Universidad Simón Bolívar
Dpto. de Cómputo Científico y Estadística
CO3321 / Estadística
Gabriela Hernández

Análisis de Varianza / Diseño de Experimentos.

Resultados

Pregunta 4. Se realiza una prueba para conocer si existen diferencias entre camadas.

Se define,

$$\mu_i = \text{Peso Promedio en la camada i, } i = 1, ..., 8.$$

El modelo general considerado es:

$$Y_{ij} = \mu_i + \varepsilon_{ij}\ , \quad \begin{array}{l} \text{i} = 1, \ldots, 8; \\ \text{j} = 1, \ldots, n_i \end{array}$$

Las hipótesis a contrastar son:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_7 = \mu_8 \text{ vs. } H_1 : \text{algún } \mu_i \text{ distinto.}$$

El modelo considerado bajo R viene dado por, Modelo: (peso $\sim$ camada)

Tabla de Análisis de Varianza

```
             gl  Sum Sq  Mean Sq   F value   Pr(>F)
camada        7  7.4889  1.0698    2.9916    0.01092 *
Residuals    48 17.1654  0.3576
---
Signif. codes:  0  ‘***’  0.001  ‘**’  0.01  ‘*’  0.05  ‘.’  0.1  ‘ ’  1
```

El estadístico de contraste es $F^* = 2.9916 > F_{(7,48)}$. Luego se rechaza la hipótesis nula, y podemos decir que existen diferencias en los pesos al nacer (en libras) de cerdos pertenecientes a diferentes camadas. Para conocer si existen diferencias entre los pesos de los cerdos pertenecientes a camadas grandes y los pesos de los cerdos en camadas pequeñas se realiza un contraste por mínima diferencia significativa. Es decir probaremos la hipótesis,

$$H_0 : \mu_i = \mu_j \text{ vs. } H_1 : \mu_i \neq \mu_j \ i \neq j$$

Los resultados al aplicar M.D.S. con $\alpha = 0.05$ viene dada por:

```
 > pairwise.t.test(peso,camada,p.adjust.method="none")

        Pairwise comparisons using t tests with pooled SD

data:  peso and camada

        camada1 camada2 camada3 camada4 camada5 camada6 camada7
camada2 0.53444    -       -       -       -       -       -
camada3 0.20974 0.07432    -       -       -       -       -
camada4 0.63629 0.30119 0.47337    -       -       -       -
camada5 0.13190 0.36424 0.01133 0.06568    -       -       -
camada6 0.86604 0.51972 0.43258 0.83859 0.17348    -       -
camada7 0.00786 0.04074 0.00032 0.00351 0.27241 0.02161    -
camada8 0.17245 0.39770 0.02315 0.09434 0.96574 0.19958 0.34693

P value adjustment method: none
```

Se puede resumir los resultados anteriores colocando los promedios de los pesos por camada ordenados de menor a mayor, y subrayando aquellas que no pueden rechazar la hipótesis de igualdad, de la manera siguiente

Promedio por Camada							
Camada1	Camada2	Camada3	Camada4	Camada5	Camada6	Camada7	Camada8
2.84	2.6625	3.18	2.975	2.367	2.9	1.983	2.35

```
Camadas pequeñas    Camadas Grandes
 7   8   5   2       1   6   4   3
 ---------
     -------------------------
                     -------------
```

A partir de estos resultados, la camada 7 se comporta en forma diferente a casi todas las camadas exceptuando las 5 y 8 que son, al igual que esta, una camada pequeña. Así mismo, la camada 3, que es una camada grande es diferente a las camadas pequeñas 5 y 8. Esto puede indicarnos que existen diferencias entre los pesos de camadas pequeñas y grandes.

Pregunta.6

(a). Contraste de hipótesis para probar la igualdad en los pesos promedios por tipo de dieta suministrada.

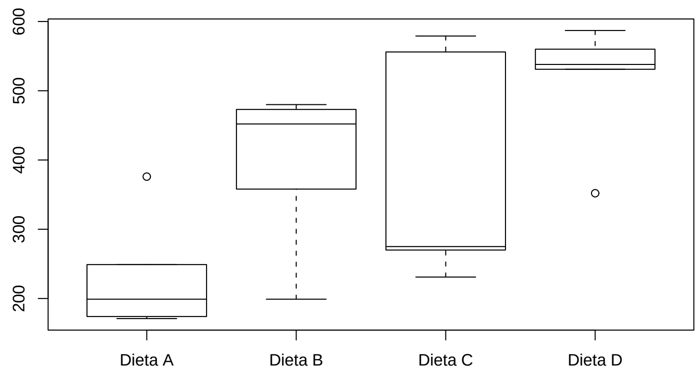

Figure 1: Diagrama de Caja para los efectos de 4 tipos de dietas en los pesos de pollos con ocho semana de nacidos

$$peso_i = \text{Peso Promedio en la Dieta i; } i = A, B, C, D.$$

$H_0 : peso_A = peso_B = peso_C = peso_D$ vs. H_1 : algún $peso_i$ distinto ; $i = A, B, C, D$

```
modelo1<-aov(peso~dieta)

summary(modelo1)
            Df Sum Sq Mean Sq F value  Pr(>F)
dieta        3 196905   65635  4.4513 0.01865 *
Residuals   16 235924   14745
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

F * $= 4.4513 > F_{(3,15)} = 2.49$. Luego se rechaza la hipótesis nula, y podemos decir que existen diferencias entre los pesos en pollos de ocho semanas de nacidos por el tipo de dieta aplicada, es decir, el tipo de dieta influye en el peso de los pollos.

(b). Suponga ahora que las repeticiones dadas en el diseño realmente representan distintos grupos de huevos, cada grupo de una gallina distinta. Considere cada grupo como un bloque y hacer el análisis de varianza adecuado.

Las hipótesis a contrastar son la igualdad entre medias por el tipo de dietas y la igualdad entre medias por grupo de huevos.

El modelo general considerado es:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \ , \quad \begin{array}{l} \text{i} = 1, \dots ,4; \\ \text{j} = 1, \dots ,5 \end{array}$$

El modelo considerado bajo R viene dado por, Modelo: (pesos $\sim$ dieta + huevo)

```
>summary(modelo1)
            Df Sum Sq Mean Sq F value   Pr(>F)
dieta        3 196905   65635  8.8479 0.002286 **
huevo        4 146907   36727  4.9509 0.013666 *
Residuals   12  89018    7418
---
Signif. codes:  0  ‘***’  0.001  ‘**’  0.01  ‘*’  0.05  ‘.’  0.1  ‘ ’  1
```

Igualdad de Dietas.

$F^* = 8.85 > F_{(3,12)} = 2.61$. Luego se rechaza la hipótesis nula, y podemos decir que existen diferencias entre los pesos en pollos de ocho semanas de nacidos por el tipo de dieta aplicada, es decir, el tipo de dieta influye en el peso de los pollos.

Igualdad de Grupo de Huevos.

$F^* = 4.95 > F_{(4,12)} = 2.48$. Luego se rechaza la hipótesis nula, y podemos decir que existen diferencias entre el grupo de huevos a los cuales pertenezcan los pollos y su peso.

(c). Explicar las diferencias entre los análisis hechos en a) y b).
El primer análisis solo hace referencia a un factor con 5 réplicas cada uno. El segundo análisis considera que una variable (grupo de huevos) puede influir en la variabilidad de pesos en los pollos. Por ello se toma esta variable como bloque.

Pregunta 10. (a). En base al gráfico ¿Cuál cree Ud. que es el mejor modelo para los datos?

$$Y_{ijh} = \mu + \alpha_i + \beta_j + \varepsilon_{ijh}$$

α_i = Tiempo de recuperación para el sexo i, i = 1, 2.

β_j = Tiempo de recuperación para el hospital j, j = A,B,C,D.

(b). La decisión es el hospital D, ya que tiene el menor tiempo en recuperación. No parece existir interacción porque las líneas son prácticamente paralelas.

Pregunta 17. (a). Realizar el gráfico de Interacciones
Observamos (figura 2) que al cambiar de tratamiento los tiempos de supervivencia para los tres tipos de veneno cambian en la misma dirección pero estas variaciones no son iguales, lo cual podemos notar por la falta de paralelismo entre los tres venenos. Esto sugiere la existencia de interacción entre los factores veneno y tratamiento .

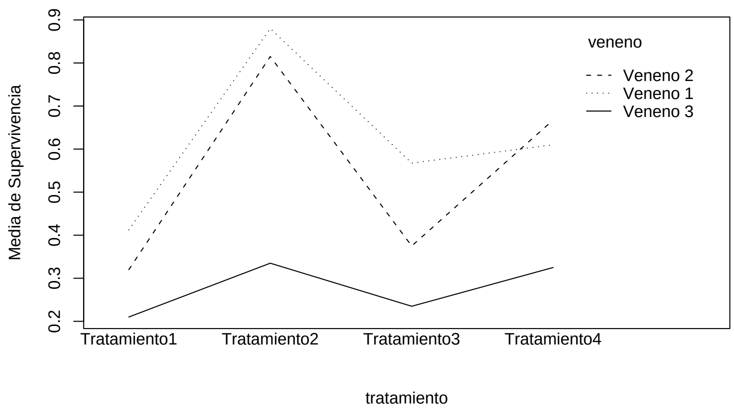

Figure 2: Interacción del Veneno por el Tipo de Tratamiento

(b - d). La hipótesis a considerar es la existencia de interacción entre los factores veneno y tratamiento.

$$H_0\text{: } \alpha\beta_{ij} = 0 \ \forall \text{ i,j vs. } H_1\text{: algún } \alpha\beta_{ij} \neq 0$$

El estadístico de contraste $F^* = 1.8743 < F_{(3,36)} = 2.866266$ lo cual no permite rechazar H_0. Luego la interacción entre estos factores no es significativa.
Se ajusta un nuevo modelo sin la interacción, sólo con efectos principales y se prueba la significancia para estos.

$$H_0\text{:}\alpha_i = 0 \ \forall\text{i vs. } H_1\text{: algún } \alpha_i \neq 0$$

$$H_0\text{:}\beta_j = 0 \ \forall\text{j vs. } H_1\text{: algún } \beta_j \neq 0$$

Los resultados en las dos pruebas rechazan a H_0 lo cual indica que los factores son significativos.

<u>Tabla de Análisis de Varianza</u>

```
> summary(modelo1)
            Df  Sum Sq Mean Sq F value    Pr(>F)
veneno       2 1.03301 0.51651  20.643 5.704e-07 ***
tratamiento  3 0.92121 0.30707  12.273 6.697e-06 ***
Residuals   42 1.05086 0.02502
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
```

El modelo ajustado viene dado por los factores tratamiento y veneno, no considera significativa la interacción.

(e). Los residuos presentan una variabilidad que no es constante (Heterocedasticidad), esto es, los residuos no cumplen con uno de los supuestos (varianza constante).

(f). Debemos determinar una transformación para los datos originales y solucionar el problema de heterocedasticidad. Una alternativa es estimar una transformación a partir de la suposición:

$$\text{Log(Desviación Estándar i)} = \log q + \alpha \text{Log(Media i)}$$

Los resultados son los siguientes:

```
lm(formula = log(des) ~ log(media))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.9390     0.2653  -3.539  0.00537 **
media         2.0203     0.2794   7.231 2.82e-05 ***
```

$\alpha = 2.0203$
Luego 1 - $\alpha \cong$ -1 = λ
La transformación sugerida es la inversa de los datos.

(g). Análisis de Varianza para los datos Transformados
Se ajusta un nuevo modelo considerando los datos transformados, la tabla de análisis de varianza viene dada por:

```
modelo3<-aov(datosa~veneno*tratamiento)
> summary(modelo3)
                   Df Sum Sq Mean Sq F value    Pr(>F)
veneno              2 34.877  17.439 72.6347 2.310e-13 ***
tratamiento         3 20.414   6.805 28.3431 1.376e-09 ***
veneno:tratamiento  6  1.571   0.262  1.0904    0.3867
Residuals          36  8.643   0.240
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1
' '  1
```

La hipótesis a considerar es la existencia de interacción entre los factores veneno y tratamiento para los datos transformados.

$$H_0\text{: } \alpha\beta_{ij} = 0 \ \forall \text{ i,j vs. } H_1\text{: algún } \alpha\beta_{ij} \neq 0$$

El p-valor es 0.3867 lo cual no permite rechazar H_0. Luego la interacción entre estos factores no es significativa.
Esto lo podemos observar en la figura 3 los niveles de venenos son aproximadamente paralelos al cambiar de un tratamiento a otro.

Se ajusta un nuevo modelo sin la interacción, sólo con efectos principales y se prueba la significancia de estos. El modelo arroja la siguiente tabla de análisis de varianza

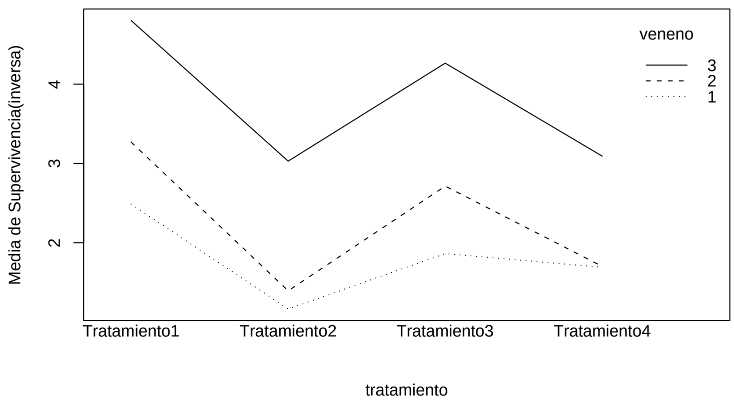

Figure 3: Interacción para los datos transformados

```
summary(modelo3)
            Df Sum Sq Mean Sq F value    Pr(>F)
veneno       2 34.877  17.439  71.708 2.864e-14 ***
tratamiento  3 20.414   6.805  27.982 4.192e-10 ***
Residuals   42 10.214   0.243
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Los factores veneno y tratamiento son significativos para explicar el tiempo de supervivencia. La figura 3 muestra el tiempo promedio de supervivencia para los datos transformados, en la cual podemos observar variaciones en la respuesta en la misma dirección a cambios en los tratamientos. La falta de paralelismo está presente sin embargo no es tan pronunciada como en los datos originales, razón por la cual la interacción entre los factores resultó ser no significativa al ajustar el modelo.

(h). Realizamos una mínima diferencia significativa para conocer si los venenos se comportan en forma similar; este análisis arrojó el siguiente resultado,

```
pairwise.t.test(datosa,veneno)

        Pairwise comparisons using t tests with pooled SD

data:  datosa and veneno

    1       2
2   0.12    -
3   5.2e-08 8.3e-06
```

La prueba indica que no existen diferencias significativas entre el veneno 1 y el 2, mientras que no se comportan igual con respecto al veneno 3. Por lo tanto, al comparar los promedios de supervivencia por cada tipo de veneno, podemos concluir que efectivamente el veneno con menor tiempo de supervivencia es el # 3, es decir es el más efectivo.

(j). El gráfico de residuos arroja el siguiente resultado

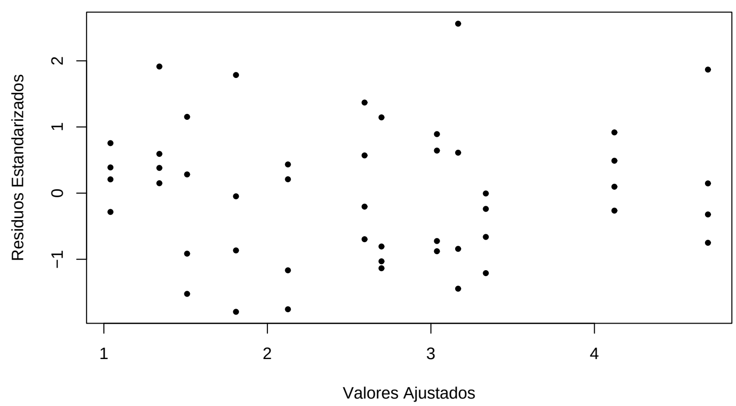

Figure 4: Gráfico de residuos para el modelo ajustado

El gráfico de residuos no parece tener patrón que indique heterocedasticidad.