Stat 135 Lab 12

GSI: Yutong Wang

April 24, 2020

Source: tomas del amo / shutterstock

Lanikai Beach

https://www.thecrazytourist.com/15-best-beaches-oahu/

To-do Today

1. Statistical properties of the regression coefficients
2. Estimating sigma squared
3. Practice problems

Recap of Linear Regression from last week

$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ where $\mathbf{Y} \in \mathbb{R}^n$, $\mathbf{X}$ is the design matrix.
β is an unknown constant vector, and $\mathbf{e}$ is the noise term.

- $n \geq k+1$, and the **design matrix X** spans a $k+1$ dimension subspace of $\mathbb{R}^n$.
- $\mathbf{X}$ is of full rank, i.e., the columns of $\mathbf{X}$ are independent.
- $e_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$.
- **Homoscedasticity**: Each y_i is of the same variance σ^2, and independent of $\mathbf{X}$.

$$\begin{bmatrix} y_1 \\ \dots \\ \dots \\ \dots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n\times(k+1)} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} e_1 \\ \dots \\ \dots \\ \dots \\ e_n \end{bmatrix}_{n\times 1}$$

Recap of Linear Regression from last week

How to find $\hat{\beta}_0$ and $\hat{\beta}_1$?
The optimal fitting line is found by minimizing the residual sum of squares (RSS) which is

$$\text{RSS} = \|\hat{\mathbf{e}}\|_2^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 = \|\mathbf{y} - \mathbf{X}\hat{\beta}\|_2^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

To minimize RSS, we can

- either use least squares (set the partial derivatives of $\|\mathbf{y} - \mathbf{X}\hat{\beta}\|$ w.r.t. β_0, β_1 equal to 0),
- or apply the orthogonal projection $(\mathbf{X}'_{(k+1)\times n} \cdot (\mathbf{y} - \mathbf{X}\hat{\beta})_{n\times 1} = \mathbf{0}_{(k+1)\times 1})$

$$\Longrightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

In simple linear regression case, we have that

$$\hat{\beta}_1 = \frac{\text{Cov}(\mathbf{x}, \mathbf{y})}{\text{Var}(\mathbf{x})} = \frac{\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}})}{\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})^2}$$

Berkeley
UNIVERSITY OF CALIFORNIA

Review: Statistical properties of regression coefficients

Unbiasness of $\hat{\beta}$:

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\mathbf{X}\beta + \epsilon] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta\end{aligned}$$

Variance of $\hat{\beta}$: Recall $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$,

$$\begin{aligned}\hat{\beta} - \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\ \mathrm{Var}[\hat{\beta}] &= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\epsilon\epsilon']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

In simple linear regression, you may compute the result in matrix form as well and derive

$$\mathrm{Var}[\hat{\beta}_1] = \frac{\sigma^2}{n\mathrm{Var}(X)}, \quad \mathrm{Var}[\hat{\beta}_0] = \sigma^2[\frac{1}{n} + \frac{(\bar{X})^2}{n\mathrm{Var}(X)}]$$

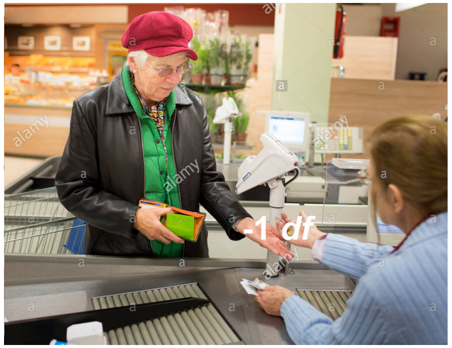
$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\sigma^2\bar{X}}{n\mathrm{Var}(X)}$$

Review: Estimating sigma squared

To measure the variance of $\hat{\beta}$, we need to estimate σ^2 which is the variance of the noise term. The fact is that,

$$\frac{\sum e_i^2}{\sigma^2} \sim \chi_n^2, \text{ because } e_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

me: i'd like to buy an additional parameter
model: that'd be one degree of freedom please
me: okay here you go
model: thank you



Thus,

$$\frac{\text{RSS}}{\sigma^2} = \frac{\sum \hat{e}_i^2}{\sigma^2} \sim \chi_{n-2}^2$$

The detailed proof can be found in lecture slides (Lec 35).
The intuition is that we have estimated $\hat{\beta}_0$ and $\hat{\beta}_1$,
thus the two estimated parameters "consumed" 2 degrees of freedom in RSS.
As $\mathbb{E}[\chi_n^2] = n, \text{Var}(\chi_n^2) = 2n$, we know

$$\mathbb{E}[\frac{\text{RSS}}{\sigma^2}] = n - 2$$

$$\implies \mathbb{E}[\frac{\text{RSS}}{n-2}] = \sigma^2$$

Q1: Simple Linear Regression (R)

Rice 14.9.37: Dissociation pressure for a reaction involving barium nitride was recorded as a function of temperature (Orcutt 1970). The second law of thermodynamics gives the approximate relationship

$$\ln(\text{pressure}) = A + B/T$$

where T is absolute temperature. From the data in the file barium, estimate A and B and their standard errors. Form approximate 95% confidence intervals for A and B. Examine the residuals and comment.

(See data on bCourses)

Q2: Regression and MLE

Rice 14.9.18:

Suppose that

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \ldots, n$$

where the e_i are independent and normally distributed with mean zero and variance σ^2. Find the mle's of β_0 and β_1 and verify that they are the least squares estimates. (Hint: Under these assumptions, the Y_i are independent and normally distributed with mean $\beta_0 + \beta_1 x_i$ and variance σ^2. Write the joint density function of the Y_i and thus the likelihood.