1 Distribution Theory for $b_0, b_1, \ldots, b_{p-1}$

1.1 Means and Variances of $b_0, b_1, \ldots, b_{p-1}$

- It can be shown that the least squares estimators $b_0, b_1, \ldots, b_{p-1}$ are **unbiased** estimators for $\beta_0, \beta_1, \ldots, \beta_{p-1}$, i.e.

$$\mathrm{E}(b_k) \;=\; \beta_k \qquad \text{for } k = 0, 1, \ldots, p-1$$

so that

$$\mathrm{E}(\mathbf{b}) \;=\; \boldsymbol{\beta}$$

where $\mathbf{b}$ is the $p \times 1$ **estimated coefficient vector** and $\boldsymbol{\beta}$ is the (true) **coefficient vector**.

- The $p \times p$ **variance-covariance matrix** of the vector of coefficient estimates $\mathbf{b}$ is defined as

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} \;=\; \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} & \ldots & \sigma\{b_0, b_{p-1}\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} & \ldots & \sigma\{b_1, b_{p-1}\} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma\{b_{p-1}, b_0\} & \sigma\{b_{p-1}, b_1\} & \ldots & \sigma^2\{b_{p-1}\} \end{bmatrix}$$

where the diagonal elements $\sigma^2\{b_k\}$ are the **variances** of the b_k's and the off-diagonal elements $\sigma\{b_j, b_k\}$ are the covariances between the two estimates b_j and b_k.

It can be shown, as was true in simple linear regression, that

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} \;=\; \sigma^2 \cdot (\mathbf{X}^T\mathbf{X})^{-1}\,. \tag{1}$$

1.2 Normality of $b_0, b_1, \ldots, b_{p-1}$

- It turns out that each $b_0, b_1, \ldots, b_{p-1}$ is a **linear combination** of $Y_1, Y_2, \ldots, Y_n$, and so, because linear combinations of normal random variables are again normal, we have the following.

> **Sampling Distribution of b_k**: Under the multiple linear regression model, with the ϵ_i's independent $N(0, \sigma)$,
>
> $$b_k \;\sim\; N\left(\beta_k,\, \sigma^2\{b_k\}\right) \tag{2}$$
>
> for each $k = 0, 1, \ldots, p-1$, where $\sigma^2\{b_k\}$ is the kth diagonal element of the variance-covariance matrix (1).

2 Inference About Regression Parameters

- Usually we don't know the value of σ^2, so we estimate it by the MSE. This gives the **(estimated)** ***variance-covariance matrix*** of $\mathbf{b}$, denoted $\boldsymbol{s^2\{\mathbf{b}\}}$:

> **(Estimated) Variance-Covariance Matrix of b**:
>
> $$s^2\{\mathbf{b}\} \;=\; \text{MSE}\cdot(\mathbf{X}^T\mathbf{X})^{-1}\,.$$

  The diagonal elements of $s^2\{\mathbf{b}\}$ are the *squares* of the (estimated) **standard errors** $\boldsymbol{s\{b_0\}, s\{b_1\}, \ldots, s\{b_{p-1}\}}$ of $b_0, b_1, \ldots, b_{p-1}$ reported by statistical software when a multiple regression analysis is carried out.

- When we replace $\sigma\{b_k\}$ in (2) by its estimate $s\{b_k\}$, the resulting random variable follows a $\boldsymbol{t}$ **distibution with** $\boldsymbol{n-p}$ **degrees of freedom**, i.e.

> **Fact 2.1** Under the multiple linear regression model, with the ϵ_i's independent $N(0,\sigma)$,
>
> $$\frac{b_k-\beta_k}{s\{b_k\}} \;\sim\; t(n-p)$$
>
> for each $k = 0, 1, \ldots, p-1$.

2.1 Confidence Intervals for $\beta_0, \beta_1, \ldots, \beta_{p-1}$

- ***Confidence interval for*** $\boldsymbol{\beta_k}$ with level of confidence $100(1-\alpha)\%$:

> **Confidence Interval for** $\boldsymbol{\beta_k}$: Under the multiple linear regression model, with the ϵ_i's independent $N(0,\sigma)$, a $\mathbf{100(1-\alpha)\%}$ **confidence interval for** $\boldsymbol{\beta_k}$ is
>
> $$b_k \;\pm\; t(\alpha/2, n-p)s\{b_k\}$$
>
> for each $k = 0, 1, \ldots, p-1$, where $t(\alpha/2, n-p)$ is the $100(1-\alpha/2)$th percentile of the $t(n-p)$ distribution.

  We can be $100(1-\alpha)\%$ confident that the true (unknown) value of β_k will be contained in this interval.

2.2 Confidence Interval for a Mean Response and Prediction Interval

- A confidence interval for a mean response $\mathrm{E}(Y_h)$ and prediction interval for an individual response $Y_{h(\text{new})}$ can also be computed. See the textbook.

2.3 Hypothesis Tests for $\boldsymbol{\beta_0, \beta_1, \ldots, \beta_{p-1}}$

- ***Hypothesis Test for β_k***: For any $k = 0, 1, \ldots, p-1$, to test

$$\begin{aligned} H_0 : \beta_k &= 0 \\ H_a : \beta_k &\neq 0 \end{aligned} \tag{3}$$

the ***t* test statistic** is

> **Test Statistic**: The test statistic for the t *test for a coefficient* is
>
> $$t = \frac{b_k - 0}{s\{b_k\}}.$$

When H_0 is true, by Fact 2.1, the test statistic t follows a $t(n-p)$ distribution. **P-values** are tail areas beyond the observed t value under the $t(n-p)$ distribution.

The test of (3) (for $k \neq 0$) is a test of whether the mean response $\mathrm{E}(Y)$ changes with X_k **while the other predictors in the model are held constant**.

- Other null-hypothesized values for β_k could be used in place of zero, and one-sided alternative hypotheses can be tested.

2.4 Lack of Fit Test

- A lack of fit F test can be performed too. See the textbook.

3 Model-Checking Diagnostics

- To assess the adequacy of the multiple linear regression model, we check:

  - ▷ Linearity of the relationship between Y and each $X_1, X_2, \ldots, X_{p-1}$.
  - ▷ The normality assumption for the error term ϵ.
  - ▷ The independence assumption for the error term.
  - ▷ The constant variance assumption for the error term.
  - ▷ The presence or absence of outliers.

  using plots of the residuals (or the semi-studentized residuals).

3.1 Some Useful Plots for Checking Model Assumptions

- Some good plots for checking assumptions are:
  - ▷ For checking for **linearity** between Y and individual predictors $X_1, X_2, \ldots, X_{p-1}$:
    - * ***Scatterplot matrix*** showing scatterplots of the response and predictor variables taken pairwise.
  - ▷ For checking for **linearity** between Y and individual predictors $X_1, X_2, \ldots, X_{p-1}$ and checking for **constant variance**:
    - * Scatterplots of residuals versus individual predictors $X_1, X_2, \ldots, X_k$ (curved patterns indicate non-linearity, uneven vertical spread indicates non-constant variance).
    - * Scatterplot of residuals versus fitted (predicted) values $\hat{Y}$ (curved patterns indicate non-linearity, uneven vertical spread indicates non-constant variance).
  - ▷ For checking **normality** of the errors:
    - * Histogram of residuals (should be roughly bell-shaped).
    - * Normal probability plot of residuals (should follow a straight line).
  - ▷ For checking **whether more predictors should be included** in the model:
    - * Scatterplot of residuals versus the other potential predictors not included in the model (linear and curved patterns suggest the other predictor should be included in the model).
  - ▷ For checking **independence** of the errors (and deciding whether time should be included in the model as a predictor):
    - * Scatterplot of residuals versus the time order in which the observations were recorded (if nearby points tend to be more alike than points farther apart, either the errors are dependent or time should be included in the model).

- **Comments**:
  - ▷ All of the above plots are also useful for identifying outliers.
  - ▷ Any of the above plots of residuals could also be made using studentized residuals.
  - ▷ We can perform hypothesis tests for normality, independence, constant variance, outliers, etc. using the residuals. See the textbook.