stat151A HW1

1. (a) $\hat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ $\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x} = \bar{y} - \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\bar{x}$

(b) $E(y_i|x_i) = \hat{\beta_0} + \hat{\beta_1}x_i$

$E(x_i|y_i) = \frac{1}{\beta_1}y_i - \frac{\beta_0}{\beta_1} = \alpha_1 y_i + \alpha_0$

thus. $\alpha_1 = \frac{1}{\beta_1} = \frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})\cdot y_i}$

(c) yes, As can be seen in (b) $\hat{\alpha_1} = \frac{1}{\hat{\beta_1}}$

(d). See R code.

3. (b). $math10 = \beta_0 + \beta_1 \log_{10}(expand) + \varepsilon$

if expand increases by 10%.

$math10 = \beta_0 + \beta_1 \log_{10}(1.1\,expand) + \varepsilon$

$= math10 + 0.1\beta_1$

Scanned by CamScanner

4. (a) $\hat{\beta_0}+\hat{\beta_1}a = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}a + \bar{y} - \hat{\beta_1}\bar{x}$

$= \bar{y} - \hat{\beta_1}\bar{x} + \hat{\beta_1}a$

$= \bar{y} - \hat{\beta_1}(a-\bar{x}) = \frac{\sum y_i}{n} + \frac{\sum(a-\bar{x})(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$

(b) $var(\hat{\beta_0}+\hat{\beta_1}a) = var\sum_{i=1}^{n} y_i\left(\frac{1}{n} + \frac{(a-\bar{x})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$

$= \sum_{i=1}^{n} var(y_i)\left[\frac{1}{n} + \frac{(a-\bar{x})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2$

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

$= \sigma^2\sum_{i=1}^{n}\left(\frac{1}{n^2} + \frac{2(a-\bar{x})(x_i-\bar{x})}{n\sum_{i=1}^{n}(x_i-\bar{x})^2} + \frac{(a-\bar{x})(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^4}\right]$

$= \sigma^2\sum_{i=1}^{n}\left(\frac{1}{n^2} + 0 + \frac{(a-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$

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

(c). let $t = var(\hat{\beta_0}+\hat{\beta_1}a)$

when $a=\bar{x}$, $var(\hat{\beta_0}+\hat{\beta_1}a)$ can acheive its smallest value.

since $\frac{\sigma^2(a-\bar{x})}{\sum(x_i-\bar{x})^2} = 0$.

and $var(\hat{\beta_0}+\hat{\beta_1}a) = \frac{\sigma^2}{n}$

Scanned by CamScanner

6. (a) $\hat{y} = \hat{\beta_1} x$

$$Q(\hat{\beta_1}) = \sum_{i=1}^{n} (y_i - \hat{\beta_1} x_i)^2$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta_1} x_i) \cdot x_i = 0.$$

$$\Rightarrow \quad \sum (y_i x_i - \hat{\beta_1} x_i^2) = 0.$$

$$\sum y_i x_i = \sum \hat{\beta_1} x_i^2$$

$$\sum y_i x_i = \hat{\beta_1} \sum x_i^2$$

$$\hat{\beta_1} = \frac{\sum y_i x_i}{\sum x_i^2}$$

(b). To show $\hat{\beta_1}$ is unbiased. $E(\hat{\beta_1} | x) = \beta_1$.

$$E(\hat{\beta_1} | x) = E\left(\frac{\sum y_i x_i}{\sum x_i^2}\right).$$

$$= \frac{\sum x_i E(y_i)}{\sum x_i^2} = \frac{\sum x_i E(\beta_1 x_i)}{\sum x_i^2} = \frac{\beta_1 \sum x_i x_i}{\sum x_i^2}$$

$$= \frac{\beta_1 \sum x_i^2}{\sum x_i^2} = \beta_1$$

thus proved.

(c) $var(y|x) = \sigma^2$ by assumption.

$$var(\hat{\beta_1}) = var(\hat{\beta_1} | x) = var\left(\frac{\sum y_i x_i}{\sum x_i^2} \,\middle|\, x\right)$$

$$= \frac{\sum x_i^2 var(y_i)}{(\sum x_i^2)^2} = \frac{\sum x_i^2}{(\sum x_i^2)^2} \cdot \sigma^2 = \frac{1}{\sum x_i^2} \cdot \sigma^2.$$

7. (a). $E(\hat{\beta_0}|x) = E(\bar{y} - \hat{\beta_1}\bar{x})$.

$$E(\hat{\beta_1}|x) = E\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\Big|x\right)$$

$$= \frac{\sum(x_i-\bar{x})E(y_i|x_i)}{\sum(x_i-\bar{x})^2}$$

$$E(y_i|x_i) = \beta_0 + \beta_1 x$$

$$E(\hat{\beta_1}|x) = \frac{\sum(x_i-\bar{x})(\beta_0+\beta_1 x_i)}{\sum(x_i-\bar{x})^2} = \beta_1$$

thus $\hat{\beta_1}$ is unbiased estimator.

$$E(\hat{\beta_0}|x) = E\left(\frac{1}{n}\sum y_i\right) - \hat{\beta_1}\bar{x}$$

$$= \frac{1}{n}\sum E(y_i) - \beta_1\bar{x}$$

$$= \frac{1}{n}\sum(\beta_0+\beta_1 x_i) - \beta_1\bar{x}$$

$$= \beta_0 + \beta_1 \cdot \frac{1}{n}\sum x_i - \beta_1\bar{x}$$

$$= \beta_0 + \beta_1 \cdot \cancel{\frac{1}{n}\sum x_i - \beta_1 \cdot \frac{1}{n}\sum x_i} = \beta_0.$$

thus $\hat{\beta_0}$ is unbiased estimator.

(c) $\begin{cases} N(0,25) & x_i \le 65 \\ T_3(0,10) & 65 < x_i \le 70 \\ \text{uniform}[-8,8] & x_i > 70 \end{cases}$

variance from the three intervals are not consistent.

The assumption of homoskedasticity suggests that the variance is consistent for all x_is. Thus in this case, the assumption of homoskedasticity is not valid.

Scanned by CamScanner