Least Squares Estimator

If we define the following notations,

$$\bar{x} = \frac{1}{n}\sum_i x_i \qquad \bar{y} = \frac{1}{n}\sum_i y_i$$

$$\text{SXX} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n\,\bar{x}^2$$
$$\text{SYY} = \sum_i (y_i - \bar{y})^2 = \sum_i y_i^2 - n\,\bar{y}^2$$
$$\text{SXY} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - n\,\bar{x}\,\bar{y}$$

then the estimators are given by

They are random variables! => The one from data set: an estimate

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\,\bar{x} \\ \hat{\beta}_1 = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \dfrac{\text{SXY}}{\text{SXX}} \\ \hat{\sigma}^2 = \dfrac{\sum_i \hat{e}_i^{\,2}}{n-2} = \dfrac{\text{SYY} - \text{SXY}^2\,/\,\text{SXX}}{n-2}. \end{cases}$$

In lecture notes.

Under our estimated model, we therefore have the fitted values given x_i and the residual, i.e. the difference between the fitted value and the realised value.

$$\hat{y}_i = \hat{\mathbb{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1\,x_i$$
$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1\,x_i$$

Remark 2.1. For the detailed derivation, please refer to the lecture notes. You should be familiar with their derivation as they may be tested in midterm and final exam. ■

Remark 2.2. $\hat{\beta}_0$ and $\hat{\beta}_1$ can be written as linear combinations of y_i:

$\hat{\beta}_1 = \sum_i \left(\frac{x_i - \bar{x}}{SXX}\right) y_i - \sum_i \cancel{\left(\frac{x_i - \bar{x}}{SXX}\right)}^{\,0} \bar{y}$

$$\hat{\beta}_1 = \sum_i \left(\frac{x_i - \bar{x}}{\text{SXX}}\right) y_i \quad \text{and} \quad \hat{\beta}_0 = \sum_i \left[\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{\text{SXX}}\right)\right] y_i.$$

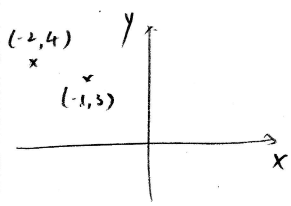
use $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ + linear form of $\hat{\beta}_1$

This is useful when deriving the distribution and consistency of the estimators in Exercise 4.1. ■

Remark 2.3. By the derivative condition of RSS with respect to β_0, we have

$$\sum_i \hat{e}_i = 0 \quad \text{and} \quad \bar{y} = \hat{\beta}_0 + \hat{\beta}_1\,\bar{x}\,. \qquad ■$$

Exercise 2.1. (2012 Fall Midterm #3) Use the simple linear regression model to fit a straight line on two data points: $(-2, 4), (-1, 3)$. What are the values of $\hat{\beta}_0$ and $\hat{\beta}_1$?



There are 2 points => st. line

$$\frac{y-4}{x+2} = \frac{y-3}{x+1}$$

$$xy - 4x + y - 4 = xy - 3x + 2y - 6$$

$$y = -x + 2$$

Exercise 2.2. Show that

$$\sum_i x_i \hat{e}_i = 0 \qquad \text{and, therefore} \qquad \sum_i \hat{y}_i \hat{e}_i = 0.$$

By $RSS_{\beta_1} = 0$, we have $\sum_i x_i \underbrace{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}_{\hat{e}_i} = 0$.

$\therefore \sum_i x_i \hat{e}_i = 0$.

$$\sum_i \hat{y}_i \hat{e}_i = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_i)\hat{e}_i$$
$$= \hat{\beta}_0 \cancelto{0}{\sum_i \hat{e}_i} + \hat{\beta}_1 \cancelto{0}{\sum_i x_i \hat{e}_i}$$
$$= 0$$

Exercise 2.3. Show that $\hat{\sigma^2}$ is an unbiased estimator of σ^2, i.e.

$$\mathbb{E}(\hat{\sigma^2}) = \sigma^2.$$

Refer to Chapter 3 notes.

Remark 3.2. It should be noticed that the $\hat{\sigma}^2$ here is under the model in H_1. The reason of using this instead of $\hat{\sigma}^2$ under H_0 is examined in Exercise 3.1. ■

Exercise 3.1. What is the estimator of σ^2 under H_0? Explain why the use of it in the denominator makes no sense.

If H_0 is true, $\hat{\sigma}^2$ should be sampled variance of y, i.e. $\frac{1}{n-1}SYY$.

Consider using it in the denominator of F, i.e. let

$F' = \frac{SSreg}{SYY/(n-1)}$ * If F is large, it could be.
① $SSreg$ is large $\Rightarrow H_1$ is likely to be true
② $SYY = RSS_{H_0}$ is small $\Rightarrow H_0$ is likely to be valid

∴ There will be NO valid rejection criteria (∵ No conclusion).

But for $F = \frac{SSreg}{\hat{\sigma}^2} = \frac{SSreg}{RSS_{H_1}/(n-2)}$ * If F is large, it could be
① $SSreg$ is large $\Rightarrow H_1$ is likely to be true
② RSS_{H_1} is small $\Rightarrow H_1$ is likely to be true.

∴ F is large $\Rightarrow H_1$ is true.

Exercise 3.2. Show that

$$\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2$$

$$\Sigma_i (y_i - \bar{y})^2 = \Sigma_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$
$$= \Sigma_i (y_i - \hat{y}_i)^2 + \Sigma_i (\hat{y}_i - \bar{y})^2 + 2\Sigma_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})^2$$

Consider $2\Sigma_i (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 2\Sigma_i \hat{e}_i (\hat{y}_i - \bar{y})$
$$= 2\underbrace{\Sigma_i \hat{e}_i \hat{y}_i}_{0} - 2\bar{y}\underbrace{\Sigma_i \hat{e}_i}_{0}$$
$$= 0$$

$\therefore \Sigma_i (y_i - \bar{y})^2 = \Sigma_i (y_i - \hat{y}_i)^2 + \Sigma_i (\hat{y}_i - \bar{y})^2$

3.2 ANOVA Table

Thanks to the result of Exercise 3.2, we have a neat and tidy representation of ANOVA, which is called the ANOVA table.

Source	df	SS	MS	F	p-value
Regression	1	SSreg	SSreg /1	SSreg / $\hat{\sigma}^2$	$P(F_{1,n-1} > F)$
Residual	$n-1$	RSS_{H_1}	$\hat{\sigma}^2 = RSS/(n-2)$		
Total	$n-2$	SYY			

You should be extremely familiar with the above table because it appears in every midterm. We will practice this in Exercise 4.3 and 4.4.

Remark 3.3. Therefore, we can define the "Coefficient of Determination" to be

$$R^2 = \frac{\text{SSreg}}{\text{SYY}} \in [0, 1].$$

The realised value summarises the strength of relationship between the sampled response and predictors. ■

4 Intervals, Tests and Band

Besides testing the mean functions by ANOVA, we will also want to perform test on individual parameters. Therefore, we need the distributions of the estimator. We begin this section with an exercise.

Exercise 4.1. Prove that $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased and find their variance and asymptotic distribution.

$$\mathbb{E}(\hat{\beta}_1) = \sum_i \left(\frac{x_i - \bar{x}}{SXX}\right) \mathbb{E}(y_i) = \frac{1}{SXX} \sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i) = \frac{1}{SXX}\left(\beta_0 \sum_i (x_i - \bar{x}) + \beta_1 \sum x_i (x_i - \bar{x})\right) = \beta_1$$

$$\mathbb{E}(\hat{\beta}_0) = \mathbb{E}(\bar{y} - \hat{\beta}_1 \bar{x}) = \mathbb{E}(\bar{y}) - \bar{x}\,\mathbb{E}(\hat{\beta}_1) = \frac{1}{n}\sum_i (\beta_0 + \beta_1 x_i) - \bar{x}\beta_1 = \beta_0$$

$$Var(\hat{\beta}_1) = \sum_i \left(\frac{x_i - \bar{x}}{SXX}\right)^2 Var(y_i) = \frac{\sum_i (x_i - \bar{x})^2}{SXX^2}\sigma^2 = \frac{SXX}{SXX^2}\sigma^2 = \frac{\sigma^2}{SXX}$$

$$Var(\hat{\beta}_0) = \sum_i \left[\frac{1}{n} - \bar{x}\left(\frac{x_i - \bar{x}}{SXX}\right)\right]^2 Var(y_i) = \sigma^2 \sum_i \left[\frac{1}{n^2} - \frac{2\bar{x}(x_i - \bar{x})}{nSXX} + \frac{\bar{x}^2 (x_i - \bar{x})^2}{SXX^2}\right]$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\bar{y}^2}{SXX}\right)$$

By CLT, $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX^2}\right)\right)$ & $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{SXX}\right)$

4.1 Confidence Intervals and Tests for Intercept and Slope

With the distributions of the estimator and some facts in statistics, we can construct the test statistics from the distribution derived in Exercise 4.1.

Confidence Interval and Test for Intercept

If we want to test whether the intercept is a certain value β_0^*, i.e.

X, $\mathbb{E}X = \mu$
$Var(X) = \sigma^2$

Test stat
$\frac{\bar{X} - \mu}{\sigma}$ ← $\bar{x}$ estimator of μ
← If we do not know this, we use the estimator

$$H_0 : \beta_0 = \beta_0^* \quad \text{vs} \quad H_1 : \beta_0 \neq \beta_0^*,$$

then the test statistic is

$$t = \frac{\hat{\beta}_0 - \beta_0^*}{\text{se}(\hat{\beta}_0)} \sim t(n-2) \quad \text{where} \quad \text{se}(\hat{\beta}_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\text{SXX}}}.$$

Now
$\hat{\beta}_0$, $\mathbb{E}\hat{\beta}_0 = \beta_0$
$Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right)$

Therefore, for significance level α, we reject H_0 when $|t| > t_{1-\frac{\alpha}{2}}(n-2)$. Also, the $(1-\alpha) \times 100\%$ confidence interval of β_0 is given by

$$\hat{\beta}_0 - t_{1-\frac{\alpha}{2}}(n-2)\,\text{se}(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{1-\frac{\alpha}{2}}(n-2)\,\text{se}(\hat{\beta}_0).$$

← Explain the construction of confidence interval.

Confidence Interval and Test for Slope

Similarly, for the test of slope, i.e.

$$H_0 : \beta_1 = \beta_1^* \quad \text{vs} \quad H_1 : \beta_1 \neq \beta_1^*,$$

the test statistic is

$$t = \frac{\hat{\beta}_1 - \beta_1^*}{\text{se}(\hat{\beta}_1)} \sim t(n-2) \quad \text{where} \quad \text{se}(\hat{\beta}_0) = \frac{\hat{\sigma}}{\sqrt{\text{SXX}}}.$$

Therefore, for significance level α, we reject H_0 when $|t| > t_{1-\frac{\alpha}{2}}(n-2)$. Also, the $(1-\alpha) \times 100\%$ confidence interval of β_1 is given by

$$\hat{\beta}_1 - t_{1-\frac{\alpha}{2}}(n-2)\,\text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{1-\frac{\alpha}{2}}(n-2)\,\text{se}(\hat{\beta}_1).$$

Remark 4.1. Obviously, a test of zero slope, i.e.

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0,$$

is equivalent to testing

$$H_0 : \mathbb{E}(Y|X = x) = \beta_0 \quad \text{vs} \quad H_1 : \mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x,$$

which is our ANOVA F-test in Section 3. Therefore, they should give the same result. Mathematically, if we look at the t-statistics,

$$t = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{\text{SXX}}}$$

$$t^2 = \frac{\hat{\beta}_1^{\,2}}{\hat{\sigma^2}\,/\,\text{SXX}} = \frac{\hat{\beta}_1^{\,2}\,\text{SXX}}{\hat{\sigma^2}} = \frac{\text{SXY}^2}{\hat{\sigma^2}\,\text{SXX}} = \frac{\text{SSreg}}{\hat{\sigma^2}} = F.$$

In general, we have

$$F(1, m) = \frac{\chi^2(1)}{\chi^2(m)/m} = \frac{Z^2}{\chi^2(m)/m} = \left(\frac{Z}{\sqrt{\chi^2(m)/m)}}\right)^2 = t(m)^2 \qquad \blacksquare$$

Exercise 4.2. Construct a 95% confidence interval for the slope from the data set $\{(1, 1), (4, 9), (10, 10)\}$, given $t_{0.975}(1) = 12.7062$. Bosco argues that the confidence interval you construct has a 95% probability of including the true slope. Explain whether he is correct.

$SXX = 42$, $SYY = \frac{146}{3}$, $SXY = 37$

$\therefore \hat{\beta}_1 = \frac{SXY}{SXX} = \frac{37}{42}$, $\hat{\sigma}^2 = \frac{1}{n-2}\left(SYY - \frac{SXY^2}{SXX}\right) = \frac{1}{3-2}\left(\frac{146}{3} - \frac{37^2}{42}\right) = 16.0714$

$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SXX}} = \frac{\sqrt{16.0714}}{\sqrt{42}} = 0.6186$

A 95% C.I. for β_1 is

$\hat{\beta}_1 \pm t_{0.975}(1)\, se(\hat{\beta}_1) = \frac{37}{42} \pm 7.8599$.

Bosco is WRONG. The above interval is NOT random so we cannot talk about probability of it. In fact, it is just a realisation of the random interval. If there are many of these intervals, 95% of them will include the true β_1.

Exercise 4.3. (2013 Fall Midterm #1) Fill in the missing values in the following tables of regression output from a data set of size 100.

ANOVA Table

Source	df	SS	MS	F
Regression				
Residual				
Total				

Coefficient Table

Variable	Coefficient	s.e.	t-statistics	p-value
Constant	0.5854			0.2188
X		0.4927		
$n =$	$\hat{\sigma} = 4.714$	$R^2 = 0.03294$		

Exercise 4.4. (2012 Spring Midterm #1) Fill in the missing values in the following tables of regression output. In R, it is found that $qf(1-9.5e^{-9}, 1, 6) = 1917.3$. Also, $\bar{x} = 5.125, \bar{y} = -9.1974$, SXX $= 54.875$.

ANOVA Table

Source	df	SS	MS	F	p-value
Regression	1	228.9	228.9	1917.3	9.5e-09
Residual	6	0.7162	0.1194		
Total	7	229.6			

Coefficient Table

Variable	Coefficient	s.e.	t-statistics	p-value
Constant	1.2702	0.2684	4.7325	0.00322
X	-2.04245	0.04665	43.79	9.5e-0.9
$n =$ 8	$\hat{\sigma} =$ 0.3455	$R^2 =$ 0.9970		

By $qf(1-9.5e^{-9}, 1, 6) = 1917.3$, $F = 1917.3$, $df_{SSreg} = 1$, $df_{Res} = 6$

$df_{Total} = df_{SSreg} + df_{Res} = 1 + 6 = 7$, $n = SS_{res} + 2 = 6 + 2 = 8$

$t_{\beta_1}^2 = F \Rightarrow |t_{\beta_1}| = \sqrt{1917.3} \Rightarrow P_{\beta_1} = 9.5e\text{-}0.9$

$se(\hat{\beta}_1) = \frac{|\hat{\beta}_1|}{t_{\beta_1}} = \frac{2.04245}{\sqrt{1917.3}} = 0.04665$

$\hat{\sigma} = \sqrt{SXX} \cdot se(\hat{\beta}_1) = 0.3455 \Rightarrow MS_{res} = \hat{\sigma}^2 = 0.3455^2 = 0.1194$

$\Rightarrow SS_{res} = 6 \times 0.3455^2 = 0.7162$

$MS_{reg} = SS_{reg} = F \times MS_{reg} = 1917.3 \times 0.1194 = 228.9286$

$SSY = SS_{res} + SS_{reg} = 228.9286 + 0.7162 = 229.6418$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -9.1974 + 2.04245 \times 5.125 = 1.2702$

$se(\hat{\beta}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SXX}} = 0.3455 \sqrt{\frac{1}{8} + \frac{5.125^2}{54.875}} = 0.2684$

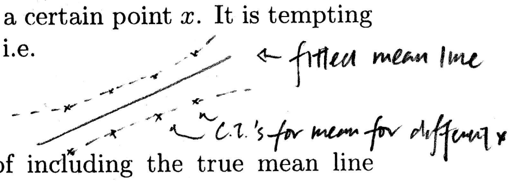
$|t_{\beta_0}| = \frac{|\hat{\beta}_0|}{se(\hat{\beta}_0)} = \frac{1.2702}{0.2684} = 4.7325$

$R^2 = \frac{SS_{reg}}{SSY} = \frac{228.9}{229.6} = 0.9970$

4.3 Confidence Band

In the previous subsection, we construct confidence interval of mean for a certain point x. It is tempting to connect all the upper limits and lower limits of confidence intervals, i.e.

$$(\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{1-\frac{\alpha}{2}} \operatorname{sefit}(\hat{y}|x), \quad \forall x$$



and say that this random band has a $(1-\alpha) \times 100\%$ probability of including the true mean line $\mathbb{E}(Y|X=x) = \beta_0 + \beta_1 x$. However, this is wrong (see Remark 4.3 and Exercise 4.5). The correct band is given below.

Confidence Band for Mean Function

The $(1-\alpha) \times 100\%$ confidence band of the mean function is given by

$$C(x) = (\hat{\beta}_0 + \hat{\beta}_1 x) \pm \sqrt{2F_{1-\alpha}(2, n-2)}\, \operatorname{sefit}(\hat{y}|x), \quad \forall x.$$

Therefore, it is true that

$$\Pr(\text{The mean line lies in the confidence band}) = \Pr\Big(\forall x,\ \mathbb{E}(Y|X=x) \in C(x)\Big) = 1 - \alpha.$$

Remark 4.3. For confidence interval $C(x) = (\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{1-\frac{\alpha}{2}} \operatorname{sefit}(\hat{y}|x_*)$, we have by definition

$$\forall x,\ \Pr\Big(\mathbb{E}(Y|X=x) \in C(x)\Big) = 1 - \alpha.$$

This relationship holds for each point, i.e. pointwise. While for the confidence band, we have

$$\Pr\Big(\forall x,\ \mathbb{E}(Y|X=x) \in C(x)\Big) = 1 - \alpha.$$

Here, the inclusion is for the entire line. The two cases are different. ■

Exercise 4.5. Explain, why it is wrong to say the band,

$$(\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{1-\frac{\alpha}{2}} \operatorname{sefit}(\hat{y}|x), \quad \forall x$$

has a $(1-\alpha) \times 100\%$ probability of including the mean line $\mathbb{E}(Y|X=x) = \beta_0 + \beta_1 x$.

Let $C(x) = (\hat{\beta}_0 + \hat{\beta}_1 x) \pm t_{1-\frac{\alpha}{2}} \operatorname{sefit}(\hat{y}|x)$, we know that

$$\forall x,\ P(\mathbb{E}(Y|X=x) \in C(x)) = 1-\alpha$$

Consider $P(\mathbb{E}(Y|X=x) \in C(x), \forall x) = 1-\alpha$, note the event

$$\{\mathbb{E}(Y|X=x) \in C(x), \text{ for some } x_1 \neq x_2\} \supset \{\mathbb{E}(Y|X=x) \in C(x), \forall x\}$$

$$P(\mathbb{E}(Y|X=x) \in C(x), \forall x) \leq P(\mathbb{E}(Y|X=x) \in C(x), \text{ for } x_1 \neq x_2)$$

$$= P(\mathbb{E}(Y|X=x_1) \in C(x))\, P(\mathbb{E}(Y|X=x_2) \in C(x) \mid \mathbb{E}(Y|X=x_1) \in C(x))$$

$$< 1-\alpha$$

as $P(\mathbb{E}(Y|X=x_2) \in C(x) \mid \mathbb{E}(Y|X=x_1) \in C(x))$

Exercise 4.6. For the data set $\{(1,1),(4,9),(10,10)\}$, construct

1. a 95% confidence interval and a 95% prediction interval for the point $x = 3$, and
2. a 95% confidence band.
3. What is the value of the band when $x = 3$?

You are given $t_{0.975}(1) = 12.7062$ and $F_{0.95}(2,1) = 199.5$.

$\hat\beta_1 = \frac{37}{42}$. $\hat\beta_0 = \frac{20}{3} - \frac{37}{42}\cdot 5 = \frac{95}{42}$, $\hat\sigma = \sqrt{16.0714} = 4.0089$

$se_{fit}(\hat y(3)) = 4.0089\sqrt{\frac{1}{3} + \frac{(3-5)^2}{42}} = 2.6245$, $se_{pred}(\hat y(3)) = 4.0089\sqrt{1+\frac{1}{3}+\frac{(3-5)^2}{42}} = 4.7916$

$\therefore$ 95% C.I. for $x=3$ is $\frac{95}{42} + \frac{37}{42}\cdot 3 \pm 12.7062\times 2.6245 = 4.9048 \pm 33.347$

95% P.I. for $x=3$ is $4.9048 \pm 12.7062 \times 4.7916 = 4.9048 \pm 60.883$

95% CB is $\frac{95}{42} + \frac{37}{42}x \pm \sqrt{2\times 199.5}\cdot 4.0089\sqrt{\frac{1}{3}+\frac{(x-5)^2}{42}}$

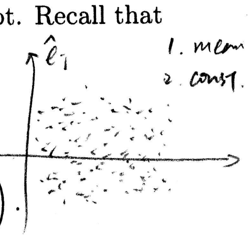
The value of CB at $x=3$ is

$\frac{95}{42} + \frac{37}{42}\cdot 3 \pm 19.975 \cdot 2.6245 = 4.9048 \pm 52.424$.

5 Residuals

To check whether our model assumption is valid, a good way is to look at the residual plot. Recall that the residuals

$$\hat e_i = y_i - \hat y_i = y_i - \hat\beta_0 - \hat\beta_1 x_i$$



1. mean zero
2. const. variance

so this gives

$$\mathbb{E}(\hat e_i) = \mathbb{E}(y_i) - \beta_0 - \beta_1 x_i = 0 \qquad \text{and} \qquad \mathrm{Var}(\hat e_i) = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_i-\bar x)^2}{\mathrm{SXX}}\right).$$

The data set gives a set of realised $\hat e_i$. According to our observation above, these realised residuals should

- have mean close to zero, and
- have constant variance for all value x_i. if n is large

A plot that satisfies the above criteria is a null plot, which indicates that the model assumption is valid and the regression is a good fit.

6 Appendix

For more reference, you may refer to the following text books.

References

[1] DOUGLAS C. MONTGOMERY, ELIZABETH A. PECK AND G. GEOFFREY VINING (2006). *Introduction to Linear Regression Analysis*, Wiley.

[2] ROBERT V. HOGG AND ALLEN T. CRAIG *Introduction to Mathematical Statistics*, Pearson.