

MODELOS LINEALES

ANÁLISIS DE VARIANZA

Fernando Massa; Bruno Bellagamba

4 de junio 2023



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

UESTA INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



- 1 INTRODUCCIÓN
- 2 ONE WAY ANOVA
- 3 RESTRICCIONES
- 4 REPARAMETRIZACIONES
- 5 COMPARACIONES MÚLTIPLES
- 6 PRÓXIMA CLASE



- Veremos un uso particular de los modelos lineales como generalización de la prueba de comparación de 2 medias para grupos independientes.
- Identificaremos los problemas que surgen al considerar variables cualitativas como variables explicativas.
- Aprenderemos algunas de las soluciones disponibles y consideraciones cada una en cuanto a la interpretación de los parámetros del modelo.
- Llevaremos a cabo comparaciones múltiples para responder algunas preguntas que suelen surgir al emplear estos modelos.

Hasta aquí nos enfocamos en el uso de modelos lineales en situaciones más propias de estudios *observacionales*, donde si bien asumimos que los valores de las variables explicativas (X) eran fijos, perfectamente podían haber sido considerados aleatorios.

Hasta aquí nos enfocamos en el uso de modelos lineales en situaciones más propias de estudios *observacionales*, donde si bien asumimos que los valores de las variables explicativas (X) eran fijos, perfectamente podían haber sido considerados aleatorios.

En muchas situaciones *experimentales*, un investigador aplica varios tratamientos (o combinaciones) a unidades experimentales seleccionadas al azar y luego desea comparar los promedios de cada tratamiento. En estos casos, los valores de las variables x_j son fijadas por el investigador.

Hasta aquí nos enfocamos en el uso de modelos lineales en situaciones más propias de estudios *observacionales*, donde si bien asumimos que los valores de las variables explicativas (X) eran fijos, perfectamente podían haber sido considerados aleatorios.

En muchas situaciones *experimentales*, un investigador aplica varios tratamientos (o combinaciones) a unidades experimentales seleccionadas al azar y luego desea comparar los promedios de cada tratamiento. En estos casos, los valores de las variables x_j son fijadas por el investigador.

En el análisis de varianza (ANOVA), se utilizan modelos lineales para obtener una comparación de estos promedios, sobre todo en situaciones donde se comparan **más** de 2 grupos.

El llamado *One way model* considera una situación (típica pero no necesariamente experimental) donde se quiere determinar si el promedio de la variable de respuesta Y difiere entre los grupos que surgen de considerar los niveles de la variable (cualitativa) x_1 .

El llamado *One way model* considera una situación (típica pero no necesariamente experimental) donde se quiere determinar si el promedio de la variable de respuesta Y difiere entre los grupos que surgen de considerar los niveles de la variable (cualitativa) x_1 .

Si bien podríamos considerar escribir el modelo de la siguiente manera:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

El llamado *One way model* considera una situación (típica pero no necesariamente experimental) donde se quiere determinar si el promedio de la variable de respuesta Y difiere entre los grupos que surgen de considerar los niveles de la variable (cualitativa) x_1 .

Si bien podríamos considerar escribir el modelo de la siguiente manera:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

Dado que la variable x_1 es de índole **cualitativa**, este modelo suele escribirse de la siguiente forma.

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_j \quad \forall j = 1, \dots, J$$

Lo cual cambia sustancialmente la notación, e incluso la interpretación de los parámetros.

En el modelo

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$

Decimos que:

En el modelo

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$

Decimos que:

- μ es la media general de Y .
- τ_j es el *efecto* del j -ésimo tratamiento (grupo). $j = 1, \dots, J$
- y_{ij} es el i -ésimo valor del j -ésimo tratamiento (grupo). $i = 1, \dots, n_j$
- ε_{ij} es el error correspondiente al i -ésimo valor del j -ésimo tratamiento (grupo) y (al igual que antes) es quien brinda aleatoriedad al modelo.

En el modelo

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$

Decimos que:

- μ es la media general de Y .
- τ_j es el *efecto* del j -ésimo tratamiento (grupo). $j = 1, \dots, J$
- y_{ij} es el i -ésimo valor del j -ésimo tratamiento (grupo). $i = 1, \dots, n_j$
- ε_{ij} es el error correspondiente al i -ésimo valor del j -ésimo tratamiento (grupo) y (al igual que antes) es quien brinda aleatoriedad al modelo.

No obstante, planteado de esta manera, este modelo tiene un inconveniente que dificulta la estimación de sus parámetros.

Suponga que tiene 3 datos que reciben el tratamiento (grupo 1), 5 que reciben otro tratamiento distinto (grupo 2) y 3 que reciben un tercer tratamiento (grupo 3).

Niveles del factor A	$i=1$	$Y_{11} \ Y_{21} \ Y_{31}$
	$i=2$	$Y_{12} \ Y_{22} \ Y_{32} \ Y_{42} \ Y_{52}$
	$i=3$	$Y_{13} \ Y_{23} \ Y_{33} \ Y_{43}$

FIGURA: Modelo a 1 vía

MODELO a una vía

Suponga que tiene 3 datos que reciben el tratamiento (grupo 1), 5 que reciben otro tratamiento distinto (grupo 2) y 3 que reciben un tercer tratamiento (grupo 3).

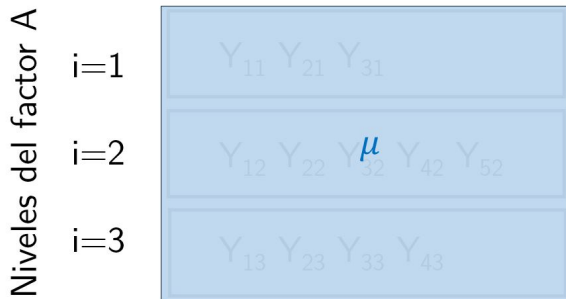


FIGURA: Efecto global

Suponga que tiene 3 datos que reciben el tratamiento (grupo 1), 5 que reciben otro tratamiento distinto (grupo 2) y 3 que reciben un tercer tratamiento (grupo 3).

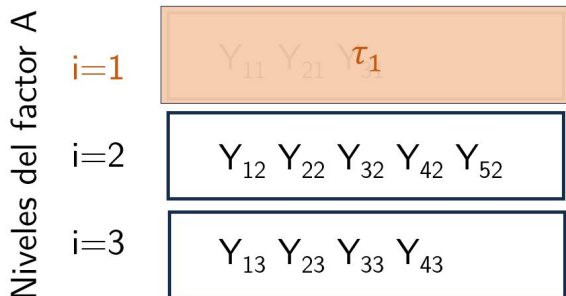


FIGURA: Efecto del tratamiento 1 (del factor A)

Suponga que tiene 3 datos que reciben el tratamiento (grupo 1), 5 que reciben otro tratamiento distinto (grupo 2) y 3 que reciben un tercer tratamiento (grupo 3).

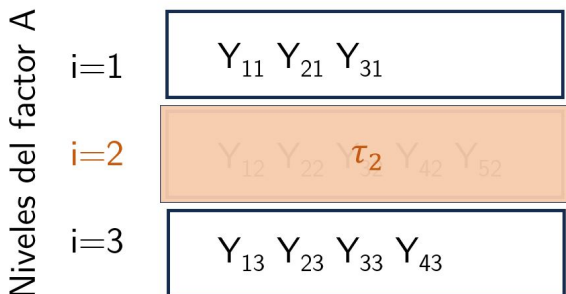


FIGURA: Efecto del tratamiento 2 (del factor A)

Suponga que tiene 3 datos que reciben el tratamiento (grupo 1), 5 que reciben otro tratamiento distinto (grupo 2) y 3 que reciben un tercer tratamiento (grupo 3).

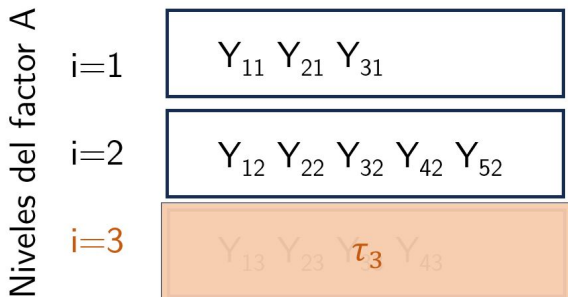


FIGURA: Efecto del tratamiento 3 (del factor A)

Para construir la matriz de diseño correspondiente a un conjunto de datos a analizar con el modelo a 1 vía, observamos que se necesita una variable *dummy* por cada tratamiento. A modo de ejemplo supongamos que se dispone de 3 tratamientos con 2 observaciones cada uno. La matriz X sería:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Siendo esta matriz conformable con el vector de parámetros $(\mu, \tau_1, \tau_2, \tau_3)$.

Para construir la matriz de diseño correspondiente a un conjunto de datos a analizar con el modelo a 1 vía, observamos que se necesita una variable *dummy* por cada tratamiento. A modo de ejemplo supongamos que se dispone de 3 tratamientos con 2 observaciones cada uno. La matriz X sería:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Siendo esta matriz conformable con el vector de parámetros $(\mu, \tau_1, \tau_2, \tau_3)$.

Sin embargo, esta matriz tiene un defecto importante...

Para construir la matriz de diseño correspondiente a un conjunto de datos a analizar con el modelo a 1 vía, observamos que se necesita una variable *dummy* por cada tratamiento. A modo de ejemplo supongamos que se dispone de 3 tratamientos con 2 observaciones cada uno. La matriz X sería:

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Siendo esta matriz conformable con el vector de parámetros $(\mu, \tau_1, \tau_2, \tau_3)$.

Sin embargo, esta matriz tiene un defecto importante... **NO** es de rango completo.

Ante el problema que surge en este modelo, existen varias alternativas que nos permiten obtener estimaciones únicas.

- En lugar de trabajar con parámetros a estimar, es posible trabajar con **funciones estimables**.
- Por otro lado, es posible **reparametrizar** el modelo, de forma de solucionar el problema del rango de X .
- La solución más común en la práctica consiste en imponer **restricciones** al vector de parámetros. Típicamente se utiliza alguna de las siguientes.


Ante el problema que surge en este modelo, existen varias alternativas que nos permiten obtener estimaciones únicas.


- En lugar de trabajar con parámetros a estimar, es posible trabajar con **funciones estimables**.
- Por otro lado, es posible **reparametrizar** el modelo, de forma de solucionar el problema del rango de X .
- La solución más común en la práctica consiste en imponer **restricciones** al vector de parámetros. Típicamente se utiliza alguna de las siguientes.

1 $\tau_1 = 0$

2 $\tau_J = 0$

3 $\sum_{j=1}^J \tau_j = 0$

Consideremos el caso donde se impone $\tau_1 = 0$. La cual es la opción que por defecto asume el .

Consideremos el caso donde se impone $\tau_1 = 0$. La cual es la opción que por defecto asume el .

De adoptar esta alternativa en nuestro ejemplo de $J = 3$, no vale la pena incluir la primera variable *dummy* por lo que la matriz de diseño pasaría a ser:

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

Obsérvese entonces, que en este caso desaparece el problema del rango de X pero se dificulta (solo un poco) la interpretación de los parámetros.

INTERPRETACIÓN DE LOS PARÁMETROS

Las primeras 2 filas de X aún corresponden a observaciones que fueron asignadas al primer tratamiento, las siguientes 2 corresponden al segundo tratamiento y las últimas 2 al tercer tratamiento. Si planteáramos el valor esperado de las observaciones de cada tratamiento obtendríamos:

$$\mathbb{E}(Y_{i1}) = \mu$$

$$\mathbb{E}(Y_{i2}) = \mu + \tau_2$$

$$\mathbb{E}(Y_{i3}) = \mu + \tau_3$$

INTERPRETACIÓN DE LOS PARÁMETROS

Las primeras 2 filas de X aún corresponden a observaciones que fueron asignadas al primer tratamiento, las siguientes 2 corresponden al segundo tratamiento y las últimas 2 al tercer tratamiento. Si planteáramos el valor esperado de las observaciones de cada tratamiento obtendríamos:

$$\mathbb{E}(Y_{i1}) = \mu$$

$$\mathbb{E}(Y_{i2}) = \mu + \tau_2$$

$$\mathbb{E}(Y_{i3}) = \mu + \tau_3$$

- μ pasa a ser el valor esperado de las observaciones del tratamiento 1, también llamado, *“tratamiento (o grupo) de referencia”*.
- Dado, que el valor esperado en el tratamiento 2 es $\mu + \tau_2$, entonces es posible decir que τ_2 es la *“diferencia esperada”* entre los promedios del tratamiento 2 con respecto al tratamiento de referencia.
- En el grupo 3, se da una situación análoga al anterior.

Como resultado general, bajo la restricción $\tau_1 = 0$, se soluciona el problema del rango de la matriz X y en cuanto a la interpretación se puede afirmar que:

Como resultado general, bajo la restricción $\tau_1 = 0$, se soluciona el problema del rango de la matriz X y en cuanto a la interpretación se puede afirmar que:

INTERPRETACIÓN

- El parámetro μ pasa a ser el promedio del grupo de referencia.
- Los parámetros $\tau_2, \tau_3, \dots, \tau_J$ corresponden a la diferencia esperada de cada tratamiento con respecto al grupo de referencia.

Como resultado general, bajo la restricción $\tau_1 = 0$, se soluciona el problema del rango de la matriz X y en cuanto a la interpretación se puede afirmar que:

INTERPRETACIÓN

- El parámetro μ pasa a ser el promedio del grupo de referencia.
- Los parámetros $\tau_2, \tau_3, \dots, \tau_J$ corresponden a la diferencia esperada de cada tratamiento con respecto al grupo de referencia.

Esta situación, se da de manera análoga ante la restricción τ_J (que es la adoptada por defecto en SAS).

De optar por esta alternativa, es importante tener en cuenta *qué* se está testeando al realizar la prueba F de significación global y al realizar las pruebas t de significación individual.

De optar por esta alternativa, es importante tener en cuenta *qué* se está testeando al realizar la prueba F de significación global y al realizar las pruebas t de significación individual.

Significación global: El estadístico F contrasta la hipótesis

$$H_0) \quad \tau_1 = \tau_2 = \dots = \tau_J = 0$$

$$H_1) \quad \tau_j \neq 0 \text{ para algún } j$$

A través de esta prueba estamos decidiendo si es razonable pensar que el promedio de Y es igual entre los grupos (H_0) o si difiere en al menos uno de ellos (H_1).

Significación individual: Estas pruebas deben ser interpretadas con cautela, porque dependen del grupo tomado como referencia.

La prueba correspondiente al parámetro asociado a la columna de unos, contrasta la hipótesis:

$$H_0) \quad \mu = 0$$

$$H_1) \quad \mu \neq 0$$

La cual corresponde a que el promedio del grupo de referencia sea cero (bajo H_0) o diferente de 0 (bajo H_1). Por lo general, esta prueba tiene poco interés.

Distinto es el caso de las $J - 1$ pruebas asociadas a los τ_j .

$$H_0) \quad \tau_j = 0$$

$$H_1) \quad \tau_j \neq 0$$

Sin embargo, debe recordarse que estas pruebas contrastan la media de cada grupo con respecto al grupo de referencia y **NO** entre ellos.

Otra de las soluciones al problema del rango de X consiste en reparametrizar el modelo.
Si partimos del modelo original.

$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$

Y nombramos $\mu_j = \mu + \tau_j$, obtenemos:


Otra de las soluciones al problema del rango de X consiste en reparametrizar el modelo.
Si partimos del modelo original.


$$y_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$


Y nombramos $\mu_j = \mu + \tau_j$, obtenemos:

$$y_{ij} = \mu_j + \varepsilon_{ij} \quad \forall i = 1, \dots, n_{ij} \quad \forall j = 1, \dots, J$$

Cuya matriz de diseño solo contiene 3 variables *dummies* pero **NO** la columna de unos.
Este modelo cuenta además con la virtud de que sus parámetros $(\mu_1, \mu_2, \dots, \mu_J)$ son la media de Y en cada tratamiento.

Al momento de llevar a cabo la inferencia en el  debemos ser cuidadosos ya que el modelo no incluye la constante, lo cual puede traer problemas computacionales (los mismos que nombramos sobre el R^2 la clase anterior.)

Al momento de llevar a cabo la inferencia en el  debemos ser cuidadosos ya que el modelo no incluye la constante, lo cual puede traer problemas computacionales (los mismos que nombramos sobre el R^2 la clase anterior.)


Signifiación global: La prueba de hipótesis que realiza el  tiene como hipótesis nula


$$\mu_1 = \mu_2 = \dots = \mu_J = 0$$

Pero la que nos interesa a nosotros es:

$$\mu_1 = \mu_2 = \dots = \mu_J$$

Para obtener el estadístico y p-valor de esta prueba debemos construir el estadístico comparando la *SCR*es de este modelo con la de otro donde **NO** existan diferencias entre grupos, es decir, un modelo que solo incluya la constante.

Al momento de llevar a cabo la inferencia en el  debemos ser cuidadosos ya que el modelo no incluye la constante, lo cual puede traer problemas computacionales (los mismos que nombramos sobre el R^2 la clase anterior.)

Signifiación global: La prueba de hipótesis que realiza el  tiene como hipótesis nula

$$\mu_1 = \mu_2 = \dots = \mu_J = 0$$

Pero la que nos interesa a nosotros es:

$$\mu_1 = \mu_2 = \dots = \mu_J$$


Para obtener el estadístico y p-valor de esta prueba debemos construir el estadístico comparando la *SCR*es de este modelo con la de otro donde **NO** existan diferencias entre grupos, es decir, un modelo que solo incluya la constante.

Signifiación individual: En este caso, las pruebas simplemente contrastan el valor promedio contra cero, lo cual, suele ser de poca utilidad.



Consideremos una situación donde se quiere comparar la vida útil de los neumáticos de 4 marcas. En los datos del archivo *anova_1_via.xlsx* se presenta el kilometraje (medido en miles de km) de 15 neumáticos de cada marca sometidos a desgaste en un laboratorio.



Vayamos al  y veamos como ajustar estos modelos considerando distintas restricciones y parametrizaciones y observemos qué resultados cambian y cuales se mantienen.

Típicamente, este tipo de análisis no sólo se centran en determinar si existe algún tratamiento mejor o peor que los demás sino que adicionalmente interesa conocer qué tratamiento es diferente a cual otro.

Típicamente, este tipo de análisis no sólo se centran en determinar si existe algún tratamiento mejor o peor que los demás sino que adicionalmente interesa conocer qué tratamiento es diferente a cual otro.

En situaciones donde el factor experimental (variable presente en X) solo tiene dos niveles, esto puede realizarse tanto con una prueba t como dentro el ámbito del modelo lineal.

Típicamente, este tipo de análisis no sólo se centran en determinar si existe algún tratamiento mejor o peor que los demás sino que adicionalmente interesa conocer qué tratamiento es diferente a cual otro.

En situaciones donde el factor experimental (variable presente en X) solo tiene dos niveles, esto puede realizarse tanto con una prueba t como dentro el ámbito del modelo lineal. Cuando el factor experimental tiene 3 o más niveles la situación cambia:

- O bien se llevan a cabo $J(J-1)/2$ pruebas t de manera de comparar todos los promedios entre sí.
- O bien se lleva a cabo la prueba F el el modelo ANOVA a una vía.

Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

Por otro lado, la prueba F mantiene el nivel *global* de la prueba, pero nos da una respuesta parcial.

Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

Por otro lado, la prueba F mantiene el nivel *global* de la prueba, pero nos da una respuesta parcial. Debe haber algo “a medio camino”.

Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

Por otro lado, la prueba F mantiene el nivel *global* de la prueba, pero nos da una respuesta parcial. Debe haber algo “a medio camino”.

Data snooping

En algunas situaciones, los investigadores disminuyen artificialmente la probabilidad de error de tipo 1, seleccionando el número de comparaciones de antemano.

Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

Por otro lado, la prueba F mantiene el nivel *global* de la prueba, pero nos da una respuesta parcial. Debe haber algo “a medio camino”.

Data snooping

En algunas situaciones, los investigadores disminuyen artificialmente la probabilidad de error de tipo 1, seleccionando el número de comparaciones de antemano.

Si esta selección se lleva a cabo **ANTES** de ver los datos, no hay problema.

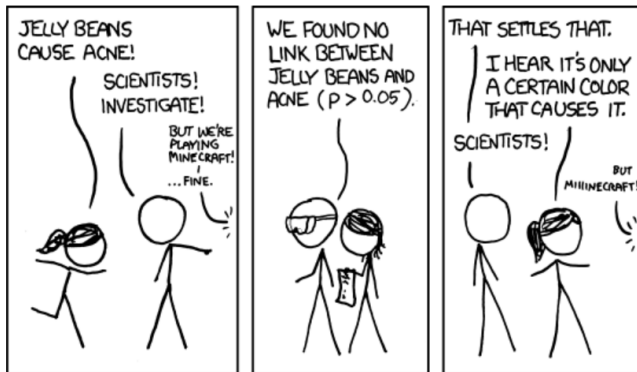
Realizar muchas pruebas t nos da la respuesta que realmente queremos conocer, pero a mayor J , mayor es la probabilidad de cometer error de tipo 1 (*false discovery rate*).

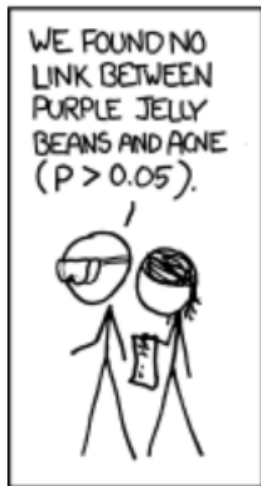
Por otro lado, la prueba F mantiene el nivel *global* de la prueba, pero nos da una respuesta parcial. Debe haber algo “a medio camino”.

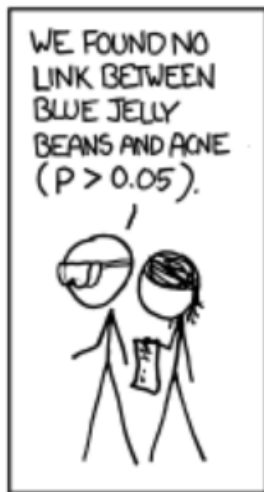
Data snooping

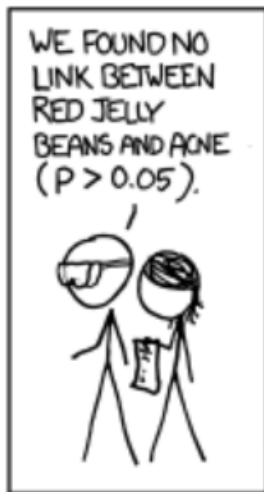
En algunas situaciones, los investigadores disminuyen artificialmente la probabilidad de error de tipo 1, seleccionando el número de comparaciones de antemano.

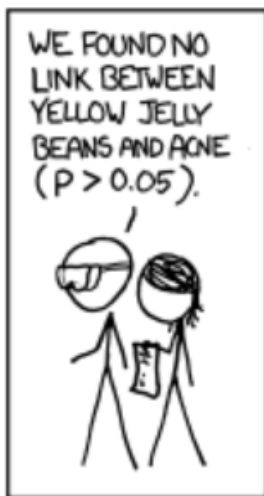
Si esta selección se lleva a cabo **ANTES** de ver los datos, no hay problema. Si seleccionamos los grupos cuyos promedios difieren más, estamos realizando solo las comparaciones que nos convienen, descartando las que creemos que no serán significativas.



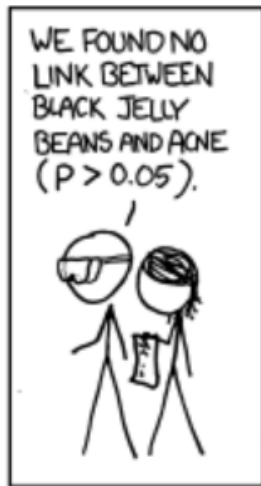


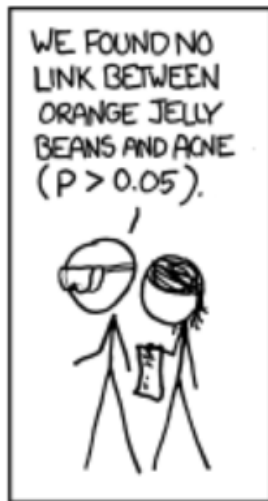


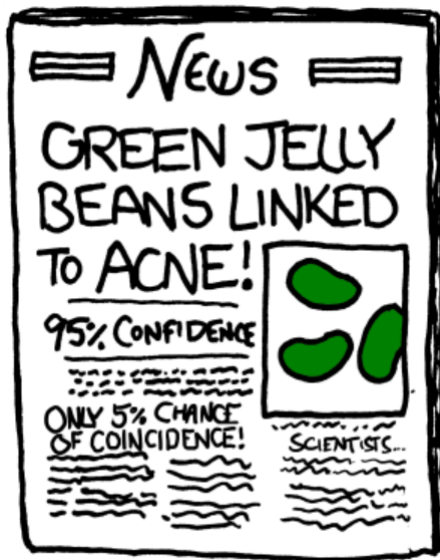












COMPARACIONES MÚLTIPLES

NIVEL GLOBAL

Al llevar a cabo un cierto número de pruebas (por ejemplo M), todas con un cierto nivel α_k , es posible cometer error de tipo I en cada una.

NIVEL GLOBAL

Al llevar a cabo un cierto número de pruebas (por ejemplo M), todas con un cierto nivel α_k , es posible cometer error de tipo I en cada una. Asumiendo que estas pruebas se llevan a cabo de forma independiente es posible obtener una expresión para el nivel global:

$$\alpha_g = 1 - (1 - \alpha_k)^M$$

NIVEL GLOBAL

Al llevar a cabo un cierto número de pruebas (por ejemplo M), todas con un cierto nivel α_k , es posible cometer error de tipo I en cada una. Asumiendo que estas pruebas se llevan a cabo de forma independiente es posible obtener una expresión para el nivel global:

$$\alpha_g = 1 - (1 - \alpha_k)^M$$

A partir de la fórmula anterior, asumiendo que siempre empleamos un nivel de 5% y recordando que para comparar J tratamientos realizaríamos $M = J(J-1)/2$ pruebas:

COMPARACIONES MÚLTIPLES

NIVEL GLOBAL

Al llevar a cabo un cierto número de pruebas (por ejemplo M), todas con un cierto nivel α_k , es posible cometer error de tipo I en cada una. Asumiendo que estas pruebas se llevan a cabo de forma independiente es posible obtener una expresión para el nivel global:

$$\alpha_g = 1 - (1 - \alpha_k)^M$$

A partir de la fórmula anterior, asumiendo que siempre empleamos un nivel de 5% y recordando que para comparar J tratamientos realizaríamos $M = J(J - 1)/2$ pruebas:

J	$J(J - 1)/2$	α_g
2	1	0,050
3	3	0,142
4	6	0,265
5	10	0,401
6	15	0,536

La solución intermedia entre las $J(J-1)/2$ pruebas t y la prueba F son las *comparaciones múltiples*.

La solución intermedia entre las $J(J-1)/2$ pruebas t y la prueba F son las *comparaciones múltiples*.

La idea consiste en 2 pasos.

- 1 Realizar las comparaciones **sólo** cuando la prueba F es significativa.
- 2 Realizar las comparaciones con algún mecanismo que controle el nivel global.

La solución intermedia entre las $J(J-1)/2$ pruebas t y la prueba F son las *comparaciones múltiples*.

La idea consiste en 2 pasos.

- 1 Realizar las comparaciones **sólo** cuando la prueba F es significativa.
- 2 Realizar las comparaciones con algún mecanismo que controle el nivel global.

Para controlar el nivel global existen varias alternativas, las más conocidas son:

- Tukey.
- Bonferroni.
- Scheffé.

La solución intermedia entre las $J(J-1)/2$ pruebas t y la prueba F son las *comparaciones múltiples*.

La idea consiste en 2 pasos.

- 1 Realizar las comparaciones **sólo** cuando la prueba F es significativa.
- 2 Realizar las comparaciones con algún mecanismo que controle el nivel global.

Para controlar el nivel global existen varias alternativas, las más conocidas son:

- Tukey.
- Bonferroni.
- Scheffé.

Pero hay varias más...

Los procedimientos que se presentan a continuación parten de un marco común donde se busca rechazar H_{0j} cuando.

$$|t_0| = \frac{\text{estimación}}{\text{desvío de la estimación}} \geq \text{valor crítico}$$

Los procedimientos que se presentan a continuación parten de un marco común donde se busca rechazar H_{0j} cuando.

$$|t_0| = \frac{\text{estimación}}{\text{desvío de la estimación}} \geq \text{valor crítico}$$

Donde las estimaciones y sus desvíos se construyen a partir de $\hat{\beta}$ y $\hat{\sigma}^2 (X'X)^{-1}$.

Los procedimientos que se presentan a continuación parten de un marco común donde se busca rechazar H_{0j} cuando.

$$|t_0| = \frac{\text{estimación}}{\text{desvío de la estimación}} \geq \text{valor crítico}$$

Donde las estimaciones y sus desvíos se construyen a partir de $\hat{\beta}$ y $\hat{\sigma}^2 (X'X)^{-1}$.

Los métodos difieren en sus supuestos y proveen de diferentes valores críticos o diferentes formas de ajustar los p-valores.

Para llevar a cabo la comparación entre los promedios de los grupos A y B el estadístico a emplear sería:

$$t = \frac{\hat{\mu}_A - \hat{\mu}_B}{s_{\hat{\mu}_A - \hat{\mu}_B}} = \frac{\hat{\mu}_A - \hat{\mu}_B}{\hat{s}(1/n_A + 1/n_B)}$$

Para llevar a cabo la comparación entre los promedios de los grupos A y B el estadístico a emplear sería:

$$t = \frac{\hat{\mu}_A - \hat{\mu}_B}{s_{\hat{\mu}_A - \hat{\mu}_B}} = \frac{\hat{\mu}_A - \hat{\mu}_B}{\hat{s}(1/n_A + 1/n_B)}$$

Si bien, típicamente llevaríamos a cabo esta prueba empleando la distribución t , la propuesta de Tukey consiste en emplear como referencia la distribución del *rango studentizado*.

DISTRIBUCIÓN DEL RANGO STUDENTIZADO

Sean Y_1, Y_2, \dots, Y_M V.A. $N(\mu, \sigma)$ y sea $R = \max Y_i - \min Y_i$, entonces:

$$q = \frac{R}{\hat{\sigma}}$$

es el estadístico del rango studentizado de la muestra.

En nuestro caso, el rango que se *studentiza* no es el de las observaciones sino el de los promedios que se están comparando. Por ende, la distribución de referencia es $q(J, n - J)$

En nuestro caso, el rango que se *studentiza* no es el de las observaciones sino el de los promedios que se están comparando. Por ende, la distribución de referencia es $q(J, n - J)$

Siendo esta la distribución del rango studentizado para:

- una muestra de n V.A. normales pertenecientes a J grupos
- con una media y varianza común a todas
- donde se construye un estimador de σ^2 con $n - J$ grados de libertad.

En nuestro caso, el rango que se *studentiza* no es el de las observaciones sino el de los promedios que se están comparando. Por ende, la distribución de referencia es $q(J, n - J)$

Siendo esta la distribución del rango studentizado para:

- una muestra de n V.A. normales pertenecientes a J grupos
- con una media y varianza común a todas
- donde se construye un estimador de σ^2 con $n - J$ grados de libertad.

Esta distribución no tiene forma cerrada, pero es posible obtener valores críticos $q_\alpha(J, n - J)$ y p-valores mediante métodos computacionales o basados de simulación

En nuestro caso, el rango que se *studentiza* no es el de las observaciones sino el de los promedios que se están comparando. Por ende, la distribución de referencia es $q(J, n - J)$

Siendo esta la distribución del rango studentizado para:

- una muestra de n V.A. normales pertenecientes a J grupos
- con una media y varianza común a todas
- donde se construye un estimador de σ^2 con $n - J$ grados de libertad.

Esta distribución no tiene forma cerrada, pero es posible obtener valores críticos $q_\alpha(J, n - J)$ y p-valores mediante métodos computacionales o basados de simulación (o empleando algún software creado con este propósito).

Para lograr significancia a un nivel global α_g , la diferencia promedio entre un par de tratamientos A y B debe ser al menos:

$$\frac{q_{\alpha}(J, n - J)}{\sqrt{2}} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}$$

A este valor comúnmente se lo llama *Tukey's Honest Significant Difference* (*Tukey's HSD*).

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

$$\alpha_g = \mathbb{P}(\text{rechazar al menos 1 } H_{0j} \text{ cuando son todas ciertas})$$

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

$$\begin{aligned}\alpha_g &= \mathbb{P}(\text{rechazar al menos 1 } H_{0j} \text{ cuando son todas ciertas}) \\ &= \mathbb{P}(E_1 \cup E_1 \cup \dots \cup E_M)\end{aligned}$$

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

$$\begin{aligned}\alpha_g &= \mathbb{P}(\text{rechazar al menos 1 } H_{0j} \text{ cuando son todas ciertas}) \\ &= \mathbb{P}(E_1 \cup E_1 \cup \dots \cup E_M) \\ &\leq \sum_k \mathbb{P}(E_j) = \sum_k \alpha = M\alpha\end{aligned}$$

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

$$\begin{aligned}\alpha_g &= \mathbb{P}(\text{rechazar al menos 1 } H_{0j} \text{ cuando son todas ciertas}) \\ &= \mathbb{P}(E_1 \cup E_1 \cup \dots \cup E_M) \\ &\leq \sum_k \mathbb{P}(E_j) = \sum_k \alpha = M\alpha\end{aligned}$$

Este procedimiento es de índole conservador ya que propone controlar el nivel global, disminuyendo el nivel de cada prueba de manera proporcional al número de pruebas a realizar.

Este método parte de la base de que se quieren realizar M pruebas (no necesariamente todas las comparaciones 2 a 4) con un nivel global α_g . Sea E_j el evento correspondiente a rechazar la hipótesis nula en la j -ésima prueba (H_{0j}) siendo cierta, entonces:

$$\begin{aligned}\alpha_g &= \mathbb{P}(\text{rechazar al menos 1 } H_{0j} \text{ cuando son todas ciertas}) \\ &= \mathbb{P}(E_1 \cup E_1 \cup \dots \cup E_M) \\ &\leq \sum_k \mathbb{P}(E_j) = \sum_k \alpha = M\alpha\end{aligned}$$

Este procedimiento es de índole conservador ya que propone controlar el nivel global, disminuyendo el nivel de cada prueba de manera proporcional al número de pruebas a realizar.

Así, para lograr un nivel global de α_g , Bonferroni propone emplear un nivel de $\frac{\alpha_g}{M}$ en cada comparación.

El análogo de Bonferroni al *HSD* de Tukey se obtiene cambiando el valor crítico de la distribución del rango estudentizado por un valor crítico obtenido a partir de la distribución *t* de Student pero empleando el nivel $\frac{\alpha_g}{M}$.

$$t_{1-\frac{\alpha}{2M}, n-J} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Y para no ser menos que Tukey, a este valor podríamos llamarlo *Bonferroni's Honest Significant Difference* (*Bonferroni's HSD*).

A diferencia de los 2 métodos anteriores, este funciona para casos más generales, llamados contrastes.

A diferencia de los 2 métodos anteriores, este funciona para casos más generales, llamados contrastes.

Sean $\mu_1, \mu_2, \dots, \mu_J$ los promedios de Y en los J grupos bajo estudio. Denominamos contraste a:

$$C = \sum_{j=1}^J w_j \mu_j \quad \sum_{j=1}^J w_j = 0$$

A diferencia de los 2 métodos anteriores, este funciona para casos más generales, llamados contrastes.

Sean $\mu_1, \mu_2, \dots, \mu_J$ los promedios de Y en los J grupos bajo estudio. Denominamos contraste a:

$$C = \sum_{j=1}^J w_j \mu_j \quad \sum_{j=1}^J w_j = 0$$

El cual es estimado a través de su equivalente muestral (obtenido a través de las estimaciones del modelo lineal).

$$\hat{C} = \sum_{j=1}^J w_j \hat{\mu}_j$$

Cuyo desvío es:

$$s_{\hat{C}} = \hat{\sigma}^2 \sum_{j=1}^J \frac{w_j^2}{n_j}$$

Para poner a prueba la hipótesis nula $C = 0$ empleamos el mismo estadístico t , pero ...

Para poner a prueba la hipótesis nula $C = 0$ empleamos el mismo estadístico t , pero ...

- asumiendo que el número posible de comparaciones es infinito
- que no hemos planeado ningún contraste de antemano
- y que incluso podemos haber caído en la tentación de hacer *data snooping*

Para poner a prueba la hipótesis nula $C = 0$ empleamos el mismo estadístico t , pero ...

- asumiendo que el número posible de comparaciones es infinito
- que no hemos planeado ningún contraste de antemano
- y que incluso podemos haber caído en la tentación de hacer *data snooping*

Entonces:

$$t_0 = \frac{\hat{C}^2}{(J-1)s_{\hat{C}}^2} \sim F_{J-1, n-J}$$

¿QUÉ MÉTODO USAR?

- Use el método de Bonferroni cuando tenga planeado de ANTEMANO un cierto conjunto de comparaciones.
- Use el método de Tukey cuando le interese llevar a cabo todas las comparaciones 2 a 2.
- Use el método de Scheffé cuando no tenga ningún plan de antemano.

¿QUÉ MÉTODO USAR?

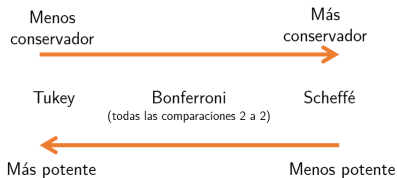
- Use el método de Bonferroni cuando tenga planeado de ANTEMANO un cierto conjunto de comparaciones.
- Use el método de Tukey cuando le interese llevar a cabo todas las comparaciones 2 a 2.
- Use el método de Scheffé cuando no tenga ningún plan de antemano.

El enfoque de Scheffé es más exploratorio mientras que los 2 anteriores son de índole confirmatorio.

¿QUÉ MÉTODO USAR?

- Use el método de Bonferroni cuando tenga planeado de ANTEMANO un cierto conjunto de comparaciones.
- Use el método de Tukey cuando le interese llevar a cabo todas las comparaciones 2 a 2.
- Use el método de Scheffé cuando no tenga ningún plan de antemano.

El enfoque de Scheffé es más exploratorio mientras que los 2 anteriores son de índole confirmatorio.





Volvamos al ejemplo anterior para determinar cuál (o cuales) es la mejor marca en términos de la vida útil más larga.



Volvamos al .

Pese a que el propósito del modelado haya cambiado levemente, la herramienta que se emplea para realizar este tipo de análisis continua siendo el modelo lineal. Por este motivo, es una buena práctica complementar el análisis con la etapa de diagnóstico.

Pese a que el propósito del modelado haya cambiado levemente, la herramienta que se emplea para realizar este tipo de análisis continua siendo el modelo lineal. Por este motivo, es una buena práctica complementar el análisis con la etapa de diagnóstico.

En estos casos, los análisis de linealidad y multicolinealidad no tienen mucho sentido debido a la *forma* de la matriz X . No obstante, es bueno validar las inferencias corroborando los demás supuestos.

- Homoscedasticidad
- Normalidad
- Observaciones atípicas

Pese a que el propósito del modelado haya cambiado levemente, la herramienta que se emplea para realizar este tipo de análisis continua siendo el modelo lineal. Por este motivo, es una buena práctica complementar el análisis con la etapa de diagnóstico.

En estos casos, los análisis de linealidad y multicolinealidad no tienen mucho sentido debido a la *forma* de la matriz X . No obstante, es bueno validar las inferencias corroborando los demás supuestos.

- Homoscedasticidad
- Normalidad
- Observaciones atípicas

Ante el no cumplimiento de alguno de estos supuestos, es posible llevar a cabo alguno de los procedimientos vistos en clases anteriores o incluso llevar a cabo el equivalente no paramétrico del ANOVA a una vía, la prueba de *Kruskal-Wallis*.



La próxima hablaremos de:

- Veremos los aspectos teóricos del modelo a 2 vías.
- Consideraremos situaciones donde se aplica este modelo.
- Introduciremos el concepto de interacción.



Carmona, Francesc (2003). *Modelos Lineales (notas de curso)*. Departament d'Estadística.



Peña, Daniel (2010). *Regresión y Diseño de Experimentos*. Alianza Editorial.



Rencher, Alvin y Bruce Schaalje (2008). *Linear Models in Statistics, second edition*. John Wiley Sons, Inc.

¿Preguntas?

Muchas Gracias