Statistics
MATH 324

McGill University, Montréal, Canada

Fall 2018

In this Chapter we will discuss:

(1) Simple Linear regression model

(2) The method of least-squares for parameter estimation

(3) Properties of the parameter estimators and statistical inference

(4) Prediction

Introduction

- Often we are interested in studying (potential) relationship between variables. More specifically, we would like to study how one variable depends on other variables. This is the topic of regression analysis.
- The main variable of interest is called dependent or response variable, y, say; the remaining variables are called independent variables (explanatory variables, predictors, covariates, or features) which are represented by $x_1, x_2, \ldots, x_d$.
- Here, we focus on the case of $d = 1$, i.e. studying the relationship between y and one explanatory variable, say x.

Introduction

- Often we are interested in studying (potential) relationship between variables. More specifically, we would like to study how one variable depends on other variables. This is the topic of regression analysis.
- The main variable of interest is called dependent or response variable, y, say; the remaining variables are called independent variables (explanatory variables, predictors, covariates, or features) which are represented by $x_1, x_2, \ldots, x_d$.
- Here, we focus on the case of $d = 1$, i.e. studying the relationship between y and one explanatory variable, say x.

Introduction

- Often we are interested in studying (potential) relationship between variables. More specifically, we would like to study how one variable depends on other variables. This is the topic of regression analysis.

- The main variable of interest is called dependent or response variable, y, say; the remaining variables are called independent variables (explanatory variables, predictors, covariates, or features) which are represented by $x_1, x_2, \ldots, x_d$.

- Here, we focus on the case of $d = 1$, i.e. studying the relationship between y and one explanatory variable, say x.

Example 1

- The data is from a study in central Florida where 15 alligators were captured and two measurements were made on each of the alligators. The weight (in pounds) was recorded with the snout vent length (in inches – this is the distance between the back of the head to the end of the nose).

  The goal: to determine whether there is a relationship between the weight (y) and snout vent length (x). The data are on the log scale (natural logarithms).

  The R code posted on myCourses. (will be discussed in class).

Example 2

- The data here is called “Prestige” and comes from the “car” package library(car) in R. The Prestige data has 102 rows and 6 columns. Each row is an observation that relates to an occupation. The columns relate to predictors such as average years of education, percentage of women in the occupation, prestige of the occupation, etc. Our focus is to investigate the relationship between income (y) and average years of education (x).

  The R code posted on myCourses. (will be discussed in class).

Our convention

- In this course, we assume the x is a fixed variable and the response variable is random, Y.
- If the explanatory variable is also random, X, our statistical analysis will be conditional on $X = x$, where x is fixed.

Our convention

- In this course, we assume the x is a fixed variable and the response variable is random, Y.
- If the explanatory variable is also random, X, our statistical analysis will be conditional on $X = x$, where x is fixed.

Simple linear regression

- Defintion:

  In a simple linear regression model we assume that

$$E(Y|x) = \beta_0 + \beta_1 x$$

  or equivalently,

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

  where ε is a random variable with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$.

Simple linear regression

- Defintion:

In a simple linear regression model we assume that

$$E(Y|x) = \beta_0 + \beta_1 x$$

or equivalently,

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

where ε is a random variable with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$.

Simple linear regression

- Defintion:

  In a simple linear regression model we assume that

  $$E(Y|x) = \beta_0 + \beta_1 x$$

  or equivalently,

  $$Y = \beta_0 + \beta_1 x + \varepsilon$$

  where ε is a random variable with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$.

...

- Unknown parameters: $(\beta_0, \beta_1, \sigma^2)$, where (β_0, β_1) are called regression parameters, and σ^2 is called the error variance.
- ε is also called "error" or "noise" term in the model.
- Note:

  if we remove the error term ε from the model then we will get a "deterministic" linear model. Such model does not allow for any error, for example, when predicting the value of Y based on x. This is not our interest in this course.

Interpretation of the slope β_1 and intercept β_0

- The slope β_1 can be interpreted as the change in the mean or expected value of Y if x increases by one unit:

$$E(Y|x+1) - E(Y|x) = \Big\{\beta_0 + \beta_1(x+1)\Big\} - \Big\{\beta_0 + \beta_1 x\Big\} = \beta_1.$$

- The intercept β_0 can be interpreted as the expected value of Y at level $x = 0$ of the explanatory variable:

$$E(Y|x=0) = \beta_0 + \beta_1 \times 0 = \beta_0.$$

- In some applications β_0 may not be even meaningful ! (Examples 1 and 2).

Interpretation of the slope β_1 and intercept β_0

- The slope β_1 can be interpreted as the change in the mean or expected value of Y if x increases by one unit:

$$E(Y|x+1) - E(Y|x) = \Big\{\beta_0 + \beta_1(x+1)\Big\} - \Big\{\beta_0 + \beta_1 x\Big\} = \beta_1.$$

- The intercept β_0 can be interpreted as the expected value of Y at level $x = 0$ of the explanatory variable:

$$E(Y|x=0) = \beta_0 + \beta_1 \times 0 = \beta_0.$$

- In some applications β_0 may not be even meaningful ! (Examples 1 and 2).

Data

- Pre-experiment data:

$$(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$$

which we assume they follow the model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, 2, \ldots, n.$$

- Post-experiment data:

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

which also follow

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, 2, \ldots, n$$

Parameter estimation

- Given the data $(x_i, y_i), i = 1, 2, \ldots, n$, how do we estimate the unknown parameters $(\beta_0, \beta_1, \sigma^2)$ in a simpel linear regression?
- Answer: the method of least-squares.

Parameter estimation

- Given the data $(x_i, y_i), i = 1, 2, \ldots, n$, how do we estimate the unknown parameters $(\beta_0, \beta_1, \sigma^2)$ in a simpel linear regression?
- Answer: the method of least-squares.

The idea of the least-squares (LS)

- Assume that $(\hat{\beta}_0, \hat{\beta}_1)$ are the estimates of the regression parameters obtained using the least-squares (LS) method. The fitted line through the data is then given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad , \quad i = 1, 2, \ldots, n.$$

- The LS method obtains $(\hat{\beta}_0, \hat{\beta}_1)$ by minimizing the sum of squares of the vertical deviations (called residual errors) of the observed values y_i from fitted line or fitted values $\hat{y}_i$, for $i = 1, 2, \ldots, n$.

- residual errors are: $y_i - \hat{y}_i$, for $i = 1, 2, \ldots, n$.

The idea of the least-squares (LS)

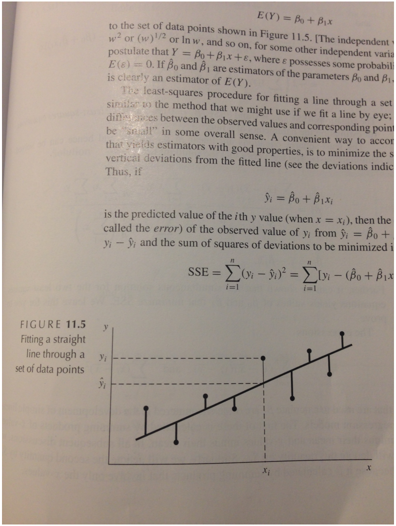
- Assume that $(\hat{\beta}_0, \hat{\beta}_1)$ are the estimates of the regression parameters obtained using the least-squares (LS) method. The fitted line through the data is then given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \ , \quad i = 1, 2, \ldots, n.$$

- The LS method obtains $(\hat{\beta}_0, \hat{\beta}_1)$ by minimizing the sum of squares of the vertical deviations (called residual errors) of the observed values y_i from fitted line or fitted values $\hat{y}_i$, for $i = 1, 2, \ldots, n$.

- residual errors are: $y_i - \hat{y}_i$, for $i = 1, 2, \ldots, n$.

Picture taken from the book



$$E(Y) = \beta_0 + \beta_1 x$$

to the set of data points shown in Figure 11.5. [The independent
w^2 or $(w)^{1/2}$ or $\ln w$, and so on, for some other independent varia
postulate that $Y = \beta_0 + \beta_1 x + \varepsilon$, where ε possesses some probabili
$E(\varepsilon) = 0$. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of the parameters β_0 and β_1,
is clearly an estimator of $E(Y)$.

The least-squares procedure for fitting a line through a set
similar to the method that we might use if we fit a line by eye;
differences between the observed values and corresponding point
be "small" in some overall sense. A convenient way to accor
that yields estimators with good properties, is to minimize the s
vertical deviations from the fitted line (see the deviations indic
Thus, if

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

is the predicted value of the ith y value (when $x = x_i$), then the
called the *error*) of the observed value of y_i from $\hat{y}_i = \hat{\beta}_0 +$
$y_i - \hat{y}_i$ and the sum of squares of deviations to be minimized i

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x$$

FIGURE **11.5** Fitting a straight line through a set of data points

Sum of squares of residual errors

Definition:

The sum of squares of the residual errors (SSE) is given by

$$\mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left\{ y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right\}^2 .$$

The Least squares estimates of the regression parameters

Defintion:

The least squares estimates of β_0 and β_1 are the values that minimize the sum of squares of the residual errors, viz.

$$(\hat{\beta}_0, \hat{\beta}_1) = \text{argmin}_{(\beta_0, \beta_1) \in \mathbb{R}^2} \sum_{i=1}^{n} \left\{ y_i - (\beta_0 + \beta_1 x_i) \right\}^2$$

$$= \text{argmin}_{(\beta_0, \beta_1) \in \mathbb{R}^2} \text{SSE}(\beta_0, \beta_1).$$

Least squares equations

- The least squares estimates are the solutions of the least squares equations:

$$\frac{\partial \mathrm{SSE}(\beta_0, \beta_1)}{\partial \beta_0} = 0 \quad ; \quad \frac{\partial \mathrm{SSE}(\beta_0, \beta_1)}{\partial \beta_1} = 0.$$

or equivalently,

$$\begin{aligned} \frac{\partial \mathrm{SSE}(\beta_0, \beta_1)}{\partial \beta_0} &= -2\sum_{i=1}^{n}\Big\{y_i - (\beta_0 + \beta_1 x_i)\Big\} = 0, \\ \frac{\partial \mathrm{SSE}(\beta_0, \beta_1)}{\partial \beta_1} &= -2\sum_{i=1}^{n} x_i\Big\{y_i - (\beta_0 + \beta_1 x_i)\Big\} = 0. \end{aligned}$$

- Consider the quantities:

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \quad , \quad S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$$

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

The least squares estimates of parameters

Theorem 1:

The least squares estimates of β_0 and β_1 are given by

$$\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n \quad , \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

And, the estimate of error variance is given by

$$\hat{\sigma}^2 = \frac{\mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1)}{n-2} = \frac{\sum_{i=1}^{n} \left\{ y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right\}^2}{n-2}.$$

Computational formula

- We may compute S_{xx} and S_{xy} using:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n \times (\bar{x}_n)^2 \quad , \quad S_{xy} = \sum_{i=1}^{n} x_i y_i - n \times (\bar{x}_n \bar{y}_n).$$

- And,

$$\hat{\sigma}^2 = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2} = \frac{S_{yy} - (\hat{\beta}_1)^2 S_{xx}}{n-2}.$$

Estimation of the mean response

- Given the fitted model, what is the estimated mean or expected value of Y for a fixed pre-specified value $x = x_0$?

  i.e., we wish to estimate

$$E(Y|x_0) = \mu(x_0) = \beta_0 + \beta_1 x_0.$$

- Answer:

$$\hat{\mu}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0,$$

  where $(\hat{\beta}_0, \hat{\beta}_1)$ are LS estimates.

Estimation of the mean response

- Given the fitted model, what is the estimated mean or expected value of Y for a fixed pre-specified value $x = x_0$?

  i.e., we wish to estimate

$$E(Y|x_0) \equiv \mu(x_0) = \beta_0 + \beta_1 x_0.$$

- Answer:

$$\hat{\mu}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0,$$

  where $(\hat{\beta}_0, \hat{\beta}_1)$ are LS estimates.

Estimation of the mean response

- Given the fitted model, what is the estimated mean or expected value of Y for a fixed pre-specified value $x = x_0$?

  i.e., we wish to estimate

$$E(Y|x_0) \equiv \mu(x_0) = \beta_0 + \beta_1 x_0.$$

- Answer:

$$\hat{\mu}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0,$$

  where $(\hat{\beta}_0, \hat{\beta}_1)$ are LS estimates.

Prediction of the response

- We wish to predict the future response

$$Y^* = \beta_0 + \beta_1 x^* + \varepsilon^*$$

at a given value $x = x^*$ of the explanatory variable.

- Answer: For any given value $x = x^*$, a predictor of Y^* is

$$\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

Prediction of the response

- We wish to predict the future response

$$Y^* = \beta_0 + \beta_1 x^* + \varepsilon^*$$

at a given value $x = x^*$ of the explanatory variable.

- Answer: For any given value $x = x^*$, a predictor of Y^* is

$$\widehat{Y^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

Properties of the LS estimators

- We now investigate statistical properties of the estimators $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$.
- To determine the properties of the LS estimators, we need to to make certain assumptions about the errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

Properties of the LS estimators

- We now investigate statistical properties of the estimators $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$.
- To determine the properties of the LS estimators, we need to to make certain assumptions about the errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$.

Assumption 1

Recall the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \quad i = 1, 2, \ldots, n.$$

- We assume that, for all $i = 1, 2, \ldots, n$,

$$E(\varepsilon_i) = 0 \ , \ Var(\varepsilon_i) = \sigma^2$$

for some unknown σ^2.

Furthermore, the errors are uncorrelated, viz.

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad , \quad \text{for any} \quad i \neq j \in \{1, 2, \ldots, n\}.$$

Assumption 1

Recall the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \quad i = 1, 2, \ldots, n.$$

- We assume that, for all $i = 1, 2, \ldots, n$,

$$E(\varepsilon_i) = 0 \ , \ \ Var(\varepsilon_i) = \sigma^2$$

for some unknown σ^2.

Furthermore, the errors are uncorrelated, viz.

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \ , \quad \text{for any} \ \ i \neq j \in \{1, 2, \ldots, n\}.$$

Consequences of Assumption 1

- We have that, for all $i = 1, 2, \ldots, n$,

$$E(Y_i|x_i) = \beta_0 + \beta_1 x_i \quad , \quad Var(Y_i|x_i) = \sigma^2.$$

And, the Y_i will also be uncorrelated, i.e.

$$Cov(Y_i, Y_j) = 0 \quad , \quad \text{for any} \quad i \neq j \in \{1, 2, \ldots, n\}.$$

Consequences of Assumption 1

- We have that, for all $i = 1, 2, \ldots, n$,

$$E(Y_i|x_i) = \beta_0 + \beta_1 x_i \quad , \quad Var(Y_i|x_i) = \sigma^2.$$

And, the Y_i will also be uncorrelated, i.e.

$$Cov(Y_i, Y_j) = 0 \quad , \quad \text{for any} \quad i \neq j \in \{1, 2, \ldots, n\}.$$

Properties of the LS estimators

Theorem 2:

Under Assumption 1, the LS estimators $(\hat{\beta}_0, \hat{\beta}_1)$ are unbiased, i.e.

$$E(\hat{\beta}_0) = \beta_0\ ,\ E(\hat{\beta}_1) = \beta_1.$$

And

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\ S_{xx}}\ ,\ Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}\ ,\ Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}_n\ \frac{\sigma^2}{S_{xx}}.$$

Finally,

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

Properties of the LS estimators

Theorem 2:

Under Assumption 1, the LS estimators $(\hat{\beta}_0, \hat{\beta}_1)$ are unbiased, i.e.

$$E(\hat{\beta}_0) = \beta_0 \ , \ E(\hat{\beta}_1) = \beta_1 .$$

And

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \ S_{xx}} \ , \ Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \ , \ Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}_n \frac{\sigma^2}{S_{xx}} .$$

Finally,

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

Confidence interval and hypothesis testing

- As before, we could also construct confidence interval and preform hypothesis testing, particularly about the regression parameters (β_0, β_1).
- We need to make more assumptions about the model.

Confidence interval and hypothesis testing

- As before, we could also construct confidence interval and preform hypothesis testing, particularly about the regression parameters (β_0, β_1).
- We need to make more assumptions about the model.

Assumption 2

Recall the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \quad i = 1, 2, \ldots, n.$$

- We assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid random variables from the Normal distribution $N(0, \sigma^2)$.

Assumption 2

Recall the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \quad i = 1, 2, \ldots, n.$$

- We assume that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid random variables from the Normal distribution $N(0, \sigma^2)$.

Consequences of Assumption 2

- Assumption 2 implies Assumption 1.
- Assumption 2 implies that the Y_i are independent and

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

Consequences of Assumption 2

- Assumption 2 implies Assumption 1.
- Assumption 2 implies that the Y_i are independent and

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

Properties of the LS estimators

Theorem 3:

(i) Under Assumption 2, we have that

$$\hat{\beta}_0 \sim N\Big(\beta_0, Var(\hat{\beta}_0)\Big) \ , \ \hat{\beta}_1 \sim N\Big(\beta_1, Var(\hat{\beta}_1)\Big)$$

where the variances $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$ are given in Theorem 2.

(ii) Also, $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$.

(iii) Moreover, $\hat{\sigma}^2$ is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$.

Properties of the LS estimators

Theorem 3:

(i) Under Assumption 2, we have that

$$\hat{\beta}_0 \sim N\Big(\beta_0, Var(\hat{\beta}_0)\Big) \quad , \quad \hat{\beta}_1 \sim N\Big(\beta_1, Var(\hat{\beta}_1)\Big)$$

where the variances $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$ are given in Theorem 2.

(ii) Also, $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$.

(iii) Moreover, $\hat{\sigma}^2$ is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$.

Properties of the LS estimators

Theorem 3:

(i) Under Assumption 2, we have that

$$\hat{\beta}_0 \sim N\Big(\beta_0, Var(\hat{\beta}_0)\Big) \ \ , \ \ \hat{\beta}_1 \sim N\Big(\beta_1, Var(\hat{\beta}_1)\Big)$$

where the variances $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$ are given in Theorem 2.

(ii) Also, $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-2)}$.

(iii) Moreover, $\hat{\sigma}^2$ is independent of both $\hat{\beta}_0$ and $\hat{\beta}_1$.

Confidence intervals for the regression parameters

- By Theorem 3, we have that

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}(\hat{\beta}_j)}} \sim t_{(n-2)} \quad , \quad j = 0, 1$$

where

$$\widehat{Var}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_{i=1}^{n} x_i^2}{n\, S_{xx}} \quad , \quad \widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

where $\hat{\sigma}^2$ is given in Theorem 1.

- Hence, we have Pivotal quantities that will be used to construct confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Confidence intervals for the regression parameters

- By Theorem 3, we have that

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}(\hat{\beta}_j)}} \sim t_{(n-2)} \quad , \quad j = 0, 1$$

where

$$\widehat{Var}(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_{i=1}^n x_i^2}{n\, S_{xx}} \quad , \quad \widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

where $\hat{\sigma}^2$ is given in Theorem 1.

- Hence, we have Pivotal quantities that will be used to construct confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Confidence intervals for β_0 and β_1

- A $100(1-\alpha)\%$ confidence interval for each β_0:

$$\left[\hat{\beta}_0 - t(n-2;\alpha/2)\times\sqrt{\widehat{Var(\hat{\beta}_0)}}\ ,\ \ \hat{\beta}_0 + t(n-2;\alpha/2)\times\sqrt{\widehat{Var(\hat{\beta}_0)}}\right]$$

- A $100(1-\alpha)\%$ confidence interval for each β_1:

$$\left[\hat{\beta}_1 - t(n-2;\alpha/2)\times\sqrt{\widehat{Var(\hat{\beta}_1)}}\ ,\ \ \hat{\beta}_1 + t(n-2;\alpha/2)\times\sqrt{\widehat{Var(\hat{\beta}_1)}}\right].$$

Confidence intervals for β_0 and β_1

- A $100(1-\alpha)\%$ confidence interval for each β_0:

$$\left[\hat{\beta}_0 - t(n-2;\alpha/2) \times \sqrt{\widehat{Var(\hat{\beta}_0)}}\ ,\ \ \hat{\beta}_0 + t(n-2;\alpha/2) \times \sqrt{\widehat{Var(\hat{\beta}_0)}}\right]$$

- A $100(1-\alpha)\%$ confidence interval for each β_1:

$$\left[\hat{\beta}_1 - t(n-2;\alpha/2) \times \sqrt{\widehat{Var(\hat{\beta}_1)}}\ ,\ \ \hat{\beta}_1 + t(n-2;\alpha/2) \times \sqrt{\widehat{Var(\hat{\beta}_1)}}\right].$$

McGill

Hypothesis testing for β_0

- We wish to test any of the following (pair of) hypotheses:

$$\text{Problem 1}: \quad \mathcal{H}_0 : \beta_0 \leq \beta_{00}, \quad \mathcal{H}_1 : \beta_0 > \beta_{00}$$

$$\text{Problem 2}: \quad \mathcal{H}_0 : \beta_0 \geq \beta_{00}, \quad \mathcal{H}_1 : \beta_0 < \beta_{00}$$

$$\text{Problem 3}: \quad \mathcal{H}_0 : \beta_0 = \beta_{00}, \quad \mathcal{H}_1 : \beta_0 \neq \beta_{00}$$

  where β_{00} is a pre-specified known value.

- Test Statistic:

$$T_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\widehat{Var(\hat{\beta}_0)}}}$$

Hypothesis testing for β_0

- We wish to test any of the following (pair of) hypotheses:

$$\text{Problem 1}: \quad \mathcal{H}_0 : \beta_0 \leq \beta_{00}, \quad \mathcal{H}_1 : \beta_0 > \beta_{00}$$

$$\text{Problem 2}: \quad \mathcal{H}_0 : \beta_0 \geq \beta_{00}, \quad \mathcal{H}_1 : \beta_0 < \beta_{00}$$

$$\text{Problem 3}: \quad \mathcal{H}_0 : \beta_0 = \beta_{00}, \quad \mathcal{H}_1 : \beta_0 \neq \beta_{00}$$

where β_{00} is a pre-specified known value.

- Test Statistic:

$$T_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{\widehat{Var}(\hat{\beta}_0)}}$$

Rejection regions

- At a significance level α, the rejection regions are:

$$\text{Problem 1}: \quad T_0 \geq t(n-2;\alpha)$$

$$\text{Problem 2}: \quad T_0 \leq -t(n-2;\alpha)$$

$$\text{Problem 3}: \quad |T_0| \geq t(n-2;\alpha/2)$$

where $t(n-2;\alpha)$ and $t(n-2;\alpha/2)$ are the upper-quantile of Student t distribution with $n-2$.

Rejection regions

- At a significance level α, the rejection regions are:

$$\text{Problem 1}: \quad T_0 \geq t(n-2;\alpha)$$

$$\text{Problem 2}: \quad T_0 \leq -t(n-2;\alpha)$$

$$\text{Problem 3}: \quad |T_0| \geq t(n-2;\alpha/2)$$

where $t(n-2;\alpha)$ and $t(n-2;\alpha/2)$ are the upper-quantile of Student t distribution with $n-2$.

Hypothesis testing for β_1

- We wish to test any of the following (pair of) hypotheses:

$$\text{Problem 1}: \quad \mathcal{H}_0 : \beta_1 \leq \beta_{10}, \quad \mathcal{H}_1 : \beta_1 > \beta_{10}$$

$$\text{Problem 2}: \quad \mathcal{H}_0 : \beta_1 \geq \beta_{10}, \quad \mathcal{H}_1 : \beta_1 < \beta_{10}$$

$$\text{Problem 3}: \quad \mathcal{H}_0 : \beta_1 = \beta_{10}, \quad \mathcal{H}_1 : \beta_1 \neq \beta_{10}$$

  where β_{10} is a pre-specified known value.

- Test Statistic:

$$T_1 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\widehat{Var}(\hat{\beta}_1)}}$$

Hypothesis testing for β_1

- We wish to test any of the following (pair of) hypotheses:

$$\text{Problem 1}: \quad \mathcal{H}_0 : \beta_1 \leq \beta_{10}, \quad \mathcal{H}_1 : \beta_1 > \beta_{10}$$

$$\text{Problem 2}: \quad \mathcal{H}_0 : \beta_1 \geq \beta_{10}, \quad \mathcal{H}_1 : \beta_1 < \beta_{10}$$

$$\text{Problem 3}: \quad \mathcal{H}_0 : \beta_1 = \beta_{10}, \quad \mathcal{H}_1 : \beta_1 \neq \beta_{10}$$

  where β_{10} is a pre-specified known value.

- Test Statistic:

$$T_1 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\widehat{Var(\hat{\beta}_1)}}}$$

McGill

Rejection regions

- At a significance level α, the rejection regions are:

$$\text{Problem 1}: \quad T_1 \geq t(n-2;\alpha)$$

$$\text{Problem 2}: \quad T_1 \leq -t(n-2;\alpha)$$

$$\text{Problem 3}: \quad |T_1| \geq t(n-2;\alpha/2)$$

where $t(n-2;\alpha)$ and $t(n-2;\alpha/2)$ are the upper-quantile of Student t distribution with $n-2$.

Rejection regions

- At a significance level α, the rejection regions are:

$$\text{Problem 1}: \quad T_1 \geq t(n-2;\alpha)$$

$$\text{Problem 2}: \quad T_1 \leq -t(n-2;\alpha)$$

$$\text{Problem 3}: \quad |T_1| \geq t(n-2;\alpha/2)$$

where $t(n-2;\alpha)$ and $t(n-2;\alpha/2)$ are the upper-quantile of Student t distribution with $n-2$.

- In the above hypothesis testing problems, the most common values for the pre-specified values are:

$$\beta_{00} = 0 \quad , \quad \beta_{10} = 0.$$

Sum of squares

- Total sum of squares or total variation in the response):

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y}_n)^2 = \text{TSS}$$

- Sum of squares of the residual errors:

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

- Total variation in the response explained by the fitted model:

$$\text{SSM} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_n)^2$$

- Note that: $y_i - \bar{y}_n = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}_n)$.
- If we squared both side and sum over $i = 1, 2, \ldots, n$, we get:

$$\text{SST} = \text{SSE} + \text{SSM}$$

- Note that: $y_i - \bar{y}_n = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}_n)$.

- If we squared both side and sum over $i = 1, 2, \ldots, n$, we get:

$$\text{SST} = \text{SSE} + \text{SSM}$$

Coefficient of determination

- The proportion (percentage) of the total variability in the response (y) explained or captured by the fitted regression model is:

$$R^2 = \frac{\text{SSM}}{\text{SST}} = \frac{\text{SSM}}{\text{SSE} + \text{SSM}}$$

Note: $0 \leq R^2 \leq 1$.

- It explains how strong the linear relationship between y and x is.

Coefficient of determination

- The proportion (percentage) of the total variability in the response (y) explained or captured by the fitted regression model is:

$$R^2 = \frac{\text{SSM}}{\text{SST}} = \frac{\text{SSM}}{\text{SSE} + \text{SSM}}$$

Note: $0 \leq R^2 \leq 1$.

- It explains how strong the linear relationship between y and x is.

Coefficient of determination

- The proportion (percentage) of the total variability in the response (y) explained or captured by the fitted regression model is:

$$R^2 = \frac{\text{SSM}}{\text{SST}} = \frac{\text{SSM}}{\text{SSE} + \text{SSM}}$$

  Note: $0 \leq R^2 \leq 1$.

- It explains how strong the linear relationship between y and x is.

- It can be shown that

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)\right]^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y}_n)^2\right]}.$$