16.ANCOVA

Estadística
Grau en Matemàtiques

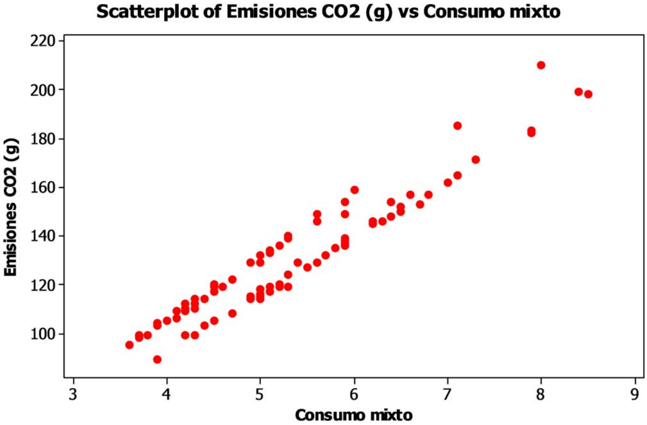
Josep A. Sanchez
Dept. Estadística i I.O.(UPC)

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

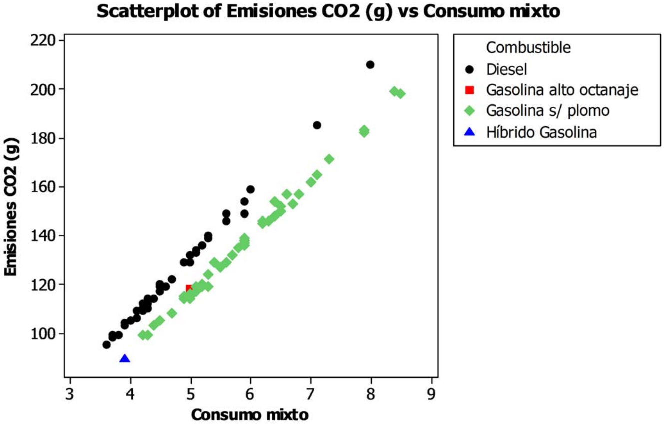
Exemple: Considerem que estem interessats en analitzar el CO_2 en funció del consum en ciutat i carretera segons tipus de combustible que es fa servir (diesel o benzina).

- Resposta: Emissions de CO_2 (v. numèrica)
- Explicativa 1: Consum mixt (v. numèrica)
- Explicativa 2: Tipus de combustible (v. categòrica)

ANCOVA: Exemple



ANCOVA: Exemple



ANCOVA

Considerem el cas en que hi ha una variable explicativa de tipus numèric (covariable) i una de tipus categòric amb 2 nivells (factor)

Un possible model seria:

$$Y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \epsilon_{ij}, \qquad i \in \{1,2\}, j \in \{1,2,\cdots,n_i\}$$

en forma matricial

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

ANCOVA

Un altre model

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + \beta_3 \mathbb{I}_{\{i=2\}} + \epsilon_{ij}, \qquad i \in \{1,2\}, j \in \{1,2,\cdots,n_i\}$$

en forma matricial,

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 \\ 1 & x_{12} & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{1n_1} & 0 \\ 1 & x_{21} & 1 \\ 1 & x_{22} & 1 \\ \vdots & \vdots & \vdots \\ 1 & x_{2n_2} & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2n_2} \end{pmatrix}$$

Aquet model assumeix que l'*intercept* pot ser diferent en els dos nivells, però la pendent és la mateixa

ANCOVA. Exemple

Equació de Regressió:

$$CO2 = 8.11 + 24.4\,Consum - 14.9\,Combustible$$

Combustible=0 si Diesel, Combustible=1 si Benzina (Combustible=$\mathbb{I}_{\{Benzina\}}$)

Un cop verificada la significació dels coeficients del model,l'equació del model ve expressada per dues rectes paral · leles:

$$\begin{cases} CO2 = 8.11 + 24.4\,Consum & \text{si Combustible=Diesel} \\ CO2 = -6,79 + 24,4\,Consum & \text{si Combustible=Benzina} \end{cases}$$

ANCOVA. Exemple

Un altre model possible:

$$Y_{ij} = \beta_1 + \beta_2 x_{ij} + \beta_3\, \mathbb{I}_{\{i=2\}} + \beta_4 \mathbb{I}_{\{i=2\}}\, x_{ij} + \epsilon_{ij}, \qquad i \in \{1,2\},\, j \in \{1,2,\cdots,n_i\}$$

en forma matricial

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{12} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n_1} & 0 & 0 \\ 1 & x_{21} & 1 & x_{21} \\ 1 & x_{22} & 1 & x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n_2} & 1 & x_{2n_2} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_2 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2n_2} \end{pmatrix}$$

Aquest model assumeix que tant l'*intercept* com la pendent són diferents en els dos nivells del factor.

ANCOVA. Exemple

Equació de Regressió:

$$CO2 = 0.82 + 26.0 Consum - 2.56 Combustible - 2.45 Consum * Combustible$$

Un cop verificada la significació dels coeficients del model,l'equació del model ve expressada per dues rectes paral · leles:

$$\begin{cases} CO2 = 0.82 + 26.0\ Consum & \text{si Combustible=Diesel} \\ CO2 = -1.74 + 23.55\ Consum & \text{si Combustible=Benzina} \end{cases}$$

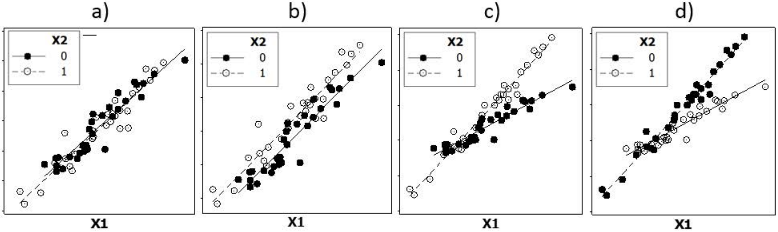
Un altre exemple:

Indica quina de les següents situacions representa el model ajustat:

```
The regression equation is
Y = 2,58 + 1,01 X1 + 5,16 X2 + 0,0318 X1*X2
```

Predictor	Coef	SE Coef	T	P
Constant	2,5828	0,7518	3,44	0,001
X1	1,00993	0,06676	15,13	0,000
X2	5,163	1,061	4,87	0,000
X1*X2	0,03177	0,08988	0,35	0,725

X1 es una variable quantitativa, X2 és qualitativa a dos nivells (0 i 1). L'eix vertical correspon a la resposta Y.



Variables explicatives del Model ANCOVA

- X: Covariable numèrica
- F: Factor categòric amb k nivells

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_i \mathbb{I}_{\{F=i\}} + \gamma_i \mathbb{I}_{\{F=i\}} x_{ij} + \epsilon_{ij}, \qquad i \in \{1,\ldots,k\},\ j \in \{1,2,\cdots,n_i\}$$

Suposem restriccions (tipus *treatment*): $\alpha_1 = 0 \quad \gamma_1 = 0$

Paràmetres:

- β_0: *Intercept* del model per al grup *baseline*
- β_1: Pendent del model per al grup *baseline*
- $\alpha_i \quad i = 2,\ldots,k$: Canvi en l'*intercept* si enlloc de ser del grup 1, és del grup i-éssim
- $\gamma_i \quad i = 2,\ldots,k$: Canvi en la pendent si enlloc de ser del grup 1, és del grup i-éssim

ANCOVA. Forma Matricial

Contrast *treatment*: $\alpha_1 = 0 \quad \gamma_1 = 0$

$$
\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix}
=
\begin{pmatrix}
1 & x_{11} & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
1 & x_{1n_1} & 0 & \dots & 0 & 0 & \dots & 0 \\
1 & x_{21} & 1 & \dots & 0 & x_{21} & \dots & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
1 & x_{2n_2} & 1 & \dots & 0 & x_{2n_2} & \dots & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
1 & x_{k1} & 0 & \dots & 1 & 0 & \dots & x_{k1} \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
1 & x_{kn_k} & 0 & \dots & 1 & 0 & \dots & x_{kn_k}
\end{pmatrix}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \\ \gamma_2 \\ \vdots \\ \gamma_k \end{pmatrix}
+
\begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{k1} \\ \vdots \\ \epsilon_{kn_k} \end{pmatrix}
$$

ANCOVA. Forma Matricial

Contrast *sum*:

$$\sum_{i=1}^{n} \alpha_i = 0 \quad \sum_{i=1}^{n} \gamma_i = 0$$

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{i1}* \\ \vdots \\ Y_{in_i} \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 1 & \dots & 0 & x_{11} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{1n_1} & 1 & \dots & 0 & x_{1n_1} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{i1} & 0 & \dots & 1 & 0 & \dots & x_{i1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{in_i} & 0 & \dots & 1 & 0 & \dots & x_{in_i} \\ 1 & x_{k1} & -1 & \dots & -1 & -x_{k1} & \dots & -x_{k1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{kn_k} & -1 & \dots & -1 & -x_{kn_k} & \dots & -x_{kn_k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \alpha_1 \\ \vdots \\ \alpha_{k-1} \\ \gamma_1 \\ \vdots \\ \gamma_{k-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2n_2} \\ \vdots \\ \epsilon_{k1} \\ \vdots \\ \epsilon_{kn_k} \end{pmatrix}$$

*i=k-1

ANCOVA. Models per categoria

Contrast *treatment*: $\alpha_1 = 0 \quad \gamma_1 = 0$

Factor	Model
1	$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$
2	$Y_{ij} = (\beta_0 + \alpha_2) + (\beta_1 + \gamma_2) x_{ij} + \epsilon_{ij}$
$\vdots$	$\vdots$
k	$Y_{ij} = (\beta_0 + \alpha_k) + (\beta_1 + \gamma_k) x_{ij} + \epsilon_{ij}$

ANCOVA. Models per categoria

Contrast *sum*: $\sum_{i=1}^{n} \alpha_i = 0 \quad \sum_{i=1}^{n} \gamma_i = 0$

$$\alpha_k = -\alpha_1 - \cdots - \alpha_{k-1} \qquad \gamma_k = -\gamma_1 - \cdots - \gamma_{k-1}$$

Factor	Model
1	$Y_{ij} = (\beta_0 + \alpha_1) + (\beta_1 + \gamma_1)x_{ij} + \epsilon_{ij}$
2	$Y_{ij} = (\beta_0 + \alpha_2) + (\beta_1 + \gamma_2)x_{ij} + \epsilon_{ij}$
$\vdots$	$\vdots$
k	$Y_{ij} = (\beta_0 - \alpha_1 - \cdots - \alpha_{k-1}) + (\beta_1 - \gamma_1 - \cdots - \gamma_{k-1})x_{ij} + \epsilon_{ij}$

El model $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$ fa referència a un model mig amb totes les dades, com si no hi haguessin diferències per grup

ANCOVA. Inferència

- Test de significació de la covariable X:

  Si no es rebutja $H_0 : \beta_1 = 0$

  "La pendent de la recta del grup de referència no és significativa i per tant no hi ha relació entre resposta i covariable en aquest grupo"

- Test de significació del factor F:

  Si no es rebutja $H_0 : \alpha_1 = \cdots = \alpha_k = 0$

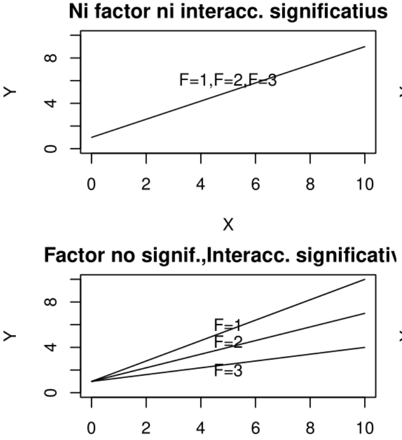
  "Els termes independients de les rectes de cada grup no difereixen significativament. Tenen ordenada a l'origen comuna"

- Test de significació de la interacció:

  Si no es rebutja $H_0 : \gamma_1 = \cdots = \gamma_k = 0$

  "Les pendents de les rectes de cada grup no difereixen significativament. Es a dir, són rectes paral · leles"

ANCOVA. Configuracions



En els dos darrers casos, per comparar els tractaments (factor F) s'ha d'especificar en quin valor de la covariable X.