1. Realice un cuadro con todas las variables a trabajar y clasifíquelas en cualitativas ordinales, nominales ocuantitativasdiscretas o continuas. (+ 0.5)

Antes de comenzar con el análisis descriptivo de los datos, organizamos de la siguiente manera: (+ 0.5)

a. Cambiar los nombres de las categorías que aparecen codificadas numéricamente y nombrarlas comoapareceen la encuesta. (utilizando Rcommander, adjuntar captura de pantalla con el código)

1. Microemprendimientos\_CH\_2023 <-

2. readXL("C:/Users/USER/Desktop/Universidad de Antioquia/Semestre 3/Teoria de probabilidad y colas/Trabajo Descriptiva/Datos\_Trab\_Descriptiva.xlsx",

3. rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)

4.

5. Microemprendimientos\_CH\_2023$f2\_h<-ifelse(Microemprendimientos\_CH\_2023$f2\_h==88,"No sabe",Microemprendimientos\_CH\_2023$f2\_h)

6. Microemprendimientos\_CH\_2023$f2\_h<-ifelse(Microemprendimientos\_CH\_2023$f2\_h==99,"No responde",Microemprendimientos\_CH\_2023$f2\_h)

7. Microemprendimientos\_CH\_2023$f2\_i<-ifelse(Microemprendimientos\_CH\_2023$f2\_i==86,"Trabaja de manera ocasional",Microemprendimientos\_CH\_2023$f2\_i)

8. Microemprendimientos\_CH\_2023$f2\_i<-ifelse(Microemprendimientos\_CH\_2023$f2\_i==87,"No tiene horario regular",Microemprendimientos\_CH\_2023$f2\_i)

9. Microemprendimientos\_CH\_2023$f2\_i<-ifelse(Microemprendimientos\_CH\_2023$f2\_i==88,"No sabe",Microemprendimientos\_CH\_2023$f2\_i)

10. Microemprendimientos\_CH\_2023$f2\_i<-ifelse(Microemprendimientos\_CH\_2023$f2\_i==99,"No responde",Microemprendimientos\_CH\_2023$f2\_i)

11.

12. Microemprendimientos\_CH\_2023$f2\_h <- factor(Microemprendimientos\_CH\_2023$f2\_h)

13. Microemprendimientos\_CH\_2023$f2\_i <- factor(Microemprendimientos\_CH\_2023$f2\_i)

14.

15. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

16. f2\_b <- factor(f2\_b, labels=c('Si','No'))

17. })

18.

19. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

20. f2\_c <- factor(f2\_c, labels=c('Hombre','Mujer'))

21. })

22. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

23. f2\_e <- factor(f2\_e, labels=c('Ninguno','Primaria','Secundaria',

24. 'Superior(Técnica, Universitaria o postítulo)','No sabe'))

25. })

26. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

27. f2\_f <- factor(f2\_f, labels=c('Trabajador asalariado',

28. 'Socios trabajadores','Familiar sin pago','Aprendices o practicantes',

29. 'No sabe','No responde'))

30. })

31. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

32. f2\_g <- factor(f2\_g, labels=c('Acuerdo de palabra','Escrito',

33. 'Escrito y firmado ante notario','No sabe'))

34. })

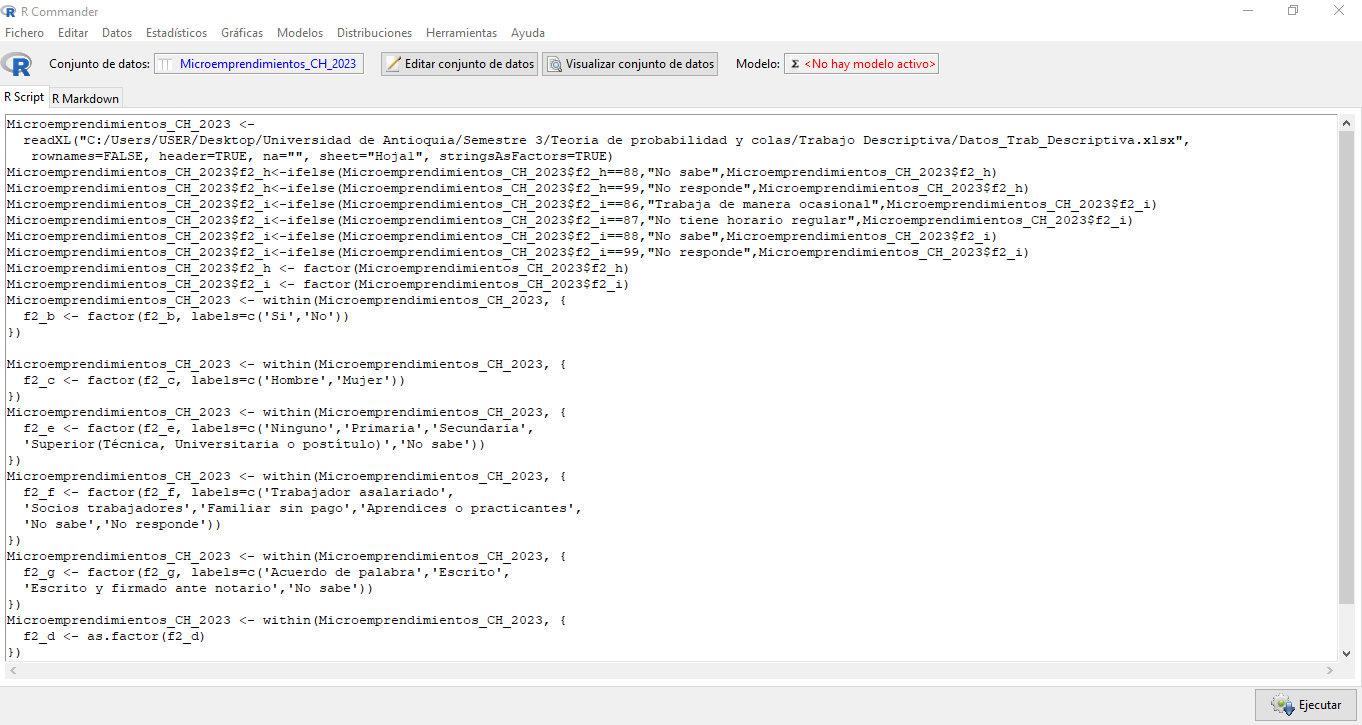
35.

36. Microemprendimientos\_CH\_2023 <- within(Microemprendimientos\_CH\_2023, {

37. f2\_d <- as.factor(f2\_d)

38. })

39.



b. Asegurese de que una vez subida la base de datos las variables correspondan a su tipo. Genere un resumen de todas las variables.

1. summary(Microemprendimientos\_CH\_2023)

Texto

Descripción generada automáticamente con confianza baja

3. Responda a las preguntas planteadas en el cuestionario

**CUESTIONARIO**

1. ¿La mayoría de trabajadores son hombres o mujeres?

Texto, Carta

Descripción generada automáticamente

1. local({

2. .Table <- with(Microemprendimientos\_CH\_2023, table(f2\_c))

3. cat("\ncounts:\n")

4. print(.Table)

5. cat("\npercentages:\n")

6. print(round(100\*.Table/sum(.Table), 2))

7. })

Gráfico, Gráfico circular

Descripción generada automáticamente

1. with(Microemprendimientos\_CH\_2023, piechart(f2\_c, xlab="", ylab="",

2. main="Sexo de trabajadores", col=rainbow\_hcl(2), scale="percent"))

**Análisis**: [Aquí iría una respuesta a la pregunta]

2. Entre los hombres, ¿qué porcentaje es integrante del hogar?

Texto

Descripción generada automáticamente

1. local({

2. .Table <- xtabs(~f2\_b+f2\_c, data=Microemprendimientos\_CH\_2023)

3. cat("\nFrequency table:\n")

4. print(.Table)

5. cat("\nColumn percentages:\n")

6. print(colPercents(.Table))

7. .Test <- chisq.test(.Table, correct=FALSE)

8. print(.Test)

9. })

**Análisis**: [Aquí iría una respuesta a la pregunta]

3. Compare con una gráfica la edad promedio entre los hombres y las mujeres. ¿Qué se puede concluir?

3.1 Primero a la variable cuantitativa discreta “Edad”, se descartan los valores que corresponden a variables categoricas (“No sabe”,”No responde”) y se crea un nuevo Conjunto de datos llamado df\_new1 para no modificar la base de datos original debido a razones de integridad de datos, auditoría, la reproducibilidad y la trazabilidad:

1. df1 <- data.frame(Microemprendimientos\_CH\_2023)

2. df\_new1<-df1[df1$f2\_d!="No sabe" & df1$f2\_d!="No responde",]

4. with(df\_new1, plotMeans(as.numeric(as.character(Microemprendimientos\_CH\_2023$f2\_d)), f2\_c, error.bars="se", xlab="Sexo",

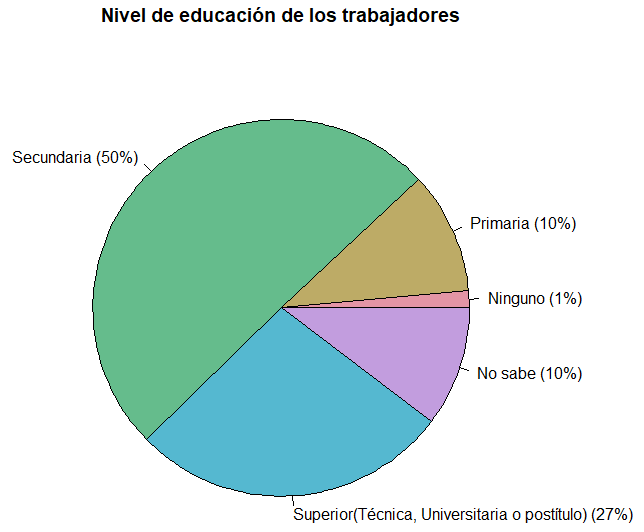
5. ylab="Promedio de edad", main="Sexo Vs Promedio de edad", connect=TRUE))

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**Análisis**: [Aquí iría una respuesta a la pregunta]

4. ¿Cuál es el nivel de educación que tienen la mayoría de trabajadores?



1. with(Microemprendimientos\_CH\_2023, piechart(f2\_e, xlab="", ylab="",

2. main="Nivel de educación de los trabajadores", col=rainbow\_hcl(5), scale="percent"))

**Análisis**: [Aquí iría una respuesta a la pregunta]

5. Entre los que tienen una educación superior ¿es mayor el porcentaje de hombre o de mujeres?

Gráfico, Gráfico de barras

Descripción generada automáticamente

1. with(Microemprendimientos\_CH\_2023, Barplot(f2\_e, by=f2\_c, style="divided",

2. legend.pos="above", xlab="Nivel de educación",

3. ylab="% Sexo del trabajador",

4. main="Nivel de educación vs Sexo de los trabajadores", scale="percent",

5. label.bars=TRUE))

**Análisis**: [Aquí iría una respuesta a la pregunta]

6. Entre las mujeres, ¿cual es el nivel de educación más frecuente?

Gráfico

Descripción generada automáticamente

1. with(Microemprendimientos\_CH\_2023, Barplot(f2\_c, by=f2\_e, style="divided",

2. legend.pos="above", xlab="Sexo del trabajador",

3. ylab="% del nivel de educación",

4. main="Sexo de los trabajadores vs Nivel de educación", scale="percent",

5. label.bars=TRUE))

**Análisis**: [Aquí iría una respuesta a la pregunta]

7. ¿Cuál es el tipo de acuerdo más popular? 

1. with(Microemprendimientos\_CH\_2023, piechart(f2\_g, xlab="", ylab="",

2. main="Tipo de acuerdo", col=rainbow\_hcl(4), scale="percent"))

**Análisis**: [Aquí iría una respuesta a la pregunta]

8. Realice un gráfico comparativo de las horas trabajadas por hombres y por mujeres. Comente al respecto

8.1 Primero a la variable cuantitativa continua “Horas semanales que trabaja” se descartan los valores que corresponden a variables categoricas (“Trabaja de manera ocasional”,”No tiene horario regular”, “No sabe”, “No responde”) y se crea un nuevo Conjunto de datos llamado df\_new para no modificar la base de datos original debido a razones de integridad de datos, auditoría, la reproducibilidad y la trazabilidad:

1. df <- data.frame(Microemprendimientos\_CH\_2023)

2. df\_new<-df[df$f2\_i!=86 & df$f2\_i!=87 & df$f2\_i!=88 & df$f2\_i!=99,]

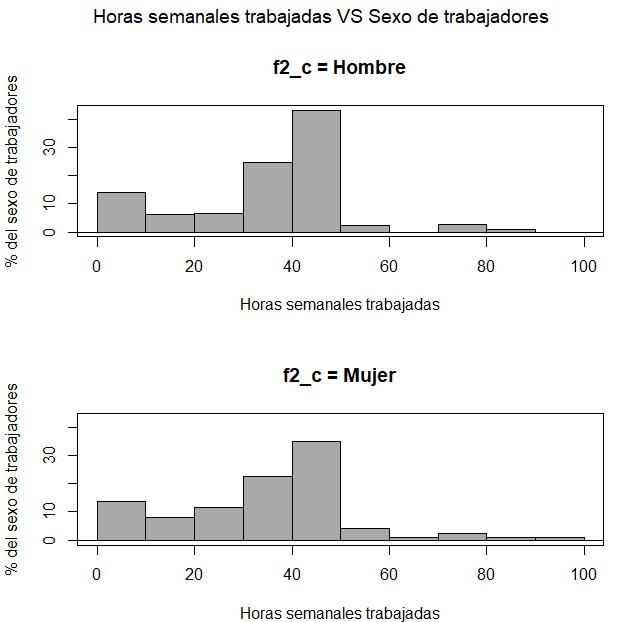
3.

4. with(df\_new, Hist(as.numeric(as.character(df\_new$f2\_i)), groups=f2\_c, scale="percent", breaks="Sturges",

5. col="darkgray", xlab="Horas semanales trabajadas",

6. ylab="% del sexo de trabajadores",

7. main="Horas semanales trabajadas VS Sexo de trabajadores"))



**Análisis**: [Aquí iría una respuesta a la pregunta]

9. El monto total bruto esta en pesos chilenos, cree una nueva variable donde el monto esté en pesos colombianos

9.1 El equivalente a 1 peso Chileno a peso Colombiano, en la fecha 2 de septiembre del 2023 a las 10:52 A.M UTC es de 4.80.

9.2 8.1 Primero a la variable cuantitativa continua “Monto total bruto” se descartan los valores que corresponden a variables categoricas (“No sabe”, “No responde”) y se crea un nuevo Conjunto de datos llamado df\_new2 para no modificar la base de datos original debido a razones de integridad de datos, auditoría, la reproducibilidad y la trazabilidad:

1. df2 <- data.frame(Microemprendimientos\_CH\_2023)

2. df\_new2<-df2[df2$f2\_h!="No sabe" & df2$f2\_h!="No responde",]

3. df\_new2$monto\_bruto\_COP<-as.numeric(as.character(df\_new2$f2\_h))\*4.80

10. Realice un boxplot para el monto total bruto y de las conclusiones pertinentes.

1. Boxplot( ~ monto\_bruto\_COP, data=df\_new2, id=list(method="y"),

2. ylab="monto total bruto mensual", main="Boxplot de monto total bruto mensual")

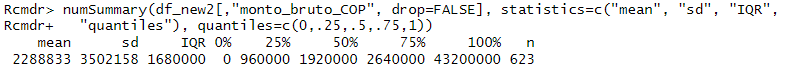
Imagen que contiene Tabla

Descripción generada automáticamente

**Nota**: Al notar un diagrama de cajas bastante particular, se decidió realizar un resumen numerico del monto total bruto mensual

1. numSummary(df\_new2[,"monto\_bruto\_COP", drop=FALSE], statistics=c("mean", "sd", "IQR",

2. "quantiles"), quantiles=c(0,.25,.5,.75,1))



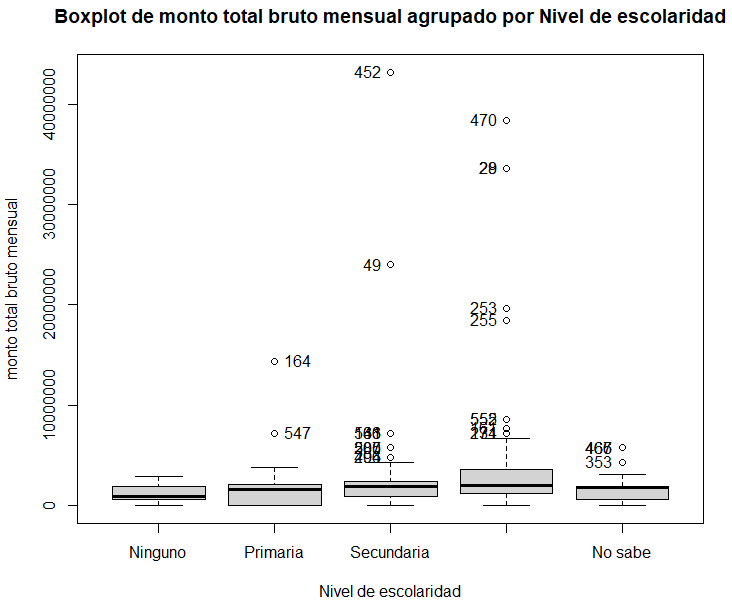
**Análisis**: [Aquí iría una respuesta a la pregunta]

11. Realice un boxplot entre el monto total bruto y el nivel de educación. De comentarios al respecto

1. Boxplot(monto\_bruto\_COP ~ f2\_e, data=df\_new2, id=list(method="y"),

2. xlab="Nivel de escolaridad", ylab="monto total bruto mensual",

3. main="Boxplot de monto total bruto mensual agrupado por Nivel de escolaridad")



**Análisis**: [Aquí iría una respuesta a la pregunta]

12. Realice un histograma de la las horas semanales que trabajan, si se obtuviera una curva de frecuencias, ¿quécaracterísticas tendría?

12.1 Primero a la variable cuantitativa continua “Horas semanales que trabaja” se descartan los valores que corresponden a variables categoricas (“Trabaja de manera ocasional”,”No tiene horario regular”, “No sabe”, “No responde”) y se crea un nuevo Conjunto de datos llamado df\_new para no modificar la base de datos original debido a razones de integridad de datos, auditoría, la reproducibilidad y la trazabilidad:

1. df3 <- data.frame(Microemprendimientos\_CH\_2023)

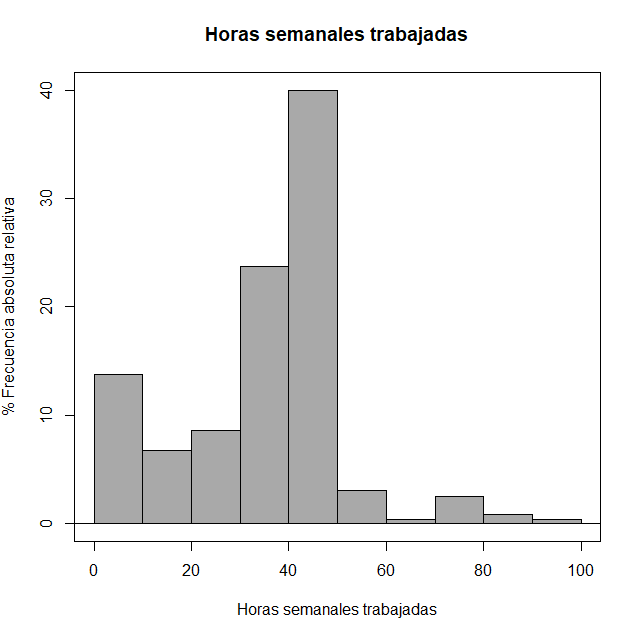
2. df\_new3<-df3[df3$f2\_i!=86 & df3$f2\_i!=87 & df3$f2\_i!=88 & df3$f2\_i!=99,]

4. with(df\_new3, Hist(as.numeric(as.character(df\_new$f2\_i)), scale="percent", breaks="Sturges",

5. col="darkgray", xlab="Horas semanales trabajadas",

6. ylab="% Frecuencia absoluta relativa",

7. main="Horas semanales trabajadas"))



**Análisis**: [Aquí iría una respuesta a la pregunta]

13. Obtenga un resumen numérico de todas las variables para los trabajadores que son integrantes del hogar.

13.1 Primero se eliminan los registros que no sean integrantes del hogar:

1. df8 <- data.frame(Microemprendimientos\_CH\_2023)

2. df\_new8<-df8[df8$f2\_b!="No",]

3. df\_new8$f2\_b<-as.numeric(as.character(df\_new8$f2\_b))

4. df\_new8$f2\_b[is.na(df\_new8$f2\_b)] <- 1

13.2 Se calcula un resumen númerico para cada una de las variables con los trabajadores que son integrandes del hogar.

**Nota**: Para calcular un resumen númerico, se procede a trabajar con una copia de la base de datos original en la que se limpiarán para las variables cuantitativas toda variable cualitativa.

13.2.1 Importamos primero la base de datos original:

1. Microemprendimientos\_COPY <- readXL("C:/Users/USER/Desktop/Universidad\_de\_Antioquia/Semestre\_3/Teoria\_de\_probabilidad\_y\_colas/Trabajo Descriptiva/Datos\_Trab\_Descriptiva.xlsx",

3. rownames=FALSE, header=TRUE, na="", sheet="Hoja1", stringsAsFactors=TRUE)

13.2.2 Eliminamos los registros de trabajadores que no sean integrantes del hogar:

1. Microemprendimientos\_COPY<- Microemprendimientos\_COPY[Microemprendimientos\_COPY$f2\_b!=2,]

13.2.3 Convertimos la variable númerica “Integrante del hogar” en factor, en este caso por el valor de “Si”:

1. Microemprendimientos\_COPY <- within(Microemprendimientos\_COPY, {

2. f2\_b <- factor(f2\_b, labels=c('Si'))

3. })

Ahora ya podemos empezar a realizar los distintos resumenes numericos:

**13.2.1 Resumen númerico de Sexo vs Integrantes del hogar:**

1. numSummary(Microemprendimientos\_COPY[,"f2\_c", drop=FALSE],

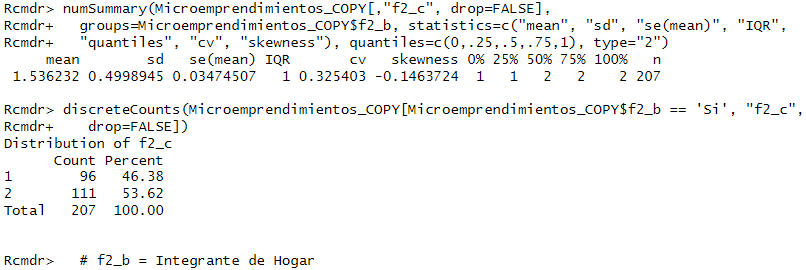
2. groups=Microemprendimientos\_COPY$f2\_b, statistics=c("mean", "sd", "se(mean)", "IQR",

3. "quantiles", "cv", "skewness"), quantiles=c(0,.25,.5,.75,1), type="2")

4. discreteCounts(Microemprendimientos\_COPY[Microemprendimientos\_COPY$f2\_b == 'Si', "f2\_c",

5. drop=FALSE])

6. # f2\_b = Integrante de Hogar



**13.2.2 Resumen númerico de Edad vs Integrantes del hogar:**

Antes de calcular un resumen numerico, se procede a eliminar de esta variable cuantitativa cualquier tipo de dato cualitativo.

1. dfAge <- data.frame(Microemprendimientos\_COPY)

2. df\_Age<-dfAge[dfAge$f2\_d!=888 & dfAge$f2\_d!=999,]

1. numSummary(df\_Age[,"f2\_d", drop=FALSE], groups=df\_Age$f2\_b, statistics=c("mean", "sd", "se(mean)", "IQR", "quantiles",

2. "cv"), quantiles=c(0,.25,.5,.75,1))

3. binnedCounts(df\_Age[df\_Age$f2\_b == 'Si', "f2\_d", drop=FALSE])

4. # f2\_b = Integrante de Hogar

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**13.2.3 Resumen númerico de Nivel educacional vs Integrantes del hogar:**

1. numSummary(Microemprendimientos\_COPY[,"f2\_e", drop=FALSE], groups=Microemprendimientos\_COPY$f2\_b, statistics=c("mean",

2. "sd", "se(mean)", "IQR", "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

3. discreteCounts(Microemprendimientos\_COPY[Microemprendimientos\_COPY$f2\_b == 'Si', "f2\_e", drop=FALSE])

4. # f2\_b = Integrante de Hogar

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**13.2.4 Resumen númerico de Categoría ocupacional vs Integrantes del hogar:**

1. numSummary(Microemprendimientos\_COPY[,"f2\_f", drop=FALSE], groups=Microemprendimientos\_COPY$f2\_b, statistics=c("mean",

2. "sd", "se(mean)", "IQR", "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

3. discreteCounts(Microemprendimientos\_COPY[Microemprendimientos\_COPY$f2\_b == 'Si', "f2\_f", drop=FALSE])

4. # f2\_b = Integrante de Hogar

Aplicación

Descripción generada automáticamente con confianza baja

**13.2.5 Resumen númerico de Tipo de acuerdo vs Integrantes del hogar:**

1. numSummary(Microemprendimientos\_COPY[,"f2\_g", drop=FALSE], groups=Microemprendimientos\_COPY$f2\_b, statistics=c("mean",

2. "sd", "se(mean)", "IQR", "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

3. discreteCounts(Microemprendimientos\_COPY[Microemprendimientos\_COPY$f2\_b == 'Si', "f2\_g", drop=FALSE])

4. # f2\_b = Integrante de Hogar

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**13.2.6 Resumen númerico de Monto total bruto vs Integrantes del hogar:**

Antes de calcular un resumen numerico, se procede a eliminar de esta variable cuantitativa cualquier tipo de dato cualitativo.

1. dfSalaryM <- data.frame(Microemprendimientos\_COPY)

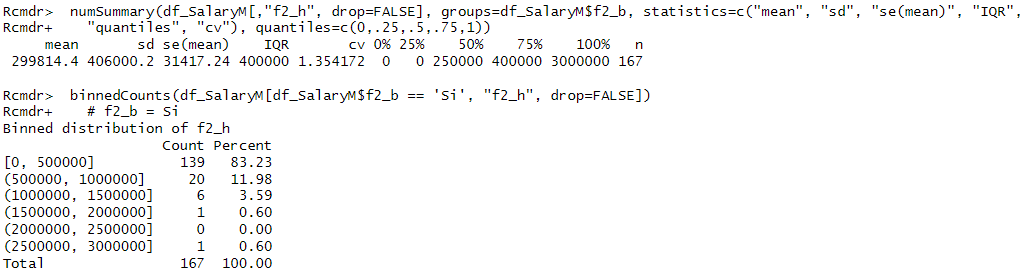
2. df\_SalaryM<-dfSalaryM[dfSalaryM$f2\_h!=88 & dfSalaryM$f2\_h!=99,]

1. numSummary(df\_SalaryM[,"f2\_h", drop=FALSE], groups=df\_SalaryM$f2\_b, statistics=c("mean", "sd", "se(mean)", "IQR",

2. "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

3. binnedCounts(df\_SalaryM[df\_SalaryM$f2\_b == 'Si', "f2\_h", drop=FALSE])

4. # f2\_b = Integrante de Hogar



**13.2.7 Resumen númerico de Horas semanales trabajadas vs Integrantes del hogar:**

Antes de calcular un resumen numerico, se procede a eliminar de esta variable cuantitativa cualquier tipo de dato cualitativo.

1. dfHours <- data.frame(Microemprendimientos\_COPY)

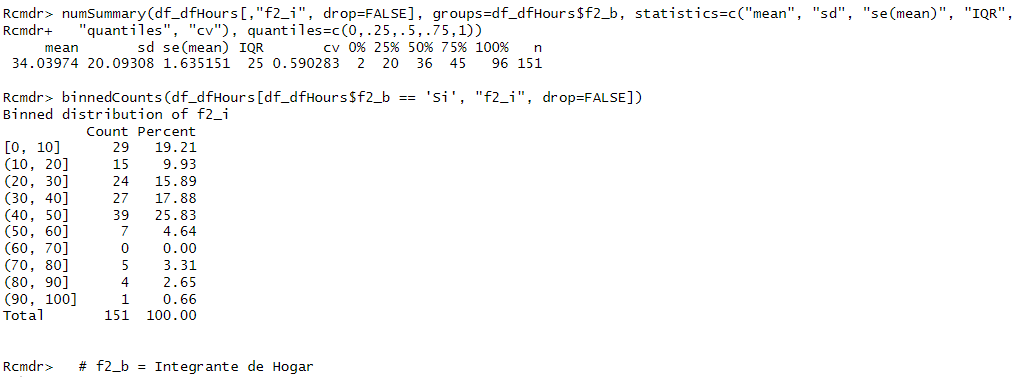
2. df\_dfHours<-dfHours[dfHours$f2\_i!=86 & dfHours$f2\_i!=87& dfHours$f2\_i!=88& dfHours$f2\_i!=99,]

1. numSummary(df\_dfHours[,"f2\_i", drop=FALSE], groups=df\_dfHours$f2\_b, statistics=c("mean", "sd", "se(mean)", "IQR",

2. "quantiles", "cv"), quantiles=c(0,.25,.5,.75,1))

3. binnedCounts(df\_dfHours[df\_dfHours$f2\_b == 'Si', "f2\_i", drop=FALSE])

4. # f2\_b = Integrante de Hogar



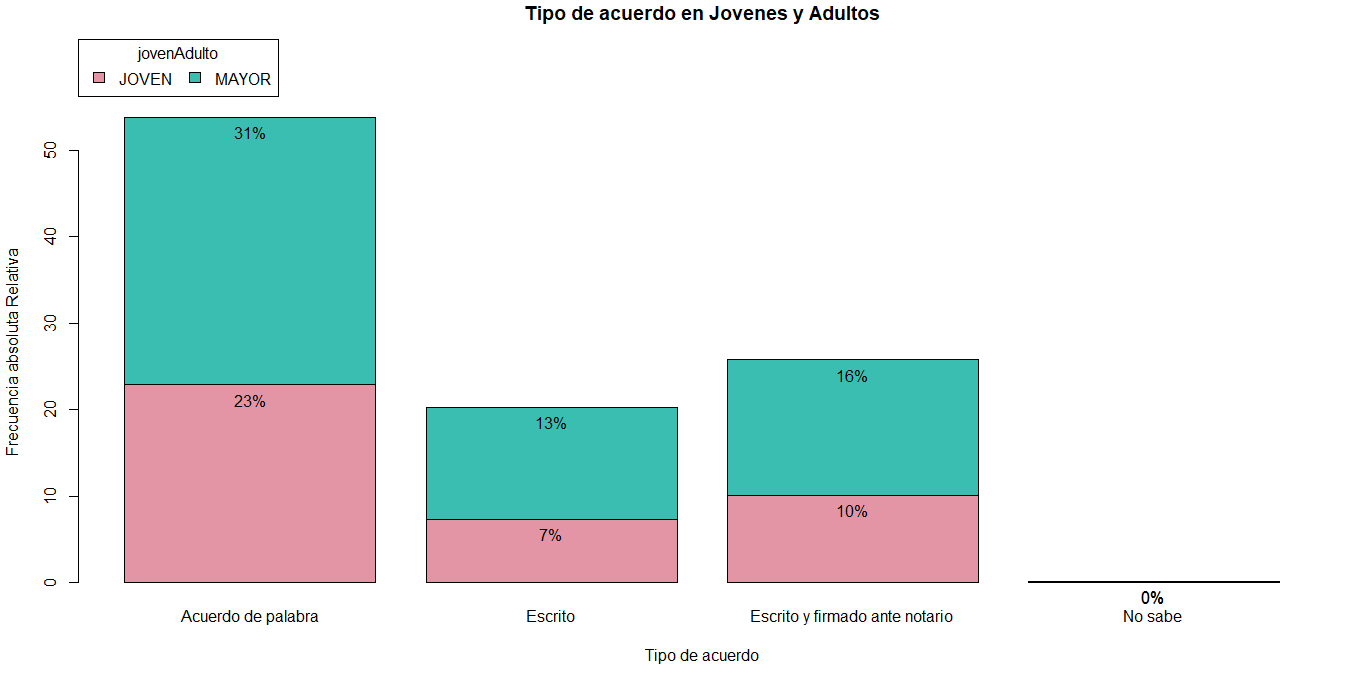
**Análisis**: [Aquí iría una respuesta a la pregunta]

14. Cree una variable jóvenes y adultos, donde jóvenes son aquellos empleados menores a 35 años, y adultos sonaquellos mayores o iguales a 36 años. ¿los acuerdos de escritos y firmados ante notario son más frecuentes conlosjóvenes o con los adultos?

14.1 Primero se realiza una copia de la base de datos original y se crea una columna con una variable cualitativa donde si es mayor a 35 el valor de la columna f2\_d (Edad), es MAYOR sino es MENOR:

1. dfAdultosJovenes <- data.frame(Microemprendimientos\_CH\_2023)

2. dfAdultosJovenes$jovenAdulto <- ifelse(as.numeric(as.character(dfAdultosJovenes$f2\_d)) < 35, "JOVEN", "MAYOR")



1. with(dfAdultosJovenes, Barplot(f2\_g, by=jovenAdulto, style="divided", legend.pos="above", xlab="Tipo de acuerdo",

2. ylab="Frecuencia absoluta Relativa", main="Tipo de acuerdo en Jovenes y Adultos", scale="percent", conditional=FALSE,

3. label.bars=TRUE))

4.

**Análisis**: [Aquí iría una respuesta a la pregunta]