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 cuadro de MODELOS se muestran los resultados del Modelo 1 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 preba $t - student$, es decir, todos los coeficientes son significativos al 5%, pues se rechaza la hipótesis nula de que en lo individual sean iguales a cero.

En el siguiente Cuadro se pueden observar las pruebas de Shapiro-Wilk, Breusch-Pagan y Durbin-Watson para el Modelo 1, que plantean la hipótesis nulas de normalidad, homoscedasticidad y no autocorrelación, respectivamente. Se concluye que el Modelo 1 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

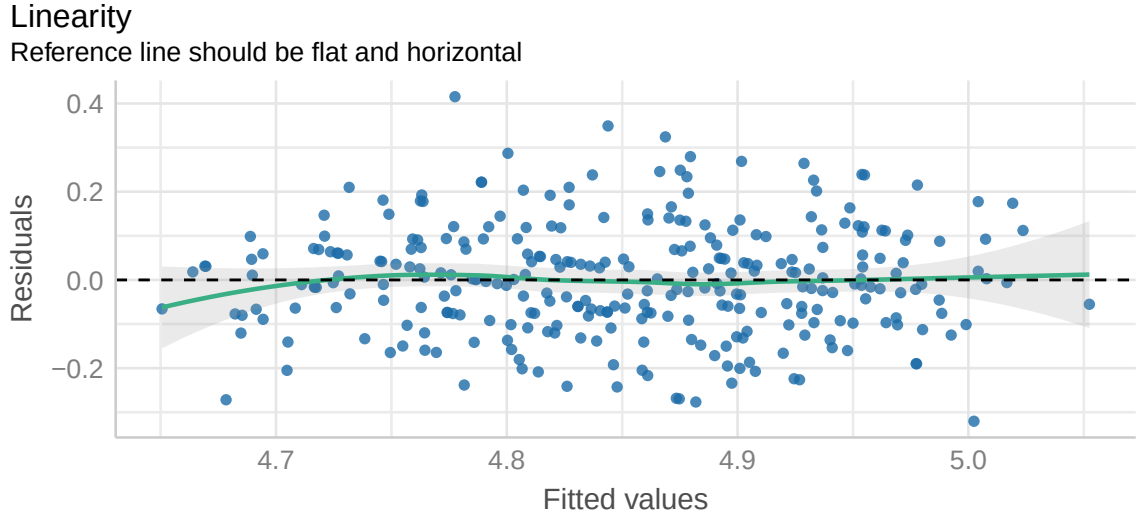
ii) Modelo adecuado con transformación de datos.

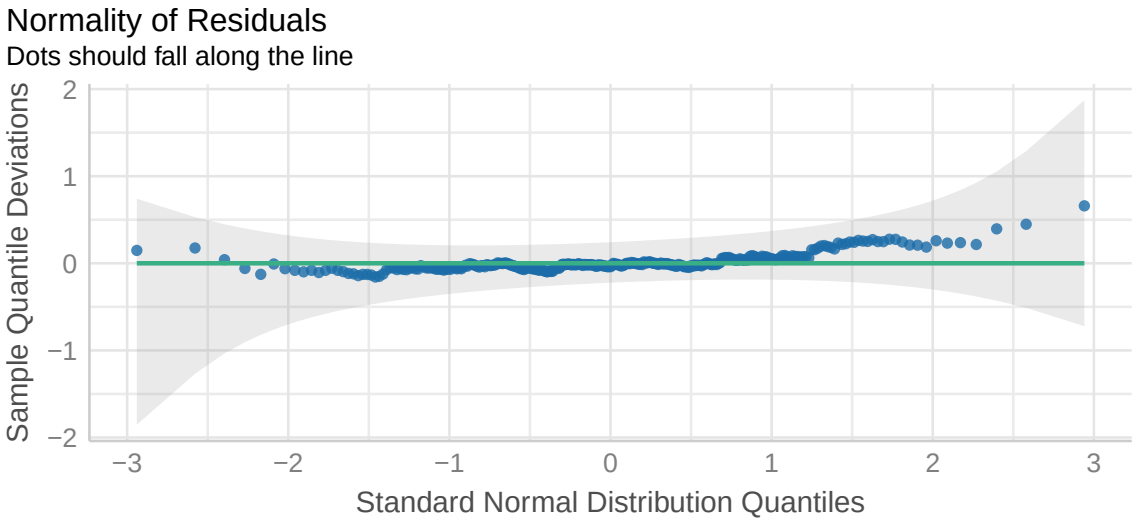
Como tenemos un problema con la normalidad, procederemos a hacer primero una transformación a la variable dependiente, probaremos con una transformación más usual que es la logarítmica, la cual se puede interpretar más fácilmente. Por simplicidad no consideraremos en el Modelo 2 interacciones entre las variables y se propone una transformación Box Cox logarítimica de la variable dependiente. Para este Modelo 2, se observa en el Cuadro de MODELOS que la prueba global F rechaza la hipótesis nula, por lo que al menos un coeficiente estimado es distinto de cero, y las pruebas $t - student$ individuales de los coeficientes estimados también rechazan las hipótesis nulas analizados individualmente. Notemos que *** implica que se rechaza la hipótesis nula incluso con un nivel de confianza del 99%, el p-value es menor a 0.01. Además al comparar los AIC, tenemos para el Modelo 1 es de 2507.213 y para el Modelo 2, considerando la transformación inversa del logaritmo, es de $e^{-376.2099}$ lo cual es cercano a 0, esto favorece la alección del Modelo 2.

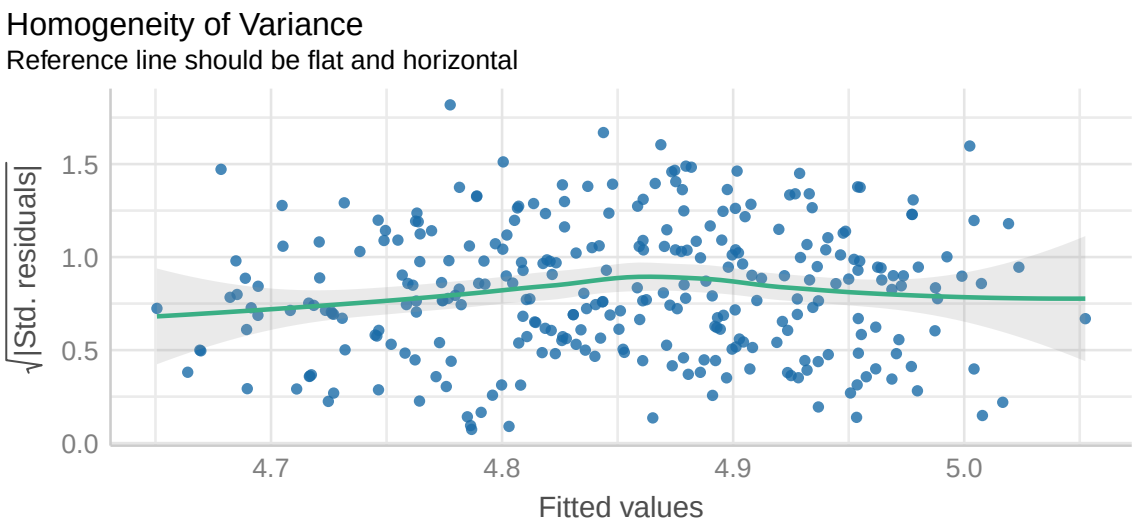
Para poder tener una interpretación válida de los coeficientes, veremos que el Modelo 2 cumple con los supuestos del modelo de regresión lineal. Primero se mostrarán algunas gráficas. La Gráfica **Residuals vs Fitted Values**, se utiliza para comprobar los supuestos de relación lineal, una línea horizontal, sin patrones distintos, es indicación de una relación lineal, lo que es bueno en nuestro caso. La Gráfica **Sample Q Deviation vs Standard Normal Distribution Q**, se utiliza para examinar si los residuos se distribuyen normalmente, es bueno que los puntos residuales sigan la línea recta, en nuestro caso parece que todo se ajusta bien, pues tenemos muchos valores que siguen la linea. La Gráfica **Scale-Location: Sqrt(|Std. Residuals|) vs Fitted values**, 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 homocedasticidad, este es el caso en nuestro modelo, donde no tenemos un problema de heterocedasticidad. La Gráfica **Std. 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.

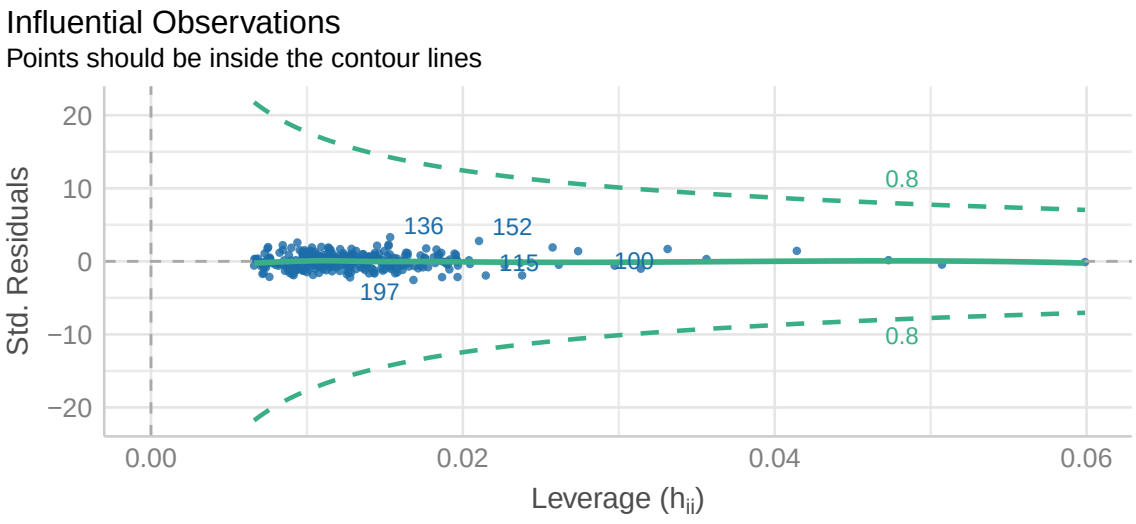
Table 2: MODELOS

	Dependent variable:	
	bpsystol	I(log(bpsystol))
	(1)	(2)
bmi	1.208***	0.009***
	(0.202)	(0.002)
sex2	−5.664***	−0.049***
	(1.964)	(0.015)
age	0.484***	0.004***
	(0.059)	(0.0004)
Constant	78.496***	4.461***
	(5.510)	(0.042)
Observations	295	295
R^2	0.310	0.321
Adjusted R^2	0.302	0.314
Residual Std. Error (df = 291)	16.784	0.127
F Statistic (df = 3; 291)	43.497***	45.922***
AIC:	2507.213	−376.2099
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 para el Modelo 2, 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)	0.596
Homoscedasticity (Breusch-Pagan)	0.254
Autocorrelation of residuals (Durbin-Watson)	0.975

iii) Asociación entre masa corporal y presión arterial sistólica.

Se puede concluir que para una persona de cierta edad y sexo, tener un índice de masa corporal alto se asocia con una alta presión arterial sistólica. Tomando en cuanta en el Cuadro anterior que se rechaza la hipótesis nula $H_0 : \beta_1 = 0$ contra la anternativa de que $H_a : \beta_1 \neq 0$, a continuación se plantea la prueba de hipótesis con dirección, en donde la hipótesis nula es $H_0 : \beta_1 < 0$ y la alternativa $H_a : \beta_1 > 0$.

El resultado de la prueba con dirección `Simultaneous Tests for General Linear Hypotheses` con el ajuste $lm(formula = I(log(bpsystol)) \sim bmi + sex + age, data = datos)$ muestra un p-value de $1.2e-09$, lo cual rechaza la hipótesis nula planteada. Por lo tanto, para una persona de cierta edad y sexo, tener un índice de masa corporal alto se asocia con una alta 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.