

ANOVA con Bloques

Sebastián Sánchez Sandí

2025-08-05

Preparación

Particularmente para el tema de bloques solamente se va a trabajar la primera parte, ya que los demás ejemplos involucran temas fuera del tópico básico de ANOVA y requieren técnicas más complejas. Los datos para este análisis están en *burbujas.Rdata*.

```
load("burbujas.Rdata")
str(base1)

## 'data.frame':   12 obs. of  3 variables:
## $ persona  : int  1 2 3 4 5 6 1 2 3 4 ...
## $ glicerina: int  60 60 60 60 60 60 120 120 120 120 ...
## $ tiempo   : num  7.25 7.15 4.66 5.46 12.85 ...
```

Cambiamos los tipos de glicerina y persona para poder utilizar las herramientas de R.

```
base1$glicerina = factor(base1$glicerina)
base1$persona = factor(base1$persona)
```

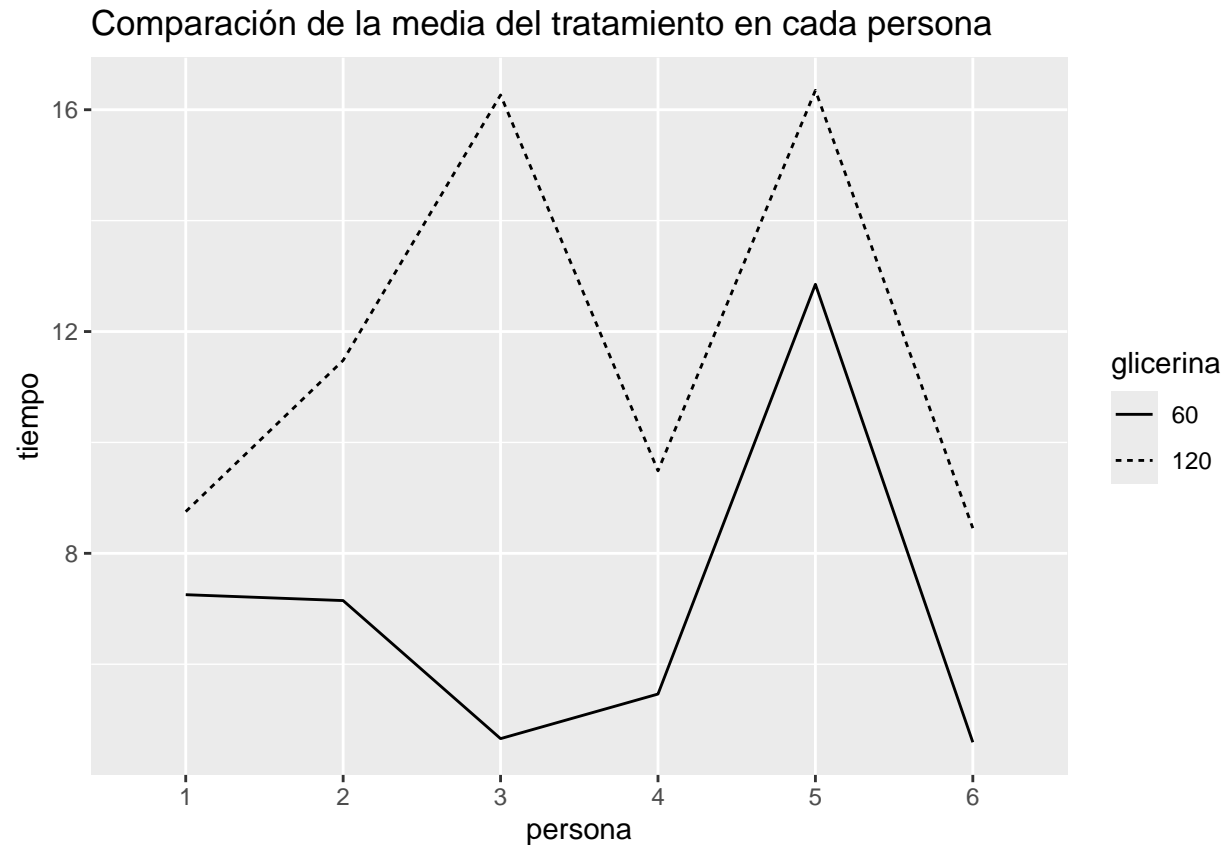
En este caso las personas son los bloques.

Visualización de Datos de Glicerina

Comenzamos con una visualización para observar el comportamiento del tratamiento en cada persona.

```
library(ggplot2)

ggplot(data = base1, aes(x = persona, y = tiempo, group = glicerina)) +
  stat_summary(fun = "mean", geom = "line", aes(linetype = glicerina)) +
  labs(title = "Comparación de la media del tratamiento en cada persona")
```



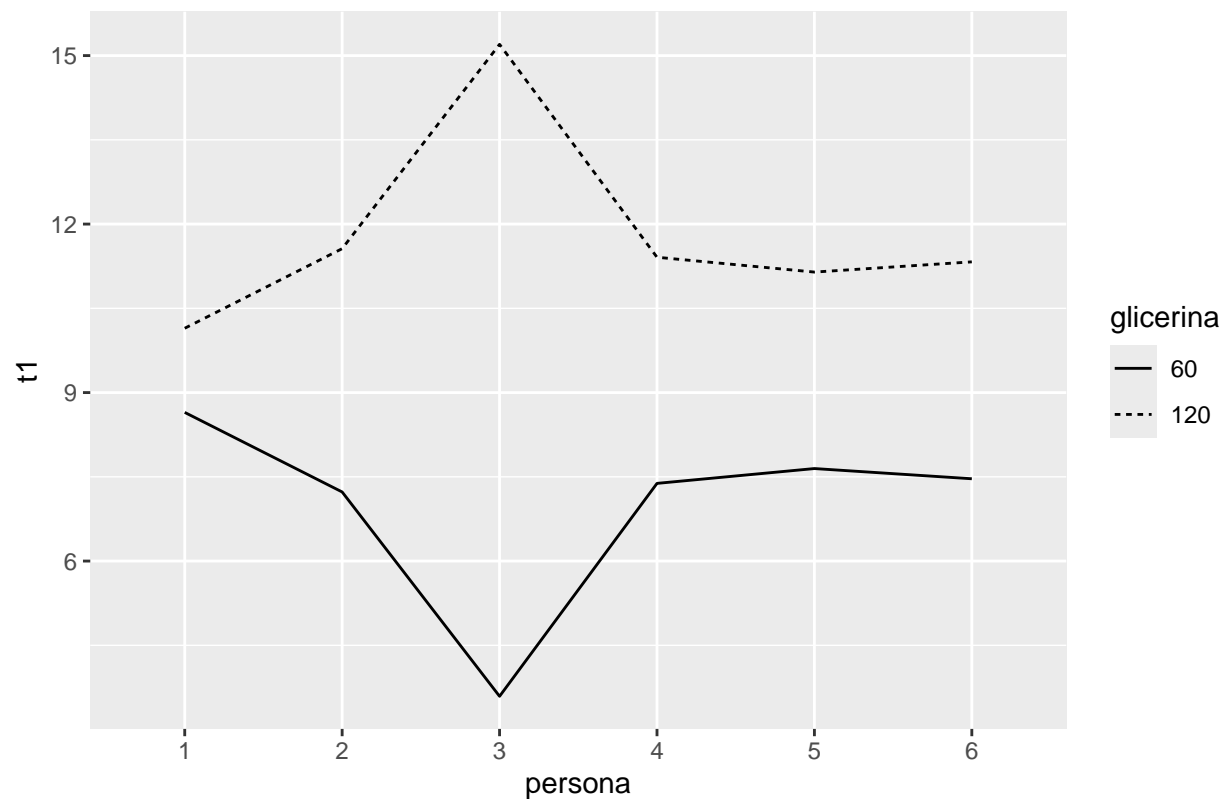
Podemos ver que parece que el nivel 120 de glicerina es mayor al de 60 en promedio en todas las personas. Sin embargo, parece que hay valores más altos y más bajos dependiendo de la persona, lo que puede sesgar el estudio. Debemos remover esta variabilidad extra que añaden las personas. Este proceso de remover la variabilidad extra que añaden las personas lo denominamos centrar los datos.

```
mod1 = lm(tiempo~persona, data = base1)
pre = predict(mod1)
t1 = base1$tiempo - pre + mean(base1$tiempo)
```

Ahora podemos construir el gráfico de líneas y de cajas con los datos centrados.

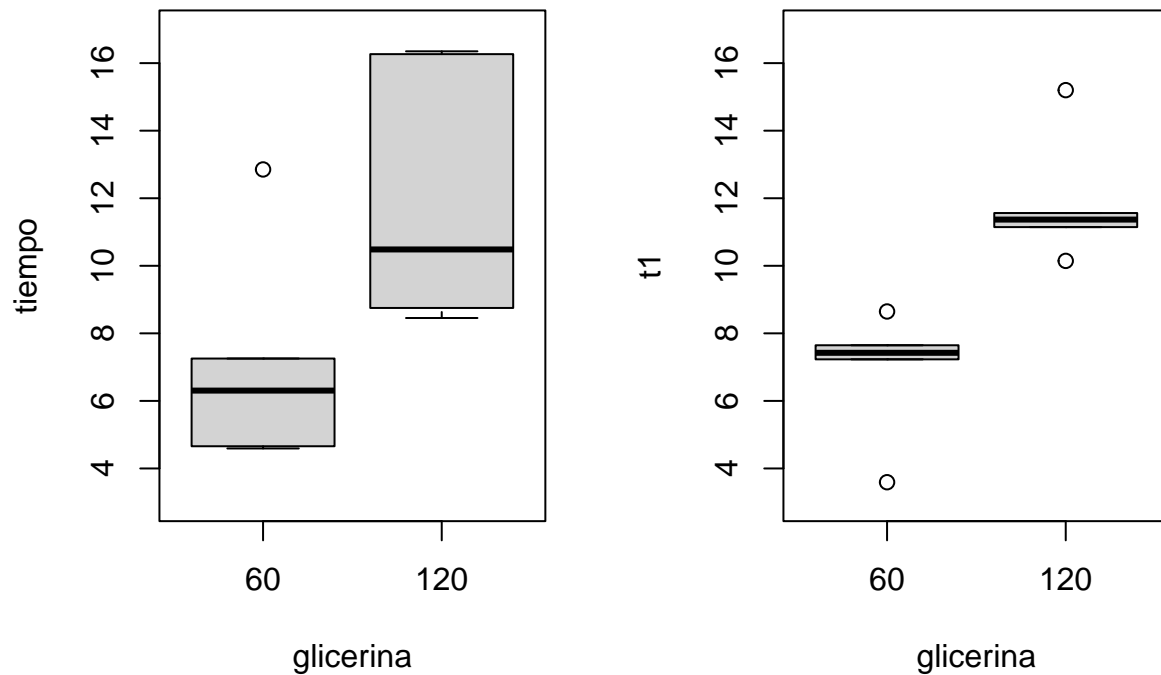
```
ggplot(data = base1, aes(x = persona, y = t1, group = glicerina)) +
  stat_summary(fun = "mean", geom = "line", aes(linetype = glicerina)) +
  labs(title = "Comparación de la media centrada del tratamiento en cada persona")
```

Comparación de la media centrada del tratamiento en cada persona



Ahora es más evidente la diferencia entre el tratamiento 60 y 120 de glicerina. Construimos los gráficos de cajas para visualizar como disminuye la variabilidad al utilizar bloques.

```
par(mfrow = c(1, 2))
boxplot(tiempo~glicerina, ylim = c(3, 17), data = base1)
boxplot(t1~glicerina, ylim = c(3, 17), data = base1)
```



Se puede apreciar que la variabilidad dentro de cada tratamiento se reduce drásticamente.

Varianza Residual

Es muy importante destacar que para realizar los gráficos se deben utilizar los datos centrados, pero para realizar la prueba de análisis de varianza se deben utilizar los datos no centrados e incluir el bloque como un factor.

```
mod2 = lm(tiempo~glicerina+persona, data = base1)
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: tiempo
##           Df Sum Sq Mean Sq F value Pr(>F)
## glicerina  1  69.235   69.235  11.4165 0.0197 *
## persona    5  84.240   16.848   2.7781 0.1432
## Residuals  5  30.323    6.065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como la probabilidad de cometer error tipo 1 es menor a 0.05, entonces rechazamos la hipótesis. Se espera que existan diferencias entre el tiempo promedio cuando se usa glicerina 120 y cuando se usa glicerina 60.

Homocedasticidad

Por último, pero no menos importante, cuando se verifica el supuesto de homocedasticidad este se debe verificar con los residuales del modelo.

```
bartlett.test(mod2$residuals~base1$glicerina)
```

```
##  
## Bartlett test of homogeneity of variances  
##  
## data: mod2$residuals by base1$glicerina  
## Bartlett's K-squared = 1.6149e-15, df = 1, p-value = 1
```

Como la probabilidad de cometer error tipo 1 es mayor a 0.05, entonces no se rechaza la hipótesis nula. Se espera que se cumpla el supuesto de homocedasticidad.