

# Modelo Lineal Generalizado ANOVA

Alicia Giménez

31 de octubre de 2018

*Resumen: El presente trabajo es un resumen de las características del modelo lineal generalizado ANOVA, en el cual la componente aleatoria presenta una distribución normal, un enlace identidad y la componente sistemática es categórica. Además a modo de ejemplificar su aplicación, realizamos un análisis en el programa estadístico Rstudio y la redacción del material utilizando Rmarkdown. No se utilizó ningún paquete en específico sino funciones por defecto del programa, trabajamos con matrices (Algunos de los códigos más importantes se encuentran en el anexo).*

## Modelos Lineales Generalizados (GLM)

Los modelos lineales generalizados son una amplia clase de modelos que incluyen regresión lineal, ANOVA, regresión de Poisson, modelos log-lineales, etc. La siguiente tabla proporciona un buen resumen de los GLM después de Agresti (cap 4, 2013):

Modelo	Aleatorio	Enlace	Sistemático
Regresión lineal	Normal	Identidad	Continuo
ANOVA	Normal	Identidad	Categórico
ANCOVA	Normal	Identidad	Mezclado
Regresión logística	Binomial	Logit	Mezclado
Loglineal	Poisson	Log	Categórico
Regresión de Poisson	Poisson	Log	Mezclado
Respuesta Multinomial	Multinomial	Logit General	Mezclado

Hay tres componentes para cualquier GLM:

- *Componente aleatorio*: se refiere a la distribución de probabilidad de la variable de respuesta (Y); por ejemplo, la distribución normal para Y en la regresión lineal, o la distribución binomial para Y en la regresión logística binaria. También se llama modelo de ruido o modelo de error.
- *Componente sistemático*: especifica las variables explicativas ( $X_1, X_2, \dots, X_k$ ) en el modelo, más específicamente su combinación lineal para crear el llamado *predictor lineal*; por ejemplo,  $\beta_0 + \beta_1 x_1 + \beta_2 x_2$  como hemos visto en una regresión lineal, o una regresión logística.
- *Función de enlace,  $\eta$  o  $g(\mu)$* : especifica el enlace entre componentes aleatorios y sistemáticos. Dice cómo el valor esperado de la respuesta se relaciona con el predictor lineal de las variables explicativas; por ejemplo,  $\eta = g(E(Y_i)) = E(Y_i)$  para regresión lineal, o  $\eta = \text{logit}(\pi)$  para regresión logística.

*Supuestos:*

- Los datos  $Y_1, Y_2, \dots, Y_k$  se distribuyen de forma independiente, es decir, los casos son independientes.
- La variable dependiente Y no necesita estar distribuida normalmente, pero generalmente asume la distribución de una familia exponencial (por ejemplo, binomial, poisson, multinomial, normal, etc. . .)
- GLM no asume una relación lineal entre la variable dependiente y las variables independientes, pero asume una relación lineal entre la respuesta transformada en términos de la función de enlace y las variables explicativas; por ejemplo, para de regresión logística binaria  $\text{logit}(\pi) = \beta_0 + \beta X$ .
- Las variables independientes pueden ser transformaciones no lineales de las variables independientes originales.

- La homogeneidad de varianza no necesariamente debe cumplirse. De hecho, en muchos casos ni siquiera es posible dada la estructura del modelo, y la *sobredispersión* (cuando la varianza observada es mayor de lo que supone el modelo) posiblemente presente.
- Los errores deben ser independientes pero no normalmente distribuidos.

-Se utiliza el método de máxima verosimilitud (MLE) en lugar de mínimos cuadrados ordinarios (OLS) para estimar los parámetros  $\beta$ , por lo tanto, se basa en aproximaciones de grandes muestras.

- Las medidas de bondad de ajuste se basan en muestras suficientemente grandes, donde una regla heurística es que no más del 20% de los recuentos esperados de las celdas sea menor a 5.

Los siguientes son ejemplos de componentes GLM para modelos que ya conocemos, como la regresión lineal, regresión logística y regresión log-lineal.

**Regresión lineal simple** modela cómo el valor esperado de una variable de respuesta continua depende de un conjunto de variables explicativas, donde el índice  $i$  representa cada punto de datos.

$$Y_i = \beta_0 + \beta x_i + \epsilon_i$$

o

$$E(Y_i) = \beta_0 + \beta x_i + \epsilon_i$$

- *Componente aleatorio*:  $Y$  es una variable de respuesta y tiene una distribución normal, generalmente asumimos errores,  $\epsilon \sim N(0, \sigma^2)$ .
- *Componente sistemático*:  $X$  es la variable explicativa (puede ser continua o discreta) y es lineal en los parámetros  $\beta_0 + \beta x_i$ . Observe que con una regresión lineal múltiple donde tenemos más de una variable explicativa, por ejemplo  $(X_1, X_2, \dots, X_k)$ , tendríamos una combinación lineal de estas  $X$  en términos de los parámetros de regresión  $\beta$ , pero las variables explicativas en sí podrían transformarse, por ejemplo  $X^2$  o  $\log(X)$ .
- *Función de enlace*: enlace de identidad,  $\eta = g(\mu) = E(Y_i)$  porque estamos modelando la media directamente; esta es la función de enlace más simple.

**Regresión logística binaria**: modela cómo la variable de respuesta binaria  $Y$  depende de un conjunto de  $k$  variables explicativas,  $X = (X_1, X_2, \dots, X_k)$ .

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta x_i + \dots + \beta_0 + \beta x_k$$

que modela las probabilidades de registro de probabilidad de “éxito” en función de las variables explicativas.

- *Componente aleatorio*: se supone que la distribución de  $Y$  es binomial  $(n, \pi)$ , donde  $\pi$  es una probabilidad de “éxito”.
- *Componente sistemático*: las  $X$  son variables explicativas (pueden ser continuas, discretas o ambas) y son lineales en los parámetros, por ejemplo,  $\beta_0 + \beta x_i + \dots + \beta_0 + \beta x_k$ . Nuevamente, la transformación de las mismas  $X$  se permite como en regresión lineal; esto vale para cualquier GLM.
- *Función de enlace*:  $\eta = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$ .

Más generalmente, el enlace logit modela las probabilidades de registro de la media, y la media aquí es  $\pi$ . Los modelos de regresión logística binaria también se conocen como modelos logit cuando los predictores son todos categóricos.

**Modelo Log-linear**: modela los recuentos esperados en función de los niveles de las variables categóricas; por ejemplo, para una tabla de dos entradas el modelo saturado.

$$\log(\mu_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

donde  $\mu_{ij} = E(n_{ij})$  los recuentos de celdas esperadas anteriormente (media en cada celda de la tabla de doble entrada), A y B representan dos variables categóricas, y  $\lambda_{ij}$  's son los parámetros del modelo, que están modelando el Registro natural de los conteos esperados.

- *Componente aleatorio*: la distribución de cuentas, que son las respuestas, es *Poisson*.
- *Componente sistemático*: las X son variables discretas utilizadas en la clasificación cruzada y son lineales en los parámetros

$$\lambda + \lambda_i^{x_1} + \lambda_j^{x_2} + \dots + \lambda_k^{x_k} + \dots$$

- *Función de enlace*: enlace Log,  $\eta = \log(\mu)$  porque estamos modelando el logaritmo de la media de las celdas.

Los modelos log-lineales son más generales que los modelos logit, y algunos modelos logit son equivalentes a ciertos modelos log-lineales. El modelo log-lineal también es equivalente al modelo de regresión de Poisson cuando todas las variables explicativas son discretas.

### Resumen de las ventajas de los GLM sobre la regresión tradicional (OLS)

- No necesitamos transformar la respuesta Y para tener una distribución normal.
- La elección del enlace es independiente de la elección del componente aleatorio, por lo que tenemos más flexibilidad en el modelado.
- Si el enlace produce efectos aditivos, no necesitamos una variación constante.
- Los modelos se ajustan mediante la estimación de máxima verosimilitud; Por lo tanto propiedades óptimas de los estimadores.
- Todas las herramientas de inferencia y la verificación de modelos que analizaremos para los modelos de regresión log-lineal y logística también se aplican a otros GLM; por ejemplo, pruebas de relación de Wald y Probabilidad, desviación, residuos, intervalos de confianza, sobredispersión.
- A menudo hay un procedimiento en un paquete de software para capturar todos los modelos mencionados anteriormente, por ejemplo, PROC GENMOD en SAS o `glm()` en R, etc... con opciones para variar los tres componentes.

Pero también hay algunas limitaciones de los GLM, como

- Función lineal, por ejemplo, solo puede tener un predictor lineal en el componente sistemático
- Las respuestas deben ser independientes.

Hay maneras de evitar estas restricciones; por ejemplo, considere el análisis de datos coincidentes, o use NLMIXED en SAS, o el paquete `{nlme}` en R, o considere otros modelos, otros paquetes de software.

### Análisis de varianza

Análisis de varianza es el término utilizado para el método estadístico de comparación de medias de grupos con observaciones continuas, donde los grupos están definidos por niveles de factores. En este caso todas las variables explicativas son categóricas y todos los elementos de la matriz de diseño **X** son variables dummificadas.

#### Análisis de varianza para un factor

Si las unidades experimentales se encuentran ubicadas aleatoriamente dentro de los grupos correspondientes a J niveles de un factor, se denomina experimento completamente aleatorio. Los datos pueden presentarse en una tabla en la que las columnas representan los niveles del factor y las filas representan a las observaciones. Las respuestas en el nivel j,  $Y_{j1}, \dots, Y_{jn_j}$ , todas poseen el mismo valor esperado y son denominadas réplicas. En general puede haber una cantidad distinta de observaciones  $n_j$  dentro de cada nivel.

Para simplificar los cálculos, suponemos que todas las muestras son del mismo tamaño  $n_j = K$  para  $j = 1, \dots, J$ . La respuesta  $y$  es un vector columna con  $N = JK$  mediciones.

$$y = [Y_{11}, Y_{12}, \dots, Y_{1K}, Y_{21}, \dots, Y_{J1}, \dots, Y_{JK}]^T$$

Consideraremos tres especificaciones distintas para los modelos y de esa manera probar la hipótesis de que la respuesta media es distinta para cada nivel del factor.

(a) La especificación más sencilla

$$E(Y_{jk}) = \mu_j$$

para  $j = 1, \dots, K$ .

Lo que también podría ser escrito como:

$$E(Y_i) = \sum x_{ij} \mu_j$$

$i = 1, \dots, N$ .

Donde  $x_{ij} = 1$  si la respuesta  $Y_i$  corresponde al nivel  $A_j$ , de lo contrario  $x_{ij} = 0$ .

De esta forma,  $E(y) = X\beta$  con

$$\beta = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_J \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & O & \cdot & \cdot \\ \cdot & O & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

Donde 0 y 1 son vectores de longitud  $K$  de ceros y unos respectivamente, y  $O$  indica los términos restantes de la matriz son ceros. Entonces  $X^T X$  es una matriz diagonal de  $J \times J$ .

$$X^T X = \begin{pmatrix} K & \cdot & \cdot & \cdot & \cdot \\ \cdot & K & \cdot & O & \cdot \\ \cdot & \cdot & K & \cdot & \cdot \\ \cdot & O & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & K \end{pmatrix}, X^T y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_J \end{pmatrix}$$

Lo que forma la ecuación

$$\beta = \frac{1}{K} \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_J \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \cdot \\ \cdot \\ \bar{Y}_J \end{pmatrix}$$

y

$$b^T X^T y = \frac{1}{K} \sum Y_j^2$$

Los valores ajustados son  $\hat{y} = [\bar{y}_1, \bar{y}_1, \dots, \bar{y}_1, \bar{y}_2, \dots, \bar{y}_J]^T$ . La desventaja de este modelo tan sencillo es que no puede ser extendido para más de un factor. Para generalizar esto, primero debemos especificar el modelo para que los parámetros de los niveles y combinaciones de los mismos puedan reflejar distintos efectos más allá de la media o respuestas específicas.

(b) La formulación del segundo modelo es

$$E(Y_{jk}) = \mu_j + \alpha_j$$

para  $j = 1, \dots, K$ . Donde  $\mu$  el efecto medio para todos los niveles y  $\alpha_j$  es el efecto adicional debido al nivel  $A_j$ . Para esta parametrización hay  $J+1$  parámetros.

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \cdot \\ \cdot \\ \alpha_J \end{pmatrix}, X = \begin{pmatrix} 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 1 & & & & \\ \cdot & & & & & O & \\ \cdot & O & & & & & \\ 1 & & & & & & 1 \end{pmatrix}$$

donde **0** y **1** son vectores de longitud K y **O** representa una matriz de ceros.

Entonces

$$X^T X = \begin{pmatrix} N & K & \cdot & \cdot & K \\ K & K & \cdot & \cdot & \cdot \\ \cdot & \cdot & K & \cdot & \cdot \\ \cdot & O & \cdot & \cdot & \cdot \\ K & \cdot & \cdot & \cdot & K \end{pmatrix}, X^T y = \begin{pmatrix} Y_{..} \\ Y_{1.} \\ \cdot \\ \cdot \\ Y_{J.} \end{pmatrix}$$

Las primeras dos filas (o columnas) de la matriz  $X^T X$  es la suma de las filas (o columnas) restantes, por lo que  $X^T X$  es singular y no tiene solución única para la ecuación normal  $X^T X b = X^T y$ . La solución general podría escribirse como

$$\beta = \begin{pmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \cdot \\ \cdot \\ \hat{\alpha}_J \end{pmatrix} = \frac{1}{K} \begin{pmatrix} 0 \\ Y_{1.} \\ \cdot \\ \cdot \\ Y_{J.} \end{pmatrix} - \lambda \begin{pmatrix} -1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

donde  $\lambda$  es una constante arbitraria. Normalmente se impone de forma adicional la restricción

$$\sum \alpha_j = 0$$

de esta manera

$$\frac{1}{K} \sum Y_{j.} - J\lambda = 0$$

Por ende

$$\lambda = \frac{1}{JK} \sum Y_{j.} = \frac{Y_{..}}{N}$$

Lo que nos da la siguiente solución

$$\hat{\mu} = \frac{Y_{..}}{N}, \hat{\alpha}_j = \frac{Y_{j.}}{K} - \frac{Y_{..}}{N}$$

para  $j = 1, \dots, J$ . Por ende

$$b^T X^T y = \frac{Y_{..}^2}{N} + \sum Y_{j.} \left( \frac{Y_{j.}}{K} - \frac{Y_{..}}{N} \right) = \frac{1}{K} \sum Y_{j.}^2$$

Lo que resulta igual a la primera versión del modelo y los valores ajustados  $\hat{y} = [\bar{Y}_{11}, \bar{Y}_{12}, \dots, \bar{Y}_{1K}, \bar{Y}_{21}, \dots, \bar{Y}_{J1}, \dots, \bar{Y}_{JK}]^T$  son los mismos. La restricción de sumatoria igual a cero es utilizada de forma estándar en la mayoría de los programas estadísticos.

- (c) La tercera versión del modelo es  $E(Y_{jk}) = \mu + \alpha_j$  con la restricción de  $\alpha_1 = 0$ . Por ende  $\mu$  representa el efecto del primer nivel y  $\alpha_j$  mide la diferencia entre el primer nivel y el  $j$ -ésimo nivel del factor. Para esta versión hay J parámetros

$$\beta = \begin{pmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_J \end{pmatrix}, X = \begin{pmatrix} 1 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 1 & & & & \\ \cdot & & & & & O & \\ \cdot & O & & & & & \\ 1 & & & & & & 1 \end{pmatrix}$$

entonces

$$X^T X = \begin{pmatrix} N & K & \cdot & \cdot & K \\ K & K & \cdot & O & \cdot \\ \cdot & \cdot & K & \cdot & \cdot \\ \cdot & O & \cdot & \cdot & \cdot \\ K & \cdot & \cdot & \cdot & K \end{pmatrix}, X^T y = \begin{pmatrix} Y_{\cdot\cdot} \\ Y_2 \\ \cdot \\ \cdot \\ Y_J \end{pmatrix}$$

La matriz  $X^T X$  es no singular por lo que tiene solución única

$$\beta = \frac{1}{K} \begin{pmatrix} Y_{1\cdot} \\ Y_{2\cdot} - Y_{1\cdot} \\ \cdot \\ \cdot \\ Y_{J\cdot} - Y_{1\cdot} \end{pmatrix}$$

También  $b^T X^T y = \frac{1}{K} [Y_{\cdot\cdot} Y_{1\cdot} + \sum Y_{j\cdot} (Y_{j\cdot} - Y_{1\cdot})] = \frac{1}{K} \sum Y_{j\cdot}^2$  y los valores ajustados  $\hat{y} = [\bar{Y}_{11}, \bar{Y}_{12}, \dots, \bar{Y}_{1K}, \dots, \bar{Y}_{JK}]^T$  son los mismos de antes.

Aunque las tres especificaciones para el modelo difieran, el valor de  $b^T X^T y$  y por ende

$$D_1 = \frac{1}{\sigma^2} (y^T y - b^T X^T y) = \frac{1}{\sigma^2} \left( \sum \sum Y_{jk}^2 - \frac{1}{K} \sum Y_{j\cdot}^2 \right)$$

es el mismo en cada caso.

Estas tres versiones del modelo, todas corresponden a la hipótesis  $H_1$ , la respuesta media de cada nivel es distinta. Para compararlo con la hipótesis nula  $H_0$  de que las medias son iguales, consideraremos el modelo  $E(Y_{jk}) = \mu$  para que  $\beta = [\mu]$  y  $X$  sea un vector de  $N$  unos. De esa forma  $X^T X = N$ ,  $X^T y = Y_{\cdot\cdot}$ , por lo tanto  $b = \hat{\mu} = \frac{Y_{\cdot\cdot}}{N}$  así  $b^T X^T y = \frac{Y_{\cdot\cdot}^2}{N}$  y

$$D_0 = \frac{1}{\sigma^2} \left( \sum \sum Y_{jk}^2 - \frac{1}{N} \sum Y_{j\cdot}^2 \right)$$

también

$$F = \frac{D_0 - D_1}{J - 1} / \frac{D_1}{N - J} \sim F(J - 1, N - J)$$

Si  $H_0$  no es correcta, entonces  $F$  se espera que  $F$  sea mayor al de la distribución  $F(J-1, N-J)$ .

Para el caso de dos factores se realiza un procedimiento similar, el cual expondremos con un ejemplo de aplicación. **Ejemplo**

La siguiente tabla presenta datos para un ANOVA de dos factores con igual cantidad de observaciones para cada subgrupo.

<i>Niveles del factor A</i>	<i>Niveles del factor B</i>		
	$B_1$	$B_2$	Total
$A_1$	6.8 , 6.6	5.3 , 6.1	24.8
$A_2$	7.5 , 7.4	7.2 , 6.5	28.6
$A_3$	7.8 , 9.1	8.8 , 9.1	34.8
Total	45.2	43.0	88.2

La hipótesis principales son:

$H_I$  : No hay efectos de interacción, los efectos A y B son aditivos.

$H_A$  : No existen diferencias asociadas a los distintos niveles del factor A.

$H_B$  : No existen diferencias asociadas a los distintos niveles del factor B.

Para este efecto utilizamos el modelo saturado y otros tres modelos resultantes de la omisión de de varios términos del modelo saturado.

### Modelo Saturado

$$E(Y_{jkl}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$$

Donde los términos  $(\alpha\beta)_{jk}$  corresponden a los efectos de interaccion de  $\alpha_j$  y  $\beta_k$  siendo estos los efectos principales de los factores.

### Modelo Aditivo

$$E(Y_{jkl}) = \mu + \alpha_j + \beta_k$$

El cual compararemos con el modelo saturado para probar la hipótesis 1.

### Modelo A

$$E(Y_{jkl}) = \mu + \alpha_j$$

Es el creado eliminando los efectos del factor B, lo compararemos con el modelo aditivo para probar la hipótesis B.

### Modelo B

$$E(Y_{jkl}) = \mu + \beta_k$$

Es el creado eliminando los efectos del factor A, lo compararemos con el modelo aditivo para probar la hipótesis A.

Todos lo modelos mencionados anteriormente se encuentran sobreparametrizados debido a que las réplicas en cada subgrupo poseen el mismo valor esperado, por lo que podrían haber cuanto mucho JK valores esperados independientes, pero el modelo saturado posee  $1 + J + K + JK = (J + 1)(K + 1)$  parámetros. Para sobreparar esta dificultad( que nos llevaría a una matrix  $X^T X$  singular) impondremos algunas restricciones.

$$\alpha_1 = \beta_1 = (\alpha\beta)_{11} = (\alpha\beta)_{12} = (\alpha\beta)_{21} = (\alpha\beta)_{31} = 0$$

El vector de respuesta es el siguiente

$$y = \begin{pmatrix} 6.8 \\ 6.6 \\ 5.3 \\ 6.1 \\ 7.5 \\ 7.4 \\ 7.2 \\ 6.5 \\ 7.8 \\ 9.1 \\ 8.8 \\ 9.1 \end{pmatrix}$$

Para el modelo saturado, con las restricciones citadas anteriorente tendremos:

$$\beta = \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{32} \end{pmatrix}, X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}, X^T y = \begin{pmatrix} Y_{...} \\ Y_{2..} \\ Y_{3..} \\ Y_{12.} \\ Y_{22.} \\ Y_{32.} \end{pmatrix} = \begin{pmatrix} 88.2 \\ 28.6 \\ 34.8 \\ 43.0 \\ 13.7 \\ 17.9 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 12 & 4 & 4 & 6 & 2 & 1 \\ 4 & 4 & 0 & 2 & 2 & 0 \\ 4 & 0 & 4 & 2 & 0 & 2 \\ 6 & 2 & 2 & 6 & 2 & 2 \\ 2 & 2 & 0 & 2 & 2 & 0 \\ 2 & 0 & 2 & 2 & 0 & 2 \end{pmatrix}, b = \begin{pmatrix} 6.7 \\ 0.75 \\ 1.75 \\ -1.0 \\ 0.4 \\ 1.5 \end{pmatrix}$$

y  $b^T X^T y = 662.62$

Para el modelo aditivo con las restricciones  $\alpha_1 = \beta_1 = 0$  la matriz se obtiene eliminando las ultimas dos columnas de la matriz para el modelo saturado. Obtenemos de esa manera:

$$\beta = \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \end{pmatrix}, X^T X = \begin{pmatrix} 12 & 4 & 4 & 6 \\ 4 & 4 & 0 & 2 \\ 4 & 0 & 4 & 2 \\ 6 & 2 & 2 & 6 \end{pmatrix}, X^T y = \begin{pmatrix} 88.2 \\ 28.6 \\ 1.75 \\ 34.8 \end{pmatrix}$$

De la misma forma obtenemos

$$b = \begin{pmatrix} 6.383 \\ 0.950 \\ 2.500 \\ -0.367 \end{pmatrix}$$

y  $b^T X^T y = 661.4133$



Para el modelo A, en el cual omitimos los niveles del factor B, utilizamos la restricción  $\alpha_1 = 0$ , la matriz de diseño se obtiene eliminando las tres ultimas columnas de la matriz de diseño para el modelo saturado.

$$\beta = \begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, X^T X = \begin{pmatrix} 12 & 4 & 4 \\ 4 & 4 & 0 \\ 4 & 0 & 4 \end{pmatrix}, X^T y = \begin{pmatrix} 88.2 \\ 28.6 \\ 34.8 \end{pmatrix}$$

de la misma forma que antes obtenemos

$$b = \begin{pmatrix} 6.20 \\ 0.95 \\ 2.50 \end{pmatrix}$$

$$\text{y } b^T X^T y = 661.01$$

Para el modelo B en el cual eliminamos los efectos del factor A, utilizamos la restricción  $\beta_1 = 0$ , seleccionamos solo la primera y cuarta columna, de esa forma obtenemos:

$$\beta = \begin{pmatrix} \mu \\ \beta_2 \end{pmatrix}, X^T X = \begin{pmatrix} 12 & 6 \\ 6 & 6 \end{pmatrix}, X^T y = \begin{pmatrix} 88.2 \\ 43.0 \end{pmatrix}$$

de la misma forma que antes obtenemos

$$b = \begin{pmatrix} 7.533 \\ -0.367 \end{pmatrix}$$

$$\text{y } b^T X^T y = 648.6733$$

Finalmente para el modelo con la media como Único efecto  $E(Y_{ijk}) = \mu$ , la estimación de  $b = [\mu] = 7.35$  y  $b^T X^T y = 648.27$

Los resultados se resumen en la tabla inferior. Los sub índices S, I, A, B y M se refieren al modelo saturado, modelos correspondientes a  $H_I$ ,  $H_A$ ,  $H_B$  y el de la media, respectivamente. La deviancia a escala spn los terminos  $\sigma^2 D = y^T y - b^T X^T y$ . Los grados de libertad, d.f, son dados por N menos el número de parametros en el modelo.

Modelo	d.f	$b^T X^T y$	Deviancia a escala
$E(Y_{jkl}) = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$	6	662.6200	$\sigma^2 D_S = 1.4800$
$E(Y_{jkl}) = \mu + \alpha_j + \beta_k$	8	661.4133	$\sigma^2 D_I = 2.6867$
$E(Y_{jkl}) = \mu + \alpha_j$	9	661.0100	$\sigma^2 D_B = 3.0900$
$E(Y_{jkl}) = \mu + \beta_k$	10	648.6733	$\sigma^2 D_A = 15.4267$
$E(Y_{jkl}) = \mu$	11	648.2700	$\sigma^2 D_M = 15.8300$

Para probar  $H_I$  asumimos que el modelo saturado es correcto y de esa forma  $D_S \sim X^2(6)$ . Si  $H_I$  también es correcta, entonces  $D_I \sim X^2(8)$  de esa forma  $D_I - D_S \sim X^2(2)$  y

$$F = \frac{D_I - D_S}{2} / \frac{D_S}{6} \sim F(2, 6)$$

El valor de

$$F = \frac{2.6867 - 1.48}{2\sigma^2} / \frac{1.48}{6\sigma^2} = 2.45$$

no es significativa por lo que los datos no proveen evidencia contra  $H_I$ . Como  $H_I$  no es rechazada procedemos a probar  $H_A$  y  $H_B$ . Para  $H_B$  consideraremos la diferencia entre el ajuste del modelo A y el modelo aditivo, comparandolo con el modelo saturado.

$$F = \frac{D_B - D_I}{1} / \frac{D_S}{6} = \frac{3.09 - 2.6867}{\sigma^2} / \frac{1.48}{6\sigma^2} = 1.63$$

Lo que resulta no significativo si lo comparamos con una distribución  $F(1,6)$ , lo que sugiere que no hay diferencias debido a los niveles del factor B. La prueba correspondiente para  $H_A$  arroja una  $F = 25.82$  lo que es significativo en comparación a una distribución  $F(2,6)$ .

Por lo tanto, concluimos que la respuesta media se ve afectada únicamente por las diferencias en los niveles del factor A.

## Anexo

```
#Matriz de datos y
y= matrix(c(6.8,6.6,5.3,6.1,7.5,7.4,7.2,6.5,7.8,9.1,8.8,9.1),nrow = 12,ncol=1)
yt=t(y)
#Matriz de variables X
X= matrix(c(1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1,0,0,0,
,0,0,0,0,0,0,0,0,1,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,0,0,
0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1),
          nrow=12,ncol=6)
Xt=t(X)
#Matriz Xty
B=Xt%*%y
#Matriz XtX
A=Xt%*%X
AA= solve(A)
#Matriz (XtX)-1Xty
beta=AA%*%B
#Error
error= t(beta)%*%B
```

Utilizamos los datos anteriores para los siguientes modelos

```
A2= A[1:4,1:4]
A2A=solve(A2)
B2= B[1:4]
beta2=A2A%*%B2

error2= t(beta2)%*%B2
```

```
A3= A[1:3,1:3]
A3A=solve(A3)
B3= B[1:3]
beta3=A3A%*%B3

error3= t(beta3)%*%B3
```

```
A4= A[-(2:3),-(2:3)]
A5= A4[-(3:4),-(3:4)]
A5A=solve(A5)
B4= B[-(2:3)]
B5= B4[-(3:4)]
beta4=A5A%*%B5

error4= t(beta4)%*%B5
```

## Bibliografia

- Dobson, Annette. (2002). An Introduction to Generalized Linear Models, Second Edition.
- <https://bookdown.org/yihui/rmarkdown/beamer-presentation.html>
- <https://www.statmethods.net/advstats/glm.html>
- <https://onlinecourses.science.psu.edu/stat504/node/216/>