# Regression Analysis Quiz-1

Likhit Garimella
lg836

A1) To prove : $\Sigma(Y_i - \bar{Y}) = 0$

$\Rightarrow$ from def'n of mean $(\bar{Y})$

$$\bar{Y} = \frac{Y_1 + Y_2 + Y_3 + \cdots + Y_n}{n}$$

$\Rightarrow$ rewrite for $\Sigma(Y_i - \bar{Y})$

$$\Sigma(Y_i - \bar{Y}) = (Y_1 - \bar{Y}) + (Y_2 - \bar{Y}) + (Y_3 - \bar{Y}) \cdots + (Y_n - \bar{Y})$$

$\Rightarrow$ expand

$$\Sigma(Y_i - \bar{Y}) = Y_1 - \bar{Y} + Y_2 - \bar{Y} + Y_3 - \bar{Y} + \cdots + Y_n - \bar{Y}$$

$\Rightarrow$ rearrange

$$\Sigma(Y_i - \bar{Y}) = (Y_1 + Y_2 + Y_3 + \cdots + Y_n) - (\bar{Y} + \bar{Y} + \bar{Y} + \cdots + \bar{Y})$$

$\Rightarrow$ simplify

$$\Sigma(Y_i - \bar{Y}) = (Y_1 + Y_2 + \cdots + Y_n) - n * (\bar{Y})$$

$\Rightarrow$ substitute $\bar{Y}$

$$\Sigma(Y_i - \bar{Y}) = (Y_1 + Y_2 + \cdots + Y_n) - n * \left(\frac{Y_1 + Y_2 + \cdots + Y_n}{n}\right)$$

⇒ simplify

$$\Sigma(Y_i - Y) = Y_1 + Y_2 + \cdots + Y_n - (Y_1 + Y_2 + \cdots + Y_n)$$

⇒ terms cancel out

$$\Rightarrow \boxed{\Sigma(Y_i - Y) = 0}$$

⇒ sum of deviations from mean = 0.

---

A4) Obtaining an estimator for $\sigma^2$ (variance of a population) is important because:-

⇒ In statistical analyses, we need to perform hypothesis tests to determine the difference between various groups or b/w variables in a relationship. The estimation of $\sigma^2$ is required to calculate such test statistics such as t-statistic or F-statistic.

⇒ In regression analysis, obtaining an estimator for $\sigma^2$ is required for model assessment. The estimated variance helps evaluate the goodness of fit of model. It allows us to assess how well the model explains the observed data.

⇒ Estimating $\sigma^2$ is required for constructing confidence intervals. These provide a range of values for a population parameter. Confidence intervals are used for precision of our estimates.

⇒ Estimating $\sigma^2$ allows us to obtain more efficient estimators for other parameters of interest. With this, we can improve the precision of our parameter estimates & reduce uncertainity.

---

A8) When p is low, $H_0$ must go.

---

A6) → computed R-square value represents the proportion of total variation in the dependent variable.

→ t-ratio is used to test the significance of the regression relationship between the independent variable & dependent variable.

In this case :-

⇒ from the given information — to the t-ratio for testing whether a regression relation exists is not significant at $\alpha = 0.05$.

⇒ Hence, there is no linear relationship b/w the 2 variables.

---

A5) $Y_i = \beta X_i + \varepsilon_i$

$\beta$ value that minimize SSE

$\beta(-1, 0, 1)$

(-2, -6)
(-1, -4)
(0, 3)
(0, 0)
(1, 2)
(2, 11)

For $\beta = -1$:

$SSE(-1) = 117$ (after calc)

For $\beta = 0$:

$SSE(0) = 186$ (after calc)

For $\beta = 1$:

$SSE(1) = 165$ (after calc)

SSE is smallest when $\beta = 1$

$\therefore$ least sq. est. of $\beta = 1$.

A3)

→ Max. $R_2$ that can be computed is <1 due to limited amount of data.

→ With only 3 obs. at $x_4$, model's ability to acc. capture reln & explain var. is limited.

→ Need More comprehensive data set & larger no of observations.

⇒ limited amount of data leads to inability to capture the complete relationship.

---

A2)

$$\hat{Y_i} = \beta_0 + \beta_1 X_i$$

$$e_i = Y_i - \hat{Y_i}$$

$$SSR(\beta_1) = \Sigma (e_i)^2$$

$$SSR(\beta_1) = \Sigma (Y_i - \hat{Y_i})^2$$

$$SSR(\beta_1) = \Sigma (Y_i - (\beta_0 + \beta_1 X_i))^2$$

$$d(SSR(\beta_1)) / d\beta_1 = -2 \Sigma X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\Sigma X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\sum X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum (X_i)^\wedge 2 = 0$$

$$\beta_1 \sum (X_i)^2 = \sum X_i Y_i - \beta_0 \sum X_i$$

Least square est. of $\beta_1 \Rightarrow$

$$\beta_1 = \frac{\left(\sum X_i Y_i - \beta_0 \sum X_i\right)}{\sum (X_i)^2}$$