Diseño de Experimentos

Diseños completamente aleatorizados DCA

UTB

Universidad Tecnologica de Bolívar

Posgrado UTB 2022

Ejemplo 1 Calidad de las suelas de un calzado

-  *Un fabricante de calzado desea mejorar la calidad de las suelas, las cuales se pueden hacer con uno de los cuatro tipos de cuero A, B, C y D disponibles en el mercado. Para ello, prueba los cueros con una máquina que hace pasar los zapatos por una superficie abrasiva. La suela de los zapatos se desgasta al pasarla por dicha superficie. Como criterio de desgaste se usa la pérdida de peso después de un número fijo de ciclos. Se prueban en orden aleatorio 24 zapatos, seis de cada tipo de cuero. Al hacer las pruebas en orden completamente al azar se evitan sesgos y las mediciones en un tipo de cuero resultan independiente de las demás. Los datos (en mg) sobre el desgaste de cada tipo de cuero se muestran en la tabla siguiente:*

Ejemplo 1 Calidad de las suelas de un calzado

Tipo de cuero	Desgaste (mg)					
A	264	260	258	241	262	255
B	208	220	216	200	213	206
C	220	263	219	225	230	228
D	217	226	215	224	220	222

- *¿Existen diferencias en el desgaste promedio de los diferentes tipos de cuero?*

Ejemplo 1 Calidad de las suelas de un calzado

- Factor de interés: Tipo de Cuero

Ejemplo 1 Calidad de las suelas de un calzado

- Factor de interés: Tipo de Cuero
- Niveles del Factor: Tipos de Cuero: A, B, C y D (cuatro niveles $K = 4$).

Ejemplo 1 Calidad de las suelas de un calzado

- Factor de interés: Tipo de Cuero
- Niveles del Factor: Tipos de Cuero: A, B, C y D (cuatro niveles $K = 4$).
- Variable de interés: $Y =$ *Desgaste en gramos de una zuela de zapato*.

Ejemplo 1 Calidad de las suelas de un calzado

- Factor de interés: Tipo de Cuero
- Niveles del Factor: Tipos de Cuero: A, B, C y D (cuatro niveles $K = 4$).
- Variable de interés: $Y =$ *Desgaste en gramos de una zuela de zapato*.
- Replicas por nivel: $n = 6$.

Ejemplo 1 Calidad de las suelas de un calzado

- Modelo estadístico:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

donde $Y_{ij} :=$ *Desgaste en gramos de la zuela* j con el cuero tipo i, $\tau_i :=$ es el efecto medio del cuero tipo i sobre el desgaste $\varepsilon_{ij} :=$ es el error aleatorio y $\mu :=$ es la media global real de todos los tratamientos.

Ejemplo 1 Calidad de las suelas de un calzado

- Modelo estadístico:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

donde $Y_{ij} :=$ *Desgaste en gramos de la zuela* j con el cuero tipo i, $\tau_i :=$ es el efecto medio del cuero tipo i sobre el desgaste $\varepsilon_{ij} :=$ es el error aleatorio y $\mu :=$ es la media global real de todos los tratamientos.

- Hipótesis del problema:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

Ejemplo 1 Calidad de las suelas de un calzado

- Modelo estadístico:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

donde Y_{ij} := *Desgaste en gramos de la zuela* j con el cuero tipo i, τ_i := es el efecto medio del cuero tipo i sobre el desgaste ε_{ij} := es el error aleatorio y μ := es la media global real de todos los tratamientos.

- Hipótesis del problema:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D = \mu$$
$$H_A : \mu_i \neq \mu_j \text{ para algún } i \neq j$$

- Significancia de la Prueba: $\alpha = 0.05$.

Solución ejemplo 1 Cálculos

-  De la tabla de datos

Tipo de cuero	Desgaste (mg)						$Y_{i.}$
A	264	260	258	241	262	255	1540
B	208	220	216	200	213	206	1263
C	220	263	219	225	230	228	1385
D	217	226	215	224	220	222	1324
							$Y_{..} = 5512$

Solución ejemplo 1 Cálculos

- Suma de Cuadrados

Solución ejemplo 1 Cálculos

- Suma de Cuadrados
  - $SCT = \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{..}^2}{N} = 1275024 - \frac{5512^2}{24} = 9101.3$

Solución ejemplo 1 Cálculos

- Suma de Cuadrados
  - $SCT = \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{..}^2}{N} = 1275024 - \frac{5512^2}{24} = 9101.3$
  - $SCTr = \sum_{i=1}^{K} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N} = \frac{1540^2+1263^2+1385^2+1324^2}{6} - \frac{5512^2}{24} = 7072.3$

Solución ejemplo 1 Cálculos

-  Suma de Cuadrados
  - $SCT = \sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{..}^2}{N} = 1275024 - \frac{5512^2}{24} = 9101.3$
  - $SCTr = \sum_{i=1}^{K} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{N} = \frac{1540^2+1263^2+1385^2+1324^2}{6} - \frac{5512^2}{24} = 7072.3$
  - $SCE = SCT - SCTr = 9101.3 - 7072.3 = 2029.0$

Solución ejemplo 1 Cálculos

-

Tabla 5. ANOVA para el DCA con Statgraphics					
Variabilidad	SC	g.l	CM	F valor	$valor - p$
Tipo de cuero	7072.3	3	2357.44	23.24	0.0000
Error	2029.0	20	101.45		
Total	9101.3	23			

Solución ejemplo 1 Cálculos

-

Tabla 5. ANOVA para el DCA con Statgraphics					
Variabilidad	SC	g.l	CM	F valor	$valor - p$
Tipo de cuero	7072.3	3	2357.44	23.24	0.0000
Error	2029.0	20	101.45		
Total	9101.3	23			

- La conclusión de la hipótesis se lee en la última columna de la tabla de ANOVA. Como el $valor\ p = 0.0000$ es menor que la significancia prefijada $\alpha = 0.05$, se rechaza H_0 y se acepta que al menos un par de cueros tienen un desgaste promedio diferente.

Solución ejemplo 1

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*

Solución ejemplo 1

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*
-  Para responder a esta pregunta, se realizan todas las comparaciones posibles dos a dos entre las medias de los tratamientos:

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*
-  Para responder a esta pregunta, se realizan todas las comparaciones posibles dos a dos entre las medias de los tratamientos:
  - usando *métodos de comparaciones múltiples*

Solución ejemplo 1

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*
-  Para responder a esta pregunta, se realizan todas las comparaciones posibles dos a dos entre las medias de los tratamientos:
  - usando *métodos de comparaciones múltiples*
  - usando gráfico de medias (means plot)

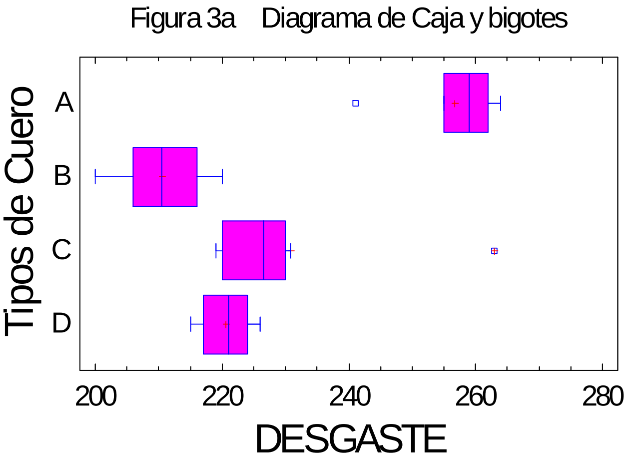
Solución ejemplo 1

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*
-  Para responder a esta pregunta, se realizan todas las comparaciones posibles dos a dos entre las medias de los tratamientos:
  - usando *métodos de comparaciones múltiples*
  - usando gráfico de medias (means plot)
  - y el diagrama de cajas (box-and-whisker plot) como se presentan a continuacion en las figuras 3a y 3b.

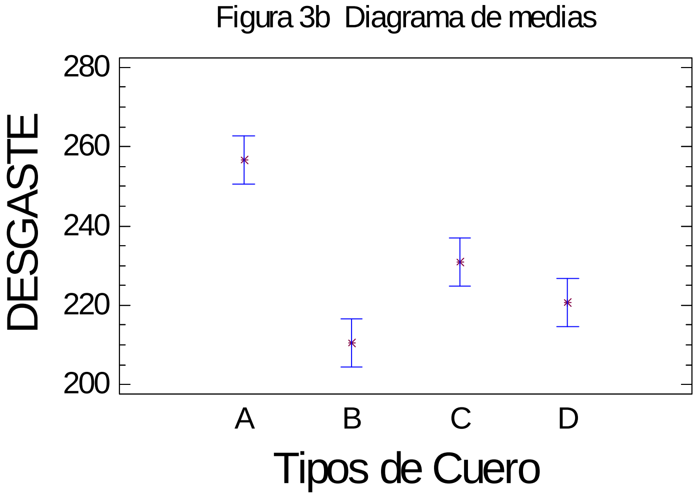
Solución ejemplo 1

-  Si al menos un cuero se desgasta diferente de otro, entonces *¿cuáles tipos de cuero son diferentes entre sí?*
-  Para responder a esta pregunta, se realizan todas las comparaciones posibles dos a dos entre las medias de los tratamientos:
  - usando *métodos de comparaciones múltiples*
  - usando gráfico de medias (means plot)
  - y el diagrama de cajas (box-and-whisker plot) como se presentan a continuacion en las figuras 3a y 3b.
-  En ambas gráficas se observa que existen almenos dos tratamientos diferentes, siendo el tratamiento A quien más difiere del resto con un 95% de confianza.

Solución ejemplo 1



Solución ejemplo 1



Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
    - A<-c(264, 260, 258, 241, 262, 255)
    - B<-c(208, 220, 216, 200, 213, 206)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)
  - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)
  - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
  - split(longitud, T.Cuero)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)
  - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
  - split(longitud, T.Cuero)
  - tapply (longitud, T.Cuero, summary)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)
  - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
  - split(longitud, T.Cuero)
  - tapply (longitud, T.Cuero, summary)
  - plot(longitud ~T.Cuero,col=c("red", "blue", "yellow", "orange"))

Solución ejemplo 1 usando el software R

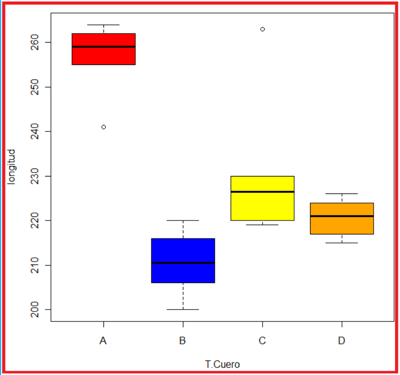
- Consideremos el script para los diferentes tipos de cuero
    - A<-c(264, 260, 258, 241, 262, 255)
    - B<-c(208, 220, 216, 200, 213, 206)
    - C<-c(220, 263, 219, 225, 230, 228)
    - D<-c(217, 226, 215, 224, 220, 222)
    - longitud<-c(A, B, C, D)
    - T.Cuero<-rep(1:4, each=6)
    - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
    - split(longitud, T.Cuero)
    - tapply (longitud, T.Cuero, summary)
    - plot(longitud ~T.Cuero,col=c("red", "blue", "yellow", "orange"))
    - p.aov<-aov(longitud ~T.Cuero)

Solución ejemplo 1 usando el software R

- Consideremos el script para los diferentes tipos de cuero
  - A<-c(264, 260, 258, 241, 262, 255)
  - B<-c(208, 220, 216, 200, 213, 206)
  - C<-c(220, 263, 219, 225, 230, 228)
  - D<-c(217, 226, 215, 224, 220, 222)
  - longitud<-c(A, B, C, D)
  - T.Cuero<-rep(1:4, each=6)
  - T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
  - split(longitud, T.Cuero)
  - tapply (longitud, T.Cuero, summary)
  - plot(longitud ~T.Cuero,col=c("red", "blue", "yellow", "orange"))
  - p.aov<-aov(longitud ~T.Cuero)
  - summary(p.aov)

Solución ejemplo 1 usando el software R

```
Mensajes de aviso perdidos
package "methods" in options("defaultPackages") was not found
> A<-c(264, 260, 258, 241, 262, 255)
> B<-c(208, 220, 216, 200, 213, 206)
> C<-c(220, 263, 219, 225, 230, 228)
> D<-c(217, 226, 215, 224, 220, 222)
> longitud<-c(A, B, C, D)
> T.Cuero<-rep(1:4, each=6)
> T.Cuero<-gl(4,6,labels=c("A","B","C","D"))
> split(longitud, T.Cuero)
$A
[1] 264 260 258 241 262 255

$B
[1] 208 220 216 200 213 206

$C
[1] 220 263 219 225 230 228

$D
[1] 217 226 215 224 220 222

> tapply (longitud, T.Cuero, summary)
$A
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  241.0   255.8   259.0   256.7   261.5   264.0

$B
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  200.0   206.5   210.5   210.5   215.2   220.0

$C
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  219.0   221.2   226.5   230.8   229.5   263.0

$D
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  215.0   217.8   221.0   220.7   223.5   226.0

> plot(longitud ~ T.Cuero)
> p.aov<-aov(longitud ~ T.Cuero)
> summary(p.aov)
            Df Sum Sq Mean Sq F value    Pr(>F)
T.Cuero      3 7072.3 2357.44  23.238 1.002e-06 ***
Residuals   20 2029.0  101.45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Solución ejemplo 1 usando el software R

- Otra posibilidad es definir el modelo lineal y obtener la tabla con la instruccióon anova.

Solución ejemplo 1 usando el software R

- Otra posibilidad es definir el modelo lineal y obtener la tabla con la instruccióon anova.
    - > g.lm <- lm(longitud ~T.Cuero)

Solución ejemplo 1 usando el software R

- Otra posibilidad es definir el modelo lineal y obtener la tabla con la instrucciíon anova.
  - $>$ g.lm $<$- lm(longitud ~T.Cuero)
  - $>$ anova(g.lm)

```
Analysis of Variance Table

Response: longitud
          Df Sum Sq Mean Sq F value    Pr(>F)
T.Cuero    3 7072.3 2357.44  23.238 1.002e-06 ***
Residuals 20 2029.0  101.45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Solución ejemplo 1 usando el software R

- Otra posibilidad es definir el modelo lineal y obtener la tabla con la instrucciíon anova.
    - > g.lm <- lm(longitud ~T.Cuero)
    - > anova(g.lm)

```
Analysis of Variance Table

Response: longitud
          Df Sum Sq Mean Sq F value    Pr(>F)
T.Cuero    3 7072.3 2357.44  23.238 1.002e-06 ***
Residuals 20 2029.0  101.45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

- Como el *p-valor* es muy pequeño se concluye que hay diferencias muy significativas entre las tres los tipos de cuero.

Solución ejemplo 1 usando el software R

- Las estimaciones de los paráametros se obtienen con:

Solución ejemplo 1 usando el software R

- Las estimaciones de los paráametros se obtienen con:
    - model.tables(p.aov)

Solución ejemplo 1 usando el software R

- Las estimaciones de los paráametros se obtienen con:
  - model.tables(p.aov)
  - model.tables(p.aov, type = "mean")

```
> model.tables(p.aov)
Tables of effects

 T.Cuero
T.Cuero
      A       B       C       D
 27.000 -19.167   1.167  -9.000


> model.tables(p.aov, type = "mean")
Tables of means
Grand mean

229.6667

 T.Cuero
T.Cuero
     A      B      C      D
256.67 210.50 230.83 220.67
```

Solución ejemplo 1 usando el software R

- El modelo lineal contiene mucha información que se puede obtener con la instrucción summary.

Solución ejemplo 1 usando el software R

- El modelo lineal contiene mucha información que se puede obtener con la instrucción summary.
  - summary(g.lm)

```
> summary(g.lm)

Call:
lm(formula = longitud ~ T.Cuero)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.667  -4.792  -0.750   3.833  32.167 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  256.667      4.112  62.419  < 2e-16 ***
T.CueroB     -46.167      5.815  -7.939 1.31e-07 ***
T.CueroC     -25.833      5.815  -4.442  0.00025 ***
T.CueroD     -36.000      5.815  -6.191 4.78e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.07 on 20 degrees of freedom
Multiple R-squared: 0.7771,	Adjusted R-squared: 0.7436 
F-statistic: 23.24 on 3 and 20 DF,  p-value: 1.002e-06
```

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
    - model.matrix(g.lm)

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
    - model.matrix(g.lm)
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
  - model.matrix(g.lm)
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es
  - ECM <- deviance(p.aov)/p.aov$df.residual

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
  - model.matrix(g.lm)
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es
  - ECM <- deviance(p.aov)/p.aov$df.residual
  - ECM

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
    - `model.matrix(g.lm)`
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es
    - `ECM <- deviance(p.aov)/p.aov$df.residual`
    - `ECM`
- Esta estimacióon también se obtiene directamente del modelo lineal

Solución ejemplo 1 usando el software R

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
  - model.matrix(g.lm)
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es
  - ECM <- deviance(p.aov)/p.aov$df.residual
  - ECM
- Esta estimacióon también se obtiene directamente del modelo lineal
  - > summary(g.lm)$sigma^2

- Sin embargo, las estimaciones que se obtienen aquí corresponden al modelo con la restricción que el parámetro de la primera especie es cero. Esta es la opción por defecto para los modelos lineales en R. Se puede ver la matriz del diseño en esta situación:
  - model.matrix(g.lm)
- El error cuadráatico medio o estimacióon insesgada de la varianza del modelo es
  - ECM <- deviance(p.aov)/p.aov$df.residual
  - ECM
- Esta estimacióon también se obtiene directamente del modelo lineal
  - > summary(g.lm)$sigma^2
  - [1] 0.3648254

Comparaciones de rangos múltiples

- Cuando se rechaza $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$ y se acepta $H_1 : \mu_i \neq \mu_j$, para algún $i \neq j$; es necesario investigar cuáles tratamientos resultaron diferentes o cuáles tratamientos provocan la diferencia.

Comparaciones de rangos múltiples

- Cuando se rechaza $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$ y se acepta $H_1 : \mu_i \neq \mu_j$, para algún $i \neq j$; es necesario investigar cuáles tratamientos resultaron diferentes o cuáles tratamientos provocan la diferencia.
- La respuesta consiste en hacer la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 : &\quad \mu_i = \mu_j \\ H_1 : &\quad \mu_i \neq \mu_j \text{ para toda } i \neq j \end{aligned} \tag{1}$$

la cuál se prueba haciendo uso de los Métodos de comparaciones de rángos múltiples.

- Cuando se rechaza $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$ y se acepta $H_1 : \mu_i \neq \mu_j$, para algún $i \neq j$; es necesario investigar cuáles tratamientos resultaron diferentes o cuáles tratamientos provocan la diferencia.
- La respuesta consiste en hacer la siguiente prueba de hipótesis:

$$\begin{aligned} H_0 &: \quad \mu_i = \mu_j \\ H_1 &: \quad \mu_i \neq \mu_j \text{ para toda } i \neq j \end{aligned} \tag{1}$$

la cuál se prueba haciendo uso de los Métodos de comparaciones de rángos múltiples.

- Entre ellos los más usados son:

Método LSD (diferencia mínima significativa)

-  Para k tratamientos se tienen en total $k(k-1)/2$ pares de medias, entonces, se rechaza $H_0 : \mu_i = \mu_j$ si ocurre $\left|\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right| > LSD$, donde:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{CME\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Método LSD (diferencia mínima significativa)

-  Para k tratamientos se tienen en total $k(k-1)/2$ pares de medias, entonces, se rechaza $H_0 : \mu_i = \mu_j$ si ocurre $\left|\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right| > LSD$, donde:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{CME\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- LSD se llama diferencia mínima significativa (least significant difference).

Método LSD (diferencia mínima significativa)

-  Para k tratamientos se tienen en total $k(k-1)/2$ pares de medias, entonces, se rechaza $H_0 : \mu_i = \mu_j$ si ocurre $\left|\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}\right| > LSD$, donde:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{CME\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

- LSD se llama diferencia mínima significativa (least significant difference).
  - Es la diferencia mínima que debe haber entre dos medias muestrales para poder considerar que los tratamientos son significativamente diferentes.

-  Si el diseño es *balanceado*, es decir, si $n_i = n_j = n$ la diferencia mínima significativa se reduce a:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME}$$

Método LSD (diferencia mínima significativa)

-  Si el diseño es *balanceado*, es decir, si $n_i = n_j = n$ la diferencia mínima significativa se reduce a:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME}$$

  - Esta prueba se usa en comparaciones simultáneas cuando:

Método LSD (diferencia mínima significativa)

-  Si el diseño es *balanceado*, es decir, si $n_i = n_j = n$ la diferencia mínima significativa se reduce a:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME}$$

  - Esta prueba se usa en comparaciones simultáneas cuando:
    - la prueba F es significativa, y

-  Si el diseño es *balanceado*, es decir, si $n_i = n_j = n$ la diferencia mínima significativa se reduce a:

$$LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME}$$

- Esta prueba se usa en comparaciones simultáneas cuando:
  - la prueba F es significativa, y
  - las comparaciones fueron planeadas antes de ejecutar el experimento.

- $LSD = t_{\alpha/2,N-k} \cdot \sqrt{\frac{2}{n}CME} = t_{0.025,20}\sqrt{\frac{2}{6}\,(101.45)} = 2.08597\sqrt{33.817} = 12.13$

Tests LSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	$210,500$	12.13	X		
CUERO D	6	$220,667$	12.13	X	X	
CUERO C	6	$230,833$	12.13		X	
CUERO A	6	$256,667$	12.13			X

Método LSD (diferencia mínima significativa)

Ejemplo 1

- $LSD = t_{\alpha/2,N-k} \cdot \sqrt{\frac{2}{n} CME} = t_{0.025,20} \sqrt{\frac{2}{6}(101.45)} = 2.08597\sqrt{33.817} = 12.13$

Tests LSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	210, 500	12.13	X		
CUERO D	6	220, 667	12.13	X	X	
CUERO C	6	230, 833	12.13		X	
CUERO A	6	256, 667	12.13			X

- El tipo de cuero A es significativamente diferente de los otros tipos diferentes de cuero.

- $LSD = t_{\alpha/2, N-k} \cdot \sqrt{\frac{2}{n} CME} = t_{0.025,20} \sqrt{\frac{2}{6}(101.45)} = 2.08597\sqrt{33.817} = 12.13$

Tests LSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	210, 500	12.13	X		
CUERO D	6	220, 667	12.13	X	X	
CUERO C	6	230, 833	12.13		X	
CUERO A	6	256, 667	12.13			X

- El tipo de cuero A es significativamente diferente de los otros tipos diferentes de cuero.
- Se observa que los cuero tipo B y D; y los cueros D y C no son diferentes estadísticamente.

Método de Tukey (HSD)

- En este procedimiento se usa la distribución de probabilidad de rango estudentizado, que representamos con $Q_{\alpha;m,n}$, donde m son los grados de libertad del numerador y n los grados de libertad del denominador.

Método de Tukey (HSD)

- En este procedimiento se usa la distribución de probabilidad de rango estudentizado, que representamos con $Q_{\alpha;m,n}$, donde m son los grados de libertad del numerador y n los grados de libertad del denominador.
- Se rechaza $H_0 : \mu_i = \mu_j$, si ocurre $\left|\overline{Y}_{i.} - \overline{Y}_{j.}\right| > Tukey$, donde

$$Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{CME}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.$$

Método de Tukey (HSD)

-  Cuando el diseño es *balanceado*, es decir, si $n_i = n_j = n$

$$Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME}.$$

Método de Tukey (HSD)

-  Cuando el diseño es *balanceado*, es decir, si $n_i = n_j = n$

$$Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME}.$$

-  Este procedimiento es llamado también *Diferencia Significativa Honesta,* se utiliza para realizar comparaciones múltiples de medias cuando:

Método de Tukey (HSD)

-  Cuando el diseño es *balanceado*, es decir, si $n_i = n_j = n$

$$Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME}.$$

-  Este procedimiento es llamado también *Diferencia Significativa Honesta*, se utiliza para realizar comparaciones múltiples de medias cuando:
  - *aposteriori* el diseño evidencia diferencia entre los tratamientos, y

Método de Tukey (HSD)

-  Cuando el diseño es *balanceado*, es decir, si $n_i = n_j = n$

$$Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME}.$$

-  Este procedimiento es llamado también *Diferencia Significativa Honesta*, se utiliza para realizar comparaciones múltiples de medias cuando:
    - *aposteriori* el diseño evidencia diferencia entre los tratamientos, y
    - trabaja con un error α muy cercano al declarado por el experimentador.

-  $Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME} = Q_{0.05;4;20} \sqrt{\frac{1}{6}\,(101.45)} = (3.96)\,\sqrt{16.908} = 16.283$

Tests HSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	$210,500$	16.283	X		
CUERO D	6	$220,667$	16.283	X	X	
CUERO C	6	$230,833$	16.283		X	
CUERO A	6	$256,667$	16.283			X

-  $Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n}CME} = Q_{0.05;4;20}\sqrt{\frac{1}{6}(101.45)} = (3.96)\sqrt{16.908} = 16.283$

Tests HSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	210, 500	16.283	X		
CUERO D	6	220, 667	16.283	X	X	
CUERO C	6	230, 833	16.283		X	
CUERO A	6	256, 667	16.283			X

- El tipo de cuero A es significativamente diferente de los otros tipos diferentes de cuero.

Método de Tukey (HSD)

Ejemplo 1

-  $Tukey = Q_{\alpha,k,N-k} \cdot \sqrt{\frac{1}{n} CME} = Q_{0.05;4;20} \sqrt{\frac{1}{6}(101.45)} = (3.96)\sqrt{16.908} = 16.283$

Tests HSD $\alpha = 0.05$						
METODO	n_i	Mean	LSD	Grupos		
CUERO B	6	210, 500	16.283	X		
CUERO D	6	220, 667	16.283	X	X	
CUERO C	6	230, 833	16.283		X	
CUERO A	6	256, 667	16.283			X

- El tipo de cuero A es significativamente diferente de los otros tipos diferentes de cuero.
- Se observa que los cuero tipo B y D; y los cueros D y C no son diferentes estadísticamente.

- MONTGOMERY C. DOUGLAS. Diseño y Análisis de Experimentos. Segunda Edición. LIMUSA WILEY
- GUTIERREZ H y DE LA VARA R. Análisis y Diseño de Experimentos. Segunda Edición. Mc Graw Hill.
- KUEHL ROBERT. Diseño de Experimentos. Segunda Edición. Thomson.
- VICENTE M, GIRON P, NIETO C, PÉREZ T. Diseño de Experimentos. Pearson Prentice Hall.
- DEVORE JAY. Probabilidad y Estadística para Ingeniería y Ciencias. Sexta Edición.
- WALPOLE MYERS. Probabilidad y Estadística. Cuarta Edición. Mc Graw Hill.