

Análisis de varianza

Andrés Añez Brito

4/29/2021

El análisis de varianza es la técnica fundamental en el desarrollo de diseños de experimentos, la idea principal de esta técnica es separar la varianza total entre las fuentes con las que contribuye cada fuente de variación en el experimento. (Humberto Gutierrez, 2015)

La utilidad de esta técnica es transversal en todas las áreas de la ciencia, a nivel industrial su importancia radica en poder comparar diferentes métodos, procesos, maquinas, proveedores, materiales y cualquier componente que influya en el sistema de producción de nuestra empresa.

Humberto Gutierrez, en su libro “Análisis y diseños de experimento” nos propone un ejemplo de una aplicación a nivel industrial, el cual consiste en lo siguiente: En una empresa de manufactura se propone un tratamiento para reducir el porcentaje de productos defectuosos. Para validar la propuesta se diseñó un experimento que se producía con o sin la propuesta de mejora. Cada corrida experimental consistió en producir un lote y la variable de respuesta es el porcentaje de producto defectuoso. Se hicieron 25 replicas para cada proceso.

Se cargan los datos del modelo

```
datos <- read.csv("/Users/andresanez/OneDrive - Universidad del Magdalena/PROYECTOS/Análisis de varianza/datos.csv", header=TRUE, sep=",")
```

Se convierte la variable “Porcentaje_de_PNC” en decimal

```
datos$Porcentaje_de_PNC<- as.numeric(datos$Porcentaje_de_PNC)
PNC <- datos$Porcentaje_de_PNC
procesos <- datos$Proceso
#El resumen del dataset es el siguiente
summary(datos)
```

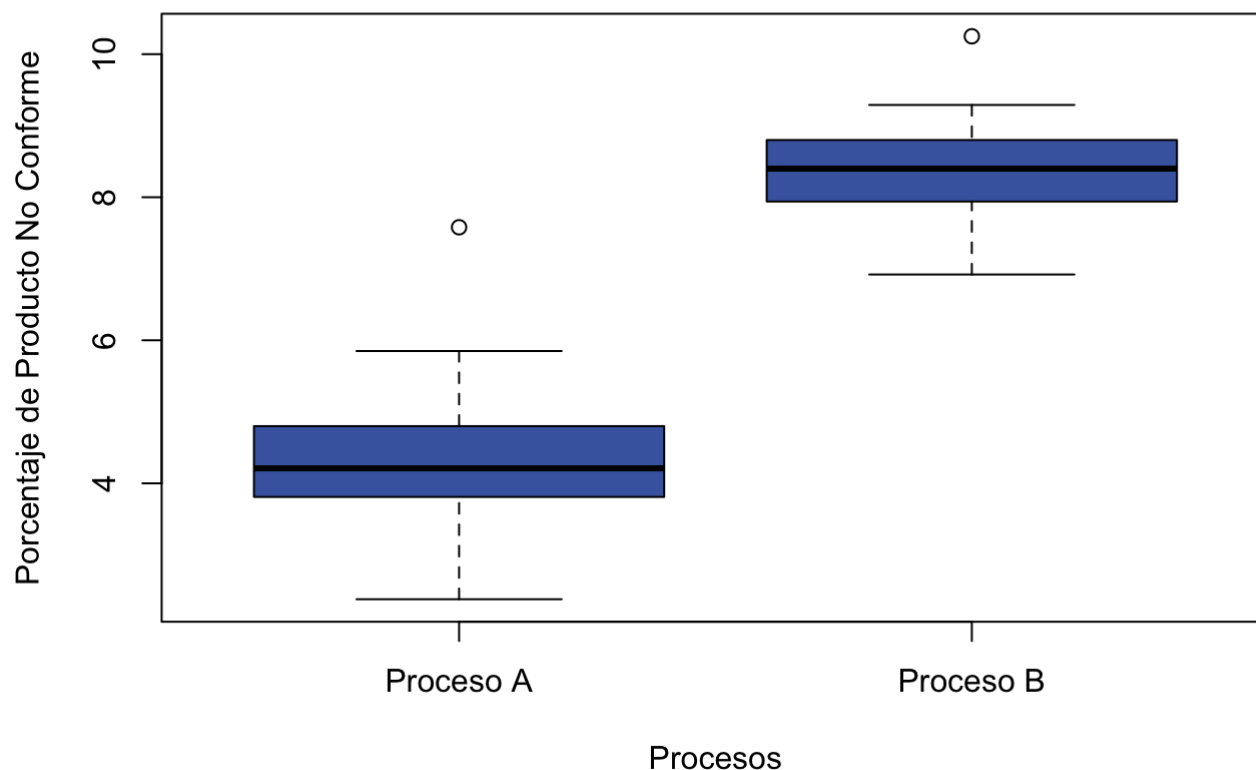
```
##      Proceso      Porcentaje_de_PNC
## Length:50      Min.   : 2.380
## Class :character 1st Qu.: 4.263
## Mode  :character Median : 6.940
##                Mean   : 6.327
##                3rd Qu.: 8.363
##                Max.   :10.250
```

El propósito del experimento es determinar si hay diferencias en los porcentajes de producto no conforme obtenidos en los dos procesos

Exploración Visual

Se realiza un box plot para explorar el comportamiento de producto no conforme por cada uno de los procesos, se puede denotar visualmente que el proceso B pareciera tener mayor incidencia en el porcentaje de producto no conforme.

```
boxplot(PNC ~ procesos, xlab="Procesos", ylab = "Porcentaje de Producto No Conforme", col = "#4767AE")
```



Modelo ANOVA

Para corroborar que el análisis visual es el correcto, se procede a realizar la prueba analítica:

```
modelo = aov(lm(PNC ~ procesos))
summary(modelo)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## procesos    1 196.69   196.69   225.5 <2e-16 ***
## Residuals   48  41.87     0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Puesto el valor p (2×10^{-16}) obtenido es menor a un nivel de significancia de 5% se puede concluir que los procesos generan diferentes niveles de producto no conforme, siendo el Proceso A el que genera niveles más bajos para el sistema productivo.

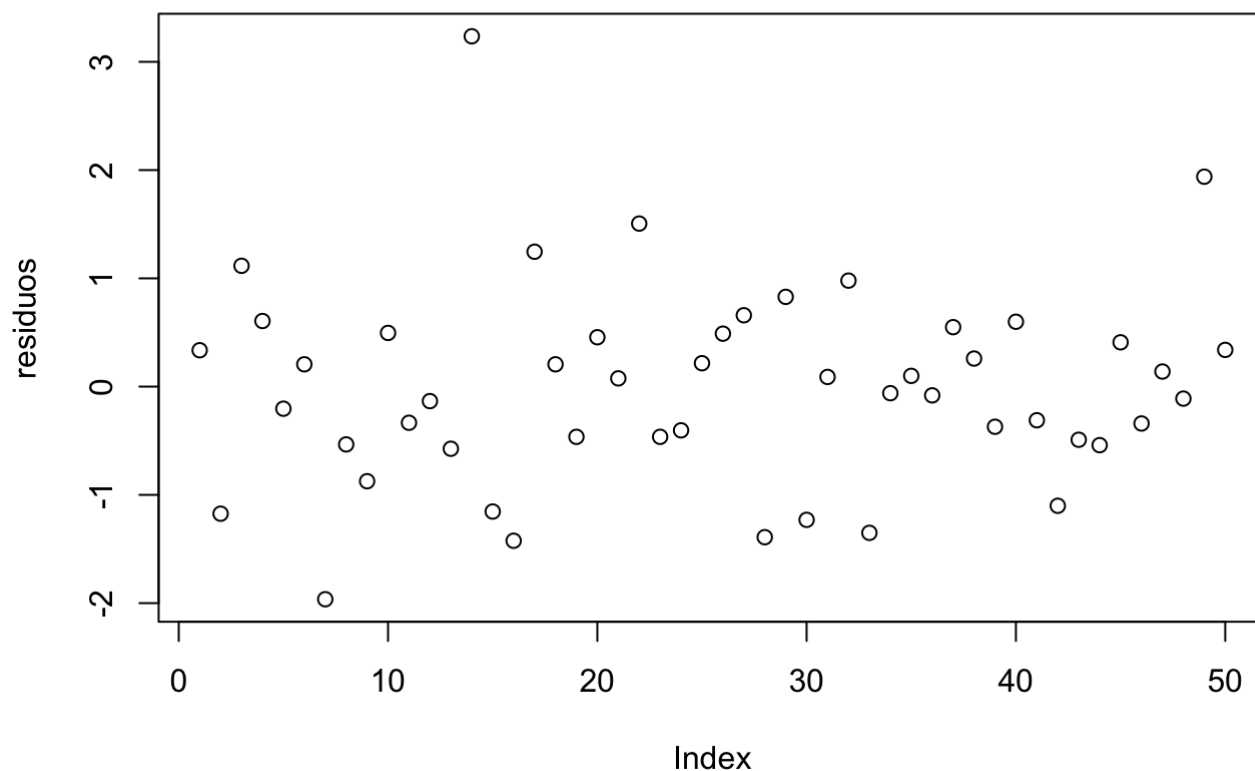
Validación del Modelo

Cada vez que se realiza un análisis de varianza es importante validar el modelo en cuanto a normalidad, independencia y homocedasticidad. Para esto, se piden los residuos del modelo de la siguiente manera:

```
residuos <- modelo$residuals
```

Independencia

```
plot(residuos)
```



No se evidencian patrones, ni tendencias, por lo que se puede concluir que las corridas cumplen con la prueba de independencia

Normalidad

Para evidenciar si los residuos siguen una distribución normal, se realiza un prueba de Shapiro Wilks

```
shapiro.test(residuos)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  residuos  
## W = 0.95784, p-value = 0.07219
```

Puesto el valor p (0,072) obtenido es mayor a un nivel de significancia de 5% no se puede rechazar la idea de que los datos provienen de una distribución normal, cumpliendo de ese modo con el presente supuesto.

Homocedasticidad

```
bartlett.test(residuos ~ procesos)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data:  residuos by procesos  
## Bartlett's K-squared = 2.3175, df = 1, p-value = 0.1279
```

Puesto el valor p (0,1279) obtenido es mayor a un nivel de significancia de 5% no se puede rechazar la idea de que los residuos presentan igualdad de varianza, cumpliendo de ese modo con el presente supuesto.

Conclusión

De esta forma, se puede concluir que el proceso A y el proceso B presentan diferencias significativas ante la generación de producto no conforme, siendo el primero el cual mejores resultados arroja, al cumplir con los 3 supuestos se puede denotar que la prueba cumple con los estándares y el desarrollo del experimento fue idóneo.