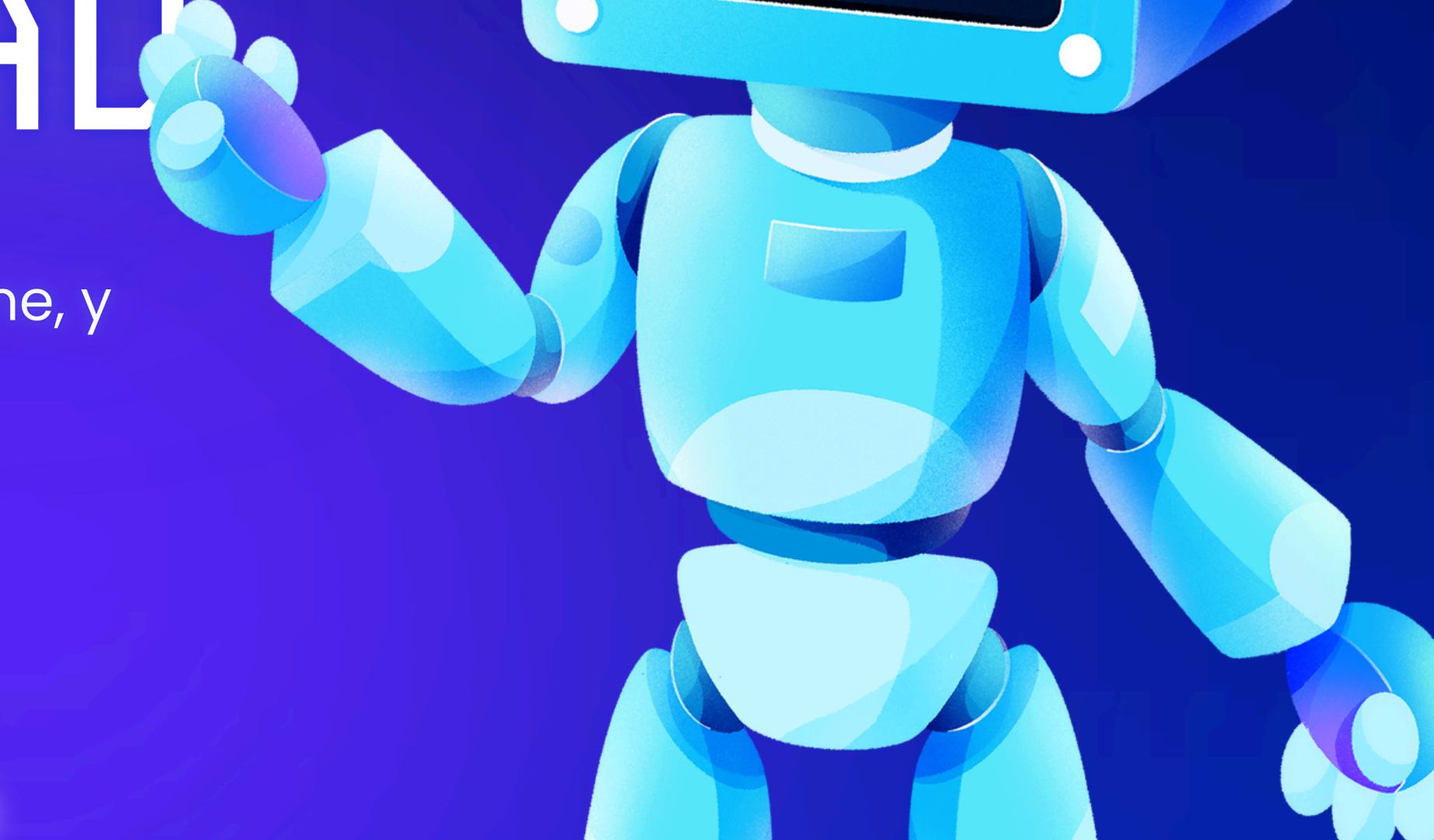


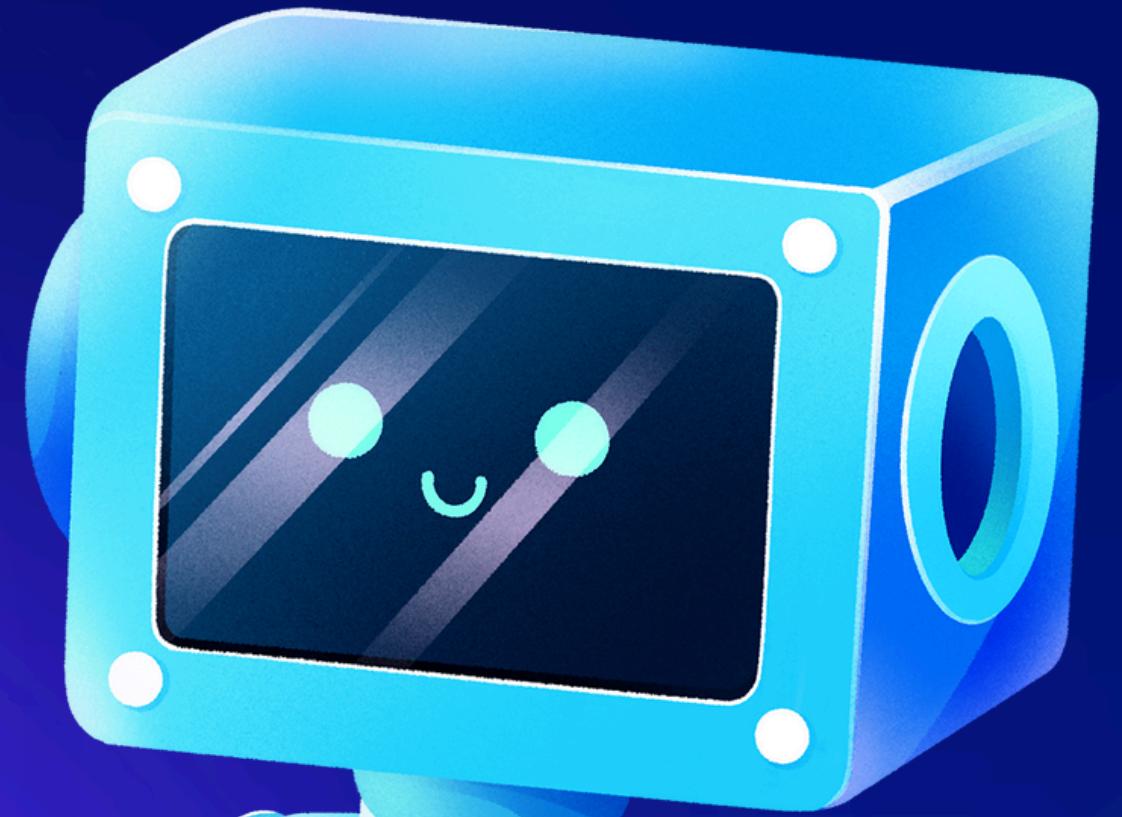
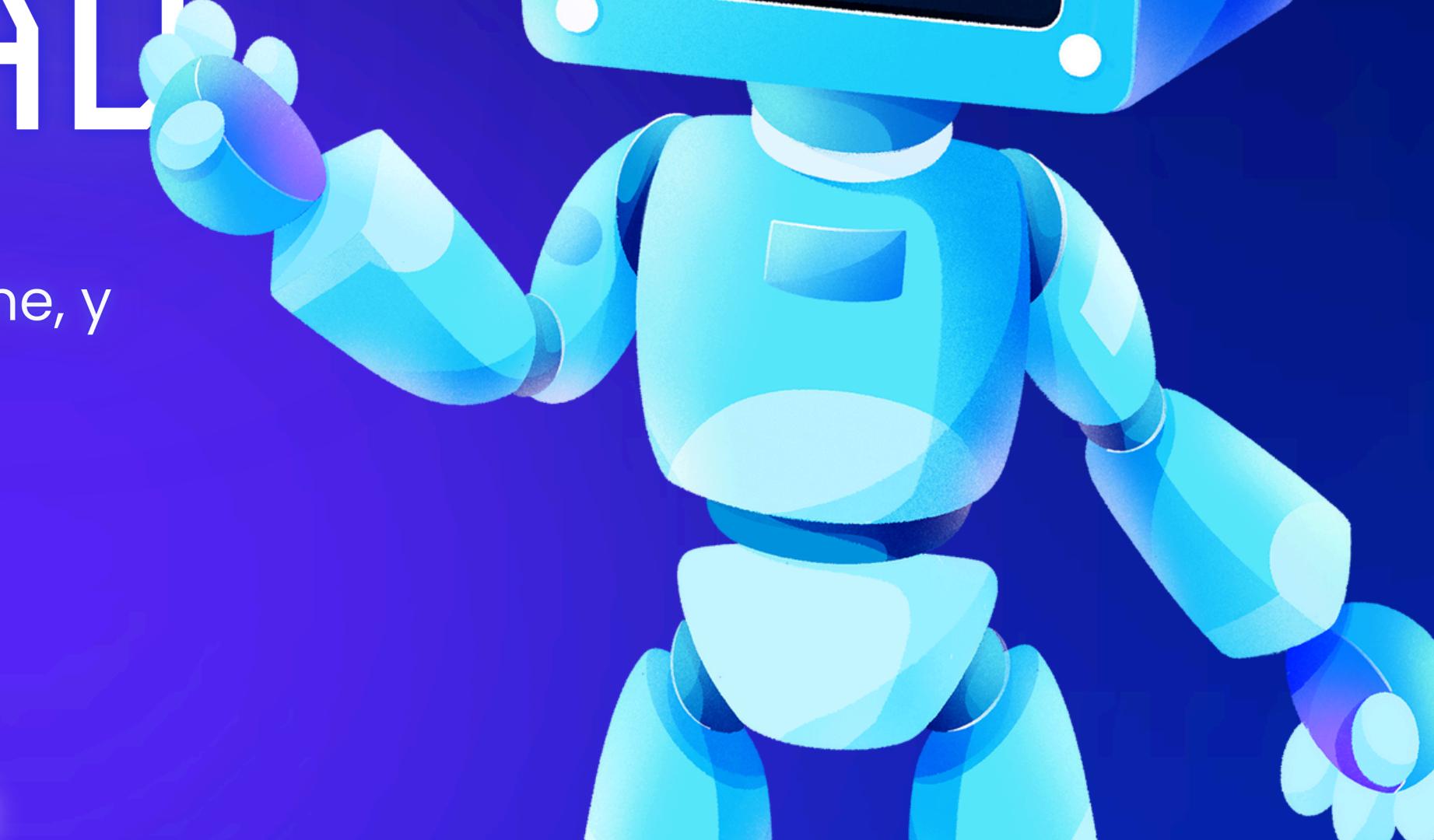


REGRESIÓN LINEAL MÚLTIPLE

FELICIDAD



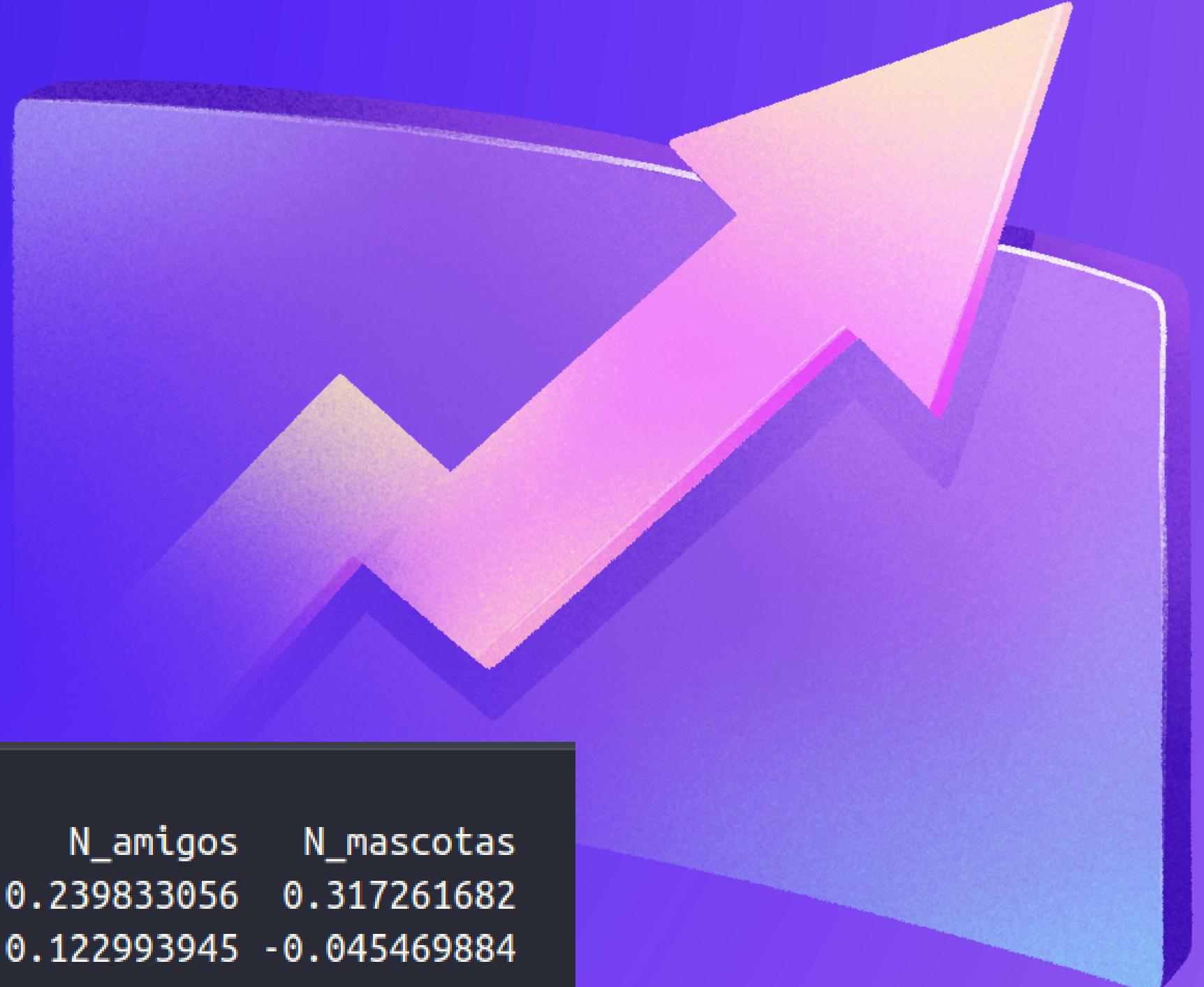
Gonzalo Arévalo, Sebastián Cassone, y
Nicolás Sánchez.



SOLUCIÓN AL PROBLEMA

- Se obtienen los 70 casos a través de la semilla 25 (Edad + mes del integrante del grupo más joven)
- Las variables categóricas como Sexo se reemplaza por 1 si es Hombre y 0 si es Mujer.
- Se obtuvo la matriz de correlaciones.

	Felicidad	Edad	Sexo	N_amigos	N_mascotas
Felicidad	1.00000000	-0.17242563	-0.03526412	0.239833056	0.317261682
Edad	-0.17242563	1.00000000	0.09184183	-0.122993945	-0.045469884
Sexo	-0.03526412	0.09184183	1.00000000	-0.070397889	0.246070505
N_amigos	0.23983306	-0.12299395	-0.07039789	1.000000000	-0.001237681
N_mascotas	0.31726168	-0.04546988	0.24607051	-0.001237681	1.000000000



Comparamos los cuatro modelos

Analysis of Variance Table

```
Model 1: Felicidad ~ 1
Model 2: Felicidad ~ N_amigos
Model 3: Felicidad ~ N_amigos + N_mascotas
Model 4: Felicidad ~ N_amigos + N_mascotas + Edad
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	69	173.49			
2	68	163.51	1	9.9789	4.6024 0.035611 *
3	67	146.01	1	17.4949	8.0689 0.005982 **
4	66	143.10	1	2.9121	1.3431 0.250665

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

- Se comienza a realizar una selección hacia Adelante sin antes comenzar a verificar las condiciones para RLM.
- Al comenzar la selección hacia adelante se comienza desde el modelo nulo hasta agregar Edad.

La primera condición se cumple puesto que las variables predictoras de números de amigos y mascotas son de tipo cuantitativo
La variable de respuesta es la felicidad es cuantitativa y debido a que los valores discretos son un subconjunto de las continuas entonces cumple esta condición

varianza de amigos: 5.649896

varianza de mascotas: 1.267909

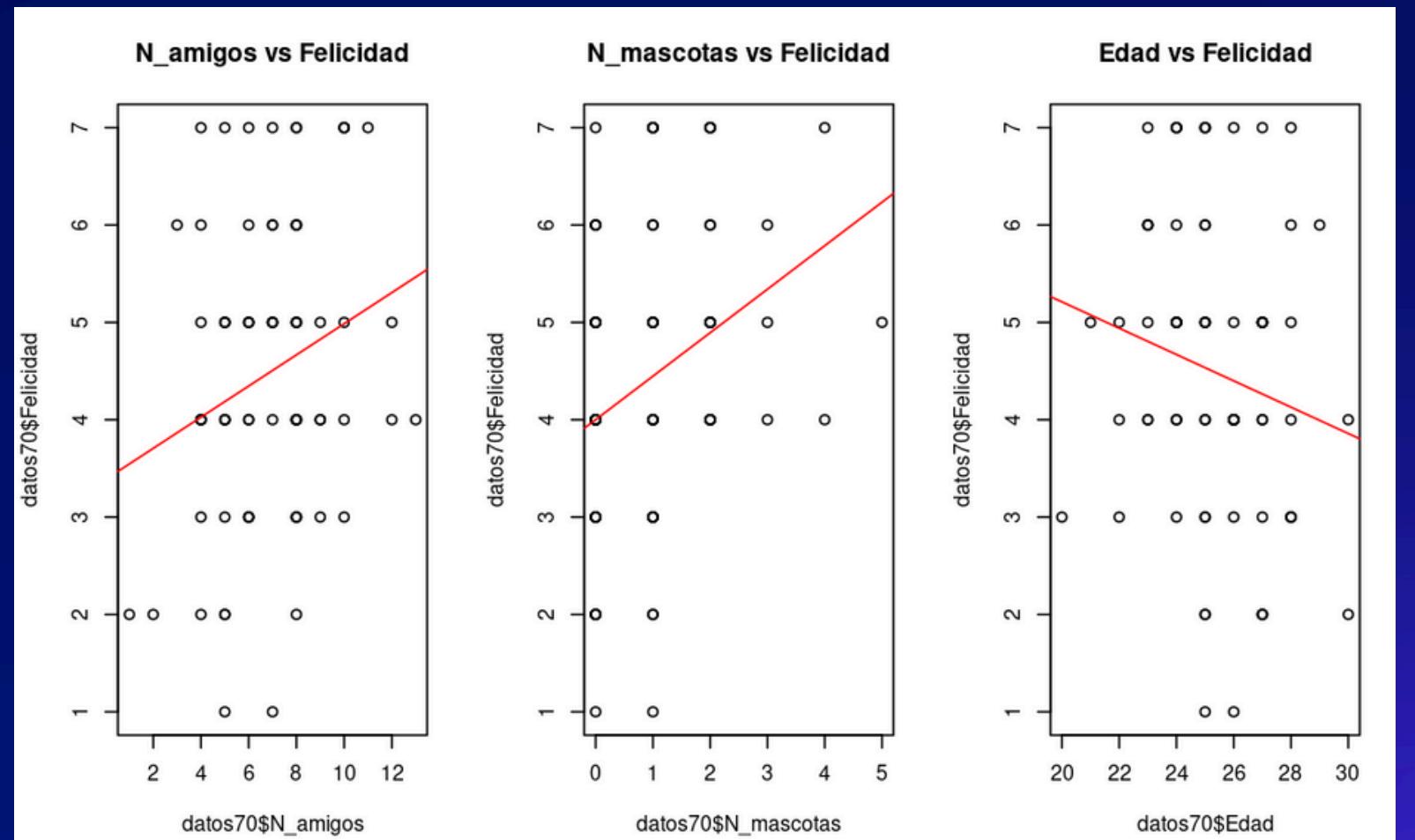
se puede ver anteriormente que las variables predictoras no tienen varianza cero, y por ende se cumple la tercera condición debido a los cálculos de correlación hechos para observar las variables predictoras y de control, se puede observar que para las variables predictoras los valores varian entre 0.2 y 0.35, por lo que se tiene un coeficiente de relación débil y uno moderado pero ninguno fuerte y se cumple la cuarta condición

studentized Breusch-Pagan test

data: modeloK

BP = 0.17485, df = 2, p-value = 0.9163

debido a que el p-value de la prueba bptest tiene un valor de 0.9163 y este es superior a 0.05, entonces se puede decir que no existe suficiente evidencia para rechazar la hipótesis de homocedasticidad por lo que se cumple la quinta condición



- Se termina de verificar las condiciones del modelo que se escogió.

6ta condición Los residuos deben seguir una distribución cercana a la normal centrada en cero.

```
35 63
64 34
```

Shapiro-Wilk normality test

```
data: modeloJ2$residuals
W = 0.98157, p-value = 0.3919
```

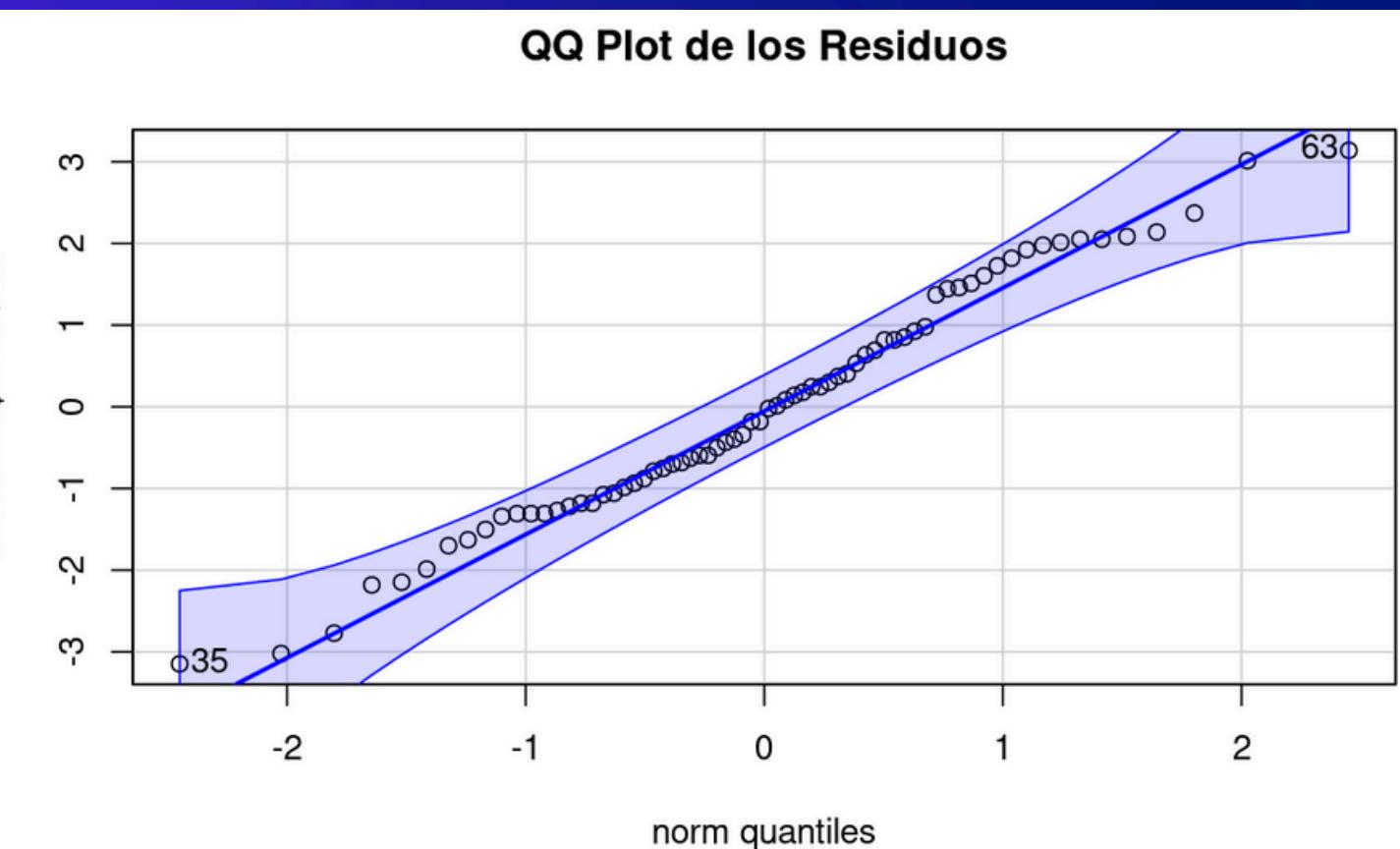
Dado que el p-valor de la prueba de Shapiro-Wilk es 0.3919, que es mayor que 0.05, no rechazamos la hipótesis nula de que los residuos siguen una distribución normal. Por lo tanto, podemos concluir que la 6ta condición de normalidad de los residuos se cumple.

7ma condición Los valores de la variable de respuesta son independientes entre sí. (utilizando la prueba de Durbin-Watson)

```
lag Autocorrelation D-W Statistic p-value
 1      0.1047277    1.781757    0.37
Alternative hypothesis: rho != 0
```

Interpretación: El valor de 1.781757, que está cerca de 2, indica que no hay evidencia significativa de autocorrelación de primer orden en los residuos. Un p-valor de 0.318 sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay autocorrelación.

8va condición Cada predictor se relaciona linealmente con la variable de respuesta.



```
111  
112 Sobrantes <- setdiff(1:filas,aleatorio)  
113 datosSobrantes <- datos[Sobrantes, ]  
114  
115 predJ2 <- predict(modeloJ2, datosSobrantes)  
116 print(cbind(datosSobrantes, predJ2))  
117  
118 #el error cuadratico medio es el siguiente  
119 mediaJ2 <- mean((datosSobrantes$Felicidad - predJ2)^2)  
120 print(mediaJ2)  
121  
122  
123 ...
```

[1] 1.812678

The screenshot shows the RStudio interface. On the left is the code editor with the R code listed above. To the right is the Data View pane displaying a data frame named 'data.frame' with 30 rows and 6 columns: Felicidad, Ed., Sexo, Habilidades, N_marcaciones, and predJ2. The data includes various numerical values. Below the Data View is the R Console pane, which shows the output of the last line of code, '[1] 1.812678'.

- Se termina de verificar las condiciones del modelo que se escogió.
- Se termina de verificar el modelo siendo capaz de predecir correctamente datos nuevos.