, necesitamos calcular los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$, que se obtienen a partir de la covarianza de ambas variables, la media de ambas variables, y la varianza de X . Afortunadamente, calculamos estos valores en el paso anterior.

$$\hat{\beta}_1 = 0,149 ; \hat{\beta}_0 = 15,67$$

Por lo que nuestro modelo lineal queda como $y = 15,67 + 0,149x$.

Solución problema 2. g)

Para comprobar la calidad de nuestro modelo, podemos calcular el parámetro R^2 , que es lo mismo que elevar al cuadrado el coeficiente de correlación. Como ya tenemos calculado éste último, basta elevar a 2.

$$R^2 = r^2 = 0,0061$$

Como $R^2 \approx 0$ consideramos que el modelo no es bueno.