

# Universidad de la Sabana



## Probabilidad y Estadística II

### Proyecto de Estadística

#### Presentado por:

Brayan Ardila Rodriguez, Samuel Campo Decastro, Pablo Campuzano

#### Profesor:

Sergio Arciniegas Alarcón

***Periodo 2022-2***  
**Chía, Cundinamarca**

Esta muestra es resultado de la iniciativa de estudiantes de Probabilidad y Estadística II de la Universidad de La Sabana para conocer conceptos básicos sobre los estudiantes que salen de la materia "Probabilidad y Estadística I" para el semestre 2022-1 . Se realiza una selección de 7 variables en total , las variables escogidas se dan de manera natural en el entendimiento ,y relación de los valores de esta variables con el promedio final de la materia. Durante el el presente trabajo expondremos las principales características del conjunto de datos y pondremos en juicio algunas de las preguntas presentadas por el equipo que tomó la muestra.

### **Planteamiento del Problema**

¿Existe una relación intrínseca entre la nota promedio final obtenida por los estudiantes y las variables escogidas para la materia "Probabilidad y Estadística I" para el semestre 2022-1?

### **Objetivo General**

-Identificar las variables que intervienen activamente en la obtención del puntaje promedio de la materia "Probabilidad y Estadística I" para el semestre 2022-1

### **Objetivo Especifico**

-Analizar las variables y sus relaciones para poder saber si existe una correlación entre las notas , otras variables y el promedio como variable respuesta.

## Definición de Variables

### Variables Cualitativas

Escala Nominal:

- Genero :F-femenino ,M-masculino
- ¿Realiza actividades extracurriculares? :Si , NO

Escala Ordinal:

- ¿Repasa los temas vistos en clase? : Nunca, Pocas Veces,Algunas Veces,Casi siempre , Siempre

### Variables Cuantitativas

Discreta :

- Edad (en años) : mínimo 15 años y máximo 25 años [razón]
- Tiempo de estudio semanal (horas) : mínimo 30 horas y máximo 55 horas [razón]
- ¿A cuántas tutorías asistió durante el semestre? : mínimo 1 y máximo 10 [razón]

Continua :

- Promedio Primer Semestre : entre 3.0 y 4.9 [razón]

## Poblacion y muestra

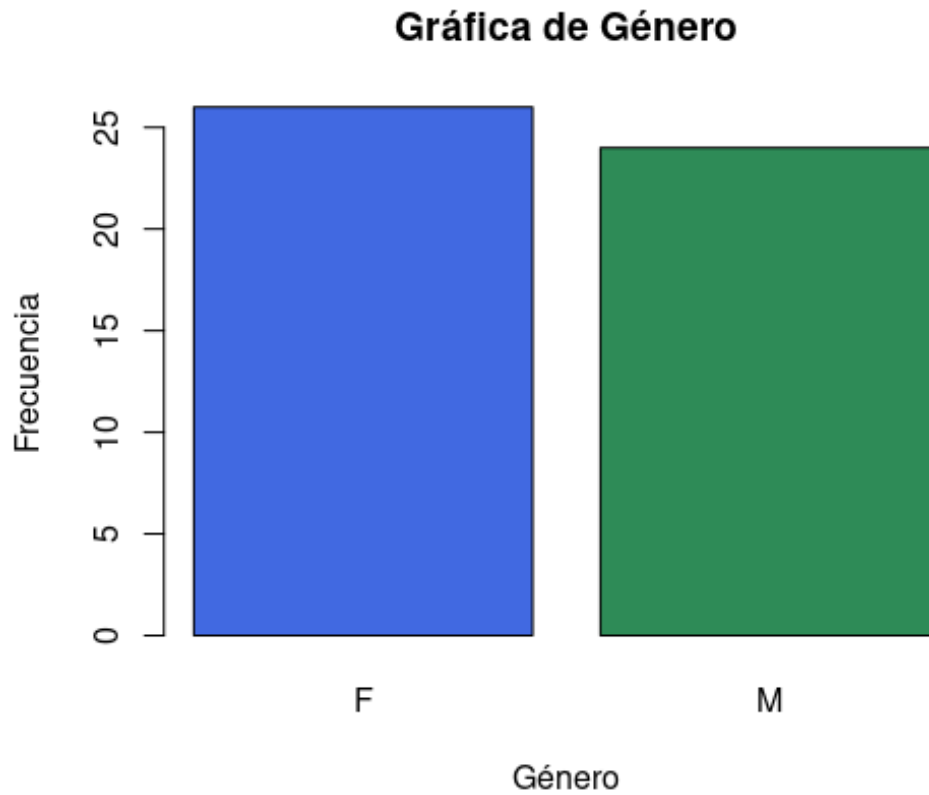
la poblacion objetivo serian todos los estudiantes de la materia "Probabilidad y Estadística I" de 2020'-1 los cuales se toma una muestra aleatoria de los estudiantes , la estrategia planteada es usar la lista obtenida de los estudiantes bajo un muestreo aleatorio simple .

1. se toma la lista de los estudiantes de todos los cursos , luego se enumeran en excel luego esta lista se guarda como un archivo CSV.
- 2.se abre el archivo csv en R y se usa la funcion SAMPLE que toma un vector de datos y nos devuelve un vector con el tamaño de muestra que se requiere en este caso de 70 correos ya que si se toman los 50 exactos se corre el peligro de no alcanzar a la muestra necesaria pues no todos contestarán la encuesta .
3. a cada correo se asocian las respuestas y se colocan dentro de la base de datos .

## Análisis de datos culitativos

¿cuál es su Género ?

Revisemos el gráfico de barras para la variable género

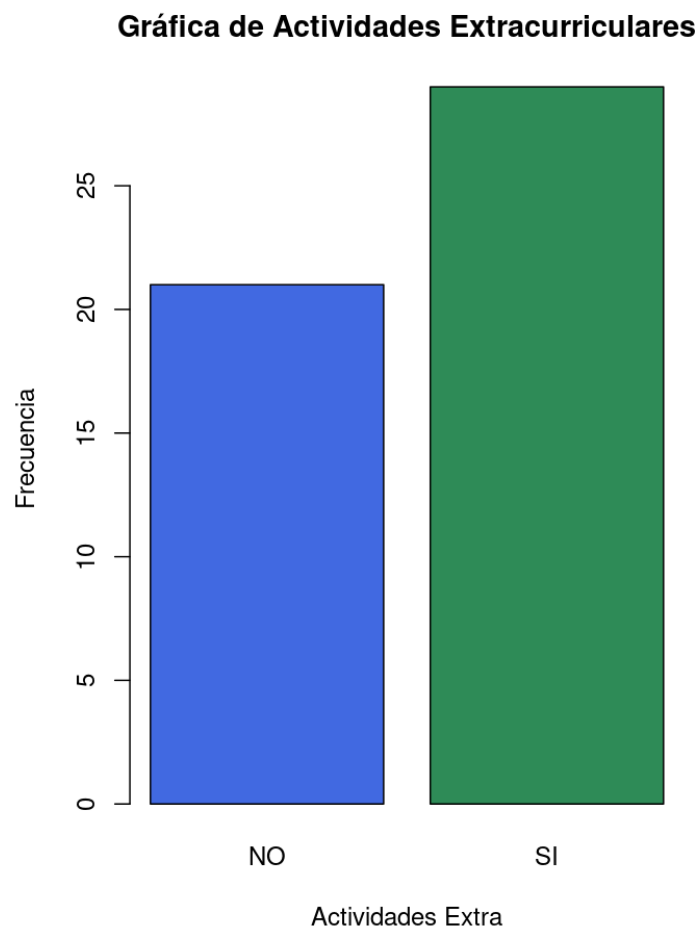


Númericamente hablando encontramos que la proporción de las personas en el género son casi iguales .Ecnontrando que 26 personas son mujere y 24 personas son hombres.

Frecuencia	
F	26
M	24

**¿Realiza Actividades Curriculares?**

Encontramos que la cantidad de personas que contestaron que “SI a la realización de actividades exrtacurricualres es mayor que el número de personas han dicho “NO”



La cantidad de personas que contestaron “SI” es de 29 personas supernado al número de personas que contestaron “NO”con una cantidad de 21.

Frecuencia	
NO	21
SI	29

## ¿Repasa los temas en clase?

La mayor cantidad de personas contestaron que “casi siempre” repasan lo temas de clase , y la menor cantidad de personas han dicho que “pocas veces” repasa los temas de clase.



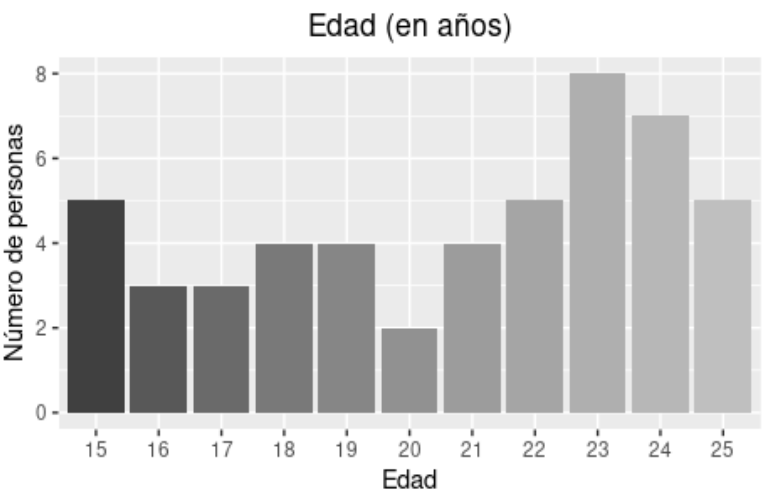
La cantidad de personas que respondieron a “Casi siempre “ que es la mayoría son 15.

	Frecuencia
Algunas veces	11
Casi siempre	15
Nunca	9
Pocas veces	7
Siempre	8

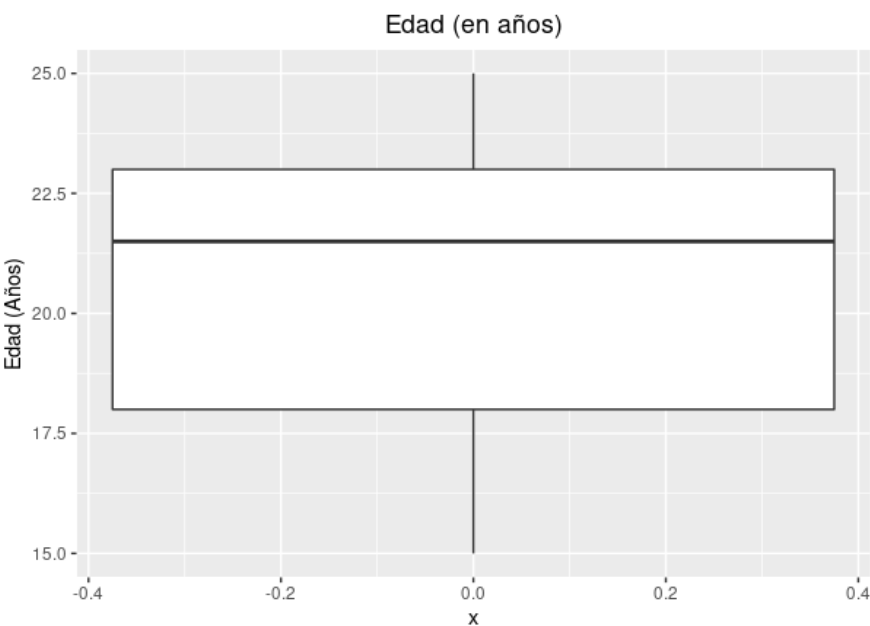
**Análisis de datos Cuantitativos**

Se establecerán las principales características de las variables cuantitativas encontradas en la investigación .

*Edad (en años)*



Se observa uniformidad en la cantidad de estudiantes por cada edad además de que al ser universitarios están entre los 15 y 25 años.

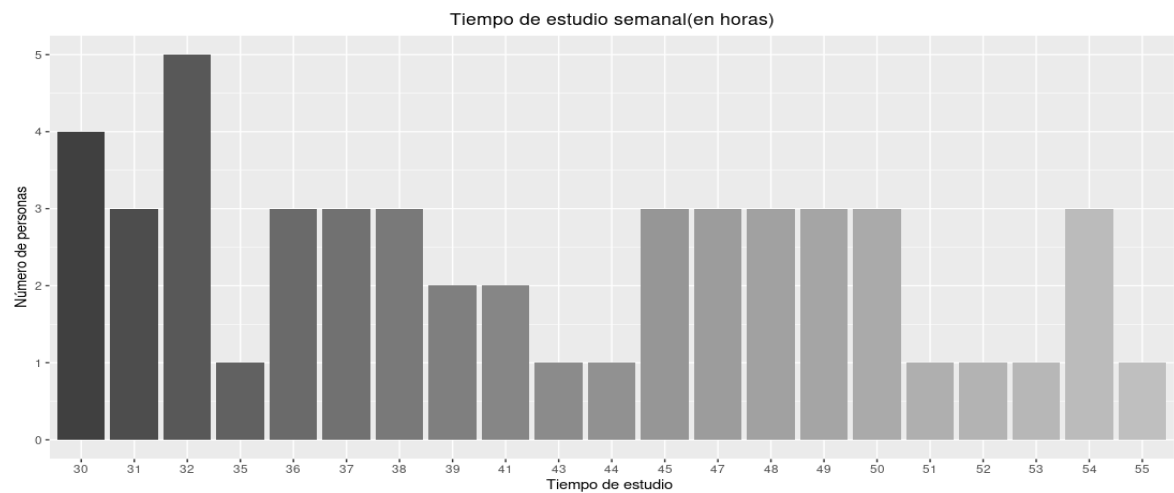


Descripción numérica de Edad

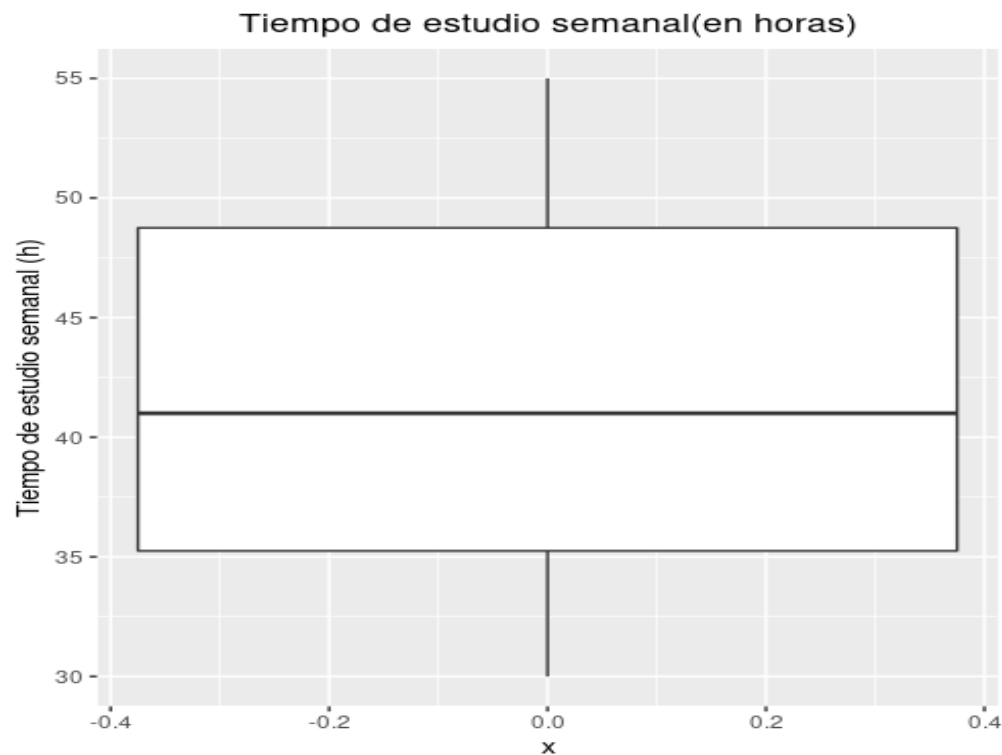
	Datos
Promedio	20.66
Mediana	21.50
Varianza	10.84
SD	3.29
CV	15.94
Q1	18.00
Q2	21.50
Q3	23.00
Minimo	15.00
Maximo	25.00

la distribución de los datos parece no tener datos atípicos .Además CV es menor al 20% la media será una buena medida descriptiva de la centralidad edad de los estudiantes universitarios.

Tiempo de estudio semanal (en horas)



Se observa uniformidad en la cantidad de tiempo dedicado al estudio por los estudiantes .Es decir el histograma del tiempo de estudio no presenta una tendencia aparente ni sesgos distinguibles .



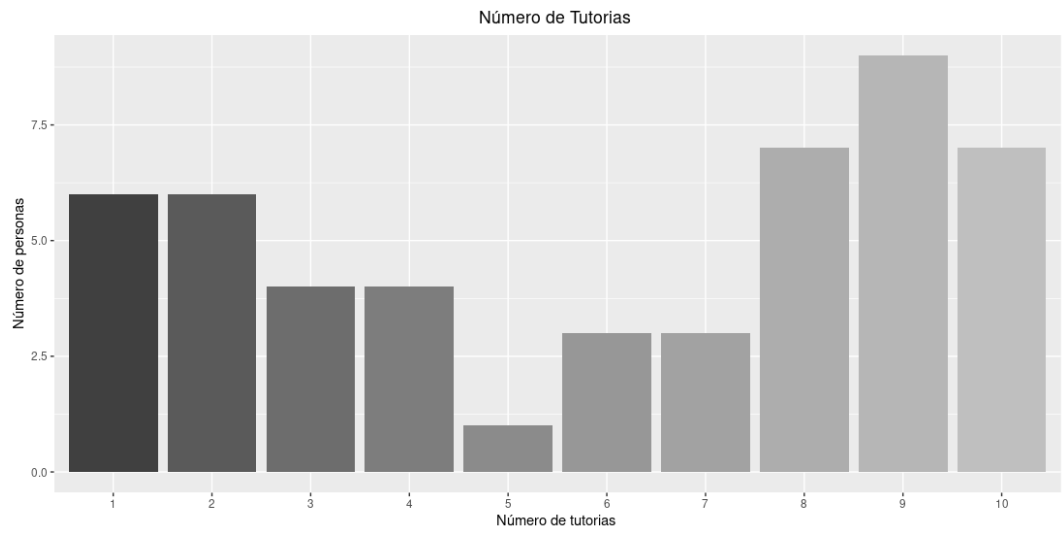
Descripción numérica para la variable Tiempo de estudio semanal.

	Datos
Promedio	41.56
Mediana	41.00
Varianza	65.03
SD	8.06
CV	19.40
Q1	35.25
Q2	41.00
Q3	48.75
Minimo	30.00
Maximo	55.00

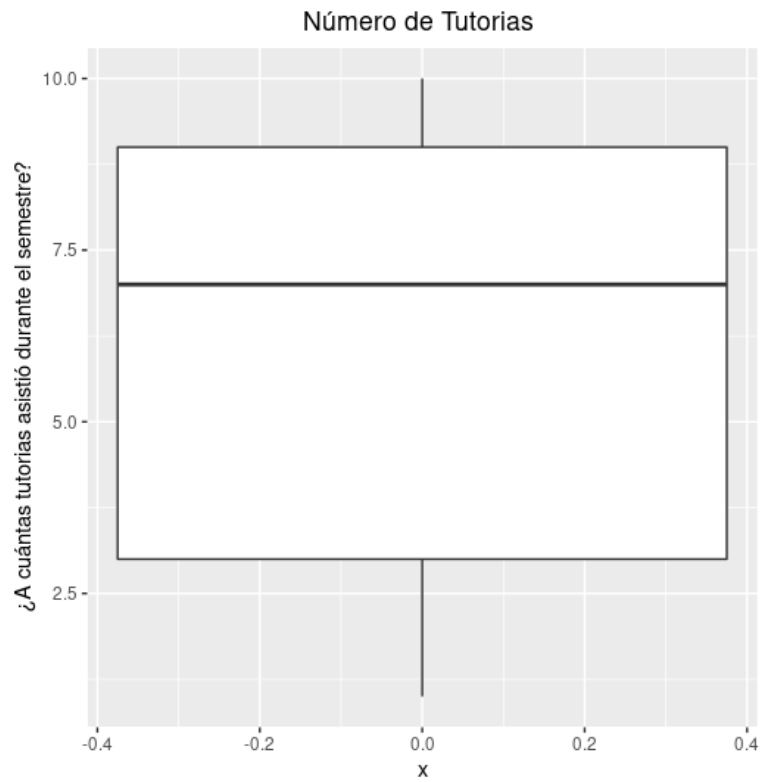
la distribución de los datos parece no tener datos atípicos . Además, el CV es menor al 20% la media será una buena medida descriptiva de la centralidad del tiempo de estudio semanal de los estudiantes universitarios.



Número de tutorías



Se observa dos focos de acumulación en la cantidad de tiempo dedicado al estudio por los estudiantes .

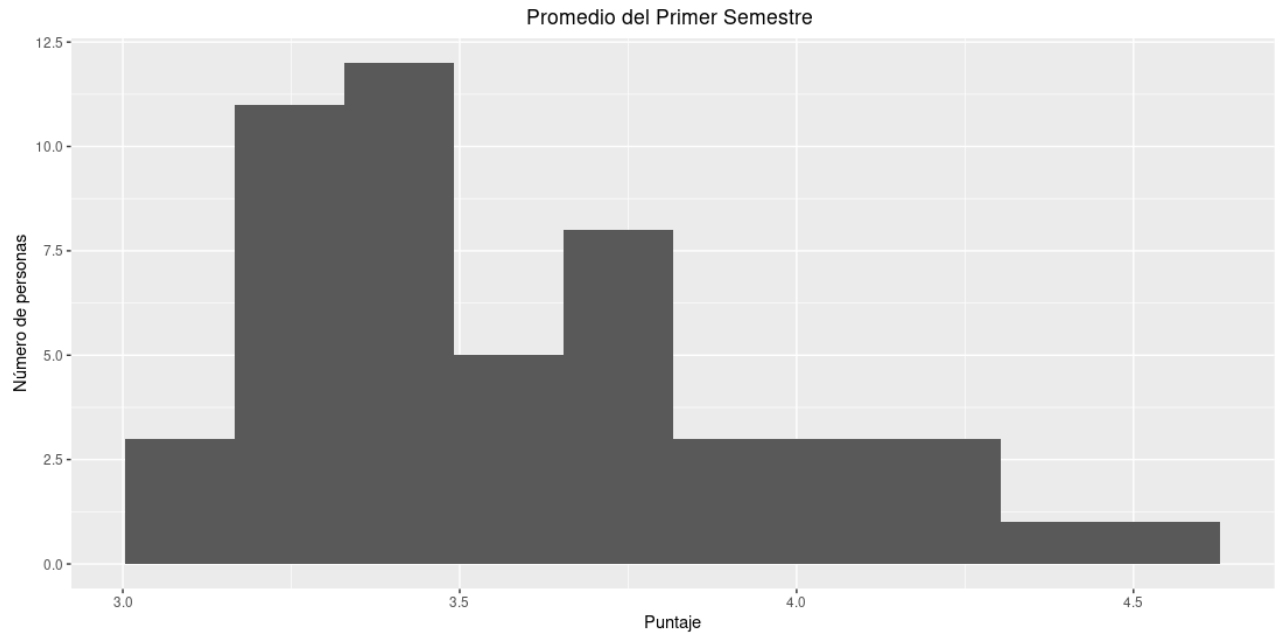


Descripción numérica  
para la variable del número  
de horas de estudio

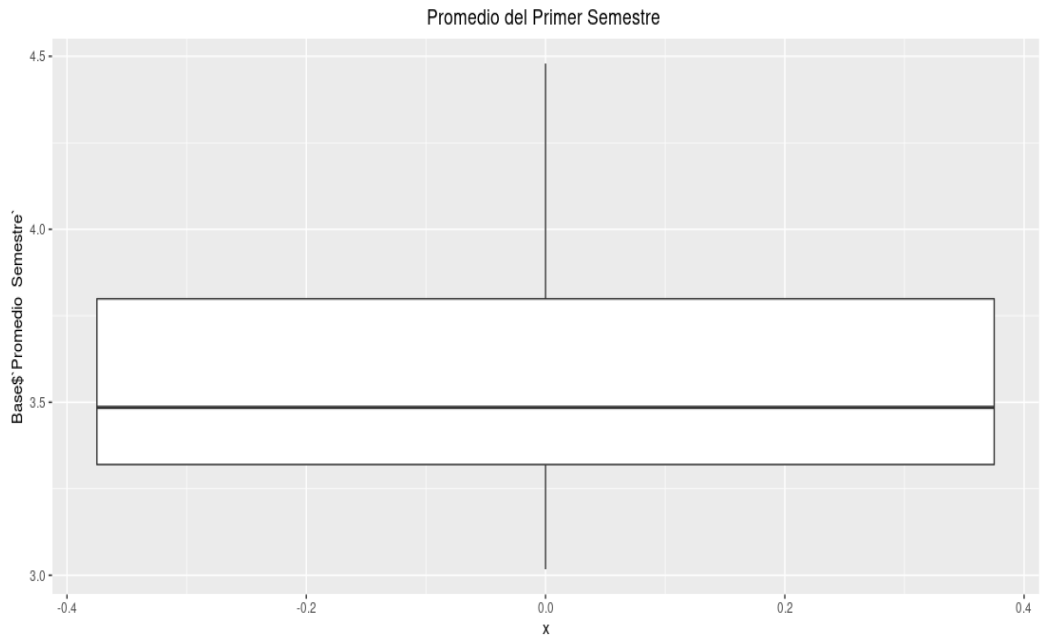
	Datos
Promedio	5.94
Mediana	7.00
Varianza	10.67
SD	3.27
CV	54.99
Q1	3.00
Q2	7.00
Q3	9.00
Minimo	1.00
Maximo	10.00

La distribución de los datos parece no tener datos atípicos .Además CV es mayor al 20% la media será tan buena medida descriptiva de la centralidad del número de tutorías que toman los estudiantes universitarios.

### Promedio del Primer Semestre



Es observable que el puntaje tiende a acercarse a la media haciendo las colas menos pesadas .



Descripción numérica para el promedio del primer semestre

	Datos
Promedio	3.66
Mediana	3.50
Varianza	0.19
SD	0.43
CV	11.78
Q1	3.31
Q2	3.50
Q3	3.92
Minimo	3.00
Maximo	4.90

la distribución parece ser apuntada un punto de acumulació y no posee valores atípicos.ademas, CV es menor al 20% la media será una buena medida descriptiva del promedio del primer semestre de los estudiantes universitarios cabe notar la poca variación de los datos frente a la media .

## Intervalo de confianza para Tiempo de estudio semanal (horas) :

- mínimo 30 horas y máximo 55 horas[razón]

```
[1] 39.26826 43.85174
attr(,"conf.level")
[1] 0.95
```

- Interpretación: al extraer 100 veces la muestra cada una con 50 individuos , se espera que aproximadamente 95 de estas muestras tendrán una media para el tiempo de estudio semanal entre 39.26 y 43.85 horas cabe aclarar que la distribución del tiempo semanal resulto aproximadamente uniforme .

## Intervalo de confianza para la comparación de medias

el equipo en psicología desea saber si hay un diferencia significativa entre el promedio de las personas que si realizaron actividades curriculares y las que no .

Primero creamos un intervalo de confianza para ver si las varianzas son iguales

```
95 percent confidence interval:
0.543101 2.868198
```

como el valor de 1 pertenece al intervalo (0.5431 ,2.8681) al 95% de confianza las poblaciones de donde provienen las muestras tienen igual varianza

Ahora construimos el intervalo de confianza para comparación de medias de igual varianza

```
Two Sample t-test

data: mSi$`Promedio Semestre` and mNO$`Promedio Semestre`
t = 0.85653, df = 48, p-value = 0.396
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.1185681 0.2945605
sample estimates:
mean of x mean of y
3.624926 3.536930
```

como se puede observar al realizar 100 veces las muestras con un tamaño de 50 individuos ,en aproximadamente 95 de estas la diferencia entre la media de las personas que si realizan actividades extracurriculares y las que no estará entre -0.1185 y 0.29 .