

Linear Regression Models

P8111

Lecture 12

Jeff Goldsmith
March 1, 2016




THE DEPARTMENT OF
BIostatISTICS

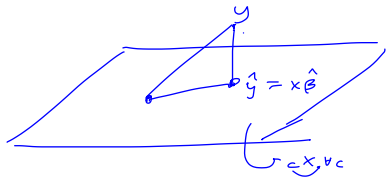


Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Today's Lecture

- 
- Gauss-Markov theorem
 - Maximum likelihood inference
 - Regression diagnostics

What's so great about LSEs?



$$LSE = \underbrace{(X^T X)^{-1}}_A \underbrace{X^T y}_{GM}$$

- Nice projection-space interpretation
- They're the "best" linear unbiased estimators
- They're maximum likelihood estimators under Normally-distributed errors

$$e \sim N(0, \sigma^2 I)$$

Gauss-Markov theorem

Assume the model

$$\underline{y} = \underline{X}\beta + \epsilon$$

where $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I$. Also assume \underline{X} is a full rank design matrix.

- Among all unbiased linear estimators $\underline{C}\underline{y}$ of the regression coefficients β , the LSE has minimum variance and is unique.

We call the LSEs “BLUE”.

Gauss-Markov theorem – proof

$$\text{unbiased: } Cy = \underbrace{(X^T X)^{-1} X^T + A} y$$

$$E(Cy) = \beta$$

$$\Rightarrow E((X^T X)^{-1} X^T + A)(X\beta + e)$$

$$\Rightarrow E(\underbrace{(X^T X)^{-1} X^T X \beta}_{\beta} + AX\beta + \underbrace{(X^T X)^{-1} X^T e}_{0} + Ae)$$

$$= \beta + AX\beta = \beta$$

$$\Rightarrow AX = 0$$

Gauss-Markov theorem – proof

$$\begin{aligned}
 & c^T y \\
 \text{Var}(c^T y) &= c^T \underbrace{\text{Var}(y)}_{\sigma^2 I} c^T \\
 &= \sigma^2 c^T c^T \\
 &= \sigma^2 \left((X^T X)^{-1} X^T + A \right) \left((X^T X)^{-1} X^T + A \right)^T \\
 &= \sigma^2 \left[\cancel{(X^T X)^{-1} X^T (X^T X)^{-1} X^T} + \cancel{(X^T X)^{-1} X^T A} \right. \\
 &\quad \left. + \cancel{A X (X^T X)^{-1} X^T} + A A^T \right] \quad \cancel{(A X)^T} \\
 &= \sigma^2 \left[(X^T X)^{-1} + A A^T \right] \quad \left| \begin{array}{l} x + y^2 \\ \geq x \end{array} \right. \\
 &\Rightarrow \underbrace{\sigma^2 (X^T X)^{-1}} \\
 &= \text{if } A=0
 \end{aligned}$$

Gauss-Markov theorem – caveats

The Gauss-Markov theorem is great, but notice the details:

- Assumed $Var(\epsilon) = \sigma^2 I$
- Only talking about unbiased linear estimators

Maximum likelihood estimation

Continue assuming the model

$$\underline{y = X\beta + \epsilon}$$

where $E(\epsilon) = 0$ and $\text{Var}(\epsilon) = \sigma^2 I$.

- Additionally, assume $\epsilon \sim N(0, \sigma^2 I)$
- Put differently, we're imposing the model

$$\underline{y \sim N(X\beta, \sigma^2 I)}$$

- y is multivariate Normal with uncorrelated entries; the y_i are each independently Normally distributed

$$y_i \sim N(x_i \beta, \sigma^2)$$

$$\begin{array}{l} P(\text{data} | \text{param}) \\ P(y | \beta) \\ \downarrow \\ L(\beta; y) \end{array}$$

Maximum likelihood estimation

Using independently Normal y_i 's:

$$L(\beta; y) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i\beta)^2\right\}$$
$$(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum (y_i - x_i\beta)^2\right\}$$

$$l(\beta; y) = \log(\quad) + \underbrace{-\frac{1}{2\sigma^2} \sum (y_i - x_i\beta)^2}$$

$$\frac{\partial l}{\partial \beta} \propto \frac{\partial}{\partial \beta} \text{RSS}$$

Maximum likelihood estimation

Using matrix notation:

$$L(\beta; y) = \underbrace{(2\pi)^{-n/2} (c^{-2}I)^{-1/2}}_{\text{constant}} \exp\left\{-\frac{1}{2} \underbrace{(y - X\beta)^T (c^{-2}I)^{-1} (y - X\beta)}_{\substack{\downarrow \\ \frac{-1}{2\sigma^2} (y - X\beta)^T (y - X\beta) \\ \text{RSS}(\beta)}}}\right\}$$

$$\underbrace{(X^T X)^{-1} X^T y}_{\text{MLE of } \beta}$$

Regression diagnostics

$$y = X\beta + \epsilon$$
$$\epsilon \sim (0, \sigma^2 I)$$

- Regression diagnostics are tools used to determine whether a given model is consistent with the data
- Usually focus on residuals
- Recall that fitted values are given by $\hat{y} = Hy$ where H is the hat matrix

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

- Residuals are defined as $y - \hat{y} = (I - H)y$

$$\hat{\epsilon} = (I - H)y$$

$\hat{\epsilon}$ and ϵ

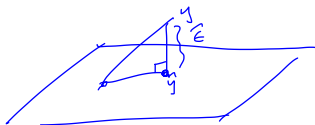
- $E(\hat{\epsilon}) = 0$
 $E((I-H)y)$
- $Var(\hat{\epsilon}) = \frac{\sigma^2(I-H)}{\dots}$
- Residuals are mean zero, but don't have constant variance nor are they uncorrelated.

$\hat{\epsilon}$ and ϵ

$$\hat{y} = Hy$$

- $\hat{\epsilon} = (I - H)y = \dots = (I - H)\epsilon$
- If ϵ is Normally distributed, so are the residuals
- Also note residuals sum to zero

Residuals and fitted values

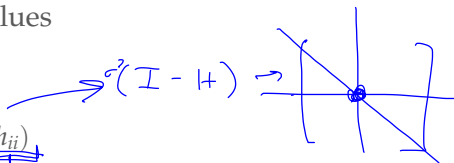


$$\begin{aligned}\underline{\text{Cov}(\hat{\boldsymbol{\epsilon}}^T, \hat{\mathbf{y}})} &= \text{Cov}((\mathbf{I} - \mathbf{H})\mathbf{y}, \mathbf{H}\mathbf{y}) \\ &= (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}\mathbf{H}^T \\ &= \sigma^2(\mathbf{H} - \mathbf{H}) \\ &= \mathbf{0}\end{aligned}$$

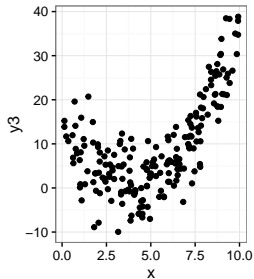
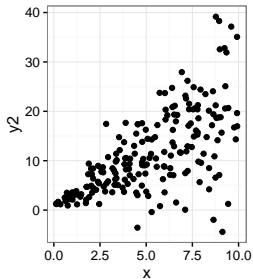
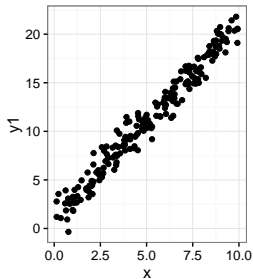
- So residuals and fitted values are uncorrelated

Residuals when model is correct

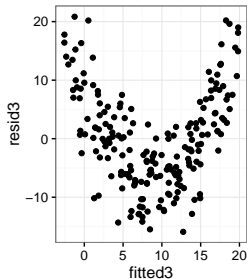
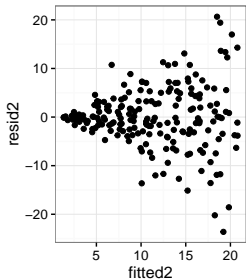
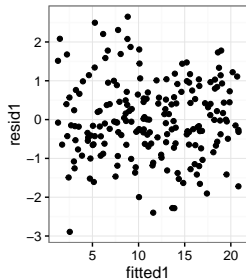
- Often we plot the residuals against one of the predictors or against the fitted values
- What we look for:
 - ✓ ▶ $E(\hat{\epsilon}|x) = 0$
 - ✓ ▶ $V(\hat{\epsilon}|x) = \sigma^2(1 - h_{ii})$
- If the model is incorrect, you may be able to spot:
 - ▶ Patterns in the residuals
 - ▶ Clear non-constant variance



Some data plots



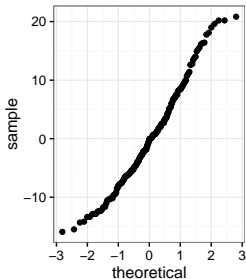
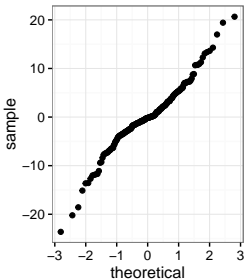
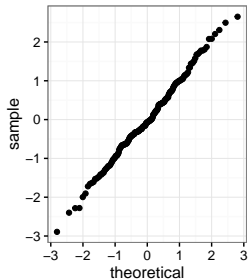
Some residual plots



Checking Normality assumption

- We often assume Normality for the errors
- Useful to check Normality of residuals
- Try a QQ plot:
 - ▶ Compute the sample quantiles of the residuals
 - ▶ Compute the quantiles of a standard Normal of size n
 - ▶ Plot these against each other
- Can also use the Shapiro-Wilk test based on correlation between observed and theoretical quantiles

Checking Normality assumption



Checking model structure

- You can plot residuals against each of the predictors, or plot outcomes against predictors
- Keep in mind the MLR uses adjusted relationships; scatterplots don't show that adjustment
- Adjusted variable plots (partial regression plots, added variable plots) can be useful

Adjusted variable plots

- Regress y on everything but x_j ; take residuals $r_{y|-x_j}$
- Regress x_j on everything but x_j ; take residuals $r_{x_j|-x_j}$
- Regress $r_{y|-x_j}$ on $r_{x_j|-x_j}$; slope of this line will match β_j in the full MLR
- Plot of $r_{y|-x_j}$ against $r_{x_j|-x_j}$ shows the “adjusted” relationship

What should you do ...

if your assumptions are violated?

- Depends on the assumption
- For problems with the errors, use LSE anyway; maybe use bootstrap for inference
- For non-linearity, try an augmented model

Today's big ideas

- Gauss-Markov, MLE, regression diagnostics
-

- Suggested reading: Faraway Ch 2.8, Ch 7