STAT 231

Nov 21, 2016

Tutorial Quiz: 30^{th}

Tutorial this week → 6-00 (DC 1351)

Roadmap

- 5 min recap
- CI for β, HT for β
- Confidence Interval for the mean response $\mu(x)$, given x.
- Prediction interval for Y_{new}, given $x = x_{new}$

---

* Checking for model assumptions

Simple Linear Regression Model

Y = Response Variate = STAT 231 score.

X = Explanatory " = STAT 230 score.

Assumptions (i) Given X, Y_i's are Normally distributed with constant variance σ^2.

(ii) The average value of Y

$$\mu(x) = \alpha + \beta X$$

STATISTICAL Model

$$Y_i \sim G(\alpha + \beta x_i, \sigma) \quad \text{—— (1)}$$

$i = 1, \ldots, n$

$\Updownarrow$

Y_i's independent

$$Y_i = \underbrace{\alpha + \beta x_i}_{} + R_i \,,$$

$R_i \sim G(0, \sigma)$

R_i's independent

Important questions in a regression model.

(i) Given your data, find the least square equation. (OR Find $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, s)

(ii) Find the C.I (95%) for β.

(iii) Test the hypothesis that

$$H_0 : \beta = \beta_0$$

$H_0 : \beta = 0 \Rightarrow$ Testing for linear relationship between X and Y.

(iv) Given the value of x, find the C.I for $\alpha + \beta x = \mu(x)$

Example: $x = 75$

Objective: To find a range for the average STAT 231 score $(\alpha + \beta \cdot 75)$ with a high degree of confidence:

(v). Predict the value of the r.v. Y_{new} when $x = x_{new}$.

PREDICTION INTERVAL.

Example $x_{new} = 75$

To find an interval for a student's STAT 231 score.

(vi) Are the assumptions made about the model validated by the data?

Sample of $\underline{n = 30}$ students

$(x_1, y_1), \ldots \quad (x_{30}, y_{30})$ were recorded.

$\underline{\bar{x} = 76.7333}$ $\qquad$ $\underline{\bar{y} = 72.2333}$

$S_{yy} = 5135.8667$ $\qquad$ $S_{xx} = 5106.8667$

S_{xy}

$S_{yy} = 7585.3667$

$S_{xy} = 5106.8667$

$S_{xx} = 5135.8667$

Data

$S_{xx} = \sum (x_i - \bar{x})^2$

$S_{yy} = \sum (y_i - \bar{y})^2$

$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$

Q1 Find $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$, s, and the least squared equation

$$\hat{\beta} = S_{xy} / S_{xx} = 0.9944$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$= 72.233 - 0.\frac{5106.8667}{5135.8667} \cdot 76.333$$

$$= -4.0667.$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\left[S_{yy} - \hat{\beta} S_{xy}\right]}$$

$$s = \sqrt{\frac{1}{n-2}\left[S_{yy} - \hat{\beta} S_{xy}\right]}$$

$$= \underline{9.4630}.$$

<u>Least Square equation</u>

$$y = \hat{\alpha} + \hat{\beta} x.$$

$$\boxed{y = -4.0667 + 0.9944\, x}$$

Q2 Find the 95% Confidence Interval for β.

$$\tilde{\beta} \sim G\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right) \quad \text{(Result)}$$

$$\frac{\tilde{\beta} - \beta}{\frac{\sigma}{\sqrt{S_{xx}}}} = Z$$

Pivotal Quantity

Pivotal Distribution

$$\boxed{\frac{\tilde{\beta} - \beta}{S/\sqrt{S_{xx}}} \sim T_{n-2}}$$

Confidence Interval.

$$\boxed{\hat{\beta} \pm t^{\alpha} \frac{s}{\sqrt{S_{xx}}}}$$

$\rightarrow df = n-2.$

What is the value of t^{α}?

$$\left.\begin{array}{l} \text{Row} = n-2 = 30-2 = 28 \\ \text{Column} = 0.975 \end{array}\right\} \Rightarrow t^{\alpha} = 2.048$$

CI: $0.9944 \pm 2.0484 \cdot \frac{9.4630}{\sqrt{5135.8667}}$

$= [0.7239, 1.2648]$

Q3: Test $H_0: \beta = 0$

Test statistic $D = \left| \dfrac{\hat{\beta} - \beta_0}{S/\sqrt{S_{xx}}} \right|$.

Calculate the value of d:

$$d = \left| \frac{\hat{\beta} - \beta_0}{S/\sqrt{S_{xx}}} \right| = 7.5304.$$

p-value:

$$P(D \geq d) = P(|T_{28}| \geq 7.5304)$$
$$\approx 0$$

Strong evidence against the null hypothesis

$H_0 : \beta = 1$

We have no evidence against $H_0 : \beta = 1$

Q4 : Suppose $x = 75$. Find a 95% C.I for the mean STAT 231 score $\mu(75) = \alpha + \beta \cdot 75$.

Define $\tilde{\mu}(75) = \tilde{\alpha} + \tilde{\beta} \cdot 75$.

$\tilde{\alpha} = \bar{Y} - \tilde{\beta}\,\bar{x}$

$\tilde{\beta} = \dfrac{S_{xY}}{S_{xx}}$.

$\tilde{\mu}$ is a linear function of $\tilde{\alpha}$ and $\tilde{\beta}$, it will be Gaussian.

Result

$$\tilde{\mu}(x) \sim G\left(\mu(x), \sigma\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}\right)$$

$x = 75$ in this case.

$$\cancel{\frac{\tilde{\mu}(x)-\mu(x)}{S\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} = T_{n-2}}$$

$$\frac{\tilde{\mu}(x)-\mu(x)}{\sigma\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} = Z.$$

$$\frac{\tilde{\mu}(x)-\mu(x)}{S\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \sim T_{n-2}$$

P.Q

P.D.

Confidence interval:

$$\boxed{(\hat{\alpha}+\hat{\beta}x) \pm t^{\alpha} s\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}}$$

$[66.9, 74.1]$