Last week we revised/learned

Random variables

Statistics of a sample

Methods of estimation

Distributions related to the normal distribution

Mean and Variance

- Let Y be a discrete random variable. Its mean and variance are

$$\mu := \mathbb{E}(Y) = \sum_{y \in P} y\, f(y)$$

$$\sigma^2 := \mathbb{E}((Y-\mu)^2) = \sum_{y \in P} (y-\mu)^2 f(y)$$

where $f(y) = P(Y = y)$ is a probability mass function.

- Random variables Y_1 and Y_2 are independent if

$$\mathrm{Cov}(Y_1, Y_2) := \mathbb{E}((Y_1-\mu_1)(Y_2-\mu_2)) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\,\mathbb{E}(Y_2) = 0$$

Statistics of a Sample

- Let $Y_1, \dots, Y_n$ be i.i.d. random variables with unknown mean μ and variance σ^2. Then

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

are unbiased estimators of μ and σ^2, that is

$$\mathbb{E}(\bar{Y}) = \mu \qquad \mathbb{E}(S^2) = \sigma^2$$

- Let $y_1, \dots, y_n$ be a sample of observations of $Y_1, \dots, Y_n$. Then

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

are estimates of μ and σ^2, called the sample mean and sample variance.

Some Distribution Theory Relating to the Normal Distribution

- Let $Y_i \overset{ind}{\sim} N(\mu_i, \sigma_i^2)$ and $a_i \in \mathbb{R}$ for $1 \leq i \leq n$ and let

$$Z = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n$$

- Then

$$\mathbb{E}(Z) = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

$$\text{Var}(Z) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

- Furthermore, Z is normally distributed

$$Z = \sum_{i=1}^{n} a_i Y_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Some distribution theory relating to the normal distribution

- Let $Y_i \overset{ind}{\sim} N(0,1)$ for $i = 1, \ldots, n$, then

$$Z = Y_1^2 + Y_2^2 + \ldots + Y_n^2 \sim \chi_n^2$$

  is distributed according to the chi-squared distribution with n d.o.f.

- Let $Z \sim \chi_p^2$ and $V \sim \chi_q^2$ be independent random variables, then

$$W = \frac{Z/p}{V/q} \sim F_{p,q}$$

  is distributed according to the F-distribution with p and q d.o.f.

- Let $Y \sim N(0,1)$ and $Z \sim \chi_n^2$ be independent random variables, then

$$W = \frac{Y}{\sqrt{Z/n}} \sim t_n$$

  is distributed according to the Student's t-distribution with n d.o.f.

Lecture 2

The simple linear regression model

Aim: to introduce the SLR model and its basic properties

1. The model
2. Least squares estimation
3. Properties of the slope and the intercept
4. Estimating variance of the random error term
5. Testing hypotheses for the slope and intercept

- Consider a situation with one response variable Y and one predictor X:
  - We will always assume that X can be controlled – it is known
  - The response Y can only be observed – it is unknown

- For instance, a company wants to investigate how its sales depend on the day of the week, then:
  - X = Day of the week
  - Y = Sales at the company

- We want to predict (or estimate) the mean value of Y for given values of X working from a sample on n pairs of observations

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

The SLR model

- Mathematically, the regression of a random variable Y on variable X is

$$\mathbb{E}(Y|X = x)$$

i.e. the expected value of Y when X takes the specific value x.

- The regression of Y on X is linear if

$$\mathbb{E}(Y|X = x) = \beta_0 + \beta_1 x$$

where the unknown parameters β_0 and β_1 are the intercept and the slope of a specific straight line.

- Suppose that Y_1, Y_2, ..., Y_n are independent instances of the random variable Y that are observed at the values x_1, x_2, ..., x_n of X. If the regression of Y on X is linear, then for $i = 1, 2, \ldots, n$

$$Y_i = \mathbb{E}(Y|X = x) + \varepsilon_i = \beta_0 + \beta_1 x + \varepsilon_i$$

where ε_i is the random error in Y_i and is such that $\mathbb{E}(\varepsilon_i|X) = 0$.

The SLR model

- The random error ε_i represents variation in Y strictly due to random phenomenon that cannot be predicted or explained. In other words, all unexplained variation in Y is called random error:

- The random error ε_i does not depend on X, nor does it contain any information about Y. Otherwise it would be a systematic error.

- We will assume that the random errors ε_i are independent identically distributed normal random variables with mean 0 and common variance σ^2

$$\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

- This in turn implies that

$$\boxed{Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2) \quad \text{with} \quad \mu_i = \beta_0 + \beta_1 x_i.}$$

With these assumptions in place the model is called the normal SLR model. The parameters β_0 and β_1 are called regression coefficients.

The SLR model

Lecture 2

The simple linear regression model

1. The model
2. Least squares estimation
3. Properties of the slope and the intercept
4. Estimating variance of the random error term
5. Testing hypotheses for the slope and intercept

Least squares estimation

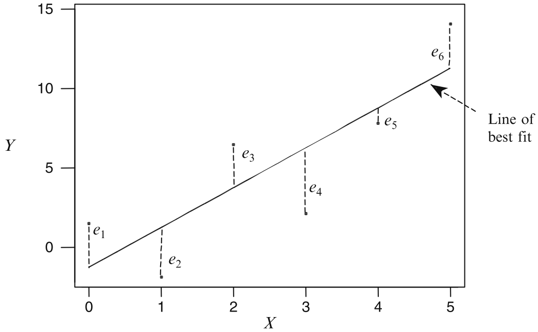
- We want to estimate β_0 and β_1 by finding the line which "best" fits our data, that is, we want to choose $\hat{\beta}_0$ and $\hat{\beta}_1$ such that fitted (predicted) values

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

  are as "close" as possible to observations y_i.

- This can be done by minimising the difference between y_i and $\hat{y}_i$. These differences a called residuals:

$$e_i = y_i - \hat{y}_i$$



Claim. The least squares estimates of β_0 and β_1 for the SLR model are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{s_{xy}}{s_{xx}}$$

where

$$s_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad s_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

and $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ are the mean values of y_i and x_i.

Proof. The sum $SS_e = \sum_i e_i^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$ is a function of two parameters. To find its minimum we differentiate it with respect to β_0 and β_1:

$$\frac{\partial SS_e}{\partial \beta_0} = -2\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$
$$\frac{\partial SS_e}{\partial \beta_1} = -2\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)\, x_i$$

Requiring

$$\left.\frac{\partial SS_e}{\partial \beta_0}\right|_{\beta_0=\hat\beta_0,\, \beta_1=\hat\beta_1} = \left.\frac{\partial SS_e}{\partial \beta_1}\right|_{\beta_0=\hat\beta_0,\, \beta_1=\hat\beta_1} = 0$$

yields the so-called normal equations

$$n\hat\beta_0 + \hat\beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$
$$\hat\beta_0 \sum_{i=1}^n x_i + \hat\beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

Solving these equations for $\hat\beta_0$ and $\hat\beta_1$ yields the wanted answer. Full details are in the lecture notes.

Example

- A manufacturer wants to investigate the time it takes (in minutes) to produce individual orders of different sizes. The relation between the time and size is expected to be linear. Data from 20 randomly selected orders is given below:

Order	1	2	3	4	5	6	7	8	9	10
Run Time	195	215	243	162	185	231	234	166	253	196
Run Size	175	189	344	88	114	338	271	173	284	277

Order	11	12	13	14	15	16	17	18	19	20
Run Time	220	168	207	225	169	215	147	230	208	172
Run Size	337	58	146	277	123	227	63	337	146	68

- The means of observations are $\bar{x} = 201.75$ (Run Size) and $\bar{y} = 202.05$ (Run Time). Thus

$$s_{xx} = \sum_{i=1}^{20}(x_i - \bar{x})^2 = 191473.80, \quad s_{xy} = \sum_{i=1}^{20}(x_i - \bar{x})(y_i - \bar{y}) = 49638.25$$

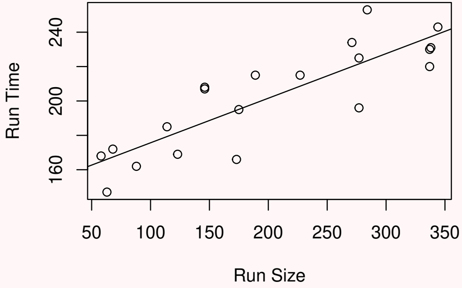
giving

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{sxx}} = 0.259, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 149.797$$

Example

- We have found that the equation of the fitted regression line is

$$y = 149.797 + 0.259x$$



- The intercept is $\beta_0 = 149.797$. We interpret this value as the average set up time, that is 149.797 minutes.
- The slope of the line is $\beta_1 = 0.259$. Thus, we say that each additional unit to be produced is predicted to add 0.259 minutes to the run time.

Lecture 2

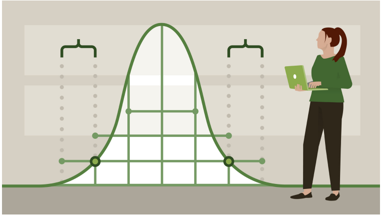
The simple linear regression model

1. The model
2. Least squares estimation
3. Properties of the slope and the intercept
4. Estimating variance of the random error term
5. Testing hypotheses for the slope and intercept

- Every time we fit the model to a different sample from the same population we obtain *different estimates* $\hat{\beta}_1$ and $\hat{\beta}_0$.

  Question. How well $\hat{\beta}_1$ and $\hat{\beta}_0$ do estimate the unknown true values of β_1 and β_0?

- To answer this question we need to determine distributions of $\hat{\beta}_1$ and $\hat{\beta}_0$.
  We can then use statistical tools to make informed decisions.



Properties of the slope and the intercept

- **Idea.** We need to rewrite $\hat{\beta}_1$ and $\hat{\beta}_0$ as

$$\hat{\beta}_1 = \sum_{i=1}^{n} c_i y_i \qquad \hat{\beta}_0 = \sum_{i=1}^{n} d_i y_i$$

  for suitable $c_i = c_i(x)$ and $d_i = d_i(x)$.

- Replacing observations y_i with random variables Y_i gives

$$\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i \qquad \hat{\beta}_0 = \sum_{i=1}^{n} d_i Y_i$$

  where $\hat{\beta}_1$ and $\hat{\beta}_0$ are now least squares estimators of β_1 and β_0.

- To be consistent with the notation, we should write "capital $\hat{\beta}_1$" and "capital $\hat{\beta}_0$" but there are no such letters, thus perhaps $\hat{B}_1$ and $\hat{B}_0$ could be a better notation.

- We have assumed that $Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2)$ with unknown μ_i and σ^2. Hence $\hat{\beta}_1$ and $\hat{\beta}_0$ are also normally distributed random variables, i.e. $\hat{\beta}_1 \sim N(?, ?)$ and $\hat{\beta}_0 \sim N(?, ?)$.

Properties of the slope and the intercept

- We have found that

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

  Since

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0$$

  the sum in the numerator is

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i - \bar{y}\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})y_i$$

- Thus we can rewrite $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \sum_{i=1}^n c_i y_i \quad \text{with} \quad c_i = \frac{x_i - \bar{x}}{s_{xx}}$$

Properties of the slope and the intercept

Claim. $\hat{\beta}_1$ is an unbiased estimator of β_1, that is $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/s_{xx})$.

Proof. Recall that $Y_i \overset{ind}{\sim} N(\mu_i, \sigma^2)$ with $\mu_i = \beta_0 + \beta_1 x_i$. Thus

$$\begin{aligned}\mathbb{E}(\hat{\beta}_1) &= \mathbb{E}\Big(\sum_{i=1}^n c_i Y_i\Big) = \sum_{i=1}^n c_i\,\mathbb{E}(Y_i) \\ &= \sum_{i=1}^n c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i\end{aligned}$$

where $c_i = (x_i - \bar{x})/s_{xx}$.

But $\sum_{i=1}^n c_i = 0$ and $\sum_{i=1}^n c_i x_i = 1$ since $\sum_{i=1}^n (x_i - \bar{x})x_i = s_{xx}$. Hence $\mathbb{E}(\hat{\beta}_1) = \beta_1$.

Next, since Y_i's are independent,

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\sum_{i=1}^n c_i Y_i\right) = \sum_{i=1}^n c_i^2 \cdot \mathrm{Var}(Y_i) = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{s_{xx}^2} \cdot \sigma^2 = \frac{\sigma^2}{s_{xx}}$$

A linear combination of normally distributed random variables is also a normally distributed random variable, and the claim follows.

- We now repeat the same analysis for $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \bar{x}\sum_{i=1}^{n} c_i y_i = \sum_{i=1}^{n}\left(\frac{1}{n} - c_i\,\bar{x}\right) y_i.$$

where $c_i = (x_i - \bar{x})/s_{xx}$.

Claim. $\hat{\beta}_0$ is an unbiased estimator of β_0, that is

$$\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)\sigma^2\right)$$

Proof. Your homework. (See lecture notes.)

Summary

- The (normal) simple linear regression model is

$$Y_i = \mathbb{E}(Y_i|X = x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

where Y_i are the response variables, X is an explanatory variable, and ε_i are random errors, usually assumed to be independent and normally distributed

$$\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2) \quad \Longrightarrow \quad Y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

- The LSE of β_0 and β_1 minimizing the $SS_E = \sum_{i=1}^n e_i^2$ are

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \qquad \hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- $\hat{\beta}_1$ and $\hat{\beta}_0$ are unbiased estimators of β_1 and β_0 distributed normally by

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{s_{xx}}\right), \qquad \hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)\sigma^2\right)$$

Lecture 2

The simple linear regression model

1. The model
2. Least squares estimation
3. Properties of the slope and the intercept
4. Estimating variance of the random error term
5. Testing hypotheses for the slope and intercept

Estimating variance of the random error term

- We found that estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ are parametrised by the unknown true values of β_1, β_0 and σ^2:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{s_{xx}}\right), \qquad \hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)\sigma^2\right)$$

  This means that making predictions about the slope and intercept requires knowing σ^2.

- Recall that errors in the model are

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 x_i) = Y_i - (\textit{unknown regression line at } x_i)$$

- Since β_0 and β_1 are unknown, we replace them with estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, giving the residuals

$$E_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) = Y_i - (\textit{estimated regression line at } x_i)$$

  These residuals can now be used to estimate σ^2.

Estimating variance of the random error term

- Let $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Then it can be shown that

$$\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sigma^2} \sim \chi^2_{n-2}$$

and

$$\mathbb{E}\left[\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right] = (n-2)\,\sigma^2$$

- Denote the sum of residuals squared by

$$SS_E := \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Then

$$\hat{\sigma}^2 := MS_E = \frac{SS_E}{n-2}$$

is an unbiased estimate of σ^2.

Estimating variance of the random error term

- **Question.** What is meaning of $n-2$ in the formula below mean?

$$\hat{\sigma}^2 = MS_E = \frac{SS_E}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- The denominator $n-2$ represents the fact that there are only $n-2$ linearly independent residuals. Indeed, the normal equations imply that

$$\sum_{i=1}^n e_i = \sum_{i=1}^n e_i x_i = 0$$

  which allow us to express e_{n-1} and e_n as linear combinations of $e_1, \ldots, e_{n-2}$.

- The quantity $\nu_E = n-2$ is called the number of degrees of freedom of SS_E, and the quantity MS_E is called the mean residual (or error) sum of squares.

- A very similar formula is used to define the sample variance

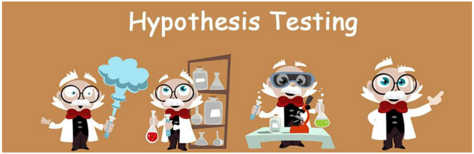
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

  In this case there are $n-1$ independent y_i's around $\bar{y}$, in other words, there are $n-1$ degrees of freedom of variation around $\bar{y}$.

Lecture 2

The simple linear regression model

1. The model
2. Least squares estimation
3. Properties of the slope and the intercept
4. Estimating variance of the random error term
5. Testing hypotheses for the slope and intercept



- Hypothesis testing is a systematic way to test claims or ideas about a group or population. It involves the following steps:
  1. Identify a hypothesis (claim)
  2. Select a decision criterion
  3. Collect a random sample
  4. Compare the observed result against expectation if the claim was true
  5. Accept or reject the hypothesis

Testing hypotheses for the slope and intercept

- Suppose we wish to test the hypothesis that the slope β_1 equals a certain value, say β_1^*, and the true regression model is $Y_i = \beta_0 + \beta_1^* x_i + \varepsilon_i$.

- The appropriate hypotheses are

$$H_0 : \beta_1 = \beta_1^* \qquad H_1 : \beta_1 \neq \beta_1^*$$

  where H_0 is the null hypothesis and H_1 is a two-sided alternative.

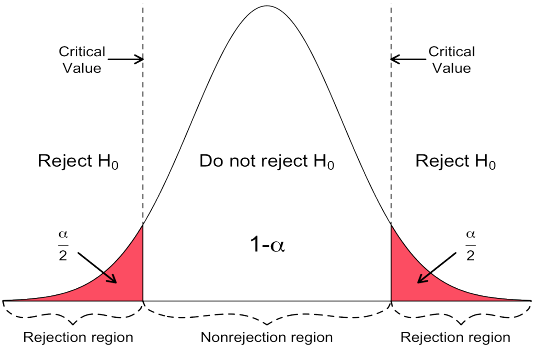
- If H_0 is true, then $\hat{\beta}_1 \sim N(\beta_1^*, \sigma^2/s_{xx})$ and

$$Z = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\sigma^2/s_{xx}}} \sim N(0,1)$$

  If σ^2 was known, we could use a Z-test to test the hypotheses. However, typically, σ^2 is unknown. Replacing σ^2 with its estimate $\hat{\sigma}^2 = MS_E$ causes $N(0,1)$ result in a Student's t-distribution with $\nu = n-2$ d.o.f.:

$$\boxed{T = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\hat{\sigma}^2/s_{xx}}} \sim t_{n-2}}$$

- The test procedure computes the value t of T for a given sample, and compares with the upper $\alpha/2$ percentage point of the t_{n-2} distribution, $t_{\alpha/2,n-2}$, where α is an a priori chosen significance level:
  - We reject the null hypothesis H_0 if $|t| \geq t_{\alpha/2,n-2}$
  - We say that there is not sufficient evidence to reject H_0 if $|t| < t_{\alpha/2,n-2}$



Testing hypotheses for the slope and intercept

- A similar procedure can be used to test hypotheses about the intercept

$$H_0 : \beta_0 = \beta_0^* \qquad H_1 : \beta_0 \neq \beta_0^*$$

- If H_0 is true, then $\hat{\beta}_0 \sim N\big(\beta_0^*, (1/n + \bar{x}^2/s_{xx})\,\sigma^2\big)$ and the test statistic is

$$\boxed{T = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{\hat{\sigma}^2\big(1/n + \bar{x}^2/s_{xx}\big)}} \sim t_{n-2}}$$

- It is convenient to denote the estimated standard error via "se", for instance

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2/s_{xx}} \qquad \mathrm{se}(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\big(1/n + \bar{x}^2/s_{xx}\big)}$$

Example

- Consider the manufacturer production data. We want to test the hypothesis

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

assuming $\alpha = 5\% = 0.05$.

- From the sample data we compute $\hat{\beta}_1 = 0.259$, $\hat{\sigma}^2 = 264.14$, $s_{xx} = 191473.75$ and

$$\text{se}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2/s_{xx}} = \sqrt{264.14/191473.75} = 0.037.$$

Thus the calculated value of the test statistic is

$$t_{cal} = t = \frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} = \frac{0.259}{0.037} = 7.0$$

- The critical value of the test statistic is

$$t_{crit} = t_{\alpha/2, 20-2} = t_{0.025, 18} = 2.1$$

Since $t_{cal} > t_{crit}$ we reject the null hypothesis, i.e. $\beta_1 \neq 0$.

Summary

- A test statistic to test the null hypothesis $H_0 : \beta_i = \beta_i^*$ is

$$T = \frac{\hat{\beta}_i - \beta_i^*}{\mathrm{se}(\hat{\beta}_i)} \sim t_{n-2}$$

We reject H_0 if $t_{cal} = |t| > t_{\alpha/2,n-2} = t_{crit}$.

- The estimated standard errors of β_1 and β_0 are

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 / s_{xx}} \qquad \mathrm{se}(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2 \big(1/n + \bar{x}^2 / s_{xx}\big)}$$

where

$$\hat{\sigma}^2 \equiv MS_E = \frac{SS_E}{n-2} = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

is an estimate of σ^2.

Next week

Further inference and significance of regression