



TRABAJO PRACTICO INTEGRADOR

PROBABILIDAD Y ESTADISTICA

ALUMNO: FABIAN IGNACIO CARDOZO

TRABAJO PRÁCTICO INTEGRADOR

Consignas a resolver primera entrega

1. Haciendo uso del vocabulario técnico, identificar con relación al caso propuesto.
 - a. Población y Muestra
 - b. Unidad Elemental
 - c. Variables en estudio, detallando su tipo.

a. Población y muestra:

En el contexto del caso planteado, se entiende por **Población** al conjunto total de estudiantes que cursan el segundo año de la Tecnicatura Universitaria en Programación en la Universidad INNOVA XXII; de dicho grupo se desea obtener conclusiones generales.

Por otro lado, la **Muestra** corresponde al subconjunto de esa población, ósea a los alumnos que participaron en la encuesta realizada por la consultora STUDIO X. Este subconjunto fue seleccionado para relevar información específica y, mediante técnicas estadísticas, permitirá inferir resultados válidos para la población en su totalidad.

b. Unidad elemental:

La unidad elemental es el objeto individual de estudio a partir del cual se recolectan los datos. En este caso, cada estudiante encuestado del segundo año de la carrera constituye una unidad elemental, ya que sobre él se registran las características personales, académicas y socioeconómicas incluidas en el cuestionario.

c. Variables en estudio y su tipología:

El cuestionario aplicado permite identificar diversas variables estadísticas, cada una con naturaleza y escala de medición específicas. A continuación, se detallan junto con su clasificación técnica:

- Edad (en años cumplidos): variable cuantitativa discreta (numérica entera).
- Género (Femenino, Masculino, Otro): variable cualitativa nominal (categórica sin orden).
- Peso (kg): variable cuantitativa continua (magnitud con posibles valores en un rango real).
- Estatura (cm): variable cuantitativa continua.
- Número de hermanos: variable cuantitativa discreta.

- Condición de fumador (Sí/No): variable cualitativa nominal dicotómica.
- Horas de estudio semanal en el hogar: variable cuantitativa continua.
- Condición laboral (Trabaja: Sí/No): variable cualitativa nominal dicotómica.
- Gastos semanales en alimentación (pesos): variable cuantitativa continua.
- Cantidad de materias aprobadas: variable cuantitativa discreta.
- Nivel de satisfacción con la carrera (Muy satisfecho, Satisfecho, Insatisfecho, Muy insatisfecho): variable cualitativa ordinal (posee categorías con un orden implícito).

2. Construir la/s Tabla/s de Frecuencias y calcular todas las frecuencias de las siguientes variables:

- a. Tiempo en horas semanales dedicadas al estudio. (Determinar la cantidad optima de intervalos a utilizar)

```
CARDOZO FABIAN IGNACIO - P1y2.R*
library(readxl)
# a). Vamos a determinar la cantidad optima de intervalos a utilizar
# Empezamos guardando archivo Excel en una Variable y Leyendola
Archivo <- file.choose()
datos <- read_excel(Archivo)
# En una variable guardamos los datos extraidos del vector "TIEMPO SEMANAL en HS. DEDIC. EST"
tiempo <- datos$`TIEMPO SEMANAL en HS. DEDIC. EST.`
# Usamos la fórmula de Sturges
n <- length(tiempo)
k <- ceiling(1 + 3.322*log10(n))
k
```

c. A partir de la tabla obtenida en el punto a. realizar la interpretación de todas las frecuencias correspondientes al cuarto intervalo en el contexto del caso planteado.

```
8 # c).Determinaremos el rango de valores de la variable "Tiempo de estudio semanal"
9 rango <- range(tiempo)
10
11 # Calculamos la amplitud (ancho de clase) utilizando la fórmula:
12 # (máximo - mínimo) / k , donde k es el número de intervalos según Sturges
13 # Usamos ceiling() para redondear hacia arriba y asegurar que todos los datos queden cubiertos
14 amplitud <- ceiling((rango[2] - rango[1]) / k)
15
16 # Visualizamos el rango y la amplitud obtenida
17 rango
18 amplitud
19
20 # Construimos los puntos de corte (breaks) para los intervalos de clase
21 # seq() genera una secuencia desde el valor mínimo redondeado hacia abajo (floor)
22 # hasta el máximo redondeado hacia arriba (ceiling) + amplitud,
23 # con un paso igual a la amplitud calculada
24 breaks <- seq(floor(rango[1]), ceiling(rango[2]) + amplitud, by= amplitud)
25 # Clasificamos los datos en intervalos utilizando cut()
26 # right = FALSE indica que los intervalos son de la forma [a, b)
27 # es decir, incluyen el límite inferior y excluyen el superior (excepto el último)
28 clases <- cut(tiempo, breaks = breaks, right = FALSE)
29 # Visualizamos las primeras clasificaciones de intervalos para verificar el resultado
30 head(clases)
31
32 # Calculamos la tabla de frecuencias absolutas (fa) para cada intervalo
33 Tabla_Tiempo <- table(clases)
34 # Calculamos la frecuencia absoluta acumulada (faa)
35 f_acum <- cumsum(Tabla_Tiempo)
36 # Calculamos la frecuencia relativa (fr) como proporción sobre el total
37 f_rel <- prop.table(Tabla_Tiempo)
38 # Calculamos la frecuencia relativa acumulada (fra)
39 f_rel_acum <- cumsum(f_rel)
40
41 # Construimos un Data Frame que consolida toda la información en una tabla de frecuencias
42 Tiempo_Semanal <- data.frame(
43   Tiempo_Estad = levels(clases),           # Intervalos de clase
44   Frecuencia = as.vector(Tabla_Tiempo),    # Frecuencia absoluta
45   Frec_Acum = as.vector(f_acum),           # Frecuencia absoluta acumulada
46   Frec_Relativa = round(as.vector(f_rel), 4), # Frecuencia relativa
47   Frec_Rel_Acum = round(as.vector(f_rel_acum), 4) # Frecuencia relativa acumulada
48 )
49 Tiempo_Semanal[4, ]
```

	Tiempo_Estad	Frecuencia	Frec_Acum	Frec_Relativa	Frec_Rel_Acum
1	[1,5)	23	23	0.092	0.092
2	[5,9)	52	75	0.208	0.300
3	[9,13)	73	148	0.292	0.592
4	[13,17)	47	195	0.188	0.780
5	[17,21)	22	217	0.088	0.868
6	[21,25)	12	229	0.048	0.916
7	[25,29)	12	241	0.048	0.964
8	[29,33)	9	250	0.036	1.000

A partir del cuarto Intervalo [13, 17] podemos Interpretar que, según la tabla:

- La Frecuencia es de: 47
- La Frecuencia acumulada es de: 195
- La Frecuencia relativa es de: 0.188
- La Frecuencia relativa acumulada es de: 0.780

Podemos observar que el tiempo dedicado al estudio semanal, correspondiente al rango de (13,17) horas, registra 47 estudiantes, los que representan el 18,8% del total encuestado. Este grupo constituye una proporción significativa, lo que sugiere que muchos estudiantes mantienen una dedicación moderada-alta al estudio. Al observar la frecuencia acumulada (195) y la frecuencia relativa acumulada (78%), se puede inferir que la mayoría de los estudiantes dedican menos de 17 horas semanales al estudio, lo cual podría estar vinculado a la carga académica .

b. Nivel de satisfacción con la Carrera.

d. A partir de la tabla obtenida en el punto “b” realizar la interpretación de todas las frecuencias correspondientes a la categoría “Satisfecho”.

```
62 # d). Transformamos la variable "SATISFACCIÓN CON LA CARRERA" en un factor con etiquetas descriptivas
63 # levels = define los códigos originales (1, 2, 3, 4)
64 # labels = asigna el nombre real de cada categoría (Muy satisfecho, Satisfecho, etc.)
65 satisfaccion <- factor(datos$'SATISFACCIÓN CON LA CARRERA',
66                       levels = c(1, 2, 3, 4),
67                       labels = c("Muy satisfecho", "Satisfecho", "Insatisfecho", "Muy insatisfecho"))
68
69
70 # b). Calculamos la tabla de frecuencias absolutas (fa) para cada categoría de satisfacción
71 Tabla_satisfaccion <- table(satisfaccion)
72 # Calculamos la frecuencia absoluta acumulada (faa) sumando las frecuencias categoría por categoría
73 frec_acum <- cumsum(Tabla_satisfaccion)
74 frec_rel <- prop.table(Tabla_satisfaccion)
75 frec_rel_acum <- cumsum(frec_rel)
76
77 # Construimos un Data Frame que consolida toda la información en una tabla de frecuencias
78 Tabla_final <- data.frame(
79   Satisfaccion= names(Tabla_satisfaccion),
80   Frecue = as.vector(Tabla_satisfaccion),
81   Frecu_Acum = as.vector(frec_acum),
82   Frecu_Rela = round(as.vector(frec_rel), 4),
83   Frecu_Rela_Acum = round(as.vector(frec_rel_acum), 4)
84 )
85
86
87 # Mostramos la tabla de frecuencias final
88 Tabla_final
89
```

d. A partir de la tabla obtenida en el punto “b” realizar la interpretación de todas las frecuencias correspondientes a la categoría “Satisfecho”.

	Satisfaccion	Frecue	Frecu_Acum	Frecu_Rela	Frecu_Rela_Acum
1	Muy satisfecho	143	143	0.572	0.572
2	Satisfecho	76	219	0.304	0.876
3	Insatisfecho	16	235	0.064	0.940
4	Muy insatisfecho	15	250	0.060	1.000

A partir del análisis de la categoría “Satisfecho”, según tu tabla obtenida observamos que:

- La Frecuencia es de : 76
- La Frecuencia relativa es de: 0.304
- La Frecuencia acumulada es de: 219
- La Frecuencia relativa acumulada es de: 0.876

La categoría “Satisfecho” agrupa a 76 estudiantes, lo que representa el 30.4% del total encuestado. Este valor indica que una proporción considerable de alumnos tiene una percepción positiva de la carrera, aunque no extrema. Al observar la frecuencia relativa acumulada (87.6%), se puede inferir que la gran mayoría de los estudiantes se encuentran entre satisfechos y muy satisfechos, lo cual refleja una tendencia general favorable hacia la propuesta académica.

3. Calcular, para las variables definidas en el punto 2, las medidas descriptivas de tendencia central, posición y dispersión, interpretando sus resultados en términos del problema planteado. (Para la variable categórica solo calcular Moda, Mediana y Cuartiles).

Cálculo de las medidas descriptivas de tendencia central, posición y dispersión.

Variable Discreta Satisfacción.

```
93
94 # 3. Cálculo de las Medidas Descriptivas de Tendencia Central.(Satisfacción)
95
96 # Convertimos la variable ordinal en numérica para poder calcular estadísticas.
97 # Se transforma una variable categórica ordinal en numérica para aplicar medidas estadísticas.
98
99 datos$Satisfaccion <- as.numeric(factor(datos$`SATISFACCIÓN CON LA CARRERA`,
100                                     levels = c(1, 2, 3, 4),
101                                     labels = c("Muy satisfecho", "Satisfecho", "Insatisfecho", "Muy insatisfecho")))
102
103 # Definimos la variable y calculamos medidas de tendencia central.
104 # Se calculan media, mediana, moda y cuartiles para la variable discreta.
105 variable_discreta <- "Satisfaccion"
106 media_discreta <- mean(datos[[variable_discreta]], na.rm= TRUE)
107 mediana_discreta <- median(datos[[variable_discreta]], na.rm= TRUE)
108 moda_discreta <- mlv(datos[[variable_discreta]], method = "mfv")
109 cuartiles <- quantile(datos[[variable_discreta]], probs=c(0.25,0.5,0.75), na.rm=TRUE)
110
111 # Creamos las tablas para poder mostrar los resultados
112
113 message("\nEstadísticos Descriptivos: ")
114 Tabla_Discreta <- data.frame(
115   Media = round(media_discreta, 4),
116   Mediana = round(mediana_discreta, 4),
117   Moda = moda_discreta
118 )
119
120 # Tabla para cuartiles.
121 Tabla_Cuartil <- data.frame(
122   Cuartil = c("Q1 (25%)", "Q2 (Mediana)", "Q3 (75%)"),
123   Valor = as.numeric(cuartiles)
124 )
125
126 print(Tabla_Discreta)
127 print(Tabla_Cuartil)
128
```


Variable Continua Tiempo.

```
131 # Medidas para Variable Continua (Tiempo de estudio)
132
133
134 datos$tiempo <- tiempo
135 variable_continua <- "tiempo"
136 k <- ceiling(1 + 3.322*log10(nrow(datos))) # Obtenemos el número óptimo de clases a formar, es decir que Se determina
137 # el número de intervalos para agrupar los datos.
138
139 # Se calculan los cortes de clase para agrupar los datos. Se define el rango y amplitud de cada clase.
140 min_valor <- floor(min(datos[[variable_continua]], na.rm=TRUE))
141 max_valor <- ceiling(max(datos[[variable_continua]], na.rm=TRUE))
142 amplitud <- ceiling((max_valor - min_valor)/k)
143 max_topes <- min_valor + amplitud * k
144 cortes <- seq(min_valor, max_topes, by = amplitud)
145
146
147 datos$clases <- cut(datos[[variable_continua]], breaks = cortes,
148   rigth = FALSE, include.lowest= TRUE)
149
150 marca_clases <- (head(cortes, -1) + tail(cortes, -1)) /2
151
152 # Se calcula frecuencia absoluta, acumulada, relativa y relativa acumulada.
153 tabla_clases <- table(datos$clases)
154 fr_acum <- cumsum(tabla_clases)
155 fr_rel <- prop.table(tabla_clases)
156 fr_rel_acum <- cumsum(fr_rel)
157
158 # Se construye la Tabla de Frecuencia
159 tabla_de_frecuencia <- data.frame(
160   Intervalo = names(tabla_clases),
161   Marca = as.vector(marca_clases),
162   Frec_Abs = as.vector(tabla_clases),
163   Frec_Acum = as.vector(fr_acum),
164   Frec_Rel = round(as.vector(fr_rel), 4),
165   Frec_Rel_Acum = round(as.vector(fr_rel_acum), 4)
166 )
167 print(tabla_de_frecuencia)
168
169
170
171
172 # Cálculos Estadísticos Agrupados para Variable Continua
173
174 #Cálculo de Media
175
176 frecuencias <- as.vector(tabla_clases)
177 media_continua <- sum(marca_clases * frecuencias) / sum(frecuencias)
178
179 # Cálculo de la Moda
180
181 ind_modal <- which.max(frecuencias)
182 Lim_inf <- cortes[ind_modal]
183 f_m <- frecuencias[ind_modal]
184 f_1 <- ifelse(ind_modal == 1, 0, frecuencias[ind_modal-1])
185 f_2 <- ifelse(ind_modal== length(frecuencias), 0,frecuencias[ind_modal+1] )
186 moda_continua <- Lim_inf + ((f_m - f_1) / ((f_m - f_1) + (f_m - f_2))) * amplitud
187
188 #Cálculo de Mediana
189
190 n_total <- sum(frecuencias)
191 n_2 <- n_total / 2 # Mitad del total
192 clase_mediana_index <- which(tabla_de_frecuencia$Frec_Acum >= n_2)[1] # Índice de la clase mediana
193
194 L <- cortes[clase_mediana_index] # Límite inferior de la clase mediana
195 F_anterior <- ifelse(clase_mediana_index == 1, 0, tabla_de_frecuencia$Frec_Acum[clase_mediana_index - 1])
196 f_mediana <- tabla_de_frecuencia$Frec_Abs[clase_mediana_index] # Frecuencia de la clase mediana
197
198 mediana_continua <- L + ((n_2 - F_anterior) / f_mediana) * amplitud
199
200 |
201 # Medidas de dispersión
202 varianza_continua <- sum(frecuencias * (tabla_de_frecuencia$marca_clase - media_continua)^2) / (n_total - 1)
203 desvio_continua <- sqrt(varianza_continua) # Desvío estándar continuo
204 coef_var_continua <- (desvio_continua / media_continua) * 100 # Coeficiente de variación
```



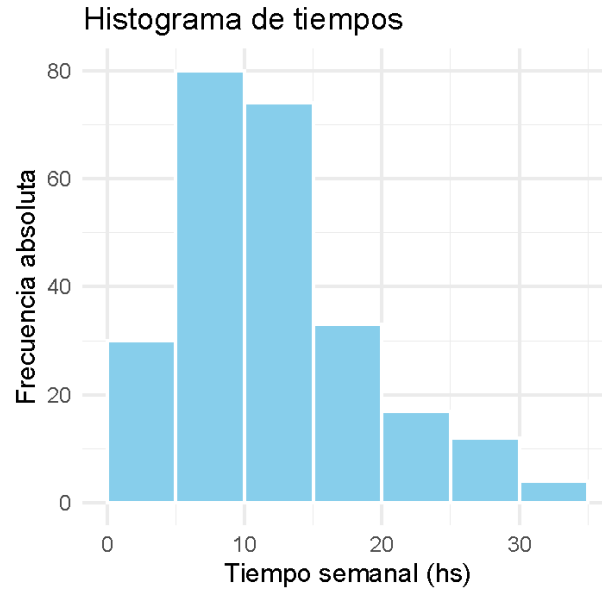
```
206 # Mostrar resultados agrupados
207
208 message("\n* Resultados AGRUPADOS - VARIABLE CONTINUA (", variable_continua, ")")
209 continua_stats <- data.frame(
210   Media = round(media_continua, 4),
211   Moda = round(modas_continua, 4),
212   Mediana = round(mediana_continua, 4)
213 )
214 print(continua_stats, row.names = FALSE)
215
216 # Cálculo de cuantiles
217
218
219 cuantiles <- quantile(datos[[variable_continua]], probs=c(0.25,0.5,0.75), na.rm=TRUE)
220 Tabla_Cuartil_Continua <- data.frame(
221   Cuartil = c("Q1 (25%)", "Q2 (Mediana)", "Q3 (75%)"),
222   Valor = as.numeric(cuantiles)
223 )
224
225 print(Tabla_Cuartil_Continua)
```

4. Representar gráficamente las variables definidas en el punto 2 y realizar el correspondiente análisis.
- Elegir una frecuencia (Frecuencia Absoluta o Frecuencia Relativa) para la variable Tiempo en horas semanales dedicadas al estudio y construir un Histograma.
 - Construir un Diagrama Circular que represente porcentualmente el Nivel de satisfacción con la carrera.
 - Realizar el análisis de los gráficos obtenidos en el contexto del caso planteado.

Histograma de tiempos semanales de estudio.

En el siguiente histograma podemos ver la distribución de horas de estudio semanal.

- La Forma de la distribución parece tener una concentración entre 10 y 20 horas, con una caída hacia los extremos. Esto sugiere una tendencia central clara, posiblemente entre 15 y 20 horas.
- En la Frecuencia absoluta su pico más alto indica que la mayoría de los casos se ubican en ese rango medio, lo que podría representar un patrón típico de dedicación semanal.



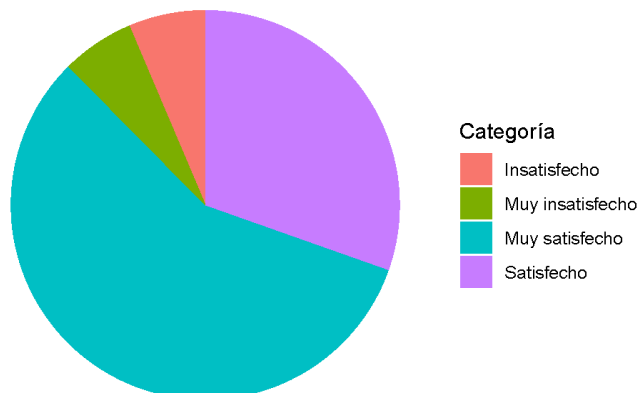
Nivel de satisfacción con la carrera.

Este gráfico muestra niveles de satisfacción:

- Categorías: “Muy insatisfecho”, “Insatisfecho”, “Satisfecho”, “Muy satisfecho”.
- Podemos ver que predominan las categorías positivas (“Satisfecho” y “Muy satisfecho”), podría inferir una buena alineación entre expectativas y realidad académica.

Este tipo de gráfico es útil para cruzar con variables como rendimiento académico, carga horaria o tipo de carrera. En este caso, podríamos vincularlo con rutinas optimizadas o incluso con simulaciones de entrevistas para evaluar motivación.

Nivel de Satisfacción con la Carrera



Link Repositorio de GitHub: <https://github.com/fabian24cf/Repositorio-Estad-stica.git>