b_0 and b_1 are unbiased (p. 42)

Recall that least-squares estimators (b_0, b_1) are given by:

$$b_1 = \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} = \frac{\sum x_i Y_i - n\bar{Y}\bar{x}}{\sum x_i^2 - n\bar{x}^2},$$

and

$$b_0 = \bar{Y} - b_1 \bar{x}.$$

Note that the numerator of b_1 can be written

$$\sum x_i Y_i - n\bar{Y}\bar{x} = \sum x_i Y_i - \bar{x} \sum Y_i = \sum (x_i - \bar{x}) Y_i.$$

Then the expectation of b_1's numerator is

$$\begin{aligned} E\left\{\sum(x_i-\bar{x})Y_i\right\} &= \sum(x_i-\bar{x})E(Y_i) \\ &= \sum(x_i-\bar{x})(\beta_0+\beta_1 x_i) \\ &= \beta_0\sum x_i - n\bar{x}\beta_0 + \beta_1\sum x_i^2 - n\bar{x}^2\beta_1 \\ &= \beta_1\left(\sum x_i^2 - n\bar{x}^2\right) \end{aligned}$$

Finally,

$$\begin{aligned} E(b_1) &= \frac{E\left\{\sum(x_i-\bar{x})Y_i\right\}}{\sum x_i^2 - n\bar{x}^2} \\ &= \frac{\beta_1\left(\sum x_i^2 - n\bar{x}^2\right)}{\sum x_i^2 - n\bar{x}^2} \\ &= \beta_1. \end{aligned}$$

Also,

$$
\begin{aligned}
E(b_0) &= E(\bar{Y} - b_1\bar{x}) \\
&= \frac{1}{n}\sum E(Y_i) - E(b_1)\bar{x} \\
&= \frac{1}{n}\sum[\beta_0 + \beta_1 x_i] - \beta_1\bar{x} \\
&= \frac{1}{n}[n\beta_0 + n\beta_1\bar{x}] - \beta_1\bar{x} \\
&= \beta_0.
\end{aligned}
$$

As promised, b_1 is unbiased for β_1 and b_0 is unbiased for β_0.

Example in book (p. 15): $x =$ is age of subject, Y is number attempts to accomplish task.

x	20	55	30
y	5	12	10

For these data, the least squares line is

$$\hat{Y} = 2.81 + 0.177x.$$

- A $x = 20$ year old is estimated to need $\hat{Y} = 2.81 + 0.177(20) = 6.35$ times to accomplish the task on average.

- For each year increase in age, the mean number of attempts increases by 0.177 attempts.
- For every $1/0.177 = 5.65$ years increase in age on average one more attempt is needed.
- $b_0 = 2.81$ is only interpretable for those who are zero years old.

Residuals & fitted values, Section 1.6

- The ith **fitted value** is $\hat{Y}_i = b_0 + b_1 x_i$.
- The points $(x_1, \hat{Y}_1), \ldots, (x_n, \hat{Y}_n)$ fall on the line $y = b_0 + b_1 x$, the points $(x_1, Y_1), \ldots, (x_n, Y_n)$ do not.
- The ith **residual** is

  $$e_i = Y_i - \hat{Y}_i = Y_i - (b_0 + b_1 x_i), \quad i = 1, \ldots, n,$$

  the difference between observed and fitted values.
- e_i estimates ϵ_i.

Properties of the residuals (pp. 23–24):

1. $\sum_{i=1}^{n} e_i = 0$ (from normal equations)

2. $\sum_{i=1}^{n} x_i e_i = 0$ (from normal equations)

3. $\sum_{i=1}^{n} \hat{Y}_i e_i = 0$ (?)

4. Least squares line always goes through $(\bar{x}, \bar{Y})$ (easy to show).

Estimating σ^2, Section 1.7

σ^2 is the error variance. If we *observed* the $\epsilon_1, \ldots, \epsilon_n$, a natural estimator is $S^2 = \frac{1}{n}\sum_{i=1}^n (\epsilon_i - 0)^2$. If we replace each ϵ_i by e_i we have $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n e_i^2$. However,

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n}\sum_{i=1}^n E(Y_i - b_0 - b_1 x_i)^2 \\ &= \text{...a lot of hideous algebra later...} \\ &= \frac{n-2}{n}\sigma^2. \end{aligned}$$

So in the end we use the unbiased *mean squared error*

$$MSE = \frac{1}{n-2}\sum_{i=1}^n e_i^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - b_0 - b_1 x_i)^2.$$

So an estimate of $\text{var}(Y_i) = \sigma^2$ is

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} \left(= \frac{\sum_{i=1}^{n} e_i^2}{n-2} \right).$$

Then $E(MSE) = \sigma^2$. MSE is automatically given in SAS and R.

$s = \sqrt{MSE}$ is an estimator of σ, the standard deviation of Y_i.

Example: page 15. $MSE = 5.654$ and $\sqrt{MSE} = 2.378$ attempts. (Verify this for practice.)

Chapter 2

So far we have only assumed $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$. We can *additionally* assume

$$\epsilon_1, \ldots, \epsilon_n \stackrel{iid}{\sim} N(0, \sigma^2).$$

This allows us to make *inference* about β_0, β_1, and obtain prediction intervals for a new Y_{n+1} with covariate x_{n+1}. The model is, succinctly,

$$Y_i \stackrel{ind.}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2), \quad i = 1, \ldots, n.$$

Fact: Under the assumption of normality, the least squares estimators (b_0, b_1) are also *maximum likelihood estimators* (pp. 27–30) for (β_0, β_1).

The *likelihood* of $(\beta_0, \beta_1, \sigma^2)$ is the density of the data given these parameters:

$$
\begin{aligned}
\mathcal{L}(\beta_0, \beta_1, \sigma^2) &= f(y_1, \ldots, y_n | \beta_0, \beta_1, \sigma^2) \\
&\overset{ind.}{=} \prod_{i=1}^{n} f(y_i | \beta_0, \beta_1, \sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left(-0.5 \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\right).
\end{aligned}
$$

$(\beta_0, \beta_1, \sigma^2)$ is maximized when $\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ is as small as possible $\Rightarrow$ least-squares estimators are MLEs too!

Note that the MLE of σ^2 is, instead, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n e_i^2$; the denominator changes.

Section 2.1: Inferences on β_1

From slide 1, b_1 is

$$b_1 = \frac{\sum(x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2} = \sum_{i=1}^{n} \left[\frac{(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \right] Y_i.$$

Thus, b_1 is a weighted sum of n independent normal random variables $Y_1, \ldots, Y_n$. Therefore

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right).$$

We computed $E(b_1)$ before and can use standard result for the variance of the weighted sum of independent random variables.

So,

$$sd(b_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Take b_1, subtract off its mean, and divide by its standard deviation and you've got...

$$\frac{b_1 - \beta_1}{sd(b_1)} \sim N(0, 1).$$

We will never know $sd(b_1)$; we estimate it by

$$se(b_1) = \sqrt{\frac{MSE}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Question: How do we make $\text{var}(b_1)$ *as small as possible*?

(If we do this, we cannot actually check the assumption of linearity...)

CI and $H_0 : \beta_1 = \beta_{10}$

Fact:

$$\frac{b_1 - \beta_1}{se(b_1)} \sim t_{n-2}.$$

A $(1 - \alpha)100\%$ CI for β_1 has endpoints

$$b_1 \pm t_{n-1}(1 - \alpha/2)se(b_1).$$

Under $H_0 : \beta_1 = \beta_{10}$,

$$t^* = \frac{b_1 - \beta_{10}}{se(b_1)} \sim t_{n-2}.$$

P-values are computed as usual.

Note: Of particular interest is $H_0 : \beta_1 = 0$, that $E(Y_i) = \beta_0$ *and does not depend on* x_i. That is H_0: x_i is useless in predicting Y_i.

Regression output typically produces a table like:

Parameter	Estimate	Standard error	t^*	p-value
Intercept β_0	b_0	se(b_0)	$t_0^* = \frac{b_0}{se(b_0)}$	$P(\|T\| > \|t_0^*\|)$
Slope β_1	b_1	se(b_1)	$t_1^* = \frac{b_1}{se(b_1)}$	$P(\|T\| > \|t_1^*\|)$

where $T \sim t_{n-p}$ and p is the number of parameters used to estimate the mean, here $p = 2$: β_0 and β_1. Later p will be the number of predictors in the model plus one.

The two p-values in the table test $H_0 : \beta_0 = 0$ and $H_0 : \beta_1 = 0$ respectively. The test for zero slope is usually not of interest.

[Prof. Hitchcock's SAS and R examples]

Inference about the intercept β_0

The intercept usually is not very interesting, but just in case...

Write b_0 as a linear combination of $Y_1, \ldots, Y_n$ as we did with the slope:

$$b_0 = \bar{Y} - b_1 \bar{x} = \sum_{i=1}^{n} \left[\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \right] Y_i.$$

After some slogging, this leads to

$$b_0 \sim N\left(\beta_0, \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \right).$$

Define se$(b_0) = \sqrt{MSE\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}$ and you're in business:

$$\frac{b_0 - \beta_0}{se(b_0)} \sim t_{n-2}.$$

Obtain CIs and tests about β_0 as usual...

Estimating $E(Y_h) = \beta_0 + \beta_1 x_h$

(e.g. inference about the regression line)

Let x_h be *any predictor*; say we want to estimate the mean of all outcomes in the *population* that have covariate x_h. This is given by

$$E(Y_h) = \beta_0 + \beta_1 x_h.$$

Our estimator of this is

$$\begin{aligned}
\hat{Y}_h &= b_0 + b_1 x_h \\
&= \sum_{i=1}^{n} \left[\frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} + \frac{(x_i - \bar{x})x_h}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \right] Y_i \\
&= \sum_{i=1}^{n} \left[\frac{1}{n} + \frac{(x_h - \bar{x})(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \right] Y_i
\end{aligned}$$

Again we have a linear combination of independent normals as out estimator. This leads, after slogging through some math (pp. 53–54), to

$$b_0 + b_1 x_h \sim N\left(\beta_0 + \beta_1 x_h, \sigma^2\left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]\right).$$

As before, this leads to a $(1-\alpha)100\%$ CI for $\beta_0 + \beta_1 x_h$

$$b_0 + b_1 x_h \pm t_{n-2}(1 - \alpha/2) se(b_0 + b_1 x_h),$$

where $\text{se}(b_0 + b_1 x_h) = \sqrt{MSE\left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]}$.

Question: For what value of x_h is the CI narrowist? What happens when x_h moves away from $\bar{x}$?

Prediction **intervals**

We discussed constructing a CI for the unknown mean at x_h, $\beta_0 + \beta_1 x_h$.

What if we want to find an interval that the actual *value* Y_h is in (versus only it's mean) with fixed probability?

If we knew β_0, β_1, and σ^2 this is easy:

$$Y_h = \beta_0 + \beta_1 x_h + \epsilon_h,$$

and so, for example,

$$P(\beta_0 + \beta_1 x_h - 1.96\sigma \leq Y_h \leq \beta_0 + \beta_1 x_h + 1.96\sigma) = 0.95.$$

Unfortunately, we don't know β_0 and β_1. We don't even know σ, but we can estimate all three of these.

An interval that contains Y_h with $(1-\alpha)$ probability needs to account for

1. The variability of the estimators b_0 and b_1; i.e. we don't know exactly where $\beta_0 + \beta_1 x_h$ is, and
2. The natural variability of each response built into the model; $\epsilon_h \sim N(0, \sigma^2)$.

We have

$$\begin{aligned}
\text{var}(b_0 + b_1 x_h + \epsilon_h) &= \text{var}(b_0 + b_1 x_h) + \text{var}(\epsilon_h) \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] + \sigma^2 \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + 1 \right]
\end{aligned}$$

Soooo.........

Estimating σ^2 by MSE we obtain a $(1-\alpha/2)100\%$ *prediction interval* (PI) for Y_h is

$$b_0 + b_1 x_h \pm t_{n-2}(1-\alpha/2)\sqrt{MSE\left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + 1\right]}.$$

Note: As $n \to \infty$, $b_0 \xrightarrow{P} \beta_0$, $b_1 \xrightarrow{P} \beta_1$, $t_{n-2}(1-\alpha/2) \to \Phi^{-1}(1-\alpha/2)$, and $MSE \xrightarrow{P} \sigma^2$. That is, as the sample size grows, the prediction interval converges to

$$\beta_0 + \beta_1 x_h \pm \Phi^{-1}(1-\alpha/2)\sigma.$$

Example: Toluca data.

- Find a 90% CI for the mean number of work hours for lots of size $x_h = 65$ units.
- Find a 90% PI for the number of work hours for a lot of size $x_h = 65$ units.
- Repeat both for $x_h = 100$ units.
- See SAS/R examples.

An aside: Working & Hotelling developed $100(1-\alpha)\%$ *confidence bands* for the entire regression line; see Section 2.6 for details. Scheffe's method can also be used here.