ASSIGNMENT - 3

MATH_564-Regression

(1) (6 points) Recalled that S_{xy} is defined as

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

Show that S_{xy} can also be written as

$$S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})y_i$$

therefore

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{S_{xx}} = \sum_{i=1}^{n} c_i y_i$$

where

$$c_i = \frac{x_i - \overline{x}}{S_{xx}}, \quad i = 1, \cdots, n.$$

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

Homework -3

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

we have to show that S_{xy} can also be written as

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

we know that from $\hat{\beta_1}$

$$\hat{\beta_1} = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n} \frac{\cancel{(x_i - \bar{x}) y_i}}{\cancel{S_{xx}}}$$

$$\hat{\beta_1} = \frac{S_{xy}}{S_{xx}} \Rightarrow \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{S_{xx}}$$

$$\Rightarrow \sum_{i=1}^{n} c_i y_i$$

where we know that c_i is written as

$$c_i = \frac{x_i - \bar{x}}{S_{xx}}$$

where $i = 1, 2, \ldots . n.$

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

Rewrithe the S_{xy} as

$$S_{xy} = \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i - (x_i - \bar{x}) \bar{y} \right]$$

we know that $\boxed{\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}}$

Substitue $\bar{y}$ is expression.

$$S_{xy} = \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i - (x_i - \bar{x}) \sum_{i=1}^{n} y_i/n \right]$$

$$\Rightarrow \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i \right] - \sum_{i=1}^{n} (x_i - \bar{x}) \sum_{i=1}^{n} y_i/n$$

$$\Rightarrow \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i \right] - \bar{y} \sum_{i=1}^{n} (x_i - \bar{x})$$

$$\bar{x} = \sum_{i=1}^{n} \frac{1}{n} x_i = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\Rightarrow n\bar{x} = \sum_{i=1}^{n} x_i$$

$$\Rightarrow \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i \right] - \bar{y} \sum_{i=1}^{n} \left[x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right]$$

$$\Rightarrow \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i \right] - \bar{y} \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \frac{1}{n} \sum_{i=1}^{n} x_i \right]$$

$$\Rightarrow \sum_{i=1}^{n} \left[(x_i - \bar{x}) y_i \right] - \bar{y} \left[\sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \cdot n \right]$$

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i - \bar{y} \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \right]$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i - \bar{y}(0) \rightarrow \text{becomes zero.}$$

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

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{S_{xx}} \Rightarrow \sum_{i=1}^{n} c_i y_i$$

$$c_i = \frac{x_i - \bar{x}}{S_{xx}} \quad \text{where } i = 1, 2, 3 \ldots n.$$

(2) (6 points) In our class discussion and on pages 37 and 38 of the August 19th version of the notes, we established that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are unbiased estimators of β_0 and β_1, respectively. This means that:

$$E(\widehat{\beta}_0) = \beta_0 \quad \text{and} \quad E(\widehat{\beta}_1) = \beta_1.$$

In the proof, we assumed that $E(\epsilon_i) = E(\epsilon \mid x_i) = 0$. However, what if $E(\epsilon_i) = E(\epsilon \mid x_i) = \alpha$, where α is a non-zero constant? How would this assumption affect the expectations of $\widehat{\beta}_0$ and $\widehat{\beta}_1$?

To receive full credit, you need to support your claim with a clear and rigorous mathematical proof.

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

2nd Solution

The true simple linear regression Model

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

$$i = 1, 2, 3 \ldots n$$

where y_i = dependent

x_i = Independent

β_0 = Intercept

β_1 = Slope

ε_i = Random error term

A20554334
AKHIL KUMAR MARNI

Assume the $E(\varepsilon_i)$ is 0.

$\hat{\beta_0}$ and $\hat{\beta_1}$ are unbiased estimators of β_0 and β_1

know that $E(\varepsilon_i) = E(\varepsilon(x_i)) = \alpha$

α is non zero constant

$$E(y_i) = E[\beta_0 + \beta_1 X_i + \varepsilon_i]$$

$$\Rightarrow E(\beta_0) + E(\beta_1 x_i) + E(\varepsilon_i)$$

β_0 & β_1 are constants

so Expectation of constants remains constants.

$$\Rightarrow E(\beta_0) = \beta_0$$

$$\Rightarrow E(\beta_1 x_i) = \beta_1 x_i \quad \text{and}$$

weknow $E(\varepsilon_i) = \alpha$

$$\boxed{y_i = \beta_0 + \beta_1 X_i + \alpha}$$

A20554334
AKHIL KUMAR MARNI

$$\hat{\beta_1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x}) x_i}$$

$$E\left[\hat{\beta_1}\right] = E\left[\frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x}) x_i}\right]$$

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

$$\Rightarrow \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x}) x_i} \sum_{i=1}^{n} (x_i - \bar{x}) \left[\beta_0 + \beta_1 x_i + \alpha\right]$$

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

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

$$\Rightarrow (0) \beta_0 + \beta_1 (1) + \alpha(0)$$

$E(\hat{\beta_1}) \Rightarrow \beta_1$

So $\hat{\beta_1}$ unbiased estimator of β_1.

A20554334

AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

when we know that. $\hat{\beta_0} = (\bar{y} - \hat{\beta_1} x)$

$\Rightarrow E(\hat{\beta_0}) \Rightarrow E[\bar{y} - \hat{\beta_1}\bar{x}]$

$\Rightarrow E(\bar{y}) - \bar{x} \; E[\hat{\beta_1}]$

$\Rightarrow \beta_0 + \cancel{\beta_1 \bar{x}} + \alpha - \cancel{\bar{x}\beta_1}$

$\Rightarrow \beta_0 + \alpha$

So $\hat{\beta_0}$ is a biased estimator

Because of Biased $E(\varepsilon_i) = \alpha$.

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

5. (6 points) Consider the true relationship between x and y given by

$$y = \beta_0 + \beta_1 x + \epsilon.$$

When we collect a data set (x_i, y_i), $i = 1, \cdots n$, this relationship can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i.$$

What assumptions about the error terms ϵ_i are necessary to obtain the following results?

(a) $E(\widehat{\beta}_0) = \beta_0$

(b) $\text{Var}(\widehat{\beta}_1) = \sigma^2 / S_{xx}$

(c) $\widehat{\beta}_1 \sim \text{N}(\beta_1, \sigma^2 / S_{xx})$

A20554334
AKHIL KUMAR MARNI

5th Solution

(i) $$y = \beta_0 + \beta_1 X + \varepsilon$$

The expected value of error term is zero

$$E(\varepsilon_i) = E\left[\varepsilon_i / x = x_i\right] = 0$$

we are assuming that error term do not systematically overestimate or underestimate values of y.

error terms ε_i have zero mean.

This ensures that the expected value of the residuals is zero

$$E(\hat{\beta_0}) = \beta_0$$

(ii) $Var(\hat{\beta_1}) = \sigma^2 / S_{xx}$

The variance of error (ε_i) term is constant for all $i = 1, 2 \ldots n$.

$$Var(\varepsilon_i) = \sigma^2 \text{ for all } i = 1, 2 \ldots n.$$

The error terms (ε_i) are uncorrelated

This means that error terms (ε_i) have same variance for all the data points (or) observations that are properly calculated as σ^2 / S_{xx}.

A20554334
AKHIL KUMAR MARNI

Due September 20th 2024, Friday, 11:59pm

(iii) The error terms (ϵ_i) are normally distributed

(i.e) $\epsilon_i \sim N(0, \sigma^2)$

and the error terms are uncorelated.

$$\text{cov}(\epsilon_i, \epsilon_j) = 0 \quad \text{for } i \neq j$$

The assumption ensures that the error terms (ϵ_i) are independent.

which needed to make valid inferences and calculate the variance correctly.