

Practica

Agustin Riquelme y Heriberto Espino

2023-09-06

Con el objetivo de sintetizar las líneas de código lo más posible, se ha creado una función que requerirá como argumentos la variable de respuesta y la variable predictora para el modelo, así como el nivel de significancia α . Esta función se utilizará a lo largo del presente documento. Con el fin de visualizar los datos de manera más amigable, el código de la función se ocultará en el informe, pero se mantendrá en formato markdown para su correcta ejecución.

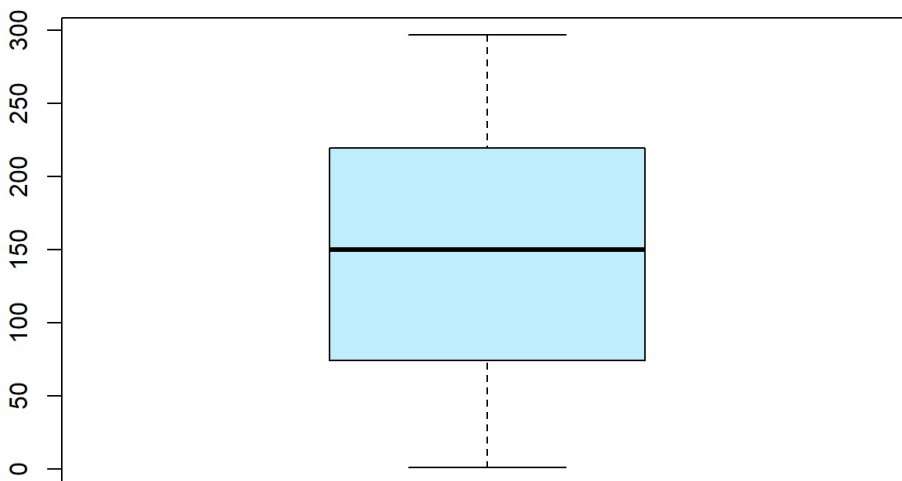
```
## New names:
## • `` -> `...1`
```

A continuación se presentarán los resultados obtenidos para cada modelo de regresión lineal simple. En estos análisis, la variable “sales” se considera como la variable de respuesta, mientras que los gastos de marketing en cada una de las plataformas se tratan como las variables predictoras. Esto se debe a que las ventas pueden estar relacionadas con el medio a través del cual se realiza el marketing. En este contexto, podemos utilizar estos modelos para predecir las ventas en función de una campaña de marketing ya propuesta y determinar si es óptima o no.

Para el modelo de regresión que involucra las ventas y el marketing en televisión, los resultados son los siguientes:

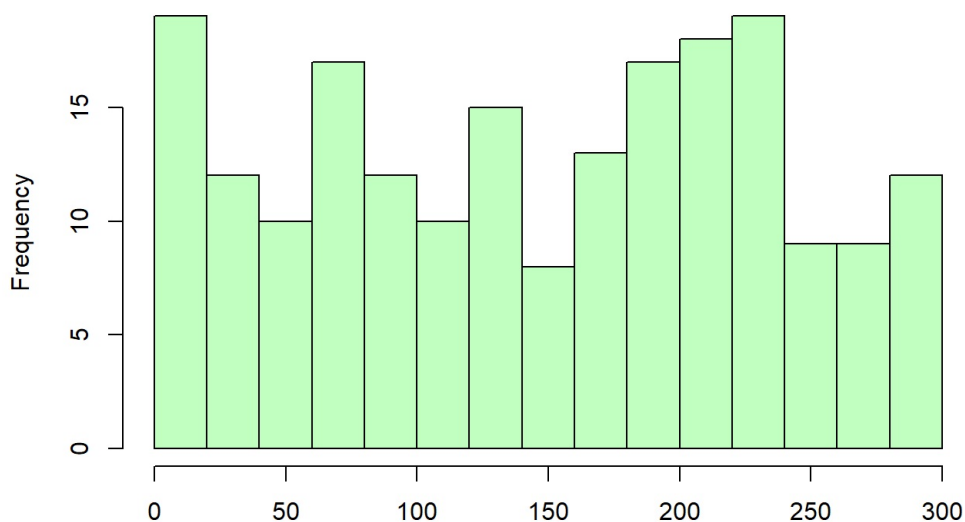
```
linealmodel(tv, sales, 0.05)
```

Boxplot x



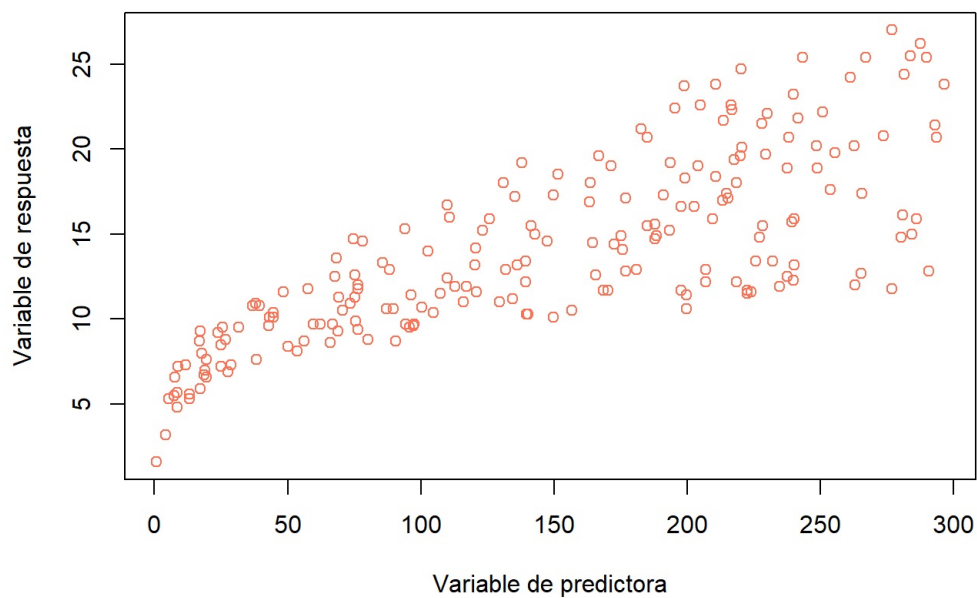
Variable de predictora

Histograma x

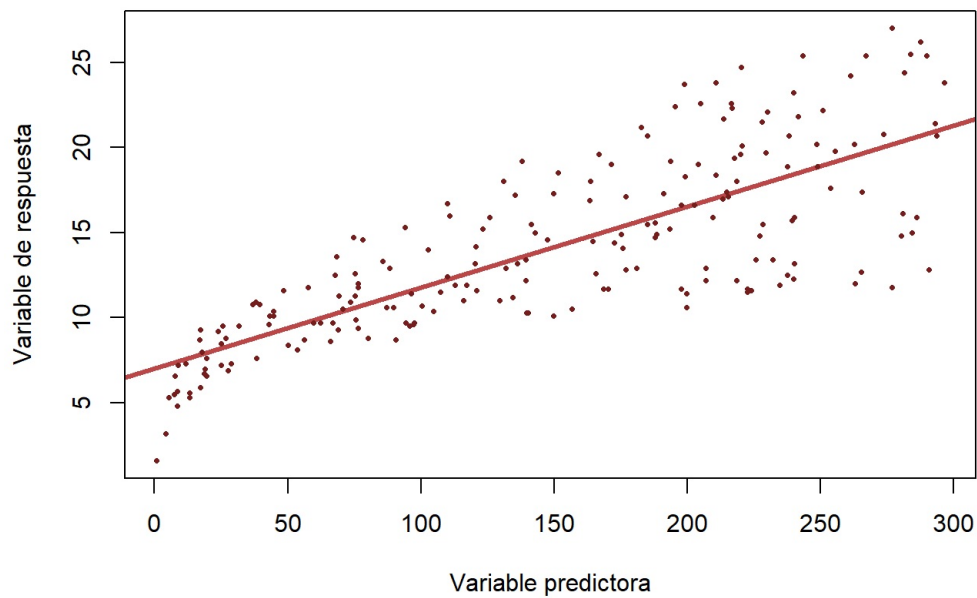


Variable de predictora

Gráfico dispersión



Recta de regresión



Intervalos de confianza

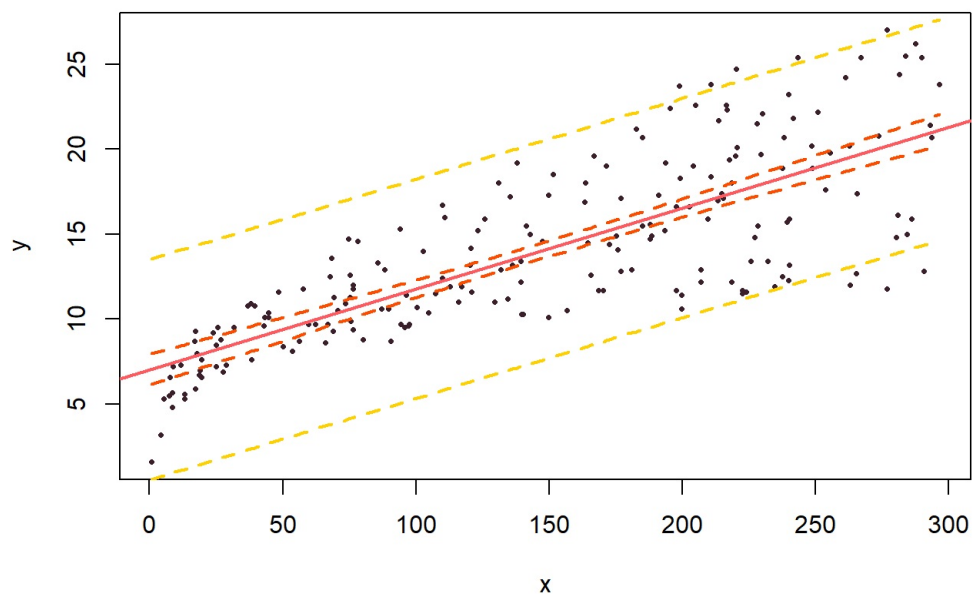
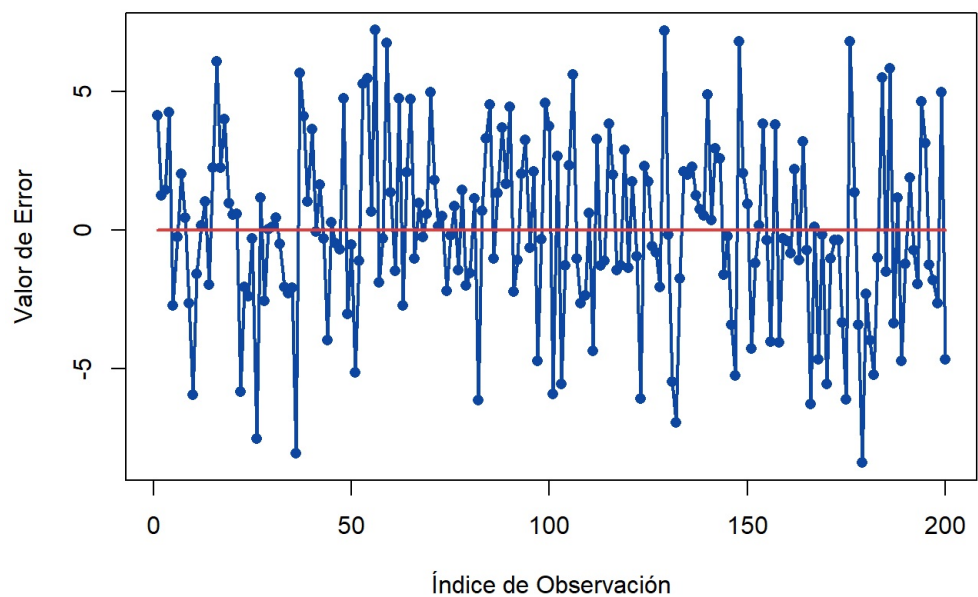
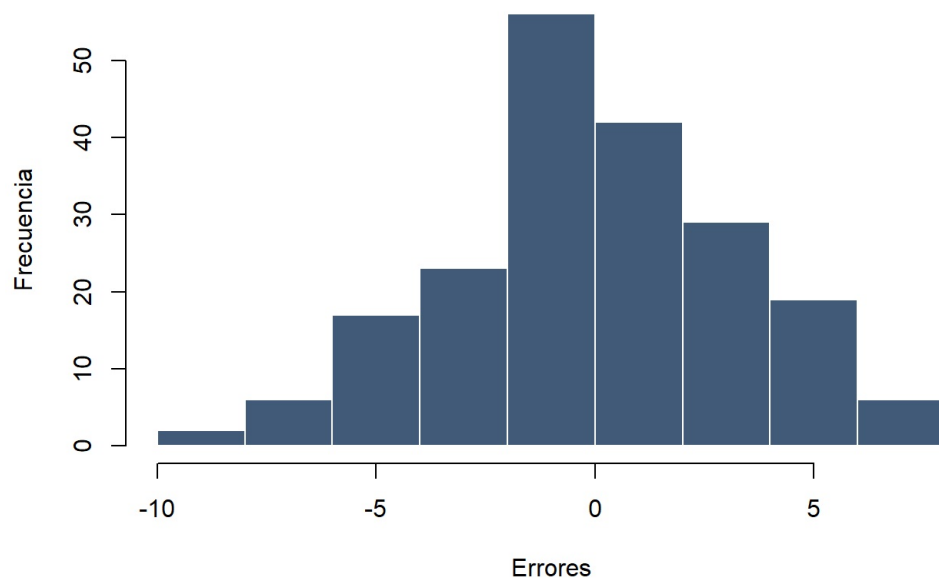


Gráfico de Errores



Histograma de Errores



Boxplot

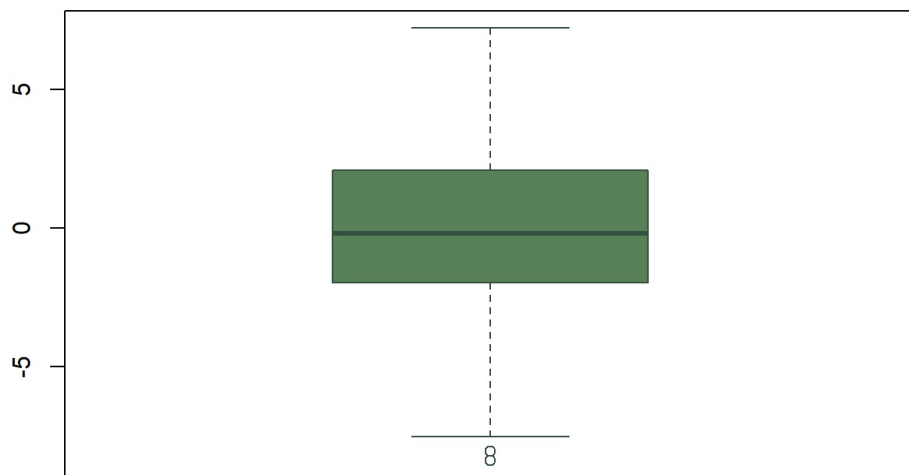
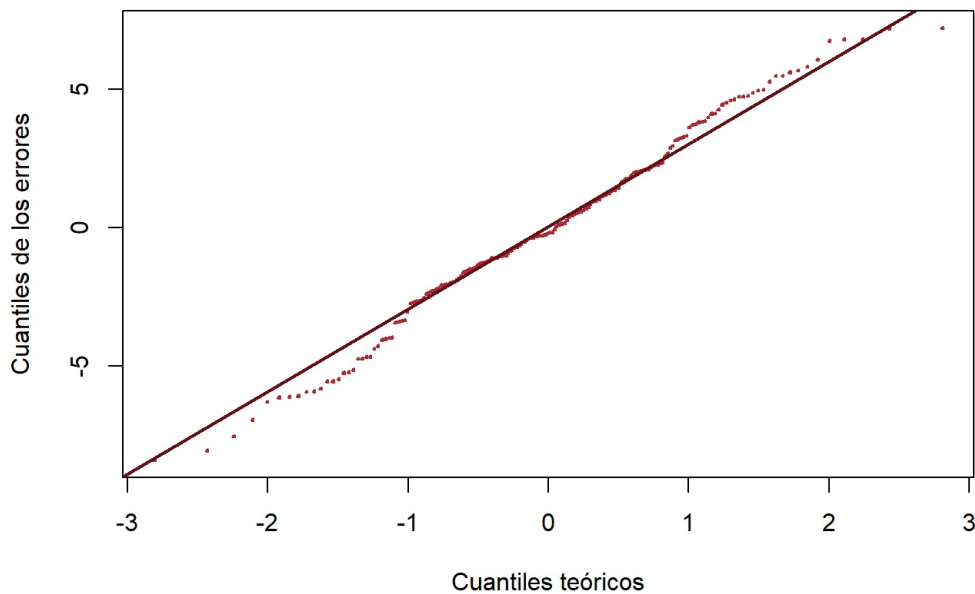


Gráfico QQ de Errores



Informacion del modelo

Datos	Valores
β_0	7.0325935491277
β_1	0.0475366404330198
S	3.25865636865046
S cuad	10.6188413289462
cv	2.21613232136999
df	198
qti	-1.97201747783631
qtd	1.97201747783631
tc	17.6676256008755
p value	1.46738970019465e-42
Int. conf.	0.0422307160326923
	0.0528425648333472
Prom. err.	2.54980603892749
Var. err.	4.03129857236752
P. Hip. β_1 Est.	Rechazamos H0
P. Hip. β_1 Pvalue	Rechazamos H0

Las tres primeras gráficas representan nuestro vector que almacena todos los presupuestos publicitarios (en miles de dólares) para televisión. Al observar solo el boxplot, podríamos inferir que la distribución de los datos es normal, ya que el 50% de los datos se encuentra en el centro, la mediana está en el centro del cuadrado y el valor máximo y mínimo están a la misma distancia que los cuantiles. Sin embargo, al realizar el histograma de los datos, comprobamos que no se trata de la distribución que inicialmente pensábamos. No obstante, en el tercer gráfico, podemos afirmar que existe una correlación lineal positiva entre las ventas y el presupuesto destinado a la televisión. Finalmente, en el cuarto gráfico, podemos corroborar que la recta de regresión se ajusta a nuestras expectativas.

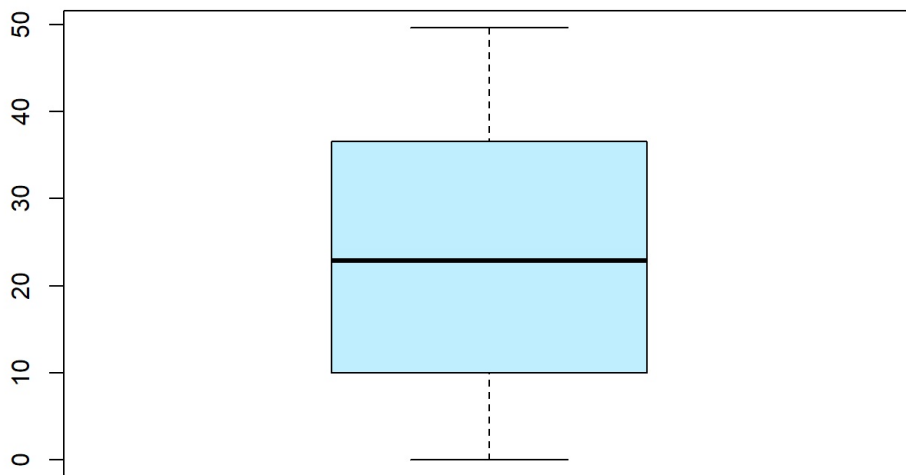
Después de estos gráficos, tenemos el de los intervalos de confianza para los valores medios y las predicciones de la variable de respuesta. Podemos observar que a medida que los valores crecen, el modelo ya no se ajusta de manera consistente, ya que algunos datos en la muestra se encuentran fuera de los intervalos de confianza. Si observamos el gráfico de dispersión que muestra la relación entre las variables, podemos notar que si trazáramos dos líneas rectas que encerraran todos los puntos, estas líneas no tendrían la misma pendiente y tendrían forma de cono. Esto indica que a medida que los valores del modelo aumentan, se separan más entre sí, lo que sugiere que el modelo no se ajusta adecuadamente para valores altos.

Por otro lado, los gráficos de los errores muestran un error máximo aproximado de 8, como se puede apreciar en el gráfico de líneas. Al calcular el promedio de los errores, obtenemos un valor de 2.54, que es relativamente bajo. Esto podría indicar que, a pesar de la falta de ajuste en los datos más grandes, el modelo sigue generando resultados precisos para los demás valores. La distribución de los errores muestra un sesgo hacia la derecha, y en el gráfico qq, los residuos intermedios del modelo se ajustan más a una distribución normal.

Nuestros coeficientes beta son la base para crear nuestra línea de regresión. Con un valor bajo de beta 1, indicamos que a medida que nuestra variable predictora aumenta, la variable de respuesta también lo hace, pero en una proporción mucho menor. La varianza residual es de 10, y tenemos un coeficiente de variación de 2, lo que sugiere que no hay una gran variabilidad relativa en comparación con nuestra media. El intervalo de confianza de beta 1 nos informa sobre el intervalo en el que se encuentra beta 1 con un nivel de significancia alpha. Como podemos observar, este intervalo no incluye el valor cero, lo que nos permite concluir con mayor confianza que rechazamos la hipótesis nula, indicando que beta 1 es diferente de cero.

```
linealmodel(radio, sales, 0.05)
```

Boxplot x



Variable de predictora

Histograma x

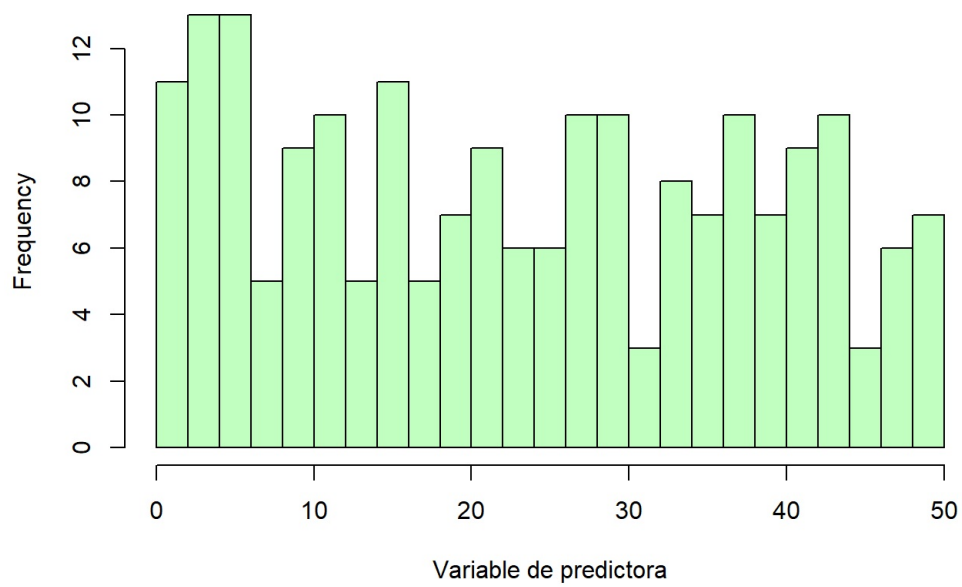
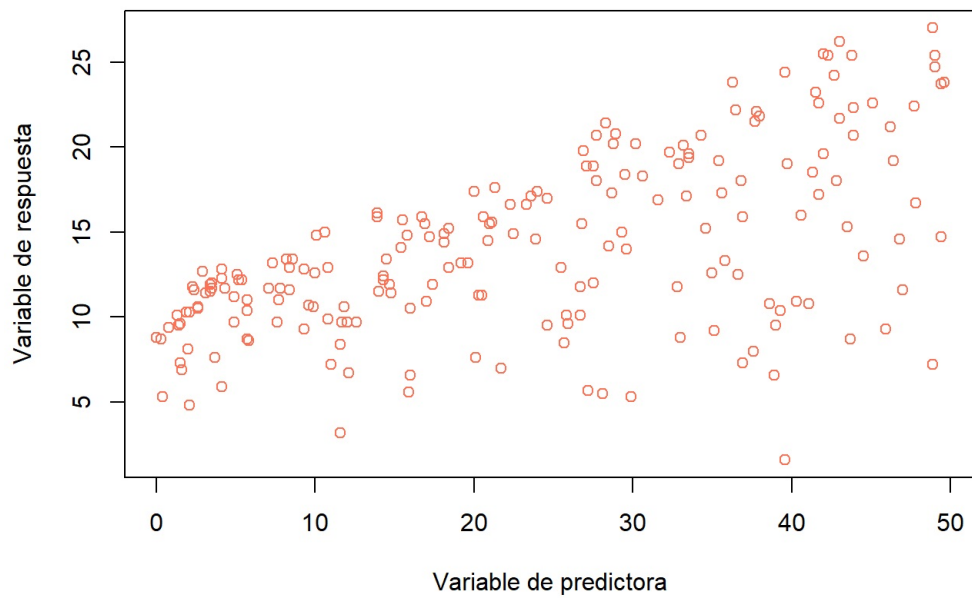
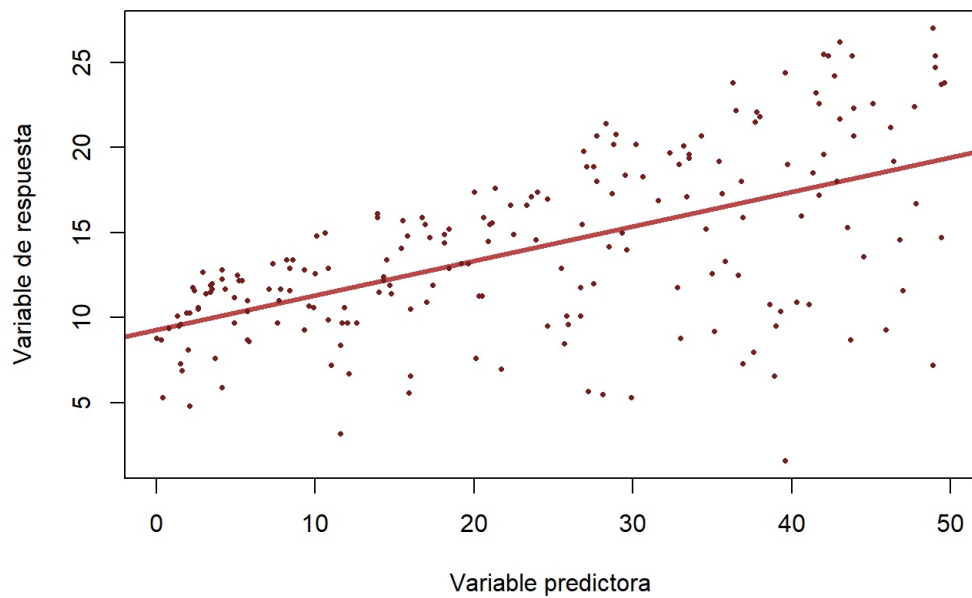


Gráfico dispersión



Recta de regresión



Intervalos de confianza

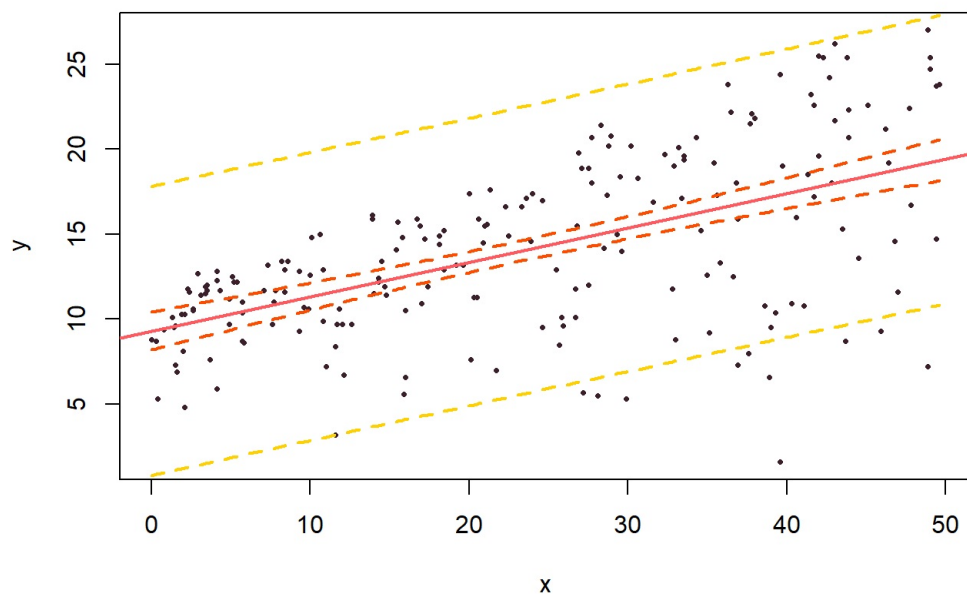
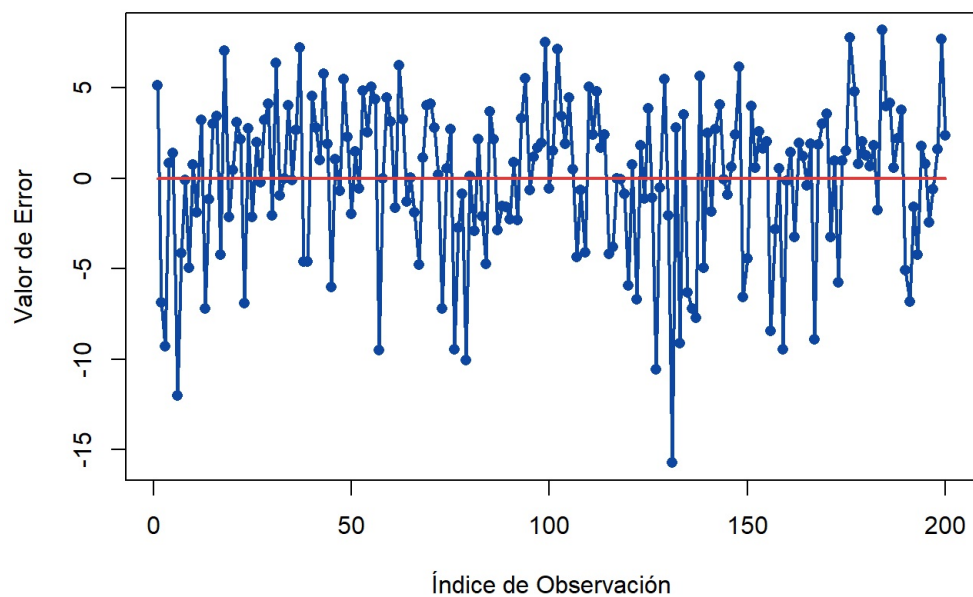
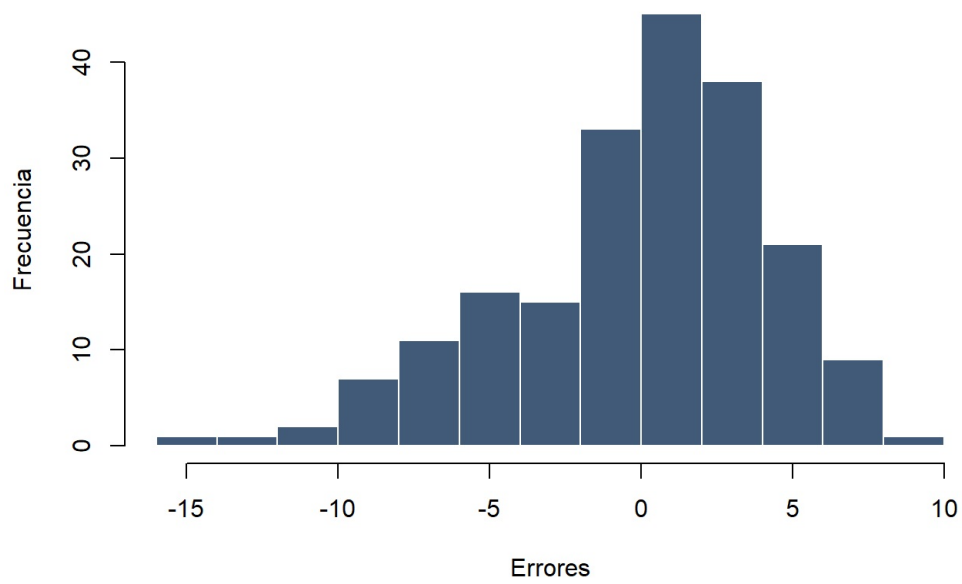


Gráfico de Errores



Histograma de Errores



Boxplot

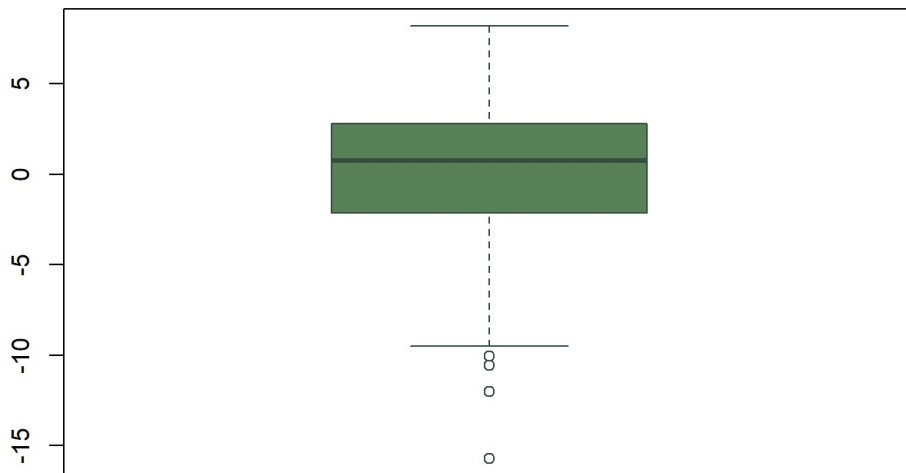
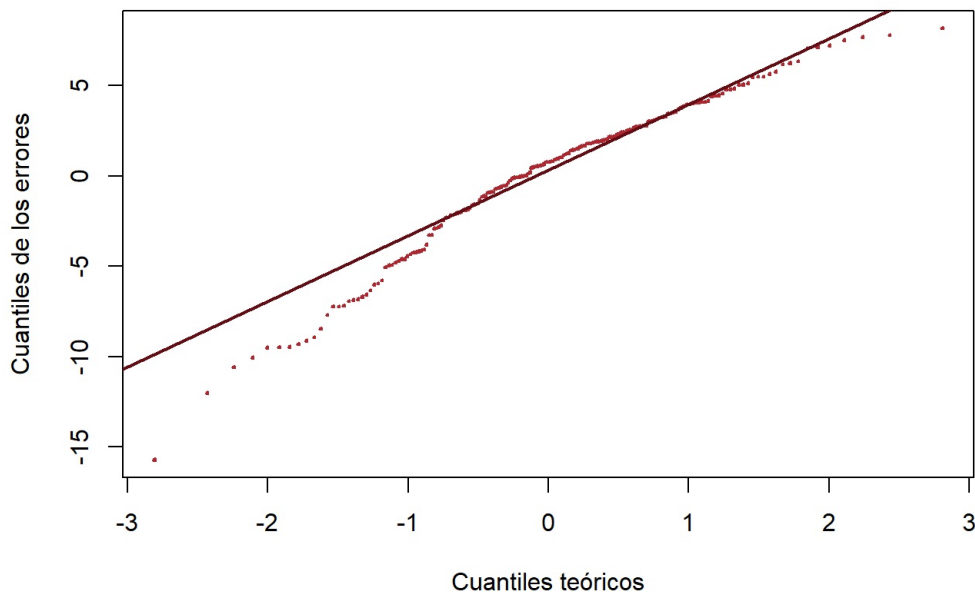


Gráfico QQ de Errores



Informacion del modelo

Datos	Valores
β_0	9.31163809515829
β_1	0.20249578339244
S	4.27494435490106
S cuad	18.2751492375005
cv	18.3757924471332
df	198
qti	-1.97201747783631
qtd	1.97201747783631
tc	9.92076547282495
p value	4.35496600176698e-19
Int. conf.	0.162244330504869
	0.24274723628001
Prom. err.	3.32021879764207
Var. err.	7.10406520694444
P. Hip. β_1 Est.	Rechazamos H0
P. Hip. β_1 Pvalue	Rechazamos H0

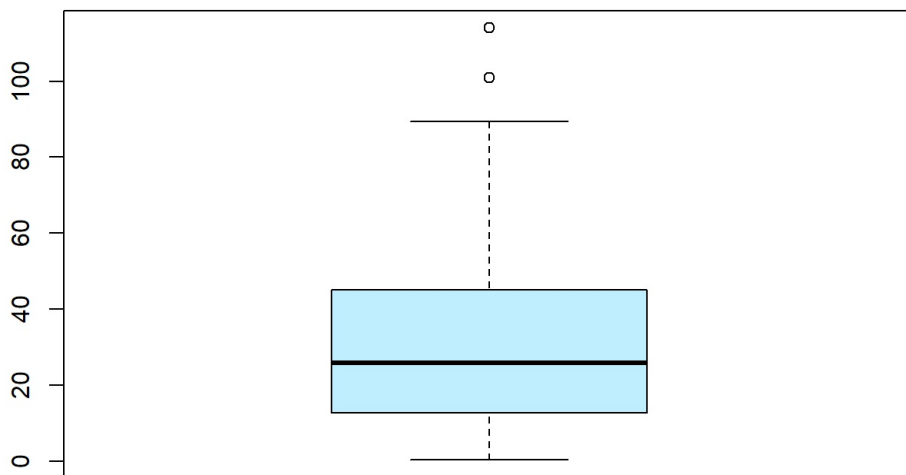
Con respecto a nuestros datos del nuevo modelo con "radio" como variable predictora, podemos observar en el boxplot que los datos están sesgados hacia la izquierda. El histograma también muestra una concentración mayor de datos en el lado izquierdo. Sin embargo, no se asemejan a una distribución conocida. En comparación con el modelo anterior, notamos que en este caso los datos están más dispersos, incluso con algunos valores atípicos en la relación. Al graficar la línea de regresión, podemos observar que en esta ocasión los puntos están más alejados de la línea, lo que sugiere un promedio de errores mayor que en el modelo anterior.

En cuanto a los intervalos de confianza, podemos observar que, en su mayoría, contienen los datos, excepto por algunos valores atípicos del modelo. Sin embargo, estos intervalos parecen ser más amplios que en el modelo anterior, ya que los datos se alejan más de la línea de regresión. En los gráficos de errores, notamos que los errores están más alejados en la parte inferior, mientras que en la parte superior de la línea, observamos un comportamiento de errores similar al del modelo anterior. En este caso, el histograma de errores también muestra un sesgo hacia la derecha, y en lo que respecta al ajuste a una distribución normal, los errores tienden a aproximarse más a una normal a medida que aumentan los presupuestos.

En este caso, nuestro valor de beta 1 es más alto, lo que indica que un aumento en la variable predictora tiene un impacto mucho mayor en los resultados de nuestra variable de respuesta. Sin embargo, también notamos una varianza mayor, ya que los datos están más dispersos entre sí, casi el doble que en el modelo anterior. Del mismo modo, el coeficiente de variación indica una alta variabilidad relativa en comparación con la media. El intervalo de confianza para beta 1 se sitúa entre 0.16 y 0.25 con un nivel de significancia alpha. Esto, junto con el valor p, nos lleva a rechazar la hipótesis nula que sugiere que beta 1 podría ser igual a cero. Como resultado, los errores son mayores, con un promedio de 3.32, aunque se esperaba que fueran mayores dada la forma de la gráfica y la aproximación de la línea.

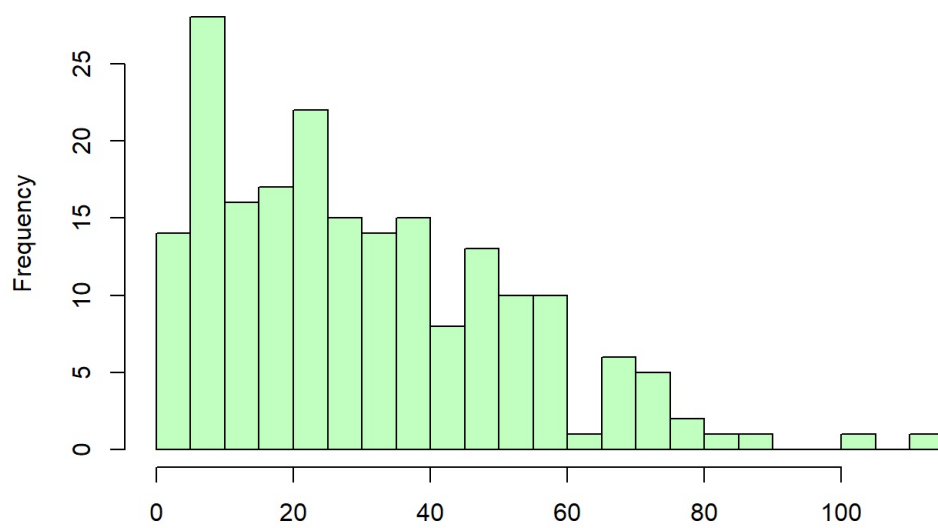
```
linealmodel(newsp, sales, 0.05)
```

Boxplot x



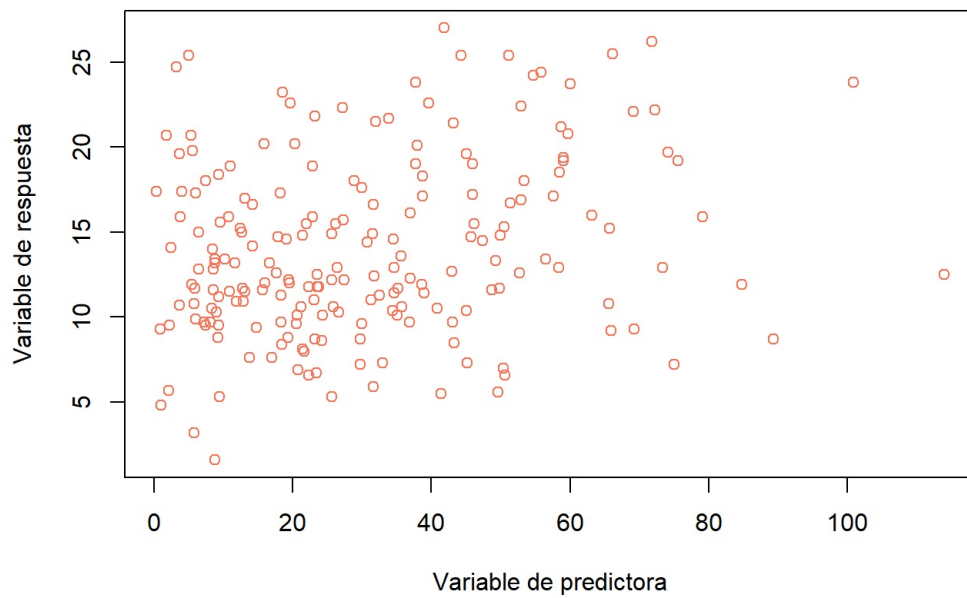
Variable de predictora

Histograma x

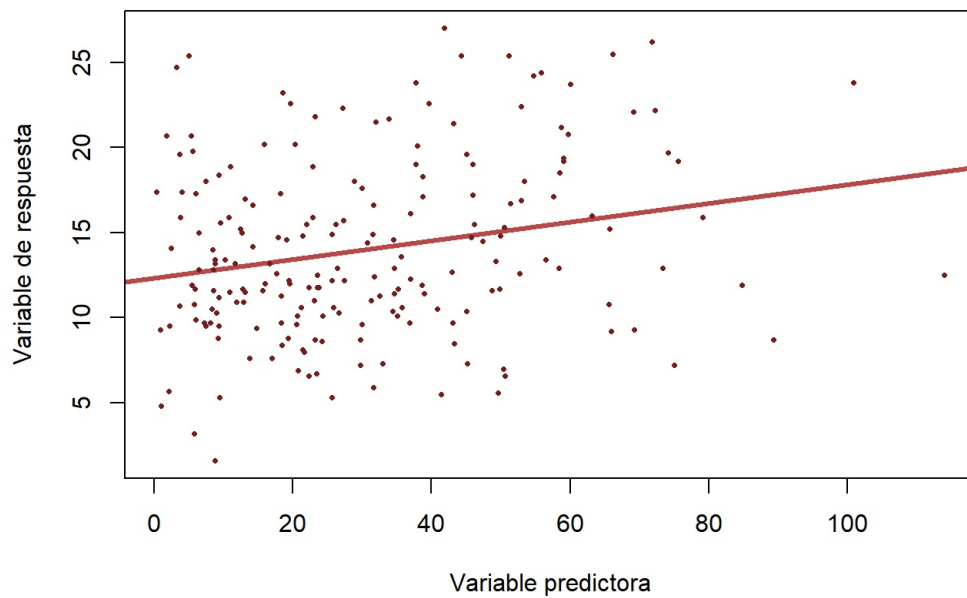


Variable de predictora

Gráfico dispersión



Recta de regresión



Intervalos de confianza

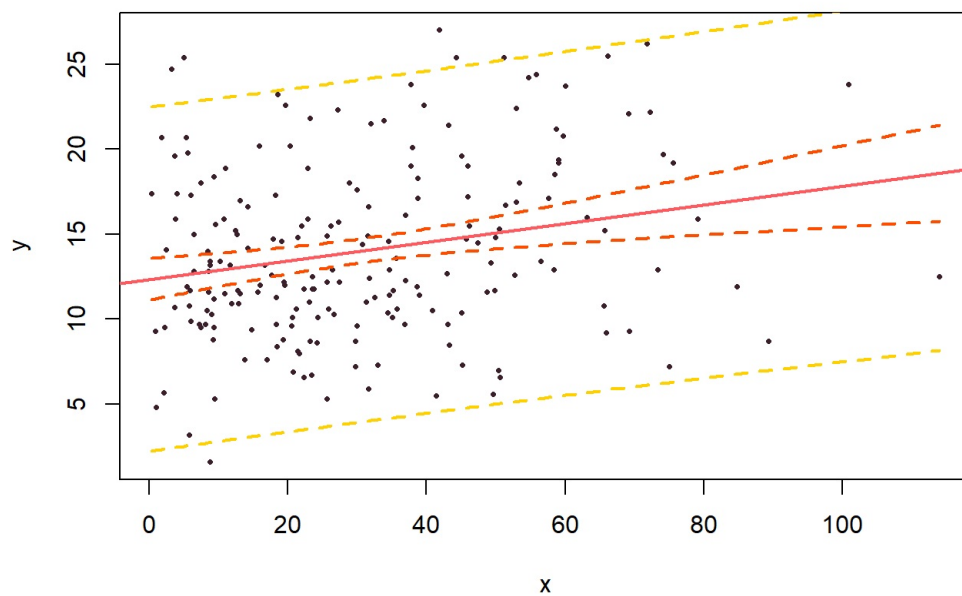
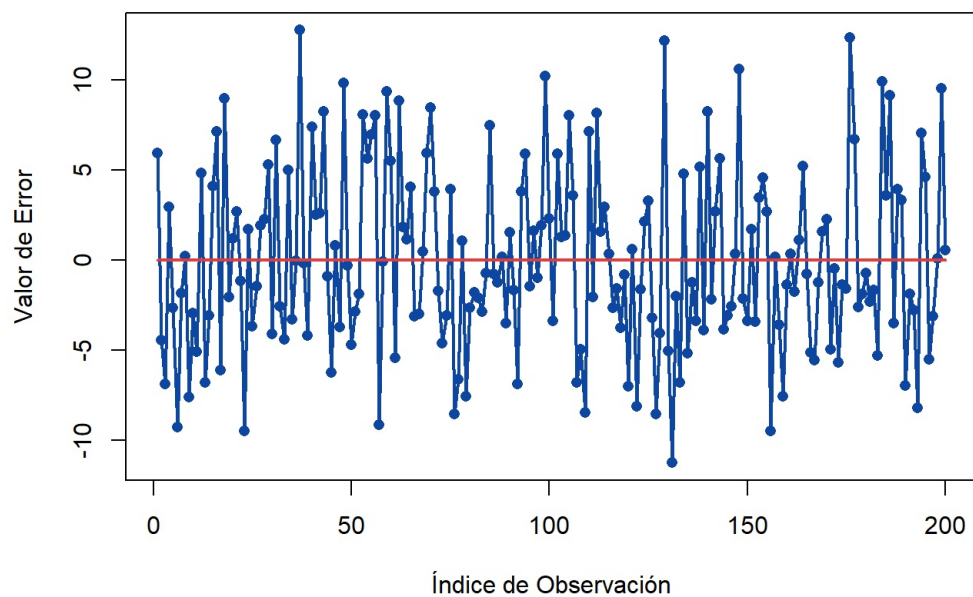
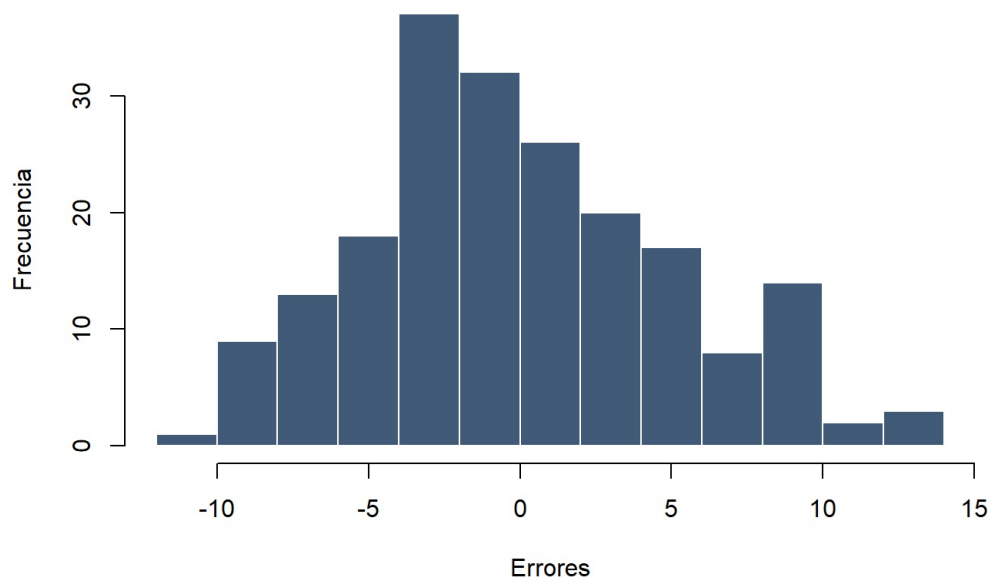


Gráfico de Errores



Histograma de Errores



Boxplot

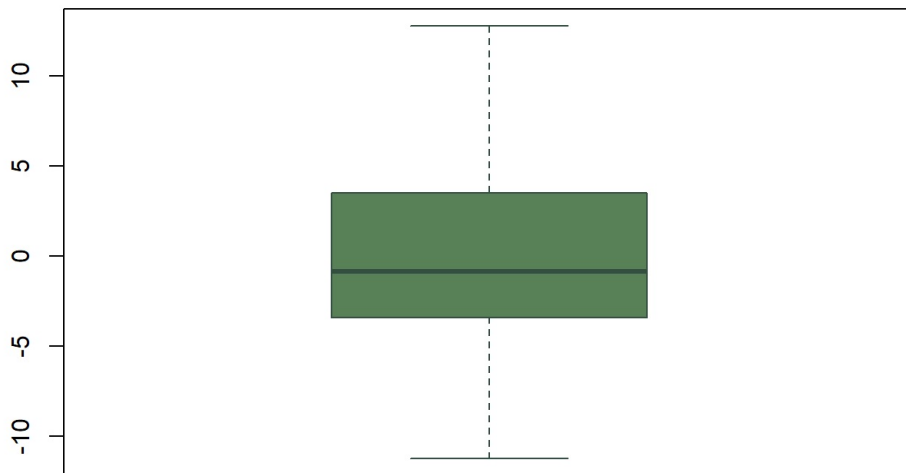
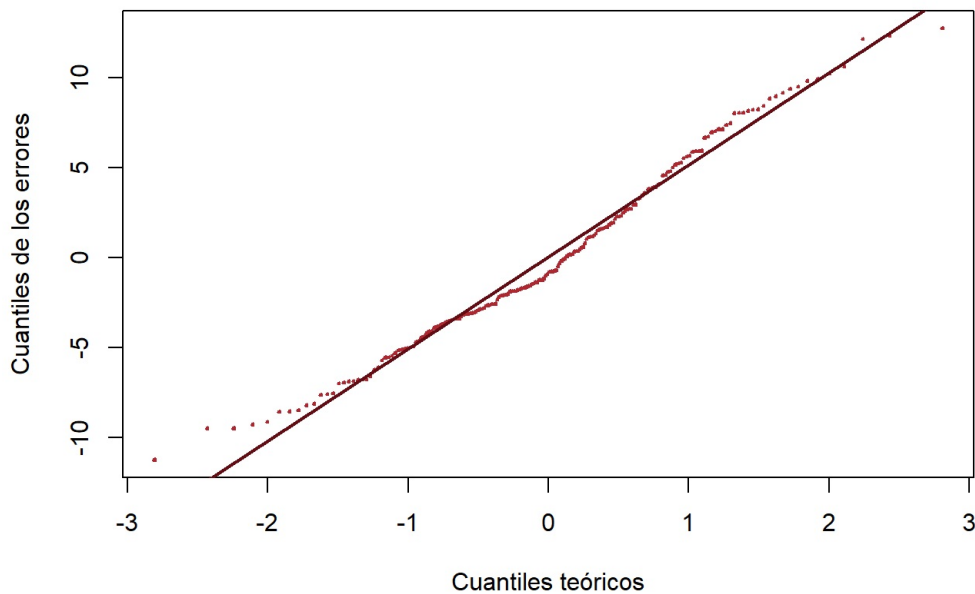


Gráfico QQ de Errores



Informacion del modelo

Datos	Valores
β_0	12.3514070692782
β_1	0.0546930984722733
S	5.09248036652019
S cuad	25.9333562833936
cv	16.667147890686
df	198
qti	-1.97201747783631
qtd	1.97201747783631
tc	3.29959074363342
p value	0.00114819586888821
Int. conf.	0.0220054852243413
	0.0873807117202053
Prom. err.	4.14655974383849
Var. err.	8.52267840335456
P. Hip. β_1 Est.	Rechazamos H0
P. Hip. β_1 Pvalue	Rechazamos H0

En relación a nuestro último modelo, donde la variable predictora corresponde a los gastos publicitarios en periódicos, podemos observar dos outliers notables en el box plot que están muy alejados de los demás datos. Además, notamos que los datos están sesgados hacia la izquierda, y esto se confirma en el histograma, donde la mayoría de los presupuestos publicitarios son bajos para este tipo de medio. En este caso, se evidencia una mayor variabilidad en los datos, ya que los puntos están más dispersos que en los modelos anteriores, incluso más dispersos que en los modelos previos. Esto sugiere que este modelo probablemente tendrá errores más altos debido a la distancia entre los puntos y la línea de regresión. Los outliers que observamos indican que algunos gastos publicitarios son significativamente mayores de lo que generalmente se observa.

En cuanto a los intervalos de confianza, notamos que estos intervalos aún se ajustan a los datos en su mayoría, pero dejan algunos valores superiores fuera de los intervalos, como habíamos previsto. Los errores son bastante altos, llegando a superar los 10, y el promedio de errores es de 4.14, que es el error promedio máximo entre los tres modelos. El histograma de estos errores muestra un sesgo hacia la izquierda, y al observar el gráfico qq, se confirma que este modelo es el que menos se ajusta a una distribución normal.

Como en los casos anteriores, la varianza de este modelo es la más alta, lo que sugiere que los datos no dependen mucho del modelo y se ajustan de manera deficiente a la línea de regresión. La variación promedio relativa con respecto a la media es también bastante alta, lo que indica una baja confiabilidad en el modelo. El valor de beta 1 es pequeño, lo que significa que la variable de respuesta no cambia significativamente a medida que aumenta la variable predictora. Además, el valor p nos lleva a rechazar la hipótesis nula, lo que indica que beta 1 es diferente de cero.