Some details to tidy up

ST552 Lecture 7

Charlotte Wickham

2019-01-23

Next week: No Trevor

- Charlotte will lead lab
- Thu 11am–12 Office Hour 255 Weniger

Summary of last week

For the linear regression model

$$Y = X\beta + \epsilon$$

where $\mathrm{E}(\epsilon) = 0_{n\times 1}$, $\mathrm{Var}(\epsilon) = \sigma^2 I_n$, and the matrix $X_{n\times p}$ is fixed with rank p.

The least squares estimates are

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

Furthermore, the least squares estimates are BLUE, and

$$\mathrm{E}\left(\hat{\beta}\right) = \beta, \qquad \mathrm{Var}\left(\hat{\beta}\right) = \sigma^2 (X^T X)^{-1}$$

We have not used any Normality assumptions to show these properties.

Today

- Verify:

$$\mathrm{E}\left(\hat{\sigma}^2\right) = \mathrm{E}\left(\frac{1}{n-p}\sum_{i=1}^{n} e_i^2\right) = \sigma^2$$

- Add Normal assumption to get inference on regression coefficents.

Go over the estimation of σ

Strategy: Write e_i^2 as a linear combination of uncorrelated variables, ϵ_i.

Write correlated residuals as combination of uncorrelated errors

Claim:

$$\underset{\text{residuals}}{||e||^2} = \epsilon^T(I - H)\underset{\text{errors}}{\epsilon}$$

Your turn at home:

1. Show $(I - H)\epsilon = e$. Hint: substitute $\epsilon = Y - X\beta$, expand and use properties of H.
2. Show $||e||^2 = e^T e = \epsilon^T(I - H)\epsilon$. Hint: substitute in $e = (I - H)\epsilon$ from above and use properties of $(I - H)$.

Find expected value of $||e||^2$ in terms of trace$(I - H)$

Show $\mathrm{E}\left(\overbrace{\epsilon^T(I-H)\epsilon}^{\|e\|^2}\right) = \sigma^2 \mathrm{trace}(I - H)$

Hint

$$\underbrace{\overset{\text{vector}}{x}{}^T \overset{\text{matrix}}{A} x}_{\text{quadratic form}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \underbrace{x_i x_j A_{ij}}_{\text{elements}}$$

where

$$x = (x_1, x_2, \ldots, x_n)^T, \quad A = \begin{pmatrix} A_{11} & A_{12} & \ldots \\ A_{21} & A_{22} & \ldots \\ \vdots & & \end{pmatrix}_{n \times n}$$

Find expected value of $||e||^2$ in terms of trace$(I - H)$

$$E\left(\epsilon^T(I-H)\epsilon\right) = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \varepsilon_i \cdot \varepsilon_j \cdot (I-H)_{ij}\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E(\varepsilon_i \varepsilon_j)(I-H)_{ij}$$

by linearity of expectation

$$= \sum_{i=1}^{n} \sigma^2 (I-H)_{ii}$$

$$= \sigma^2 \,\mathrm{trace}(I-H)$$

$i = j$ $\quad E(\varepsilon_i^2) = \sigma^2$

$i \neq j$ $\quad E(\varepsilon_i \varepsilon_j) = 0$

uncorrelated

Find trace($I - H$)

Show

$$\text{trace}(I - H) = n - p$$

Hint:

$$\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B) \qquad 1$$

$$\text{trace}(AB) = \text{trace}(BA)$$

$$\text{trace}(I - H) = \text{trace}(I_{n \times n}) - \text{trace}(H) \qquad \text{by } 1$$

$$= n - \text{trace}\left(X(X^T X)^{-1} X^T\right) \qquad \text{substitute in } H$$

$$= n - \text{trace}\left(\underbrace{(X^T X)^{-1}}_{p \times n \; n \times p} \underbrace{X^T X}_{p \times n \; n \times p}\right) \qquad X_{n \times p}$$

$$= n - \text{trace}(I_{p \times p})$$

$$= n - p$$

Put it all together

$$E\left(\hat{\sigma}^2\right) = E\left(\frac{1}{\boxed{n-p}} \|e\|^2\right) = \frac{1}{n-p}\sigma^2(n-p)$$

$$= \sigma^2, \quad \text{unbiased}$$

degrees of freedom

Exercises

See handout.

Exercises

ST552 Winter 2016

January 20, 2016

In a study of cheddar cheese from the LaTrobe Valley of Victoria, Australia, samples of cheese were analyzed for their chemical composition and were subjected to taste tests. Overall taste scores were obtained by combining the scores from several tasters.

cheddar is a data frame with 30 observations on the following 4 variables:

taste, a subjective taste score
Acetic, concentration of acetic acid (log scale)
H2S, concentration of hydrogen sulfice (log scale)
Lactic, concentration of lactic acid

The following model:

$$\textbf{Full Model:}\quad \text{taste}_i = \beta_0 + \beta_1 \text{Acetic}_i + \beta_2 \text{H2S}_i + \beta_3 \text{Lactic}_i + \epsilon_i$$

was fit in R and the output is shown below.

```
data(cheddar, package = "faraway")
fit <- lm(taste ~ . , data = cheddar)
summary(fit)
```

```
##
## Call:
## lm(formula = taste ~ ., data = cheddar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.390  -6.612  -1.009   4.908  25.449
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28.8768    19.7354  -1.463  0.15540
## Acetic        0.3277     4.4598   0.073  0.94198
## H2S           3.9118     1.2484   3.133  0.00425 **
## Lactic       19.6705     8.6291   2.280  0.03108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.13 on 26 degrees of freedom
## Multiple R-squared:  0.6518, Adjusted R-squared:  0.6116
## F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

$\hat{\sigma}$

1. Write down the form of the β vector and the $\hat{\beta}$ vector.

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} -28.8 \\ 0.33 \\ 3.91 \\ 19.7 \end{pmatrix}$$

2. What are the values of $\hat{\sigma}$, n, $\hat{\sigma}^2$, $\sum_{i=1}^{n} e_i^2$, σ?

$n = 30$ $\qquad\qquad$ σ = unknown!

$\hat{\sigma}^2 = (10.13)^2$

$\sum_{i=1}^{n} e_i^2 = 26\,(10.13)^2$ $\qquad\qquad$ $\hat{\varepsilon} = e$ ↑ residual

$$
\begin{array}{cc}
 & \begin{array}{cccc} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \end{array} \\
(X^TX)^{-1} = & \begin{pmatrix} 3.80 & -0.76 & 0.09 & -0.07 \\ -0.76 & \boxed{0.19} & \boxed{-0.02} & -0.13 \\ 0.09 & -0.02 & 0.02 & -0.05 \\ -0.07 & -0.13 & -0.05 & 0.73 \end{pmatrix}
\end{array}
$$

3. Verify the reported value for $SE(\hat{\beta}_1)$.

$SE(\hat{\beta}_1) = \hat{\sigma}\sqrt{(X^TX)^{-1}_{(1,1)}}$

start conting at zero

$= 10.13 \times \sqrt{0.19}$

4. What is the value of $\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$?

$= 10.13^2 \times -0.02$

Inference on the regression coefficients

Normality assumption

multivariate normal, dimension n

Assume $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 I)$. $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

univariate Normal

Important reminders:

- Linear combinations of Normal r.v's are also Normal

  $x_{n\times 1} \sim N(\mu, \Sigma)$ $A_{p\times n}$ fixed

  $Ax \sim N(A\mu, A\Sigma A^T)$

- Uncorrelated elements are independent

Leads to:

$$\underset{n\times 1}{Y} \sim N(X\beta, \sigma^2 I) \qquad Y = X\beta + \varepsilon$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

$$\hat{\beta} = \underbrace{(X^T X)^{-1} X^T}\, Y$$

Normality: $\hat{\beta}$ are also MLE of β (MLE of σ^2 is not $\hat{\sigma}^2$)

Inference on individual parameters

With the addition of the Normal assumption, it can be shown that

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim t_{n-p}$$

leads to the usual construction of tests and confidence intervals for single parameters.