Outline

Review

- Recall the simple linear regression (SLR) model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim i.i.d.\ N(0, \sigma^2),$$

  for all $i = 1, \ldots, n$.
- The least-squares (LS) estimates:

$$\begin{aligned} \hat{\beta}_1 &= \underline{\hspace{2cm}} \\ \hat{\beta}_0 &= \underline{\hspace{2cm}} \\ \hat{\sigma}^2 &= \underline{\hspace{2cm}} \end{aligned}$$

- What are the sampling distributions of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$?

Sampling distribution of SLR estimation

Under a simple linear regression model,

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{bmatrix} \right).$$

Furthermore, let $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ be the residual mean square. Then

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2},$$

and is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

We will prove the first part of the theorem, i.e., the sampling distribution of $(\hat{\beta}_0, \hat{\beta}_1)^T$.

Outline

Random Vector and Matrix

- A **random vector** or a **random matrix** contains

  ____________.

- SLR: The response variables $Y_1, \ldots, Y_n$ can be written in the form of a random vector

$$\boldsymbol{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

- Alternative notation: ______ .

Expectation of Random Vector/Matrix

- The expectation of an $n \times 1$ random vector $\boldsymbol{Y}$ is

$$E(\boldsymbol{Y})_{n\times 1} = [E(Y_i) : i = 1, \ldots, n] = \begin{bmatrix} E(Y_1) \\ \vdots \\ E(Y_n) \end{bmatrix}$$

- SLR: What is $E(\boldsymbol{Y}|\boldsymbol{X})$?

- In general, the expectation of an $n_1 \times n_2$ random matrix $\boldsymbol{Y}$ is

Variance-Covariance Matrix of Random Vector

- The **variance-covariance matrix** of an $n \times 1$ random vector $\boldsymbol{Y}$ is

$$\begin{aligned} Var(\boldsymbol{Y}) &= E\left[(\boldsymbol{Y} - E(\boldsymbol{Y}))(\boldsymbol{Y} - E(\boldsymbol{Y}))'\right] \\ &= [\underline{\hspace{6cm}}] \end{aligned}$$

- Note: Var($\boldsymbol{Y}$) is symmetric.
  Why?
- SLR: What is $Var(\boldsymbol{Y}|\boldsymbol{X})$? ________________________

Variance-Covariance Matrix of Random Vector

- The random errors $\varepsilon_1, \ldots, \varepsilon_n$ can be written in the form of a random vector

$$\boldsymbol{\varepsilon}_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- SLR: What is $E(\boldsymbol{\varepsilon})$?

- SLR: What is the variance-covariance matrix of $\boldsymbol{\varepsilon}$?

Multivariate Normal Distribution

- Let $\boldsymbol{Y}_{n\times 1} = (Y_1, \ldots, Y_n)'$ follow a **multivariate normal distribution** with mean

$$\boldsymbol{\mu}_{n\times 1} = (\mu_1, \ldots, \mu_n)'$$

and variance

$$\boldsymbol{\Sigma}_{n\times n} = [\sigma^2_{ii'} : i = 1, \ldots, n; i' = 1, \ldots, n].$$

- We denote this by

$$\boldsymbol{Y} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

- The probability density function is

Preliminaries (Rencher and Schaalj, Chapter 4.4)

Properties of random vectors

For $\boldsymbol{Y}$ ($n \times 1$ random vector), $\boldsymbol{A}$ ($n \times n$ non-random matrix), and $\boldsymbol{b}$ ($n \times 1$ non-random vector), we have

$$\begin{aligned} E(\boldsymbol{AY} + \boldsymbol{b}) &= \boldsymbol{A}E(\boldsymbol{Y}) + \boldsymbol{b} \\ Var(\boldsymbol{AY} + \boldsymbol{b}) &= \boldsymbol{A}Var(\boldsymbol{Y})\boldsymbol{A}' \end{aligned}$$

Properties of Derivative

For $\boldsymbol{\theta}$ ($p \times 1$ vector of parameters), $\boldsymbol{c}$ ($p \times 1$ vector of variables), and $\boldsymbol{C}$ ($p \times p$ symmetric matrix of variables), we have

$$\begin{aligned} \frac{\partial(\boldsymbol{\theta}'\boldsymbol{c})}{\partial\boldsymbol{\theta}} &= \boldsymbol{c} \\ \frac{\partial(\boldsymbol{\theta}'\boldsymbol{C}\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= 2\boldsymbol{C}\boldsymbol{\theta} \end{aligned}$$

Outline

Notation

- Let $\boldsymbol{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ denote the $n \times 1$ vector of response variables.
- Let $\boldsymbol{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$ denote the $n \times 2$ design matrix of predictor variables.
- Let $\boldsymbol{\varepsilon}_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ denote the $n \times 1$ vector of random errors.
- Let $\boldsymbol{\beta}_{2\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ denote the 2×1 vector of regression coefficients.

Simple Linear Regression in Matrix Terms

- The simple linear regression model in matrix terms is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{\varepsilon} \sim \mathcal{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

- Equivalently, we have

$$\boldsymbol{Y} \sim \mathcal{MVN}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}).$$

Why?

Least Squares Method

- Recall that the least squares method minimizes

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

- In matrix terms,

$$\begin{aligned} Q(\boldsymbol{\beta}) &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{Y}'\boldsymbol{Y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \end{aligned}$$

Normal Equations

- Let

$$\left(\frac{\partial Q}{\partial \boldsymbol{\beta}}\right)_{2\times 1} = \left[\begin{array}{c} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \end{array}\right].$$

- Differentiate Q with respect to $\boldsymbol{\beta}$ to obtain:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.$$

- Set the equation above to $\mathbf{0}_{2\times 1}$ and obtain a set of normal equations:

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{Y}.$$

Estimated Regression Coefficients $\hat{\boldsymbol{\beta}}$

- Let $\hat{\boldsymbol{\beta}}_{2\times 1} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ denote the least squares estimate of $\boldsymbol{\beta}$.
- Thus the least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'}_{\text{non-random}}\boldsymbol{Y}$$

  assuming that the 2×2 matrix $\boldsymbol{X}'\boldsymbol{X}$ is nonsingular and thus invertible.
- What is the distribution of $\boldsymbol{Y}$ based on SLR?
- What is the distribution of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$?

Mean and Variance of $\hat{\boldsymbol{\beta}}$

- Recall that $\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'}_{\text{non-random}}\boldsymbol{Y}$
- What is the expectation of $\hat{\boldsymbol{\beta}}$?
- What is the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$?
- That is,

$$\begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) \end{bmatrix} = \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{bmatrix}$$

- What is the distribution of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$?

We have proved the first part of the following theorem.

Sampling distribution of SLR estimators

Under a simple linear regression model,

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{bmatrix} \right).$$

Furthermore, let $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ be the residual mean square. Then

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2},$$

and is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

- How to use the above result to perform the hypothesis testing?

$$H_0 : \beta_1 = 0, \quad \text{v.s.} \quad H_A : \beta_1 \neq 0$$

Sampling distribution of $\hat{\beta}_1$

- We have known that

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) \iff \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} \sim N(0,1).$$

- But, we do not know σ^2. A natural (unbiased) estimator of σ^2 is

  and $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$ by previous theorem.
- Consider the following test statistic:

  ______

- The denominator is also referred to as the estimated standard error of $\hat{\beta}_1$.

Hypothesis Testing for β_1

A test of interest is:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_A : \beta_1 \neq 0.$$

- The test statistic is:

- Under the $H_0 : \beta_1 = 0$,

- p-value =

- Similar procedure for CI.

Example: Wetland Species Richness

- In the wetland species richness example, the summary statistics are:
  $\bar{x} = 0.5210$, $\bar{y} = 7.9483$, $n = 58$
  $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -10.7775$, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 2.3316$
  $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 479.03$, $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 528.84$
- The least squares estimated slope is:
- The least squares estimated intercept is:
- The estimated error variance is:

Example: Wetland Species Richness

- The estimated standard error of $\hat{\beta}_1$ is:

- Note that $t_{n-2,\alpha/2} = t_{56,0.025} = 2.003$. Thus, a 95% CI for β_1 is

- Interpretation:

Example: Wetland Species Richness

- To test whether there is a linear relationship between the number of species and the percent forest cover:

- The observed test statistic is

$$t^* = \frac{\hat{\beta}_1}{\widehat{se}(\hat{\beta}_1)} =$$

- Compared with T_{56}, the p-value is

- Interpretation:

Recall:

Sampling distribution of SLR estimators

Under a simple linear regression model,

$$\begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \end{pmatrix} \sim \mathcal{MVN}\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar X^2}{\sum_{i=1}^n (X_i - \bar X)^2} & \frac{-\bar X}{\sum_{i=1}^n (X_i - \bar X)^2} \\ \frac{-\bar X}{\sum_{i=1}^n (X_i - \bar X)^2} & \frac{1}{\sum_{i=1}^n (X_i - \bar X)^2} \end{bmatrix} \right).$$

Furthermore, let $\hat\sigma^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat Y_i)^2$ be the residual mean squared. Then

$$\frac{(n-2)\hat\sigma^2}{\sigma^2} \sim \chi^2_{n-2},$$

and is independent of $\hat\beta_0$ and $\hat\beta_1$.

- We have used the above results to perform the inference on β_1.
- How to conduct the inference on β_0?

$$H_0 : \beta_0 = 0, \quad \text{v.s.} \quad H_A : \beta_0 \neq 0$$

Sampling distribution of $\hat{\beta}_0$

- What is the distribution of $\hat{\beta}_0$?

$$\hat{\beta}_0 \sim N\left(\beta_0, \underbrace{\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)}_{\text{sampling variance of } \hat{\beta}_0}\right).$$

- Cannot use Z-test, because we do not know σ^2.
- Consider T-test.

- The CI and hypothesis testing for β_0 follow similarly.

Example: Wetland Species Richness

- The estimated standard error of $\hat{\beta}_0$ is:

$$\widehat{se}(\hat{\beta}_0) \stackrel{\text{def}}{=}$$

- A 95% CI for β_0 is

$$\hat{\beta}_0 \pm t_{n-2,\alpha/2}\widehat{se}(\hat{\beta}_0)$$

- Interpretation:

Example: Wetland Species Richness

- To test whether there is zero species of wetlands with zero forest cover around

$$H_0 : \beta_0 = 0 \quad \text{vs.} \quad H_A : \beta_0 \neq 0.$$

- The observed test statistic is

- Interpretation:

Understanding the R output

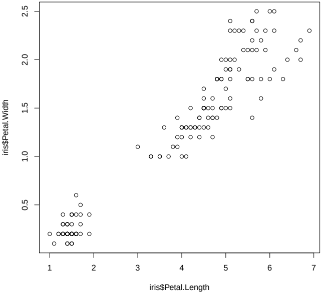
```
> fit=lm(iris$Petal.Width~iris$Petal.Length)
> summary(fit)

Call:
lm(formula = iris$Petal.Width ~ iris$Petal.Length)

Residuals:
     Min       1Q   Median       3Q      Max
-0.56515 -0.12358 -0.01898  0.13288  0.64272

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.363076   0.039762  -9.131  4.7e-16 ***
iris$Petal.Length  0.415755   0.009582  43.387  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2065 on 148 degrees of freedom
Multiple R-squared:  0.9271,	Adjusted R-squared:  0.9266
F-statistic:  1882 on 1 and 148 DF,  p-value: < 2.2e-16
```



Outline

Estimation of $E(Y_h)$

- $X_h =$ the level of X for which we want to estimate the mean response.
- X_h could be observed or not, but should be within the range of $\{X_i\}$.
- $\mu_h = E(Y_h) = \beta_0 + \beta_1 X_h =$ the mean response at X_h.
- The estimate of μ_h is

$$\hat{\mu}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h.$$

- $\hat{\mu}_h \sim N(\mu_h, \sqrt{\text{Var}(\hat{\mu}_h)})$. Why?

Estimation of $E(Y_h)$

- The variance of $\hat{\mu}_h$ is

- The estimated variance of $\hat{\mu}_h$ is

- A useful test statistic is

- A $(1-\alpha)$ CI for μ_h is

Example: Wetland Species Richness

- The estimated mean number of species at $x_h = 0.10$ is

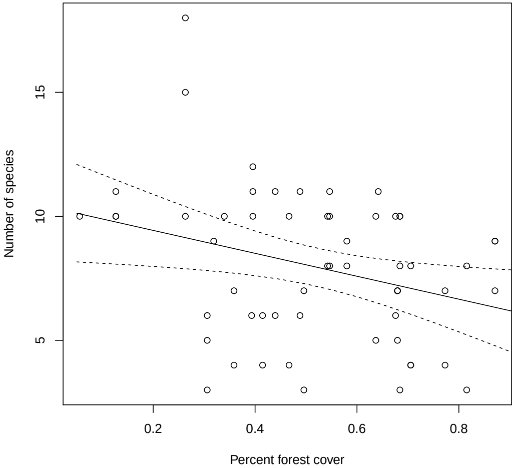
- The estimated variance of $\hat{\mu}_h$ is

$$\widehat{\mathrm{Var}(\hat{\mu}_h)} \quad =$$

- The 95% CI for the mean number of species at $X_h = 0.10$ is

- Interpretation:

Example: Wetland Species Richness



Outline

Example: Wetland Species Richness

- The fitted regression line is $\hat{y} = 10.357 - 4.622x$.
- The estimated error variance is $\hat{\sigma}^2 = \frac{479.03}{56} = 8.554$.
- Questions of interest:

  1. What is the population mean number of species for a 10% forest cover around the wetland?
  2. What is the number of species for a 10% forest cover around a wetland yet to be sampled?

- In both cases, the estimated/predicted value is:

$$\hat{y} = 10.357 - 4.622 \times 0.10 = 9.895.$$

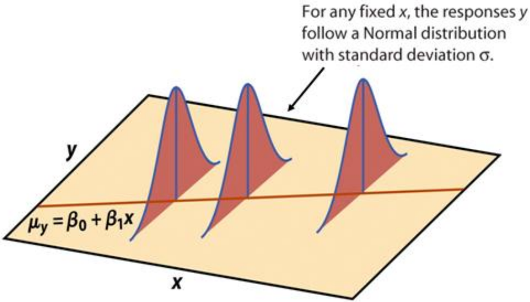
- Q: Which quantity has larger uncertainty?

Estimation vs. Prediction

Simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2), \; i = 1, \ldots, n.$$

- mean response at $X = 0.1$: $\beta_0 + \beta_1 \times 0.1$
- “new" response at $X = 0.1$: $\beta_0 + \beta_1 \times 0.1 + \varepsilon$
- sub-population vs. single observation



Estimation vs. Prediction

Consider a simple model (with covariate **0**)

$$Y_i = \mu + \varepsilon_i, \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2).$$

1. Then, estimate μ by

$$\hat{\mu} = \bar{Y}$$

   - What is $Var(\hat{\mu})$?

2. Also, predict a new observation Y by

$$\hat{Y}_{(\text{new})} = \bar{Y}$$

   - What is the variance of the prediction error?