Inference in Regression Analysis

Paweł Polak

September 11, 2017

Linear Regression Models - Lecture 2

Content of the Lecture: ALRM Book Chap. 2 (Sec. 2.1-2.6)

- Inference concerning β_1.
- Inference concerning β_0.
- Interval estimation of $\mathbb{E}\left(Y_h\right)$.
- Prediction of new observation.
- Confidence Bands for regression line.

Inference in the Normal Error Regression Model

$$Y_i = \beta_0 + \beta_1 X_i \varepsilon_i$$

- Y_i value of the dependent variable, for $i = 1, \ldots, N$,
- β_0 and β_1 unknown parameters,
- X_i is a known constant, the value of the independent variable, for $i = 1, \ldots, N$,
- $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}\left(0, \sigma^2\right)$, for $i = 1, \ldots, N$,

Maximum Likelihood Estimator(s)

- $$\widehat{\beta}_0 = b_0$$
  (the same as in least squares case)
- $$\widehat{\beta}_1 = b_1$$
  (the same as in least squares case)
- $$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2}{N} = \frac{N-2}{N}s^2,$$
  where $s^2 = \frac{\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2}{N-2}$
  (ML estimator is biased as s^2 is unbiased)

Inference Concerning β_1

Tests concerning β_1 (the slope) are often of interest, particularly

$$\begin{aligned} H_0 :& \beta_1 = 0 \\ H_1 :& \beta_1 \neq 0 \end{aligned} \tag{1}$$

Under the null hypothesis the model is

$$Y_i = \beta_0 + 0X_i + \varepsilon.$$

Hence, under H_0

- there is no linear relationship between Y and X,
- the means of all Y_i's are equal at all levels of X_i.

Sampling Distribution of b_1

- The point estimator for b_1 is

$$b_1 = \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum \left(X_i - \bar{X}\right)^2}$$

- The sampling distribution for b_1 is the distribution of b_1 that arises from the variability of b_1 when the predictor variables X_i are held fixed and the observed outputs are repeatedly sampled
- Note that the sampling distribution of b_1 will depend on our model assumptions.

Samping Distribution of b_1 In Normal Regression Model

- For a normal error regression model the sampling distribution of b_1 is normal, with mean and variance given by

$$\mathbb{E}(b_1) = \beta_1 \qquad \mathrm{Var}(b_1) = \frac{\sigma^2}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}$$

- To show this we need to go through a number of algebraic steps.

b_1 is a Linear Combination of the Observations Y_i

$$
\begin{aligned}
b_1 &= \frac{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2} \\
&\overset{(*)}{=} \frac{\sum_{i=1}^{N} \left(X_i - \bar{X}\right) Y_i}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2} \\
&= \sum_{i=1}^{N} k_i Y_i, \quad \text{where} \quad k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}.
\end{aligned}
$$

(*)

$$
\begin{aligned}
\sum_{i=1}^{N} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) &= \sum_{i=1}^{N} \left(X_i - \bar{X}\right) Y_i - \sum_{i=1}^{N} \left(X_i - \bar{X}\right) \bar{Y} \\
&= \sum_{i=1}^{N} \left(X_i - \bar{X}\right) Y_i - \bar{Y} \sum_{i=1}^{N} \left(X_i - \bar{X}\right) \\
&= \sum_{i=1}^{N} \left(X_i - \bar{X}\right) Y_i.
\end{aligned}
$$

Properties of the k_i's

It can be shown that

(i) $\sum_{i=1}^{N} k_i = 0$

(ii) $\sum_{i=1}^{N} k_i X_i = 1$

(iii) $\sum_{i=1}^{N} k_i^2 = \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$

We will use these properties to prove various properties of the sampling distribution of b_0 and b_1.

Proof (directly from the definition of k_i's):

$$\text{(i)} \quad \sum_{i=1}^{N} k_i = \sum_{i=1}^{N} \frac{X_i - \bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} X_i - N\frac{1}{N}\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = 0.$$

$$\text{(ii)} \quad \sum_{i=1}^{N} k_i X_i = \sum_{i=1}^{N} \frac{X_i - \bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} X_i \overset{\text{(i)}}{=} \frac{\sum_{i=1}^{N}(X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = 1.$$

$$\text{(iii)} \quad \sum_{i=1}^{N} k_i^2 = \sum_{i=1}^{N} \left(\frac{X_i - \bar{X}}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right)^2 = \frac{1}{\left(\sum_{i=1}^{N}(X_i - \bar{X})^2\right)^2} \sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{\sum_{i=1}^{N}(X_i - \bar{X})^2}.$$

Normality of b_1 Sampling Distribution

$$b_1 = \sum_{i=1}^{N} k_i Y_i = \frac{\sum_{i=1}^{N} \left(X_i - \bar{X}\right) Y_i}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}$$

Since b_1 is a linear combination of the Y_i's and each Y_i is an independent normal random variable, then b_1 is distributed normally as well, and

$$\mathbb{E}\left(b_1\right) = \sum_{i=1}^{N} k_i \mathbb{E}\left(Y_i\right) \quad \text{and} \quad \mathrm{Var}\left(b_1\right) = \sum_{i=1}^{N} k_i^2 \mathrm{Var}\left(Y_i\right).$$

Proof:
Follows from the fact that when $Y_1, \ldots, Y_N$ are independent normal random variables, then the linear combination

$$\sum_{i=1}^{N} a_i Y_i \sim \mathrm{N}\left(\sum_{i=1}^{N} a_i \mathbb{E}\left(Y_i\right), \sum_{i=1}^{N} a_i^2 \mathrm{Var}\left(Y_i\right)\right)$$

b_1 is an Unbiased Estimator

This can be seen using two of the properties

$$\begin{aligned}
\mathbb{E}(b_1) &= \mathbb{E}\left(\sum_{i=1}^{N} k_i Y_i\right) \\
&= \sum_{i=1}^{N} k_i \mathbb{E}(Y_i) \\
&= \sum_{i=1}^{N} k_i (\beta_0 + \beta_1 X_i) \\
&= \beta_0 \sum_{i=1}^{N} k_i + \beta_1 \sum_{i=1}^{N} k_i X_i \\
&= \beta_0 0 + \beta_1 1 \\
&= \beta_1 .
\end{aligned}$$

Variance of b_1

Since the Y_i are independent random variables with variance σ^2 and the k_i's are constant we get

$$\begin{aligned}
\mathrm{Var}\left(b_1\right) &= \mathrm{Var}\left(\sum_{i=1}^{N} k_i Y_i\right) \\
&= \sum_{i=1}^{N} k_i^2 \mathrm{Var}\left(Y_i\right) \\
&= \sum_{i=1}^{N} k_i^2 \sigma^2 \\
&= \sigma^2 \sum_{i=1}^{N} k_i^2 \quad (\text{ from (iii) for } k_i\text{'s}) \\
&= \sigma^2 \frac{1}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}.
\end{aligned}$$

Estimated Variance of b_1

- When we do not know σ^2, then we have to replace it with the MSE estimate (from the Least Square estimation)
- Let

$$s^2 = \text{MSE} = \frac{\text{SSE}}{N-2},$$

where

$$\text{SSE} = \sum_{i=1}^{N} e_i^2 \text{ and } e_i = Y_i - \widehat{Y}_i.$$

Plugging in we get

$$\widehat{\text{Var}}(b_1) = \frac{s^2}{\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2}.$$

Recap

- We now have an expression for the sampling distribution of b_1 when σ^2 is known

$$b_1 \sim \mathrm{N}\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right)$$

- When σ^2 is unknown, we have an unbiased point estimator of σ^2

$$\widehat{\mathrm{Var}}\left(b_1\right) = \frac{s^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}$$

Notation: for shorter notation we will use the following

$$\sigma^2\left\{b_0\right\} = \mathrm{Var}\left(b_0\right) \text{ and } \sigma^2\left\{b_1\right\} = \mathrm{Var}\left(b_1\right)$$

$$s^2\left\{b_0\right\} = \widehat{\mathrm{Var}}\left(b_0\right) \text{ and } s^2\left\{b_1\right\} = \widehat{\mathrm{Var}}\left(b_1\right)$$

- b_1 is normally distributed so

$$\frac{b_1 - \beta_1}{\sqrt{\sigma^2\{b_1\}}} \sim \mathrm{N}(0, 1)$$

- We do not know $\sigma^2\{b_1\}$ so it must be estimated from data. We have already derived its estimate.
- Using the estimate $s^2\{b_1\}$ it can be shown that

$$\frac{b_1 - \beta_1}{s\{b_1\}} \sim t(N-2)$$

where

$$s\{b_1\} = \sqrt{\widehat{\mathrm{Var}}(b_1)}.$$

Where Does This Come From?

- We need to rely upon the following theorem:

Theorem

For the normal error regression model

$$\frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2}{\sigma^2} \sim \chi^2(N-2)$$

and is independent of b_0 *and* b_1.

- Intuitively this follows the standard result for the sum of squared normal random variables.
- Here there are two linear constraints imposed by the regression parameter estimation that each reduce the number of degrees of freedom by one.
- We will revisit this subject soon.

Distribution of the Studentized Statistic

To derive the distribution of this statistic, first we write

$$\frac{b_1 - \beta_1}{s\{b_1\}} = \left(\frac{b_1 - \beta_1}{\sigma\{b_1\}}\right) / \left(\frac{s\{b_1\}}{\sigma\{b_1\}}\right)$$

Since $\frac{s\{b_1\}}{\sigma\{b_1\}} = \sqrt{\frac{\widehat{\mathrm{Var}}(b_1)}{\mathrm{Var}(b_1)}}$ and

$$\frac{\widehat{\mathrm{Var}}(b_1)}{\mathrm{Var}(b_1)} = \frac{\frac{\mathrm{MSE}}{\sum_{i=1}^N (X_i - \bar{X})^2}}{\frac{\sigma^2}{\sum_{i=1}^N (X_i - \bar{X})^2}} = \frac{\mathrm{MSE}}{\sigma^2} = \frac{\mathrm{SSE}}{\sigma^2(N-2)},$$

where we know (by the given theorem) the distribution of the last term is χ^2 and independent of b_1 and b_0

$$\frac{\mathrm{SSE}}{\sigma^2(N-2)} \sim \frac{\chi^2(N-2)}{(N-2)}$$

Studentized Statistic Final

Putting everything together we can see that

$$\frac{b_1 - \beta_1}{s\left\{b_1\right\}} \sim \frac{Z}{\sqrt{\frac{\chi^2(N-2)}{N-2}}},$$

and by the definition of the t-distribution given below we have our result that

$$\frac{b_1 - \beta_1}{s\left\{b_1\right\}} \sim t\left(N-2\right).$$

Definition

Let Z and $\chi^2(\nu)$ be independent random variables (standard normal and χ^2 respectively). The following random variable is a t-distributed random variable

$$t(\nu) = \frac{Z}{\sqrt{\chi^2(\nu)/\nu}}.$$

This version of the t-distribution has one parameter, the degrees of freedom ν.

Confidence Intervals and Hypothesis Testing

So we have shown that

$$\frac{b_1 - \beta_1}{s\{b_1\}} \sim \frac{Z}{\sqrt{\frac{\chi^2(N-2)}{N-2}}} = t(N-2).$$

Some things to think about:

- What is $\left(t(\nu) \mid \chi^2(\nu)\right) = \left(\frac{Z}{\sqrt{\chi^2(\nu)/\nu}} \mid \chi^2(\nu)\right)$ distributed like?
- What does the t-distribution look like?
- Why is the estimator distributed according to a t-distribution rather than a normal distribution?
- When and why is it safe to use a normal approximation?

Now that we know the sampling distribution of b_1 (t with $N-2$ degrees of freedom) we can construct *confidence intervals* and *hypothesis tests* easily.

Confidence Intervals for β_1

Since $(b_1 - \beta_1)/s\{b_1\}$ follows a t-distribution, we can make the following probability statement

$$P\left(\ t(\alpha/2; N-2) \le (b_1 - \beta_1)/s\{b_1\} \le t(1-\alpha/2; N-2)\ \right) = 1-\alpha.$$

Here, $t(\alpha/2; N-2)$ denotes the $100\alpha/2$ percentile of the t-distribution with $N-2$ degrees of freedom. By the symmetry of the t-distribution

$$t(\alpha/2; N-2) = -t(1-\alpha/2; N-2).$$

Hence, after rearranging, we obtain

$$P\left(b_1 - t(1-\alpha/2; N-2)\, s\{b_1\} \le \beta_1 \le b_1 + t(1-\alpha/2; N-2)\, s\{b_1\}\right) = 1-\alpha.$$

Since it holds for all possible values of β_1, the $1-\alpha$ confidence limits for β_1 are:

$$b_1 \pm t(1-\alpha/2; N-2)\, s\{b_1\}.$$

Test Concerning β_1: Two-Sided Test

$$\begin{aligned} H_0 : &\beta_1 = 0 \\ H_1 : &\beta_1 \neq 0 \end{aligned} \tag{2}$$

The explicit test of the alternatives H_1 is based on the test statistic:

$$t^* = \frac{b_1}{s\left\{b_1\right\}}$$

The decision rule with this test statistic for controlling the level of significance at α is:

$$\begin{aligned} &\text{If } \left|t^*\right| \leq t\left(1-\alpha/2; N-2\right), \text{ conclude } H_0 \\ &\text{If } \left|t^*\right| > t\left(1-\alpha/2; N-2\right), \text{ conclude } H_1 \end{aligned}$$

Test Concerning β_1: One-Sided Test

$$\begin{aligned} H_0 :&\beta_1 \leq 0 \\ H_1 :&\beta_1 > 0 \end{aligned} \tag{3}$$

The explicit test of the alternatives H_1 is the same as in two-sided test, and it is based on the test statistic:

$$t^* = \frac{b_1}{s\left\{b_1\right\}}$$

The decision rule with this test statistic for controlling the level of significance at α is:

$$\text{If } t^* \leq t\left(1-\alpha; N-2\right), \text{ conclude } H_0$$
$$\text{If } t^* > t\left(1-\alpha; N-2\right), \text{ conclude } H_1$$

General Test Concerning β_1

$$
\begin{aligned}
H_0 :& \beta_1 = {\beta_1}^* \\
H_1 :& \beta_1 \neq {\beta_1}^*
\end{aligned}
\tag{4}
$$

The explicit test of the alternatives H_1 is the same as in two-sided test, and it is based on the test statistic:

$$
t^* = \frac{b_1 - {\beta_1}^*}{s\left\{b_1\right\}}
$$

The decision rule with this test statistic for controlling the level of significance at α is:

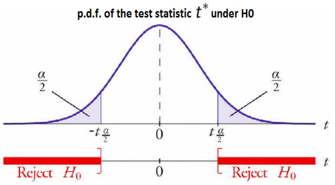
$$
\begin{aligned}
&\text{If } \ |t^*| \leq t\left(1 - \alpha/2; N - 2\right), \ \text{conclude } H_0 \\
&\text{If } \ |t^*| > t\left(1 - \alpha/2; N - 2\right), \ \text{conclude } H_1
\end{aligned}
$$

Quick Review: Hypothesis Testing

(i) Significance level α is a probability threshold below which the null hypothesis will be rejected. Common values are $\alpha = 5\%$ or 1%.

(ii) The p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

	H_0 is true $\beta_1 = {\beta_1}^*$	H_1 is true $\beta_1 \neq {\beta_1}^*$
Accept null hypothesis $\lvert t^* \rvert \leq t(1-\alpha/2; N-2)$	Right decision	Wrong decision Type II Error
Reject null hypothesis $\lvert t^* \rvert > t(1-\alpha/2; N-2)$	Wrong decision Type I Error	Right decision



(iii) A type I error is made if H_0 is rejected when H_0 is true. The probability of a type I error is denoted by α. The value of α is called the level of the test.

(iv) A type II error is made if H_0 is accepted when H_1 is true. The probability of a type II error is denoted by β.

(v) Power of the test is the probability that the decision rule will lead to conclusion H_1 when H_1 holds true, i.e., in the general test of $\beta_1 = {\beta_1}^*$ it is given by:

$$\text{Power} = P\left(\lvert t^* \rvert > t(1-\alpha/2; N-2) \mid H_1\right),$$

where $t^* = \frac{b_1 - {\beta_1}^*}{s\{b_1\}}$.

Recall $b_0 = \bar{Y} - b_1\bar{X}$, we will show that

$$b_0 \sim \mathrm{N}\left(\beta_0, \sigma^2\{b_0\}\right),$$

where $\sigma^2\{b_0\} = \sigma^2\left[\frac{1}{N} + \frac{\bar{X}^2}{\sum_{i=1}^N (X_i - \bar{X})^2}\right]$.

Proof:

- Normality comes from the fact that b_1 and $\bar{Y}$ are normal.
- $$\mathbb{E}(b_0) = \mathbb{E}\left(\bar{Y} - b_1\bar{X}\right) = \frac{1}{N}\sum_{i=1}^N \mathbb{E}(Y_i) - \bar{X}\mathbb{E}(b_1)$$ $$= \frac{1}{N}\sum_{i=1}^N (\beta_0 + \beta_1 X_i) - \bar{X}\beta_1 = \beta_0 + \beta_1\bar{X} - \bar{X}\beta_1 = \beta_0.$$
- $$\mathrm{Var}(b_0) = \mathrm{Var}\left(\bar{Y} - b_1\bar{X}\right) = \mathrm{Var}\left(\bar{Y}\right) + \bar{X}^2\mathrm{Var}(b_1) - 2\bar{X}\mathrm{Cov}\left(\bar{Y}, b_1\right)$$ $$= \frac{\sigma^2}{N} + \bar{X}^2\frac{\sigma^2}{\sum_{i=1}^N (X_i - \bar{X})^2} - 0 = \sigma^2\left[\frac{1}{N} + \frac{\bar{X}^2}{\sum_{i=1}^N (X_i - \bar{X})^2}\right].$$

Sampling Distribution of b_0

- When error variance is known:

$$b_0 \sim \mathrm{N}\left(\beta_0, \sigma^2\{b_0\}\right), \text{ where } \sigma^2\{b_0\} = \sigma^2\left[\frac{1}{N} + \frac{\bar{X}^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right].$$

- When error variance is unknown:

$$b_0 \sim t\left(\beta_0, s^2\{b_0\}\right), \text{ where } s^2\{b_0\} = MSE\left[\frac{1}{N} + \frac{\bar{X}^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right].$$

- Hence, analogously to b_1,

$$\frac{b_0 - \beta_0}{s\{b_0\}} \sim t(N-2).$$

- Hence, confidence intervals for β_0 are obtained in the same manner as those for β_1, they are given by

$$b_0 \pm t(1-\alpha/2; N-2)\, s\{b_0\}.$$

Some Considerations on Making Inferences on β_0 and β_1

(i) Effects of departures from Normality:

- Even if the distributions of Y are far from normal, the estimators b_0 and b_1 generally have the property of *asymptotic normality* - their distribution approach normality under very general conditions as the sample size increases.
- For large samples, the t value is replaced by the Z value for the standard normal distribution.

(ii) Spacing of the X levels:

- Variances of b_0 and b_1 depend on $\sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2$. Hence, in experiments where spacing of X can be controlled, we can reduce the variance of the estimators. We will revisit this topic.

Sampling distribution of $\widehat{Y}_h$

- The goal is to estimate the mean of the probability distribution of Y.
- Let X_h denote the level of X for which we wish to estimate the mean response.
- X_h may be a value which occurred in the sample or it may be some other value of the predictor variable within the scope of the model.
- The mean response when $X = X_h$ is denoted by $\mathbb{E}(Y_h)$.
- The point estimator $\widehat{Y}_h$ of $\mathbb{E}(Y_h)$ is given by

$$\widehat{Y}_h = b_0 + b_1 X_h.$$

Sampling distribution of $\widehat{Y}_h$

- We have
$$\widehat{Y}_h = b_0 + b_1 X_h$$
- Since this quantity is itself a linear combination of the Y_i's its sampling distribution is itself normal.
- The mean of $\widehat{Y}_h$ is
$$\mathbb{E}\left(\widehat{Y}_h\right) = \mathbb{E}\left(b_0\right) + \mathbb{E}\left(b_1\right) X_h = \beta_0 + \beta_1 X_h.$$

So, $\widehat{Y}_h$ is an unbiased estimator of the mean of Y_h
$$\mathbb{E}\left(\widehat{Y}_h\right) = \beta_0 + \beta_1 X_h = \mathbb{E}\left(Y_h\right).$$

Sampling distribution of $\widehat{Y}_h$

- To derive the sampling distribution variance of the mean response we first show that b_1 and $\frac{1}{N}\sum_{i=1}^{N} Y_i$ are uncorrelated and, hence, for the normal error regression model independent.
- We start with the definitions

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \text{ and } b_1 = \sum_{i=1}^{N} k_i Y_i, \text{ where } k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)}.$$

- We want to show that the mean response and the estimate b_1 are uncorrelated:

$$\operatorname{Cov}\left(\bar{Y}, b_1\right) = \sum_{i=1}^{N}\frac{k_i}{N}\operatorname{Cov}\left(Y_i, Y_i\right) = \sum_{i=1}^{N}\frac{k_i}{N}\sigma^2 = \frac{\sigma^2}{N}\sum_{i=1}^{N} k_i = 0.$$

Sampling distribution of $\widehat{Y}_h$

- This means that we can write the variance

$$\mathrm{Var}\left(\widehat{Y}_h\right) = \mathrm{Var}\left(b_0 + b_1 X_h\right) = \mathrm{Var}\left(\underbrace{\bar{Y} - b_1\bar{X}}_{=b_0} + b_1 X_h\right)$$

$$= \mathrm{Var}\left(\bar{Y} + b_1\left(X_h - \bar{X}\right)\right) = \mathrm{Var}\left(\bar{Y}\right) + \mathrm{Var}\left(b_1\right)\left(X_h - \bar{X}\right)^2$$

- Recall, $\sigma^2\{b_1\} = \frac{\sigma^2}{\sum_{i=1}^N (X_i - \bar{X})^2}$ and $s^2\{b_1\} = \frac{MSE}{\sum_{i=1}^N (X_i - \bar{X})^2}$.
- The variance of $\bar{Y}$ is $\sigma^2\{\bar{Y}\} = \frac{1}{N^2}\sum_{i=1}^N \sigma^2\{Y_i\} = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$.
- So, plugging in, we get

$$\mathrm{Var}\left(\widehat{Y}_h\right) = \frac{\sigma^2}{N} + \frac{\sigma^2\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^N \left(X_i - \bar{X}\right)^2} = \sigma^2\left[\frac{1}{N} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^N \left(X_i - \bar{X}\right)^2}\right].$$

Sampling distribution of $\widehat{Y}_h$

- Since we often won't know σ^2 we can, as usual, plug in $s^2 = \frac{SSE}{N-2}$, our estimate for it to get our estimate of this sampling distribution variance

$$s^2\left\{\widehat{Y}_h\right\} = s^2\left[\frac{1}{N} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right].$$

Sampling distribution of $\widehat{Y}_h$

- The sampling distribution of our point estimator for the output is distributed as a t-distribution with $N-2$ degrees of freedom

$$\frac{\widehat{Y}_h - \mathbb{E}(Y_h)}{s\left\{\widehat{Y}_h\right\}} \sim t(N-2).$$

- We can construct confidence intervals in the same manner as before.
- The $1-\alpha$ confidence interval for $\mathbb{E}(Y_h)$ are

$$\widehat{Y}_h \pm t(1-\alpha/2; N-2)\, s\left\{\widehat{Y}_h\right\}.$$

- From this hypothesis tests can be constructed as usual.
- The variance of the estimator for $\mathbb{E}(Y_h)$ is smallest near the mean of X. Designing studies such that the mean of X is near X_h will improve inference precision.
- When X_h is zero the variance of the estimator for $\mathbb{E}(Y_h)$ reduces to the variance of the estimator b_0 for β_0.

Prediction Interval for Single New Observation

- Essentially follows the sampling distribution arguments for $\mathbb{E}(Y_h)$.
- One difference: we want to predict an individual outcome not the mean.
- Great majority of individual outcomes deviates from the mean response and this must be taken into account by the procedure for predicting $Y_{h(\text{new})}$.

Prediction Interval for Single New Observation

- If all regression parameters are known, then

$$\frac{Y_{h(\text{new})} - \mathbb{E}(Y_h)}{\sigma} \sim \mathrm{N}(0,1),$$

where $\mathbb{E}(Y_h) = \beta_0 + \beta_1 X_h$.

- Hence, the $1-\alpha$ prediction interval for a new observation $Y_{h(\text{new})}$ is

$$\mathbb{E}(Y_h) \pm z(1-\alpha,)\sigma.$$

Prediction Interval for Single New Observation

- If the regression parameters are unknown, then, for a normal regression model, the prediction limits for a new observation $Y_{h(\text{new})}$ at a given level X_h are obtained by means of the following theorem

$$\frac{Y_{h(\text{new})} - \widehat{Y}_h}{s\{\text{pred}\}} \sim t(N-2).$$

- Numerator: $Y_{h(\text{new})} - \widehat{Y}_h$
  - represents how far the new observation $Y_{h(\text{new})}$ will deviate from the estimated mean $\widehat{Y}_h$ based on the original N observations in the study.
  - It is a prediction error, with $\widehat{Y}_h$ the best point estimate of the value of the new observation $Y_{h(\text{new})}$.
- Denominator: $s\{\text{pred}\}$
  - The variance of this prediction error can be obtained by utilizing the independence of the new observation $Y_{h(\text{new})}$ and the original N sample cases on which $\widehat{Y}_h$ is based.

Prediction Interval for Single New Observation

We have

$$\sigma^2\{\text{pred}\} = \sigma^2\left\{Y_{h(\text{new})} - \widehat{Y}_h\right\} = \sigma^2\left\{Y_{h(\text{new})}\right\} + \sigma^2\left\{\widehat{Y}_h\right\} = \sigma^2 + \sigma^2\left\{\widehat{Y}_h\right\}$$

An unbiased estimator of $\sigma^2\{\text{pred}\}$ is $s^2\{\text{pred}\} = MSE + s^2\left\{\widehat{Y}_h\right\}$, which is given by

$$s^2\{\text{pred}\} = MSE\left[1 + \frac{1}{N} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right]$$

Prediction Interval for Single New Observation

Hence, the $1-\alpha$ prediction interval for a new observation $Y_{h(\text{new})}$ is given by the following theorem

$$\widehat{Y}_h \pm t\,(1-\alpha/2; N-2)\, s\,\{\text{pred}\},$$

where $s^2\{\text{pred}\} = s^2\left\{\widehat{Y}_h\right\} + MSE.$

Interpretation: why $s^2\{\text{pred}\} \geq s^2\left\{\widehat{Y}_h\right\}$?

- $\widehat{Y}_h$ is a point estimator of the parameter of the distribution - the mean.
- $Y_{h(\text{new})}$ is a prediction of future realization of the random variable (probably not equal to its mean).
- Confidence Interval: represents an inference on a parameter, and is an interval which is intended to cover the value of the parameter.
- Prediction Interval: a statement about the value to be taken by a random variable. Wider than confidence interval.

Prediction of Mean of m New Observations for Given X_h

Denote by $\bar{Y}_{h(\text{new})}$ the mean of the new Y observations to be predicted. Assuming that the observations are independent, we get

$$\widehat{Y}_h \pm t\,(1-\alpha/2; N-2)\, s\,\{\text{predmean}\}$$

where

$$s\,\{\text{predmean}\} = \frac{s^2}{m} + s^2\left\{\widehat{Y}_h\right\},$$

or equivalently

$$s\,\{\text{predmean}\} = s^2\left[\frac{1}{m} + \frac{1}{N} + \frac{\left(X_h - \bar{X}\right)^2}{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}\right].$$

Confidence Band for Regression Line

- At times, we want to get a confidence band for the entire regression line
$$\mathbb{E}(Y) = \beta_0 + \beta_1 X.$$
- This band enables us to see the region in which the entire regression line lies.
- The Working-Hotelling $1-\alpha$ confidence band is
$$\widehat{Y}_h \pm W\ s\left\{\widehat{Y}_h\right\},$$
where $W^2 = 2F(1-\alpha; 2, N-2)$.
- Same form as before, except the t multiple is replaced with the W multiple.

Next Lecture

- ANOVA
- General linear test approach.
- Descriptive measures of linear association between X and Y.
- Normal correlations models.