Modelos Lineales

Análisis de Varianza a 2 vías

Fernando Massa; Bruno Bellagamba

11 de junio 2024

FCEA Facultad de Ciencias Económicas y de Administración
IESTA Instituto de Estadística
Universidad de la República Uruguay

Temario

1. Introducción
2. Two way ANOVA
3. Modelo aditivo
4. Modelo con interacción
5. Próxima clase

Objetivos

- Introduciremos una variable explicativa más al modelo de la clase anterior.
- Veremos los aspectos teóricos del modelo a 2 vías.
- Consideraremos situaciones donde se aplica este modelo.
- Introduciremos el concepto de interacción.

Los modelos que consideran el efecto de una segunda variable o factor sobre Y suelen aparecer en situaciones donde su incorporación al modelo responde a considerar condiciones homogéneas para todas las unidades. Este tipo de variables se denominan *bloques*.

Los modelos que consideran el efecto de una segunda variable o factor sobre Y suelen aparecer en situaciones donde su incorporación al modelo responde a considerar condiciones homogéneas para todas las unidades. Este tipo de variables se denominan *bloques*.

No obstante, también existen situaciones donde la inclusión de una segunda variable responde al mismo objetivo que la primera e incluso existe interés por conocer si el efecto de uno de estos factores cambia a través de los niveles del otro factor.

Los modelos que consideran el efecto de una segunda variable o factor sobre Y suelen aparecer en situaciones donde su incorporación al modelo responde a considerar condiciones homogéneas para todas las unidades. Este tipo de variables se denominan *bloques*.

No obstante, también existen situaciones donde la inclusión de una segunda variable responde al mismo objetivo que la primera e incluso existe interés por conocer si el efecto de uno de estos factores cambia a través de los niveles del otro factor.

En la clase de hoy exploraremos como *ampliar* el modelo de ANOVA a 1 vía para contemplar este tipo de situaciones y como lidiar con las consideraciones teóricas que surjan en el camino.

MODELO *a dos vías*

El *"Two way model"* se emplea en situaciones donde se quiere determinar si el promedio de la variable de Y *responde* a los niveles de un cierto factor x_1 u otro factor x_2.

Modelo *a dos vías*

El *"Two way model"* se emplea en situaciones donde se quiere determinar si el promedio de la variable de Y *responde* a los niveles de un cierto factor x_1 u otro factor x_2.

Antes de postular el modelo es importante aclarar algunos detalles que pueden incidir sobre los subíndices.

- Hablamos de **réplicas** cuando se dispone de más de una observación para cada tratamiento u combinaciones de tratamientos. El el modelo a una vía de la clase pasada, el número de réplicas era n_j (15 en el ejemplo del los neumáticos).
- Al analizar el modelo a 2 vías debemos considerar 2 especificaciones:
  1. Modelo aditivo.
  2. Modelo con interacción.

Modelo *a dos vías*

El caso más simple corresponde al anális del modelo a 2 vías considerando los factores A y B con *efectos aditivos* **SIN** replicas. De esta manera:

El caso más simple corresponde al anális del modelo a 2 vías considerando los factores A y B con *efectos aditivos* **SIN** replicas. De esta manera:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b$$

Antes de imponer restricciones, los parámetros corresponden a la media general y los efectos de los niveles del los factores A y B respectivamente.

MODELO *a dos vías*

El caso más simple corresponde al anális del modelo a 2 vías considerando los factores A y B con *efectos aditivos* **SIN** replicas. De esta manera:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b$$

Antes de imponer restricciones, los parámetros corresponden a la media general y los efectos de los niveles del los factores A y B respectivamente.

En cuanto a los subíndices diremos que y_{ij} es la observación de Y correspondiente al i-ésimo nivel de A (con efecto α_i) y al j-ésimo nivel del factor B (con efecto β_j).

El caso más simple corresponde al anális del modelo a 2 vías considerando los factores A y B con *efectos aditivos* **SIN** replicas. De esta manera:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b$$

Antes de imponer restricciones, los parámetros corresponden a la media general y los efectos de los niveles del los factores A y B respectivamente.

En cuanto a los subíndices diremos que y_{ij} es la observación de Y correspondiente al i-ésimo nivel de A (con efecto α_i) y al j-ésimo nivel del factor B (con efecto β_j).

En este caso $n = ab$.

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa.

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa. No obstante, esto no siempre es posible, es una buena práctica contar con más de una observación en cada combinación.

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa. No obstante, esto no siempre es posible, es una buena práctica contar con más de una observación en cada combinación.

La versión del modelo a dos vías que sí incluye réplicas, necesita un tercer subíndice.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b \quad \forall k = 1, \ldots, n_{ij}$$

Así, y_{ijk} es la **k-ésima** réplica de Y en la combinación del i-ésimo nivel de A (con efecto α_i) y del j-ésimo nivel del factor B (con efecto β_j).

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa. No obstante, esto no siempre es posible, es una buena práctica contar con más de una observación en cada combinación.

La versión del modelo a dos vías que sí incluye réplicas, necesita un tercer subíndice.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b \quad \forall k = 1, \ldots, n_{ij}$$

Así, y_{ijk} es la **k-ésima** réplica de Y en la combinación del i-ésimo nivel de A (con efecto α_i) y del j-ésimo nivel del factor B (con efecto β_j).

Si bien en la práctica es una buena costumbre disponer del mismo número de réplicas en cada combinación de tratamientos (tratamientos *balanceados*), esto suele suceder con mayor frecuencia en estudios experimentales y no en los observacionales.

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa. No obstante, esto no siempre es posible, es una buena práctica contar con más de una observación en cada combinación.

La versión del modelo a dos vías que sí incluye réplicas, necesita un tercer subíndice.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b \quad \forall k = 1, \ldots, n_{ij}$$

Así, y_{ijk} es la **k-ésima** réplica de Y en la combinación del i-ésimo nivel de A (con efecto α_i) y del j-ésimo nivel del factor B (con efecto β_j).

Si bien en la práctica es una buena costumbre disponer del mismo número de réplicas en cada combinación de tratamientos (tratamientos *balanceados*), esto suele suceder con mayor frecuencia en estudios experimentales y no en los observacionales.

En este caso $n = \sum_{ij} n_{ij}$.

Modelo *a dos vías*

El modelo anterior es útil cuando existen razones para pensar que cada y_{ij} es lo suficientemente precisa. No obstante, esto no siempre es posible, es una buena práctica contar con más de una observación en cada combinación.

La versión del modelo a dos vías que sí incluye réplicas, necesita un tercer subíndice.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \quad \forall i = 1, \dots, a \quad \forall j = 1, \dots, b \quad \forall k = 1, \dots, n_{ij}$$

Así, y_{ijk} es la **k-ésima** réplica de Y en la combinación del i-ésimo nivel de A (con efecto α_i) y del j-ésimo nivel del factor B (con efecto β_j).

Si bien en la práctica es una buena costumbre disponer del mismo número de réplicas en cada combinación de tratamientos (tratamientos *balanceados*), esto suele suceder con mayor frecuencia en estudios experimentales y no en los observacionales.

En este caso $n = \sum_{ij} n_{ij}$. Y en caso de que $n_{ij} = r \quad \forall i, j$, entonces $n = rab$

Modelo *a dos vías*

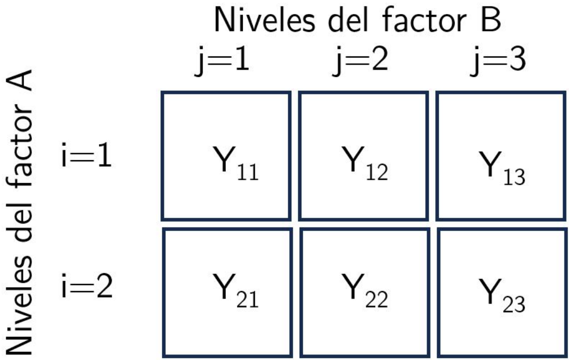

Figura: Modelo a 2 vías

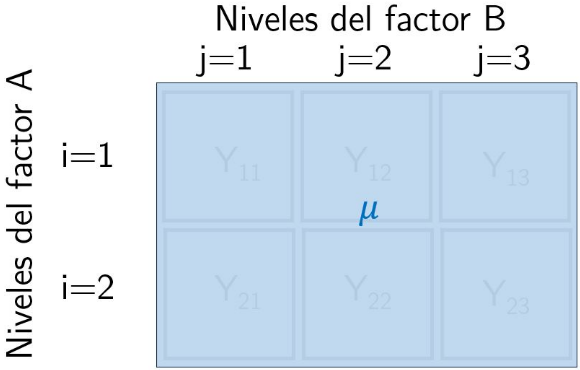

FIGURA: Efecto global

MODELO *a dos vías*

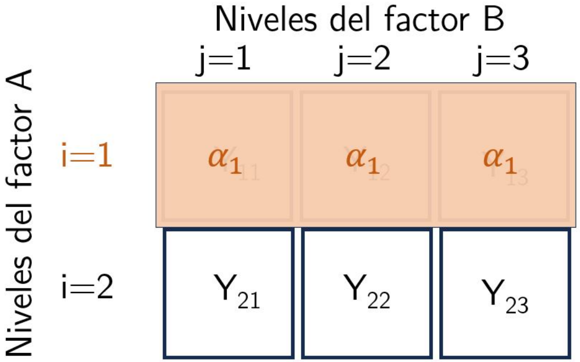

FIGURA: Efecto del tratamiento 1 (del factor A)

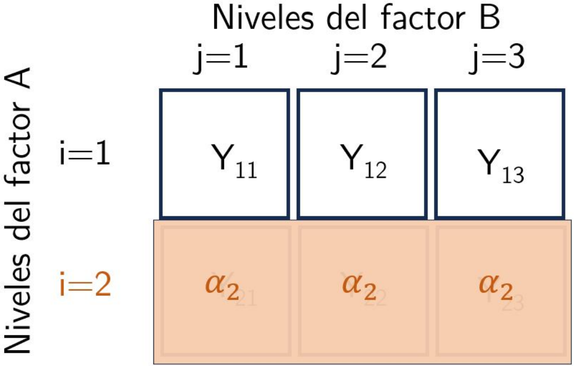

FIGURA: Efecto del tratamiento 2 (del factor A)

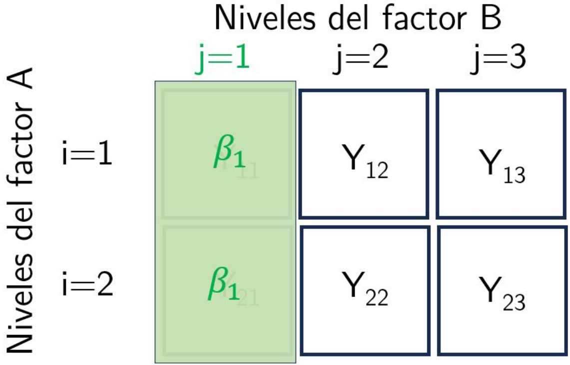

FIGURA: Efecto del tratamiento 1 (del factor B)

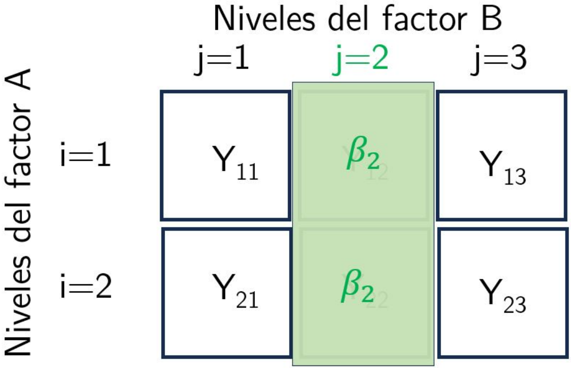

FIGURA: Efecto del tratamiento 2 (del factor B)

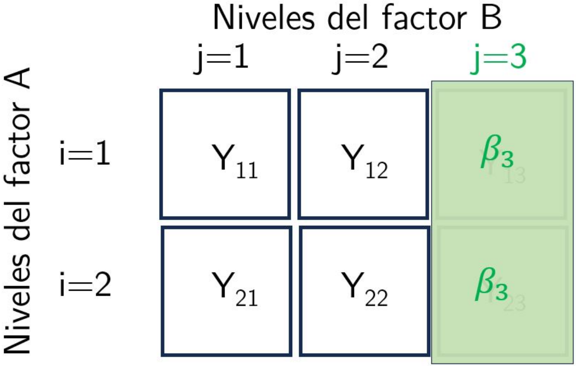

FIGURA: Efecto del tratamiento 3 (del factor B)

Restricciones

El mismo problema que surgía con el rango de la matriz X en el modelo a una vía, se intensifica en el modelo a dos vías. No obstante, las soluciones son las mismas.

RESTRICCIONES

El mismo problema que surgía con el rango de la matriz X en el modelo a una vía, se intensifica en el modelo a dos vías. No obstante, las soluciones son las mismas.

En este curso optaremos por trabajar bajo restricciones de la forma:

- Fijando en cero el primer efecto de **cada** factor: $\alpha_1 = 0$ y $\beta_1 = 0$
- Fijando en cero la suma de los efectos de **cada** factor: $\sum_{i=1}^{a} \alpha_i = 0$ y $\sum_{j=1}^{b} \beta_j = 0$

Restricciones

El mismo problema que surgía con el rango de la matriz X en el modelo a una vía, se intensifica en el modelo a dos vías. No obstante, las soluciones son las mismas.

En este curso optaremos por trabajar bajo restricciones de la forma:

- Fijando en cero el primer efecto de **cada** factor: $\alpha_1 = 0$ y $\beta_1 = 0$
- Fijando en cero la suma de los efectos de **cada** factor: $\sum_{i=1}^{a} \alpha_i = 0$ y $\sum_{j=1}^{b} \beta_j = 0$

En cuanto a las interpretaciones, en el primer caso:

- μ es al valor promedio de Y en el nivel de referencia del factor A y del factor B.
- α_i es la diferencia entre el promedio del i-ésimo tratamiento del factor A respecto del promedio del nivel de referencia (en cualquier nivel de B).
- β_j es la diferencia entre el promedio del j-ésimo tratamiento del factor B respecto del promedio del nivel de referencia (en cualquier nivel de A).

Restricciones

El mismo problema que surgía con el rango de la matriz X en el modelo a una vía, se intensifica en el modelo a dos vías. No obstante, las soluciones son las mismas.

En este curso optaremos por trabajar bajo restricciones de la forma:

- Fijando en cero el primer efecto de **cada** factor: $\alpha_1 = 0$ y $\beta_1 = 0$
- Fijando en cero la suma de los efectos de **cada** factor: $\sum_{i=1}^{a} \alpha_i = 0$ y $\sum_{j=1}^{b} \beta_j = 0$

En cuanto a las interpretaciones, en el segundo caso:

- μ es el promedio global de Y.
- α_i es la diferencia entre el promedio del i-ésimo tratamiento del factor A respecto del promedio global (en cualquier nivel de B).
- β_j es la diferencia entre el promedio del j-ésimo tratamiento del factor B respecto del promedio global (en cualquier nivel de A).

Al momento de realizar inferencia, las pruebas de significación individual deben llevarse a cabo con las mismas consideraciones que en el caso del modelo a 1 vía.

Al momento de realizar inferencia, las pruebas de significación individual deben llevarse a cabo con las mismas consideraciones que en el caso del modelo a 1 vía.

Por otro lado, la prueba de significación global del modelo se parece más a la prueba global de los modelos de regresión.

Al momento de realizar inferencia, las pruebas de significación individual deben llevarse a cabo con las mismas consideraciones que en el caso del modelo a 1 vía.

Por otro lado, la prueba de significación global del modelo se parece más a la prueba global de los modelos de regresión. En este caso, Un valor significativo del estadístico F indicaría que el promedio de la variable Y varía en al menos uno de los niveles de al menos uno de los 2 factores.

$$\begin{array}{ll} H_0) & \alpha_1 = \alpha_2 = \ldots = \alpha_a = \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_1) & \text{Existe al menos un 1 efecto distinto de cero} \end{array}$$

Al momento de realizar inferencia, las pruebas de significación individual deben llevarse a cabo con las mismas consideraciones que en el caso del modelo a 1 vía.

Por otro lado, la prueba de significación global del modelo se parece más a la prueba global de los modelos de regresión. En este caso, Un valor significativo del estadístico F indicaría que el promedio de la variable Y varía en al menos uno de los niveles de al menos uno de los 2 factores.

$$\begin{array}{ll} H_0) & \alpha_1 = \alpha_2 = \ldots = \alpha_a = \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_1) & \text{Existe al menos un 1 efecto distinto de cero} \end{array}$$

Dicho de otro modo, al menos uno de los dos factores tiene un efecto sobre el promedio de Y.

Pese a que la salida estándar del R nos provee información respecto de cada efecto de cada factor, no nos permite conocer la significación del factor en su conjunto. Esto se torna relevante, cuando el factor en cuestión tiene 3 o más niveles.

Inferencia

Pese a que la salida estándar del R nos provee información respecto de cada efecto de cada factor, no nos permite conocer la significación del factor en su conjunto. Esto se torna relevante, cuando el factor en cuestión tiene 3 o más niveles.

Para poner a prueba el *aporte* de cada factor, existen (al menos) 2 alternativas equivalentes.

Pese a que la salida estándar del R nos provee información respecto de cada efecto de cada factor, no nos permite conocer la significación del factor en su conjunto. Esto se torna relevante, cuando el factor en cuestión tiene 3 o más niveles.

Para poner a prueba el *aporte* de cada factor, existen (al menos) 2 alternativas equivalentes.

- Llevar a cabo pruebas de hipótesis lineales como las vistas anteriormente ($C\beta = b$).
- Realizar una prueba F que compare la $SCRes$ del modelo completo (2 factores) con la del modelo reducido (solo 1 factor).

INFERENCIA

Pese a que la salida estándar del R nos provee información respecto de cada efecto de cada factor, no nos permite conocer la significación del factor en su conjunto. Esto se torna relevante, cuando el factor en cuestión tiene 3 o más niveles.

Para poner a prueba el *aporte* de cada factor, existen (al menos) 2 alternativas equivalentes.

- Llevar a cabo pruebas de hipótesis lineales como las vistas anteriormente ($C\beta = b$).
- Realizar una prueba F que compare la $SCRes$ del modelo completo (2 factores) con la del modelo reducido (solo 1 factor).

En ambos casos, se obtiene el mismo valor del estadístico y, por ende, el mismo p-valor.

Utilicemos el modelo de ANOVA a dos vías para analizar un experimento donde se trató de determinar si la cantidad de luz que emite un tubo de luz depende del material del tubo y la temperatura a la que opera.

Vayamos al R y veamos como ajustar estos modelos.

La inclusión de este término responde a que en ciertas situaciones, el efecto de uno de los factores cambia dependiendo de los niveles del otro.

Interacción

La inclusión de este término responde a que en ciertas situaciones, el efecto de uno de los factores cambia dependiendo de los niveles del otro.

Pensemos por ejemplo, en el caso de que que un fármaco elaborado por 3 laboratorios (factor A) tenga mejor o peor efecto dependiendo de la concentración con la que se lo aplique (factor B).

La inclusión de este término responde a que en ciertas situaciones, el efecto de uno de los factores cambia dependiendo de los niveles del otro.

Pensemos por ejemplo, en el caso de que que un fármaco elaborado por 3 laboratorios (factor A) tenga mejor o peor efecto dependiendo de la concentración con la que se lo aplique (factor B).

En estos casos el modelo se especifica de la siguiente manera:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b \quad \forall k = 1, \ldots, n_{ij}$$

En estos casos es *INDISPENSABLE* contar con al menos 2 observaciones en cada combinación de los factores.

Desde un enfoque basado en el contexto de diseño de experimentos decimos que los 2 factores están *cruzados* o *anidados* cuando:

- **Cruzados**: Todos los niveles del factor A se combinaron con todos los niveles del factor B.
- **Anidados**: Los niveles que se aplicaron del factor A dependen del nivel del factor B.

CRUZAMIENTO *vs* ANIDACIÓN

Desde un enfoque basado en el contexto de diseño de experimentos decimos que los 2 factores están *cruzados* o *anidados* cuando:

- **Cruzados**: Todos los niveles del factor A se combinaron con todos los niveles del factor B.
- **Anidados**: Los niveles que se aplicaron del factor A dependen del nivel del factor B.

La correcta estimación del término de interacción entre 2 factores requiere que ambos estén cruzados.

Restricciones

La inclusión del término de interacción permite modelar situaciones más complejas, pero con el costo de añadir potencialmente ab parámetros $(\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ab})$.

Restricciones

La inclusión del término de interacción permite modelar situaciones más complejas, pero con el costo de añadir potencialmente ab parámetros $(\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ab})$.

Al igual que sucedía con el modelo a una vía o con el modelo de efectos aditivos, para obtener estimaciones únicas es necesario imponer restricciones (o reparametrizar o trabajar con funciones estimables).

Restricciones

La inclusión del término de interacción permite modelar situaciones más complejas, pero con el costo de añadir potencialmente ab parámetros $(\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ab})$.

Al igual que sucedía con el modelo a una vía o con el modelo de efectos aditivos, para obtener estimaciones únicas es necesario imponer restricciones (o reparametrizar o trabajar con funciones estimables).

Las versiones de los 2 tipos de restricciones que hemos mencionado son:

- Anulando el primer nivel: $\gamma_{1j} = 0 \quad \forall j = 1, \ldots, b$ y $\gamma_{i1} = 0 \quad \forall i = 1, \ldots, a$
- Suma cero: $\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0 \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b$

Restricciones

La inclusión del término de interacción permite modelar situaciones más complejas, pero con el costo de añadir potencialmente ab parámetros $(\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ab})$.

Al igual que sucedía con el modelo a una vía o con el modelo de efectos aditivos, para obtener estimaciones únicas es necesario imponer restricciones (o reparametrizar o trabajar con funciones estimables).

Las versiones de los 2 tipos de restricciones que hemos mencionado son:

- Anulando el primer nivel: $\gamma_{1j} = 0 \quad \forall j = 1, \ldots, b$ y $\gamma_{i1} = 0 \quad \forall i = 1, \ldots, a$
- Suma cero: $\sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0 \quad \forall i = 1, \ldots, a \quad \forall j = 1, \ldots, b$

En ambos 2 casos, las variables a incluir en X surgen de considerar los productos cruzados entre las variables ya presentes en X correspondientes a los 2 factores del modelo aditivo.

Interpretación

La interpretación de los parámetros es ligeramente más compleja. Se presenta el caso correspondiente a las restricciones que anulan primer el efecto de cada factor en un experimento donde $a = 2$, $b = 2$ y $n_{ij} = 3 \quad \forall i, j$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \forall i = 1,2 \quad \forall j = 1,2 \quad \forall k = 1,2,3$$

Interpretación

La interpretación de los parámetros es ligeramente más compleja. Se presenta el caso correspondiente a las restricciones que anulan primer el efecto de cada factor en un experimento donde $a = 2$, $b = 2$ y $n_{ij} = 3 \quad \forall i, j$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \forall i = 1,2 \quad \forall j = 1,2 \quad \forall k = 1,2,3$$

Para realizar la interpretación resulta conveniente contar con la media de Y en cada combinación de i, j.

$$\begin{aligned}
\mathbb{E}(Y_{11k}) &= \mu \\
\mathbb{E}(Y_{12k}) &= \mu + \beta_2 \\
\mathbb{E}(Y_{21k}) &= \mu + \alpha_2 \\
\mathbb{E}(Y_{22k}) &= \mu + \alpha_2 + \beta_2 + \gamma_{22}
\end{aligned}$$

Interpretación

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**

Interpretación

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.
  2. α_2 es el efecto del segundo nivel del factor A respecto del primero, **en el nivel de referencia del factor B**.

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.
  2. α_2 es el efecto del segundo nivel del factor A respecto del primero, **en el nivel de referencia del factor B**.
  3. β_2 es el efecto del segundo nivel del factor B respecto del primero, **en el nivel de referencia del factor A**.

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.
  2. α_2 es el efecto del segundo nivel del factor A respecto del primero, **en el nivel de referencia del factor B**.
  3. β_2 es el efecto del segundo nivel del factor B respecto del primero, **en el nivel de referencia del factor A**.
- **Interacción**. Este término adquiere significado de 2 formas.

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.
  2. α_2 es el efecto del segundo nivel del factor A respecto del primero, **en el nivel de referencia del factor B**.
  3. β_2 es el efecto del segundo nivel del factor B respecto del primero, **en el nivel de referencia del factor A**.
- **Interacción**. Este término adquiere significado de 2 formas.
  1. γ_{22} es el incremento en el efecto del segundo nivel del factor A **en el segundo nivel del factor B**.

A partir de la diapositiva anterior se puede interpretar que:

- **Efectos principales**
  1. μ es el valor promedio de Y en la combinación de los primeros niveles de ambos factores.
  2. α_2 es el efecto del segundo nivel del factor A respecto del primero, **en el nivel de referencia del factor B**.
  3. β_2 es el efecto del segundo nivel del factor B respecto del primero, **en el nivel de referencia del factor A**.
- **Interacción**. Este término adquiere significado de 2 formas.
  1. γ_{22} es el incremento en el efecto del segundo nivel del factor A **en el segundo nivel del factor B**.
  2. γ_{22} es el incremento en el efecto del segundo nivel del factor B **en el segundo nivel del factor A**.

Al imponer restricciones sobre los parámetros del modelo, se cuenta con:

- 1 parámetro para la media.
- $a-1$ parámetros correspondientes al primer factor.
- $b-1$ parámetros del segundo.
- $(a-1)(b-1)$ para la interacción.

Al imponer restricciones sobre los parámetros del modelo, se cuenta con:

- 1 parámetro para la media.
- $a-1$ parámetros correspondientes al primer factor.
- $b-1$ parámetros del segundo.
- $(a-1)(b-1)$ para la interacción.

Al momento de llevar a cabo la inferencia, resulta conveniente realizar primero la prueba de la significación de la interacción.

Inferencia

Al imponer restricciones sobre los parámetros del modelo, se cuenta con:

- 1 parámetro para la media.
- $a-1$ parámetros correspondientes al primer factor.
- $b-1$ parámetros del segundo.
- $(a-1)(b-1)$ para la interacción.

Al momento de llevar a cabo la inferencia, resulta conveniente realizar primero la prueba de la significación de la interacción.

$$\begin{aligned} H_0) &\quad \gamma_{11} = \gamma_{12} = \ldots = \gamma_{ab} \\ H_1) &\quad \text{no } H_0) \end{aligned}$$

Esto puede realizarse con pruebas de hipótesis lineales o comparando las *SCRes* del modelo con interacción respecto de su análoga en el modelo aditivo.

Al analizar la significación de los efectos principales debe tenerse en cuenta un detalle importante que se desprende de la interpretación de los efectos.

Al analizar la significación de los efectos principales debe tenerse en cuenta un detalle importante que se desprende de la interpretación de los efectos.

Las pruebas F realizadas sobre estos efectos **NO** contrastan la nula de que no existe un efecto del factor bajo escrutinio, sino que solo se refieren a los efectos de dicho factor *en el nivel anulado del otro*.

Al analizar la significación de los efectos principales debe tenerse en cuenta un detalle importante que se desprende de la interpretación de los efectos.

Las pruebas F realizadas sobre estos efectos **NO** contrastan la nula de que no existe un efecto del factor bajo escrutinio, sino que solo se refieren a los efectos de dicho factor *en el nivel anulado del otro*.

Por este motivo, resulta conveniente que:

- Si el F de la interacción es significativo, parar ahí y pasar a las comparaciones múltiples.
- Si el F de la interacción **NO** es significativo, pasar al modelo aditivo y realizar inferencia sobre ese.

Retomemos el experimento de los tubos de luz e indaguemos sobre la presencia de interacción entre la temperatura y el material del vidrio.

Vayamos al R.

La próxima clase:

- Iniciaremos la transición de vuelta a los modelos de regresión analizando el modelo ANCOVA.
- Veremos como varios conceptos de los modelos de ANOVA pueden surgir en los modelos de regresión.
- Haremos especial hincapié en el concepto de interacción cuando una (o ambas) de las variables es cuantitativa.

- Carmona, Francesc (2003). *Modelos Lineales (notas de curso).* Departament d'Estadística.
- Farraway, Julian (2014). *Linear Models with R, second edition.* Chapman Hall/CRC.
- Rencher, Alvin y Bruce Schaalje (2008). *Linear Models in Statistics, second edition.* John Wiley Sons, Inc.

¿Preguntas?

Muchas Gracias