Trabajo práctico 12: ANOVA y regresión

Apellido y nombre: ..... Comisión: ..... Fecha: .....

AL TERMINAR ESTA GUÍA

- Comparar más de dos medias con el ANOVA de un factor: armar y completar la tabla ANOVA, y leer la prueba F.
- Identificar qué grupos difieren con una comparación post hoc, y entender por qué no se hacen muchas pruebas t.
- Ajustar una recta de mínimos cuadrados, interpretar sus coeficientes y probar si la pendiente es distinta de cero.
- Revisar los supuestos del ANOVA y de la regresión, y reconocer en los residuos un modelo que no ajusta.

EJERCICIO 1 Los talleres de preparación para el ingreso

A mano

Una facultad ofreció dos talleres de preparación para el examen de ingreso, uno presencial y otro virtual. Se eligieron al azar 6 aspirantes que hicieron el taller presencial, 6 que hicieron el virtual y 6 que no hicieron ninguno, y se registró su nota en el examen (de 0 a 100).

Taller presencial	69	71	65	80	64	86

Taller virtual	80	73	65	56	70	66

Sin taller	52	59	57	61	68	53

Se puede suponer que las notas tienen distribución aproximadamente normal en cada grupo.
Se usa $\alpha = 0{,}05$.

a) Planteá las hipótesis para responder si la nota media difiere según la preparación.

b) Calculá la media y el desvío de cada grupo y la media general. ¿Es razonable suponer varianzas iguales?

c) Calculá la suma de cuadrados entre grupos, la suma de cuadrados dentro de los grupos y la total.

d) Armá la tabla ANOVA y tomá la decisión con la tabla de la F. Acotá el p-valor.

e) ¿Por qué no comparar los grupos de a pares con tres pruebas t al 5 %?

f) Aplicá la prueba de Tukey para saber qué grupos difieren. El valor de la tabla del rango estudentizado para 3 grupos y 15 grados de libertad es q = 3,67.

EJERCICIO 2 Completar una tabla ANOVA

A mano

Para comparar el tiempo de espera (en minutos) hasta ser atendido en 4 centros de atención primaria de una ciudad, se registró la espera de 10 pacientes elegidos al azar en cada centro. Al imprimir la salida del software se perdieron varias celdas de la tabla ANOVA:

Fuente	gl	SC	CM	F	p-valor
Entre centros	…	2.705,8	…	…	…
Dentro de los centros	…	…	…		
Total	…	7.120,4			

a) ¿Cuántos grupos y cuántas observaciones hay? Completá los grados de libertad.

b) Completá la suma de cuadrados dentro, los cuadrados medios y el estadístico F.

c) Tomá la decisión con $\alpha = 0{,}05$ y con $\alpha = 0{,}01$, y acotá el p-valor.

d) Calculá la proporción de la variabilidad total explicada por el centro (η^2). ¿Cómo la interpretás?

e) Un directivo lee la tabla y concluye: «los cuatro centros tienen tiempos de espera distintos». ¿Es correcto?

f) ¿Qué supuestos necesita el ANOVA? ¿Cómo los revisarías con los datos completos?

EJERCICIO 3 La superficie de la vivienda y el consumo eléctrico

A mano

Una distribuidora eléctrica eligió al azar 10 viviendas de una ciudad y registró la superficie cubierta (en m^2) y el consumo de electricidad del último bimestre (en kWh). Se puede usar el modo estadístico de la calculadora para las sumas.

Superficie, viviendas 1 a 5	63	48	140	147	132
Consumo, viviendas 1 a 5	341	233	399	427	364
Superficie, viviendas 6 a 10	150	100	103	105	65
Consumo, viviendas 6 a 10	518	350	338	372	329

a) Hacé el diagrama de dispersión, con la superficie en el eje horizontal. ¿Qué forma tiene la relación?

b) Calculá la recta de mínimos cuadrados.

c) Interpretá la pendiente y la ordenada al origen en el contexto del problema.

d) Calculá el coeficiente de determinación y el desvío de los residuos.

e) Probá si la pendiente poblacional es distinta de cero, con $\alpha = 0{,}05$, y acotá el p-valor.

f) Construí el intervalo de confianza del 95 % para la pendiente.

g) Estimá el consumo medio de las viviendas de 100 m². ¿Usarías la recta para una vivienda de 400 m²?

EJERCICIO 4 El precio de los autos usados

A mano

En un sitio de ventas se eligieron 12 autos usados del mismo modelo y versión, en buen estado, y se registró su antigüedad (en años) y el precio pedido (en millones de pesos).

Antigüedad	1	2	3	4	6	7	8	9	12	13	14	15

Precio, autos 1 a 6	27,1	22,7	21,7	18,6	14,4	13,1

Precio, autos 7 a 12	11,3	10,3	6,8	5,4	5,0	4,7

El software ajustó la recta de mínimos cuadrados: precio = 25,477 − 1,539 · antigüedad, con R^2 = 0,955. La pendiente es muy significativa ($p < 0{,}001$).

a) Calculá el precio predicho por la recta y el residuo de cada auto.

b) Graficá los residuos contra la antigüedad. ¿Se ve algún patrón?

c) ¿Qué dice ese patrón sobre el modelo? ¿No alcanzaba con un R^2 tan alto?

d) Según la recta, ¿cuánto costaría un auto de 20 años?

e) Un compañero propone usar el logaritmo del precio. Con software, ajustá esa recta. ¿Qué cambia?

f) Además de la linealidad, ¿qué otros supuestos de la regresión se revisan con los residuos?

EJERCICIO 5 Ingreso, nivel educativo y edad en la EPH

Con software

Se usa la base «Ocupación e ingresos en Gran Catamarca», extracto de la EPH del primer trimestre de 2026 (estadistica.ar/datos/eph-catamarca-2026t1), con las personas ocupadas que declararon ingreso (ESTADO = 1 y P21 > 0). En R: aov(log(P21) ~ grupo) para el ANOVA, pairwise.t.test(..., p.adjust.method = "bonferroni") para las comparaciones de a pares y lm(log(P21) ~ CH06) para la regresión.

Se agrupa el nivel educativo (NIVEL_ED) en tres categorías: hasta secundario incompleto (1 a 3), secundario completo o superior incompleto (4 y 5) y superior completo (6). Como en las semanas anteriores, se trata a la base como una muestra aleatoria simple, sin el ponderador.

a) Calculá el desvío del ingreso y el de su logaritmo en cada grupo. ¿En qué escala es razonable el supuesto de varianzas iguales?

b) Hacé el ANOVA del logaritmo del ingreso según el nivel educativo. Informá F, los grados de libertad, el p-valor y η^2.

c) Compará los grupos de a pares con la corrección de Bonferroni. ¿Cuánto más gana, en proporción, cada grupo que el anterior?

d) Ajustá la regresión del logaritmo del ingreso sobre la edad (CH06). Interpretá la pendiente, su prueba y el R^2.

e) Calculá el residuo medio de la regresión en cada tramo de edad. ¿Qué muestra? ¿Es adecuada la recta?

EJERCICIO 6 Lecturas de un ANOVA y de una regresión

Para discutir

Cada afirmación interpreta el resultado de un ANOVA o de una regresión. Decidí si es correcta y, si no lo es, corregila.

a) «El ANOVA dio p = 0,003: los cuatro tratamientos tienen efectos distintos entre sí.»

b) «Con dos grupos, da lo mismo hacer un ANOVA o una prueba t con varianza combinada.»

c) «La pendiente de la edad es significativa ($p < 0{,}001$), así que la edad explica buena parte del ingreso.»

d) «El R^2 dio 0,02: entre las dos variables no hay ninguna relación.»

e) «La recta predice que una vivienda de 0 m^2 consume 150 kWh: el modelo está mal.»

f) «En los departamentos con más centros de salud hay más defunciones por año: los centros de salud no mejoran la salud de la población.»

g) «Los estudiantes con las peores notas del primer parcial mejoraron mucho en el segundo: la tutoría que se les dio funcionó.»

---

estadistica.ar/docentes/trabajos-practicos/semana-12 · Las lecturas de cada tema y la hoja de fórmulas están en el sitio.