

# PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

## FACULTAD DE CIENCIAS E INGENIERÍA



### PC1

#### Curso:

ESTADÍSTICA PARA INGENIERÍA

#### Horario:

501

#### Elaborado por:

(20185391)	Rojas Rosales, Luiggi	luiggi.rojas@pucp.edu.pe
(20200315)	Camayo Arias, Jose Félix	20200315@pucp.edu.pe
(20195861)	Iman Loja, Luis Angel Luis	luis.imanl@pucp.edu.pe

#### Docente:

TARAZONA VARGAS , ENVER GERALD

Lima, 17 de enero del 2023

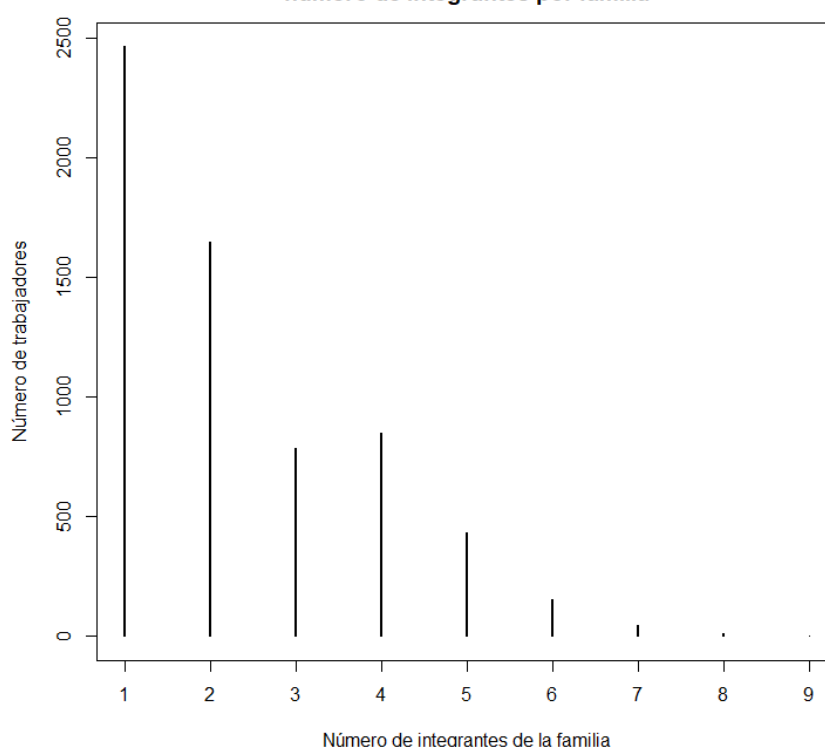
## PREGUNTA 1:

Tabla N° 1: Distribución del número de familias de trabajadores según número de integrantes.

Número de integrantes	Número de familias	Proporción de familias	Porcentaje de familias
1	2466	0.38531	38.53
2	1651	0.25797	25.8
3	786	0.12281	12.28
4	851	0.13297	13.3
5	436	0.06813	6.81
6	154	0.02406	2.41
7	45	0.00703	0.7
8	10	0.00156	0.16
9	1	0.00016	0.02

De la Tabla N° 1 se puede destacar que hay un mayor número de trabajadores que tienen familias con un integrante, los cuales forman parte del 38.53% del total de trabajadores. Por otro lado, existe solo un trabajador con nueve integrantes en su familia. Este último, a su vez, es el mayor número de integrantes por familia de los trabajadores.

**Gráfica N°1: Cantidad de trabajadores según el número de integrantes por familia**



A partir del Gráfico N°1 se puede resaltar que hay una menor cantidad de trabajadores que provienen de familias numerosas, a diferencia que hay mayor número de personas que provienen de familias de menos integrantes. Por ende, hay una tendencia a que los trabajadores tengan

menos integrantes en su familia. Sin embargo, esta tendencia se interrumpe para familias con tres integrantes.

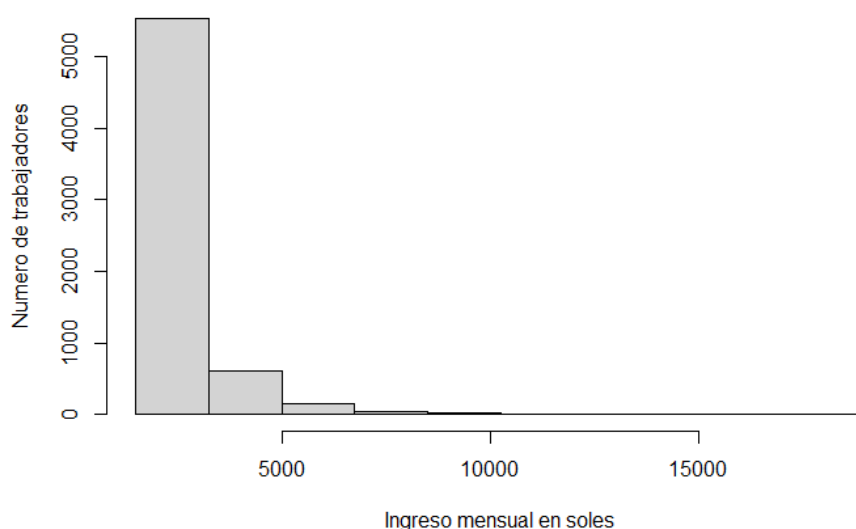
## PREGUNTA 2:

Tabla N° 2: Distribución del ingreso mensual de los trabajadores.

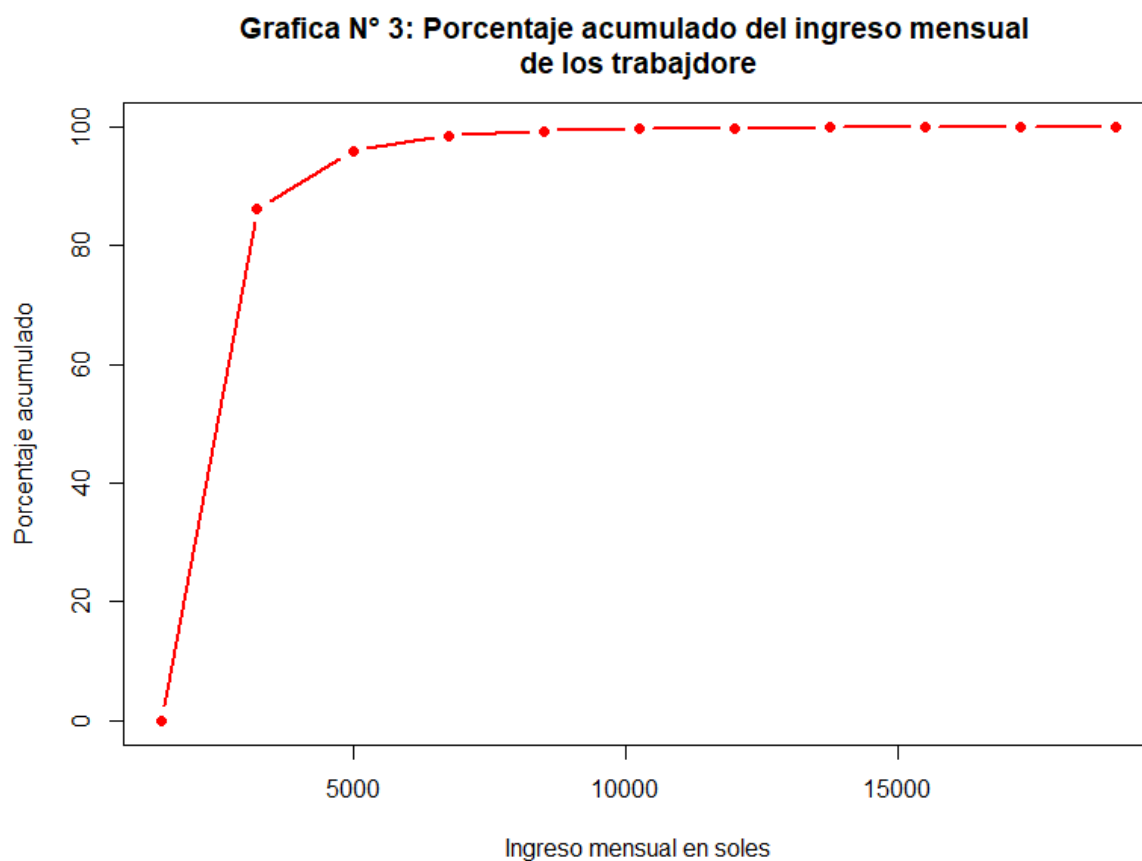
INGRESO MENSUAL EN SOLES		Marca de clase	Número de trabajadores	Porcentaje (%)	Número de trabajadores acumulado	Porcentaje acumulado (%)
Desde	Hasta					
1500	3250	2375	5524	86.3	5524	86.3
3250	5000	4125	616	9.6	6140	95.9
5000	6750	5875	164	2.6	6304	98.5
6750	8500	7625	50	0.8	6354	99.3
8500	10250	9375	27	0.4	6381	99.7
10250	12000	11125	7	0.1	6388	99.8
12000	13750	12875	3	0	6391	99.9
13750	15500	14625	4	0.1	6395	99.9
15500	17250	16375	1	0	6396	99.9
17250	19000	18125	4	0.1	6400	100

De acuerdo a los resultados de la Tabla N° 2, se puede distinguir que hay un mayor número de trabajadores con ingresos mensuales de más de 1500 hasta 3250 soles, el cual es el menor intervalo de ingresos. Por otro lado, solo un trabajador percibe ingresos entre 15 500 hasta 17 250 soles. No obstante, este último no es el mayor intervalo ya que hay cuatro trabajadores que ganan entre 17 250 a 19 000 soles.

**Grafica N° 2: Histograma del ingreso mensual de trabajadores**



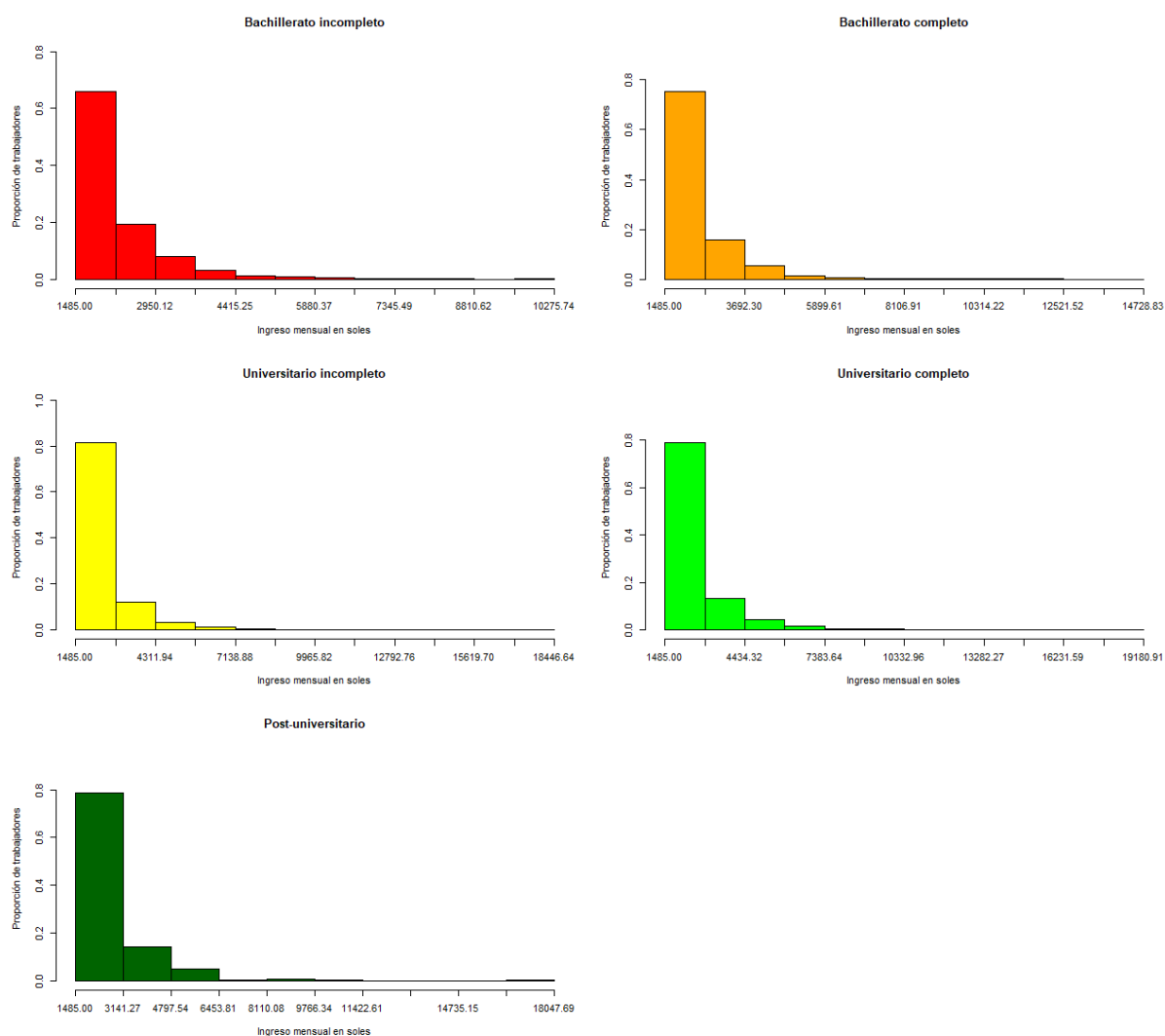
En la Grafica N° 2 se puede visualizar que hay una gran cantidad de trabajadores que perciben ingresos mensuales menores; mientras que trabajadores con mayores ingresos son inusuales.



Por último, concatenando la información de la Grafica N° 3 y la Tabla N° 2, se puede afirmar que es notoria la gran proporción de trabajadores cuyo ingreso mensual pertenece al primer intervalo de salarios mensuales conformando el 86.3% de trabajadores que ganan hasta 3250.

### PREGUNTA 3:

Gráfica N° 4: Ingreso mensual de los trabajadores según nivel educativo



De la Gráfica N° 4 se puede visualizar que hay una mayor proporción de trabajadores que perciben sueldos menores para cualquier nivel de educación. Donde a su vez, son poco usuales los trabajadores que perciben ingresos más altos de la muestra.

Tabla N° 3: Medidas de tendencia central y dispersión del ingreso mensual según nivel educativo

ESTADÍSTICOS	Bachillerato incompleto	Bachillerato completo	Universitario incompleto	Universitario completo	Post- universitario
Número de trabajadores	1390	1936	1360	1355	359
Mean	2304	2404	2466	2601	2735
Median	1974	2022	2053	2132	2290
Asimetría	3.396	4.064	4.395	4.807	4.394
Rango	8670	13080	16760	17490	16370
1st Cuartil	1753	1785	1800	1848	1911
3rd Cuartil	2448	2575	2579	2796	2962
Rango inter cuartil	695	790	779	948	1051

De los resultados de la Tabla N° 3, se puede observar que tanto la media y la mediana de ingresos aumenta conforme a mayor nivel de educación, por lo tanto, el promedio de ingresos de los trabajadores aumenta positivamente debido al nivel de educación. Los cuartiles aumentan, los cuales significan que hay una mejor posición salarial cuando se tiene un mayor nivel de educación. De la misma manera, tanto el rango como el rango inter cuartil incrementa de acuerdo al nivel de educación. El cuál significa que hay una mayor diferencia entre salarios mensuales de los trabajadores que pertenecen a niveles de educación mayores.

En resumen, existe cierta tendencia a mejores salarios cuanto mayor nivel de educación tiene el trabajador. Sin embargo, a su vez existe una mayor heterogeneidad salarial lo que supone que hay una mayor diferencia en los ingresos entre los trabajadores con mayor nivel de educación.

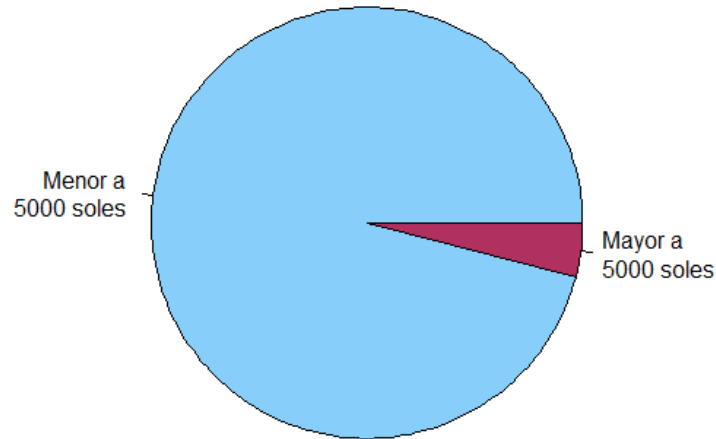
#### PREGUNTA 4:

Tabla N° 4: Proporción de trabajadores con ingreso mensual igual o mayor a 5000

Ingresos mensuales	PROPORCIÓN CALCULADO DIRECTAMENTE (%)	PROPORCION TABLA DE FRECUENCIAS (%)
Menos de 5000 soles	95.94	95.9
Más de 5000 soles	4.06	4.1

Las proporciones calculadas de forma directa y con la tabla de frecuencias coinciden. Esto se debe a la composición de Tabla N° 2, al dividir los intervalos coincide con uno de los límites superiores, por ende, se puede usar el porcentaje acumulado de forma directa para obtener la proporción.

**Grafica N°5: Proporción de trabajadores de ingresos menor o mayor a 5000 soles**



**PREGUNTA 5:**

Promedio de los ingresos mensuales atípicos superiores = 5560 soles

Se puede apreciar la gran diferencia entre la media de los datos atípicos y las medias de los trabajadores según el nivel de educación. Esto refleja la distancia del centro de gravedad de las observaciones de la muestra con respecto al centro de gravedad de las observaciones atípicas, las cuales pudieron incrementar el valor de la media. Por ende, sería más conveniente utilizar a la mediana como medida de tendencia central, ya que esta es más robusta ante la presencia de valores atípicos. Siguiendo esa analogía, se debería utilizar al rango inter cuartil como medida de dispersión puesto a que el rango se ve fuertemente afectado por los valores extremos de las observaciones.

## PREGUNTA 6:

Gráfico N° 6: Distribución de número de trabajadores según su edad y género

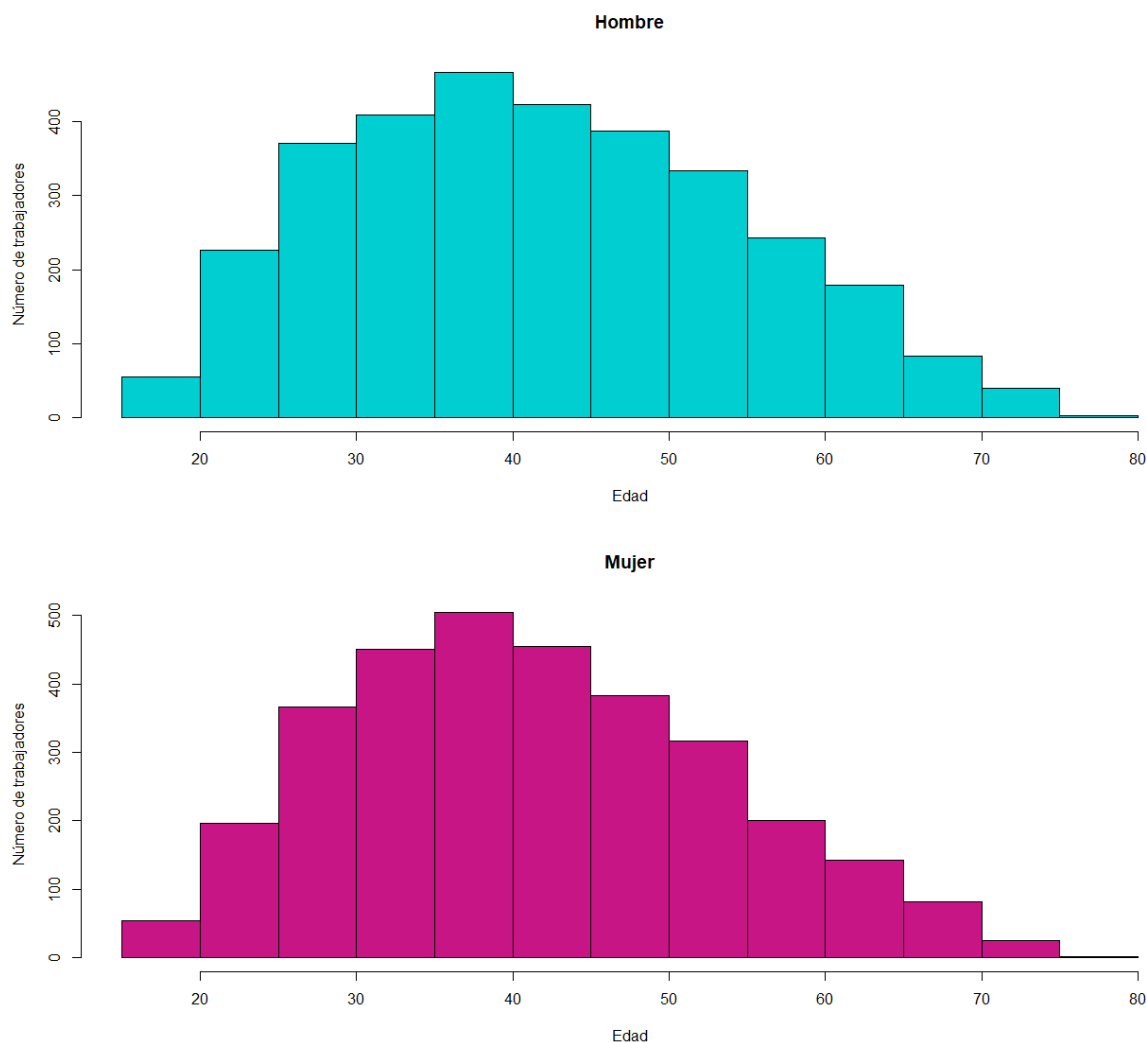


Tabla N° 5: Medidas de tendencia central y dispersión.

ESTADÍSTICOS	HOMBRE	MUJER
Media	42.37	41.74
Mediana	41.00	41.00
Asimetría	0.27	0.33
Mínimo	18.00	18.00
Máximo	77.00	76.00
Rango	59.00	58.00
1er Cuartil	32.00	33.00
3er Cuartil	51.00	50.00
Rango Inter Cuartil	19.00	17.00

En primer lugar, los cuartiles son numéricamente similares los cuales significan que la distribución de edades de trabajadores hombres y mujeres son parecidos, las cuales se pueden visualizar en la Grafica N° 6. En segundo lugar, las medianas para ambos grupos tienen el



mismo valor; mientras que la media difiere ligeramente. Por lo tanto, el promedio de edades es casi el mismo para ambos sexos. Por último, el rango y el rango inter cuartil tienen valores cercanos lo cual significa que hay una heterogeneidad similar en las edades tanto de trabajadores hombres como mujeres.

## PREGUNTA 7:

A)

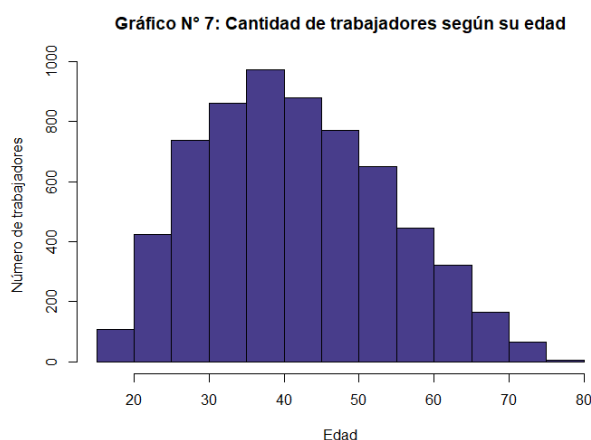


Tabla N° 6: Medidas de tendencia central y dispersión de la variable edad.

ESTADÍSTICAS	EDAD
Media	42.06
Mediana	41
Asimetría	0.2991
Mínimo	18
Máximo	77
Rango	59
Rango Inter cuartil	18

No es adecuado usar la media para representar los datos. Para empezar, hay presencia de asimetría positiva en las edades, los cuales significan que los datos extremos se ubican en mayores valores de edad, las cuales afectan a la media. Por otro lado, no sería acertado usar una medida de tendencia central para representar los datos de la edad, ya que hay una gran heterogeneidad. Por ende, sería más útil utilizar medidas de dispersión o de posición para representar mejor el gran número de valores de variable edad.

B)

No se puede afirmar estrictamente que la edad es la misma para todos los niveles de sexo puesto a que las observaciones y su frecuencia difieren. Sin embargo, podemos afirmar que la distribución de observaciones de las edades es similar para ambos sexos puesto a que, como se mencionó anteriormente, las medidas estadísticas tanto de tendencia central, posición y variabilidad se asemejan para ambos grupos de observaciones.

C)

Tabla N° 6: Porcentaje acumulado del nivel de satisfacción laboral de los trabajadores.

NIVEL DE SATISFACCIÓN	PORCENTAJE ACUMULADO DE TRABAJADORES
Muy insatisfecho	17.3
Insatisfecho	37.1
Neutro	58.9
Satisfecho	80.9
Muy satisfecho	100.0

No, el nivel de satisfacción “Satisfecho” no es la mediana de la observación, como se puede ver en la Tabla N° 6, hasta la categoría “Insatisfecho” se encuentra el 37.1% de los trabajadores y hasta el nivel “Neutro” se encuentra el 58.9% de los trabajadores. En tal sentido, el 50% de los trabajadores se encuentran hasta la categoría “Neutro”, por ende, la mediana sería esta.

ANEXO:

CODIGO R UTILIZADO PARA LA GENERACIÓN DE RESULTADOS

## Script

2023-01-18

```
d <- read.csv("Encuesta.csv",
              stringsAsFactors = TRUE)
d$educacion <- with(d, factor(educacion, levels=c('Bachillerato incompleto',
                                                'Bachillerato completo',
                                                'Universitario incompleto',
                                                'Universitario completo',
                                                'Post-universitario'), ordered=TRUE))
d$satlab <- with(d, factor(satlab, levels=c('Muy insatisfecho', 'Insatisfecho', 'Neutro',
                                            'Satisfecho', 'Muy satisfecho'), ordered=TRUE))
library(agricolae)
library(psych)
library(gmodels)
library(EnvStats)
library(fdth)
library(ggplot2)

# PREGUNTA 1

## ELABORACION DE LA TABLA

familia<-d$familia
## COLUMNAS PARA ELABORAR LA TABLA DE FRECUENCIAS
numero<-c(min(familia):max(familia))
f<-table(familia)
fr<-prop.table(table(familia))
P<-fr*100
## TABLA DE FRECUENCIAS
tabla1<-cbind(as.character(numero),as.character(f),round(fr,5),round(P,2))
colnames(tabla1)<-c("Numero de integrantes", "Numero de familias",
                  "Proporcion de familias","Porcentaje de familias")
tabla1<-as.table(tabla1)
write.table(x = tabla1, file = "TABLA1.txt", sep = ",",
           row.names = FALSE, col.names = TRUE) ##PARA EXPORTAR LA TABLA

## GRAFICO DE VARAS

plot(table(familia),xlab="Número de integrantes de la familia",
     ylab="Número de trabajadores",
```

```
main="Gráfica N°1: Cantidad de trabajadores según el \nnúmero de integrantes por familia")
```

## #PREGUNTA 2

### ##HISTOGRAMA

```
ingrmens<-d$ingrmens
##Determinacion de intervalos de misma amplitud a partir del numero de categorias
##propuesto
range<-max(ingrmens)-min(ingrmens)
Amplitud<-ceiling(range/10)
intervalos<-seq(min(ingrmens),min(ingrmens)+10*Amplitud,Amplitud)
##Histograma de frecuencias
histograma<-hist(ingrmens,breaks = intervalos,xlab = "Ingreso mensual en soles",
                  ylab="Numero de trabajadores",
                  main="Grafica N° 2: Histograma del ingreso mensual \nde trabajadores",)
```

### #TABLA DE FRECUENCIAS A PARTIR DEL HISTOGRAMA (AGRICOLAE)

```
tabla2<-table.freq(histograma)
colnames(tabla2)<-c("Desde","Hasta","Marca de clase","Numero de trabajadores",
                  "Porcentaje","Numero de trabajadores acumulado",
                  "Porcentaje acumulado")
write.table(x = tabla2, file = "TABLA2.txt", sep = ",",
            row.names = FALSE, col.names = TRUE) ##PARA EXPORTAR LA TABLA
```

### #GRAFICO DE FRECUENCIAS ACUMULADAS

```
ls<-c(min(tabla2$Desde),tabla2$Hasta)
acum=c(0,tabla2$`Porcentaje acumulado`)
plot(ls,acum,type="b",col="red",lwd=2,pch=16,xlab="Ingreso mensual en soles",
      ylab="Porcentaje acumulado",
      main="Grafica N° 3: Porcentaje acumulado del ingreso mensual \nde los trabajadores")
```

## #PREGUNTA 3

### #GRAFICOS DE FRECUENCIAS RELATIVAS

```
par(mfrow=c(3,2),main="")
```

```
## Warning in par(mfrow = c(3, 2), main = ""): "main" is not a graphical parameter

plot(fdt(d$ingrmens[d$educacion=="Bachillerato incompleto"]),type='rfh',
      xlab = "Ingreso mensual en soles",ylab="Proporción de trabajadores"
,
      main="Bachillerato incompleto",
      col="red")
plot(fdt(d$ingrmens[d$educacion=="Bachillerato completo"]),type='rfh',
      xlab = "Ingreso mensual en soles",ylab="Proporción de trabajadores"
,
      main="Bachillerato completo",
      col="orange")
plot(fdt(d$ingrmens[d$educacion=="Universitario incompleto"]),type='rfh'
,
      xlab = "Ingreso mensual en soles",ylab="Proporción de trabajadores"
,
      main="Universitario incompleto",
      col="yellow")
plot(fdt(d$ingrmens[d$educacion=="Universitario completo"]),type='rfh',
      xlab = "Ingreso mensual en soles",ylab="Proporción de trabajadores"
,
      main="Universitario completo",
      col="green")
plot(fdt(d$ingrmens[d$educacion=="Post-universitario"]),type='rfh',
      xlab = "Ingreso mensual en soles",ylab="Proporción de trabajadores"
,
      main="Post-universitario",
      col="darkgreen")
par(mfcol=c(1,1))
```

## ## MEDIDAS PARA EL ANALISIS DESCRIPTIVO

```
tabla3<-format(summaryFull(d$ingrmens~d$educacion),scientific = F)
write.table(x = tabla3, file = "TABLA3.txt", sep = ",",
            row.names = T, col.names = TRUE) ##PARA EXPORTAR LA TABLA
```

## #PREGUNTA 4

### #GENERACIÓN DE LA VARIABLE

```
vector<-ingrmens>=5000
vector<-as.numeric(vector) #FALSO COMO 0 Y VERDADERO COMO 1
```

### #PROPORCIÓN OBTENIDA DE FORMA DIRECTA

```
tabla41<-prop.table(table(vector))
write.table(x = tabla41, file = "TABLA4.1.txt", sep = ",",
            row.names = FALSE, col.names = TRUE) ##PARA EXPORTAR LA TABLA
```

A

### #PROPORCIÓN OBTENIDA DE FORMA INDIRECTA CON LA TABLA DE FRECUENCIAS

### Proporción usando La tabla de frecuencias de La Pregunta 2  
 ### el Límite superior del segundo intervalo es 5000, por ende, se puede usar

### el porcentaje acumulado para conocer La proporción

```
menos_5000<-tabla2[2,7]
mas_5000<-100-menos_5000
tabla4<-matrix(c(menos_5000,mas_5000))
rownames(tabla4)<-c("Menos de 5000","Más de 5000")
colnames(tabla4)<-c("Porcentaje (%)")
tabla42<-as.table(tabla4)
write.table(x = tabla42, file = "TABLA4.2.txt", sep = ",",
           row.names = FALSE, col.names = TRUE) ##PARA EXPORTAR LA TABLA
```

A

#GRAFICO DE SECTORES CIRCULARES

```
etiquetas<-c("Menor a \n5000 soles ", "Mayor a \n5000 soles")
pie(prop.table(table(vector))*100, labels=etiquetas, col=c("lightskyblue",
"maroon"),
    main="Grafica N°5: Proporción de trabajadores de \ningresos menor o
mayor a 5000 soles")
```

#PREGUNTA 5

#CALCULO DEL LS Y PROMEDIO

```
Q1<-quantile(ingrmens, probs = 0.25, names = FALSE)
Q3<-quantile(ingrmens, probs=0.75, names = FALSE)
RIC<-Q3-Q1
LS<-Q3+1.5*RIC
mean(ingrmens[ingrmens>LS])
```

```
## [1] 5560.013
```

#PREGUNTA 6

#GRAFICOS POR NIVEL DE GENERO

```
par(mfcol=c(2,1))
hist(d$edad[d$genero=="Hombre"], xlab="Edad", ylab="Número de trabajadores",
    main="Hombre",
    col="darkturquoise")
hist(d$edad[d$genero=="Mujer"], xlab="Edad", ylab="Número de trabajadores",
    main="Mujer",
    col="mediumvioletred")
```

```
par(mfcol=c(1,1))
```

#Estadísticas

```

tabla6<-summaryFull(d$edad~d$genero)
write.table(x = tabla6, file = "TABLA6.txt", sep = ",",
           row.names = T, col.names = TRUE) ##PARA EXPORTAR LA TABLA

```

#PREGUNTA 7

#A)

```

hist(d$edad,xlab="Edad",ylab="Número de trabajadores",
     main="Gráfico N° 7: Cantidad de trabajadores según su edad",
     col="darkslateblue")

```

```

tabla7a<-summaryFull(d$edad)
colnames(tabla7a)<-c("Estadísticas de la variable edad")
write.table(x = tabla7a, file = "TABLA7a.txt", sep = ",",
           row.names = T, col.names = TRUE) ##PARA EXPORTAR LA TABLA

```

#B)

```

tabla7b<-summaryFull(d$edad,group = d$genero)
write.table(x = tabla7b, file = "TABLA7b.txt", sep = ",",
           row.names = T, col.names = TRUE) ##PARA EXPORTAR LA TABLA

```

#C)

```

tabla7c<-prop.table(table(d$satlab))*100
tabla7c<-cumsum(tabla7c)
tabla7c<-as.table(cbind(tabla7c))
write.table(x = tabla7c, file = "TABLA7c.txt", sep = ",",
           row.names = T, col.names = TRUE) ##PARA EXPORTAR LA TABLA

```