

Modelación Estadística

Actividad 5: comparaciones múltiples en ANOVA

Dr. Edlin Guerra

Comparaciones múltiples luego del Análisis de Varianza (ANOVA)

Continuamos con el caso de la planta procesadora de celulosa. Ya hicimos la primera evaluación estadística a los valores de DBO, y conseguimos evidencias estadísticas para indicar que la DBO cambia a lo largo del río. Ahora se requiere evaluar el patrón de diferencia, ya que solo algunas combinaciones de diferencias pueden o no reflejar impacto ambiental. Es importante destacar que el resultado del ANOVA que realizamos no implica automáticamente que podemos hablar de un impacto ambiental, se debe detallar más en los resultados para poder señalar esto.

1. Identifica todas las posibles hipótesis alternativas y qué significaría cada una en el contexto del estudio.
2. En caso no tengas desplegado las gráficas y resultados del ANOVA, repite el ANOVA, tablas y gráficos.
3. Para aplicar comparaciones pareadas entre todos los pares de medias de un factor, se puede usar la función `pairwise.t.test`. Esta función toma en cuenta un valor de error común (“pooled”), y realiza todas las comparaciones posibles (como si no hubiera aplicado un ANOVA). Conviene utilizar un método de ajuste a alfa por el alto número de pruebas aplicadas (“family wise error rate”). Resuelto esto, responde:
 - a) ¿Cuántas pruebas de t habría que aplicar?
 - b) ¿Qué representan los números de esta matriz triangular?
4. Para aplicar el procedimiento Tukey’s HSD (Honest Smallest Difference) es necesario que el objeto sea el resultado de una función `aov` (una función distinta de aplicar un ANOVA e R). Toma los mismos datos de DBO analizados, y aplica la función `aov` en sustitución de la función `lm` usada antes. Al objeto resultante aplícale la función `TukeyHSD`.
 - a) ¿Qué representan los datos de la columna con el nombre `diff`? Usa la función ‘`aggregate`’ recién aprendida para ayudarte en los cálculos.
 - b) ¿Qué crees que sean los que están bajo `lwr` y `upr`?

- c) Aplica la función `plot` al objeto que resultó de aplicar la de Tukey. Estudia e interpreta el gráfico que produce.
5. Para aplicar el procedimiento SNK de la librería **GAD** es necesario establecer cuales son factores fijos y cuales aleatorios. Primero tienes que instalar el paquete **GAD**, y luego llamarlo para hacerla disponible en esta sesión de R. Luego tienes que volver explícito que el factor dietas es un factor fijo. Ajustas el modelo, y después aplicas el procedimiento. Identifica estos pasos con las líneas de código a continuación e interpreta la salida.

```
library(GAD)
datos$localidades <- as.fixed(datos$localidades)
mod.lm<-lm(datos$DBO~datos$localidades)
snk.test(mod.lm, term="datos$localidades")
```

7. Compara los resultados de los 3 métodos usados.
- a) ¿Cuál produce un mayor número de resultados significativos?
- b) ¿Cuál involucra menor número de pruebas?
- c) ¿Cuál prefieres en este caso? ¿Por qué?