Chapter 2. Simple Linear Regression - Proofs

The following lemmas are from MATH2421

Lemma

If $X_1 \sim N(\mu_1, \sigma_1^2), \cdots, X_n \sim N(\mu_n, \sigma_n^2)$ are independent normal random variables, then

$$a_1X_1 + \cdots + a_nX_n \sim N\left(a_1\mu_1 + \cdots + a_n\mu_n, a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2\right)$$

for any constants $a_1, \cdots, a_n$.

Lemma

For any random variables $X_1, \cdots, X_n, Y_1, \cdots, Y_m$ and constants $a_1, \cdots, a_n, b_1, \cdots, b_m$, we have

$$\mathrm{Cov}(a_1X_1 + \cdots + a_nX_n, b_1Y_1 + \cdots + b_mY_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathrm{Cov}(X_i, Y_j)$$

Chapter 2. Simple Linear Regression - Proofs

Distribution of $\hat{\beta}_1$

$\hat{\beta}_1$ is equivalent to $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Now, view $\hat{\beta}_1$ as an estimator for treating random variables Y_i and write $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ where $Y_1, \cdots, Y_n$ are independent normal random variables with $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Then, by the formula of linear combinations of independent normal random variables, we get

$$\hat{\beta}_1 \sim N\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{[\sum_{i=1}^n (x_i - \bar{x})]^2}\right)$$

The mean can be simplified to

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\beta_1 \sum_{i=1}^n (x_i - \bar{x}) x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \beta_1$$

and the variance is just equal to

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

Chapter 2. Simple Linear Regression - Proofs

Distribution of $\hat{\beta}_0$

By definition, we have $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ and treat $\bar{y}$ as the random variable $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$.

Then, $\hat{\beta}_0$ is a linear combination of normal random variables, so it is also a normal random variable. Its expectation is $E(\hat{\beta}_0) = E(\bar{Y}) - \bar{x}E(\hat{\beta}_1) = \beta_0 + \beta_1\bar{x} - \bar{x}\beta_1 = \beta_0$ and its variance is

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{Y} - \hat{\beta}_1\bar{x}) =& \mathrm{Var}(\bar{Y}) + \bar{x}^2\mathrm{Var}(\hat{\beta}_1) - 2\bar{x}\mathrm{Cov}(\bar{Y}, \hat{\beta}_1) \\
=& \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 2\bar{x}\mathrm{Cov}\left(\frac{Y_1 + \cdots + Y_n}{n}, \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \\
=& \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - \sum_{i=1}^n \frac{(x_i - \bar{x})\bar{x}\sigma^2}{n\sum_{i=1}^n (x_i - \bar{x})^2} \\
=& \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}
$$

Chapter 2. Simple Linear Regression - Proofs

Unbiasedness of $\hat{\sigma}^2$

By definition, $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x}))^2$.

Now, view $\hat{\sigma}^2$ as an estimator by viewing y_i as random variable Y_i, we write

$$
\begin{aligned}
(n-2)\hat{\sigma}^2 &= \sum_{i=1}^n \big(Y_i - \bar{Y} - \hat{\beta}_1(x_i - \bar{x})\big)^2 \\
&= \sum_{i=1}^n (Y_i - \bar{Y})^2 - 2\sum_{i=1}^n (Y_i - \bar{Y})\hat{\beta}_1(x_i - \bar{x}) + \sum_{i=1}^n \hat{\beta}_1^2 (x_i - \bar{x})^2 \\
&= \left(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2\right) - 2\sum_{i=1}^n Y_i\hat{\beta}_1(x_i - \bar{x}) + \sum_{i=1}^n \hat{\beta}_1^2 (x_i - \bar{x})^2
\end{aligned}
$$

Taking expectation, $E(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2) = \sum_{i=1}^n (\beta_0 + \beta_1 x_i)^2 + n\sigma^2 - [n(\beta_0 + \beta_1\bar{x})^2 + \sigma^2] = \beta_1^2\sum_{i=1}^n (x_i - \bar{x})^2 + (n-1)\sigma^2$ and

$$
\begin{aligned}
E\left[2\sum_{i=1}^n Y_i\hat{\beta}_1(x_i - \bar{x})\right] =& 2E\left[\sum_{i=1}^n \frac{Y_i(x_i - \bar{x})\sum_{j=1}^n (x_j - \bar{x})Y_j}{\sum_{j=1}^n (x_j - \bar{x})^2}\right] \\
=& 2\frac{E\big[\sum_{i=1}^n Y_i(x_i - \bar{x})\big]^2}{\sum_{j=1}^n (x_j - \bar{x})^2} = 2\frac{\big[\sum_{i=1}^n (\beta_0 + \beta_1 x_i)(x_i - \bar{x})\big]^2 + \sigma^2\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \\
=& 2\frac{\beta_1^2\big[\sum_{i=1}^n x_i(x_i - \bar{x})\big]^2}{\sum_{j=1}^n (x_j - \bar{x})^2} + 2\sigma^2 = 2\frac{\beta_1^2\big[\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})\big]^2}{\sum_{j=1}^n (x_j - \bar{x})^2} + 2\sigma^2 \\
=& 2\beta_1^2\sum_{i=1}^n (x_i - \bar{x})^2 + 2\sigma^2
\end{aligned}
$$

and $E\big[\sum_{i=1}^n \hat{\beta}_1^2(x_i - \bar{x})^2\big] = \beta_1^2\sum_{i=1}^n (x_i - \bar{x})^2 + \sigma^2$. Putting together the above terms, we end up with

$$
E[(n-2)\hat{\sigma}^2] = (n-2)\sigma^2
$$

Chapter 2. Simple Linear Regression - Proofs

Prediction Interval for Y at $X = x_0$

First, the new single response Y includes a random error ε which contributes a variance σ^2.

Second, the point prediction $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 = \bar{y} + \hat{\beta}_1(x_0 - \bar{x})$. View it as an "estimator" by viewing y_i as random variable, we write the point prediction as

$$\hat{Y}_0 = \bar{Y} - \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}(x_0 - \bar{x})$$

Then, its variance is

$$\begin{aligned}
\mathrm{Var}(\hat{Y}_0) &= \mathrm{Var}\left(\bar{Y} - \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}(x_0 - \bar{x})\right) \\
&= \mathrm{Var}(\bar{Y}) + \mathrm{Var}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}(x_0 - \bar{x})\right) - 2\mathrm{Cov}\left(\bar{Y}, \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}(x_0 - \bar{x})\right) \\
&= \frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} - 0 = \frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}$$

Putting together, the total variance is $\sigma^2\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$ and by replacing σ with $\hat{\sigma}$, we get the standard error

$$\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Chapter 2. Simple Linear Regression - Proofs

Prediction Interval for $E(Y)$ at $X = x_0$

In this case, the variance just comes from $\hat{\beta}_0 + \hat{\beta}_1 x_0$, which has been derived in the previous slide. So its standard error is

$$\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Chapter 2. Simple Linear Regression - Proofs

$$\mathrm{Cor}(Y, \hat{Y}) = \big|\mathrm{Cor}(Y, X)\big|$$

By definition, we have

$$\mathrm{Cor}(Y, \hat{Y}) = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}$$

Note that $\bar{\hat{y}} = (\hat{y}_1 + \cdots + \hat{y}_n)/n = \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_i)/n = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y}$. Then, $\hat{y}_i - \bar{\hat{y}} = \hat{y}_i - \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i - (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}) = \hat{\beta}_1 (x_i - \bar{x})$.

Therefore,

$$\mathrm{Cor}(Y, \hat{Y}) = \frac{\hat{\beta}_1 \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{|\hat{\beta}_1| \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Note that $\hat{\beta}_1$ has the same sign of $\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$. Then, $\mathrm{Cor}(Y, \hat{Y}) = |\mathrm{Cor}(Y, X)|$.

Chapter 2. Simple Linear Regression - Proofs

Orthogonality between $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}} - \mathbf{Y}$

By definition and $\hat\beta_0 = \bar y - \hat\beta_1 \bar x$, we have

$$
\begin{aligned}
\hat{\mathbf{Y}}'(\hat{\mathbf{Y}} - \mathbf{Y}) &= \sum_{i=1}^n (\hat y_i - y_i)\hat y_i = \sum_{i=1}^n (\hat\beta_0 + \hat\beta_1 x_i - y_i)(\hat\beta_0 + \hat\beta_1 x_i) \\
&= \sum_{i=1}^n [(\bar y - y_i) + \hat\beta_1(x_i - \bar x)][\bar y + \hat\beta_1(x_i - \bar x)] \\
&= \sum_{i=1}^n [(\bar y - y_i) + \hat\beta_1(x_i - \bar x)]\hat\beta_1(x_i - \bar x) \\
&= \hat\beta_1 \left(\sum_{i=1}^n (\bar y - y_i)(x_i - \bar x) + \hat\beta_1 \sum_{i=1}^n (x_i - \bar x)^2 \right) \\
&= \hat\beta_1 \cdot 0 = 0
\end{aligned}
$$

where the last equality is due to the definition of $\hat\beta_1$.

Chapter 2. Simple Linear Regression - Proofs

SST=SSR+SSE

We write $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 - 2\sum_{i=1}^{n}\bar{y}(y_i - \hat{y}_i) \\
&= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2
\end{aligned}$$

where the second equality is due the orthogonality between $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Y}} - \mathbf{Y}$, and the third equality is due to $\sum_{i=1}^{n} e_i = 0$.

Chapter 2. Simple Linear Regression - Proofs

Equivalence between $\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)}$ and $\frac{\mathrm{Cor}(Y,X)\sqrt{n-2}}{\sqrt{1-[\mathrm{Cor}(Y,X)]^2}}$

By definition, we have

$$\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = \frac{\hat{\beta}_1 \cdot \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\hat{\sigma}} = \frac{\hat{\beta}_1 \cdot \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{n-2}}{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}$$

Again, use the fact $y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = (y_i - \bar{y}) + \hat{\beta}_1(\bar{x} - x_i)$. Then, by the definition of $\hat{\beta}_1$,

$$\begin{aligned}
\sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 + 2\sum_{i=1}^n (y_i - \bar{y})\hat{\beta}_1(\bar{x} - x_i) + \hat{\beta}_1^2 \sum_{i=1}^n (\bar{x} - x_i)^2 \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \bar{y})\hat{\beta}_1(\bar{x} - x_i) \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 \cdot \left(1 - \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}\right) \\
&= \sum_{i=1}^n (y_i - \bar{y})^2 \cdot \left(1 - [\mathrm{Cor}(Y,X)]^2\right)
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} &= \hat{\beta}_1 \cdot \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \cdot \frac{\sqrt{n-2}}{\sqrt{1 - [\mathrm{Cor}(Y,X)]^2}} \\
&= \frac{\mathrm{Cor}(Y,X)\sqrt{n-2}}{\sqrt{1 - [\mathrm{Cor}(Y,X)]^2}}
\end{aligned}$$

Chapter 2. Simple Linear Regression - Proofs

$$[\mathrm{Cor}(Y,X)]^2 = [\mathrm{Cor}(Y,\hat{Y})]^2 = R^2$$

By definition and the formula $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \cdot \left(1 - [\mathrm{Cor}(Y,X)]^2\right)$ from the previous slide, we have

$$\begin{aligned} 1 - R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} =& \frac{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \left(1 - [\mathrm{Cor}(Y,X)]^2\right)}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ =& 1 - [\mathrm{Cor}(Y,X)]^2 \end{aligned}$$

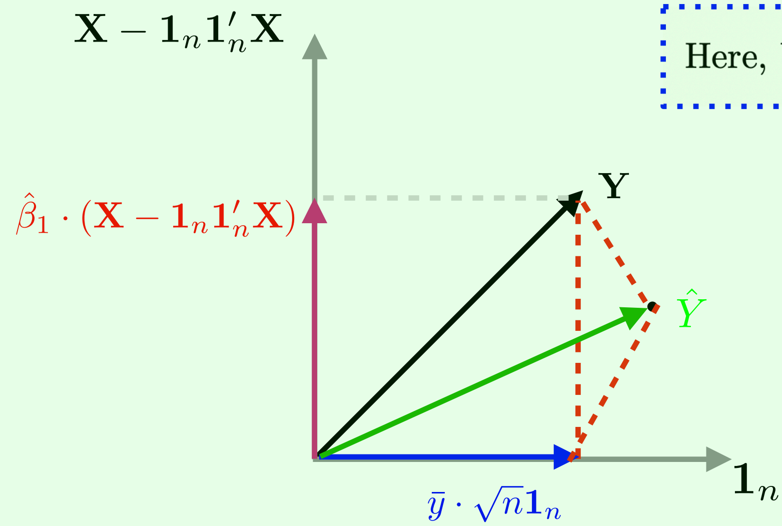
proving that $R^2 = [\mathrm{Cor}(Y,X)]^2$.

Chapter 2. Simple Linear Regression - Proofs

Understanding of No-intercept Model

Intercept Model

Define the observation vector $\mathbf{Y} = (y_1, \cdots, y_n)'$ and predictor variable $\mathbf{X} = (x_1, \cdots, x_n)'$. Then the simple linear regression is essentially the projection of $\mathbf{Y}$ onto the linear space spanned by $\mathbf{X}$ and $\mathbf{1}_n = (1, \cdots, 1)/\sqrt{n}$. By orthogonalizing, it is equivalent to project $\mathbf{Y}$ onto the two vectors $\mathbf{1}_n$ and $\mathbf{X} - \mathbf{1}_n\mathbf{1}_n'\mathbf{X}$.



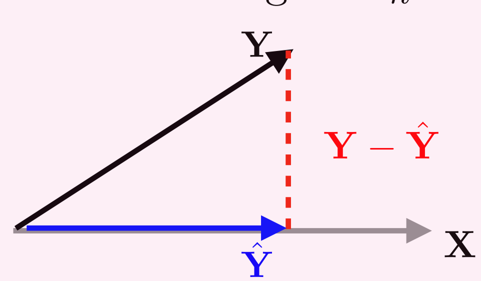
Here, $\hat{Y}$ is in the plane (generated by $\mathbf{1}_n$ and $\mathbf{X} - \mathbf{1}_n\mathbf{1}_n'\mathbf{X}$) while Y is not.

Here $\bar{y}$ is about the length of projection of Y onto $\mathbf{1}_n$

The red dashed triangle a right-angled triangle and is corresponding to the fundamental equality SST=SSR+SSE.

No-intercept Model

When there is no intercept, the linear regression is essentially the projection of $\mathbf{Y}$ onto the vector $\mathbf{X}$, or it means we are using $0 \cdot \mathbf{1}_n$ in approximating $\mathbf{Y}$, while in the intercept model this component is $\bar{y} \cdot \sqrt{n}\mathbf{1}_n$.



This right-angled triangle is corresponding to SST=SSR+SSE.

Basically, $\bar{y}$ is corresponding to the component of Y explained by the constant vector $\mathbf{1}_n$.