

Linear regression (chap 6, 7, 8, 9, 10)

simple

Recall (L1)

Response: y

← what we want to model/predict.

Covariates: $x_1, \dots, x_k$ $\vec{x}$

Covariates/
explanatory
variables...

Data consists of n observations

$(y_1, \vec{x}_1), \dots, (y_n, \vec{x}_n)$

Fixed- x assumption: our n observations are

realizations of $(y_1, \vec{x}_1), \dots, (y_n, \vec{x}_n)$

↑ random ↑ fixed

→ Repeating the 'experiment' means
using same x -s but getting new y -s.
Chap. 10: Random- x assumption

Model for observation i ($i=1, \dots, n$)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

↑
random
response

fixed, linear

↑
random
noise

Assumptions

$$1) E(Y_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

ie, fixed part is a model for the expectation of the response

$$E(\varepsilon_i) = 0$$

$$2) \text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2 \leftarrow \begin{array}{l} \text{not related} \\ \text{to } x\text{'s} \end{array}$$

homoskedasticity, constant variance

$$3) \text{Cov}(Y_i, Y_j) = \text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

$$4)^* \quad \varepsilon_i \sim N(0, \sigma^2)$$

not always $Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \sigma^2 I)$

Then, by 3), Y_i, Y_j are independent.

Matrix form

$$\vec{Y} = X \vec{\beta} + \vec{\varepsilon}$$

↑
random
response
vector

↑
design
matrix
 $n \times (k+1)$

← coefficients $\beta_0, \beta_1, \dots, \beta_k$

← random error vector

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}$$

Assumptions (again)

$$1) E(\vec{Y}) = X\vec{\beta}$$

$$2,3) \text{Cov}(\vec{Y}) = \text{Cov}(\vec{\varepsilon}) = \sigma^2 I$$

$$4)^* \vec{Y} \sim \text{MVN}_n(X\vec{\beta}, \sigma^2 I)$$

$$\vec{\varepsilon} \sim \text{MVN}_n(\vec{0}, \sigma^2 I)$$

[Discuss, from chap 1, why these models?]

Using data, we shall

* Estimate $\vec{\beta}$ (and σ^2)
"Fit model"

OK without assumptions

* Test hypotheses and
obtain confidence regions for $\vec{\beta}$

requires assumptions.

* Predict new y-s, with uncertainty

Example, simple linear regression

We forgot π

But we know, for spheres, $V = \frac{4}{3}\pi r^3$

$\downarrow$ volume $\downarrow$ radius

We have 7 spheres with known radii (cm)

Experiment: for each sphere (known r_i),
measure volume (y_i).

Model for measurement i :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \leftarrow \text{random error}$$

$\uparrow$ $\uparrow$ $\uparrow$

possible systematic error π $\frac{4}{3}r_i^3$

Design matrix

Design matrix for $n=7$ experiments

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_7 \end{bmatrix}$$

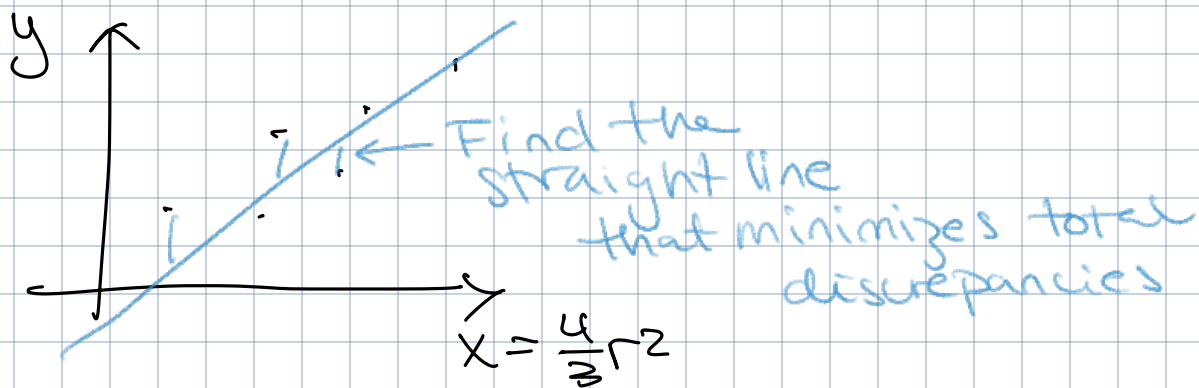
$$x_i = \frac{4}{3}r_i^3$$

$$y_1 = 13.8 \text{ cm}^3$$

$$y_2 = 181.6 \text{ cm}^3$$

Model $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$ observations: $\vec{y}, X$.

Goal: Estimate $\vec{\beta}$ using $n=7$ observations.



Find $\hat{\beta}_0, \hat{\beta}_1$ that minimize

$$\sum_{i=1}^n (y_i - \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_i)}_{\hat{y}_i = E(\hat{Y}_i) \text{ "prediction"}})^2 = SSE \quad (= RSS)$$

$y_i - \hat{y}_i = \hat{\varepsilon}_i$ "residual" estimate of error

SSE = sum of squared estimate of error

RSS = residual sum of squares

Solve $\frac{\partial SSE}{\partial \hat{\beta}_0} = 0$, $\frac{\partial SSE}{\partial \hat{\beta}_1} = 0$ simultaneously

Solution: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}$

↑ this you have seen before

cumbersome if we have multiple covariates.

Try matrix formulation

$\vec{y}$: n-vector observations

$\hat{\vec{y}} = X\hat{\beta}$ n-vector of predictions $(\hat{\vec{y}})$

$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \quad (\hat{\vec{\beta}})$

$\hat{\vec{\epsilon}} = \vec{y} - \hat{\vec{y}}$ n-vector of residuals

$$SSE = \hat{\vec{\epsilon}}^T \hat{\vec{\epsilon}} = (\vec{y} - \hat{\vec{y}})^T (\vec{y} - \hat{\vec{y}})$$

$$= \vec{y}^T \vec{y} - \underbrace{\vec{y}^T \hat{\vec{y}} + \hat{\vec{y}}^T \vec{y}}_{-2\vec{y}^T \hat{\vec{y}}} + \hat{\vec{y}}^T \hat{\vec{y}}$$

$$= \vec{y}^T \vec{y} - 2 \underbrace{\vec{y}^T X}_{\vec{a}^T} \hat{\beta} + \hat{\beta}^T (X^T X) \hat{\beta}$$

$$\frac{\partial SSE}{\partial \hat{\beta}} = 0 - 2 \underbrace{X^T \vec{y}}_{\text{thm 2.14a}} + 2 \underbrace{(X^T X) \hat{\beta}}_{\text{thm 2.14b}}$$

Solve $(X^T X) \hat{\beta} = X^T \vec{y}$

"normal equations"

$$\hat{\beta} = (X^T X)^{-1} X^T \vec{y}$$



Example (cont.)

$$\hat{\beta} = \begin{pmatrix} -0.42 \\ 3.14 \end{pmatrix}$$

$$\begin{aligned}\hat{y}_i &= -0.42 + 3.14 x_i \\ &= -0.42 + 3.14 \cdot \frac{4}{3} x_i^3\end{aligned}$$

Theorem 7.3a

If $\vec{y} = X\vec{\beta} + \vec{\epsilon}$, where X is $n \times (k+1)$ and $\text{rank}(X) = k+1 < n$, then the "value" of $\hat{\beta}$ that minimizes

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{is} \quad \hat{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$\uparrow$ $(k+1)$ -vector

NB : For observed data $\vec{y}, X$,

$\hat{\beta}$ is a vector of numerical values (estimates)

e.g. $\hat{\beta} = \begin{pmatrix} -0.42 \\ 3.14 \end{pmatrix}$.

For random response $\vec{y}$ (fixed X)

$\hat{\beta}$ is a random vector (estimators)

will take new values every time the experiment is repeated.

NB 2 : $\hat{\vec{y}}$: n -vector of predictions

$$\hat{\vec{y}} = X \hat{\beta} = X \underbrace{(X^T X)^{-1} X^T}_{\text{estimate}} \vec{y} \quad \text{Observations}$$

$[R]$

$\vec{y}$: random n -vector

$$\hat{\vec{y}} = X \hat{\beta} = X \underbrace{(X^T X)^{-1} X^T}_{\text{Estimator}} \vec{y} \quad \text{random responses}$$

$$H = X (X^T X)^{-1} X^T$$

'Hat matrix' $\rightarrow$ puts the hat on y .

$$H^T = H$$

$$H \cdot H = H$$

}

Symmetric

Idempotent

$$\text{rank}(H) = \text{tr}(H) =$$

$$\text{tr} \left(X (X^T X)^{-1} X^T \right) =$$

$$\text{tr} \left(X^T X (X^T X)^{-1} \right) =$$

$$\text{tr} \left(I_{(k+1)} \right) = k+1.$$

$$\text{tr}(AB) = \text{tr}(BA).$$