Probability & Statistics

BITS Pilani
K K Birla Goa Campus

Dr. Jajati Keshari Sahoo
Department of Mathematics

Curve of Regression

Definition(Recall): Let (X, Y) be a 2-dimensional random variable. The graph of the mean value of Y given $X = x$, denoted by $\mu_{Y|x}$ is called the curve of regression of Y on X.

Linear Curve of Regression

Assumption:

X is a mathematical variable rather random variable.

Example: Suppose we are developing a model to describe the temperature of the dept water in the sea. Since temperature depends in part on the depth of the water, 2 variables are involved; X=depth, Y=temperature. We are not interested in making inferences on the depth of the water. With X(depth) fixed, the temperature measurements(Y) at different places (of depth X) varies.

Linear Curve of Regression

In the graph of $\mu_{Y|x}$, here Y depends on X and is called the *dependent* or *response* variable. The variable X whose value is used to help predict the behaviour of $Y|_x$ is called the *independent* or *predictor* variable or the *regressor*.

Sometimes the values of X used can be preselected and in this case the study is said to be *controlled*, otherwise it is called *observational* study.

Linear Curve of Regression

Linear Curve of Regression of Y on X is given by

$$\mu_{Y|x} = \beta_0 + \beta_1 x, \quad \beta_0, \beta_1 \in \mathbb{R}$$

β_0 denotes the intercept and

β_1 denotes the slope of the regression line.

We need to estimate the value of β_0 and β_1.

Linear Curve of Regression

Let $x_1, x_2, ..., x_n$ be n values of X (these points are assumed to be measured without error) We are concerned with the n random variables,

$$Y|_{x_1}, Y|_{x_2}, ..., Y|_{x_n} .$$

A random variable varies about its mean value.

Let $E_i = Y|_{x_i} - \mu_{Y|_{x_i}}$

$\Rightarrow Y|_{x_i} = E_i + \mu_{Y|_{x_i}}$

Linear Curve of Regression

We assume that the random difference E_i has mean 0.

Since we are assuming that the regression is linear we can conclude that

$$\mu_{Y|_{x_i}} = \beta_0 + \beta_1 x_i$$

Therefore,

$$\boxed{Y_i = \beta_0 + \beta_1 x_i + E_i} \rightarrow \text{ Simple Linear Regression Model}$$

where $Y_i = Y\big|_{x_i}$

where E_i is assumed to be a random variable with mean 0.

Linear Curve of Regression

So now we have a data consisting of a collection of n pairs (x_i, y_i), where x_i is an observed value of the variable X and y_i is the corresponding observation for the random variable Y_i. The observed value usually differs from its mean value by some random amount. This idea is mathematically expressed by writing

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

ε_i corresponds to E_i when Y_i takes value y_i

Linear Curve of Regression

We draw a scattergram (plot of the points in the *xy*-planes) and if a linear regression is applicable, the points should exhibit a linear trend. Since we do not know the true values for β_0 and β_1, we shall not know the true value for ε_i (vertical distance from the point (x_i, y_i) to the regression line).

Scatter Diagram

innovate achieve lead

Hours Studied *x*	Test Score y
4	31
9	58
10	65
14	73
4	37
7	44
12	60
22	91
1	21
17	84

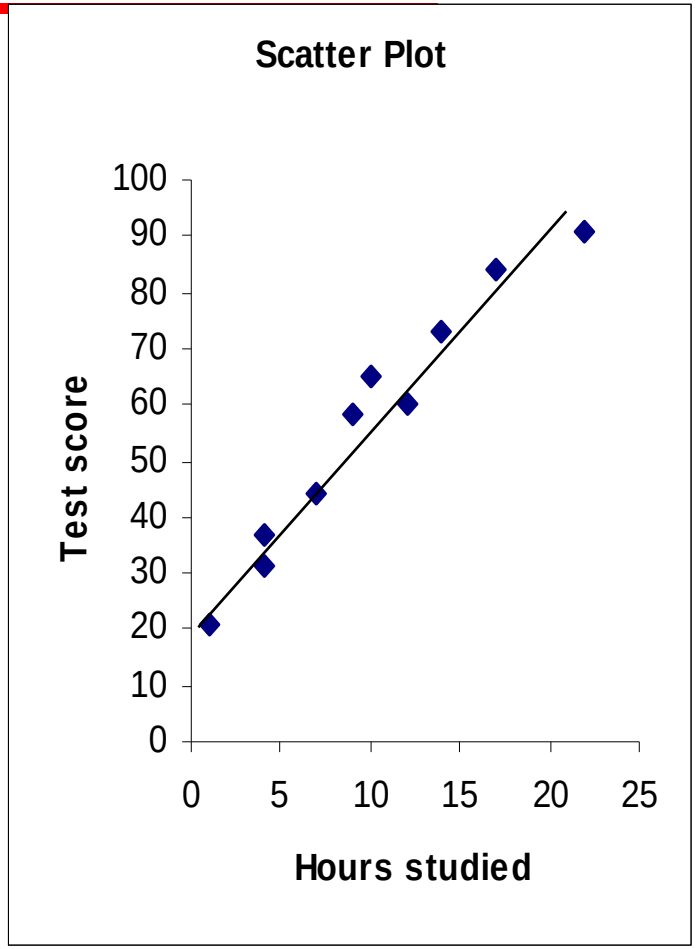

Figure 1: Data on hours studied and test scores

Linear Curve of Regression

Let b_0 and b_1 denote the estimates of β_0 and β_1 respectively. The estimated line of regression takes the form

$$\hat{\mu}_{Y|x} = b_0 + b_1 x$$

Let e_i be the vertical distance from a point (x_i, y_i) to the estimated regression line, then each data point satisfies the equation $y_i = b_0 + b_1 x_i + e_i$

The term e_i is called the *residual* or *residual error*.

Least-squares Estimation

The parameters β_0 and β_1 are estimated by method of least squares. In the sense that, from the many lines that can be drawn through a scatter diagram, we wish to pick the one that "best fits" the data. The fit is "best" when the chosen values of b_0 and b_1 minimizes the sum of the squares of the residuals. In this way we are picking the line that comes as close as possible to all the data points simultaneously.

Sum of squares of errors (SSE)

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right)^2 . \text{ Now,}$$

$$\frac{\partial (SSE)}{\partial b_0} = -2 \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right) = 0 \text{ and}$$

$$\frac{\partial (SSE)}{\partial b_1} = -2 \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right) x_i = 0$$

$$\Rightarrow \boxed{\begin{aligned} \sum_{i=1}^{n} y_i &= n b_0 + b_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i &= b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 \end{aligned}} \quad \text{(Normal equations)}$$

Solving which gives b_0 and b_1, the estimates of β_0 and β_1.

Least-Squares estimates and estimators

The least-squarea estimate for β_0 and β_1 are given by,

$$\hat{\beta}_1 = b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2} \text{ and } \hat{\beta}_o = b_0 = \bar{y} - b_1 \bar{x}.$$

The respective estimators are given by

$$B_1 = \hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2} \text{ and } B_0 = \hat{\beta}_o = \bar{Y} - B_1 \bar{x}.$$

Estimation of Linear regression line and conditional random variable

The least-squares estimator for the linear regression line is $\boxed{\hat{\mu}_{Y|x} = b_0 + b_1 x}$ and the estimator for the random variable Y given x is $\boxed{\hat{Y}_{|x} = \hat{\mu}_{Y|x} = b_0 + b_1 x.}$

Problem

Question:

Given that: n=10, $\sum x_i = 16.75$, $\sum x_i^2 = 28.64$,

$\sum y_i = 170$, $\sum y_i^2 = 2898$, $\sum x_i y_i = 285.625$,

Estimate the linear regression equation

$\mu_{Y|x} = \beta_0 + \beta_1 x$ and estimate Y when $x = 2$.

Model Assumptions:

Recall that a simple linear regression model is given by

$$Y_i = \beta_0 + \beta_1 x_i + E_i$$

We now assume the followings:

(1) The random variables Y_i are independently and normally distributed.

(2) The mean of Y_i is $\beta_0 + \beta_1 x_i$

(3) The variance of Y_i is σ^2

In otherwords: $Y_i \sim N(\mu = \beta_0 + \beta_1 x_i, \sigma^2) \Leftrightarrow E_i \sim N(0, \sigma^2)$

Properties

1. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)=0$

2. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(Y_i-\overline{Y}\right)=\sum_{i=1}^{n}\left(x_i-\overline{x}\right)Y_i$

3. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\left(Y_i-\overline{Y}\right)=\frac{\left(n\sum_{i=1}^{n}x_iY_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}Y_i\right)}{n}$

4. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\sum_{i=1}^{n}\left(x_i-\overline{x}\right)x_i$

5. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=\frac{\left(n\sum_{i=1}^{n}x_i^{\,2}-\left(\sum_{i=1}^{n}x_i\right)^2\right)}{n}$

Distribution Of B_1

Here we show that B_1 is normally distributed with mean β_1 and

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

$$B_1 = \frac{n\sum_{i=1}^{n} x_i Y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Distribution Of B_1

Since, $E(Y_i) = \beta_0 + \beta_1 x_i \Rightarrow E[B_1] = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$$\Rightarrow E[B_1] = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\beta_0 + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})x_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1$$

(by summation properties 1 and 4)

$\Rightarrow$ B_1 is an unbiased estimator of β_1.

Distribution Of B_1

$$\text{Var}\, B_1 = \text{Var}\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] = \left[\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2 \text{Var}\left[\sum_{i=1}^{n}(x_i-\bar{x})Y_i\right]$$

$$= \left[\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2 \left[\sum_{i=1}^{n}\text{Var}(x_i-\bar{x})Y_i\right]$$

$$= \left[\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2 \sum_{i=1}^{n}(x_i-\bar{x})^2\text{Var}(Y_i)$$

Distribution Of B_1

Since variance of Y_i is assumed to be σ^2 for each i, So

$$\text{Var } B_1 = \left[\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]^2 \sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\therefore B_1 \sim N\left[\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

Distribution of B_0

Note that: $\mathrm{Cov}(\bar{Y}, B_1)=0$ [see Ex. 14].

Since, $B_0 = \bar{Y} - B_1\bar{x}$ and which is a linear combination of independent normal random variables, and therefore is itself normally distributed.

$$E[B_0] = E\left[\bar{Y} - B_1\bar{x}\right] = E\left[\frac{Y_1 + Y_2 + \ldots + Y_n}{n} - B_1\bar{x}\right]$$

$$= \left[\frac{(\beta_0 + \beta_1 x_1) + (\beta_0 + \beta_1 x_2) + \ldots + (\beta_0 + \beta_1 x_n)}{n}\right] - \bar{x}\beta_1$$

$$= \beta_0 + \bar{x}\beta_1 - \bar{x}\beta_1 = \beta_0$$

$\therefore$ B_0 is an unbiased estimator for β_0.

Distribution of B_0

The variance of B_0 is given by

$$\text{Var}\, B_0 = \text{Var}[\bar{Y} - B_1\bar{x}] = \text{Var}[\bar{Y}] + \bar{x}^2\text{Var}[B_1]$$

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

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

Distribution of B_0

innovate achieve lead

Distribution of B_0

$$B_0 \sim N\left[\beta_0,\ \sigma^2\left(\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\right]$$

Estimator for σ^2

Estimator for σ^2 is given by,

$$S^2 = \hat{\sigma}^2 = \frac{SSE}{n-2}$$

Note: We divide by $(n-2)$ so that S^2 becomes an unbiased estimator for σ^2.

$$SSE = \sum_{i=1}^{n} (Y_i - B_0 - B_1 x_i)^2$$

Notations

innovate achieve lead

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

$$(2)\ S_{yy} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$(3)\ S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})$$

$$(4)\ B_1 = \frac{S_{xy}}{S_{xx}}$$

$$(5)\ SSE = S_{yy} - B_1 S_{xy}$$

$$(6)\ \sigma_{B_1}^2 = \sigma^2 / S_{xx}$$

$$(7)\ \sigma_{B_0}^2 = \frac{\left(\sum_{i=1}^{n} x_i^2 \sigma^2 \right)}{n S_{xx}}$$

Confidence interval estimation and Hypothesis testing

innovate achieve lead

Inferences about Slope:

A regresion line is said to be significant if we have sufficient evidence to conclude that the slope of the the true regression line is not zero.

Null hypothesis $H_0 : \beta_1 = 0.$

Since $B_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$

$\therefore \frac{B_1 - \beta_1}{\sigma / \sqrt{S_{xx}}}$ is standard normal.

Confidence interval estimation and Hypothesis testing

Note: The random variable $\frac{(n-2)S^2}{\sigma^2} = \frac{SSE}{\sigma^2}$

follows χ^2 distribution with $(n-2)$ degrees of freedom.

Therefore, we have

$$\frac{B_1 - \beta_1}{S \setminus \sqrt{S_{xx}}} = \frac{(B_1 - \beta_1) / (\sigma / \sqrt{S_{xx}})}{\sqrt{(n-2)S^2 / \sigma^2 (n-2)}}$$

has a T-distribution with $(n-2)$ degrees of freedom.

Confidence interval estimation for β_1

$100(1-\alpha)\%$ Confidence interval for β_1 :

$$\left[B_1 - t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}},\ B_1 + t_{\frac{\alpha}{2}} \frac{S}{\sqrt{S_{xx}}} \right]$$

Hypothesis testing on β_1

(Type-1): $H_0 : \beta_1 = \beta_1^0$ (known value)

$H_1 : \beta_1 > \beta_1^0$

(called Right -tailed test)

Example (Type-2): $H_0 : \beta_1 = \beta_1^0$

$H_1 : \beta_1 < \beta_1^0$

(called Left-tailed test)

Example (Type-3): $H_0 : \beta_1 = \beta_1^0$

$H_1 : \beta_1 \neq \beta_1^0$

(called Two-tailed test)

Hypothesis testing on β_1

Alternative hypothesis	*Critical Region:*
$\beta_1 < \beta_1^0$	$C=\{ T_{n-2} : T_{n-2} < -t_\alpha \}$
$\beta_1 > \beta_1^0$	$C=\{T_{n-2} : T_{n-2} > t_\alpha \}$
$\beta_1 \neq \beta_1^0$	$C=\{T_{n-1} : T_{n-2} < -t_{\alpha/2}$ or $T_{n-2} > t_{\alpha/2} \}$

Test statistic:

$$T_{n-2} = \frac{B_1 - \beta_1^0}{S / \sqrt{S_{xx}}}$$

Inferences about Intercept

$$B_0 \sim N\left(\beta_0,\ \frac{\sigma^2}{nS_{xx}}\sum_{i=1}^{n} x_i^2\right)$$

Then, the random variable

$$\frac{B_0 - \beta_0}{\left(\sigma\sqrt{\sum_{i=1}^{n} x_i^2}\right)/\sqrt{n.S_{xx}}}$$ is standard normal.

Thus $$\frac{B_0 - \beta_0}{\left(S\sqrt{\sum_{i=1}^{n} x_i^2}\right)/\sqrt{n.S_{xx}}}$$ follows T-distribution

with $(n-2)$ degrees of freedom.

Confidence interval estimation for β_0

$100(1-\alpha)\%$ Confidence interval for β_0 :

$$\left[B_0 - t_{\frac{\alpha}{2}} \frac{S\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{nS_{xx}}},\ B_0 + t_{\frac{\alpha}{2}} \frac{S\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{nS_{xx}}} \right]$$

Hypothesis testing on β_0

(Type-1): $H_0 : \beta_0 = \beta_0^0$ (known value)

$H_1 : \beta_0 > \beta_0^0$

(called Right -tailed test)

Example (Type-2): $H_0 : \beta_0 = \beta_0^0$

$H_1 : \beta_0 < \beta_0^0$

(called Left-tailed test)

Example (Type-3): $H_0 : \beta_0 = \beta_0^0$

$H_1 : \beta_0 \neq \beta_0^0$

(called Two-tailed test)

Hypothesis testing on β_0

Alternative hypothesis	*Critical Region:*
$\beta_0 < \beta_0^0$	$C=\{ T_{n-2} : T_{n-2} < -t_\alpha \}$
$\beta_0 > \beta_0^0$	$C=\{T_{n-2} : T_{n-2} > t_\alpha \}$
$\beta_0 \neq \beta_0^0$	$C=\{T_{n-1} : T_{n-2} < -t_{\alpha/2}$ or $T_{n-2} > t_{\alpha/2} \}$

Test statistic:

$$T_{n-2} = \frac{B_0 - \beta_0^0}{\left(S\sqrt{\sum_{i=1}^{n} x_i^2} \right) / \sqrt{nS_{xx}}}$$

Problem

Given $n = 10$, $\sum_{i=1}^{n} x_i = 16.75$, $\sum_{i=1}^{n} y_i = 170$,

$\sum_{i=1}^{n} x_i^2 = 28.64$, $\sum_{i=1}^{n} y_i^2 = 2898$, $\sum_{i=1}^{n} x_i y_i = 285.625$

(a) Estimate the linear regression equation, also estimate the average value of y when $x = 2$.

(b) Obtain 99% confidence interval for β_0

(c) Test the hypothesis: $\beta_1 = 0$ at 1% level.

Solution

$\bar{x} = 1.675,\ \bar{y} = 17,$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 28.64 - \frac{(16.75)^2}{10} = 0.584$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 0.875$$

Similarly, $S_{yy} = 8$

$$\therefore B_1 = \frac{S_{xy}}{S_{xx}} = 1.498, \quad B_0 = \bar{y} + B_1\bar{x} = 14.49$$

$\therefore \boxed{\hat{Y} = \hat{\mu}_{Y|_x} = 14.5 + 1.5x}$ and hence when $x = 2,\ \hat{\mu}_{Y|_x} = 17.5$

Solution

Confidence Interval for β_0 : $\boxed{B_0 \pm t_{\alpha/2} \frac{S\sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{nS_{xx}}}}$

$$SSE = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 6.69$$

$$S = \sqrt{\frac{SSE}{n-2}} = 0.914$$

$t_{\alpha/2} = 3.355$ corresponding to $(n-2)$.

$\therefore$ 99% Confidence interval = $[7.7, 21.3]$

Solution

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

$$\alpha = 0.01$$

$$t_{n-2} = \frac{B_1}{S / \sqrt{S_{xx}}} = 1.25$$

$t_{\alpha/2}$ at 0.01 level with $(n-2)$ degrees of freedom is $3.355 > 1.25$, therefore we accept H_0.

So the regression line is not significant.

Inferences about Estimated Mean

The point estimate for $\mu_{Y|_x}$, is given by

$$\hat{\mu}_{Y|_x} = B_0 + B_1 x = (\overline{Y} - B_1 \overline{x}) + B_1 x$$

$$\Rightarrow \boxed{\hat{\mu}_{Y|_x} = \overline{Y} + B_1 (x - \overline{x})}$$

Since $\overline{Y}$ and $B_1$are both normally distributed which implies $\hat{\mu}_{Y|_x}$ is normal.

Distribution of $\mu_{Y|x}$

$$E(\hat{\mu}_{Y|_x}) = E(B_0) + xE(B_1) = \beta_0 + \beta_1 x = \mu_{Y|_x}$$

$$\Rightarrow \boxed{\hat{\mu}_{Y|_x} \text{ is an unbiased estimator of } \mu_{Y|_x}.}$$

$$\text{Var}(\hat{\mu}_{Y|_x}) = Var[\bar{Y} + B_1(x - \bar{x})] = \text{Var}(\bar{Y}) + (x - \bar{x})^2 \frac{\sigma^2}{S_{xx}}$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right)$$

$$\therefore \hat{\mu}_{Y|_x} \sim N\left\{ \mu_{Y|_x},\ \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \right\}$$

Inferences about Estimated Mean

Since $\dfrac{\hat{\mu}_{Y|_x} - \mu_{Y|_x}}{\sigma\sqrt{\dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{S_{xx}}}}$ is standard Normal.

the random variable $\dfrac{\hat{\mu}_{Y|_x} - \mu_{Y|_x}}{S\sqrt{\dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{S_{xx}}}}$

follows T - distribution with $(n-2)$ degrees of freedom.

Confidence Interval on $\mu_{Y|_x}$

100(1-α)% Confidence interval for $\mu_{Y|x}$:

$$\left[\hat{\mu}_{Y|_x} - t_{\alpha/2} S\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}},\ \hat{\mu}_{Y|_x} + t_{\alpha/2} S\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}\right]$$

Problem

Given $n = 10$, $\sum_{i=1}^{n} x_i = 16.75$, $\sum_{i=1}^{n} y_i = 170$,

$\sum_{i=1}^{n} x_i^2 = 28.64$, $\sum_{i=1}^{n} y_i^2 = 2898$, $\sum_{i=1}^{n} x_i y_i = 285.625$

(a) Find 90% confidence interval for $\mu_{Y|_{x=2}}$.

(b) Obtain 99% confidence interval for β_1

(c) Test the hypothesis: $\beta_0 = 14.5$ at 1% level.

Solution

Confidence interval on $\mu_{Y|_x}$ is given by,

$$\hat{\mu}_{Y|_x} \pm t_{\alpha/2} S \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

$x - \bar{x} = 2 - 1.675 = 0.325, \quad S_{xx} = 0.584, \quad S = 0.914,$

$\hat{\mu}_{Y|_{x=2}} = 17.5,$

$n = 10, \; t_{\alpha/2} = 1.860$

$\therefore$ 90% Confidence interval=[16.59,18.402].