

Análisis de Métodos de Clasificación

Grupo de Estadística 2

2025-03-10

Introducción

En este estudio se comparan tres métodos de clasificación de frijoles según su eficiencia en términos de tiempo. Se analiza si hay diferencias significativas entre los métodos usando ANOVA y la prueba de LSD.

Diseño Experimental

- **Factor:** Método de clasificación.
- **Bloque:** Estudiante.
- **Niveles del factor:** Método 1, Método 2, Método 3.
- **Número de bloques:** 5 (uno por cada estudiante).
- **Diseño de efectos:** Diseño de efectos fijos.
- **Variable de respuesta:** Tiempo de clasificación (en segundos).
- **Aleatorización:** Los métodos se asignaron de manera aleatoria a los estudiantes para evitar sesgos.

Materiales

- Platos desechables pequeños.
- Frijoles de tres colores diferentes.
- Cronómetro.

Planteamiento del Modelo

El modelo de análisis de varianza se expresa como:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

Donde: - Y_{ij} es el tiempo de clasificación. - μ es la media general. - τ_i es el efecto del método i . - β_j es el efecto del estudiante j . - ϵ_{ij} es el error aleatorio, con los supuestos: normalidad, media cero, varianza constante e independencia.

Carga de Datos

```
library(readxl)
library(dplyr)
```

Table 1: Datos sin ordenar

Estudiante	1	2	3
De Leon	939	688	821
Gonzalez	649	642	714
Jorge	687	581	680
Estheban	556	500	685
Ariel	750	856	920

```
##
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(agricolae)

## Warning: package 'agricolae' was built under R version 4.4.3

library(tidyr)
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.4.3

##
## Adjuntando el paquete: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##   group_rows

mydata <- tibble::tibble(
  Estudiante = c("De Leon", "Gonzalez", "Jorge", "Estheban", "Ariel"),
  `1` = c(939, 649, 687, 556, 750),
  `2` = c(688, 642, 581, 500, 856),
  `3` = c(821, 714, 680, 685, 920)
)

# Mostrar datos sin ordenar
mydata %>%
  kable("latex", caption = "Datos sin ordenar") %>%
  kable_styling(latex_options = "striped")
```

Transformación de Datos

```
arranged_data <- pivot_longer(mydata, cols = "1":"3", names_to = "Metodo", values_to="Tiempo")
arranged_data
```

```
## # A tibble: 15 x 3
##   Estudiante Metodo Tiempo
##   <chr>      <chr>   <dbl>
## 1 De Leon    1         939
## 2 De Leon    2         688
## 3 De Leon    3         821
## 4 Gonzalez  1         649
## 5 Gonzalez  2         642
## 6 Gonzalez  3         714
## 7 Jorge     1         687
## 8 Jorge     2         581
## 9 Jorge     3         680
## 10 Estheban  1         556
## 11 Estheban  2         500
## 12 Estheban  3         685
## 13 Ariel     1         750
## 14 Ariel     2         856
## 15 Ariel     3         920
```

Datos ordenados

```
# Ordenar los datos por Estudiante
sorted_data <- arranged_data %>%
  arrange(Estudiante)

# Mostrar datos ordenados
sorted_data %>%
  kable("latex", caption = "Datos ordenados por estudiante") %>%
  kable_styling(latex_options = "striped")
```

Prueba de Hipótesis (ANOVA)

```
results <- aov(Tiempo ~ Metodo + Estudiante, data = arranged_data)
summary(results)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Metodo      2  30768   15384   2.817  0.1186
## Estudiante  4 152648   38162   6.987  0.0101 *
## Residuals   8  43696    5462
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 2: Datos ordenados por estudiante

Estudiante	Metodo	Tiempo
Ariel	1	750
Ariel	2	856
Ariel	3	920
De Leon	1	939
De Leon	2	688
De Leon	3	821
Estheban	1	556
Estheban	2	500
Estheban	3	685
Gonzalez	1	649
Gonzalez	2	642
Gonzalez	3	714
Jorge	1	687
Jorge	2	581
Jorge	3	680

Diferencia Significativa de Fisher (LSD)

```
lsdresults <- LSD.test(results, "Estudiante")
lsdresults
```

```
## $statistics
##      MSerror Df  Mean      CV  t.value      LSD
##      5461.95  8 711.2 10.39159 2.306004 139.1516
##
## $parameters
##      test p.adjusted      name.t ntr alpha
##      Fisher-LSD      none Estudiante  5  0.05
##
## $means
##      Tiempo      std r      se      LCL      UCL Min Max  Q25 Q50
## Ariel      842.0000  85.86035  3 42.66908 743.6049 940.3951 750 920 803.0 856
## De Leon    816.0000 125.57468  3 42.66908 717.6049 914.3951 688 939 754.5 821
## Estheban   580.3333  94.87009  3 42.66908 481.9383 678.7284 500 685 528.0 556
## Gonzalez   668.3333  39.70306  3 42.66908 569.9383 766.7284 642 714 645.5 649
## Jorge      649.3333  59.28181  3 42.66908 550.9383 747.7284 581 687 630.5 680
##
##      Q75
## Ariel    888.0
## De Leon   880.0
## Estheban 620.5
## Gonzalez 681.5
## Jorge     683.5
##
## $comparison
## NULL
##
## $groups
##      Tiempo groups
```

```
## Ariel      842.0000      a
## De Leon    816.0000      a
## Gonzalez   668.3333      b
## Jorge      649.3333      b
## Estheban   580.3333      b
##
## attr(,"class")
## [1] "group"
```

Cotización del Estudio

Se estima que el costo del estudio incluye: - Materiales: Q30 - Total: Q30

Recomendación a Glamour

El análisis ANOVA muestra que hay una diferencia significativa entre los estudiantes ($p = 0.0101$), pero no entre los métodos de clasificación ($p = 0.1186$). La prueba LSD indica que Ariel y De Leon tienen tiempos significativamente más altos que los otros estudiantes. Se recomienda a la empresa “Glamour” capacitar a los clasificadores con menor rendimiento o modificar los métodos para mejorar la eficiencia general.