

Trabajo regresión lineal múltiple

Estudiantes

**Rojas Martínez, Ivan Santiago
Hernandez Ruiz, Juan Sebastian
Londoño Montoya, Wilson Duván
Pérez García, Pablo**

Docente

Isabel Cristina Ramirez Guevara

Asignatura

Análisis de Regresión



UNIVERSIDAD
NACIONAL
DE COLOMBIA

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 Multiple(MRLM)	3
3.1. Tabla de parámetros ajustados	4
3.2. Ecuación Ajustada	4
3.3. Tabla Anova	4
3.4. Prueba de significancia del Modelo	4
3.5. Coeficiente de determinación R^2 : proporción de la variabilidad total de la respuesta explicada por el modelo y opiniones al respecto	5
4. Coeficientes de regresión estandarizados	5
5. Significancia individual de los parámetros del model	6
5.1. Tabla de la significancia individual de los parámetros	6
5.2. Pruebas de hipotesis	6
6. Ejercicio6	6
7. Ejercicio7	7
8. Residuales estudentizados vs. Valores ajustados	7
8.1. Gráfico de los residuales estudentizados vs. Valores ajustados	7
9. Prueba de normalidad para los residuales estudentizados	8
9.1. Gráfico q-norm residuales estudentizados	9
10.Diagnostico sobre la presencia de observaciones atipicas, de balanceo y/o influenciales y conclusiones	9
11.Ejercicio11	16

12.Ejercicio 12	19
12.1. Matriz de correlación de las variables predictoras	19
12.2. VIF's	19
12.3. Proporciones de varianza	19
13.Ejercicio13	20
13.1. Selección según el R_{adj}^2	21
13.2. Selección según el estadístico C_p	21
13.3. Stepwise	21
13.4. Selección hacia adelante o forward	21
13.5. Selección hacia atrás o backward	21
14.Selección del modelo	21

Índice de figuras

Índice de cuadros

4. Tabla ANOVA para el modelo	4
3. Resumen de los coeficientes	4
5. Multiple R-squared	5
6. Resumen de los coeficientes	6

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

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Con la intencion de validar si dicho modelo es adecuado para

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.

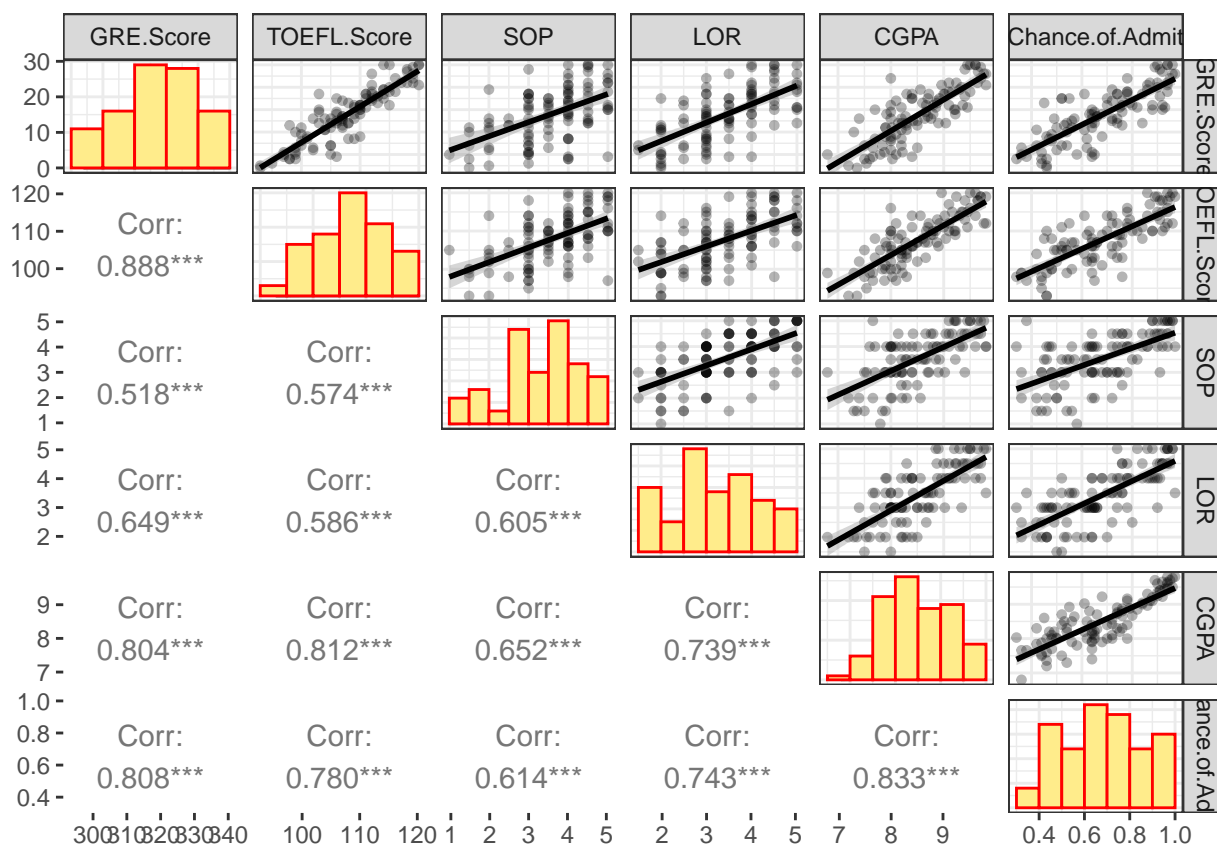
La base de datos disponible en Kaggle corresponde a puntajes de admision creados para la predicción de las admisiones de posgrado en La India. Cuenta con 400 observaciones y 9 variables. De las cuales se consideran los primeros 100 estudiantes y 6 variables de interes por indicación de la docente.

Variables	Descripción
Chance.of.Admit:	Posibilidad de ser admitido.
GRE Score:	Examen que tiene como finalidad medir la capacidad de razonamiento verbal, razonamiento cuantitativo, y habilidades para pensar y escribir de forma analítica.
TOEFL Score:	Prueba estandarizada de dominio del idioma inglés.
SOP:	Ensayo de admisión o solicitud de postgrado.
LOR:	Carta de recomendación.
CGPA:	Promedio general acumulado en el pregrado.

No se consideran	Descripción
Serial No.	Numero de serial que identifica a cada estudiante
University Rating	Calificación universitaria
Research Experience	Si tiene experiencia en investigación o no

2. Análisis descriptivo

2.1. Grafico de dispersión con Matriz de Correlaciones y conclusiones



Se puede observar en la matriz de dispersión se pueden realizar varias observaciones:

-La correlación entre la variable **GRE.Score**(Examen que tiene como finalidad medir la capacidad de razonamiento verbal, razonamiento cuantitativo, y habilidades para pensar y escribir de forma analítica) y **Chance.of.admit** (Posibilidad de ser admitido) es $r = 0.808$, por lo que se puede concluir que tienen una relación lineal positiva fuerte.

-Entre la variable **TOEFL.Score** (Prueba estandarizada de dominio del idioma inglés) y **Chance.of.admit** (Posibilidad de ser admitido) existe una relación lineal positiva moderada debido a que $r = 0.780$.

-Al igual que la anterior, entre la variable **SOP**(Ensayo de admisión o solicitud de postgrado) y **Chance.of.admit** (Posibilidad de ser admitido) existe una relación lineal positiva moderada, en este caso debido a que $r = 0.614$.

-En este caso, existe una relación lineal positiva moderada entre **LOR**(Carta de recomendación) y **Chance.of.admit** (Posibilidad de ser admitido) dado que $r = 0.743$.

-Por último, la relación más fuerte es entre **CGPA**(Promedio general acumulado en el pregrado) y **Chance.of.admit** (Posibilidad de ser admitido) con un $r = 0.833$.

De esta manera, podemos concluir que las variables más significativas son **CGPA** y **GRE.Score**, seguidas por **TOEFL.Score**. Sin embargo, es posible que haya problemas de multicolinealidad, ya que, existe una correlación alta entre estas 3 variables.

3. Modelo Ajustado de Regresion Lineal Multiple(MRLM)

```
##
## Call:
## lm(formula = data$Chance.of.Admit ~ ., data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24114 -0.04987  0.01784  0.05772  0.14308
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.772265   0.300694  -5.894 5.86e-08 ***
## GRE.Score    0.004109   0.001684   2.440  0.01656 *
## TOEFL.Score  0.002912   0.003092   0.942  0.34879
## SOP          0.012040   0.011924   1.010  0.31519
## LOR          0.042831   0.014266   3.002  0.00343 **
## CGPA         0.075708   0.026328   2.876  0.00499 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0847 on 94 degrees of freedom
## Multiple R-squared:  0.7783, Adjusted R-squared:  0.7665
## F-statistic: 65.99 on 5 and 94 DF,  p-value: < 2.2e-16
```

Cuadro 4: Tabla ANOVA para el modelo

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FO(GRE.Score, TOEFL.Score, SOP, LOR, CGPA)	5	2.3674478	0.4734896	65.99381	0
Residuals	94	0.6744272	0.0071748	NA	NA

3.1. Tabla de parámetros ajustados

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

3.2. Ecuación Ajustada

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

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

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

3.3. Tabla Anova

Pruebas de hipótesis: $H_0 : B_1 = B_2 = B_3 = B_4 = B_5 = 0$ vs $H_A : \text{Algún } B_j \neq 0$
 Para al menos un $j, j = 1, 2, 3, 4, 5$ p-value = $2.2e-16 < \alpha$. Hay evidencia para rechazar H_0 .
 Por lo tanto el modelo es globalmente significativo, al menos una de las pruebas ayuda a explicar la variabilidad de las chances de admisión

3.4. Prueba de significancia del Modelo

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

Cuadro 5: Multiple R-squared

x
0.7782857

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)

Coeficientes estimados, sus I.C, Vifs y Coeficientes estimados estandarizados

	Estimación	Limites.2.5..	Limites.97.5..	Vif	Coef.Std
(Intercept)	-1.7722651	-2.3693009	-1.1752294	0.000000	0.0000000
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**, respectivamente

5. Significancia individual de los parámetros del model

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 6: 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

5.2. Pruebas de hipotesis

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

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

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

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

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

6. Ejercicio6

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.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
97	1.0661892	NA	NA	NA	NA
94	0.6744272	3	0.391762	18.20094	0

7. Ejercicio7

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?

	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

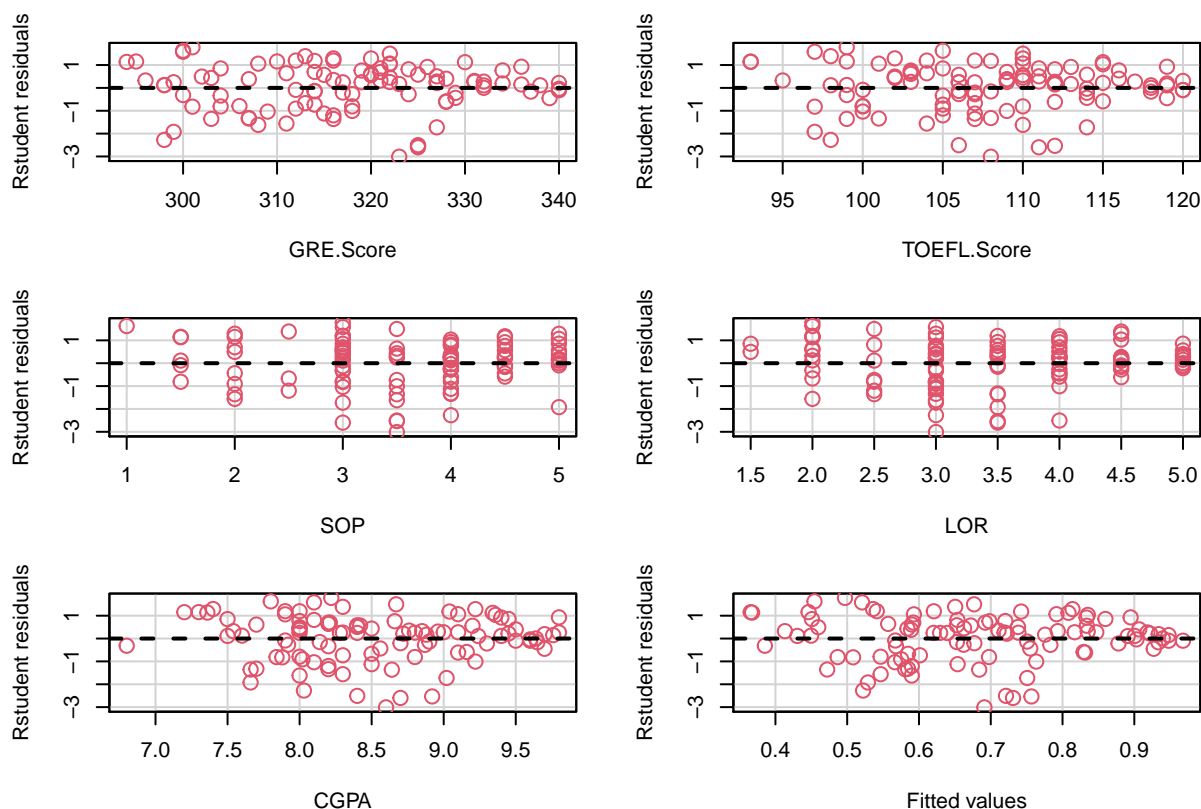
	Sum Sq	Df	F value	Pr(>F)
GRE.Score	0.0427161	1	5.9536667	0.0165615
TOEFL.Score	0.0063619	1	0.8867053	0.3487854
SOP	0.0073158	1	1.0196631	0.3151912
LOR	0.0646740	1	9.0141001	0.0034318
CGPA	0.0593270	1	8.2688553	0.0049890
Residuals	0.6744272	94	NA	NA

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

Construya una gráfica de probabilidad normal para los residuales estudentizados. ¿Existen razones para dudar de la hipótesis de normalidad sobre los errores en este modelo?

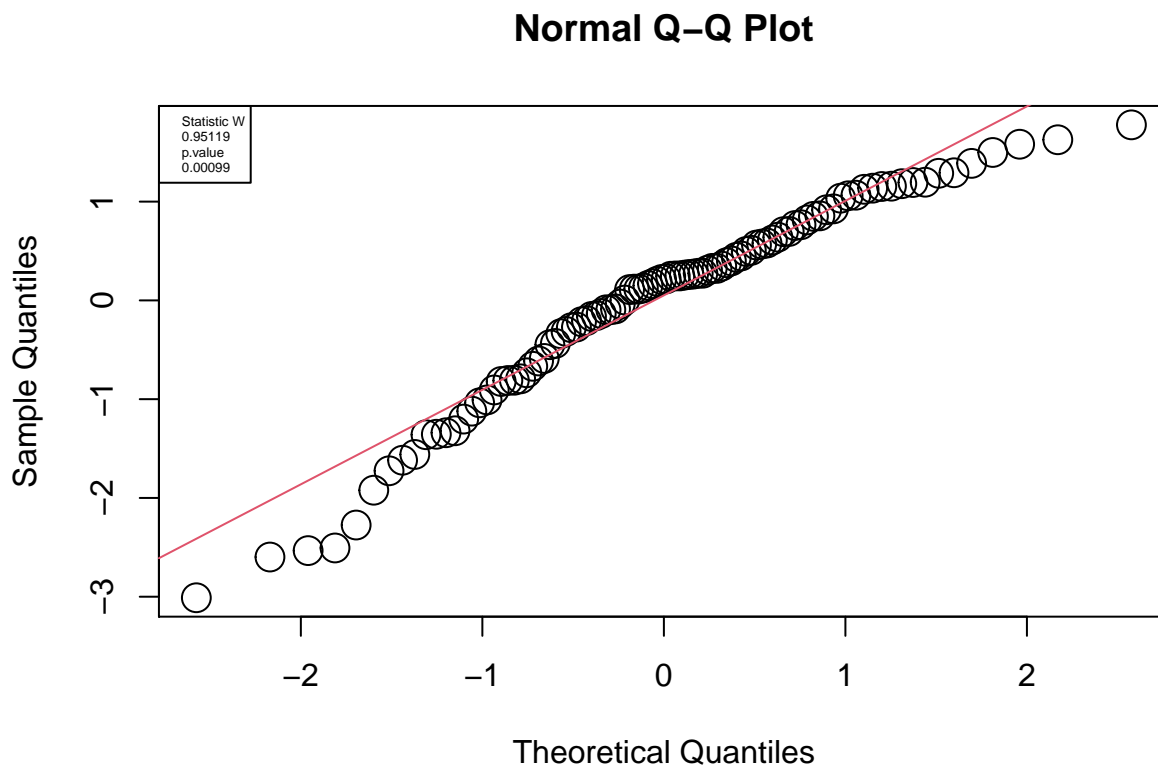


En los gráficos de las variables GRE.Score, TOEFL.Score y CGPA, además de los valores ajustados no se observa ningún tipo de patrón, por lo tanto se cumple el supuesto de varianza constante y no hay observaciones atípicas.

9. Prueba de normalidad para los residuales estudentizados

Construya una gráfica de probabilidad normal para los residuales estudentizados. ¿Existen razones para dudar de la hipótesis de normalidad sobre los errores en este modelo?

9.1. Gráfico q-norm residuales estudentizados



Aunque muchos datos 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 (los datos tienen comportamiento normal), concluyendo de esta manera que

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

```
## Influence measures of
## 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

```



```

## 71  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 72  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 73  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 74  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 75  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 76  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 77  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 78  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 79  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 80  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 81  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 82  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 83  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 84  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 85  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 86  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 87  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 88  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 89  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 90  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 91  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 92  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  TRUE  FALSE  FALSE  FALSE
## 93  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 94  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 95  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 96  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 97  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 98  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 99  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE
## 100 FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE  FALSE

```

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

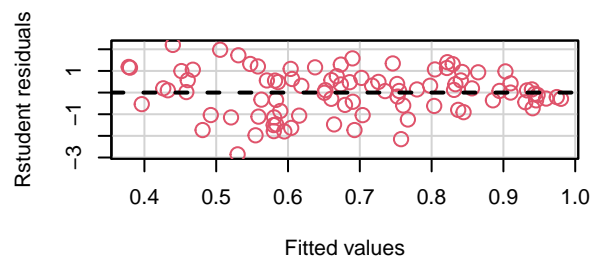
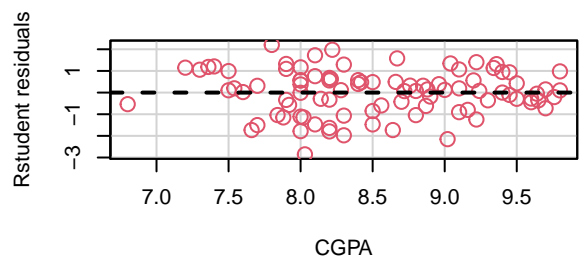
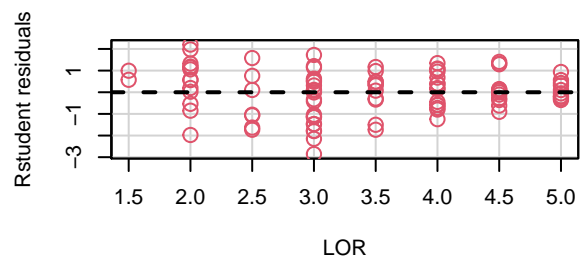
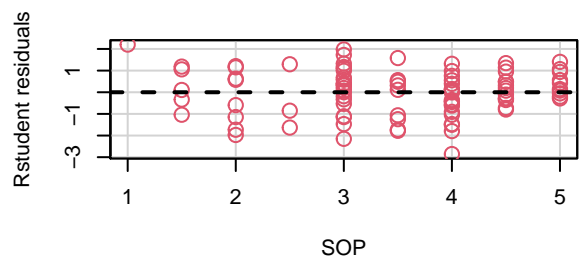
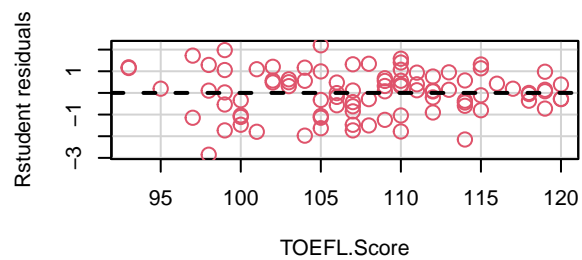
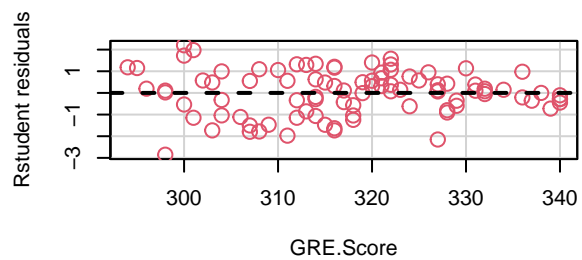
```

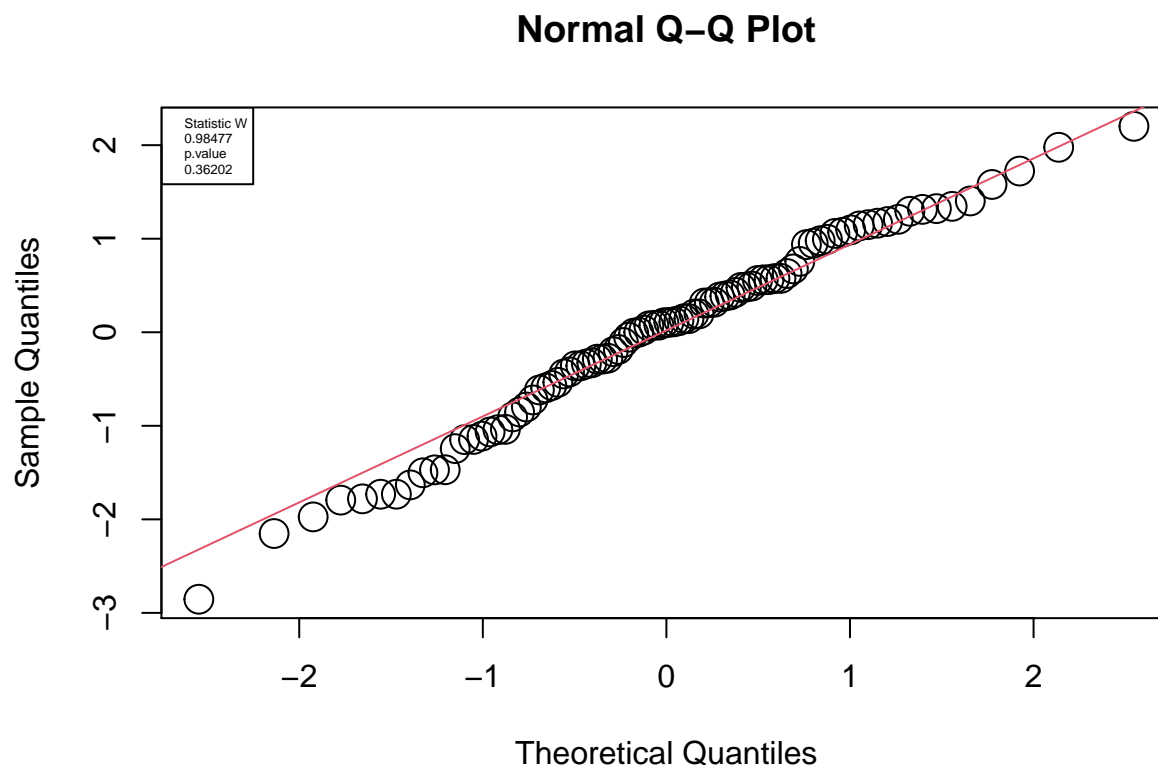
```

##                                Df  Sum Sq Mean Sq F value    Pr(>F)
## FO(GRE.Score, TOEFL.Score, SOP, LOR, CGPA)  5 2.35558 0.47112  90.573 < 2.2e-16
## Residuals                                86 0.44733 0.00520
##
## 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
## -0.189817 -0.041224  0.007218  0.045167  0.140378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.9798608  0.2948862  -6.714 1.93e-09 ***
## GRE.Score    0.0057790  0.0016903   3.419 0.000963 ***
## TOEFL.Score  0.0002077  0.0028940   0.072 0.942945
## SOP          0.0153807  0.0108883   1.413 0.161385
## LOR          0.0403174  0.0125305   3.218 0.001823 **
## CGPA         0.0728150  0.0241848   3.011 0.003420 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07212 on 86 degrees of freedom
## Multiple R-squared:  0.8404, Adjusted R-squared:  0.8311
## F-statistic: 90.57 on 5 and 86 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

12.2. VIF's

12.3. Proporciones de varianza

##	GRE.Score	TOEFL.Score	SOP	LOR	CGPA
## GRE.Score	1.0000000	0.8883788	0.5183744	0.6493865	0.8040898
## TOEFL.Score	0.8883788	1.0000000	0.5739543	0.5855924	0.8117425
## SOP	0.5183744	0.5739543	1.0000000	0.6050348	0.6515129
## LOR	0.6493865	0.5855924	0.6050348	1.0000000	0.7393324
## CGPA	0.8040898	0.8117425	0.6515129	0.7393324	1.0000000
## Chance.of.Admit	0.8078850	0.7800010	0.6136879	0.7428750	0.8326816
##	Chance.of.Admit				

```
## GRE.Score          0.8078850
## TOEFL.Score        0.7800010
## SOP                0.6136879
## LOR                0.7428750
## CGPA               0.8326816
## Chance.of.Admit    1.0000000
```

```
## GRE.Score TOEFL.Score      SOP      LOR      CGPA
## 5.691210 5.858051 1.928844 2.519579 4.615227
```

```
## GRE.Score TOEFL.Score      SOP      LOR      CGPA
## GRE.Score 1.0000000 0.9061621 0.5769656 0.6945897 0.8381136
## TOEFL.Score 0.9061621 1.0000000 0.6269824 0.6213013 0.8266527
## SOP 0.5769656 0.6269824 1.0000000 0.6153549 0.6986389
## LOR 0.6945897 0.6213013 0.6153549 1.0000000 0.7571056
## CGPA 0.8381136 0.8266527 0.6986389 0.7571056 1.0000000
## Chance.of.Admit 0.8595873 0.8077359 0.6683703 0.7845642 0.8677695
## Chance.of.Admit
## GRE.Score 0.8595873
## TOEFL.Score 0.8077359
## SOP 0.6683703
## LOR 0.7845642
## CGPA 0.8677695
## Chance.of.Admit 1.0000000
```

```
## GRE.Score TOEFL.Score      SOP      LOR      CGPA
## 7.229071 6.690216 2.147637 2.640310 5.224542
```

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):

- 13.1. Selección según el R_{adj}^2
 - 13.2. Selección según el estadístico C_p
 - 13.3. Stepwise
 - 13.4. Selección hacia adelante o forward
 - 13.5. Selección hacia atrás o backward
14. Selección del modelo

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