

# Reto 2

**Autor: Oliver Orley Rodriguez berrocal.**  
**Programa: Estadística.**

**Se trabaja con la base de datos obtenida de las presentaciones de los estudiantes de TAE. La base de datos se ve así:**

nombres	edad	programa	semestre	trabaja1	pasatiempo	sexo	trabaja
juan esteban cendales sora	21	sistemas	8	trascender global	piano y programacion competitiva	H	si
santiago ramirez zapata	22	control	9	no	bailar y hacer deporte	H	no
juan jose hurtado alvarez	26	sistemas	7	no	jugar futbol	H	no
juan pablo ortega medina	21	sistemas	8	ddb latino puerto rico inc trascender global	viajar	H	si
sebastian rendon giraldo	20	sistemas	8	freelancer	jugar videojuegos de historia	H	si
jean paul yepes	22	sistemas	10	no	leer libros	H	no
daniel alexander naranjo rios	21	sistemas	9	no	videojuegos	H	no
julian alejandro usuga ortiz	21	estadística	4	no	deporte y musica	H	no
sebastian lopez restrepo	24	sistemas	9	trascender global	jugar videojuegos	H	si
carlos daniel montoya	22	sistemas	10	experimentality	explorar tecnologias y ver como integrarlos para resolver problemas	H	si

completitud:

En la siguiente tabla podemos observar el resumen de las varibles numericas, que son el semestre y la edad, ademas se le anade las desviaciones estandares.

edad	semestre
Min. :18.00	Min. : 3.000
1st Qu.:21.00	1st Qu.: 8.000
Median :21.00	Median : 8.000
Mean :22.16	Mean : 7.971
3rd Qu.:23.00	3rd Qu.: 9.000
Max. :33.00	Max. :16.000
sd :2.305	sd :2.086

Ya que las variables son numricas podemos observar sus freuencias absolutas y relativas.

	semestre	Record Count ▾	Record Count
1.	8	31	44,93 %
2.	9	10	14,49 %
3.	7	7	10,14 %
4.	10	7	10,14 %
5.	3	4	5,8 %
6.	6	3	4,35 %
7.	4	2	2,9 %
8.	11	2	2,9 %
9.	16	1	1,45 %
10.	5	1	1,45 %
11.	12	1	1,45 %
1 - 11 / 11 < >			

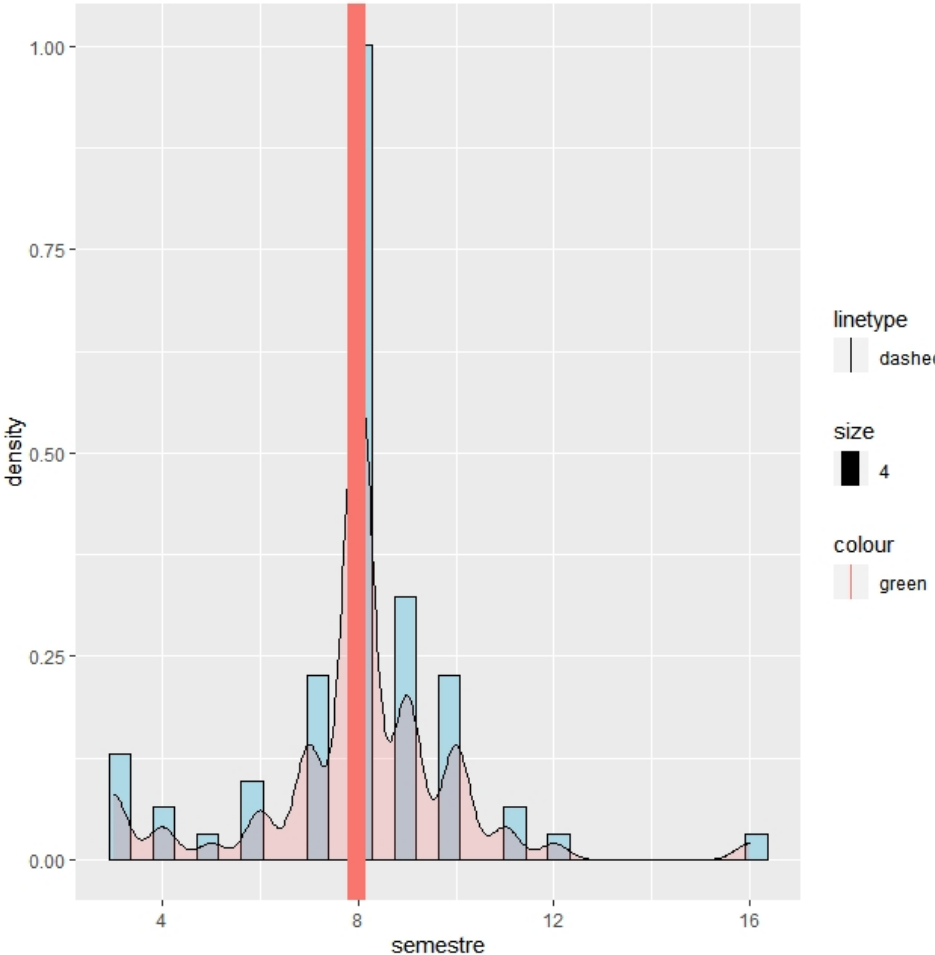
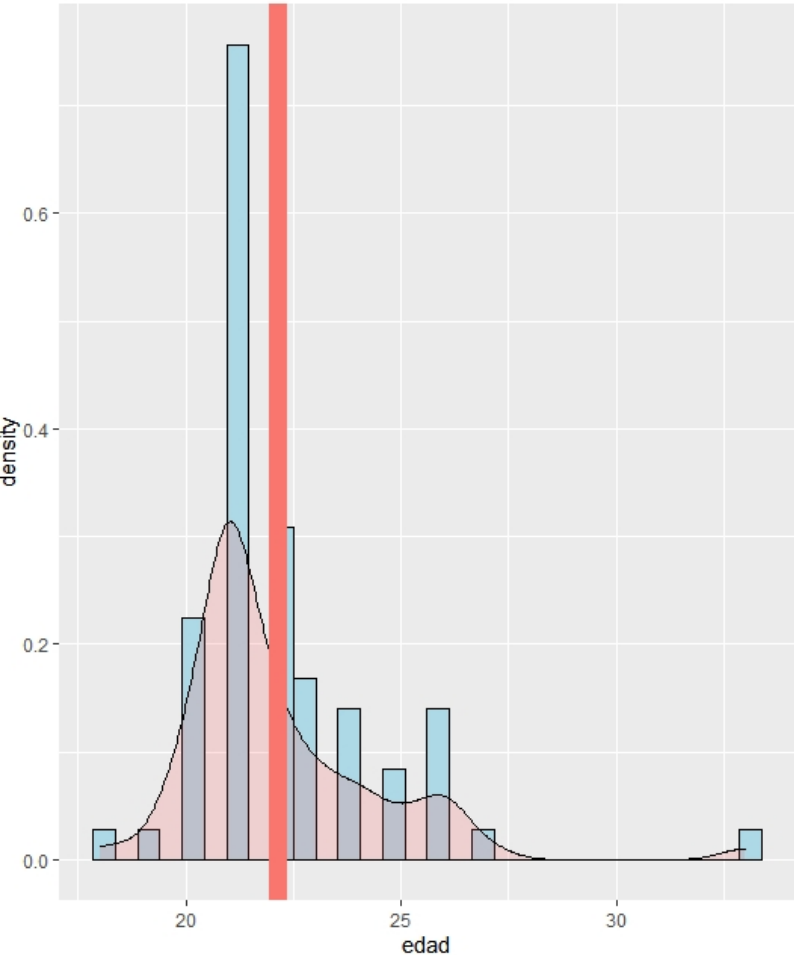
	edad	Record Count ▾	Record Count
1.	21	27	39,13 %
2.	22	11	15,94 %
3.	20	8	11,59 %
4.	23	6	8,7 %
5.	26	5	7,25 %
6.	24	5	7,25 %
7.	25	3	4,35 %
8.	19	1	1,45 %
9.	27	1	1,45 %
10.	18	1	1,45 %
11.	33	1	1,45 %
1 - 11 / 11 < >			

Aquí tenemos las medidas de escala. Podemos ver que el RIQ es pequeño para ambos casos y esto da una idea de que en este porcentaje, la mayoría de la edad de los estudiantes difiere en 2 años y que la mayoría de los estudiantes difieren en 1 semestre:

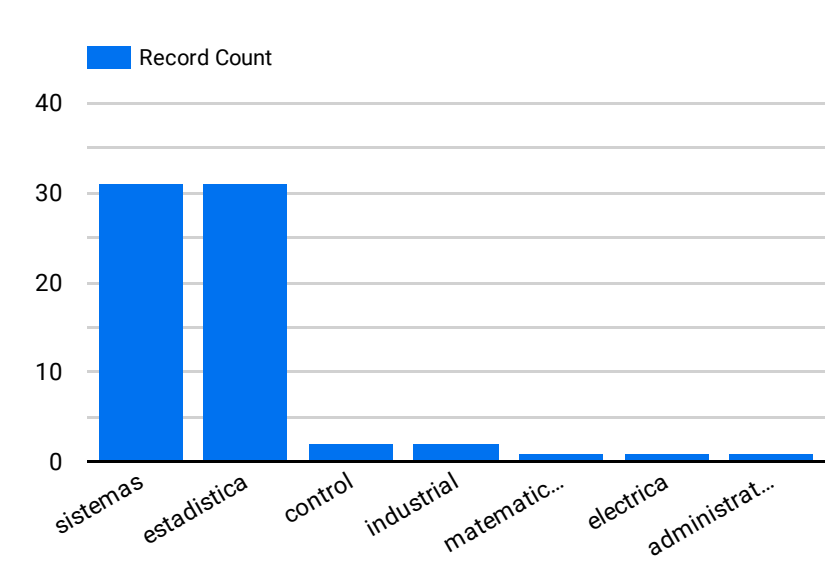
	rango	rango.intercuartil
edad	15	2
semestre	13	1

En el histograma de la edad podemos ver sus frecuencias relativas y su grafica de desnsidad, vemos que la edad más comun es 21 años, tambien se observa que el promedio es aproximadamente 22, tambien tienen algunos datos atípicos como 18 y mayor 30.

El histograma de los semestres en frecuencia relativa con su grafica de densidad nos muestra que el semetre más frecuente es el 8, y que la media es muy cercana a el octavo semestre.

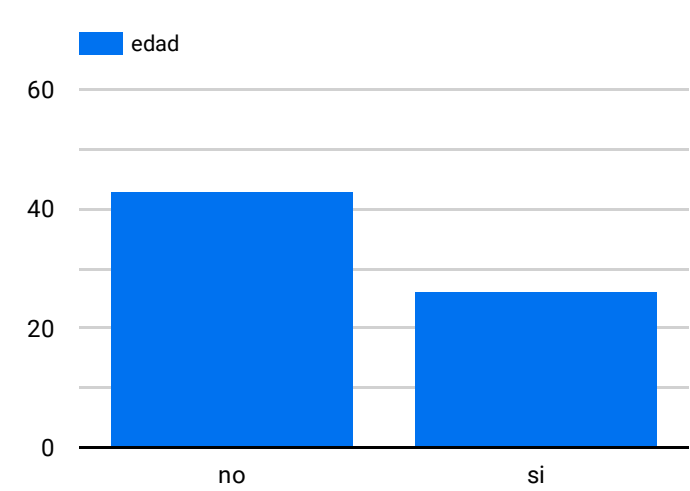


primero obsevamos la cantidad de estudiantes por carrera, en sus frecuencias absulas y con sus frecuencias relativas, donde ingenieria de sistemas y estadistica son las mas frecuentes, siendo estadistica por un estudiante de doble titulacion la carrera mas frecuente en este curso:



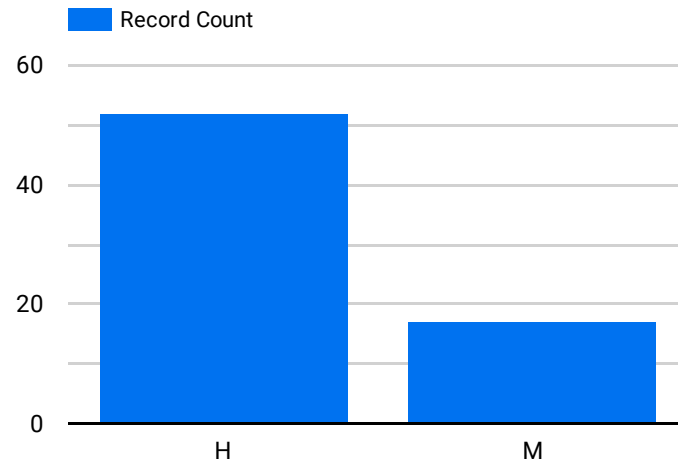
	programa	Record Count ▾	Record ...
1.	sistemas	31	44,93 %
2.	estadistica	31	44,93 %
3.	control	2	2,9 %
4.	industrial	2	2,9 %
5.	matematicas	1	1,45 %
6.	electronica	1	1,45 %
7.	administrativ...	1	1,45 %

Veamos la cantidad de estudiantes que trabajan y los que no en sus frecuencias absulas y una tabla con sus frecuencias relativas, donde observamos que aproximadamente 62% no trabajan:



	trabaja	edad ▾	Record ...
1.	no	43	62,32 %
2.	si	26	37,68 %

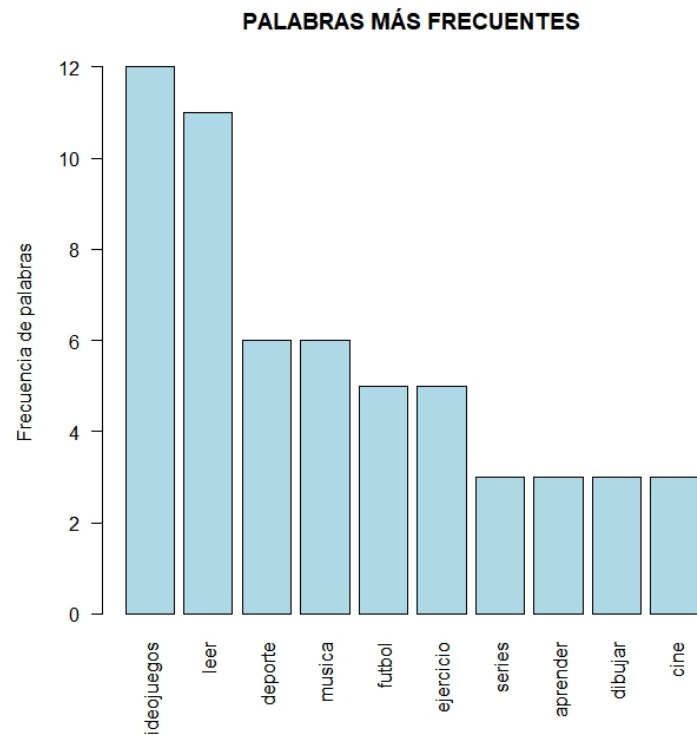
Ahora observamos la cantidad de estudiantes discriminados por sexo en sus frecuencias absolutas y una tabla con sus frecuencias relativas, aproximadamente el 75% de los estudiantes son hombres:



sexo		Record Count ▾	Record ...
1.	H	52	75,36 %
2.	M	17	24,64 %

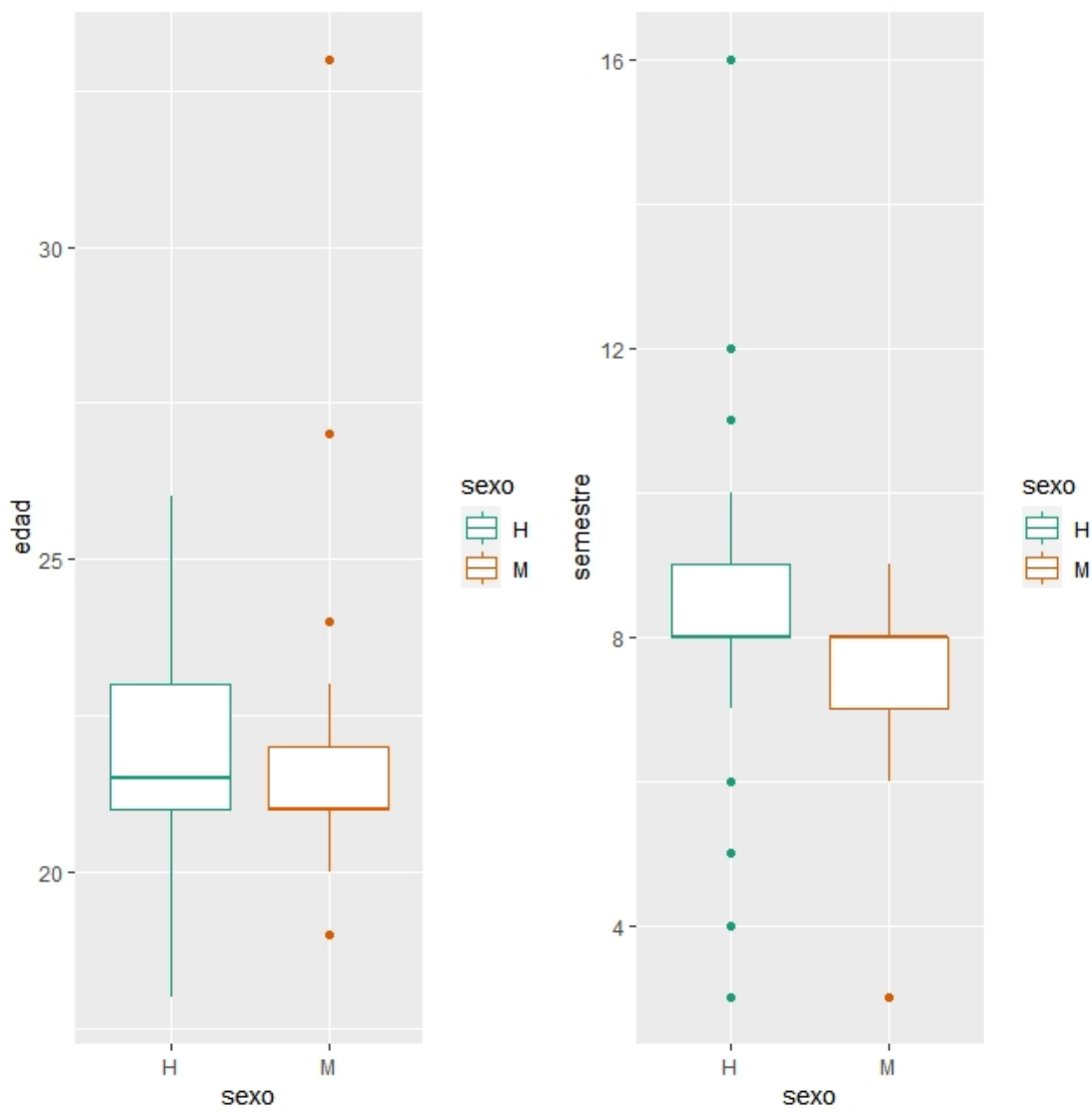
1 - 2 / 2    <    >

En estas graficas se puede obserar los pasatiempos más comunes, siento muy comun videojuegos sobre todo, luego el leer, deporte etc:

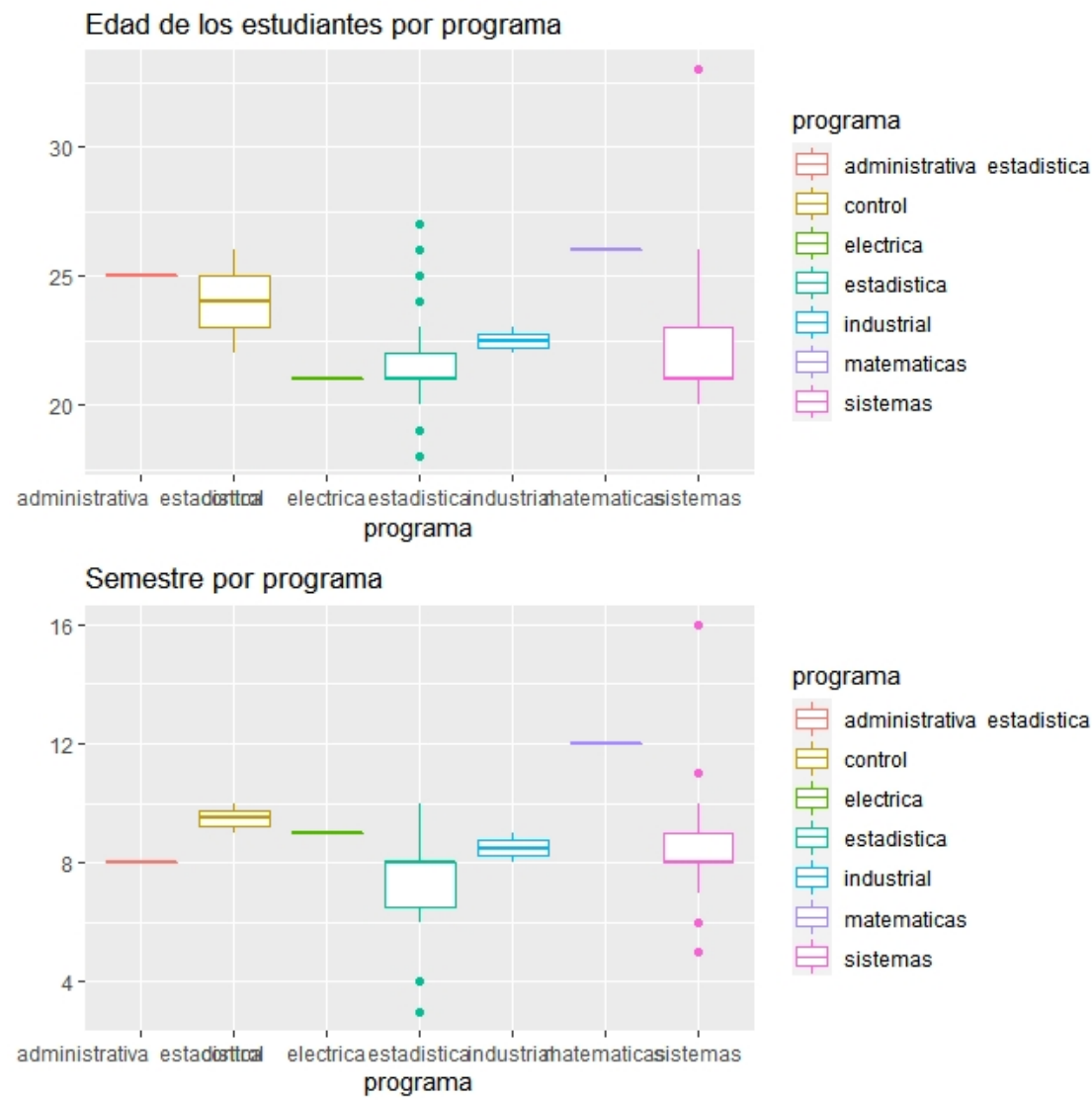


# Relaciones multivariadas:

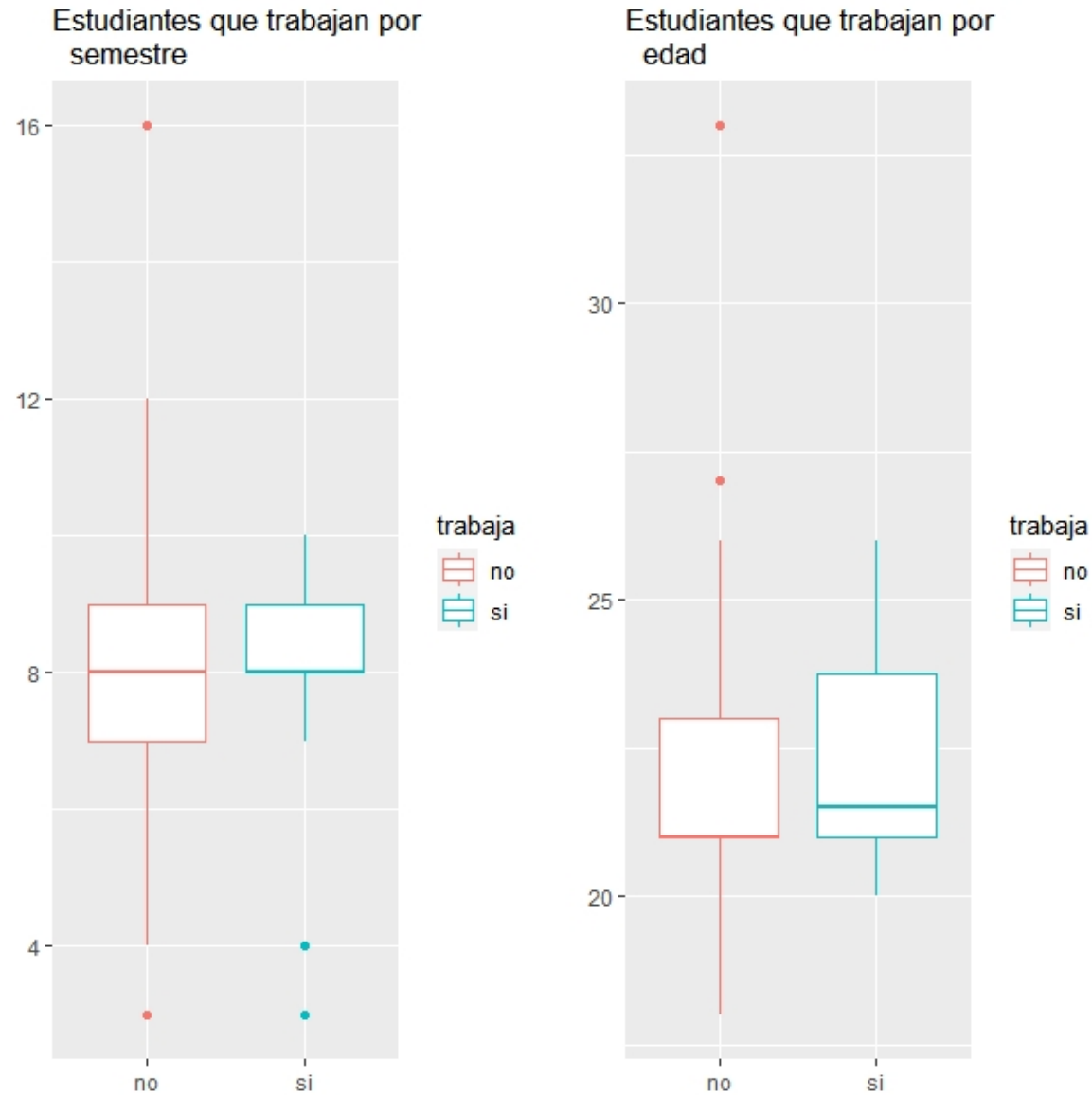
Se observa en la siguientes graficas que la edad de las mujeres es mas variable que la de los hombres, mientras que los hombres muestran mayor variabilidad en el semestre cursado.



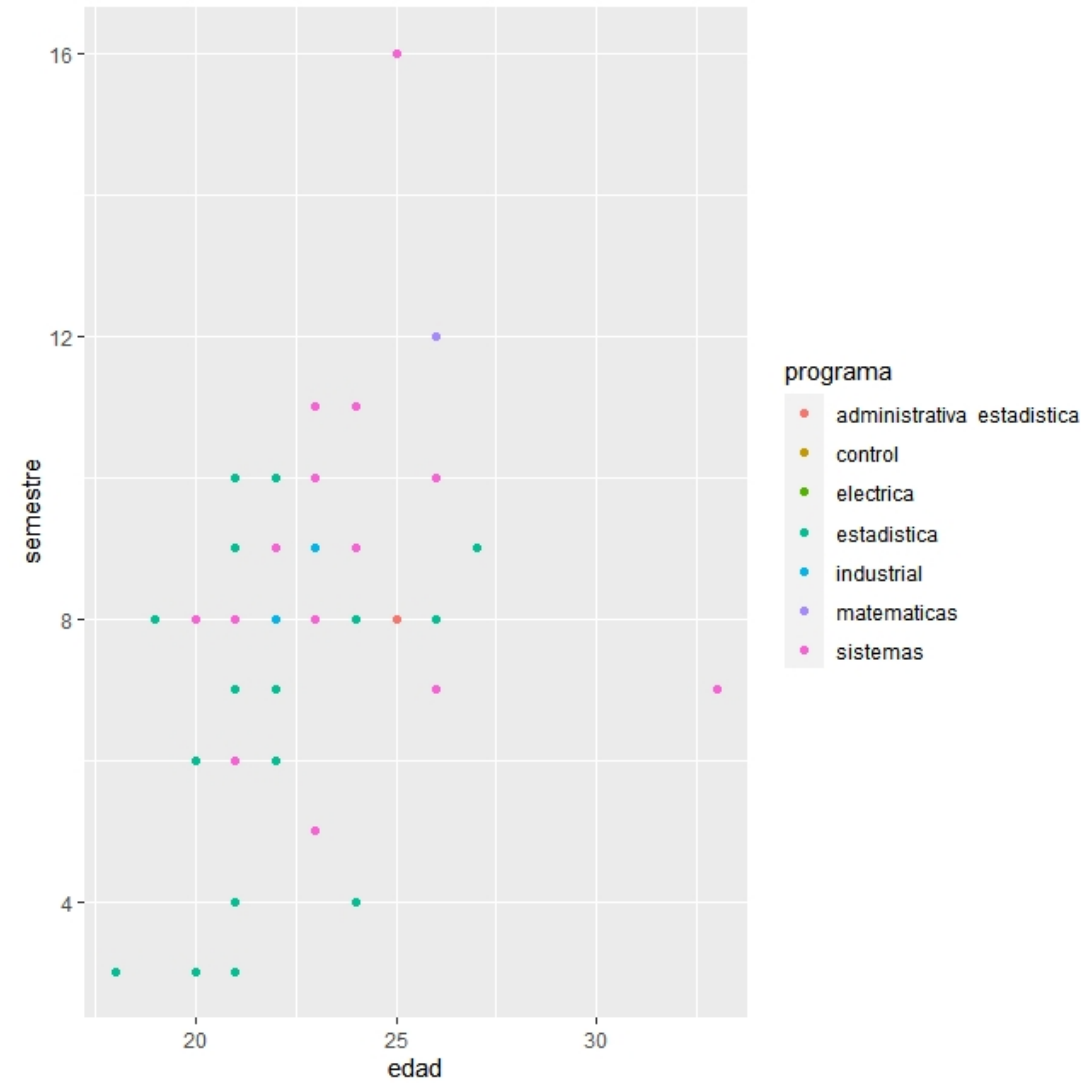
Se tienen pocos datos de algunos cursos pero podemos observar que estadsticamente se observan diferencias significativas entre ingenieria de control y estadistica en cuanto a la edad, podemos ver que la persona de mayor edad es de ingenieria de sistemas y que los estudiantes de estadistica tienen mayor variabilidad y son tambien los más jóvenes y son los de menor semestre.



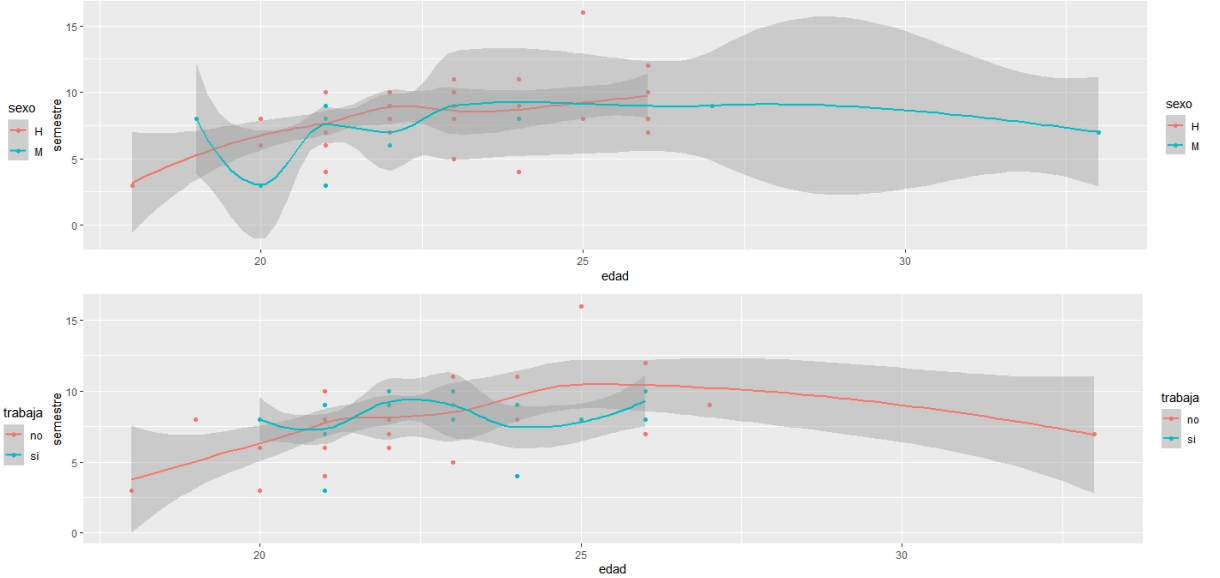
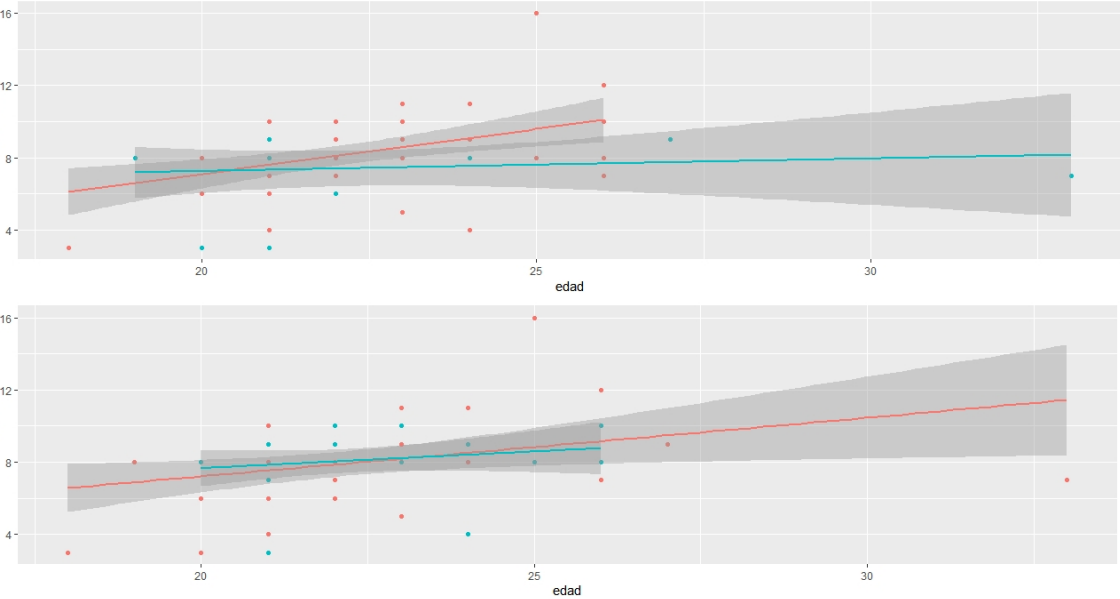
En las siguientes graficas se observa que la pesonas que no trabajan tienen mas dispercion de en cuanto a sus semestres y tambien son mas dispersos en cuanto a su edad.



En esta grafica se observa que las observaciones atipicas grandes hacen parte del programa de ingenieria de sistemas y que las atipicas pequenas son de estadisticas.



Acontinuación se observa graficas con regresion lineal y Loess discriminados por sexo y por trabajo, graficamente se puede pensar que el modelo loess tiene una curva más suave que capta mejor el comportamiento de los datos, para aclarar esto analiticamente seria conveniente compararlos con medidas numericas.



Para observar un grado de relación lineal entre las variables edad y semestre se calcula su correlacion y se obtien una realcion aprox. del 31%. Tambien se plantea un modelo que involucre estas variables y se aprecia que el semestre es una variable que es significativa para explicar la edad y que el semetre puede explicar aproximadamente el 10% de la variabilidad de las edades, aunque es poco, podria postularse posteriormente un mejor modelo.

# Formulación e hipotesis:

- Si consideramos que nuestros datos son prospectivos podemos hacernos las siguientes hipotesis:
- \*hipotesis 1: la poblacion de estudiantes de TAE son mayormente hombres?\*
  - \*hipotesis 2: la mediana de los esetudiantes son de 8avo semestre?\*
  - \*hipotesis 3: videojuegos y leer son preferidos igualmente?\*
  - \*hipotesis 4: La edad promedio de los estudiantes esta entre 22 y 23 años?\*
  - \*hipotesis 5: Los estudiantes que no trabajan son mas que los que si?\*

```
Call:
lm(formula = df$edad ~ df$semestre)

Residuals:
    Min       1Q   Median       3Q      Max
-3.1696 -1.1696 -0.8171  0.5354 11.1829

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  19.3497     1.0535  18.366  < 2e-16 ***
df$semestre   0.3525     0.1279   2.756  0.00754 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.201 on 67 degrees of freedom
Multiple R-squared:  0.1018,    Adjusted R-squared:  0.08839
F-statistic: 7.593 on 1 and 67 DF,  p-value: 0.007539
```