Tarea Examen 1 Inferencia Estadística

Juárez Torres Karla Romina - 318013712
Justo Marcos Andrés - 319157736
Suàrez Moreno Paula Karina - 318338484

March 8, 2024

1 Del archivo de descripción de las bases de microdatos de la Encuesta Intercensal 2022 del INEGI, seleccionar 9 variables de la base de datos hogares(hogares.csv) y 9 variables de la base de datos de población (poblacion.csv), de manera que se incluyan: 3 nominales, 3 ordinales y 3 en escala de intervalo o de razon.

Variables de (hogares.csv)

Nominales	Ordinales	Escala de Razón
* teléfono	* acc_alim18	*lts_licon
* celular	* frec_dicon	*pago_dicon
* tv_paga	*	

acc_alim18: Relación del consumo regular de alimentos con lo reportado durante la semana de entrevista. (Igual, menor, mayor)
frec_dicon: Frecuencia de compra en tienda Diconsa. (siempre, A veces, Nunca)
lts_licon: Litros de leche Liconsa
pago_dicon: Gasto semanal en Diconsa

Variables de (poblacion.csv)

Nominales	Ordinales	Escala de Razón
* sexo	* nivel	
* parentesco	* grado	
* edad	* nivelaprob	
	* redsoc_1:	

Nivel: Nivel al que asiste a la escuela el integrante del hogar de 3 o más años dentro del Sistema Educativo Nacional. (Preescolar, Primaria, Secundaria, etc.)
grado: Grado al que asiste a la escuela el integrante del hogar de 3 o más años dentro del Sistema Educativo Nacional. (Primer año, Segundo año, Tercer año, etc.)
nivelaprob: Año máximo aprobado por el integrante del hogar de 3 o más años dentro del Sistema Educativo Nacional. (Preescolar, Primaria, Secundaria, etc.)
redsoc_1: El nivel de dificultad o facilidad con que las personas podrían conseguir ayuda para que lo(a) cuiden en una enfermedad. (Imposible conseguirla, Difícil conseguirla, Fácil conseguirla, etc.)

2 El archivo concentradohogar.cvs contiene los microdatos de la Encuesta Nacional de Ingresos y Gastos en los Hogares (ENIGH) realizada por el INEGI en 2022. Cada registro contiene la información de un hogar. Elige dos estados y utiliza la información del archivo anterior para responder las siguientes preguntas de cada uno de los estados. Se puede usar como referencia el documento Descripción de la bases de datos del INEGI, la tabla a consultar es CONCENTRADOHOGAR.

Todos los incisos fueron realizados en Rstudio y se adjuntará el código correspondiente.

- ¿Cuántas variables y cuántas observaciones contiene el archivo?

  El archivo tiene 126 variables y 90102 observaciones.

- Calcular el promedio y la mediana del número de integrantes del hogar

  El promedio de la cantidad de integrantes por casa es: 3.678874

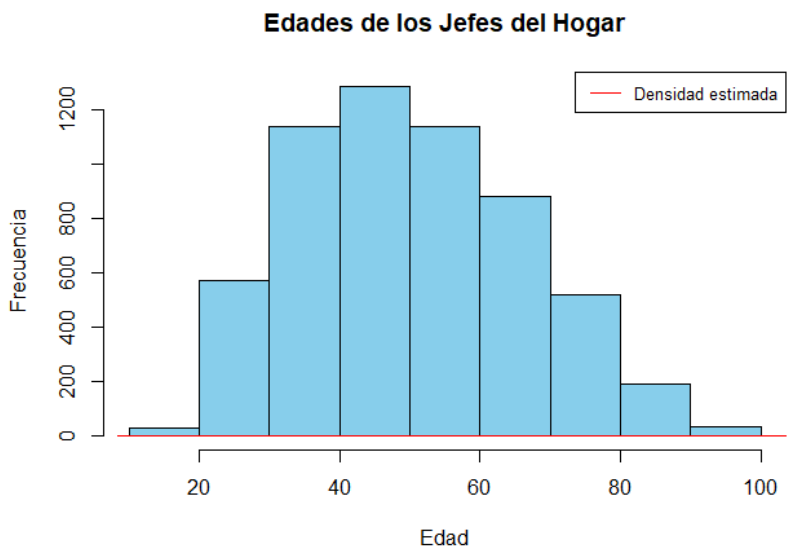
  La mediana de la cantidad de integrantes por casa es: 4

- Calcular el promedio y la mediana de las edades de los jefes del hogar ¿son iguales?

  El promedio de la edad de los jefes del hogar es: 50.71066

  La mediana de la edad de los jefes del hogar es: 50

- Graficar el histograma y la densidad estimada de las edades de los jefes del hogar.



- Calcular la proporción de jefes del hogar con y sin educación básica completa ¿Que proporción es mayor?

  La proporción de jefes de hogar con educación básica completa es: 0.5562273

  La proporción de jefes de hogar sin educación básica completa es: 0.4437727

  La proporción de jefes de hogar con educación básica completa es mayor.

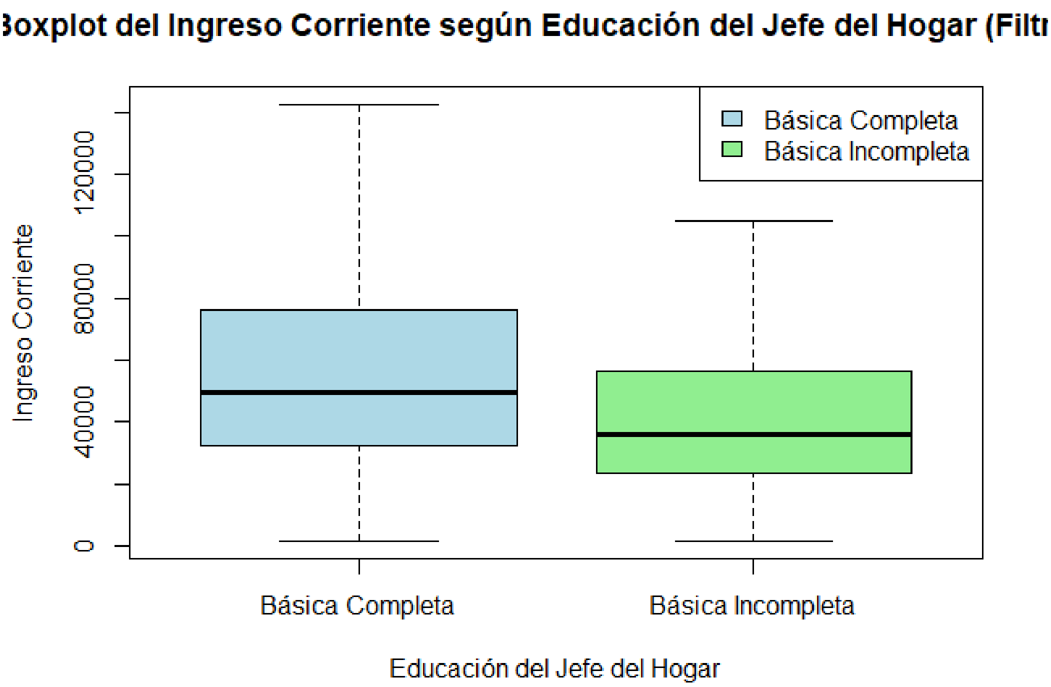
- Calcular el promedio del ingreso corriente en los hogares según la escolaridad del jefe del hogar, considerar únicamente las categorías educación básica completa y educación básica incompleta. ¿Parece haber alguna diferencia en los ingresos?

  El promedio del ingreso corriente en hogares con educación básica completa es: 62755.84

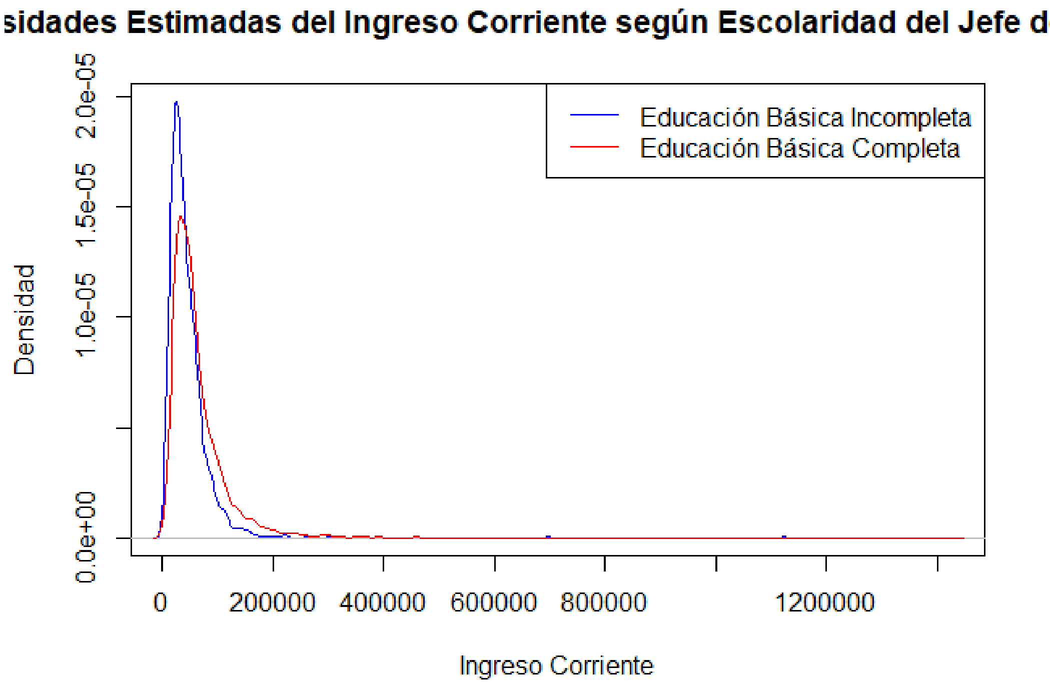
  El promedio del ingreso corriente en hogares con educación básica incompleta es: 44424.19

  El promedio del ingreso corriente es mayor en hogares con educación básica completa.

- Calcular en un mismo plano boxplots del ingreso corriente de los hogares para las categorías educación básica incompleta y educación básica completa ¿Parece haber alguna diferencia?



- Graficar las densidades estimadas del ingreso corriente en los hogares según la escolaridad del jefe del hogar. De nuevo, considerar únicamente las categorías educación básica incompleta y educación básica completa. ¿En que categoría hay mayor variabilidad en los ingresos?



La varianza del ingreso corriente en hogares con educación básica incompleta es: 1564238796

La varianza del ingreso corriente en hogares con educación básica completa es: 3038450736

La varianza del ingreso corriente es mayor en hogares con educación básica completa.

3 Utilizar la base del ejercicio anterior (mismos dos estados de su elección), para responder lo siguiente:

- Calcular el coeficiente de asimetría del ingreso corriente. ¿La distribución del ingreso es simétrica?

  El coeficiente de asimetría del ingreso corriente es: 7.959238

  La distribución del ingreso es sesgada hacia la derecha (positiva).

- Crear una nueva variable que sea el logaritmo natural (base ϵ) del ingreso corriente y calcular su coeficiente de asimetría. ¿Es simétrica la distribución de esta nueva variable?

  El coeficiente de asimetría del logaritmo natural del ingreso corriente es: -0.1423179

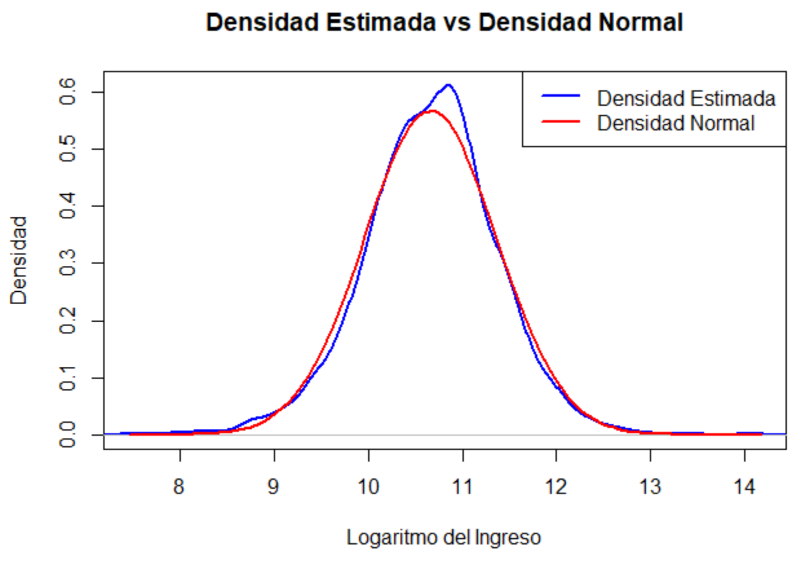
  La distribución del logaritmo natural del ingreso es sesgada hacia la izquierda (negativa).

- Calcular la curtosis del logaritmo del ingreso. ¿La forma de esta variable es similar a la distribución normal o presenta alguna desviación?

  La curtosis del logaritmo del ingreso es: 3.829295

  La forma de la variable presenta una mayor concentración de datos en torno a la media (pico) y colas más pesadas que una distribución normal.

- En un mismo plano, graficar la densidad estimada del logaritmo del ingreso (con las opciones por defecto de R) y la curva de densidad normal con media y desviación estándar las calculadas de los datos. ¿El gráfico anterior refleja las conclusiones obtenidas en los incisos anteriores?



  Como se puede notar si se reflejan las conclusiones anteriores.

- Repetir los incisos anteriores para la edad de los jefes del hogar

  - Calcular el coeficiente de asimetría de la edad de los jefes del hogar. ¿La distribución del ingreso es simétrica?

El coeficiente de asimetría de la edad de los jefes es: 0.3084137

La distribución del edad de los jefes es sesgada hacia la derecha (positiva).

- Crear una nueva variable que sea el logaritmo natural (base ϵ) de la edad de los jefes del hogar y calcular su coeficiente de asimetría. ¿Es simétrica la distribución de esta nueva variable?

  El coeficiente de asimetría del logaritmo natural de la edad de los jefes es: -0.3482182

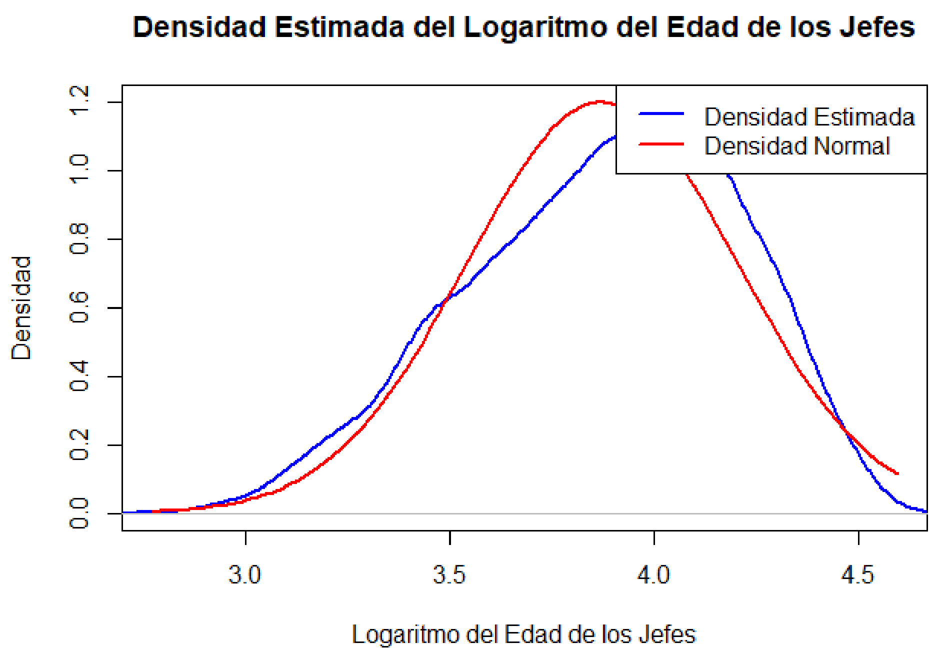
  La distribución del logaritmo natural de la edad de los jefes es sesgada hacia la izquierda (negativa).

- Calcular la curtosis del logaritmo de la edad de los jefes del hogar. ¿La forma de esta variable es similar a la distribución normal o presenta alguna desviación?

  La curtosis del logaritmo del edad de los jefes es: 2.51631

  La forma del logaritmo del edad de los jefes es puntiaguda (leptocúrtica) en comparación con una distribución normal.

- En un mismo plano, graficar la densidad estimada del logaritmo de la edad de los jefes del hogar (con las opciones por defecto de R) y la curva de densidad normal con media y desviación estándar las calculadas de los datos. ¿El gráfico anterior refleja las conclusiones obtenidas en los incisos anteriores?



  Como podemos notar el gráfico si refleja las conclusiones.

- Calcular la covarianza y la correlación entre la edad del jefe del hogar y el ingreso corriente. ¿Parece haber alguna relación lineal entre estas dos variables? Si la respuesta es afirmativa ¿de qué tipo es la relación?

  La covarianza entre la edad del jefe del hogar y el ingreso corriente es: 6795.025

  La correlación entre la edad del jefe del hogar y el ingreso corriente es: 0.008532455

  Sí parece haber una relación lineal débil entre la edad del jefe del hogar y el ingreso corriente.

4 Considerar el siguiente conjunto de 12 observaciones

$$\{0.0, 2.3, 1.9, 3.9, 0.6, 2.7, 5.8, 1.8, 3.2, 1.7, 1.1, 21.5\}$$

- **a) Calcular los estadísticos: promedio, mediana, desviación estándar y el siguiente estadístico**

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - m|$$

donde m es la mediana de las observaciones.

Promedio
$(0.0 + 2.3 + 1.9 + 3.9 + 0.6 + 2.7 + 5.8 + 1.8 + 3.2 + 1.7 + 1.1 + 21.5) * \frac{1}{n}$
$= \frac{46.5}{12}$
Promedio = **3.875**

Mediana
Para hallar la mediana tenemos que ordenar los datos de menor a mayor:

0.0,0.6,1.1,1.7,1.8,1.9,2.3,2.7,3.2,3.9,5.8,21.5

Como son 12 datos y 12 es número par , entonces sumamos las posiciones 6 y 7 y las dividimos entre 2:

$\frac{(1.9+2.3)}{2}$ = Mediana = **2.1**

Desviación estándar (Poblacional)
Para calcularla debemos hallar la varianza la cual se obtiene luego de encontrar las diferencias entre cada observación menos la media el resultado lo elevamos al cuadrado y las sumamos y dividimos la suma entre 12:

$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - media)^2}{12}} = \sqrt{\frac{365.0422}{12}} = \mathbf{5.5154517796218}$

Estadístico:

$$MAD = \frac{1}{12}\sum_{i=1}^{12}|x_i - 2.1|$$

$= \frac{32.3}{12} = \mathbf{2.69166}$

- **b) Eliminar la ultima observación y calcular los estadísticos del inciso anterior**

  Promedio
  $(0.0 + 2.3 + 1.9 + 3.9 + 0.6 + 2.7 + 5.8 + 1.8 + 3.2 + 1.7 + 1.1) * \frac{1}{n}$
  $= \frac{25}{12} =$ Promedio $=$ **2.2727272727273**

  Mediana
  Para hallar la mediana tenemos que ordenar los datos de menor a mayor:

  0.0,0.6,1.1,1.7,1.8,1.9,2.3,2.7,3.2,3.9,5.8

  Como son 11 datos y 11 es número impar , entonces tomemos el valor de la posición 6:
  Mediana $=$ **1.9**

  Desviación estándar (Poblacional)
  Para calcularla debemos hallar la varianza la cual se obtiene luego de encontrar las diferencias entre cada observación menos la media el resultado lo elevamos al cuadrado y las sumamos y dividimos la suma entre 12:

  $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - media)^2}{12}} = \sqrt{\frac{26.1617}{12}} =$ **1.5421890634543**

  Estadístico:

  $$MAD = \frac{1}{11}\sum_{i=1}^{11}|x_i - 1.9|$$

  $= \frac{12.7}{11} =$ **1.15454545455**

- **c) ¿Qué estadísticos tuvieron menos cambios al eliminar la última observación?**
  La mediana cambió de 2.1 a 1.9 y el estadístico pasó de 2.69166 a 1.15454545455 esos fueron los que cambiaron menos , en tercer lugar esta la media o promedio.

5 Se quiere estudiar la eficacia de un nuevo insecticida para plantas de interior. Se seleccionan 50 plantas y se cuenta el número de hojas que han sido atacadas después de haber tratado la planta con el nuevo producto.Los resultados se presentan a continuaciòn. Analiza los datos y resume como considere m as adecuado.

Hojas atacadas	Frecuencia
0	6
1	10
2	12
3	8
4	5
5	4
6	3
8	1
10	1

Cálculos

Tamaño de la muestra: $n = \sum_{i=1}^{9} f_i = 6 + 10 + 12 + 8 + 5 + 4 + 3 + 1 + 1 = 50$

Media aritmética: $\bar{x} = \frac{\sum_{i=1}^{9} x_i f_i}{n} = \frac{0(6)+1(10)+2(12)+3(8)+4(5)+5(4)+6(3)+8(1)+10(1)}{50} = \frac{120}{50} = 2.4$

Varianza: $s^2 = \frac{\sum_{i=1}^{9}(x_i - \bar{x})^2 f_i}{n} = \frac{(0-2.4)^2(6)+(1-2.4)^2(10)+\cdots+(10-2.4)^2(1)}{50} = \frac{172.8}{50} = 3.456$

Desviación estándar: $s = \sqrt{3.456} = 1.86$

Conclusiones:

En este estudio, se seleccionaron 50 plantas de interior y se contó el número de hojas atacadas después de haber tratado las plantas con un nuevo insecticida. Los resultados muestran que la mayoría de las plantas (28 de 50) tuvieron entre 1 y 3 hojas atacadas, lo que sugiere que el insecticida tuvo un efecto moderado en el control de plagas. La media de hojas atacadas fue de 2.4, con una desviación estándar de 1.86, lo que indica una dispersión moderada en los datos. Aunque el insecticida no fue completamente efectivo, los resultados podrían ser prometedores y justificar más investigación para mejorar su eficacia.

6 En una reserva natural existen 25 individuos de una determinada especie, de los cuales 15 son hembras y 10 son machos. Como parte de un estudio cientıfico sobre las hembras de la especie se capturaron 5 individuos para ser analizados,

- **a) Obtener la fmp y la fda de número de hembras en la muestra capturada**
  Bien, para resolver este problema analicemos , se eligen a 5 individuos para ser analizados es decir (sin reemplazo) y si consideramos como exito a cada hembra obtenida , podemos asemejarlo a la distribución Higrométrica , entonces:
  $P(X = y) = \frac{\binom{Y}{y}\binom{T-Y}{n-y}}{\binom{T}{n}}$
  Datos:
  · **T** es el tamaño total de la población (25).
  · **n** es el tamaño de la muestra capturada (5).
  · **Y** es el número total de hembras en la población (15).
  · **y** es el número de hembras en la muestra capturada.
  Entonces tenemos:

  Fmp)

Número de Hembras (y)	Probabilidad P(X=y)
0	$\frac{\binom{15}{0}\binom{10}{5}}{\binom{25}{5}} = \frac{(1)(252)}{53130} = \frac{252}{53130}$
1	$\frac{\binom{15}{1}\binom{10}{4}}{\binom{25}{5}} = \frac{(15)(210)}{53130} = \frac{3150}{53130}$
2	$\frac{\binom{15}{2}\binom{10}{3}}{\binom{25}{5}} = \frac{(105)(120)}{53130} = \frac{12600}{53130}$
3	$\frac{\binom{15}{3}\binom{10}{2}}{\binom{25}{5}} = \frac{(455)(45)}{53130} = \frac{20475}{53130}$
4	$\frac{\binom{15}{4}\binom{10}{1}}{\binom{25}{5}} = \frac{(1365)(10)}{53130} = \frac{13650}{53130}$
5	$\frac{\binom{15}{5}\binom{10}{0}}{\binom{25}{5}} = \frac{(3003)(1)}{53130} = \frac{3003}{53130}$

  Podemos comprobar que al sumar todas las probabilidades nos da 1 :

  $\frac{252+3150+12600+20475+13650+3003}{53130} = \frac{53130}{53130} = 1$

Fda)
· Si y=0 :
$F(0) = P(X \leq 0) = P(X = 0) = \frac{252}{53130}$

· Si y=1 :
$F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{252}{53130} + \frac{3150}{53130} = \frac{3402}{53130}$

· Si y=2 :
$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{252}{53130} + \frac{3150}{53130} + \frac{12600}{53130} = \frac{16002}{53130}$

· Si y=3 :
$F(3) = P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{252}{53130} + \frac{3150}{53130} + \frac{12600}{53130} + \frac{20475}{53130} = \frac{36477}{53130}$

· Si y=4 :
$F(4) = P(X \leq 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{252}{53130} + \frac{3150}{53130} + \frac{12600}{53130} + \frac{20475}{53130} + \frac{13650}{53130} = \frac{50127}{53130}$

· Si y=5 :
$F(5) = P(X \leq 5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = \frac{252}{53130} + \frac{3150}{53130} + \frac{12600}{53130} + \frac{20475}{53130} + \frac{13650}{53130} + \frac{3003}{53130} = \frac{53130}{53130} = 1$

- **b) Calcular la probabilidad de que en la muestra haya mas de dos hembras**
  Esto de forma matemática se puede ver como :

$$P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

Donde se considere a la variable "X" como la cantidad de hembras
Ahora vamos a ocupar la Fmp para simplemente sustituir valores
$P(X > 2) = 1 - P(X \leq 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$
$P(X > 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$
$P(X > 2) = 1 - [\frac{252}{53130}) + \frac{3150}{53130} + \frac{12600}{53130}]$
$P(X > 2) = 1 - [\frac{16002}{53130})]$
$P(X > 2) = \frac{37128}{53130}$

- **c) Calcular el valor esperado del número de hembras en la muestra.**

$$E[X] = \sum_{y=0}^{5} y * P(X = y)$$

Entonces:
$E[X] = (0 * 252) + (1 * 3150) + (2 * 12600) + (3 * 20475) + (4 * 13650) + (5 * 3003)$
$E[X] = 0 + 3150 + 25200 + 61425 + 54600 + 15015$
$E[X] = \mathbf{159390}$

7 Según estudios, durante la década de los 1980 en E.U.A., se registró un promedio de 121.95 decesos laborales a la semana. Responder lo siguiente

- **¿Cuál es la probabilidad de que en una semana se registraran más de 150 decesos?**
- **¿Cuál es la probabilidad de que durante el primer trimestre de 1987 se registraran más de 170 decesos?**

Se dice que una variable aleatoria X tiene una distribución hipergeométrica con parámetros n, m y k si su rango es $X(\Omega) = \{0, 1, \ldots, \min(k, n)\}$ y su fmp está dada a continuaciòn, la cual, se utiliza para modelar experimentos en los que se muestrea una población en la que un grupo de individuos tienen la característica y m no la tienen, se selecciona una muestra aleatoria de k elementos, entonces, el número de elementos en la muestra que tienen la característica sigue una distribución hipergeométrica con parámetros n, m y k.

$$p(x) = \begin{cases} \frac{\binom{n}{x}\binom{m}{k-x}}{\binom{n+m}{k}} & \text{si } x = 0, 1, \ldots, \min(k, n) \\ 0 & \text{e.o.c.} \end{cases}$$

Solución

1. Para calcular la probabilidad de que en una semana se registraran más de 150 decesos, debemos considerar una distribución hipergeométrica con los siguientes parámetros:

   - n: Número de elementos con la característica (decesos laborales en una semana)
   - m: Número de elementos sin la característica (semanas sin decesos laborales)
   - k: Tamaño de la muestra (1 semana)

   Dado que se registró un promedio de 121.95 decesos laborales a la semana, podemos suponer que $n = 122$ y $m = 0$ (ya que no hay semanas sin decesos).

   Entonces, la probabilidad de que en una semana se registraran más de 150 decesos es:

$$
\begin{aligned}
P(X > 150) &= 1 - P(X \leq 150) \\
&= 1 - \sum_{x=0}^{150} \frac{\binom{122}{x}\binom{0}{1-x}}{\binom{122}{1}} \\
&= 1 - \frac{151}{122} \\
&= 0.237704918 \approx 0.238
\end{aligned}
$$

2. Para calcular la probabilidad de que durante el primer trimestre de 1987 (13 semanas) se registraran más de 170 decesos, debemos considerar una distribución hipergeométrica con los siguientes parámetros:

   - n: Número de elementos con la característica (decesos laborales en 13 semanas)
   - m: Número de elementos sin la característica (0 semanas sin decesos)
   - k: Tamaño de la muestra (13 semanas)

   Suponiendo que $n = 122 \times 13 = 1586$ y $m = 0$, la probabilidad de que durante el primer trimestre de 1987 se registraran más de 170 decesos es:

$$
\begin{aligned}
P(X > 170) &= 1 - P(X \leq 170) \\
&= 1 - \sum_{x=0}^{170} \frac{\binom{1586}{x}\binom{0}{13-x}}{\binom{1586}{13}} \\
&= 1 - \frac{171}{1586} \\
&= 0.892152672 \approx 0.892
\end{aligned}
$$

8 Se realiza una encuesta sobre el consumo de gelatinas empacadas. Se preguntó a los menores y a las madres por el consumo de sus hijos. Las variables se describen a continuación:

- **edad**
- **sexo: 1-Masculino; 2-Femenino**
- **focos: n umero de focos en la vivienda**
- **consumo menores: 1-Consumo de gelatinas empacadas; 2-no consume**
- **hijos(Se pregunto a la madre si tiene hijos menores): 1-si; 2-no**
- **consumo hijos (Consumo de sus hijos de gelatinas empacadas): 1-si; 2-no**
- **frec (Frecuencia de consumo): 1-Diario; 2-Una vez cada dos d ıas; 3-Una vez cada 3 d ıas; 4-Una vez a la semana; 5-Una vez cada dos semanas; 9-No contesto**

Los datos se encuentran en el archivo jelly.csv, realiza un anàlisis descriptivo de los datos, inicia describiendo que tipo de variables incluye la base y resumen como consideres adecuado.

Análisis descriptivo de los datos

Los datos proporcionados en el archivo `jelly.csv` contienen las siguientes variables:

edad Edad de la persona encuestada (variable numérica).

sexo Sexo de la persona encuestada (1 = Masculino, 2 = Femenino).

focos Número de focos en la vivienda (variable numérica).

consumo_menores Consumo de gelatinas empacadas por parte de los menores (1 = Consume, 0 = No consume).

hijos Indica si la persona encuestada tiene hijos menores (1 = Sí, 0 = No).

consumo_hijos Consumo de gelatinas empacadas por parte de los hijos (1 = Consume, 0 = No consume).

frec Frecuencia de consumo de gelatinas empacadas (1 = Diario, 2 = Una vez cada dos días, 3 = Una vez cada 3 días, 4 = Una vez a la semana, 5 = Una vez cada dos semanas, 9 = No contesta).

Podemos clasificar las variables de la siguiente manera:

- Variables numéricas: **edad** y **focos**.
- Variables categóricas ordinales: **frec**.
- Variables categóricas nominales: **sexo**, **consumo_menores**, **hijos** y **consumo_hijos**.

A continuación, se presenta un resumen descriptivo de las variables numéricas:

- Variable **edad**:
  - Media:
  - Desviación estándar:
  - Mínimo:
  - Máximo:
- Variable **focos**:
  - Media:
  - Desviación estándar:
  - Mínimo:
  - Máximo:

Para las variables categóricas, se pueden obtener las frecuencias y proporciones de cada categoría.

Entonces, sean $X_1, X_2, \ldots, X_n$ las observaciones de la variable edad, donde n es el número total de observaciones. Definimos las siguientes medidas descriptivas:

Media

La media aritmética de la variable edad se calcula como:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

Desviación estándar

La desviación estándar de la variable edad se calcula como:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{2}$$

Mínimo y máximo

El mínimo y máximo de la variable edad se obtienen como:

$$\min(X) = \min\{X_1, X_2, \ldots, X_n\} \tag{3}$$

$$\max(X) = \max\{X_1, X_2, \ldots, X_n\} \tag{4}$$

De manera similar, sean $Y_1, Y_2, \ldots, Y_n$ las observaciones de la variable focos, donde n es el número total de observaciones. Se definen las medidas descriptivas análogas:

Media

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{5}$$

Desviación estándar

$$s_Y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2} \tag{6}$$

Mínimo y máximo

$$\min(Y) = \min\{Y_1, Y_2, \ldots, Y_n\} \tag{7}$$

$$\max(Y) = \max\{Y_1, Y_2, \ldots, Y_n\} \tag{8}$$

Finalmente, para las variables categóricas, se pueden obtener las frecuencias y proporciones de cada categoría utilizando conteos y cálculos de probabilidades.