DÁM ƯỚC MƠ

Problem set 2

4.2c **Problem**: Change the units of the variables => how does it affect the regression estimates? For the record, the original eqn is:

$$\hat{W} = -79.24 + 4.16h \quad , R^2 = 0.72, \; SER = 12.6.$$

Solution: Instead of converting the units and performing the calculations, it would be instructive to estimate the changes in a more systematic way, namely, by using dimensional analysis (We borrow from other sciences).

To start off, we begin with R^2. Recall its formula:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{y}_i - \hat{y})^2}{\sum(y_i - \hat{y})^2}.$$

Suppose that y has dim of γ, then R^2's dim can be derived as:

$$\boxed{[R^2] = \frac{\sum(\gamma-\gamma)^2}{\sum(\gamma-\gamma)^2} = 0}$$

which means that R^2 is dimensionless! Therefore, any change in the units of the variables won't affect R^2.

Similarly, SER is also dimensionless

$$\boxed{[SER] = [\sqrt{s_{\hat{u}}^2}] = \left[\sqrt{\frac{1}{n-2} \sum(y_i - \hat{y}_i)^2}\right] = 0}$$

Let us proceed to evaluate the changes in $\hat{\beta}_1$ and $\hat{\beta}_0$, respectively, next:

Suppose that we wanna transform (x_i^1, y_i^1) to (x_i^2, y_i^2). The two points will be related by the eqns:

$$\begin{cases} x_i^1 = m x_i^2 \\ y_i^1 = n y_i^2 \end{cases} \quad (m, n \in \mathbb{R} \setminus \{0\}) \; (*)$$

Also recall that the slope of $\hat{\beta}_1^1$ and $\hat{\beta}_1^2$ can be calculated as:

$$\hat{\beta}_1^1 = \frac{\sum(x_i^1 - \bar{x}^1)(y_i^1 - \bar{y}^1)}{\sum(x_i^1 - \bar{x}^1)^2} = \frac{mn \sum(x_i^2 - \bar{x}^2)(y_i^2 - \bar{y}^2)}{m^2 \sum(x_i^2 - \bar{y}^2)}$$

and $$\hat{\beta}_1^2 = \frac{\sum(x_i^2 - \bar{x}^2)(y_i^2 - \bar{y}^2)}{\sum(x_i^2 - \bar{x}^2)^2} \quad (\text{from eqn } (*))$$

Then, it's easy to see that $\hat{\beta}_1^1 = \frac{n}{m} \hat{\beta}_1^2$, or $\boxed{\hat{\beta}_1^2 = \frac{m}{n} \hat{\beta}_1^1}$ $(**)$

Finally, we can calculate $\hat{\beta}_0^2$:

$$\begin{cases} \hat{\beta}_0^1 = \bar{y}^1 - \hat{\beta}_1^1 \times \bar{x}^1 \\ \hat{\beta}_0^2 = \bar{y}^2 - \hat{\beta}_1^2 \bar{x}^2 \end{cases}$$

Divide $\hat{\beta}_0^2$ by $\hat{\beta}_0^1$, we obtain:

$$\frac{\hat{\beta}_0^2}{\hat{\beta}_0^1} = \frac{\bar{y}^2 - \hat{\beta}_1^2 \bar{x}^2}{\bar{y}^1 - \hat{\beta}_1^1 \bar{x}^1} =$$

$$= \frac{\bar{y}^2 - \hat{\beta}_1^2 \bar{x}^2}{n\bar{y}^2 - \frac{n}{m}\hat{\beta}_1^2 \cdot m\bar{x}^2} \quad \text{(from eqns (*) and (**))}$$

$$= \frac{1}{n}$$

$$\Rightarrow \boxed{\hat{\beta}_0^2 = \frac{\hat{\beta}_0^1}{n}}$$

5.1 $\widehat{TS} = \underset{(23.5)}{640.3} - \underset{(2.02)}{4.93}\, CS$, $R^2 = 0.11$, $SER = 8.7$ (56 samples)

a. 95% confidence interval for β_1 is $\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$

Plug in the numbers, we have:

$$CI = \{-4.93 \pm 1.96 \times 2.02\} = (-8.89, -0.97)$$

b. $$t^{act} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{-4.93}{2.02} = -2.44$$

Suppose that $\hat{\beta}_1$ is approx. normally distributed; then,

$$\begin{aligned} p\text{-value} = \Pr(|z| > |t^{act}|) &= 2\Phi(-|t^{act}|) \\ &= 2 \times \Phi(-2.44) \\ &= 2 \times 0.00734 \quad \text{(z-table.net)} \\ &= \boxed{0.015 < 0.05} \end{aligned}$$

$\Rightarrow$ reject the null hypo

DÁM ƯỚC MƠ

At the 1% level, $\boxed{\text{p-value} = 0.015 > 0.01}$ => don't reject

c. $H_0: \beta_1 = -5.0$, then $t^{act} = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \dfrac{-4.93+5}{2.02} = 0.035 \approx 0.04$

$\Rightarrow$ p-value $= 2\Phi(-|t^{act}|) = 2\Phi(-0.04) = 2 \times \cancel{0.51595}\ 0.48405$

$= \cancel{0.1684 > 0.05}$

~~=> not contained in the 95% CI for β_1~~

$= 0.9681 > 0.05$

=> ~~β_1~~ -5.0 is contained in the 95% CI for β_1

d. $P(-z \le Z \le z) = 0.9$

$\Leftrightarrow P(Z \le z) - P(Z \le -z) = 0.9$

$\Leftrightarrow 2P(Z \le z) - 1 = 0.90$

$\Leftrightarrow P(Z \le z) = 0.95 \Rightarrow z = \cancel{1.96}\ 1.64$ (z-table.net)

~~The 90% CI for β_1 is then:~~

~~$CI = \{\hat{\beta}_1 \pm 1.64\, SE(\hat{\beta}_1)\} = \{-4.93 \pm 1.64 \times 2.02\}$~~

~~$= (-8.24, -1.62)$~~

The 90% CI for β_0 is then:

$CI = \{\hat{\beta}_0 \pm 1.64\, SE(\hat{\beta}_0)\} = \{640.3 \pm 1.64 \times 23.5\}$

$= \boxed{(601.76,\ 678.84)}$

6.5 $\hat{P} = 109.7 + 0.567 BDR + 26.9 bath + 0.239 Hsize + 0.005 Lsize$
$+ 0.1 age - 56.9 poor$, $\bar{R}^2 = 0.85$, $SER = 45.8$

a. $\Delta p = 26.9\, \Delta bath = 26.9$ (thousands)

b. $\Delta p = 26.9\, \Delta bath + 0.239\, \Delta Hsize$

$= 26.9 + 0.239 \times 80 = 46.02$ (thousands)

c. $\Delta p = -56.9\,\Delta pcr = -56.9$ (thousands)

d.

$$\begin{cases} R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{SSR}{TSS} \\ \bar{R}^2 = 1 - \dfrac{n-1}{n-k-1}\dfrac{SSR}{TSS} \Rightarrow \dfrac{SSR}{TSS} = (1-\bar{R}^2)\dfrac{n-k-1}{n-1} \end{cases}$$

$$\Rightarrow R^2 = 1 - (1-\bar{R}^2)\frac{n-k-1}{n-1}$$

$$= 1 - (1-0.88^2)\frac{200-7-1}{200-1} \approx 0.$$

$$\Leftrightarrow \boxed{R^2 \approx 0.73}$$