Clase_4

2024-09-27

Ejemplo 1:

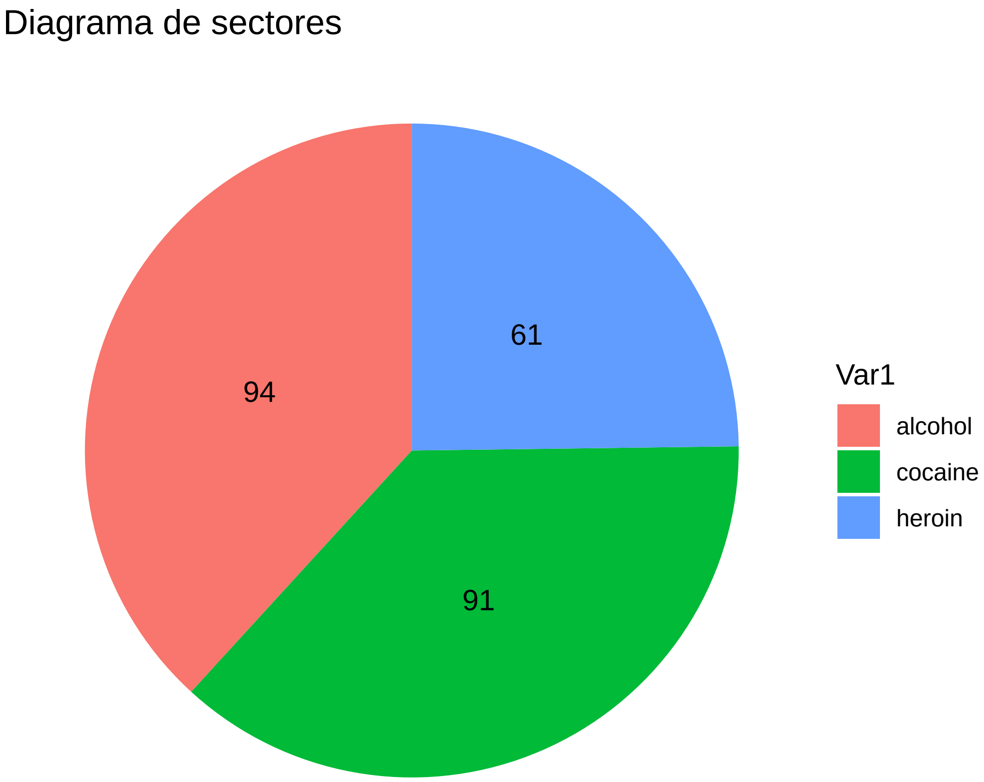
```
df2 = read.table("/home/sebastian/Documents/Maestría/Metodos_estadisticos_av/Datos/HIPOTESIS.txt", heade
```

Ejemplo 2

```
sust = table(df2$substance)
sust = data.frame(sust)
print(sust)
```

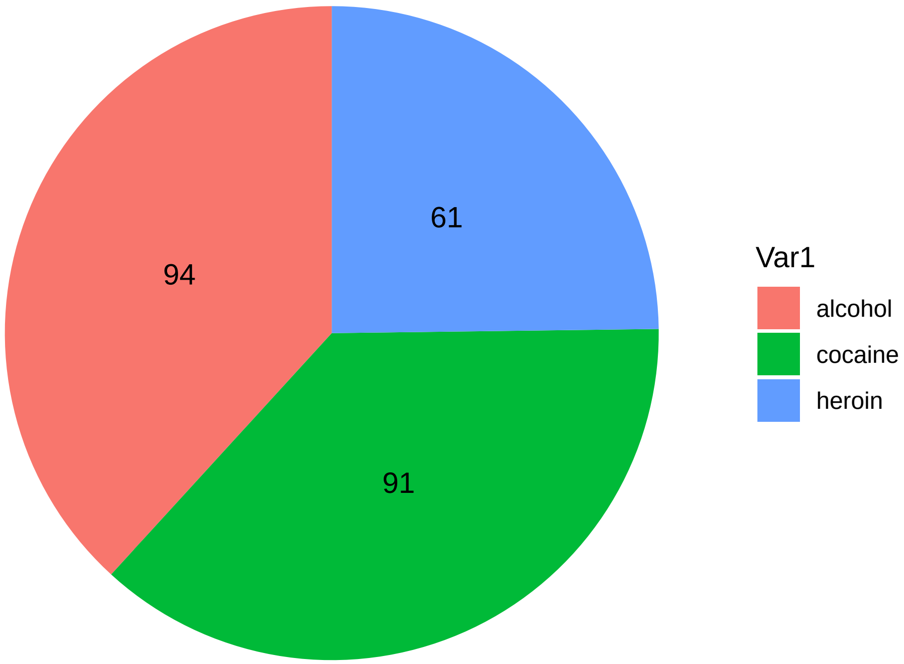
```
##      Var1 Freq
## 1 alcohol   94
## 2 cocaine   91
## 3  heroin   61
```

```
library(ggplot2)
ggplot(data = sust, aes(x = "", y = Freq, fill = Var1)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  geom_text(aes(label = Freq), position = position_stack(vjust = 0.5)) +  # Usar Freq directamente
  labs(title = "Diagrama de sectores") +
  theme_void()
```



```
library(ggplot2)
ggplot(data = sust, aes(x="", y=Freq, fill=Var1)) +
  geom_bar(stat="identity", width=1) +
  coord_polar(theta="y") +
  geom_text(aes(label=Freq), position=position_stack(vjust=0.5)) +  # Usar directamente Freq y ajustar
  labs(title="Diagrama de coso") +
  theme_void()
```

Diagrama de coso



Definir valores para el intervalo de confianza PARA PROPORCIONES, que sirve para variables cualitativas

Vamos a crear el intervalo de confianza para la muestra de 246 datos guardados en la variable n

Usaremos la función prop.test que recibe los siguientes parametros

prop.test(x, n, p = NULL, alternative = c("two.sided", "less", "greater"), conf.level = 0.95, correct = TRUE)

- Prop significa prueba de proporción en R studio
- x es el vector del valor de la tabla "sust", es decir, le especifico la ubicación en x,y y me devuelve el número que esté allí
- n el número de filas
- p es el índice o valor de confianza
- conf.level es el nivel de confianza

```
n = nrow(df2)
#Vamos a crear el intervalo de confianza para la muestra de 246 datos guardados en la variable n

#Prop significa prueba de proporción en R studio
#x es el vector del valor de la tabla "sust", es decir, le especifico la ubicación en x,y y me devuelve
#n el número de filas
#conf.level es el nivel de confianza
prop.test(sust[1,2], n , conf.level = 0.95, correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  sust[1, 2] out of n, null probability 0.5
## X-squared = 13.675, df = 1, p-value = 0.0002174
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3236478 0.4442050
## sample estimates:
##         p
## 0.3821138
```

```r
binom.test(94,n,conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  94 and n
## number of successes = 94, number of trials = 246, p-value = 0.0002627
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3211125 0.4459929
## sample estimates:
## probability of success
##              0.3821138
```

CONCLUSIÓN: Con un 95 de confianza, la proporción de personas que consumen alcohol está entre u 32% y un 44%. La probabilidad de que el verdadero valor no se encuentre en el intervalo es del 5%

Ahora, vamos a encontrar el límite inferior:

```r
p = 94/n
qnorm(0.025,lower.tail =  F)
```

```
## [1] 1.959964
```

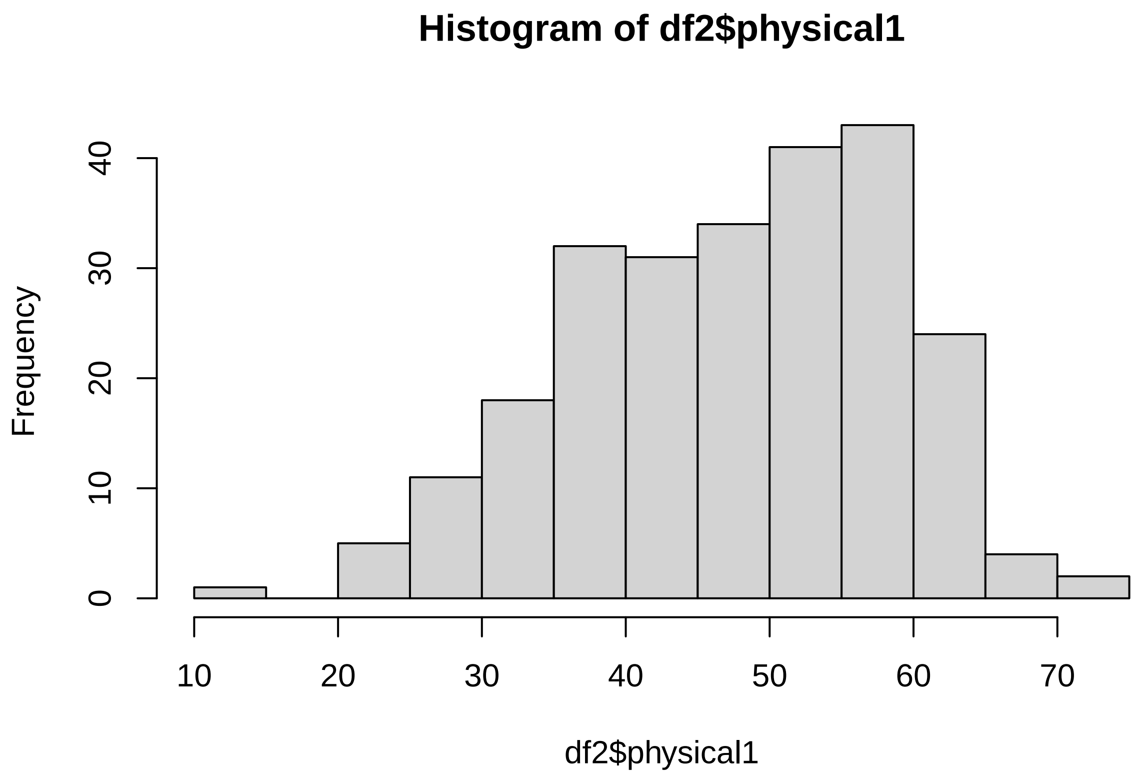
```r
Li = p-qnorm(0.025, lower.tail= F )*(sqrt(p*(1-p)/n))
```

Ahora, vamos a hacer intervalos de confianza para variables cualitativas

Usaremos la variable “physical1” para definir la media poblacional

```r
physical1 = t.test(df2$physical1, conf.level = 0.95)

hist(df2$physical1)
```



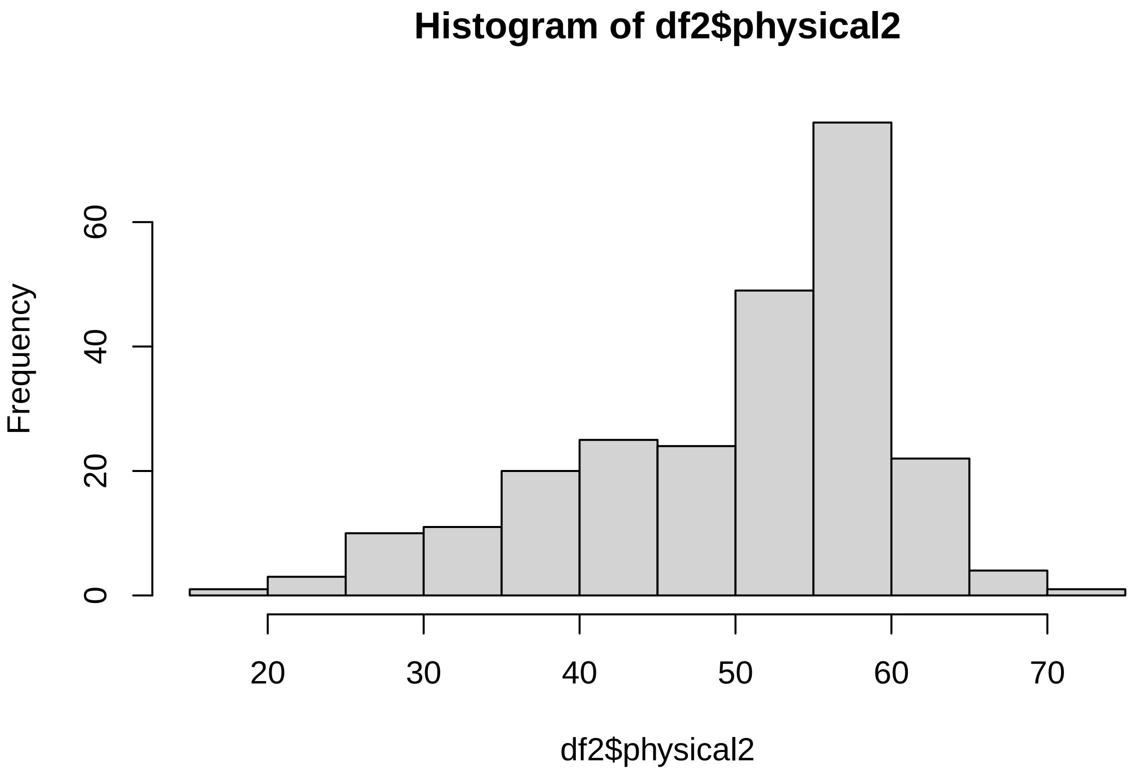

```r
print(physical1)
```

```
##
##  One Sample t-test
##
## data:  df2$physical1
## t = 66.265, df = 245, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  46.08438 48.90797
## sample estimates:
## mean of x
##  47.49617
```

```r
t.test(df2$physical2, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  df2$physical2
## t = 76.016, df = 245, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  48.85815 51.45748
## sample estimates:
## mean of x
##  50.15782
```

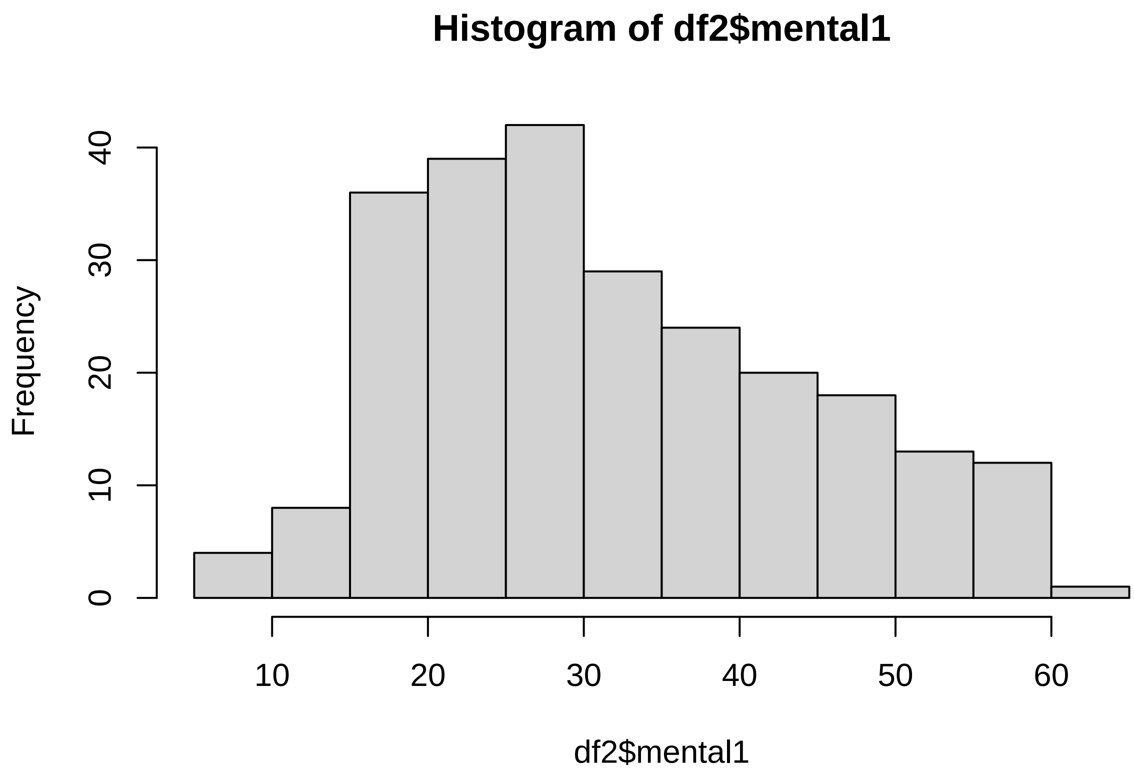
```r
hist(df2$physical2)
```




```r
t.test(df2$mental1, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  df2$mental1
## t = 39.796, df = 245, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  30.11237 33.24840
## sample estimates:
## mean of x
##  31.68038
```

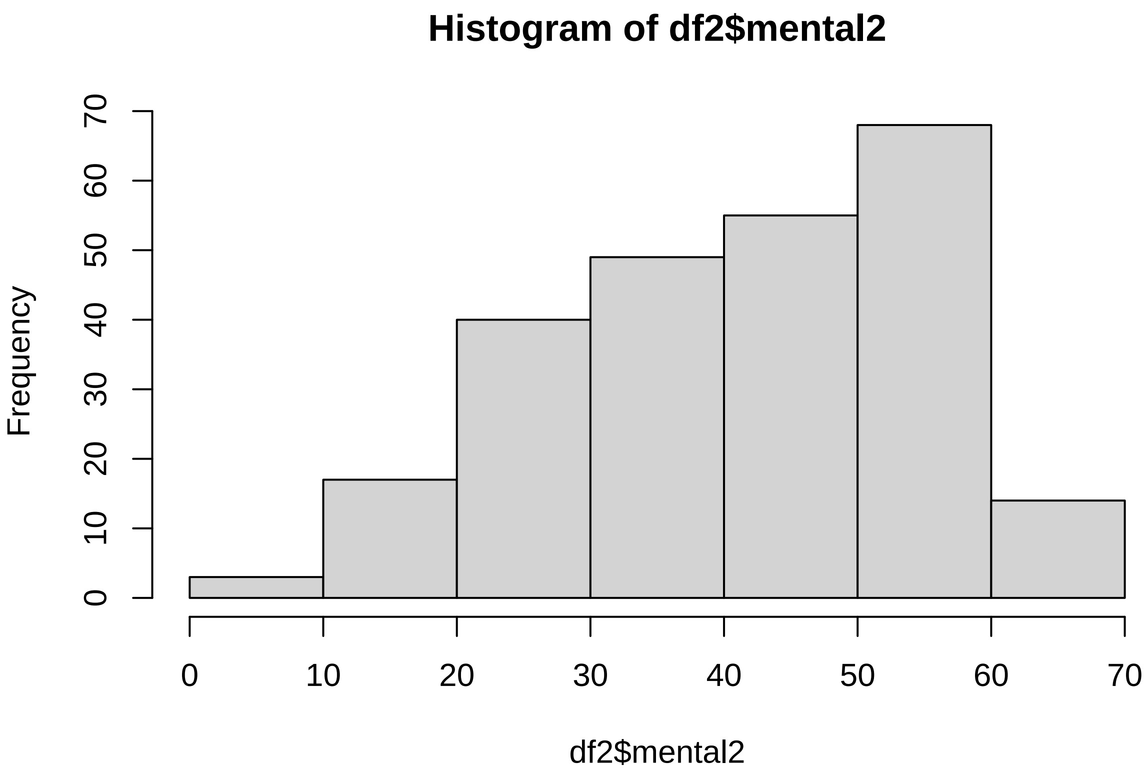
```r
hist(df2$mental1)
```




```
t.test(df2$mental2, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  df2$mental2
## t = 46.582, df = 245, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  39.25050 42.71642
## sample estimates:
## mean of x
##  40.98346
```

```
hist(df2$mental2)
```



Ahora vamos a hallar la Varianza a la variable physical1 de nuestro dataset

```r
chisq.test(df2$physical1)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  df2$physical1
## X-squared = 651.91, df = 245, p-value < 2.2e-16
```

Hallemos nuevamente el límite inferior a physical1 de nuestro dataset

```r
Li = (n-1) * var(df2$physical1)/qchisq(0.025, n-1, lower.tail = F)
Ls = (n-1) * var(df2$physical1)/qchisq(0.025, n-1, lower.tail = T)

print(Li)
```

```
## [1] 106.6786
```

```r
sqrt(Li)
```

```
## [1] 10.32853
```

```r
sqrt(Ls)
```

```
## [1] 12.33388
```

Ahorra la varianza de physical

```r
chisq.test(df2$physical2)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  df2$physical2
```

```
## X-squared = 523.15, df = 245, p-value < 2.2e-16
```

```r
Li2 = (n-1) * var(df2$physical2)/qchisq(0.025, n-1, lower.tail = F)
Ls2 = (n-1) * var(df2$physical2)/qchisq(0.025, n-1, lower.tail = T)

print(Li2)
```

```
## [1] 90.40643
```

```r
sqrt(Li2)
```

```
## [1] 9.50823
```

```r
sqrt(Ls2)
```

```
## [1] 11.35431
```