Ejercicio 6_Pinguinos Macaroni

Eduardo Niño Pedraza

Los pinguinos Macaroni ponen nidadas de dos huevos de tamaño diferente. El peso en gramos de los huevos de 11 nidadas se presenta en la tabla de abajo.

x	79	93	100	105	101	96	96	109	70	71	87
y	113	138	158	168	157	139	142	165	107	103	130

Con base en la información anteriormente proporcionada, los investigadores tienen algunas preguntas, por lo que procedemos de la siguiente manera:

Inciso 1

Ajuste la recta de regresión para estimar el peso promedio del huevo mayor (y) dado el peso del huevo menor (x). Comente sobre el ajuste del modelo, es decir, si parece correcto y si se cumplen los supuestos.

Solución: Comenzamos nuestro estudio ajustando un modelo de regresión lineal simple usando y y x, de donde observamos en primera instancia que le prueba F asociada a la tabla ANOVA indica que existe una relación de algún tipo entre los pesos de las nidadas, $p - value = 1.179e - 07 < 0.05$, posteriormente, observamos que al ser un modelo de regresión lineal simple la prueba que involucra si el coeficiente relacionado a la variable x es cero o no nos indica que es diferente de cero, *p-value = 1.18e-07<0.05*. Además podemos ver que la R cuadrada ajustada nos dan un buen indicio de la variabilidad explicada por el modelo en un 95.7% Así, nuestro modelo ajustado queda de la siguiente manera:

$$\hat{E}(y; x) = -17.616 + 1.702x$$

Del modelo ajustado podemos ver que conforme aumenta la masa en gramos de los huevos de la nidada x, los huevos de la nidada y aumentan en un factor de 1.702, lo que en efecto se observa en los datos de la base dada. Sin embargo, ahora debemos de revisar los supuestos del modelo, por lo que comenzaremos por homocedasticidad

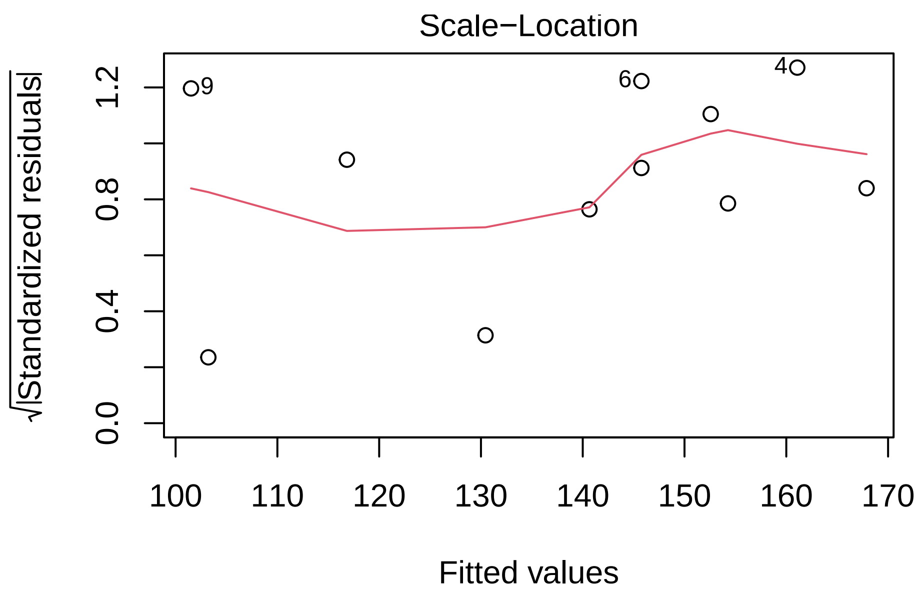

Figure 1: Residuos estandarizasos vs valores ajustados

Por parte de la grafica que verifica homocedasticidad no encontramos evidencia en contra, pero procedemos a complementar nuestro análisis visual con pruebas de hipótesis, *Breusch-Pagan test* y *Non-constant Variance Score Test* donde en ambas se contrasta varianza constante contra varianza no constante. De las pruebas de hipótesis reafirmamos las sospechas iniciales pues en ambas no se encuentra evidencia en contra de la homocedasticidad, en BP el $p-value = 0.41 > 0.05$ y en NCV el $p-value = 0.616 > 0.05$. Ahora, debemos de continuar con el supuesto de linealidad

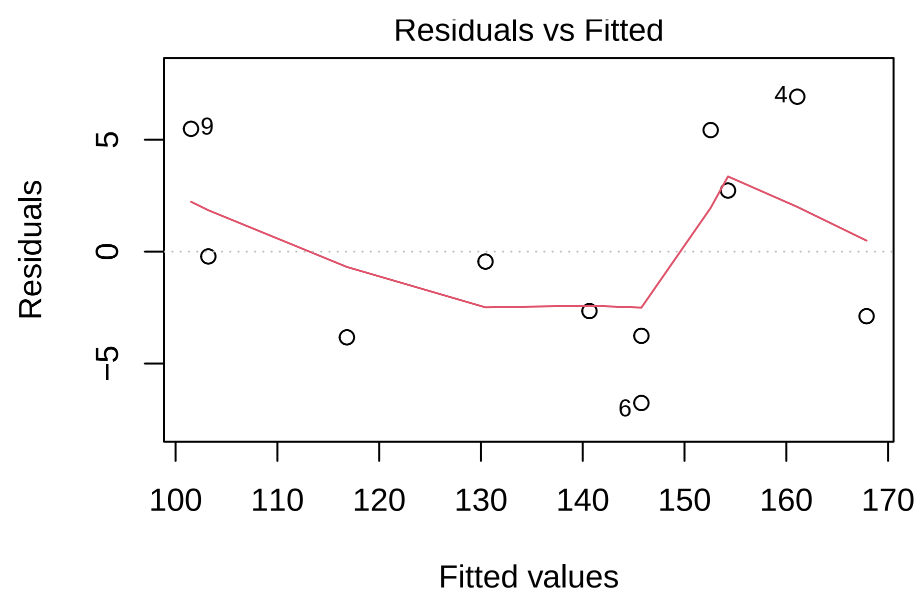

Figure 2: Residuos vs valores ajustados

De la gráfica no podríamos decir que se cumple la linealidad pues no se nota una simetría clara, dado que observamos más residuales negativos que positivos, sin embargo procedemos a hacer pruebas de hipótesis para argumentar la **no** necesidad de transformación de la variable x usando Box-TidwellDe la prueba obtenemos un $p-value = 0.2268 > 0.05$, por lo que no se encuentra evidencia en contra de no tener que usar alguna transformació para x, es decir, es plausible considerar el modelo con x directamente.

Como siguiente, analizamos la normalidad, aquí cabe hacer notar que dada que nuestra muestra es pequeña (11 observaciones), la QQ plot de la normal puede no ser del todo fiable, aunque no se observan problemas graves de normalidad, pero en dado caso hacemos uso de una QQ plot pero con una distribución $t_{11-3} = t_8$, donde observamos nuevamente un comportamiento aceptable para la normalidad

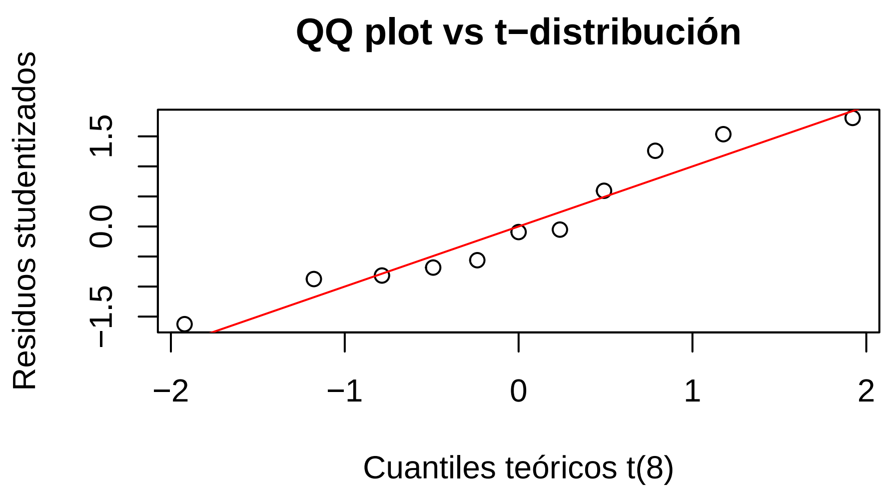

Figure 3: Q-Q plot distribucion t con DF = 8

Apoyandonos de una prueba de hipótesis Kolmogorov-Smirnov no obtenemos evidencia en contra de que nuestros residuos provienen de una distribución t, con $n-3$ grados de libertad, teniendo un $p-value = 0.9012 > 0.05$

Finalmente, procedemos a revisar el ultimo supuesto respecto a la aleatoriedad de los datos, donde no conocemos nada más que los datos, por lo que corremos las pruebas de hipótesis disponibles. De las pruebas observamos que no se encuentra evidencia en contra de la aleatoriedad, la cual se complementa con el autocorrelograma y las prueba de Durbin-Watson, $p-value = 0.2327$.

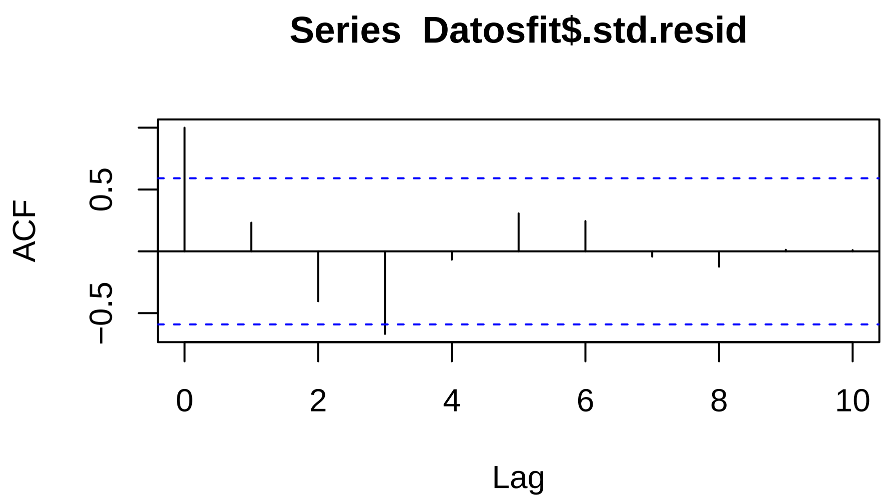

Figure 4: Autocorrelograma

Con base en el análisis anterior nuestro modelo cumple los supuestos y además nos indica una relación creciente entre los valores de x y y.

Inciso 2

Los investigadores tienen la sospecha de que en promedio se puede decir que la diferencia entre el peso mayor y el peso menor es constante (es decir, no depende del peso del huevo menor observado). Usando el modelo obtenido antes realice una prueba de hipótesis para responder la pregunta de los investigadores, describiendo con detalle las hipótesis que se contrastan.

Solución: Notemos que los investigadores se preguntan si en promedio la diferencia entre el peso del huevo mayor y el menor es constante, es decir que:

$$k = E(y; x) - x = \beta_0 + \beta_1 x - x = \beta_0 - (\beta_1 - 1)x$$

donde para que el lado derecho sea constante, $k = \beta_0$ se debe de cumplir que $\beta_1 - 1 = 0$ por lo que nuestra prueba de hipótesis queda establecida de la siguiente manera:

$$H_0 : \beta_1 = 1 \text{ vs } H_a : \beta_1 \neq 1$$

Usando el paquete de Multcomp, y una prueba glht obtenemos un $p-value = 0.0001677056 < 0.05$ como evidencia para concluir que β_1 es diferente de 1, por lo que la sospecha de los investigadores es incorrecta y la diferencia entre el peso mayor y el menor no es constante.

Inciso 3

Posteriormente se observa el peso de los huevos de una nueva nidada, observándose un peso de 80 y 125 gramos. Usando un intervalo adecuado, comente sobre la sospecha de que la nidada de huevos sí proviene de pinguinos Macaroni.

Solución: Como en este caso tenemos un peso menor de 80 y uno mayor de 125, lo que los investigadores quieren es saber si con una confianza adecuada podemos decirles si pertenece a las nidadas que fueron observadas, así usando un intervalo de confianza del 95%, buscamos el intervalo de confianza con una valor observado de $x = 80$, el cual varia de $(114.1202, 122.9458)$ con un valor puntual de 118.533. Por otro lado, si cambiamos la confianza al 90% tenemos un intervalo de $(114.9571, 122.1088)$, donde en ambos casos podemos concluir que los huevos de la nueva nidada no son de los pinguinos Macaroni.

Conclusiones.

Teniendo la información de entre dos nidadas de huevos de pinguinos Macaroni obtenemos un modelo que cumple con los supestos de linealidad, homocedasticidad, normalidad y aleatoriedad. Del modelo concluimos que existe una relación positiva entre el peso del huevo menor y el mayor, descrito por el modelo ajustado:

$$\hat{E}(y; x) = -17.616 + 1.702x$$

Además se observa que la diferencia entre el peso mayor y el menor no es constante. Y finalmente si se obtuviera un huevo de peso menor de 80 gramos y uno de peso mayor de 125 podemos concluir que no provienen de una nidada de pinguinos Macaroni.