Mini-Proyecto de Estadística (Clase Práctica 8)

Daniel Alberto García Pérez D.GARCIA@ESTUDIANTES.MATCOM.UH.CU
Grupo C412

Roberto Marti Cedeño R.MARTI@ESTUDIANTES.MATCOM.UH.CU
Grupo C412

Leonel Alejandro García López L.GARCIA3@ESTUDIANTES.MATCOM.UH.CU
Grupo C412

Tutor(es):
Msc. Dalia Diaz Sistachs, *LASO*

Resumen

Este informe corresponde a la realización de un ejercicio de la clase práctica número 8 correspondiente a dicha conferencia sobre el tema de el análisis de varianzas o **ANOVA**, técnica Estadística que permite la comparación de las medias de una característica en varias poblaciones. Para ello se nos otorgó el ejercicio número 1 de la clase práctica, el cual será presentado en la primera sección y resulto en las siguientes.

1. El Ejercicio

Se hace un estudio sobre la efectividad de tres marcas de spray para matar moscas. Para ello, cada producto se aplica a un grupo de 100 moscas, y se cuenta el número de moscas muertas expresado en porcentajes. Se hacen seis réplicas y los resultados obtenidos se muestran a continuación.

	Número de Réplica					
Marca de Spray	1	2	3	4	5	6
1	72	65	67	75	62	73
2	55	59	68	70	53	50
3	64	74	61	58	61	69

Table 1: Resultados del experimento

a) Dibuje las gráficas de medias y los diagramas de caja simultáneos, interprételos.

b) Formule la hipótesis adecuada y el modelo estadístico.

c) ¿Existe diferencia entre la efectividad promedio de los productos en spray?

d) ¿Hay algún spray mejor? Argumente su respuesta.

e) Verifique los supuestos de normalidad y de igual varianza entre las marcas

2. Gráfica de medias y diagrama de cajas simultáneo

Para la obtención de la gráfica pedida nos apoyamos en el lenguaje **R**, como ya es usual en esta asignatura. El código fuente se muestra a continuación:

```
library(lmtest)

brand <- c(rep('S1',6), rep('S2',6), rep('S3',6))

reply <- c(72, 65, 67, 75, 62, 73, 55, 59, 68, 70, 53, 50, 64, 74, 61, 58, 51, 69)

df <- data.frame(brand, reply)

plot(reply~brand, data=df)

brand.anova <- aov(reply~brand, data=df)
```

Figure 1: Código fuente

Al ejecutar dicho código obtuvimos la siguiente gráfica (figura 2) donde se muestran los diagramas de cajas y las medias muestrales. Las etiquetas S1, S2 y S3 corresponde a cada una de las marcas de spray; y el eje *reply* el porciento de moscas muertas.

Se aprecia que la media de spray número 1 es claramente superior a los dos restantes, esto es también observable en la tabla 1, donde cada uno de los valores en la primera fila son casi siempre mayores a los de su columna (réplicas) correspondiente. Sin embargo el comportamiento de los spray 2 y 3 es bastante similar. Sin todaviá haber realizado ninguna prueba, nos podemos arriesgar a decir que existe una diferencia entre la efectividad promedio de los productos en spray.

3. Hipótesis y Modelo Estadístico

El modelo escogido fue le de Clasificación Simple. Tenemos claramente el factor *tipo de spray*, para el cual queremos saber si existe diferencia entre la efectividad media de los mismos; sin embargo dados los datos del problema no existe ningún otro factor a tener en cuenta

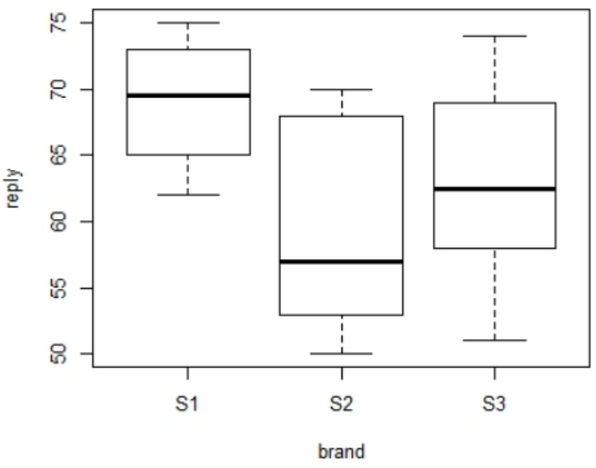

Figure 2: Diagrama de cajas simultáneo

en el modelo, pues contamos solo con cada una de las réplicas del experimento realizado con las moscas, las últimas son potencialmente un factor a tener en cuenta, pues pueden existir moscas más resistentes que otras, pero no es posible probar cada grupo de mosca con cada tipo de spray(moririan antes de realizar todos los experimentos y no llegan con la misma vitalidad a cada uno). Con lo cual ignoramos este factor y solo tendremos en cuenta el inicial, por lo que procedemos con el Modelo de Clasificación Simple.

Ahora bien, ¿existe diferencia entre la efectividad promedio de los productos en spray? La respuesta a esta pregunta es el resultado de contrastar las hipótesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \text{existen } 1 \leq i < j \leq 3 \text{ tales que } \mu_i \neq \mu_j$$

Donde μ_i denota la efectividad media del tipo de spray i. Como estudiamos en conferencias si denotamos y_{ij} como el porciento de moscas aniquiladas por el tipo de spray i en la réplica j, la misma se puede escribir como: $y_{ij} = \mu_i + e_{ij}$, siendo e_{ij} el error experimental o la perturbación.

4. Diferencias entre la efectividad promedio de los tipos de spray

Realizamos la prueba de hipótesis planteada anteriormente para un nivel de significación de $\alpha = 0.1$. Nuevamente apoyándonos en el lenguaje **R**, la forma de realizarla se muestra en la figura 1. Los resultados de la misma son los siguientes (figura 3).

```
> summary(brand.anova)
            Df Sum Sq Mean Sq F value Pr(>F)
brand        2  296.3  148.17   2.793 0.0931 .
Residuals   15  795.7   53.04
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: Prueba de Hipótesis

Los resultados arrojan un p-value de 0.0931, el cual es suficiente para nuestro nivel de significación $\alpha = 0.1$, con lo cual rechazamos la hipótesis nula y podemos afirmar que existe diferencia entre la efectividad promedio de los tipos de spray con un nivel de confianza del 90%.

El mejor spray es el primero. Dado el éxito de la prueba de hipótesis, partimos de la existencia de una diferencia, ahora bien, la evidencia visual nos muestra que la media del spray 1 es superior, incluso si tenemos en cuenta los extremos del intervalo, estos también son superiores a los de los otros dos tipos de sprays.

5. Supuestos de normalidad y de igual varianza

La validez de los resultados obtenidos en cualquier análisis de varianza queda supeditada a que los supuestos del modelo se cumplan. Estos supuestos son:

1. Los e_{ij} siguen una distribución normal con media cero.
2. Los e_{ij} son independientes entre sí.
3. Los residuos de cada tratamiento tienen la misma varianza σ^2.

Comenzando por las pruebas gráficas. Las mismas se obtuvieron nuevamente usando **R**, los resultados se muestran a continuación en las figuras 4, 5 y 6.

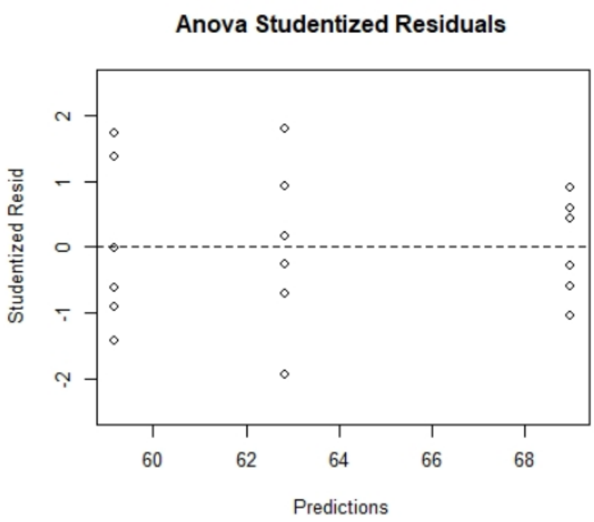

Figure 4: Residuos

Primeramente, en el gráfico estandarizado de residuos (figura 4), notamos los puntos muy dispersos, por lo que no aseguramos el supuesto de varianza constante.

En el gráfico de predichos contra los residuos (figura 5), todos los puntos tienden a estar sobre la una misma recta, o sea están bastante alineados, a excepción de dos puntos aberrantes que podemos tolerar y seguir afirmando que se cumple el supuesto de normalidad.

Finalmente, en la gráfica del histograma de residuos (figura 6), a pesar de ser pocos los datos dicha gráfica se asemeja decentemente a la normal. Con lo cual afirmamos que se cumple el supuesto de normalidad.

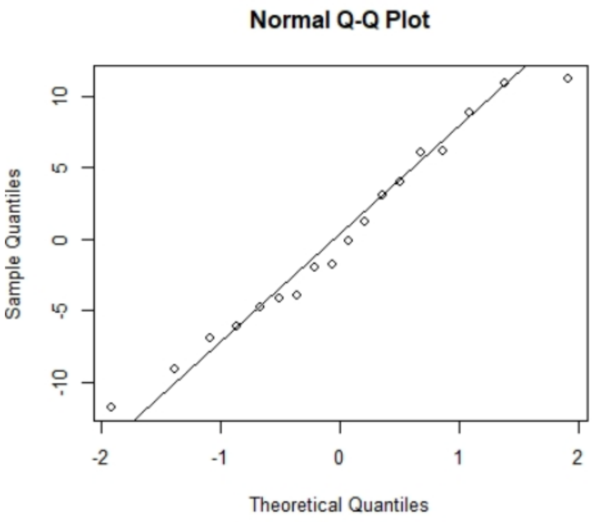

Figure 5: Predichos contra los residuos

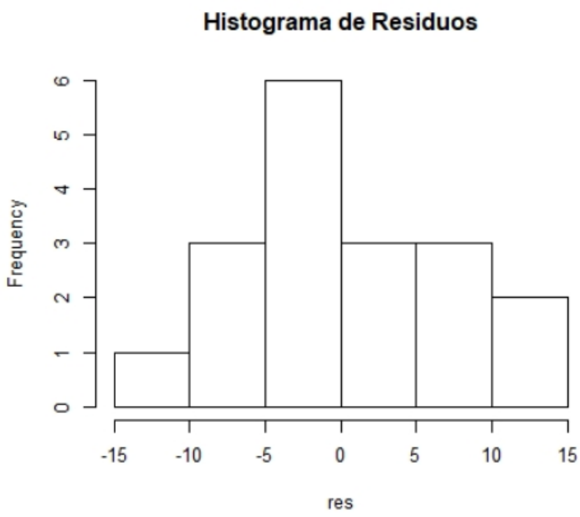

Figure 6: Histograma de residuos

Ahora contrastemos los resultados anteriores con las pruebas concluyentes correspondientes. De igual forma estas fueron realizadas utilizando **R**.

```
> shapiro.test(res)

        Shapiro-Wilk normality test

data:  res
W = 0.96797, p-value = 0.7589
```

Figure 7: Test de Normalidad Shapiro-Wilk

Los resultados arrojados por el test en la figura 7 (Test de Shapiro-Wilk) muestran que la prueba no es significativa, por lo que no podemos rechazar H_0 que es el supuesto de normalidad. Con lo cual podemos asumir que el mismo se cumple y continuar con los dos test restantes.

```
> bartlett.test(res, df$brand)

        Bartlett test of homogeneity of variances

data:  res and df$brand
Bartlett's K-squared = 1.1889, df = 2, p-value = 0.5519
```

Figure 8: Test de igual varianza Bartlett

De igual forma en la prueba de la figura 8 (Test de Bartlett), el p-value muestra que tampoco es significativa, con lo que confirmamos el supuesto de igualdad de las varianzas.

```
> dwtest(brand.anova)

        Durbin-Watson test

data:  brand.anova
DW = 2.0492, p-value = 0.3444
alternative hypothesis: true autocorrelation is greater than 0
```

Figure 9: Test Durbin-Watson

Finalmente, y de igual forma la prueba de la figura 9 (Test Durbin-Watson) muestra que el test no es significativo, por tanto asumimos el supuesto de independencia.

En conclusión, los tres supuestos los podemos tomar como válidos y por ende todos los resultados iniciales son válidos y podemos afirmar con seguridad que existe diferencia entre la efectividad promedio de los tipos de spray.