

1. Regresión lineal múltiple.

i) Modelo de RLM reducido para E(bpsystol; bmi, sex, age) con datos originales.

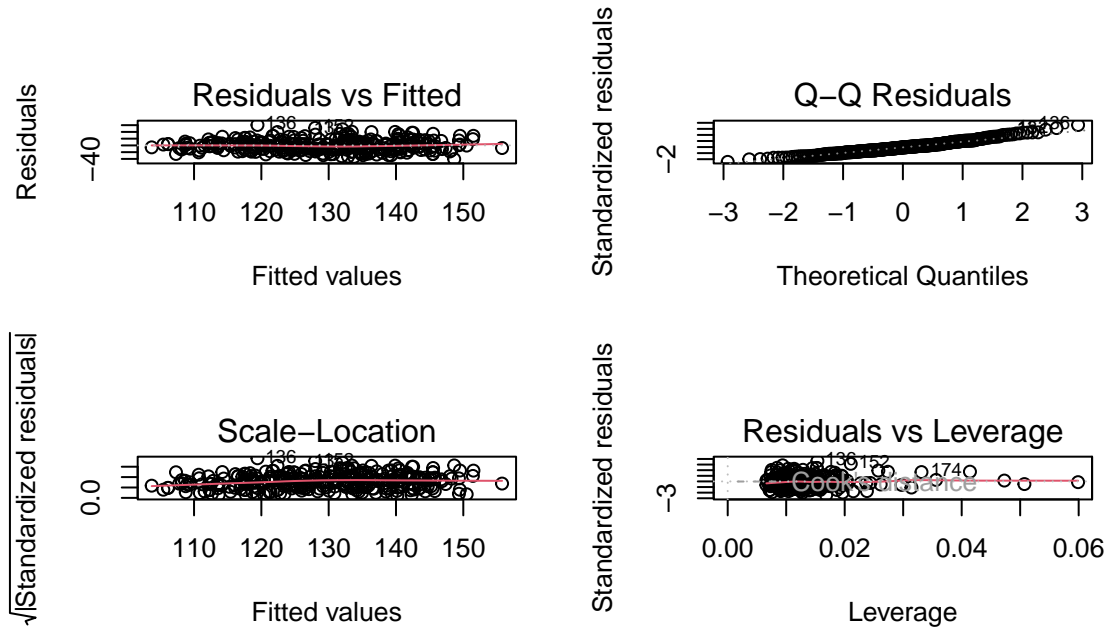
Para analizar si existe una asociación entre la presión arterial sistólica (bpsystol) como variable dependiente y el índice de masa corporal (bmi), ajustaremos un modelo de regresión lineal múltiple, considerando el sexo (sex: 1-hombre y 2-mujer con nivel de referencia hombre) y la edad (age) de los pacientes. Para ello usaremos la base de datos `reg1B.csv` con 295 pacientes, 142 hombres y 153 mujeres, de entre 20 y 74 años. En el siguiente Cuadro se muestran los resultados del modelo planteado, sin pretratamiento de los datos.

La prueba global F muestra un p-value menor a 0.05, por lo que rechazamos la hipótesis nula de que los parámetros estimados son cero, es decir, podemos decir que al menos un coeficiente estimado es distinto de cero, por lo que el modelo es estadísticamente significativo al nivel de confianza del 95%. Las pruebas individuales también rechazan la hipótesis nula con la prueba t – *student*, es decir, todos los coeficientes son significativos al 5% de significancia estadística, pues rechaza la hipótesis nula de que son en lo individual cero.

Table 1:

	Dependent variable:
	bpsystol
bmi	1.208*** s.e: (0.202) t value: 5.995 Pr(> t): 6.02e-09
sex2	-5.664*** s.e: (1.964) t value: -2.884 Pr(> t): 0.00421
age	0.484*** s.e: (0.059) t value: 8.264 Pr(> t): 5.03e-15
Constant	84.160*** s.e: (6.037) t value: 13.942 Pr(> t): < 2e-16
Observations	295
R ²	0.310
Adjusted R ²	0.302
Residual Std. Error	16.784 (df = 291)
F Statistic	43.497*** (df = 3; 291); p-value: < 2.2e-16
Note:	*p<0.1; **p<0.05; ***p<0.01

Para poder tener una interpretación válida de los coeficientes, veremos si el modelo cumple con los supuestos del modelo de regresión lineal. La Gráfica **Residuals vs Fitted**, se utiliza para comprobar los supuestos de relación lineal, una línea horizontal, sin patrones distintos, es indicación de una relación lineal, lo que es bueno en nuestro caso. La Gráfica **Normal Q-Q Residuals**, se utiliza para examinar si los residuos se distribuyen normalmente, es bueno que los puntos residuales sigan la línea recta discontinua, en nuestro caso parece que no todo se ajusta bien, pues tenemos muchos vaores que no siguen la linea. La Gráfica **Scale-Location**, se utiliza para comprobar la homogeneidad de la varianza de los residuos (homoscedasticidad), la línea horizontal con puntos igualmente distribuidos es una buena indicación de homoscedasticidad, este es el caso en nuestro ejemplo, donde no tenemos un problema de heterocedasticidad. La Gráfica **Residuals vs Leverage**, se utiliza para identificar casos de valores influyentes, es decir, valores extremos que podrían influir en los resultados de la regresión cuando se incluyen o excluyen del análisis, al parecer ningún valor sale de la distancia de Cook.



En el siguiente Cuadro se pueden observar las pruebas de Shapiro-Wilk, Breusch-Pagan y Durbin-Watson, que plantean la hipótesis nulas de normalidad, homoscedasticidad y no autocorrelación, respectivamente. Se concluye que el modelo presenta no autocorrelación y homoscedasticidad, sin embargo no presenta normalidad de los errores. Por lo que tendremos que hacer algunos ajustes al modelo, con algunos tratamientos a las variables.

	1
Normality (Shapiro-Wilk)	0.001
Homoscedasticity (Breusch-Pagan)	0.095
Autocorrelation of residuals (Durbin-Watson)	0.981

ii) Modelo adecuado con transformación de datos.

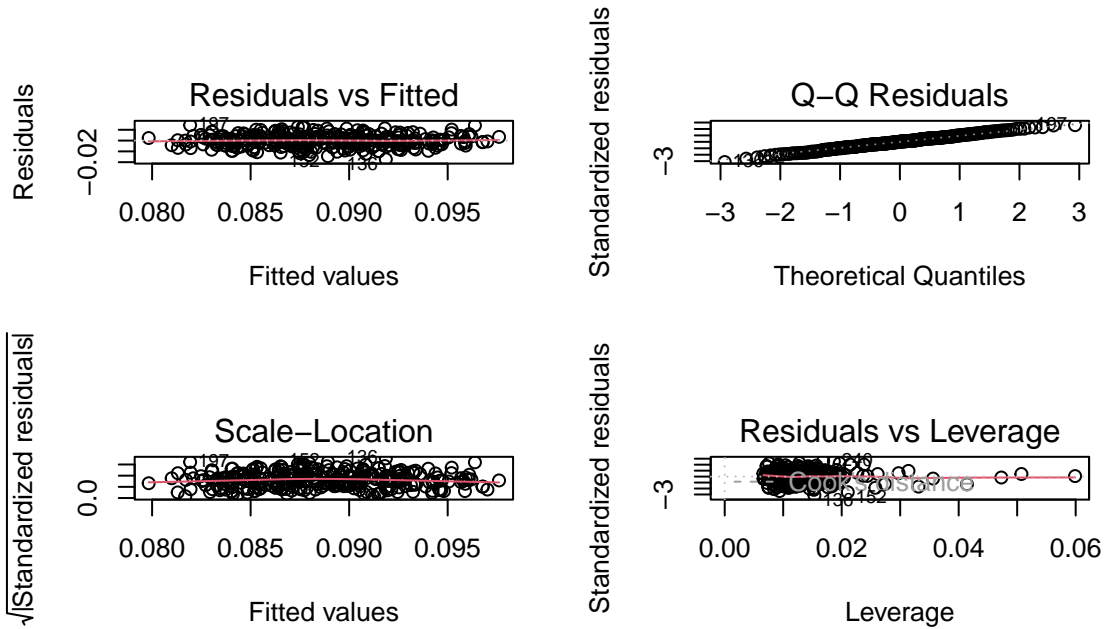
De acuerdo con la prueba `powerTransform` para transformación Box-Cox, se tiene un valor de -0.56588677520915 , por lo que podemos transformar la variable dependiente con un exponente de $-\frac{1}{2}$.

Por simplicidad no consideraremos en el modelo interacciones entre las variables y se propone una transformación Box Cox de la variable dependiente. Para este modelo, se observa en el siguiente Cuadro que la prueba global F rechaza la hipótesis nula, y las pruebas t - *student* individuales de los coeficientes estimados también rechaza las hipótesis nulas de manera individual e independiente.

En las siguientes gráficas podemos observar en **Residuals vs Fitted** que se conserva la linealidad, en **Q-Q Residuals** se observa una mejora con respecto a la normalidad de los errores, en **Scale-Location** se observa que hay homoscedasticidad, y en **Residuals vs Leverage** parece no haber outliers influyentes.

Table 3:

	Dependent variable:
	I(bpsystol ^{-1/2})
bmi	-0.0004155*** s.e: (6.686e-05) t value: -6.214 Pr(> t): 1.78e-09
sex2	0.002292*** s.e: (6.514e-04) t value: 3.518 Pr(> t): 0.000504
age	-0.0001636*** s.e: (1.943e-05) t value: -8.421 Pr(> t): 1.73e-15
Constant	0.1059*** s.e: (1.828e-03) t value: 57.927 Pr(> t): < 2e-16
Observations	295
R ²	0.325
Adjusted R ²	0.318
Residual Std. Error	0.006 (df = 291)
F Statistic	46.691*** (df = 3; 291); p-value: < 2.2e-16
Note:	*p<0.1; **p<0.05; ***p<0.01



En el siguiente Cuadro, se muestra las pruebas de Shapiro-Wilk, Breusch-Pagan y Durbin-Watson, que plantean la hipótesis nulas de normalidad, homoscedasticidad y no autocorrelación, respectivamente. En todos los casos no hay evidencia suficiente para rechazar las hipótesis nulas.

	1
Normality (Shapiro-Wilk)	0.981
Homoscedasticity (Breusch-Pagan)	0.244
Autocorrelation of residuals (Durbin-Watson)	0.972

iii) Asociación entre masa corporal y presión arterial sistólica.

iv) Gráfica resumen con la estimación puntual de la relación bpsystol y bmi.

A continuación presentaremos una gráfica resumen con la estimación puntual asociada a la relación entre bpsystol y bmi. Para esto consideremos sólo tres posibles edades: 30, 50 y 64, así como la diferenciación entre mujeres y hombres. El comportamiento en general es que los hombres tienden a tener una mayor presión arterial sistólica, comparado con las mujeres. En todos los casos al aumentar la masa corporal, la presión arterial sistólica incrementa tanto para hombres como para mujeres. Además podemos observar que a mayor edad, es mayor la presión arterial sistólica tanto para hombres como para mujeres.

