



Influencia de la dieta en el peso de los pollos: Una aproximación práctica.

PROBABILIDAD, ESTADÍSTICA Y ELEMENTOS DE BIOMATEMÁTICA

Pedro Sánchez García

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA PARA CIENCIAS DE LA SALUD.
PROFESOR: DR. RICARDO CAO ABAD.

ÍNDICE DE CONTENIDO.

INTRODUCCIÓN.....	2
ENUNCIADO DEL PROBLEMA.	2
OBJETIVOS.....	2
METODOLOGÍA.....	2
ANÁLISIS EXPLORATORIO GENERAL.....	3
ANÁLISIS EXPLORATORIO POR DIETAS.....	6
MODELO DEL ANOVA I.....	8
COMPARACIONES MÚLTIPLES.....	10
DIAGNOSIS DEL MODELO.....	11
CONTRASTE NO PARAMÉTRICO: KRUSKAL-WALLIS.....	13
CONCLUSIONES.....	14
BIBLIOGRAFÍA.....	14

INTRODUCCIÓN.

Uno de los principales aspectos a los que se enfrentan los profesionales en la avicultura es la determinación de la dieta que se ajusta de forma idónea al peso deseado en función de la raza seleccionada. Además, el papel de la dieta también resulta crucial para lograr un desarrollo óseo adecuado a lo largo del primer mes de crecimiento. No obstante, los resultados varían notablemente según la proporción de proteínas en la dieta, seguido del contenido en calcio y vitaminas. Por ejemplo, en la raza Leghorn originaria de la Toscana y seleccionada para lograr una buena producción de huevos, al suministrar una dieta con un porcentaje de proteínas comprendido entre 15 y 21%, se logra un marcado incremento de peso y de la longitud de la tibia en comparación con una dieta que presente un contenido superior al 50% en calcio y fósforo. Este proceso viene acompañado por una mayor síntesis de colágeno en la matriz ósea, lo que indica claramente la correlación entre una dieta rica en proteínas y el desarrollo óseo durante el primer mes (**Summers et al., 1990**). De esta forma, podemos apreciar la importancia de esta cuestión, así como la necesidad de emplear los análisis estadísticos oportunos, para garantizar el peso idóneo de los pollos y la posterior producción.

ENUNCIADO DEL PROBLEMA.

StatFarm es una iniciativa empresarial de bioestadística creada por un antiguo alumno del MUBICS. Su actividad se centra en la aplicación de diferentes técnicas y análisis estadísticos para la elaboración de informes en el sector agroalimentario. Recientemente, el equipo de StatFarm ha recibido la solicitud de un informe por parte de un cliente que pertenece a una cooperativa avícola. En este caso, el cliente desea conocer la posible influencia de cuatro dietas que administra a los pollos con el fin de determinar cuál es la mejor casa comercial ofrecida como servicio para los próximos años. En consecuencia, el equipo ha de aplicar las herramientas y metodologías estadísticas oportunas para proporcionar un informe de calidad al cliente que le resulte de gran ayuda para tomar la decisión.

OBJETIVOS.

El presente trabajo tiene como objetivos principales:

1. Evaluar la posible influencia de la dieta en el peso de los pollos a lo largo del tiempo.
2. Mostrar los diferentes análisis estadísticos expuestos en la materia, que se llevarían a cabo en un caso práctico enfocado al sector agroalimentario.

METODOLOGÍA.

Se parte de un conjunto de datos que presenta 578 observaciones y 4 columnas correspondientes al identificador único para cada pollo, la dieta rica en proteínas, el tiempo como número de días desde el nacimiento y el peso corporal en gramos para cada pollo.

Las medidas del peso corporal en cada pollo se llevaron a cabo desde el nacimiento cada 2 días hasta que se alcanzaron los 20 días, realizando una última medida a los 21 días.

RESUMEN NUMÉRICO.

```

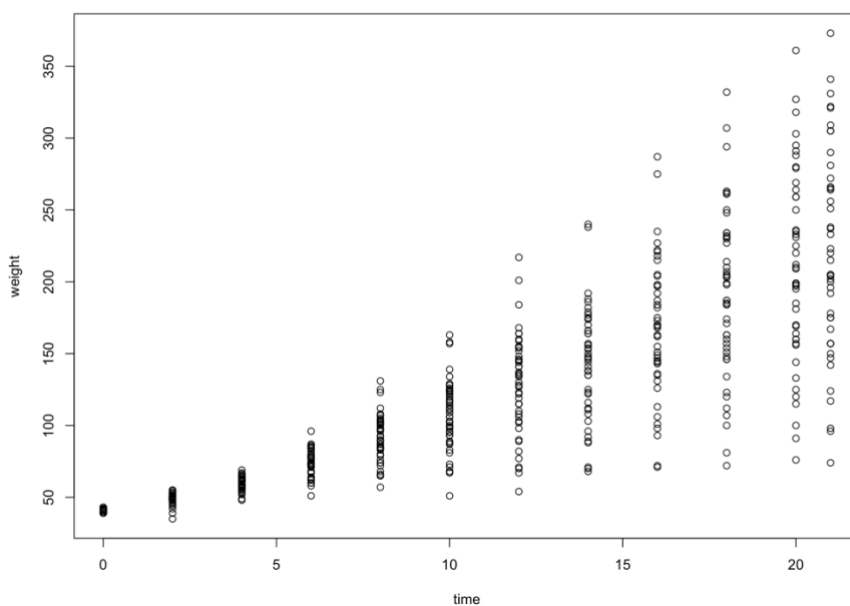
> summary(ChickWeight)
  weight      Time      Chick      Diet
Min.   : 35.0   Min.   : 0.00   13      : 12   1:220
1st Qu.: 63.0   1st Qu.: 4.00    9      : 12   2:120
Median :103.0   Median :10.00   20      : 12   3:120
Mean   :121.8   Mean   :10.72   10      : 12   4:118
3rd Qu.:163.8   3rd Qu.:16.00   17      : 12
Max.   :373.0   Max.   :21.00   19      : 12
              (Other):506
  > sd(weight)
[1] 71.07196
  > skewness(weight)
[1] 0.9587845
  > kurtosis(weight)
[1] 0.3372147

```

El resumen estadístico del conjunto de datos resulta de interés para la variable peso, donde apreciamos el valor mínimo de 35 y máximo de 373 gramos, así como un peso medio de 121.8 gramos. La asimetría presenta un valor próximo a 1, de tal forma que en la variable peso existe una marcada asimetría hacia la derecha. Por su parte, la curtosis con un valor superior a 0 nos muestra que la distribución de la variable peso es más elevada en comparación con una normal.

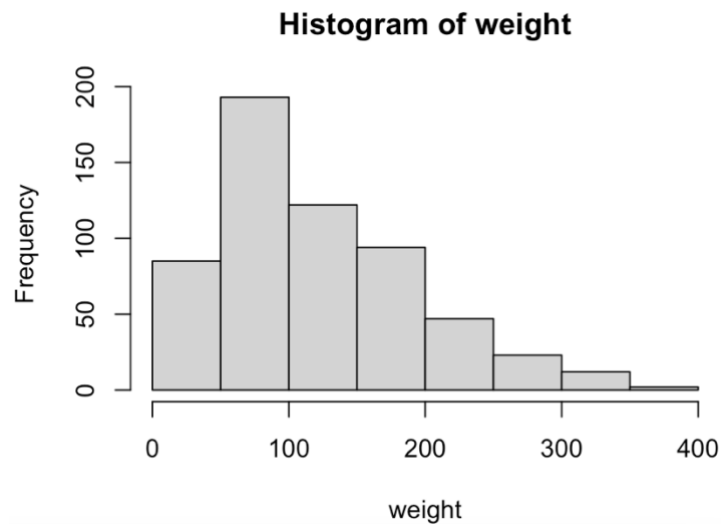
ANÁLISIS GRÁFICO.

DIAGRAMA DE DISPERSIÓN.



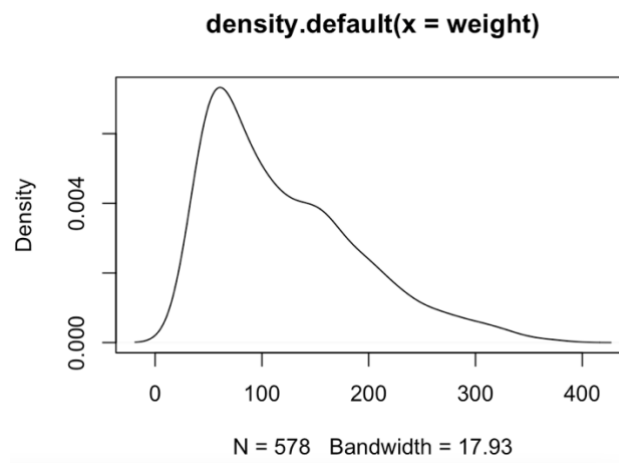
El diagrama de dispersión es una construcción gráfica bidimensional que nos indica la tendencia de los pares de observaciones que toman las variables, en este caso el tiempo y el peso corporal de los pollos. A pesar de la marcada concentración de los puntos debido a las diferentes dietas, podemos observar que el peso corporal incrementa conforme lo hace el tiempo transcurrido de los 21 días analizados.

HISTOGRAMA.



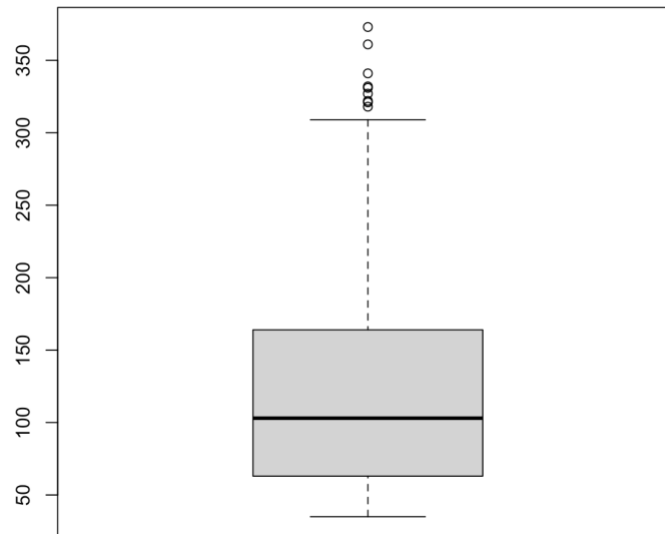
El histograma, en el que se representa la densidad de frecuencia en la variable estudiada, pone de manifiesto que gran parte de las observaciones para los pollos devuelven un peso corporal comprendido entre 50 y 100 gramos. De esta forma, para la población existe una elevada probabilidad de que el peso corporal de un determinado pollo tomado al azar se sitúe en ese rango.

FUNCIÓN DE DENSIDAD.



La estimación del conjunto de datos mediante el estimador tipo núcleo (kernel) de la densidad resulta mejor que el histograma, apreciando un valor máximo próximo a 100 gramos. Gracias a esta aproximación, se corrige la variación en el caso de intervalos de peso con misma longitud, es decir, se abre un intervalo en torno a un punto y se lleva a cabo un recuento de aquellos datos de la variable ubicados en su interior, procediendo posteriormente al cálculo de la frecuencia mostrada para las clases en la gráfica. En consecuencia, no se producen unos saltos como en el caso del histograma dado que la función de ponderación es de tipo continua con un máximo en cada recuento del intervalo real.

DIAGRAMA DE CAJAS | BOX-PLOT.



El diagrama de cajas permite visualizar la asimetría y la existencia de valores atípicos existentes en el conjunto de datos, tal y como se observa en este caso, con numerosos valores de peso concentrados. Además, en lo que respecta a la simetría, cuanto más centrada se ubique la mediana en el interior de la caja, mayor será la simetría. En este caso, existe un ligero desplazamiento de la mediana, aunque no resulta notable, hacia valores bajos en la escala, posiblemente debido a los datos atípicos.

ANÁLISIS DE NORMALIDAD.

CONTRASTE DE NORMALIDAD.

Mediante el test de normalidad de Shapiro-Wilk, se plantea el siguiente contraste de hipótesis:

$H_0 \equiv$ la variable sigue una distribución normal

$H_1 \equiv$ la variable no sigue una distribución normal

```
> shapiro.test(weight)

      Shapiro-Wilk normality test

data:  weight
W = 0.90866, p-value < 2.2e-16
```

Rechazamos la hipótesis nula de normalidad dado que el p-valor se encuentra por debajo del nivel de significación estadística $\alpha=0.05$. Por tanto, la variable peso no sigue una distribución normal.

RESUMEN NUMÉRICO POR DIETAS.

```

> tapply(weight,diet,summary)           # cuantiles por dietas
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 35.00  57.75   88.00  102.65  136.50  305.00

$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 39.0   65.5   104.5   122.6   163.0   331.0

$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 39.0   67.5   125.5   142.9   198.8   373.0

$`4`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 39.00  71.25  129.50  135.26  184.75  322.00

> tapply(weight,diet,mean)               # medias por dietas
      1      2      3      4
102.6455 122.6167 142.9500 135.2627

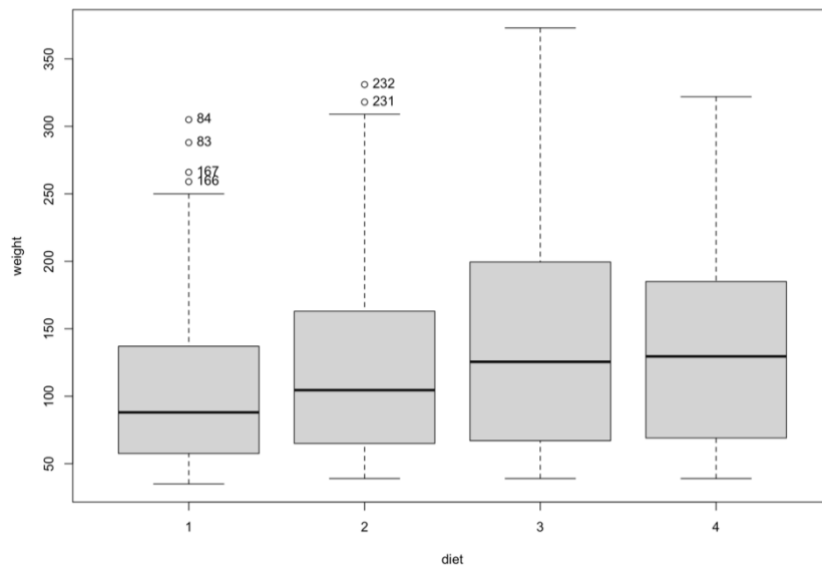
> tapply(weight,diet,sd)                 # desviaciones estándar por dietas
      1      2      3      4
56.65655 71.60749 86.54176 68.82871

```

El resumen numérico con los principales descriptivos para el peso según las dietas indica para el caso de los cuantiles según la dieta, unos valores correspondientes a aquellos que dejan np observaciones a su izquierda y $n(1-p)$ observaciones a la derecha. En este caso, apreciamos que, para las diferentes dietas, los valores del primer y tercer cuantil se ubican en rangos similares. Por otro lado, cabe destacar que no se producen diferencias muy importantes en la escala de los datos cuando observamos las medias. Esto es relevante, ya que, si se da la situación de una diferencia notable en la escala, llevaría a que la varianza de los datos grandes tuviese un peso desproporcionado en los análisis efectuados.

En lo que respecta a la desviación estándar, viendo los resultados obtenidos, podemos afirmar que se produce una marcada variación con respecto a la media aritmética y, en consecuencia, un error proporcional a esta. Este sesgo, posiblemente ocasionado por la existencia de datos atípicos, podría reducirse mediante un incremento del tamaño muestral.

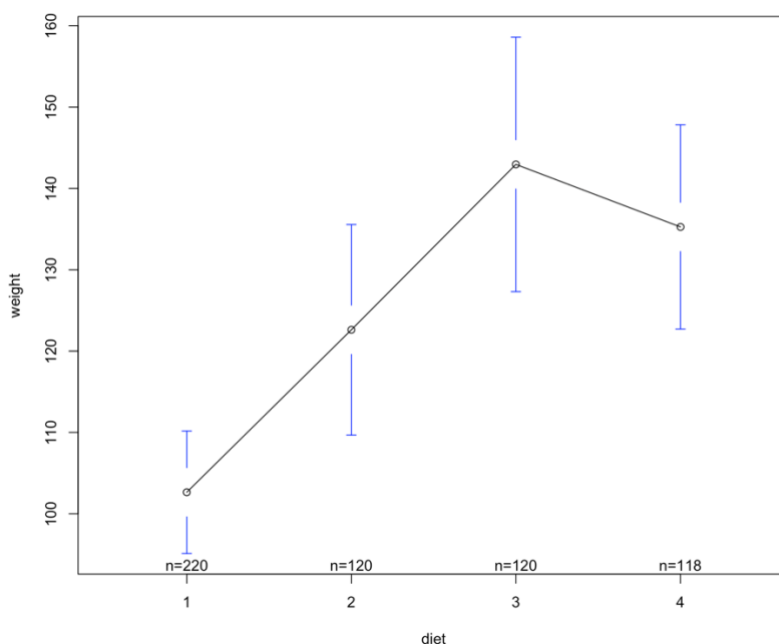
DIAGRAMA DE CAJAS | BOX-PLOT SEGÚN LAS DIETAS.



El diagrama de cajas realizado anteriormente para la variable peso en el conjunto de datos nos mostraba la asimetría y la existencia de valores atípicos. Al llevar a cabo el diagrama por las dietas, podemos observar a grandes rasgos, que el peso medio alcanzado con la dieta 1 es el más bajo, incrementando progresivamente en el caso de las dietas 2, 3 y 4. La tendencia del incremento es superior entre las dietas 1, 2 y 3 en comparación con la que se produce en las dietas 3 y 4, donde vemos unos valores de peso medio muy similares.

La existencia de los datos atípicos que se muestran en el gráfico para la dieta 1 (83, 84, 166 y 167) y dieta 2 (231 y 232) concuerda con los resultados obtenidos para la desviación estándar del peso, de tal forma que en el caso de la dieta 3, la desviación estándar obtenida se podría asociar al tamaño muestral empleado.

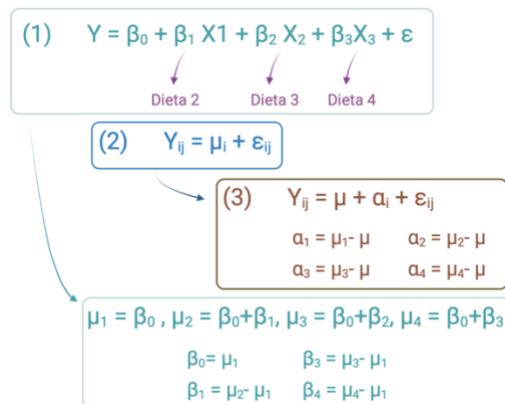
DIAGRAMA DE MEDIAS E INTERVALOS DE CONFIANZA POR DIETAS.



Este gráfico es una forma útil de ver las medias y los correspondientes intervalos de confianza para los cuatro niveles de la dieta.

La longitud de los intervalos de confianza es elevada en todos los casos, de forma que existe una elevada probabilidad con respecto a que las medias obtenidas se ubiquen en ese rango de valores.

CARACTERÍSTICAS DEL MODELO.



```
> LM.1 <- lm(weight~diet)
# indicadores
> summary(LM.1)
```

Call:
lm(formula = weight ~ diet)

Residuals:

	Min	1Q	Median	3Q	Max
	-103.95	-53.65	-13.64	40.38	230.05

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	102.645	4.674	21.961	< 2e-16 ***
diet2	19.971	7.867	2.538	0.0114 *
diet3	40.305	7.867	5.123	4.11e-07 ***
diet4	32.617	7.910	4.123	4.29e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 69.33 on 574 degrees of freedom
Multiple R-squared: 0.05348, Adjusted R-squared: 0.04853
F-statistic: 10.81 on 3 and 574 DF, p-value: 6.433e-07

Se lleva a cabo la formulación del modelo ANOVA I en un formato de regresión para la posterior realización del análisis de la varianza. Mediante la función LM se introducen unas variables artificiales (“dummys”) de tipo binario con las dietas. De esta forma, la variable respuesta Y correspondiente al peso de los pollos conformará un coeficiente independiente, mientras que para la variable dieta, la 1 es la referencia en el modelo, representándose las otras variables con las indicadoras X₁, X₂ y X₃ en referencia a las dietas 2, 3 y 4 respectivamente (**ecuación 1**). No obstante, sabemos que el valor de Y para un determinado nivel del factor i y réplica j (Y_{ij}), será la media en el nivel (μ_i) junto con un error aleatorio asociado (**ecuación 2**). Esta última ecuación se puede modificar ligeramente para generar la **ecuación 3**, donde se muestra que la media de cada nivel (μ_i) se puede descomponer en la suma de la media general y los efectos para cada nivel (α_i). Estos efectos consisten en la diferencia de media entre un nivel de factor y la media general.

Con respecto a los cambios entre las distintas ecuaciones, debemos centrarnos en que μ₁ corresponde al nivel medio de la variable peso con la dieta 1, mientras que μ₂, μ₃ y μ₄ son el sumatorio del nivel medio del peso con la dieta 1 y esas dietas respectivamente.

En el resultado del resumen para el modelo, se reflejan los coeficientes β₀, β₁ y β₂ y β₃ estimados para el modelo con esta nomenclatura inherente al modelo de regresión lineal, así como los grados de libertad (578-4 parámetros a estimar). Podemos intuir que las mayores diferencias en términos de las medias corresponden entre las dietas 3 y 4 con la dieta 1, siendo la diferencia inferior si se comparan las dietas 2 y 1.

```
> confint(LM.1,level=0.95)                                     # Obtención de intervalos de confianza al nivel de confianza del 95%
```

	2.5 %	97.5 %
(Intercept)	93.465316	111.82559
diet2	4.518737	35.42369
diet3	24.852070	55.75702
diet4	17.080279	48.15424

Los **intervalos de confianza** a un nivel del 95% nos indica que no contienen a 0 como valor, es decir, los parámetros β₀, β₁ y β₂ y β₃ del modelo anterior son distintos de 0.

Esta opción nos permite determinar cuando las variables es cuestión son prescindibles debido a que no se produce influencia, así como diferencias en los niveles del factor.

```
> # Obtención de las medias estimadas y los efectos con respecto a la media general
> model.tables(objeto, type = "means", se = TRUE) # Medias estimadas
Design is unbalanced - use se.contrast() for se's
Tables of means
Grand mean
121.8183

diet
  1      2      3      4
102.6 122.6 142.9 135.3
rep 220.0 120.0 120.0 118.0
> model.tables(objeto, type = "effects", se = TRUE) # Efectos
Design is unbalanced - use se.contrast() for se's
Tables of effects

diet
  1      2      3      4
-19.17  0.7983 21.13 13.44
rep 220.00 120.0000 120.00 118.00
```

Las **medias estimadas** para el modelo consisten en las mostradas anteriormente con la **ecuación 2**. El valor más elevado se logra con las dietas 3 y 4, con valores de 142,9 y 135,3 gramos respectivamente, mientras que la media genérica toma un valor de 121,8183. Por su parte, los **efectos** nos muestran un valor muy reducido para el nivel de la dieta 2 en comparación con el resto, donde los valores son más elevados.

TABLA DE ANÁLISIS DE LA VARIANZA DEL MODELO (ANOVA I).

```
> Anova(LM.1)
Anova Table (Type II tests)

Response: weight
      Sum Sq Df F value    Pr(>F)
diet    155863   3   10.81 6.433e-07 ***
Residuals 2758693 574
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Contraste de hipótesis:

$$H_0 \equiv \alpha_1 = \dots = \alpha_I = 0$$

$$H_1 \equiv \alpha_I \neq 0 \text{ para algún } i$$

Aludiendo a los datos obtenidos en la tabla, la suma de cuadrados residual resulta muy elevada, lo que se traduce en que hay un porcentaje considerable de variabilidad no explicada por el modelo. En este caso, si determinamos el R^2 , alcanzamos que el modelo explica en torno al 5,35% de la variabilidad de la respuesta:

$$R^2 = \left(\frac{155863}{155863 + 2758693} \right) = 0,0535$$

Dado que el p-valor es inferior al nivel de significación estadística $\alpha = 0,05$, se rechaza la hipótesis nula, por lo que $H1: \mu_i \neq \mu_j$ para algún $i \neq j$. Resulta asumible, pues, que hay diferencias para las

medias de los pesos en los distintos niveles de las dietas. Ahora la cuestión radica en estudiar entre que niveles se producen tales diferencias significativas.

COMPARACIONES MÚLTIPLES.

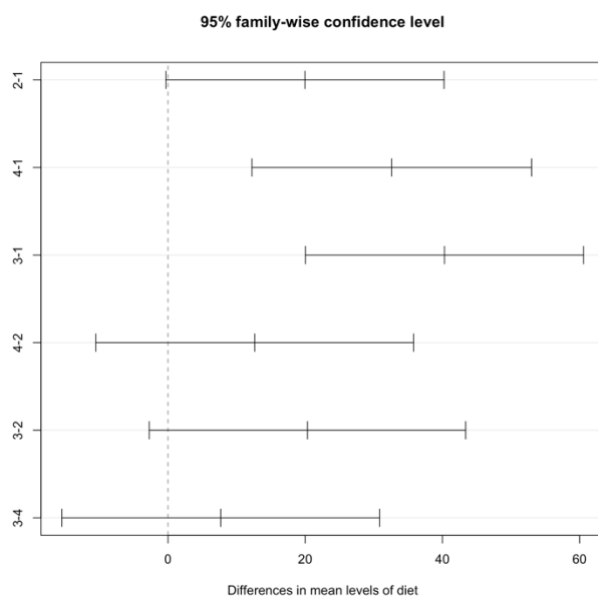
Para el estudio de los niveles entre los que se producen diferencias significativas, se procede al empleo de diferentes métodos, donde se realizará el siguiente contraste de hipótesis:

$$H_0^{ij}: \mu_i = \mu_j \equiv \alpha_i = \alpha_j$$

$$H_1^{ij}: \mu_i \neq \mu_j \equiv \alpha_i \neq \alpha_j$$

De los métodos existentes (Tukey, LSD, Bonferroni, Scheffé y Duncan), en el presente trabajo se opta por el método de Tukey dado el inconveniente que presenta el método LSD de llevar a cabo rechazos de hipótesis inadecuados y, por otra parte, la restricción en el método de Bonferroni, que lo hace demasiado conservador.

TEST DE TUKEY.



```
> out1.tuk
Tukey multiple comparisons of means
95% family-wise confidence level
factor levels have been ordered

Fit: aov(formula = weight ~ diet)

$diet
      diff      lwr      upr    p adj
2-1 19.971212 -0.2998092 40.24223 0.0552271
4-1 32.617257 12.2353820 52.99913 0.0002501
3-1 40.304545 20.0335241 60.57557 0.0000025
4-2 12.646045 -10.5116315 35.80372 0.4954239
3-2 20.333333 -2.7268370 43.39350 0.1058474
3-4  7.687288 -15.4703886 30.84496 0.8277810
```

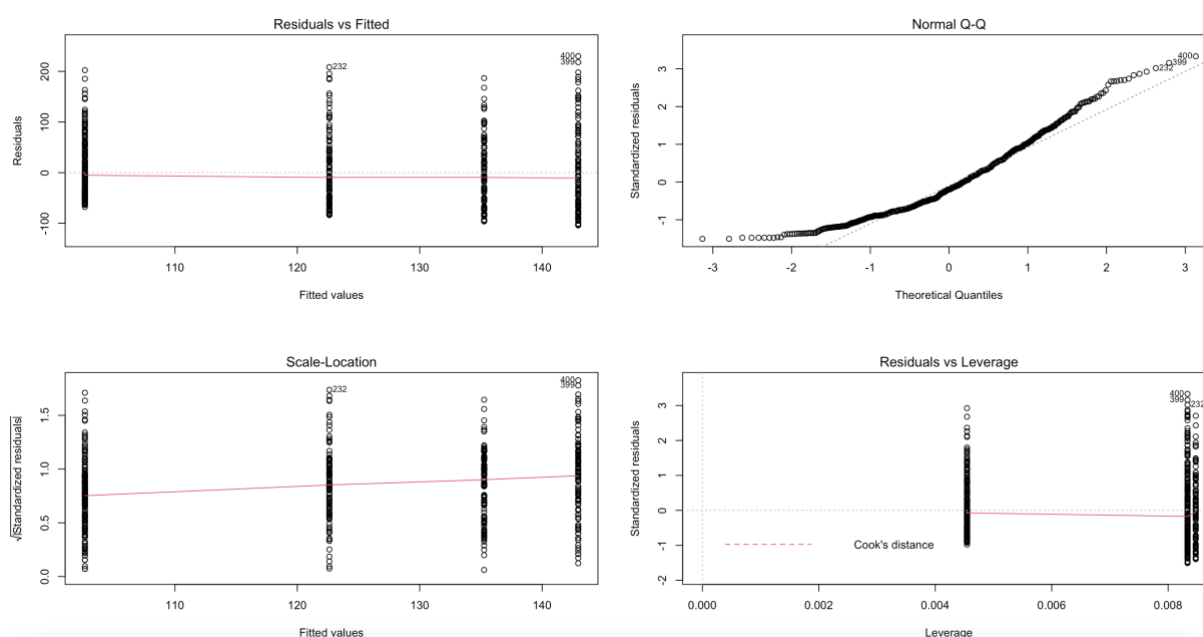
Se obtienen diferencias estadísticamente significativas entre las dietas 1-4 y 1-3, como se esperaba teniendo en cuenta los valores elevados correspondientes a las diferencias de las estimaciones obtenidas.

Gráficamente, podemos visualizar los intervalos de confianza para cada comparación, así como si estos incluyen el 0, mostrado con la línea punteada, lo que significa la ausencia de diferencias.

DIAGNOSIS DEL MODELO.

Una vez evaluadas la tabla ANOVA y las comparaciones, se realizan unas hipótesis básicas del modelo, donde si alguna no es cierta, estaríamos ante un caso en el que las conclusiones comentadas anteriormente no son fiables e incluso erróneas. En consecuencia, se procede a la realización de métodos gráficos y descriptivos aplicados sobre los residuos.

ANÁLISIS GRÁFICO DE LOS RESIDUOS.

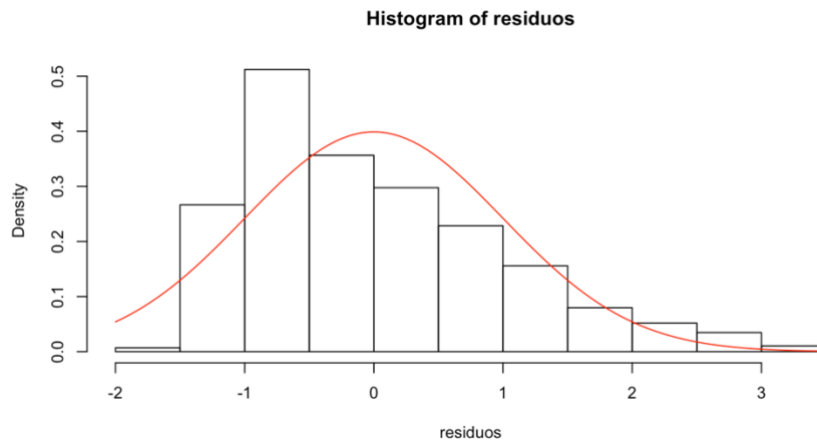


En primer lugar, se obtienen los residuos estandarizados para proceder a la muestra de gráficos. El gráfico superior izquierdo correspondiente a los residuos frente a valores ajustados nos conduce a sospechas de heterocedasticidad dada la dispersión de los valores obtenidos con respecto a 0, que consiste en la media para los residuos por construcción. Teniendo como referencia la línea roja de la estimación de la tendencia, parece que se produce una acusada variabilidad.

En el siguiente gráfico del QQ-plot, se enfrentan los valores de los residuos estandarizados frente al valor teórico de la normal, es decir, se trata de la recta en la que los residuos toman un valor correspondiente del cuantil en la normal $N(0,1)$. El ajuste de la recta indica que en los extremos se alcanza una menor proximidad y se recogen datos atípicos que alteran e influyen en el modelo, por lo que existen sospechas de falta de normalidad.

La representación de la raíz cuadrada de los residuos estandarizados frente a los valores ajustados nos permite detectar aquellos residuos más notables en las diferentes regiones, que en este caso nos indica, de nuevo, unas variaciones considerables y sospechas de heterocedasticidad.

Con el último gráfico podemos tener una idea de la distancia existente entre la nube de datos para cada uno de los niveles, es decir, las dietas. Así, apreciamos las diferencias notables en las estimaciones. Viendo el resultado y la referencia de la distancia de Cook, en general, podemos destacar que no existen puntos influyentes.



Si tenemos en cuenta el gráfico anterior del histograma de los residuos con la superposición de la distribución normal en rojo, podemos apreciar que la distribución de estos no se aproxima a una normal $N(0,1)$. No obstante, teniendo en cuenta que el conjunto de datos no presenta un número de observaciones reducido, la ausencia de normalidad no alteraría la estimación de la varianza ni el contraste ANOVA realizado.

CONTRASTE DE NORMALIDAD.

Con el test de normalidad de Shapiro-Wilk, se plantea el siguiente contraste de hipótesis:

$$H_0 \equiv \text{los residuos siguen una distribución normal}$$

$$H_1 \equiv \text{los residuos no siguen una distribución normal}$$

```
> # Normalidad: Prueba de Shapiro-Wilk
> shapiro.test(residuos)
```

Shapiro-Wilk normality test

```
data:  residuos
W = 0.94074, p-value = 2.072e-14
```

Rechazamos la hipótesis nula de normalidad en los residuos dado que el p-valor se encuentra por debajo del nivel de significación estadística $\alpha=0.05$.

CONTRASTE DE IGUALDAD DE VARIANZAS.

Con el test de Levene, se plantea el siguiente contraste de hipótesis sobre la igualdad de varianzas:

$$H_0 \equiv \sigma_1^2 = \sigma_2^2 = \dots = \sigma_I^2$$

$$H_1 \equiv \sigma_1^2 \neq \sigma_2^2 \text{ para algún } i \neq j$$

```
> # Homoscedasticidad (varianza constante):
> leveneTest(LM.1) # Prueba de Levene
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  3  9.6001 3.418e-06 ***
      574
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Rechazamos la hipótesis nula de igualdad de varianzas entre las diferentes dietas.

Para este conjunto de datos, con diseño es más o menos equilibrado a pesar de que no es balanceado, no habría un efecto destacable de las varianzas desiguales sobre la estimación de la varianza y el contraste ANOVA realizados.

TEST DE LJUNG-BOX (ALEATORIEDAD).

Con el test de Ljung-Box se verifica la independencia de residuos:

```
> Box.test(residuos, lag = 5, type = "Ljung-Box")           # Prueba de Ljung-Box

Box-Ljung test

data:  residuos
X-squared = 400.5, df = 5, p-value < 2.2e-16
```

Tal y como se esperaba dada la estructura del conjunto de datos, donde se aprecia un orden de la toma de registros, los residuos no son independientes. En este caso, a diferencia de los otros contrastes, el efecto es más grave, por lo que lo más adecuado sería repetir el experimento y aleatorizar la toma e inserción de datos.

EVALUACIÓN DE DATOS ATÍPICOS.

Llevando a cabo la prueba de Bonferroni, se alcanza el siguiente resultado:

```
> # Datos atípicos con la prueba de Bonferroni:
> outlierTest(LM.1)
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferroni p
400  3.362071      0.00082518      0.47695
```

El análisis nos muestra que la observación 400 podría resultar atípica, aunque el p-valor ajustado obtenido nos lleva a mantener la hipótesis nula de que no sucede de tal forma. En caso contrario, se debería tener en cuenta que las conclusiones alcanzadas para el modelo dependen en gran medida de esa observación, un aspecto que el experimentador debe revisar con rigurosidad.

CONTRASTE NO PARAMÉTRICO: KRUSKAL-WALLIS.

En base a los resultados obtenidos en el diagnóstico del modelo, se lleva a cabo un contraste no paramétrico de Kruskal-Wallis para verificar si el diseño de experimento en valoración es correcto con respecto a la existencia de diferencias determinadas con las comparaciones múltiples:

```
> ##### Contraste no paramétrico: Kruskal-Wallis: #####
> kruskal.test(weight~diet)

Kruskal-Wallis rank sum test

data:  weight by diet
Kruskal-Wallis chi-squared = 24.45, df = 3, p-value = 2.012e-05
```

Viendo el p-valor obtenido, se rechaza la hipótesis nula de ausencia de diferencias estadísticamente significativas entre las dietas.

CONCLUSIONES.

Atendiendo a los resultados obtenidos en el presente trabajo, se concluye que la dieta influye en el peso de los pollos analizados. En concreto, se han observado diferencias estadísticamente significativas en el peso de los pollos obtenido al comparar las dietas 1-3 y 1-4. Teniendo en cuenta el peso medio alcanzado con la dieta 3, se le recomendaría al cliente optar por esta para la compra a la casa comercial en los próximos años.

No obstante, lo más recomendable sería replantear el diseño del experimento dada la dependencia detectada entre las observaciones, junto con la ausencia de normalidad y homocedasticidad. El nuevo diseño implicaría un mayor tamaño muestral y la aleatorización de la toma de datos. Además, sería interesante la evaluación de la influencia del tiempo en el proceso. A pesar de que no se muestra nada al respecto, cabe destacar que esto último también se ha llevado a cabo con el conjunto de datos. Sin embargo, dado que se lograba una mejora del R^2 pero no había modificaciones en torno a la normalidad, homocedasticidad ni aleatoriedad, se ha optado por una elaboración de trabajo centrada en la dieta.

BIBLIOGRAFÍA.

Summers, J. D., Leeson, S., & Spratt, D. (1990). Influence of diet composition on weight gain and skeletal development of White Leghorn chicks reared to four weeks of age. *Canadian Journal of Animal Science*, 71, 185-190.