

Trabajo regresión lineal múltiple

Estudiantes

Rojas Martinez, Ivan Santiago
Hernandez Ruiz, Juan Sebastian
Londoño Montoya, Wilson Duván
Perez Garcia, Pablo

Docente

Isabel Cristina Ramirez Guevara

Asignatura

Análisis de Regresión



Sede Medellín
Enero de 2022

Índice

1. Base de datos	1
1.1. Breve Descripción de los Datos contextualizando el problema y explicando cada una de las variables involucradas en el modelo.	1
2. Análisis descriptivo	2
2.1. Grafico de dispersión con Matriz de Correlaciones y conclusiones	2
3. Modelo Ajustado de Regresion Lineal múltiple (MRLM)	3
3.1. Tabla de parámetros ajustados	3
3.2. Ecuación Ajustada	3
3.3. Tabla ANOVA	3
3.4. Prueba de significancia del Modelo	3
3.5. Coeficiente de determinación R^2 : proporción de la variabilidad total de la respuesta explicada por el modelo y opiniones al respecto	4
4. Coeficientes de regresión estandarizados	4
4.1. Tabla de coeficientes estandarizados	4
5. Significancia individual de los parámetros del modelo	5
5.1. Tabla de la significancia individual de los parámetros	5
5.2. Pruebas de hipótesis	5
6. Sumas de cuadrados extras	6
6.1. Prueba de hipótesis	6
6.2. Modelo completo y reducido	6
6.3. Estadístico de prueba	7
6.4. Tabla del Test lineal general	7
7. Sumas de cuadrados tipo I y tipo II	7
7.1. Sumas de cuadrados secuenciales	7
7.2. Tabla anova	7
7.3. Sumas de cuadrados parciales	8
7.4. Tabla Anova	8

8. Residuales estudentizados vs. Valores ajustados	8
8.1. Gráfico de los residuales estudentizados vs. Valores ajustados	9
9. Gráfico q-norm residuales estudentizados	10
9.1. Pruebas de hipótesis	10
10. Diagnostico sobre la presencia de observaciones atípicas , de balanceo y/o influenciales y conclusiones	10
11. Ejercicio 11	18
12. Ejercicio 12	20
12.1. Matriz de correlación de las variables predictoras	20
12.2. VIF's	21
12.3. Proporciones de varianza	21
13. Ejercicio 13	21
13.1. Selección según el R_{adj}^2	23
13.2. Selección según el estadístico C_p	24
13.3. Stepwise	25
13.4. Selección hacia adelante o forward	32
13.5. Selección hacia atrás o backward	37
14. Selección del modelo	39

Índice de figuras

Índice de cuadros

2. Resumen de los coeficientes	3
3. Resumen de los coeficientes	5
4. Valores ajustados VS Residuales Estudentizados	11

Se realizará una análisis de regresión lineal múltiple(RLM):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Con la intención de validar si dicho modelo es adecuado para explicar la posibilidad de ser admitido a una carrera de postgrado en la india teniendo en cuenta determinadas pruebas de aptitud.

1. Base de datos

1.1. Breve Descripción de los Datos contextualizando el problema y explicando cada una de las variables involucradas en el modelo.

La base de datos disponible en Kaggle corresponde a puntajes de admision creados para la predicción de las admisiones de posgrado en La India. Cuenta con 400 observaciones y 9 variables. De las cuales se consideran los primeros 100 estudiantes y 6 variables de interes por indicación de la docente.

Variables	Descripción
Chance.of.Admit:	Posibilidad de ser admitido. Variable numérica continua de 0-1
GRE Score:	Puntaje de Examen que proporciona a las

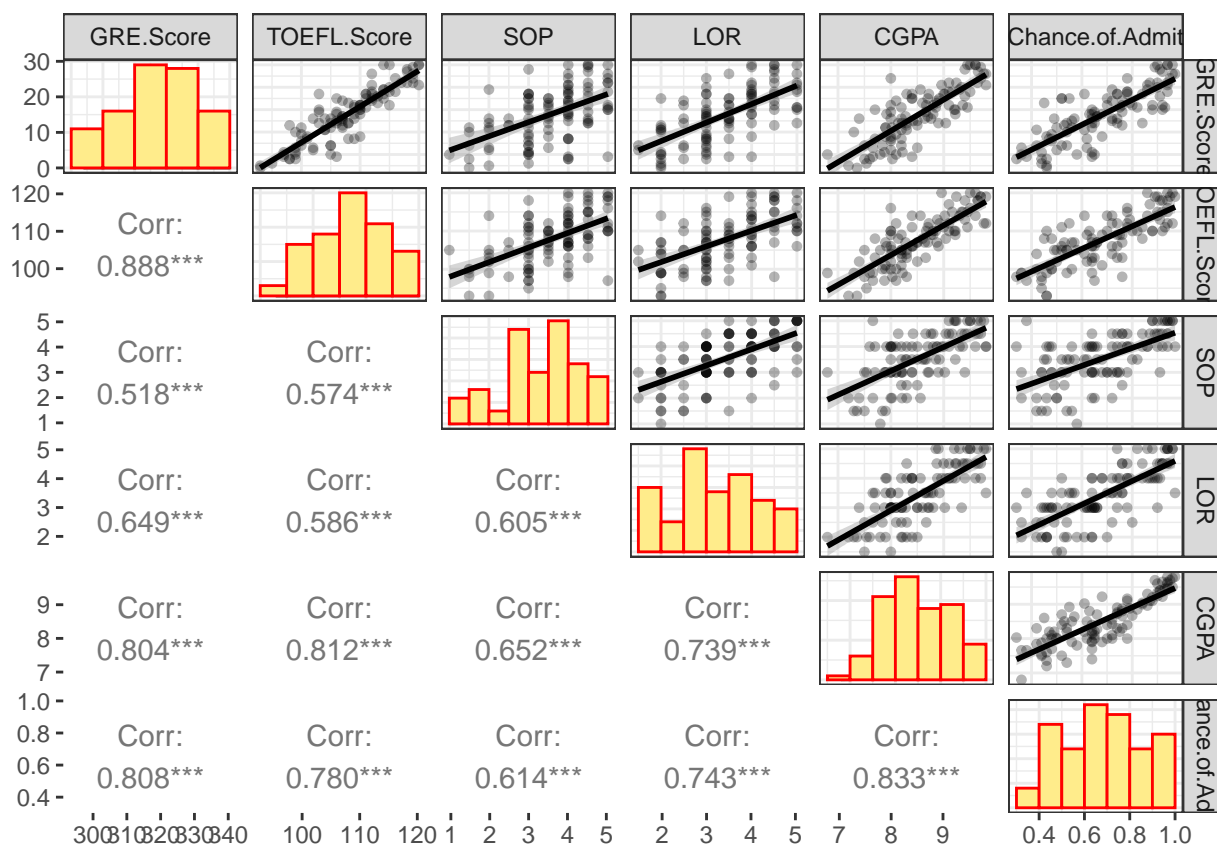
escuelas una medida común para la comparación de la capacidad de razonamiento verbal, razonamiento cuantitativo, y habilidades para pensar y escribir de forma analítica. Variable numérica que toma valores de 294 - 340.| **TOEFL Score:** |Puntaje en prueba estandarizada de dominio del idioma inglés. Variable numérica que toma valores del 93 - 120| **SOP:** |Puntaje en Ensayo de admisión o solicitud de postgrado. Variable numérica que toma valores del 1 - 5, tomando el valor medio entre cada par de enteros en el intervalo.| **LOR:** |Puntaje en Carta de recomendación. Variable numérica que toma valores del 1.5 - 5, tomando el valor medio entre cada par de enteros en el intervalo.| **CGPA:** |Promedio general acumulado en el pregrado. Variable numérica que toma valores del 6.8 - 9.8.|

Renombrando las variables:

- **GRE Score** = X_1
- **TOEFL Score** = X_2
- **SOP** = X_3
- **LOR** = X_4
- **CGPA** = X_5

2. Análisis descriptivo

2.1. Grafico de dispersión con Matriz de Correlaciones y conclusiones



-Se observan relaciones de interés.

-La variable **Chance.of.Admit** (Posibilidad de ser admitido) se encuentran altamente correlacionada con las variables **GRE.Score**, **TOEFL.Score**, **SOP**, **LOR** y **CGPA** con correlaciones de **0.808**, **0.780**, **0.614**, **0.743** y **0.833** respectivamente. Con relaciones del tipo lineales positivas.

-La variable **CGPA** (Promedio general acumulado en el pregrado) se encuentran altamente correlaciona con las variables **GRE.Score**, **TOEFL.Score**, **SOP** y **LOR** con correlaciones de **0.804**, **0.812**, **0.652** y **0.739** respectivamente. Con relaciones del tipo lineales positivas. Esto nos puede indicar redundancia en el modelo o multicolinealidad lo cual validaremos más adelante.

-La variable **GRE.Score** se encuentra altamente correlacionadas con las variables **TOEFL.Score** y **CGPA**. Y moderadamente con las variables **SOP** y **LOR**.

3. Modelo Ajustado de Regresion Lineal múltiple (MRLM)

3.1. Tabla de parámetros ajustados

Cuadro 2: Resumen de los coeficientes

	Estimación	Error estándar	T_0	Valor P
β_0	-1.7723	0.3007	-5.8939	0.0000
β_1	0.0041	0.0017	2.4400	0.0166
β_2	0.0029	0.0031	0.9417	0.3488
β_3	0.0120	0.0119	1.0098	0.3152
β_4	0.0428	0.0143	3.0023	0.0034
β_5	0.0757	0.0263	2.8756	0.0050

3.2. Ecuación Ajustada

Con base en la tabla de parámetros estimados se obtiene la ecuación de regresión ajustada:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \cdots + \hat{\beta}_5 X_{i5}, \quad i = 1, 2, \dots, 100$$

$$\hat{Y}_i = -1.7723 + 0.0041X_{i1} + 0.0029X_{i2} - 0.0120X_{i3} + 0.0428X_{i4} + 0.0757X_{i5}, \quad i = 1, 2, \dots, 100$$

3.3. Tabla ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(GRE.Score, TOEFL.Score, SOP, LOR, CGPA)	5	2.3674478	0.4734896	65.99381	
Residuals	94	0.6744272	0.0071748	NA	

Donde F-value = $F_0 = \frac{MSR}{MSE} \sim F_{5,94}$

3.4. Prueba de significancia del Modelo

$$\begin{cases} H_0 : \beta_1 = \cdots = \beta_5 = 0 \\ H_1 : \text{Al menos un } \beta_j \neq 0 \end{cases}$$

Analizando el **p-valor** = **2.2e-16** = **0** de la tabla ANOVA y con una confianza del **95 %** hay evidencia suficiente para rechazar la **hipótesis nula**. Esto quiere decir que el modelo es globalmente significativo y por lo tanto al menos una de las pruebas de aptitud ayuda a explicar la variabilidad de ser admitido a un curso de postgrado.

3.5. Coeficiente de determinación R^2 : proporción de la variabilidad total de la respuesta explicada por el modelo y opiniones al respecto

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$$R^2 = \frac{2.3674478}{2.3674478 + 0.6744272} = 0.7782857$$

El **77.83 %** de la variabilidad de la posibilidad de ser admitido es explicada por la relación con las variables GRE.Score, TOEFL.Score, SOP, LOR y CGPA.

4. Coeficientes de regresión estandarizados

Calcule los coeficientes de regresión estandarizados y concluya acerca de cuál de las variables aporta más a la respuesta según la magnitud en valor absoluto de tales coeficientes (cuidado, no confunda esto con la significancia de los coeficientes de regresión)

4.1. Tabla de coeficientes estandarizados

	Estimación	Limites.2.5..	Limites.97.5..	Vif	Coef.Std
(Intercept)	-1.7722651	-2.3693009	-1.1752294	0.000000	0.000000
GRE.Score	0.0041091	0.0007654	0.0074528	5.691210	0.0495542
TOEFL.Score	0.0029116	-0.0032277	0.0090508	5.858052	0.0194023
SOP	0.0120402	-0.0116343	0.0357148	1.928844	0.0119389
LOR	0.0428307	0.0145058	0.0711556	2.519579	0.0405706
CGPA	0.0757081	0.0234330	0.1279833	4.615227	0.0525903

Gracias a esta tabla, se puede deducir con una diferencia en el valor muy pequeña que, las variables que más aportan según el valor de sus coeficientes estandarizados son **CGPA** y **GRE.Score**

5. Significancia individual de los parámetros del modelo

Pruebe la significancia individual de cada uno de los parámetros del modelo (excepto intercepto), usando la prueba t. Establezca claramente la prueba de hipótesis y el criterio de decisión.

5.1. Tabla de la significancia individual de los parámetros

Cuadro 3: Resumen de los coeficientes

	Estimación	Error estándar	T_0	Valor P
β_0	-1.7723	0.3007	-5.8939	0.0000
β_1	0.0041	0.0017	2.4400	0.0166
β_2	0.0029	0.0031	0.9417	0.3488
β_3	0.0120	0.0119	1.0098	0.3152
β_4	0.0428	0.0143	3.0023	0.0034
β_5	0.0757	0.0263	2.8756	0.0050

De la tabla anterior, se puede observar que a nivel marginal, las variables **GRE** $\text{Score}(\beta_1)$, **LOR** (β_4) y **CGPA** (β_5) son significativas en la respuesta, con un nivel de significancia de $\alpha = 0.05$.

Con el estadístico de prueba $T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t_{94}$

Dicha afirmaciones serán contrastadas con las siguientes pruebas de hipótesis y analizando el **p-valor**.

5.2. Pruebas de hipótesis

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}$$

Analizando el **valor-p** = **0.0166** del parámetro β_1 y con una confianza del **95 %** hay evidencia para rechazar la hipótesis nula. Luego el parámetro β_1 es significativo.

$$\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

Analizando el **valor-p** = **0.3488** del parámetro β_2 y con una confianza del **95 %** no hay evidencia para rechazar la hipótesis nula. Luego el parámetro β_2 no es significativo.

$$\begin{cases} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 \neq 0 \end{cases}$$

Analizando el **valor-p** = **0.3152** del parámetro β_3 y con una confianza del **95 %** no hay evidencia para rechazar la hipótesis nula. Luego el parámetro β_3 no es significativo.

$$\begin{cases} H_0 : \beta_4 = 0 \\ H_1 : \beta_4 \neq 0 \end{cases}$$

Analizando el **valor-p** = **0.3152** del parámetro β_4 y con una confianza del **95 %** no hay evidencia para rechazar la hipótesis nula. Luego el parámetro β_4 no es significativo.

$$\begin{cases} H_0 : \beta_5 = 0 \\ H_1 : \beta_5 \neq 0 \end{cases}$$

Analizando el **valor-p** = **0.0050** del parámetro β_5 y con una confianza del **95 %** hay evidencia para rechazar la hipótesis nula. Luego el parámetro β_5 es significativo.

6. Sumas de cuadrados extras

Teniendo en cuenta los resultados anteriores, realice una prueba con sumas de cuadrados extras con test lineal general; especifique claramente el modelo reducido y completo, estadístico de la prueba, su distribución, cálculo de valor P, decisión y conclusión a la luz de los datos. Justifique la hipótesis que desea probar en este numeral.

$$SSR(X_1, X_4, X_5 | X_2, X_3) = SSR(X_1, X_2, X_3, X_4, X_5) - SSR(X_2, X_3)$$

6.1. Prueba de hipótesis

$$\begin{cases} H_0 : \beta_2 = 0, \beta_3 = 0 \\ H_1 : \beta_2 \neq 0 \vee \beta_3 \neq 0 \end{cases}$$

6.2. Modelo completo y reducido

$$MF : Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_5 X_{i5} + \varepsilon_i, \quad i = 1, 2, \dots, 100$$

$$MR: Y_i = \beta_0 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \quad i = 1, 2, \dots, 100$$

6.3. Estadístico de prueba

$$\begin{aligned} F_0 &= \frac{[SSR(X_1, X_4, X_5|X_2, X_3)]/2}{MSE} \\ &= \frac{[SSE(X_2, X_3) - SSE(X_1, X_2, X_3, X_4, X_5)]/2}{\frac{SSE(MF)}{n-k-1}} = \frac{[0.6915854 - 0.6744272]/2}{0.6744272/94} \\ &= 1.1957338 \end{aligned}$$

6.4. Tabla del Test lineal general

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
96	0.6915854	NA	NA	NA	NA
94	0.6744272	2	0.0171582	1.195733	0.3070405

Con un nivel de significancia de $\alpha = 0.05$ el valor crítico es $f_{0.05,2,94} = 3.093266$.

Como $F_0 = 1.1957338 < f_{0.05,2,94} = 3.093266$, No hay evidencia para rechazar la **hipótesis nula**. por lo tanto X_2 (**TOEFL.Score**) y X_3 (**SOP**) no ayudan a explicar la posibilidad de ser admitido a una carrera de postgrado dado que en el modelo estan presentes **GRE.Score**, **LOR** y **CGPA**.

7. Sumas de cuadrados tipo I y tipo II

Calcule las sumas de cuadrados tipo I (secuenciales) y tipo II (parciales) ¿Cuál de las variables tienen menor valor en tales sumas? ¿Qué puede significar ello?

7.1. Sumas de cuadrados secuenciales

7.2. Tabla anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
GRE.Score	1	1.9853655	1.9853655	276.715340	0.000000
TOEFL.Score	1	0.0559996	0.0559996	7.805092	0.006314
SOP	1	0.1181923	0.1181923	16.473354	0.000102
LOR	1	0.1485633	0.1485633	20.706392	0.000016
CGPA	1	0.0593270	0.0593270	8.268855	0.004989
Residuals	94	0.6744272	0.0071748	NA	NA

$$-X_1 \text{ SSR}(X_1) = 1.98537$$

$$-X_2|X_1 \text{ SSR}(X_2|X_1) = 0.05600$$

$$-X_3|X_1, X_2 \text{ SSR}(X_3|X_1, X_2) = 0.1181923$$

$$-X_4|X_1, X_2, X_3 \text{ SSR}(X_4|X_1, X_2, X_3) = 0.14856$$

$$-X_5|X_1, X_2, X_3, X_4 \text{ SSR}(X_5|X_1, X_2, X_3, X_4) = 0.05933$$

7.3. Sumas de cuadrados parciales

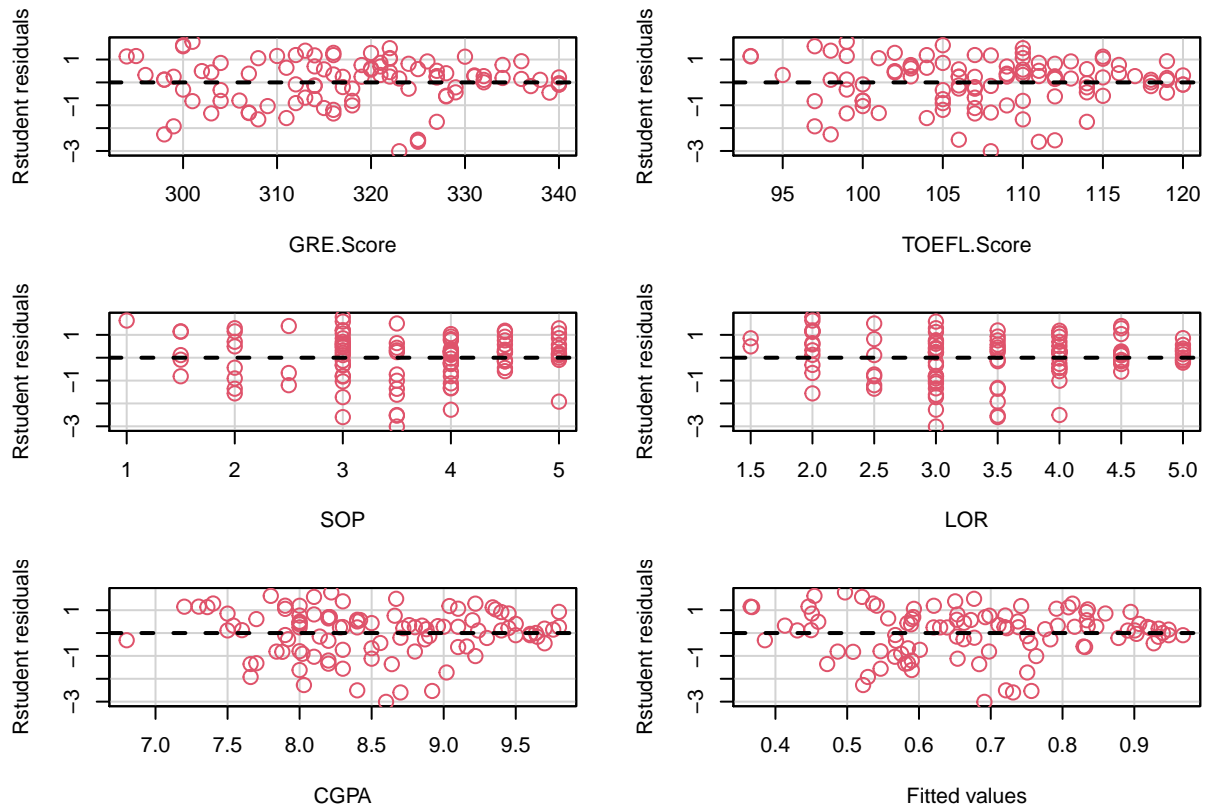
7.4. Tabla Anova

	Sum Sq	Df	F value	Pr(>F)
GRE.Score	0.0427161	1	5.9536667	0.0165615
TOEFL.Score	0.0063619	1	0.8867053	0.3487854
SOP	0.0073158	1	1.0196631	0.3151912
LOR	0.0646740	1	9.0141001	0.0034318
CGPA	0.0593270	1	8.2688553	0.0049890
Residuals	0.6744272	94	NA	NA

8. Residuales estudentizados vs. Valores ajustados

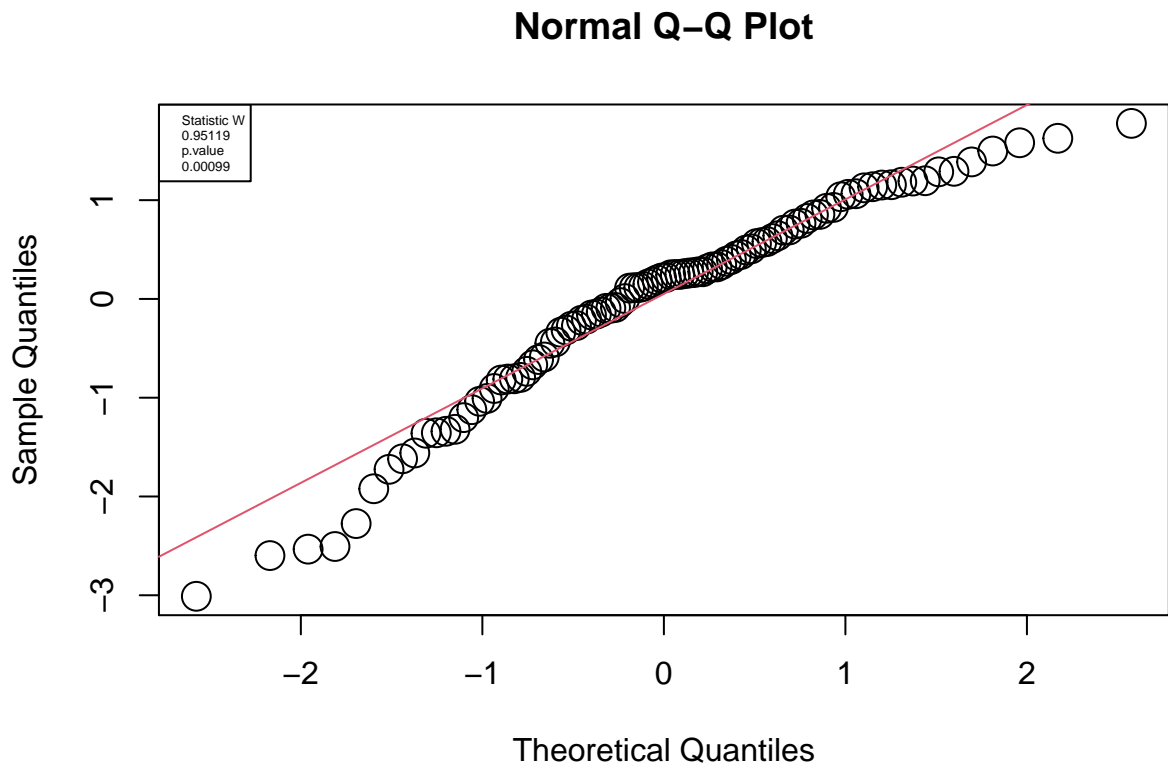
Construya y analice gráficos de los residuales estudentizados vs. Valores ajustados y contra las variables de regresión utilizadas. ¿Qué información proporcionan estas gráficas?

8.1. Gráfico de los residuales estudentizados vs. Valores ajustados



- En los gráficos de las variables **GRE.Score**, **TOEFL.Score**, **SOP**, **LOR**, **CGPA** y además de la gráfica de los valores ajustados no se observa ningún tipo de patrón, por lo tanto se cumple el supuesto de varianza constante. Se aprecian algunos valores **atípicos**, información que se verificará más adelante.

9. Gráfico q-norm residuales estudentizados



9.1. Pruebas de hipótesis

$$\begin{cases} H_0 : \varepsilon \sim Normal \\ H_1 : \varepsilon \not\sim Normal \end{cases}$$

Aunque muchos residuales se concentren cerca de la recta ajustada, se puede observar cantidad considerable que se aleja de esta generando una asimetría hacia la derecha, además al realizar la prueba de **Shapiro-Wilk** tenemos un **p valor de 0.00099** por lo que podemos rechazar la hipótesis nula, concluyendo de esta manera que hay evidencia para decir que no tienen un comportamiento normal.

10. Diagnostico sobre la presencia de observaciones atípicas , de balanceo y/o inflenciales y conclusiones

Influence measures of

Cuadro 4: Valores ajustados VS Residuales Estudentizados

	Ajustados	Errores
10	0.6911442	-3.010410
65	0.7310632	-2.598320
66	0.7566507	-2.533292
11	0.7212283	-2.506545
93	0.5221676	-2.274321
92	0.5288087	-1.921725

```
## lm(formula = data$Chance.of.Admit ~ ., data = data) :
##
##      dfb.1_ dfb.GRE. dfb.TOEFL dfb.SOP dfb.LOR dfb.CGPA      dffit cov.r
## 1  0.015973 -0.005271 -0.000885  1.31e-03  0.002340 -8.90e-03 -0.032882 1.108
## 2  0.010661 -0.024639  0.034326 -4.49e-03 -0.022518 -1.79e-03 -0.056504 1.105
## 3 -0.024402  0.089374 -0.049594 -2.27e-02  0.092619 -1.18e-01  0.213646 1.005
## 4 -0.146856  0.086964 -0.039096  3.90e-02 -0.271542  1.05e-01  0.319520 0.966
## 5 -0.021116  0.031828 -0.043631 -1.08e-01  0.005196  3.90e-02  0.150002 1.081
## 6 -0.141724  0.061166 -0.037386  9.77e-02 -0.252847  1.41e-01  0.313284 1.060
## 7 -0.000457  0.008477  0.050364 -5.48e-02  0.107665 -1.05e-01  0.157786 1.087
## 8  0.114484 -0.026751 -0.010652 -4.42e-02  0.206779 -1.07e-01  0.260972 1.052
## 9  0.030175 -0.047175  0.017687 -5.13e-02 -0.085702  7.61e-02  0.149108 1.145
## 10 0.331213 -0.336547  0.264130 -7.77e-02  0.323458 -1.01e-01 -0.519954 0.629
## 11 0.274899 -0.435231  0.332579 -3.85e-02 -0.156837  2.08e-01 -0.582385 0.759
## 12 -0.008881  0.010344 -0.008548 -3.61e-03  0.022573 -3.28e-03  0.044150 1.088
## 13 0.020921 -0.015918  0.011308  1.51e-02 -0.044372 -2.35e-03 -0.096725 1.067
## 14 0.066481 -0.074405  0.087449  3.07e-02  0.015934 -4.92e-02  0.114265 1.151
## 15 -0.020685  0.026567 -0.041123  5.93e-02 -0.128505  4.48e-02  0.151365 1.097
## 16 0.025571 -0.034946  0.043923 -4.62e-02  0.097394 -3.15e-02 -0.128213 1.062
## 17 0.004144 -0.004799  0.013229 -1.55e-02  0.024002 -1.68e-02 -0.035468 1.095
## 18 -0.019400  0.032642 -0.014545  3.59e-02 -0.011729 -3.35e-02  0.057146 1.119
## 19 -0.004595  0.032455 -0.017848 -3.99e-02  0.089071 -5.98e-02 -0.139737 1.053
## 20 0.064999 -0.057361 -0.007345  3.40e-03 -0.017908  6.56e-02  0.103144 1.112
## 21 -0.013257 -0.014360  0.095139  2.85e-02 -0.145174 -5.83e-02  0.260675 1.019
## 22 -0.077504  0.023602  0.061565 -1.24e-02 -0.115764 -1.39e-02  0.179340 1.143
## 23 0.023989 -0.039710  0.031889  1.38e-02  0.040631  5.31e-03  0.092445 1.112
## 24 -0.006252 -0.008693  0.012712  8.08e-03 -0.003257  1.02e-02  0.042630 1.116
## 25 -0.124820  0.008542  0.016489 -5.15e-02 -0.163031  1.71e-01  0.282553 1.103
## 26 0.011731 -0.004207 -0.004059  8.36e-04  0.000665 -3.30e-04 -0.021018 1.117
## 27 -0.011184  0.014573 -0.017092  3.03e-02 -0.018890  5.01e-03  0.043659 1.093
## 28 0.015479 -0.011365  0.005564 -1.95e-02  0.008679  3.29e-04  0.030514 1.140
## 29 0.089603  0.028826 -0.115292 -2.49e-02 -0.022679 -3.50e-03  0.295331 1.043
## 30 -0.113045  0.181667 -0.099887 -1.10e-01 -0.047718 -9.79e-02  0.340113 1.063
## 31 0.198657 -0.079822 -0.166260 -1.05e-02  0.006879  1.78e-01  0.375503 0.961
```

## 32	-0.051491	0.085395	-0.076129	3.43e-02	0.005311	-2.82e-02	0.100998	1.219
## 33	-0.013715	0.003431	0.007064	-2.34e-02	0.008446	2.70e-03	0.033081	1.155
## 34	0.009161	-0.008044	0.007108	4.24e-04	0.004706	-4.90e-03	-0.011655	1.171
## 35	-0.011816	0.006516	-0.031057	-2.48e-02	0.014565	4.89e-02	0.076932	1.150
## 36	0.061927	-0.049089	0.000340	4.18e-02	0.074195	1.17e-02	0.140805	1.111
## 37	0.088374	-0.089525	0.058479	-1.43e-03	0.046613	-6.19e-04	0.099095	1.232
## 38	0.289491	-0.429309	0.423333	-4.91e-01	0.031771	9.97e-02	0.694648	1.066
## 39	0.052161	-0.066482	0.122254	5.63e-02	-0.113521	-7.53e-02	0.251884	1.107
## 40	-0.255487	0.232718	-0.326200	-1.17e-01	-0.215678	3.35e-01	-0.477976	1.079
## 41	-0.277706	0.346912	-0.434158	-4.85e-03	-0.106308	2.07e-01	-0.497296	0.988
## 42	0.084137	-0.082768	0.058895	7.84e-02	0.108972	-4.59e-02	-0.218716	1.004
## 43	0.013603	0.026640	-0.011987	6.10e-02	0.123882	-1.05e-01	-0.180700	1.112
## 44	-0.018278	0.003128	0.021003	1.30e-02	-0.003311	-1.48e-02	0.051596	1.099
## 45	-0.014287	-0.021290	-0.016245	3.08e-02	-0.048117	1.02e-01	0.167914	1.044
## 46	0.007463	0.004356	-0.059844	1.46e-01	-0.041006	5.42e-02	0.213723	1.031
## 47	-0.002662	0.008571	-0.008722	1.63e-02	-0.031293	-9.99e-04	-0.046295	1.112
## 48	0.068405	-0.029890	0.004641	-5.31e-03	0.049188	-3.87e-02	-0.108445	1.114
## 49	0.029021	-0.027530	0.021785	-3.67e-02	0.075908	-1.40e-02	0.086286	1.134
## 50	-0.037297	0.035597	0.020621	-4.45e-02	0.057919	-6.79e-02	0.115376	1.103
## 51	0.026081	0.140218	-0.290453	-1.84e-01	0.285638	3.97e-02	0.516553	1.074
## 52	0.000729	-0.005882	0.007035	1.86e-02	-0.012222	1.43e-03	-0.026790	1.169
## 53	-0.160106	0.142206	0.082089	9.95e-02	-0.034187	-2.55e-01	0.327963	1.215
## 54	-0.093182	0.076644	0.053409	1.11e-01	-0.101269	-1.33e-01	0.237963	1.109
## 55	-0.018067	0.018214	0.043723	-1.96e-02	0.043928	-8.83e-02	0.109819	1.115
## 56	-0.087691	0.138606	-0.077528	3.09e-02	0.000688	-1.02e-01	0.177323	1.129
## 57	-0.118490	0.211814	-0.059031	-1.04e-01	0.122542	-2.61e-01	0.407980	1.054
## 58	0.014636	-0.005705	-0.002282	2.59e-02	-0.015036	-3.71e-03	0.037494	1.159
## 59	-0.022431	-0.005830	-0.021343	-4.07e-02	-0.005099	8.32e-02	-0.107246	1.183
## 60	0.047541	0.003836	0.048042	2.08e-01	0.240416	-2.42e-01	-0.420743	0.980
## 61	-0.012294	-0.055765	0.124552	-2.84e-03	0.018216	-4.29e-02	-0.182073	1.026
## 62	-0.081901	-0.008995	0.126685	-1.66e-01	0.069436	-4.59e-02	-0.272327	0.990
## 63	-0.057222	0.062205	-0.042738	2.05e-02	-0.012725	-1.06e-02	-0.073869	1.109
## 64	-0.012667	0.005902	0.017642	-9.38e-02	0.090919	-2.98e-02	-0.167268	1.006
## 65	0.175776	-0.087560	-0.048097	2.40e-01	0.000367	-1.04e-02	-0.407973	0.718
## 66	0.129238	-0.019008	-0.046649	1.14e-01	0.100351	-1.12e-01	-0.353366	0.728
## 67	0.155795	0.000649	-0.086083	1.89e-01	0.195107	-1.67e-01	-0.403104	0.931
## 68	-0.039722	0.037537	0.014270	2.48e-02	-0.001391	-6.70e-02	-0.157145	0.960
## 69	-0.064875	0.096831	0.005195	9.56e-02	-0.019210	-1.69e-01	-0.224471	1.049
## 70	0.014708	0.012221	-0.030933	-2.56e-02	0.009103	-2.35e-03	-0.093566	1.069
## 71	0.000075	-0.000283	0.000299	8.24e-05	0.000287	6.69e-05	0.000828	1.122
## 72	-0.020166	0.023640	-0.032527	1.13e-02	-0.001191	1.77e-02	0.046797	1.167
## 73	0.094859	-0.092866	-0.001896	4.41e-02	0.088543	7.29e-02	0.222919	1.086
## 74	0.137421	-0.136186	0.004666	6.64e-02	0.017041	1.22e-01	0.252313	1.020
## 75	-0.021091	0.018772	-0.003615	2.49e-02	-0.034008	-8.94e-03	-0.045077	1.188
## 76	0.028506	0.001118	-0.061666	1.17e-01	-0.066943	4.16e-02	-0.151314	1.180

```

## 77 -0.025741 0.014689 0.001279 -1.50e-02 -0.019196 5.39e-03 0.041335 1.104
## 78 0.127688 -0.081097 -0.172463 3.78e-02 -0.310500 3.46e-01 0.509345 0.944
## 79 0.030019 -0.000635 -0.033715 2.78e-02 -0.030049 1.39e-02 0.082230 1.127
## 80 0.111463 -0.020275 -0.100074 -1.26e-01 -0.016872 7.94e-02 0.320202 1.060
## 81 -0.026938 0.032904 -0.054656 1.47e-01 -0.053523 1.93e-02 -0.188881 1.055
## 82 0.008637 -0.003825 -0.005465 -4.43e-03 -0.006818 7.62e-03 -0.023089 1.130
## 83 0.098485 -0.084766 -0.026844 1.24e-01 0.054883 7.97e-02 0.281096 1.005
## 84 0.119641 -0.198210 0.158992 -9.28e-02 0.100847 7.59e-02 0.269832 1.063
## 85 -0.042854 0.040413 -0.028288 8.92e-03 -0.006532 3.63e-03 0.055436 1.132
## 86 -0.060438 0.124849 -0.180635 1.27e-01 -0.070342 4.58e-02 0.218606 1.113
## 87 0.016725 0.009047 -0.022808 8.29e-02 -0.008040 -2.18e-02 0.103772 1.079
## 88 -0.006190 0.008571 -0.001700 8.79e-03 -0.010627 -8.03e-03 0.030202 1.078
## 89 -0.011000 0.006249 -0.012024 -2.30e-02 -0.006780 2.29e-02 -0.035760 1.122
## 90 0.020258 -0.021429 0.007697 3.48e-02 -0.013044 1.02e-02 0.059259 1.089
## 91 -0.001317 -0.019541 -0.000741 -2.70e-02 -0.038467 6.19e-02 -0.074524 1.142
## 92 -0.342180 0.033477 0.170196 -5.82e-01 -0.145039 2.59e-01 -0.776410 0.982
## 93 -0.396022 0.201905 0.139874 -2.95e-01 0.018618 -1.16e-01 -0.605135 0.825
## 94 -0.086032 0.013961 0.078864 -1.12e-02 -0.022405 -3.13e-02 -0.172054 1.066
## 95 -0.084618 0.025508 0.023474 1.32e-01 -0.021645 -7.56e-03 -0.267881 0.985
## 96 -0.050721 0.039164 -0.001385 1.53e-01 -0.013264 -5.24e-02 -0.203312 1.086
## 97 -0.049015 0.000753 0.053381 2.65e-04 -0.008551 -1.66e-02 -0.126933 1.050
## 98 -0.002882 -0.028471 0.071999 -1.24e-02 0.017486 -3.54e-02 0.087095 1.146
## 99 -0.000438 -0.006485 0.015672 7.69e-03 0.005817 -9.51e-03 0.026547 1.128
## 100 0.004924 -0.011130 0.015630 -1.58e-03 0.008346 -4.14e-03 0.025468 1.088
##      cook.d      hat inf
## 1  1.82e-04 0.0393
## 2  5.37e-04 0.0401
## 3  7.57e-03 0.0310
## 4  1.68e-02 0.0435
## 5  3.77e-03 0.0441
## 6  1.63e-02 0.0720
## 7  4.17e-03 0.0489
## 8  1.13e-02 0.0572
## 9  3.74e-03 0.0831
## 10 4.15e-02 0.0290  *
## 11 5.35e-02 0.0512  *
## 12 3.28e-04 0.0251
## 13 1.57e-03 0.0243
## 14 2.20e-03 0.0823
## 15 3.84e-03 0.0530
## 16 2.75e-03 0.0297
## 17 2.12e-04 0.0292
## 18 5.50e-04 0.0507
## 19 3.27e-03 0.0290
## 20 1.79e-03 0.0530

```



```

## 21  1.13e-02  0.0453
## 22  5.40e-03  0.0871
## 23  1.44e-03  0.0507
## 24  3.06e-04  0.0465
## 25  1.33e-02  0.0851
## 26  7.44e-05  0.0463
## 27  3.21e-04  0.0283
## 28  1.57e-04  0.0657
## 29  1.45e-02  0.0615
## 30  1.92e-02  0.0796
## 31  2.31e-02  0.0534
## 32  1.72e-03  0.1291  *
## 33  1.84e-04  0.0778
## 34  2.29e-05  0.0892
## 35  9.96e-04  0.0769
## 36  3.33e-03  0.0594
## 37  1.65e-03  0.1379  *
## 38  7.90e-02  0.1543
## 39  1.06e-02  0.0806
## 40  3.78e-02  0.1160
## 41  4.05e-02  0.0863
## 42  7.94e-03  0.0321
## 43  5.47e-03  0.0684
## 44  4.48e-04  0.0342
## 45  4.71e-03  0.0324
## 46  7.60e-03  0.0387
## 47  3.61e-04  0.0441
## 48  1.98e-03  0.0553
## 49  1.25e-03  0.0662
## 50  2.24e-03  0.0493
## 51  4.40e-02  0.1218
## 52  1.21e-04  0.0885
## 53  1.80e-02  0.1550  *
## 54  9.47e-03  0.0786
## 55  2.03e-03  0.0563
## 56  5.28e-03  0.0776
## 57  2.75e-02  0.0907
## 58  2.37e-04  0.0810
## 59  1.94e-03  0.1044
## 60  2.91e-02  0.0678
## 61  5.52e-03  0.0299
## 62  1.23e-02  0.0397
## 63  9.18e-04  0.0458
## 64  4.65e-03  0.0218
## 65  2.61e-02  0.0241  *

```

## 66	1.97e-02	0.0191	*
## 67	2.65e-02	0.0518	
## 68	4.08e-03	0.0132	
## 69	8.40e-03	0.0473	
## 70	1.47e-03	0.0247	
## 71	1.15e-07	0.0499	
## 72	3.69e-04	0.0880	
## 73	8.31e-03	0.0635	
## 74	1.06e-02	0.0436	
## 75	3.42e-04	0.1035	
## 76	3.85e-03	0.1076	
## 77	2.88e-04	0.0367	
## 78	4.23e-02	0.0759	
## 79	1.14e-03	0.0604	
## 80	1.70e-02	0.0736	
## 81	5.96e-03	0.0416	
## 82	8.98e-05	0.0566	
## 83	1.31e-02	0.0455	
## 84	1.21e-02	0.0637	
## 85	5.17e-04	0.0608	
## 86	8.00e-03	0.0767	
## 87	1.81e-03	0.0320	
## 88	1.54e-04	0.0150	
## 89	2.15e-04	0.0510	
## 90	5.91e-04	0.0283	
## 91	9.35e-04	0.0708	
## 92	9.77e-02	0.1403	*
## 93	5.84e-02	0.0661	
## 94	4.95e-03	0.0421	
## 95	1.19e-02	0.0376	
## 96	6.91e-03	0.0591	
## 97	2.70e-03	0.0249	
## 98	1.28e-03	0.0755	
## 99	1.19e-04	0.0552	
## 100	1.09e-04	0.0221	

[illegible]


```
## 99  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE FALSE FALSE  FALSE FALSE
## 100 FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE FALSE FALSE  FALSE FALSE
```

-La observación **10** es un **outlier** ya que el valor absoluto de su residual estudentizado es **3.01**, mayor a **3**.

-De acuerdo al **COVRATIO** y el **DFFITS**, las observaciones **10,11, 32, 37, 53, 65, 66** y **92** son **influenciables**.

-Para evaluar las observaciones de **balanceo** miramos las que superan la cota de $\frac{2(k+1)}{n} = \frac{2(5+1)}{100} = 0.12$. Las observaciones **32, 37, 38, 53** y **92** superan dicha cota y por lo tanto, son de **balanceo**.

-En conclusión, las observaciones **10, 11, 53, 65** y **66** son **influenciables**; la observación **38** es de **balanceo** y las observaciones **32, 37, 53** y **92** son **influenciables** y de **balanceo**.

11. Ejercicio11

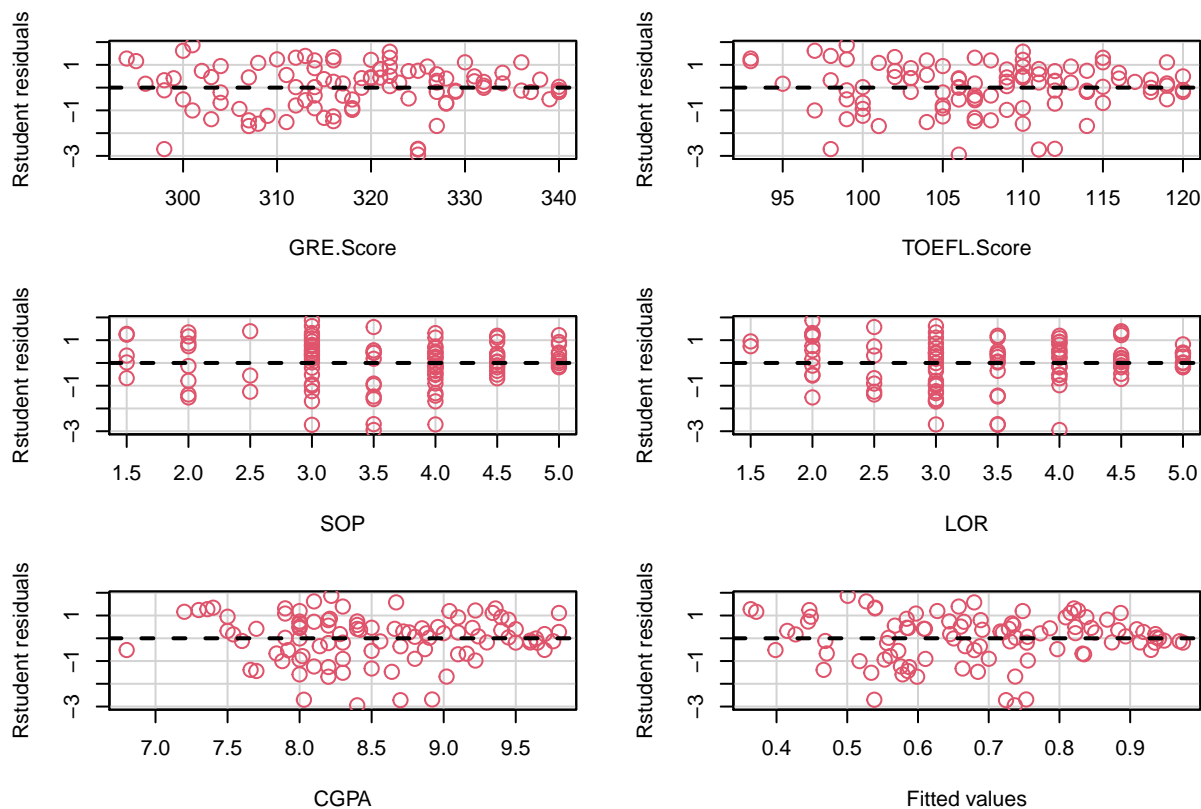
Ajuste el modelo de regresión sin las observaciones 10, 38 y 92, suponga que se establece que hay un error de digitación con estas dos observaciones, presente sólo la tabla de parámetros ajustados resultante ¿Cambian notoriamente las estimaciones de los parámetros, sus errores estándar y/o la significancia? ¿Qué concluye al respecto? Evalúe el gráfico de normalidad para los residuales estudentizados para este ajuste ¿mejoró la normalidad?

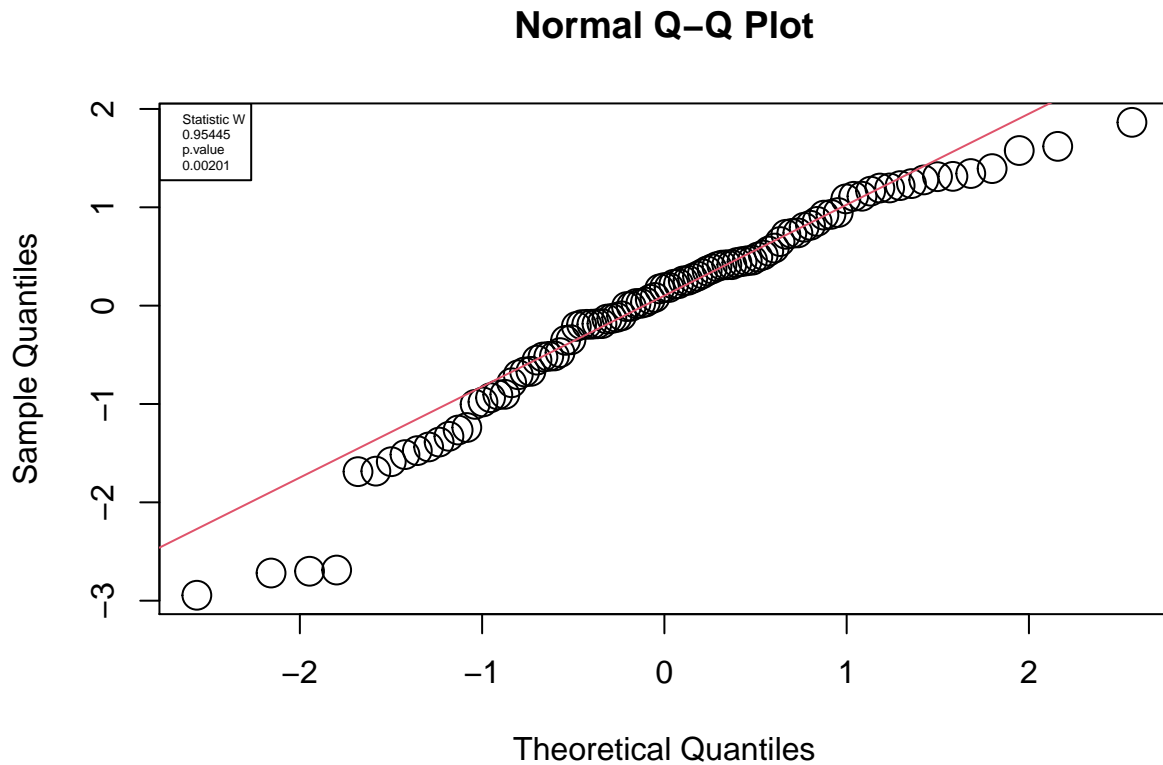
Concluya sobre los efectos de este par de observaciones.

```
## Analysis of Variance Table
##
## Response: Chance.of.Admit
##
##               Df Sum Sq Mean Sq F value    Pr(>F)
## FO(GRE.Score, TOEFL.Score, SOP, LOR, CGPA)  5 2.31585 0.46317  74.498 < 2.2e-16
## Residuals                                91 0.56577 0.00622
##
## FO(GRE.Score, TOEFL.Score, SOP, LOR, CGPA) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = Chance.of.Admit ~ ., data = AdmissionPredict_sin_influencias)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.21685 -0.03972  0.01428  0.05500  0.13928
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.8564761  0.2897498  -6.407 6.42e-09 ***
## GRE.Score    0.0053985  0.0016314   3.309  0.00134 **
## TOEFL.Score  0.0001168  0.0030020   0.039  0.96904
## SOP          0.0267458  0.0121536   2.201  0.03029 *
## LOR          0.0400539  0.0133955   2.990  0.00359 **
## CGPA         0.0681658  0.0248061   2.748  0.00723 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07885 on 91 degrees of freedom
## Multiple R-squared:  0.8037, Adjusted R-squared:  0.7929
## F-statistic: 74.5 on 5 and 91 DF, p-value: < 2.2e-16
```





12. Ejercicio 12

Para el modelo con todas las variables y sin las observaciones 10, 38 y 92, realice diagnósticos de multicolinealidad mediante

12.1. Matriz de correlación de las variables predictoras

	GRE.Score	TOEFL.Score	SOP	LOR	CGPA	Chance.of.A
GRE.Score	1.0000000	0.8907133	0.5405067	0.6538861	0.7984379	0.818
TOEFL.Score	0.8907133	1.0000000	0.6237639	0.5963160	0.8091454	0.780
SOP	0.5405067	0.6237639	1.0000000	0.5988709	0.6826010	0.670
LOR	0.6538861	0.5963160	0.5988709	1.0000000	0.7440234	0.758
CGPA	0.7984379	0.8091454	0.6826010	0.7440234	1.0000000	0.839
Chance.of.Admit	0.8184117	0.7808914	0.6706553	0.7580440	0.8390995	1.000

Entre las pruebas **GRE - TOEFL**, **GRE - CGPA**, **TOEFL - CGPA** y finalmente **LOR - CGPA** se observan **correlaciones fuertes**, esto puede indicar problemas de **multicolinealidad**.

Se observa que entre **GRE - SOP**, **GRE - LOR**, **TOEFL - SOP**, **TOEFL - LOR** se tienen **correlaciones moderadas**.

12.2. VIF's

GRE.Score	TOEFL.Score	SOP	LOR	CGPA
5.859487	6.190994	2.108352	2.497704	4.599604

- En los factores de **inflación de varianza** no se concluye que existan problemas de **multicolineadlidad**, pues nos indica que ninguna estimación **supera** el valor de **10**.

12.3. Proporciones de varianza

```
## Condition
## Index      Variance Decomposition Proportions
##              intercept GRE.Score TOEFL.Score SOP   LOR   CGPA
## 1      1.000 0.000      0.000      0.000      0.001 0.001 0.000
## 2     10.083 0.002      0.000      0.001      0.173 0.161 0.001
## 3     14.645 0.000      0.000      0.000      0.597 0.490 0.000
## 4     60.699 0.197      0.003      0.036      0.179 0.235 0.399
## 5     83.509 0.092      0.009      0.368      0.001 0.025 0.593
## 6    189.536 0.710      0.988      0.596      0.049 0.088 0.007
```

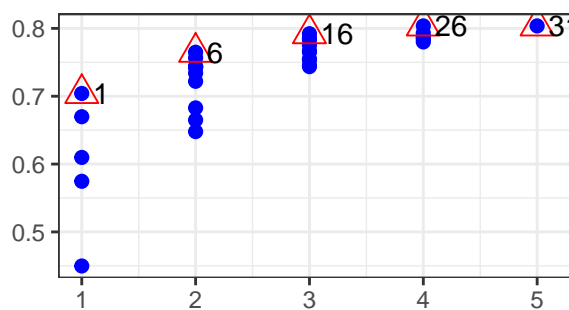
- La raíz del número condición es de **189**. Lo cual nos indica que se tienen problemas graves de **multicolinealidad**.
- Examinando la descomposición de varianza se visualiza que existe problemas de **multicolinealidad** entre las pruebas **GRE - TOEFL** y las pruebas **SOP - LOR**

13. Ejercicio13

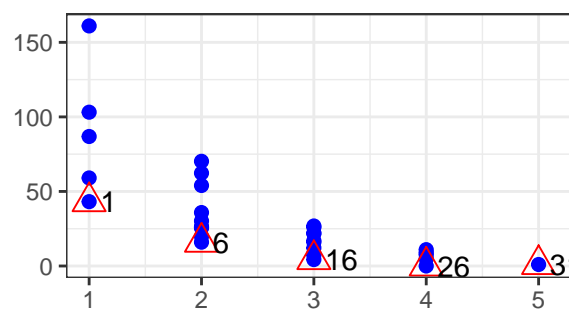
En el modelo ajustado sin las observaciones 10, 38 y 92, construya modelos de regresión utilizando los métodos de selección (muestre de cada método sólo la tabla de resumen de este y la tabla ANOVA y la de parámetros estimados del modelo finalmente resultante):

page 1 of 2

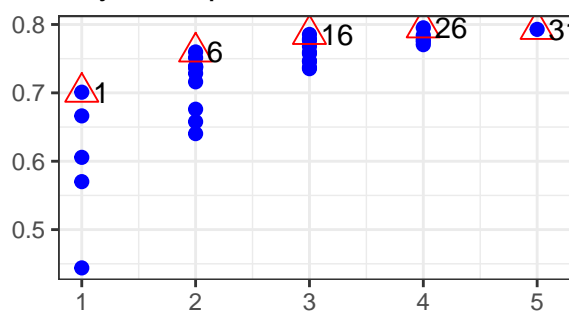
R-Square



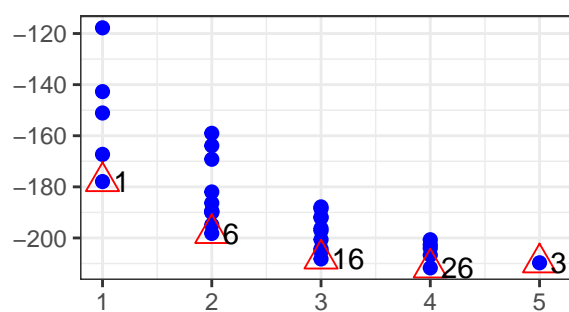
Cp



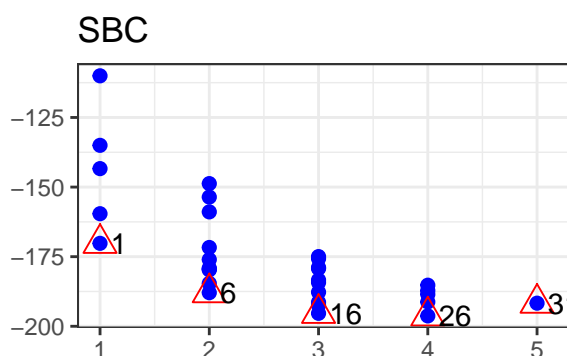
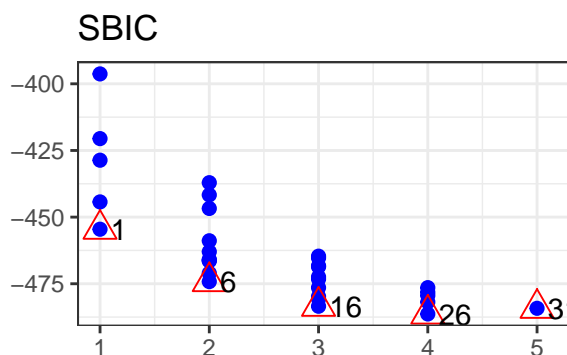
Adj. R-Square



AIC



page 2 of 2



# de covariables	modelo	R2_adj
1	(1) y~CGPA	0.75
2	(6) y~GRE.Score+CGPA	0.81
3	(16) y~GRE.Score+LOR+CGPA	0.83
4	(26) y~GRE.Score+SOP+LOR+CGPA	0.83
5	(31) y~GRE.Score+TOEFL.Score+SOP+LOR+CGPA	0.83

- De acuerdo al **principio de parsimonia** un buen modelo bajo el criterio del **R2_adj** es el modelo **(6) y~GRE.Score+CGPA**

13.1. Selección según el R_{adj}^2

```
##
## Call:
## lm(formula = Chance.of.Admit ~ GRE.Score + CGPA, data = AdmissionPredict_sin_influenc
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23031 -0.04277  0.01477  0.05736  0.13309
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.290157   0.264866  -8.646 1.38e-13 ***
## GRE.Score    0.005941   0.001205   4.930 3.53e-06 ***
## CGPA         0.127528   0.020685   6.165 1.75e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0849 on 94 degrees of freedom
## Multiple R-squared:  0.7649, Adjusted R-squared:  0.7599
## F-statistic: 152.9 on 2 and 94 DF,  p-value: < 2.2e-16
```

Ecuacion Ajustada

$$\hat{Y}_i = -2.290157 + 0.005941X_{i1} + 0.127528X_{i5}$$

Tabla anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(GRE.Score, CGPA)	2	2.2040778	1.1020389	152.8941	0
Residuals	94	0.6775387	0.0072079	NA	NA

13.2. Selección según el estadístico C_p

# de covariables	modelo	abs(Cp - p)
1	(1) $y \sim \text{CGPA}$	$45.00 - 1 = 44$
2	(6) $y \sim \text{GRE.Score} + \text{CGPA}$	$15.39 - 2 = 13.39$
3	(16) $y \sim \text{GRE.Score} + \text{LOR} + \text{CGPA}$	$4.17 - 3 = 1.17$
4	(26) $y \sim \text{GRE.Score} + \text{SOP} + \text{LOR} + \text{CGPA}$	$4.00 - 4 = 0.00$
5	(31) $y \sim \text{GRE.Score} + \text{TOEFL.Score} + \text{SOP} + \text{LOR} + \text{CGPA}$	$6 - 5 = 1$

- De acuerdo al **Cp** el mejor modelo es el **(26) $y \sim \text{GRE.Score} + \text{SOP} + \text{LOR} + \text{CGPA}$**

```
##
## Call:
## lm(formula = Chance.of.Admit ~ GRE.Score + LOR + CGPA + SOP,
##     data = AdmissionPredict_sin_influencias)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.21728 -0.03976  0.01443  0.05516  0.13902
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.860801   0.266142  -6.992 4.24e-10 ***
## GRE.Score    0.005444   0.001127   4.832 5.37e-06 ***
## LOR          0.039939   0.012995   3.073 0.00279 **
## CGPA         0.068438   0.023673   2.891 0.00479 **
## SOP          0.026878   0.011603   2.316 0.02275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07842 on 92 degrees of freedom
## Multiple R-squared:  0.8037, Adjusted R-squared:  0.7951
## F-statistic: 94.14 on 4 and 92 DF,  p-value: < 2.2e-16
```

Ecuacion ajustada

$$\hat{Y}_i = -1.860801 + 0.005444X_{i1} + 0.026878X_{i3} + 0.039939X_{i4} + 0.068438X_{i5}$$

Tabla anova

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(GRE.Score, LOR, CGPA, SOP)	4	2.3158387	0.5789597	94.14348	0
Residuals	92	0.5657778	0.0061498	NA	NA

13.3. Stepwise

```
## Stepwise Selection Method
## -----
##
## Candidate Terms:
##
## 1. GRE.Score
## 2. TOEFL.Score
## 3. SOP
## 4. LOR
## 5. CGPA
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - CGPA added
##
##                               Model Summary
## -----
```

```
## R                0.839      RMSE                0.095
## R-Squared        0.704      Coef. Var            13.749
## Adj. R-Squared   0.701      MSE                 0.009
## Pred R-Squared   0.694      MAE                 0.075
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
```

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	2.029	1	2.029	226.041	0.0000
Residual	0.853	95	0.009		
Total	2.882	96			

```
## -----
```

```
##
```

```
## Parameter Estimates
```

	model	Beta	Std. Error	Std. Beta	t	Sig.	lower	upper
(Intercept)		-1.095	0.119		-9.198	0.000	-1.331	-0.859
CGPA		0.209	0.014	0.839	15.035	0.000	0.181	0.237

```
## -----
```

```
##
```

```
##
```

```
##
```

```
## Stepwise Selection: Step 2
```

```
##
```

```
## - GRE.Score added
```

```
##
```

```
## Model Summary
```

```
## -----
```

R	0.875	RMSE	0.085
R-Squared	0.765	Coef. Var	12.321
Adj. R-Squared	0.760	MSE	0.007
Pred R-Squared	0.752	MAE	0.066

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```

## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.204          2          1.102      152.894      0.0000
## Residual        0.678         94          0.007
## Total           2.882         96
## -----
##
##              Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upp
## -----
## (Intercept)    -2.290          0.265              -8.646      0.000      -2.816      -1.7
##      CGPA       0.128          0.021           0.512      6.165      0.000      0.086      0.1
##      GRE.Score   0.006          0.001           0.410      4.930      0.000      0.004      0.0
## -----
##
##
##
##              Model Summary
## -----
## R              0.875          RMSE              0.085
## R-Squared       0.765          Coef. Var       12.321
## Adj. R-Squared  0.760          MSE           0.007
## Pred R-Squared  0.752          MAE           0.066
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##              ANOVA
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.204          2          1.102      152.894      0.0000
## Residual        0.678         94          0.007
## Total           2.882         96
## -----
##
##              Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig.      lower      upp
## -----

```

```
## (Intercept)    -2.290          0.265          -8.646      0.000      -2.816      -1.7
##           CGPA      0.128          0.021          0.512      6.165      0.000      0.086      0.1
##      GRE.Score      0.006          0.001          0.410      4.930      0.000      0.004      0.0
```

```
## -----
```

```
##
```

```
##
```

```
##
```

```
## Stepwise Selection: Step 3
```

```
##
```

```
## - LOR added
```

```
##
```

```
##                               Model Summary
```

```
## -----
```

```
## R                               0.890      RMSE                               0.080
```

```
## R-Squared                       0.792      Coef. Var                       11.645
```

```
## Adj. R-Squared                   0.786      MSE                               0.006
```

```
## Pred R-Squared                   0.775      MAE                               0.063
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
##                               ANOVA
```

```
## -----
```

```
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
```

```
## -----
```

```
## Regression      2.283          3          0.761      118.188      0.0000
```

```
## Residual        0.599         93          0.006
```

```
## Total           2.882         96
```

```
## -----
```

```
##
```

```
##                               Parameter Estimates
```

```
## -----
```

```
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
```

```
## -----
```

```
## (Intercept)    -1.931          0.271          -7.140      0.000      -2.469      -1.3
```

```
##           CGPA      0.089          0.022          0.359      3.993      0.000      0.045      0.1
```

```
##      GRE.Score      0.005          0.001          0.368      4.638      0.000      0.003      0.0
```

```
##           LOR      0.046          0.013          0.250      3.498      0.001      0.020      0.0
```

```
## -----
```

```
##
```

```
##
```

```
##
```

```
##                               Model Summary
```

```

## -----
## R                0.890      RMSE                0.080
## R-Squared        0.792      Coef. Var            11.645
## Adj. R-Squared   0.786      MSE                 0.006
## Pred R-Squared   0.775      MAE                 0.063
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##              Sum of
##              Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.283        3          0.761    118.188    0.0000
## Residual        0.599       93          0.006
## Total           2.882       96
## -----
##
##                               Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upper
## -----
## (Intercept)   -1.931      0.271          -7.140      0.000    -2.469    -1.3
## CGPA          0.089      0.022           0.359      3.993     0.000     0.045     0.1
## GRE.Score     0.005      0.001           0.368      4.638     0.000     0.003     0.0
## LOR           0.046      0.013           0.250      3.498     0.001     0.020     0.0
## -----
##
##
##
## Stepwise Selection: Step 4
##
## - SOP added
##
##                               Model Summary
## -----
## R                0.896      RMSE                0.078
## R-Squared        0.804      Coef. Var            11.381
## Adj. R-Squared   0.795      MSE                 0.006
## Pred R-Squared   0.784      MAE                 0.060
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error

```


MAE: Mean Absolute Error

##

ANOVA

##	Sum of				
##	Squares	DF	Mean Square	F	Sig.
## Regression	2.316	4	0.579	94.143	0.0000
## Residual	0.566	92	0.006		
## Total	2.882	96			

##

##

Parameter Estimates

##	model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
##	(Intercept)	-1.861	0.266		-6.992	0.000	-2.389	-1.333
##	CGPA	0.068	0.024	0.275	2.891	0.005	0.021	0.115
##	GRE.Score	0.005	0.001	0.375	4.832	0.000	0.003	0.007
##	LOR	0.040	0.013	0.219	3.073	0.003	0.014	0.066
##	SOP	0.027	0.012	0.149	2.316	0.023	0.004	0.050

##

##

##

##

Model Summary

##	R	0.896	RMSE	0.078
##	R-Squared	0.804	Coef. Var	11.381
##	Adj. R-Squared	0.795	MSE	0.006
##	Pred R-Squared	0.784	MAE	0.060

##

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

##

ANOVA

##	Sum of				
##	Squares	DF	Mean Square	F	Sig.
## Regression	2.316	4	0.579	94.143	0.0000
## Residual	0.566	92	0.006		
## Total	2.882	96			

##

```

##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp
## -----
## (Intercept)    -1.861        0.266                -6.992    0.000    -2.389    -1.3
##      CGPA       0.068        0.024         0.275      2.891    0.005     0.021    0.1
##      GRE.Score   0.005        0.001         0.375      4.832    0.000     0.003    0.0
##      LOR         0.040        0.013         0.219      3.073    0.003     0.014    0.0
##      SOP         0.027        0.012         0.149      2.316    0.023     0.004    0.0
## -----
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## -----
##
##                                     Model Summary
## -----
## R                0.896      RMSE                0.078
## R-Squared         0.804      Coef. Var          11.381
## Adj. R-Squared    0.795      MSE                0.006
## Pred R-Squared    0.784      MAE                0.060
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                                     ANOVA
## -----
##      Sum of
##      Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.316        4          0.579    94.143    0.0000
## Residual        0.566       92          0.006
## Total           2.882       96
## -----
##
##                                     Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp
## -----

```

```
## (Intercept)    -1.861      0.266      -6.992    0.000    -2.389    -1.3
##          CGPA      0.068      0.024      0.275    2.891    0.005    0.021    0.1
##    GRE.Score      0.005      0.001      0.375    4.832    0.000    0.003    0.0
##          LOR      0.040      0.013      0.219    3.073    0.003    0.014    0.0
##          SOP      0.027      0.012      0.149    2.316    0.023    0.004    0.0
## -----
```

```
##
##                               Stepwise Selection Summary
## -----
##                               Added/      Adj.
## Step    Variable    Removed    R-Square    R-Square    C(p)      AIC      RMS
## -----
##    1      CGPA      addition    0.704      0.701      44.1520   -177.9290  0.09
##    2    GRE.Score      addition    0.765      0.760      17.9770   -198.2339  0.08
##    3      LOR      addition    0.792      0.786      7.3090    -208.2209  0.08
##    4      SOP      addition    0.804      0.795      4.0020    -211.7196  0.07
## -----
```

13.4. Selección hacia adelante o forward

```
## Forward Selection Method
## -----
##
## Candidate Terms:
##
## 1. GRE.Score
## 2. TOEFL.Score
## 3. SOP
## 4. LOR
## 5. CGPA
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - CGPA
##
##                               Model Summary
## -----
## R              0.839      RMSE              0.095
## R-Squared      0.704      Coef. Var          13.749
## Adj. R-Squared 0.701      MSE              0.009
```

```
## Pred R-Squared      0.694      MAE      0.075
```

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
```

	Sum of Squares	DF	Mean Square	F	Sig.
--	-------------------	----	-------------	---	------

```
## -----
```

Regression	2.029	1	2.029	226.041	0.0000
------------	-------	---	-------	---------	--------

Residual	0.853	95	0.009		
----------	-------	----	-------	--	--

Total	2.882	96			
-------	-------	----	--	--	--

```
## -----
```

```
##
```

```
## Parameter Estimates
```

```
## -----
```

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
-------	------	------------	-----------	---	-----	-------	-------

```
## -----
```

(Intercept)	-1.095	0.119		-9.198	0.000	-1.331	-0.859
-------------	--------	-------	--	--------	-------	--------	--------

CGPA	0.209	0.014	0.839	15.035	0.000	0.181	0.237
------	-------	-------	-------	--------	-------	-------	-------

```
## -----
```

```
##
```

```
##
```

```
##
```

```
## Forward Selection: Step 2
```

```
##
```

```
## - GRE.Score
```

```
##
```

```
## Model Summary
```

```
## -----
```

R	0.875	RMSE	0.085
---	-------	------	-------

R-Squared	0.765	Coef. Var	12.321
-----------	-------	-----------	--------

Adj. R-Squared	0.760	MSE	0.007
----------------	-------	-----	-------

Pred R-Squared	0.752	MAE	0.066
----------------	-------	-----	-------

```
## -----
```

```
## RMSE: Root Mean Square Error
```

```
## MSE: Mean Square Error
```

```
## MAE: Mean Absolute Error
```

```
##
```

```
## ANOVA
```

```
## -----
```

	Sum of Squares	DF	Mean Square	F	Sig.
--	-------------------	----	-------------	---	------

```

## -----
## Regression      2.204      2      1.102    152.894    0.0000
## Residual       0.678     94      0.007
## Total          2.882     96
## -----
##
##                               Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp
## -----
## (Intercept)   -2.290      0.265              -8.646    0.000    -2.816    -1.7
##      CGPA      0.128      0.021      0.512     6.165    0.000     0.086     0.1
##      GRE.Score  0.006      0.001      0.410     4.930    0.000     0.004     0.0
## -----
##
##
##
## Forward Selection: Step 3
##
## - LOR
##
##                               Model Summary
## -----
## R              0.890      RMSE              0.080
## R-Squared      0.792      Coef. Var      11.645
## Adj. R-Squared 0.786      MSE              0.006
## Pred R-Squared 0.775      MAE              0.063
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##      Sum of
##      Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.283      3      0.761    118.188    0.0000
## Residual        0.599     93      0.006
## Total           2.882     96
## -----
##
##                               Parameter Estimates
## -----
##      model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp

```

```

## -----
## (Intercept)    -1.931          0.271          -7.140      0.000      -2.469      -1.3
##          CGPA      0.089          0.022          0.359      3.993      0.000      0.045      0.1
##          GRE.Score  0.005          0.001          0.368      4.638      0.000      0.003      0.0
##          LOR       0.046          0.013          0.250      3.498      0.001      0.020      0.0
## -----
##
##
##
## Forward Selection: Step 4
##
## - SOP
##
##                               Model Summary
## -----
## R                          0.896          RMSE                          0.078
## R-Squared                   0.804          Coef. Var                     11.381
## Adj. R-Squared              0.795          MSE                          0.006
## Pred R-Squared              0.784          MAE                          0.060
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.316          4          0.579      94.143      0.0000
## Residual        0.566          92          0.006
## Total           2.882          96
## -----
##
##                               Parameter Estimates
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp
## -----
## (Intercept) -1.861          0.266          -6.992      0.000      -2.389      -1.3
##          CGPA  0.068          0.024          0.275      2.891      0.005      0.021      0.1
##          GRE.Score  0.005          0.001          0.375      4.832      0.000      0.003      0.0
##          LOR    0.040          0.013          0.219      3.073      0.003      0.014      0.0
##          SOP    0.027          0.012          0.149      2.316      0.023      0.004      0.0
## -----
##

```

```

##
##
## No more variables to be added.
##
## Variables Entered:
##
## + CGPA
## + GRE.Score
## + LOR
## + SOP
##
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                               0.896      RMSE                0.078
## R-Squared                       0.804      Coef. Var          11.381
## Adj. R-Squared                   0.795      MSE                0.006
## Pred R-Squared                   0.784      MAE                0.060
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      2.316           4           0.579      94.143      0.0000
## Residual        0.566          92           0.006
## Total           2.882          96
## -----
##
##                               Parameter Estimates
## -----
## model      Beta      Std. Error      Std. Beta      t      Sig      lower      upp
## -----
## (Intercept) -1.861      0.266           -6.992      0.000      -2.389      -1.3
## CGPA         0.068      0.024           0.275      2.891      0.005      0.021      0.1
## GRE.Score    0.005      0.001           0.375      4.832      0.000      0.003      0.0
## LOR          0.040      0.013           0.219      3.073      0.003      0.014      0.0
## SOP          0.027      0.012           0.149      2.316      0.023      0.004      0.0

```

```
## -----
##
##                               Selection Summary
## -----
##      Variable                Adj.
## Step   Entered      R-Square  R-Square    C(p)      AIC      RMSE
## -----
##    1    CGPA          0.7041    0.7010    44.1518   -177.9290   0.0947
##    2   GRE.Score      0.7649    0.7599    17.9775   -198.2339   0.0849
##    3    LOR           0.7922    0.7855     7.3092   -208.2209   0.0802
##    4    SOP           0.8037    0.7951     4.0015   -211.7196   0.0784
## -----
```

13.5. Selección hacia atrás o backward

```
## Backward Elimination Method
## -----
##
## Candidate Terms:
##
## 1 . GRE.Score
## 2 . TOEFL.Score
## 3 . SOP
## 4 . LOR
## 5 . CGPA
##
## We are eliminating variables based on p value...
##
## - TOEFL.Score
##
## Backward Elimination: Step 1
##
## Variable TOEFL.Score Removed
##
##                               Model Summary
## -----
## R              0.896      RMSE              0.078
## R-Squared      0.804      Coef. Var        11.381
## Adj. R-Squared 0.795      MSE              0.006
## Pred R-Squared 0.784      MAE              0.060
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
```


MAE: Mean Absolute Error

##

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	2.316	4	0.579	94.143	0.0000
Residual	0.566	92	0.006		
Total	2.882	96			

##

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	-1.861	0.266		-6.992	0.000	-2.389	-1.333
GRE.Score	0.005	0.001	0.375	4.832	0.000	0.003	0.007
SOP	0.027	0.012	0.149	2.316	0.023	0.004	0.050
LOR	0.040	0.013	0.219	3.073	0.003	0.014	0.066
CGPA	0.068	0.024	0.275	2.891	0.005	0.021	0.115

##

##

##

No more variables satisfy the condition of p value = 0.05

##

##

Variables Removed:

##

- TOEFL.Score

##

##

Final Model Output

##

Model Summary

R	0.896	RMSE	0.078
R-Squared	0.804	Coef. Var	11.381
Adj. R-Squared	0.795	MSE	0.006
Pred R-Squared	0.784	MAE	0.060

RMSE: Root Mean Square Error

MSE: Mean Square Error

MAE: Mean Absolute Error

##

ANOVA

##	Sum of				
##	Squares	DF	Mean Square	F	Sig.
## Regression	2.316	4	0.579	94.143	0.0000
## Residual	0.566	92	0.006		
## Total	2.882	96			

##

##

Parameter Estimates

##	model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
## (Intercept)		-1.861	0.266		-6.992	0.000	-2.389	-1.333
## GRE.Score		0.005	0.001	0.375	4.832	0.000	0.003	0.007
## SOP		0.027	0.012	0.149	2.316	0.023	0.004	0.050
## LOR		0.040	0.013	0.219	3.073	0.003	0.014	0.066
## CGPA		0.068	0.024	0.275	2.891	0.005	0.021	0.115

##

##

##

Elimination Summary

##	Variable		Adj.			
## Step	Removed	R-Square	R-Square	C(p)	AIC	RMSE
## 1	TOEFL.Score	0.8037	0.7951	4.0015	-211.7196	0.0784

14. Selección del modelo

Con base en los anteriores numerales, ¿Cuál modelo sugiere para la variable respuesta?
¿por qué?