
Clase 179 — ANOVA (one-way, two-way)

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 13 + Bruce & Bruce, cap. 3 ANOVA. Duración estimada: 80 min.

Clase 179 — ANOVA (one-way, two-way)

Parte: 3 — Estadística Inferencial y Causal · Fuente: ISLP, cap. 13 + Bruce & Bruce, cap. 3 ANOVA.
Duración estimada: 80 min.

Objetivo

Que el alumno aplique ANOVA de una vía (≥ 3 grupos, una variable categórica) y ANOVA de dos vías (dos factores categóricos + interacción), entienda por qué no se hacen "t-tests todos contra todos" (inflación de α) y sepa hacer post-hoc con Tukey HSD. Reconocer los supuestos (independencia, normalidad por grupo, homogeneidad de varianzas) y cuándo usar la alternativa robusta Welch ANOVA o el no paramétrico Kruskal-Wallis (Clase 150).

Resultados de aprendizaje

Al finalizar, el estudiante podrá:

- Plantear $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ vs H_1 : al menos uno difiere y aplicar `scipy.stats.f_oneway` o `pingouin.anova`.
- Interpretar $F = MS_{\text{between}} / MS_{\text{within}}$ y su relación con la F-distribution ($F(k-1, n-k)$).
- Aplicar Welch's ANOVA (`pingouin.welch_anova`) cuando se viola la homogeneidad de varianzas (Levene rechaza).
- Hacer Tukey HSD post-hoc con `pingouin.pairwise_tukey` y leer los IC ajustados.
- Distinguir efectos principales de interacción en ANOVA two-way y graficar interaction plots con `seaborn.pointplot`.
- Reportar η^2 o ω^2 como effect size de ANOVA.

Temas

- ¿Por qué no t-tests múltiples? Si hacés 10 t-tests al $\alpha=0.05$, la probabilidad de al menos un falso positivo es $\approx 40\%$.
- Descomposición de varianza: $SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$.
- F-statistic: razón entre varianza explicada por los grupos y varianza residual.
- Supuestos: independencia, normalidad (Shapiro por grupo o residuos), homocedasticidad (Levene/Bartlett).
- Welch ANOVA — análogo a Welch's t-test para ≥ 3 grupos.
- Post-hoc: Tukey HSD (controla family-wise error rate), Bonferroni, Holm.
- Two-way ANOVA: efectos principales A, B, e interacción $A \times B$.
- Effect size: η^2 (eta-squared), ω^2 (omega-squared, menos sesgado).

Definiciones y características

- One-way ANOVA: testea si la media de una variable continua difiere entre los niveles de una variable categórica.
- Two-way ANOVA: dos variables categóricas. Permite testear 3 cosas: efecto principal de A, efecto principal de B, e interacción (¿el efecto de A depende del nivel de B?).

- F-statistic: $MS_{\text{between}} / MS_{\text{within}}$. Si los grupos tienen la misma media, ambos son estimadores de $\sigma^2 \rightarrow F \approx 1$. Si difieren, MS_{between} crece.
- SS (sum of squares) within: variabilidad residual dentro de cada grupo. $\sum (x_{ij} - \bar{x}_i)^2$.
- SS between: variabilidad de las medias grupales respecto a la gran media. $\sum n_i \cdot (\bar{x}_i - \bar{x})^2$.
- Homocedasticidad: igualdad de varianzas entre grupos. Test: Levene (robusto), Bartlett (sensible a normalidad).
- Welch ANOVA: no asume homocedasticidad. Es el default razonable moderno, igual que Welch's t-test.
- Tukey HSD (Honestly Significant Difference): compara todas las parejas controlando el family-wise error rate al α global.
- Interacción: cuando el efecto de un factor cambia según el nivel del otro. En el plot, las líneas no son paralelas.
- η^2 (eta-squared): $SS_{\text{between}} / SS_{\text{total}}$. Proporción de varianza explicada. Sesgado hacia arriba.
- ω^2 (omega-squared): corrige el sesgo de η^2 . Recomendado para reportar.

Dataset / recursos

- `seaborn.load_dataset('tips')`: total_bill por day (one-way), o por day \times time (two-way).
- `seaborn.load_dataset('penguins')`: body_mass_g por species (one-way claro).
- Librerías: `scipy.stats`, `pingouin`, `statsmodels.api` as `sm`, `statsmodels.formula.api` as `smf`.

Ejercicios

1. One-way: aplicá `scipy.stats.f_oneway(*[grupo for grupo in penguins.groupby('species').body_mass_g])`. Reportá F, p y `dof_between`, `dof_within`.
2. Supuestos: testá normalidad por grupo (`pingouin.normality(penguins, dv='body_mass_g', group='species')`) y homocedasticidad (`pingouin.homoscedasticity`). Si Levene rechaza, repetí con `pingouin.welch_anova`.
3. Post-hoc Tukey: `pingouin.pairwise_tukey(data=penguins, dv='body_mass_g', between='species')`. Identificá qué pares de especies difieren significativamente.
4. Two-way con interacción: `pingouin.anova(data=tips, dv='total_bill', between=['day', 'time'])`. Mirá las tres filas de la tabla (day, time, day \times time).
5. Interaction plot: `sns.pointplot(data=tips, x='day', y='total_bill', hue='time')`. Si las líneas se cruzan o no son paralelas \rightarrow hay interacción visual; cruzala con el p-value del término `day*time`.

Homework verificable

Sobre penguins (filtrando filas con NA):

1. ANOVA one-way de `flipper_length_mm` por species.
2. Chequear Levene y Shapiro; decidir entre ANOVA clásico y Welch ANOVA, justificando.
3. Tukey HSD post-hoc.
4. Reportar ω^2 (`pingouin.anova(... effsize='n2'`, y calcular ω^2 manualmente).
5. Conclusión en 3 líneas: qué pares difieren, magnitud del efecto general (η^2/ω^2), si hay alguna comparación dudosa.

Criterio de aceptación: el ANOVA debe rechazar H ($p < 0.001$), Tukey debe mostrar las tres parejas significativas, y η^2 debe ser > 0.5 (efecto grande — las tres especies tienen aletas claramente distintas).

Errores comunes

Síntoma / mensaje	Causa y cómo arreglar
Hago <code>ttest_ind</code> entre cada par de 4 grupos	Con 4 grupos son 6 comparaciones; α efecti
ANOVA da $p < 0.05$ pero ningún Tukey lo da	Puede pasar con varianzas muy distintas. F
Aplico ANOVA con varianzas muy distintas (El F-test clásico es sensible a esto, espe
Los residuos no son normales y reporto ANO	Si los grupos son grandes ($n \geq 30$ c/u), TC
Interpreto "efecto de A" sin mirar la inte	Si hay interacción, el efecto principal de

Preguntas frecuentes

¿ANOVA o regresión lineal con dummies?

Son matemáticamente equivalentes. ANOVA es la presentación tradicional; OLS con dummies (`statsmodels.formula.api.ols('y ~ C(species)', data).fit()`) da los mismos F y p. La regresión también te da los coeficientes de cada nivel respecto a la categoría de referencia, lo cual es más informativo.

¿Tukey o Bonferroni post-hoc?

Tukey es mejor para todas-las-parejas porque controla el family-wise error rate de manera específica para ANOVA (es uniformemente más poderoso que Bonferroni en ese caso). Bonferroni es más conservador (peor poder). Ver Clase 151 para alternativas modernas (Holm, BH).

¿Qué pasa si los grupos están desbalanceados (n distinto)?

ANOVA aguanta moderadamente, pero con varianzas distintas el F-test se rompe. Welch ANOVA lo maneja bien.

¿Two-way ANOVA cuando alguna celda tiene $n=0$?

Modelo no balanceado: usar Type III SS (`statsmodels` lo soporta vía `sm.stats.anova_lm(model, typ=3)`). Type I (default) depende del orden de los factores en la fórmula.

¿Y si tengo medidas repetidas (mismo sujeto en cada nivel)?

`pingouin.rm_anova` (repeated measures ANOVA). Modela la correlación intra-sujeto, lo cual ANOVA clásico ignora.

Referencias

- ISLP, cap. 13 — sección sobre ANOVA y multiple testing.
- Bruce & Bruce, cap. 3 — sección ANOVA.
- Welch, B.L. (1951), On the comparison of several mean values: an alternative approach, *Biometrika*.
- `scipy.stats.f_oneway`, `pingouin.welch_anova`, `pingouin.pairwise_tukey`.
- `statsmodels` — ANOVA (para Type II/III SS y modelos mixtos).

Siguiente clase

Clase 180 — Tests no paramétricos: Mann-Whitney, Wilcoxon, Kruskal-Wallis

Apéndice: notebook (primer bloque)

Primera celda ejecutable del notebook de la clase.

```
# Imports y configuración inicial
```

Archivos complementarios

- notebook.ipynb