Conceptos Clave: ANOVA Completo

Validación de Supuestos, Post-hoc y Alternativas

Q2003B Diseño de Experimentos — Sesión 3 — Febrero 2026

1 De ANOVA básico a ANOVA completo

En la Sesión 2 aprendimos que ANOVA nos dice si hay diferencia significativa entre grupos. Sin embargo, tiene dos limitaciones importantes:

1. No nos dice **cuáles** grupos difieren entre sí.
2. Asume que los datos cumplen ciertos **supuestos**. Si no los cumplen, los resultados pueden no ser confiables.

En esta sesión resolvemos ambos problemas.

2 Los supuestos de ANOVA

ANOVA es una prueba **paramétrica**: funciona bajo ciertas condiciones. Antes de confiar en sus resultados, debemos verificar que esas condiciones se cumplan.

2.1 Supuesto 1: Normalidad

Definición: Los datos de *cada grupo* deben seguir una distribución aproximadamente normal (forma de campana).
Analogía: Es como calibrar un instrumento de medición antes de usarlo. Si el instrumento está descalibrado (datos no normales), las mediciones pueden ser engañosas.
Prueba: Shapiro-Wilk (se aplica a cada grupo por separado).

Shapiro-Wilk

- **Hipótesis nula (H_0):** Los datos siguen una distribución normal.
- **Hipótesis alternativa (H_1):** Los datos NO siguen una distribución normal.
- **Estadístico:** W (entre 0 y 1). Valores cercanos a 1 indican normalidad.
- **Decisión:**
  - Si $p \geq 0.05$: no rechazamos H_0. Los datos son compatibles con normalidad.
  - Si $p < 0.05$: rechazamos H_0. Hay evidencia de que los datos no son normales.
- **Importante:** Se aplica a *cada grupo* por separado. Si al menos un grupo falla, el supuesto global no se cumple.

2.2 Supuesto 2: Homocedasticidad

Definición: Las varianzas de todos los grupos deben ser aproximadamente iguales. Es decir, la dispersión de los datos debe ser similar en cada grupo.
Analogía: Si un grupo tiene datos muy concentrados (varianza pequeña) y otro tiene datos muy dispersos (varianza grande), compararlos directamente sería injusto; es como comparar una báscula de precisión con una de baño.
Prueba: Levene.

Prueba de Levene

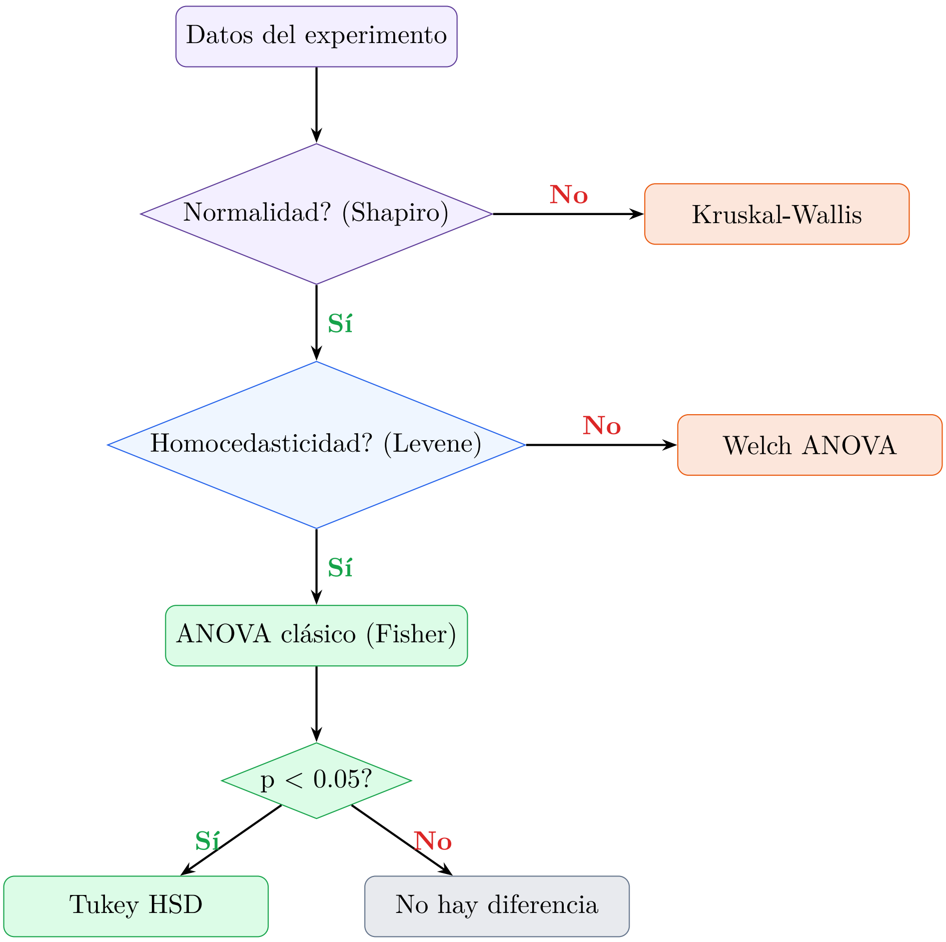
- **Hipótesis nula (H_0):** Las varianzas de los grupos son iguales.
- **Hipótesis alternativa (H_1):** Al menos un grupo tiene varianza diferente.
- **Decisión:**
  - Si $p \geq 0.05$: varianzas iguales. Se puede usar ANOVA clásico.
  - Si $p < 0.05$: varianzas desiguales. Considerar Welch ANOVA.

2.3 Supuesto 3: Independencia

Las observaciones deben ser independientes entre sí. Esto no se prueba con un test estadístico; depende de cómo se diseñó el experimento. Si cada medición se tomó de manera independiente (diferentes muestras, diferentes días, sin influencia cruzada), el supuesto se cumple.

3 Algoritmo de decisión

Cuando tenemos los resultados de Shapiro-Wilk y Levene, seguimos este flujo para decidir qué prueba usar:



4 Prueba post-hoc: Tukey HSD

Problema: ANOVA dice "hay diferencia entre al menos un par de grupos", pero no dice cuáles. **Solución:** Tukey HSD (Honestly Significant Difference) compara *todos los pares posibles* de grupos y determina cuáles diferencias son estadísticamente significativas.

Por qué no hacer múltiples pruebas t?

Si tenemos 3 grupos (A, B, C), habría que hacer 3 comparaciones: A vs B, A vs C, B vs C. Si usamos $\alpha = 0.05$ en cada una, la probabilidad de cometer al menos un error Tipo I (falso positivo) crece:

$$P(\text{al menos un error}) = 1 - (1 - 0.05)^3 = 1 - 0.857 = 0.143 \approx 14.3\%$$

Eso es casi el triple del 5% que queríamos. Tukey HSD **controla el error global** para que se mantenga en 5%.

Cómo leer la tabla de Tukey

Grupo 1	Grupo 2	Diferencia	p_{adj}	Interpretación
Control	Orgánico	+1.650	0.0001	Significativo ($p < 0.05$)
Control	Químico	+1.900	0.0000	Significativo ($p < 0.05$)
Orgánico	Químico	+0.250	0.5234	No significativo ($p \geq 0.05$)

- **Diferencia:** Cuánto difieren las medias de los dos grupos.
- **p_{adj}:** Valor p *ajustado* para comparaciones múltiples. Si $p_{adj} < 0.05$, la diferencia es significativa.
- En este ejemplo: ambos fertilizantes (Orgánico y Químico) son significativamente mejores que el Control, pero entre ellos *no hay diferencia significativa*.

5 Alternativa no paramétrica: Kruskal-Wallis

Cuándo usarla: Cuando el supuesto de normalidad falla (Shapiro-Wilk da $p < 0.05$ en al menos un grupo).
Qué hace: Compara las *medianas* (en lugar de las medias) usando rangos. No asume ninguna distribución particular.
Estadístico: H (en lugar de F).
Post-hoc: Si Kruskal-Wallis da significativo, se usa la prueba de Dunn (en lugar de Tukey).

6 Resumen: tabla de referencia rápida

Prueba	H_0	Cuándo usar	Interpretación
Shapiro-Wilk	Datos son normales	Antes de ANOVA, por grupo	$p \geq 0.05$: normal
Levene	Varianzas iguales	Antes de ANOVA, todos los grupos	$p \geq 0.05$: varianzas iguales
ANOVA (Fisher)	Medias iguales	Normalidad OK, varianzas OK	$p < 0.05$: al menos un grupo difiere
Kruskal-Wallis	Distribuciones iguales	Normalidad falla	$p < 0.05$: al menos un grupo difiere
Tukey HSD	Par de medias iguales	Después de ANOVA significativo	$p_{adj} < 0.05$: ese par difiere

7 Glosario

ANOVA
Análisis de varianza. Compara medias de 3 o más grupos.

Valor p
Probabilidad de observar datos tan extremos como los obtenidos si H_0 fuera cierta. Si $p < \alpha$, rechazamos H_0.

Error Tipo I
Rechazar H_0 cuando en realidad es verdadera (falso positivo).

Normalidad
Los datos siguen una distribución en forma de campana.

Homocedasticidad
Las varianzas de los grupos son iguales.

Paramétrico
Prueba que asume una distribución específica (normal). Ej: ANOVA, t-test.

No paramétrico
Prueba que no asume distribución. Ej: Kruskal-Wallis, Mann-Whitney.

Post-hoc
"Después del hecho". Prueba que se aplica después de encontrar significancia en ANOVA.