

Informe de Examen de Estadística y Probabilidad

Universidad Nacional Hermilio Valdizán
Facultad de Economía

23 de octubre de 2024

Estudiante: Palomino Ricaldi Antony

1. Problema 1: Derivación de la Fórmula de la Moda para Datos Clasificados

La fórmula de la moda para datos clasificados se expresa como:

$$Mo = L_i + A \cdot \left(\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right)$$

Donde:

- L_i = Límite inferior de la clase modal
- A = Amplitud del intervalo
- f_i = Frecuencia de la clase modal
- f_{i-1} = Frecuencia de la clase anterior a la modal
- f_{i+1} = Frecuencia de la clase posterior a la modal

2. Problema 2: Análisis de Ventas de Gasolina

2.1. Planteamiento del Problema

Se analizan las ventas de gasolina de una refinería de Petro-Perú en Ica, con datos clasificados en intervalos de 10,000 galones.

2.2. Solución en R

Listing 1: Análisis de Ventas de Gasolina

```
1  
2 # Crear el dataframe con los datos  
3 ventas_gasolina <- data.frame(  
4     intervalo_inferior = seq(0, 70, by = 10),  
5     intervalo_superior = seq(10, 80, by = 10),
```

```

6     frecuencia = c(10, 20, 30, 25, 15, 10, 5, 5)
7 )
8
9 # Calcular marca de clase
10 ventas_gasolina$marca_clase <- (
11     ventas_gasolina$intervalo_inferior +
12     ventas_gasolina$intervalo_superior
13 ) / 2
14
15 # a) Calculo del total de galones vendidos
16 total_galones <- sum(
17     ventas_gasolina$marca_clase *
18     ventas_gasolina$frecuencia
19 )
20
21 # b) Media de galones por operacion
22 n_operaciones <- sum(ventas_gasolina$frecuencia)
23
24 media_galones <- total_galones / n_operaciones
25
26 # c) Moda
27 clase_modal <- which.max(ventas_gasolina$frecuencia)
28 Li <- ventas_gasolina$intervalo_inferior[clase_modal]
29 A <- 10 # Amplitud del intervalo
30 fi <- ventas_gasolina$frecuencia[clase_modal]
31 fi_anterior <- ventas_gasolina$frecuencia[clase_modal - 1]
32 fi_posterior <- ventas_gasolina$frecuencia[clase_modal + 1]
33
34 moda <- Li + A * (
35     (fi - fi_anterior) / ((fi - fi_anterior) + (fi -
36         fi_posterior))
37 )
38
39 # d) Calcule la mediana de las ventas.
40
41 ventas_gasolina$f_a_acumulada <-
42     cumsum(ventas_gasolina$frecuencia)
43
44
45 intervalo_medinal <- which(
46     ventas_gasolina$f_a_acumulada >= n_operaciones / 2
47 ) [1]
48
49 Li <- ventas_gasolina$intervalo_inferior[intervalo_medinal]
50 fi <- ventas_gasolina$frecuencia[intervalo_medinal]
51 Fi_1 <- ventas_gasolina$f_a_acumulada[intervalo_medinal - 1]
52
53 mediana <- Li + A / fi * (n_operaciones / 2 - Fi_1)

```

2.3. Resultados e Interpretación

1. Total de galones vendidos: 3 900 (miles)

- *Interpretación:* La refinería vendió aproximadamente 3 900 000 galones de gasolina durante la semana.

2. Media de galones por operación: 32.5 (miles)

- *Interpretación:* En promedio, cada operación de venta fue de aproximadamente 32 500 galones.

3. Moda: 26.66667 (miles)

- *Interpretación:* El volumen de venta más frecuente se encuentra en aproximadamente 26 667 galones, lo cual está por encima de 25 000.

4. Mediana de las ventas: 30 (miles)

- *Interpretación:* La mediana de las ventas es 30 000.

3. Problema 3: Análisis de Trabajadores en Empresas

3.1. Solución en R

Listing 2: Análisis de Trabajadores en Empresas

```

1 # Obtener el valor de las variables
2 calculate_sturges_params <- function(data) {
3   R <- diff(range(data))
4   K <- ceiling(1 + 3.322 * log10(length(data)))
5   A <- ceiling(R / K)
6
7   list(
8     range = range(data),
9     num_classes = K,
10    class_width = A
11  )
12 }
13
14 create_frequency_table <- function(data) {
15   n <- length(data)
16   params <- calculate_sturges_params(data)
17
18   # Crear la secuencia de limites
19   rule <- seq(params$range[1],
20     params$range[2] + params$class_width,
21     by = params$class_width
22   )
23
24   # Calcular el ancho de clase
25   marca_de_clase <- (rule[-length(rule)] + rule[-1]) / 2
26
27   # Crear los intervalos

```

```

28 i <- cut(data, breaks = rule, right = FALSE)
29 tabla <- table(i)
30
31 # Calcular las frecuencias
32 f_absoluta <- as.numeric(tabla)
33 f_a_acumulada <- cumsum(f_absoluta)
34 f_relativa <- f_absoluta / n
35 f_r_acumulada <- cumsum(f_relativa)
36 f_porcentual <- f_relativa * 100
37 f_p_acumulada <- cumsum(f_porcentual)
38
39 # Create data frame
40 data.frame(
41   intervalos = levels(i),
42   f_absoluta = f_absoluta,
43   f_a_acumulada = f_a_acumulada,
44   f_relativa = f_relativa,
45   f_r_acumulada = f_r_acumulada,
46   f_porcentual = f_porcentual,
47   f_p_acumulada = f_p_acumulada,
48   marca_de_clase = marca_de_clase
49 )
50 }
51
52 calculate_grouped_statistics <- function(freq_table, data,
53   class_width) {
54   n <- length(data)
55
56   # Media para datos agrupados
57   media_agrupada <- sum(freq_table$marca_de_clase *
58     freq_table$f_absoluta) / n
59
60   # Moda para datos agrupados
61   intervalo_modal <- which.max(freq_table$f_absoluta)
62   Li <- as.numeric(
63     sub("\\[(.+),.*", "\\1",
64       freq_table$intervalos[intervalo_modal])
65   )
66   fi <- freq_table$f_absoluta[intervalo_modal]
67   fi_menos_1 <- if (intervalo_modal > 1) {
68     freq_table$f_absoluta[intervalo_modal - 1]
69   } else {
70     0
71   }
72   fi_mas_1 <- if (intervalo_modal < nrow(freq_table)) {
73     freq_table$f_absoluta[intervalo_modal + 1]
74   } else {
75     0
76   }
77   moda_agrupada <- Li + class_width *
78     ((fi - fi_menos_1) / ((fi - fi_menos_1) + (fi - fi_mas_1)))

```

```

76
77 # Mediana para datos agrupados
78 intervalo_medinal <- which(freq_table$f_a_acumulada >= n /
79 2)[1]
80 Li <- as.numeric(
81 sub("\\[(.+)\\.\\1", "\\1",
82 freq_table$intervalos[intervalo_medinal])
83 )
84 fi <- freq_table$f_absoluta[intervalo_medinal]
85 Fi_1 <- if (intervalo_medinal > 1) {
86 freq_table$f_a_acumulada[intervalo_medinal - 1]
87 } else {
88 0
89 }
90 mediana_agrupada <- Li + class_width / fi * (n / 2 - Fi_1)
91
92 list(
93 media = media_agrupada,
94 mediana = mediana_agrupada,
95 moda = moda_agrupada
96 )
97 }
98
99 plot_frequency_diagram <- function(freq_table, class_width) {
100 # Extraer los intervalos inferiores
101 lower_bounds <- as.numeric(sub("\\[(.+)\\.\\1", "\\1",
102 freq_table$intervalos))
103
104 # Crear diagrama
105 plot(lower_bounds, freq_table$f_a_acumulada,
106 type = "s",
107 xlab = "Numero de trabajadores",
108 ylab = "Frecuencia acumulada",
109 main = "Diagrama Escalonado con Ojiva",
110 col = "blue",
111 lwd = 2
112 )
113
114 # Agregar ojiva
115 lines(freq_table$marca_de_clase,
116 freq_table$f_a_acumulada,
117 type = "b",
118 col = "red",
119 pch = 16,
120 lwd = 2
121 )
122
123 # Agregar etiquetas
124 text(freq_table$marca_de_clase,
125 freq_table$f_a_acumulada,
126 labels = freq_table$f_a_acumulada,

```

```

124     pos = 3,
125     col = "black"
126   )
127 }
128
129 # Funcion principal para el analisis de datos
130 analyze_data <- function(data) {
131   # Obtener las variables
132   params <- calculate_sturges_params(data)
133
134   # Crear la tabla de frecuencia
135   freq_table <- create_frequency_table(data)
136
137   # Calcular los estadisticos para datos no agrupados
138   ungrouped_stats <- list(
139     media = mean(data),
140     mediana = median(data),
141     moda = as.numeric(names(sort(table(data), decreasing =
142       TRUE)[1]))
143   )
144
145   # Calcular los estadisticos para datos agrupados
146   grouped_stats <- calculate_grouped_statistics(
147     freq_table, data, params$class_width
148   )
149
150   # Retornar los resultados
151   list(
152     frequency_table = freq_table,
153     ungrouped_statistics = ungrouped_stats,
154     grouped_statistics = grouped_stats,
155     plot = function() {
156       plot_frequency_diagram(freq_table, params$class_width)
157     }
158   )
159 }

```

3.2. Fórmulas Utilizadas

1. Media Aritmética:

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

2. Mediana para datos agrupados:

$$Me = L_i + A \cdot \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right)$$

3. Moda para datos agrupados:

$$Mo = L_i + A \cdot \left(\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right)$$

4. Regla de Sturges:

$$K = 1 + 3,322 \log_{10}(n)$$

Listing 3: Numero de trabajadores encuestados por empresa

```

1
2 datos_empresas <- c(
3     73, 95, 61, 46, 70, 55, 87, 65,
4     75, 48, 69, 75, 75, 39, 63, 82,
5     58, 43, 38, 64, 69, 79, 47, 63,
6     63, 81, 59, 77, 84, 34, 75, 93,
7     67, 89, 66, 52, 59, 36, 62, 43,
8     75, 52, 59, 87, 74, 30, 95, 38,
9     50, 72, 44, 53, 68, 72, 82, 63
10 )
11
12 # Instancia de las funcion principal
13 results <- analyze_data(datos_empresas)

```

a) Elaborar la tabla de frecuencias usando la regla de Sturges. (2 puntos)

Listing 4: Mostrar tabla de frecuencia

```

1 print(results$frequency_table)

```

Tabla de Frecuencias

intervalos	f_absoluta	f_a_acumulada	f_relativa	f_r_acumulada	f_porcentual	f_p_acumulada	marca_de_clase
[30,40)	6	6	0.10714286	0.1071429	10.714286	10.71429	35
[40,50)	6	12	0.10714286	0.2142857	10.714286	21.42857	45
[50,60)	9	21	0.16071429	0.3750000	16.071429	37.50000	55
[60,70)	13	34	0.23214286	0.6071429	23.214286	60.71429	65
[70,80)	12	46	0.21428571	0.8214286	21.428571	82.14286	75
[80,90)	7	53	0.12500000	0.9464286	12.500000	94.64286	85
[90,100)	3	56	0.05357143	1.0000000	5.357143	100.00000	95

Figura 1: Tabla de distribución de frecuencia

b) Interpretar: f_5 , h_2 . (2 puntos)

Listing 5: Obtener celdas especificas

```

1 cat("f5:", results$frequency_table$f_absoluta[5], "\n")
2
3 > f5: 12

```

Representa la **frecuencia absoluta** del quinto intervalo, que corresponde al intervalo [70 - 80). Esto significa que hay 12 empresas cuyos encuestados se encuentran dentro de este intervalo específico.

Listing 6: Obtener celdas específicas

```
1 cat("h2:", results$frequency_table$f_relativa[2], "\n")
2
3 > h2: 0.1071429
```

Representa la **frecuencia relativa** del segundo intervalo, que corresponde al intervalo [40 - 50). El valor 0.107 significa que la proporción de las empresas que tiene un promedio de [40 - 50) empleados encuestados.

c) Elaborar el diagrama escalonado y la ojiva respectiva. (2 puntos)

Listing 7: Mostrar tabla de frecuencia

```
1 results$plot()
```

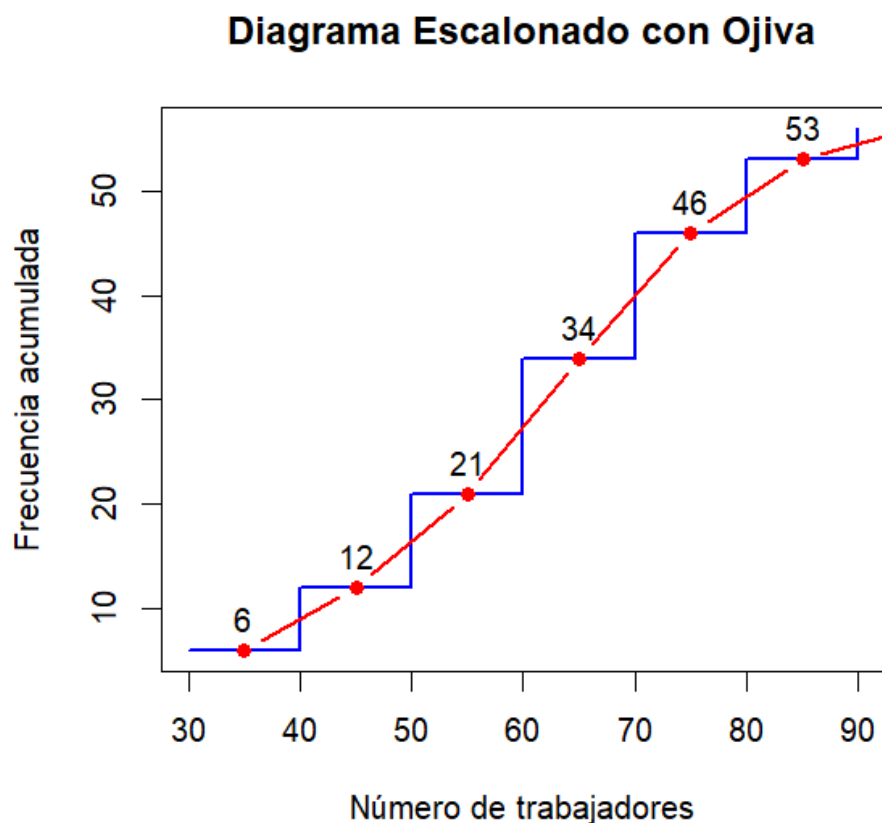


Figura 2: Diagrama escalonado

d) Calcular la media aritmética (\bar{x}) y la mediana (Me) para datos agrupados y no agrupados. (2 puntos)

Listing 8: Mostra resultados

```
1 # Media aritmetica agrupada
```



```

2 print(results$grouped_statistics$media)
3 # Media aritmetica no agrupada
4 print(results$ungrouped_statistics$media)
5
6 # Mediana agrupada
7 print(results$grouped_statistics$mediana)
8 # Mediana no agrupada
9 print(results$ungrouped_statistics$mediana)

```

Listing 9: Resultados en consola

```

1 > # Media aritmetica agrupada
2 > print(results$grouped_statistics$media)
3 [1] 64.28571
4 > # Media aritmetica no agrupada
5 > print(results$ungrouped_statistics$media)
6 [1] 64.16071
7 >
8 > # Mediana agrupada
9 > print(results$grouped_statistics$mediana)
10 [1] 65.38462
11 > # Mediana no agrupada
12 > print(results$ungrouped_statistics$mediana)
13 [1] 64.5

```

e) Calcular la moda (Mo) para datos agrupados y no agrupados. (2 puntos)

Listing 10: Mostra resultados

```

1 # Moda agrupada
2 print(results$grouped_statistics$moda)
3 # Moda no agrupada
4 print(results$ungrouped_statistics$moda)

```

Listing 11: Resultados en consola

```

1 > # Moda agrupada
2 > print(results$grouped_statistics$moda)
3 [1] 68
4 > # Moda no agrupada
5 > print(results$ungrouped_statistics$moda)
6 [1] 75

```