Modelos Lineales y Lineales Generalizados

Manuel Gijón Agudo

Octubre 2018 -

Índice

1. Regresión lineal simple

Objetivo: Nuestro objetivo será siempre explicar el comportamiento de una variable aleatoria Y en función de unos ciertos valores $X_1, ..., X_p$.

Dado un $n \in \mathbb{Z}^+$ denominaremos Y_i a la muestra de Y obtenida cuando $X_j = x_{ij} \in \mathbb{R}\ \forall i, j$.

Definición: denominamos el **Modelo Lineal** como:

$$\boxed{\forall i, \quad Y_i = \beta_0 + x_{i1}\beta_1 + ... + x_{i(p-1)}\beta_{p-1} + e_i = \mu_i + e_i}$$

donde β_0 es denominado **intercepto** (*intercept*) y los términos e_i los **errores**.

Hipótesis:

- $\forall i \ \in \{1, 2, ..., n\},\ e_i \sim N(0, \sigma_i^2)$
- **Homeodasticidad** (*homeodasticity*): $\forall i \ \in \{1, 2, ..., n\},\ \sigma_i^2 = \sigma^2$
- $\forall i, y \ \in \{1, 2, ..., n\},\ i \neq j\ e_i$ es **independiente** de e_j
- Los valores de X son fijos o variables aleatorias **independientes** de los errores.

En **forma matricial** escribiremos el modelo de la siguiente manera:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & x_{23} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \cdots & x_{n(p-1)} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Siendo así, definimos $Y_{n \cdot 1} = (Y_1, Y_2, ..., Y_n)^t,\ X_{n \cdot p} = (x_{ij}),\ \beta_{p \cdot 1} = (\beta_0, \beta_1, ..., \beta_{p-1})^t, e_{n \cdot 1} = (e_1, e_2, ..., e_n)^t$ y escribimos el modelo como:

$$Y = X\beta + e \Longleftrightarrow \mu = E(Y|X) = X\beta$$
$$Y|X \sim N(X\beta, \sigma^2 \cdot Id_n)$$

De la última línea se desprende lo siguiente:

$$E((Y - X\beta)(Y - X\beta)^t) = E\Big(\begin{pmatrix} e_1^2 & e_1e_2 & e_1e_3 & \cdots & e_1e_n \\ e_2e_3 & e_2^2 & e_2e_3 & \cdots & e_2e_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_ne_1 & e_ne_2 & e_ne_3 & \cdots & e_n^2 \end{pmatrix} \Big) = \sigma^2 \cdot Id_n$$

Observación: Las variables $X_1, X_2, ..., X_n$ pueden ser función de otro conjunto de variables. Por ejemplo, podríamos tener un conjunto $\{Z_1, Z_2, ..., Z_m\}$ tales que, para $m \in \mathbb{N}$:

$$X_i = g_i(Z_1, Z_2, ..., Z_m), \quad i \in \{1, 2, ..., n\}$$

Ejemplos de modelos lineales:

- Comparamos la presión sanguínea (Y) en dos tipos de individuos, unos que han tomado cierta medicación y un grupo de control:

  $$Y_{ij} = \mu_i + e_{ij},\ \forall i \in \{1,2\},\ j \in \{1,2,...,n\}$$

  En forma matricial lo podemos expresar de la siguiente manera:

  $$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ \vdots \\ e_{2n_2} \end{pmatrix}$$

  Los modelos conocidos como **modelos de regresión** son un caso particular de modelos lineales en los que las covariables son continuas o discretas, en ningún caso categóricas.

- Estudiamos el nivel de un determinado químico en una planta (Y) en función de su presencia en el suelo (X).

  $$Y_i = \beta_1 + x_i\beta_2 + e_i,\ i = 1,2,...,n$$

  O en su forma matricial:

  $$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

  Si las covariables son todas categóricas estaremos entonces ante un modelo **ANOVA** (*Analysis of Covariance*).

- Queremos estudiar el nivel de un determinado medicamento (Y) en función de su dosis (X_1) y del género del paciente (X_2).

  $$Y_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + e_{ij},\ \imath \in \{1,2\},\ j \in \{1,2,...,n\}$$

  En forma matricial:

  $$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{12} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_1} & 0 & 0 \\ 0 & 0 & 1 & x_{21} \\ 0 & 0 & 1 & x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{2n_2} \end{pmatrix} \cdot \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \end{pmatrix}$$

Definimos el **Modelo Nulo** (*Null Model*) al más simple, el que tiene un único parámetro.

$$Y_i = \beta_0 + e_i, \quad i = 1, ..., n$$

Equivalentemente:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \left(\beta_0\right) + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Este modelo es equivalente a estudiar una muestra de una variable aleatoria.

Denominamos **Intercepto** (*intercept*) al elemento β_0. En general cosideraremos modelos con intercepto, lo que significa que el modelo nulo será un submodelo del susodicho.

Ejemplos de modelos no lineales:

- Queremos estudiar la producción de leche de unas vacas en Litros (Y_i) en función del número de días (x_i) que hace que nacieron.

$$Y_i = e^{\beta_0 + \beta_1 x_i + \log(x_i)} + e_i, \quad e_i \sim N(0, \sigma^2)$$

- Queremos estudiar la calidad de cierto material en función de su proovedor, de cada uno de ellos elegiremos aleatoriamente una muestra de tamaño b del material de las que nos quedaremos con n muestras seleccionadas aleatoriamente.

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + e_{(ij)k}$$

$$i = 1, ..., a, \quad j = 1, ..., b, \quad k = 1, ..., n, \quad e_{(ij)k} \sim N(0, \sigma^2)$$

1.1. Estimación de parámetros

1.1.1. Mínimos cuadrados

Sea $y = (y_1, y_2, ..., y_n)^t$ una realización muestral de Y y $\widehat{\beta}$ una estimación del parámetro β.

- **Estimación por mínimos cuadrados** (*Minimum least square*): consiste en minimizar

$$S(\beta) = ||y - \widehat{y}||_2^2 = \sum_{i=1}^{n} (y - \widehat{y})^2 = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right)^2$$

  donde $\widehat{y} = \widehat{\mu} = X\widehat{\beta}$ y la **solución** es:

$$\boxed{\widehat{\beta} = (X^t X)^{-1} X^t y}$$

  Siempre que $X^t X$ sea invertible.

- ***Weighted least squares***: consisete en minimizar

$$S(\beta) = \sum_{i=1}^{n} w_i (y - \widehat{y})^2 = \sum_{i=1}^{n} w_i \left(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j \right)^2$$

  donde $w_i^{-1} = Var(Y_i)$.

  La **solución** es (obviamente, siempre que X^tX sea invertible):

$$\boxed{\widehat{\beta} = (X^tV^{-1}X)^{-1}X^tV^{-1}y}$$

  donde $V = diag(w_i)$.

Recalquemos que en ningún caso necesitamos conocer la distribución de Y.

1.1.2. Máxima verosimilitud

El estimador es el siguiente:

$$L(\beta; y) = (\sigma\sqrt{2\pi})^{-n} e^{-\sum_{i=1}^{n} \frac{(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j)^2}{2\epsilon^2}}$$

Equivalentemente (tomando logaritmos):

$$l(\beta; y) = -n \log{(\sigma\sqrt{2\pi})} - \sum_{i=1}^{n} \frac{(y_i - \sum_{j=0}^{p-1} x_{ij}\beta_j)^2}{2\epsilon^2}$$

Ahora definimos el siguiente vector:

$$U_j = \frac{\partial l}{\partial \beta_j} = \frac{1}{\sigma^2}\Big(X^t(Y - X\beta)\Big)_j, \quad \forall j$$

$$U = (U_1, U_2, ..., U_{p-1})^t$$

que denominaremos **vector de puntuaciones** (*score vector*).

$$U_j = 0 \Longleftrightarrow X^tY = X^tX\beta$$

Luego si $rank(X^tX) = p$ entonces la solución, el **estimador por máxima verosilimitud** es:

$$\boxed{\widehat{\beta} = (X^tX)^{-1}X^tY}$$

1.2. Valores predichos y residuos

Sea $\widehat{Y} = X\widehat{\beta} = (\widehat{Y}_1, \widehat{Y}_2, ..., \widehat{Y}_n)^t$ nuestro vector de valores predichos.

Si $(y_1, y_2, ..., y_n)^t$ es una realización muestral de Y, definimos el ***raw resudual*** de la siguiente manera:

$$\text{raw residual} = y_i - \widehat{y}_i = \widehat{e}_i$$

Notemos que el vector $\widehat{e} = Y - X\widehat{\beta}$ es ortogonal a las columnas de la matriz X.

$$\begin{aligned}
X^t\widehat{e} &= X^t(Y - X\widehat{\beta}) \\
&= X^t(Y - X((X^tX)^{-1}X^tY)) \\
&= X^tY - X^tX((X^tX)^{-1}X^tY) = X^tY - X^tY = 0
\end{aligned}$$

1.3. Estimación de la varianza

1.3.1. Método de los momentos

Asumimos que $p = Rank(X^tX)$, lo que implica que el rango de X^tX es máximo, luego se cumple:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \sim \chi^2_{n-p}$$

Luego inmediatamente se desprende:

$$E\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2\right) = n - p \Longleftrightarrow E(S^2) = \sigma^2$$

donde:

$$S^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$$

Con lo que S^2 es un **estimador insesgado** de σ^2. Este estimador es conocido como **Error cuadrático medio** (*mean square error*).

1.3.2. Método de máxima verosimilitud

La función de verosimilitud es la siguiente:

$$l(\sigma^2; \mu) = -n\log(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma^2} + \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}(y_i - \mu_i)^2 = 0$$

hacer los cálculos aquí

$$\Rightarrow \widehat{\sigma}^2 = \left(1 - \frac{p}{n}\right)S^2$$

Observación: si n es grande, ambos estimadores son similares pero sin embargo para p y n pequeños ambos estimadores difieren mucho.

Ejemplo: Consideremos el modelo nulo $y_i = \beta_0 + e_i, \quad i = 1, ..., n$. Demostrar que:

$$\widehat{\beta_0} = \overline{y}$$

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

- Respuesta 1
- Respuesta 2

Ejemplo: Consideremos la regresión lineal simple $y_i = \beta_0 + x_i\beta_i + e_i, \quad i = 1, ..., n$ y demostremos lo siguiente:

$$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x}$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2} = \frac{Cov(X, Y)}{Var(X)} = r_{XY}\frac{S_Y}{S_X}$$

Antes de comenzar haremos dos observaciones:

- $(\overline{x}, \overline{y})$ pertenece a la recta de regresión.
- El **coeficiente de correlación** (r_{XY}) mide la relación lineal entre X e Y.

- Respuesta 1
- Respuesta 2

1.3.3. Interpretación de los parámetros

Asumimos el siguiente modelo:

$$\forall i, \quad Y_i = \beta_0 + x_{i1}\beta_1 + +x_{i2}\beta_2 + \cdots + +x_{ip-1}\beta_{p-1} + e_i = \mu_i + e_i$$

Siendo Y_i la variable respuesta bajo la siguiente condición $X_i = (x_{i1}, x_{i2}, ..., x_{ip-1})$ y siendo Y_i^* la respuesta bajo las condiciones $X_i^* = (x_{i1}, x_{i2}, ..., x_{ij} + 1, ..., x_{ip-1})$ se tiene que:

$$Y_i - Y_i^* = \widehat{\beta}_j$$

Entonces nos queda:

- $\widehat{\beta}_j$ es el cambio medio obtenido al incrementar en una unidad el valor de x_j y mantener el resto inamovibles.
- Si $\widehat{\beta}_0$ es la estimación del intercepto, podemos interpretarlo como la respuesta media en el origen.

La **desviación residual estándard** (*residual standard deviation*) es el error asociado a nuestras predicciones, el 95 % de nuestras predicciones tendrán el error en el siguiente intervalo:

$$\Big(-t_{n-p,\frac{\alpha}{2}}\widehat{\sigma}, t_{n-p,\frac{\alpha}{2}}\widehat{\sigma}\Big) \simeq (-1{,}95\widehat{\sigma}, 1, 95\widehat{\sigma})$$

Explicar el por qué

1.4. Inferencia en los parámetros del modelo

1.4.1. Distribución de $\widehat{\beta}$

Si β_0 es el parámetro real, sabemos que:

$$\boxed{\widehat{\beta}|X \sim N\Big(\beta_0, \sigma^2(X^tX)^{-1}\Big)}$$

Por ser una combinación lineal de variables aleatorias normales.

Explicar el por qué con más detalle

$$\begin{aligned} E(\widehat{\beta}|X) &= (X^tX)^{-1}X^tE(Y|X) \\ &= (X^tX)^{-1}X^tX\beta_0 = \beta_0 \end{aligned}$$

Sabemos también que:

$$\widehat{\beta} - \beta_0 = (X^tX)^{-1}X^t(Y - X\beta_0)$$

Luego queda:

$$\begin{aligned} E\Big((\widehat{\beta}|X - \beta_0)\cdot(\widehat{\beta}|X - \beta_0)^t\Big) &= (X^tX)^{-1}X^tE\Big((Y - X\beta_0)\cdot(Y - X\beta_0)^t\Big)X(X^tX) \\ &= \sigma^2(X^tX)^{-1} \end{aligned}$$

Observación: Las componentes de $\widehat{\beta}$ no son variables aleatorias independientes.

Denominamos **Matriz de información de Fisher** (*Fisher information matrix*) a la matriz $\mathcal{J} = E(UU^t)$.

Observemos que bajo la hipótesis de la normalidad ocurre lo siguiente:

$$\begin{aligned} \mathcal{J} = E(UU^t) &= E\Big(\frac{1}{\sigma^2}X^t(Y - X\beta)\cdot(Y - X\beta)^tX\frac{1}{\sigma^2}\Big) \\ &= \frac{1}{\sigma^2}X^tE\Big((Y - X\beta)\cdot(Y - X\beta)^t\Big)X\frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2}X^tX \end{aligned}$$

Observación: La matriz de información de Fisher es la inversa de la matriz de varianzas y covarianzas de $\widehat{\beta}$.

Ejemplo: Consideremos la regresión lineal simple $y_i = \beta_0 + x_i\beta_i + e_i, \quad i = 1, ..., n$ y demostremos que, para cada coeficiente, las desviaciones estandard de los estimadores son:

$$S_{\widehat{\beta}_0} = S\cdot\left(\frac{1}{n} + \frac{1}{\sum_{i=1}^n x_i^2 - n\overline{x}^2}\right)^{1/2}$$

$$S_{\widehat{\beta}_1} = S\cdot\frac{1}{\left(\sum_{i=1}^n x_i^2 - n\overline{x}^2\right)^{1/2}}$$

- Respuesta 1
- Respuesta 2

1.4.2. Inferencia

Sabemos que $\widehat{\beta}|X \sim N(\beta_0, \sigma^2(X^tX)^{-1})$, luego cada parámetro verifica lo siguiente:

$$\boxed{\widehat{\beta}_i|X \sim N\Big(\beta_{0i}, \sigma^2[(X^tX)^{-1}]_{ii}\Big)}$$

Para cada $a \in \mathbb{R}$ podemos hacer el siguiente test:

$$H_0 : \beta_{0i} = a$$
$$H_1 : \beta_{0i} \neq a$$

Para un nivel de significación α se tiene:

$$\frac{\widehat{\beta}_i - a}{\widehat{\sigma}\sqrt{[(X^tX)^{-1}]_{ii}}}$$

Luego **rechazaremos la hipótesis nula** si:

$$|\frac{\widehat{\beta}_i - a}{\widehat{\sigma}\sqrt{[(X^tX)^{-1}]_{ii}}}| \geq t_{n-p,\alpha/2}$$

En particular, estaremos interesados en el siguiente test:

$$H_0 : \beta_{0i} = 0$$
$$H_1 : \beta_{0i} \neq 0$$

No rechazar H_0 implica que la covariable X_i no tiene una influencia significativa en Y.

Los **Intervalos de confianza** para los parámetros con un nivel de significación α son los siguiente:

$$\boxed{\widehat{\beta}_i \pm t_{n-p,\frac{\alpha}{2}} \cdot \widehat{\sigma}\sqrt{[(X^tX)^{-1}]_{ii}}}$$

Importante: Si el intervalo contiene el valor cero, la correspondiente X_i **no es estadísticamente significativa**

1.4.3. Tabla ANOVA

Dado $Y_i - \overline{Y} = (Y_i - \widehat{Y}_i) + (\widehat{Y}_i - \overline{Y})$ y que $\sum_{i=1}^n (Y_i - \widehat{Y}_i) \cdot (\widehat{Y}_i - \overline{Y})$ se tiene lo siguiente:

$$\boxed{\underbrace{\sum_{i=1}^n (Y_i - \overline{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}_{RSS} + \underbrace{\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}_{RegSS}}$$

Incluir demostración aquí

Donde cada término es:

- **TSS**: Suma de cuadrados total (*total sum of squares*)
- **RSS**: Suma de cuadrados debida a los residuos (*residual sum of squares*)
- **RegSS**: Suma de cuadrados explicada por la regresión (*regression sum of squares*)

TABLA ANOVA				
Fuente	Grados de libertad.	SS	MSS	F
Regresión	$p-1$	$\sum_{i=1}^{n}(Y_i-\overline{Y})^2$	$RegSS/p$	F_0
Residuos	$n-p$	$\sum_{i=1}^{n}(Y_i-\widehat{Y}_i)^2$	$RSS/(n-p)$	
Total	$n-1$	$\sum_{i=1}^{n}(Y_i-\widehat{Y}_i)^2$		

Denominamos **tabla ANOVA** a la siguiente construcción:

Donde $F_0 = \frac{RegSS/p}{RSS/(n-p)}$ y $\widehat{\sigma} = S^2 = RSS/(n-p)$.

Ejemplo: INCLUIREMOS AQUÍ UN EJEMPLO COMPLETO DE TODO ESTO

1.4.4. Test OMNIBUS

Si β_1 se corresponde con el intercepto, el **test OMNIBUS** se define como:

$$
\begin{aligned}
&H_0 : \beta_2 = \beta_3 = \cdots = \beta_p = 0 \\
&H_1 : \exists i : \beta_i \neq 0
\end{aligned}
$$

Observación: Este test compara un modelo con el modelo nulo.

Este test se efectúa a partir de la tabla ANOVA de la siguiente manera, si H_0 es cierto:

$$F_0 = \frac{RegSS/p}{RSS/(n-p)} \sim F_{p,n-p}$$

Lo que implica que **rechazamos la hipótesis nula** bajo un nivel de significación α si:

$$F_0 \geq F_{\frac{\alpha}{2},p,n-p}$$

Observación: El test OMNIBUS no es equivalente a comprobar si cada parámetro es equivalente a cero.

1.5. Inferencia en los valores preedichos

1.5.1. Distribución de los valores

Definimos el **vector de valores preedchos** como $\widehat{Y} = X\widehat{\beta}$.

$$\widehat{Y}|X \sim N\Big(X\beta_0, \sigma^2 X(X^tX)^{-1}X^t\Big)$$

Por ser combinación lineal de distribuciones normales se tiene:

$$E(\widehat{Y}|X) = XE(\widehat{\beta}|X) = X\beta_0$$

y también

$$\begin{aligned} E\Big((\widehat{Y}|X - X\beta_0) \cdot (\widehat{Y}|X - X\beta_0)^t \; Big) &= XE\Big((\widehat{\beta}|X - \beta_0) \cdot (\widehat{\beta}|X - \beta_0)^t\Big)X^t \\ &= X\sigma^2 (X^tX)^{-1}X^t \\ &= \sigma^2 X(X^tX)^{-1}X^t \end{aligned}$$

Denominamos a la matriz $X(X^tX)^{-1}X^t$ ***hat matrix*** debido a que:

$$\widehat{Y} = X(X^tX)^{-1}X^tY$$

Observación: σ^2 veces la *hat matrix* es la matriz de varianzas y covarianzas de las predicciones.

1.5.2. Intervalo de predicción

1.5.3. Intervalo de confianza

Llegados a este punto es importante hablar de los diferentes tipos de modelos lineales con los que nos encontraremos en función del tipo de dato que sean las variables explicativas.

2. ANOVA

En este caso alguna de las covariables si no todas serán **factores** (también llamadas **variables categóricas**). Estos modelos se denominan **análisis de la varianza (ANOVA)**.

2.0.1. Ejemplos

2.0.2. Ejercicios

2.1. ANCOVA

En el caso de que los coeficientes

2.1.1. Ejemplos

Example 2.1 *Tenemos datos sobre coche que utilizan diesel o no.*

2.1.2. Ejercicios

3. Regresión lineal generalizada

3.1. Genealidades

3.2. Binomial Response Models

Una variable aleatorio es tal que $Y \sim B(p)$ (Bernulli), $0 \leq p \leq 1$, si y solo sí toma valores 1 ó 0 con las siguientes probabilidades:

$$P(Y = 1) = p \text{ and } P(Y = 0) = 1 - p$$

Una variable aleatoria es tal que $Y \sim Bin(n, p)$ (Binomial), con parámetros.... CONTINUAR AQUÍ

BALA BLA BLA BLA

Definimos los **odds** de una variable aleatoria Binomial como Odd $= \frac{p}{1-p} \in (0, +\infty)$, tal que verifica:

$$\text{Odd} = \begin{cases} 5 & si \quad x \leq 2 \\ x^2 - 6x + 10 & si \quad 2 < x < 5 \\ 4x - 15 & si \quad x \geq 5 \end{cases}$$

SUSTITUIR APROPIADAMENTE LA MIERDA DE AQUÍ ARRIBA IMPORTANTE!!

Para comparar p_1 con $p_2 \in (-1, 1)$ CONTINUAR A1UÍ BLA BLA BLA BLA

$$\begin{cases} H_0 : p_1 = p_2 & \Longleftrightarrow H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 \neq p_2 & \Longleftrightarrow H_1 : p_1 - p_2 \neq 0 \end{cases}$$

$$Y_i \sim Bin(m_i, p_i)$$

$$g(\mu) = X\beta \Leftrightarrow g(mp) = X\beta$$

Recordemos el link canónico, el parámetro de ddispersione y la función de varianza son respectivamente:

$$\theta_i = \log\left(\frac{p_i}{1 - p_i}\right)$$

$$\Phi = 1$$

$$V(\mu_i) = \mu_i\Big(1 - \frac{\mu_i}{m_i}\Big)$$

3.3. Poisson Response Models

La principal característica de estos modelos es que la variable respuesta sigue una distribución de Poisson.

$$Y_i \sim Poiss(\mu_i)$$

$$g(\mu) = X\beta$$

Donde recordemos: $Poiss(Y = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad , E(Y) = \lambda, \quad , V(Y) = \lambda.$

El **índice de dispersión** será:

$$I(Y) = \frac{V(Y)}{E(Y)} = 1$$

—- EXPLICAR MEJOR ESTA MIERDA Ejemplo de las placas de petri overdispersión: tendencia al agrupamiento Poisson: randomness Underdisperssion: uno por casilla —- Estimamos el parámetro $\Phi = 1$

- $\widehat{\Phi} = \frac{X^2}{n-p} >> 1 \Rightarrow$ overdispersión
- $<< 1 \Rightarrow$ underdispersión
- ≈ 1 Poisson

In general $V(Y) = \Phi \cdot \mu$ quasi posisson lo usamos en overdispersion

Estimamos Φ con los datos and like this we will have a more accurate estimation of the true variance.

Una vez tengamos una tabla, si fijamos el parámetro n, lo que nos encontramos es una distribución multinomial. Si no lo fijamos seguimos contando con la distribucuín de poisson

Ejemplo de la tabla $\log(\mu_{ij}) = \beta_0 + \beta_1\text{Factor } 1 + \beta_2\text{Factor } 2$, en total $1 + a - 1 + b - 1 = a + b - 1$. Tenemos factores, luego puede haber interacciónes, sea pues el siguiente modelo a considerar:

$$\beta_0 + \beta_1\text{Factor } 1 + \beta_2\text{Factor } 2 + \beta_3\text{Interacción}$$

con un total de $a + b - 1 + (a-1)(b-1) = RELLENAR = ab$ luego se corresponde con el modelo completo (*full model*), lo que no tiene sentido considerar.