# DSO530 Applied Statistical Learning

## Lecture 2: linear Regression additional exercises

Dr. Xin Tong
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California

Q1

$\log y_{new} = 1 + 3\log x_{new}$ — ①

$\log y_{old} = 1 + 3\log x_{old}$ — ②

① − ② $\log y_{new} - \log y_{old} = 3(\log x_{new} - \log x_{old})$

$\Rightarrow \log\left(\frac{y_{new}}{y_{old}}\right) = \log\left(\frac{x_{new}}{x_{old}}\right)^3$

- Given the regression equation

$$\log y = 1 + 3\log x\,,$$

How is change in $y$ associated with change in $x$?

- How about $\log y = 1 + x$?
- And $y = 1 + 2\log x$?

$\Rightarrow \frac{y_n}{y_0} = \left(\frac{x_n}{x_0}\right)^3$

If $x \Uparrow 1\%$

$\frac{x_n}{x_0} = 1.01$

$\Rightarrow y_n/y_0 = (1.01)^3 = 1.03$

$\Rightarrow y \Uparrow$ by $3\%$ for $1\%$ increase in $x$.

$\log y_n = 1 + x_n$

$\log y_0 = 1 + x_0$

$\log\left(\frac{y_n}{y_{old}}\right) = x_n - x_0$

If $x \uparrow$ by 1, $\log\left(\frac{y_n}{y_0}\right) = 1$

$\Rightarrow y_n = e \cdot y_0 \quad \therefore y \uparrow 171.8\%$

# Q2

$y_n = 1 + 2\log x_n$

$y_0 = 1 + 2\log x_0$

$y_n - y_0 = 2 \cdot \log \frac{x_n}{x_0}$

If $x \uparrow$ by 1% then $y_n - y_0 = 2\log(1.01)$

$= 2.02$

$\{\log(1+x) \simeq x \text{ if } x \sim 0\}$

$\therefore y \uparrow$ by 0.02

- What is the meaning of $\underline{1 - R^2}$?

Variability in data not explained by model.

# Q3

- True or False: "$R^2$ is never bigger than 2"

True, $R^2 = 1 - \frac{RSS}{TSS}$

this can never be negative as it is $^2$.

$R^2 \leq 1$

# Q4

- Suppose $cor(X_1, Y) = -0.5$, $cor(X_2, X_1) = 0.5$, $cor(X_2, Y) = 0.8$. Which regression will have the second largest $R^2$?
  - regress $Y$ on $X_1$ 0.25
  - regress $Y$ on $X_2$ 0.64 $\longrightarrow R^2 = r^2$
  - regress $Y$ on $X_1$ and $X_2$ largest (because of most variables)

# Q5

$R^2_{y_1|x_1} = 0.5 = 1 - \frac{RSS}{TSS} = 1 - \frac{5}{\cancel{TSS} \to 10}$

$TSS_2 = 4 \cdot TSS_1$ $\{$ as $TSS = (y - \bar{y})^2$

$= 40$

co-relation remains same even if multiplied by constant.

- Regress $Y_1$ on $X_1$, we have $R^2 = 0.5$, $RSS = 5$. Let $X_2 = 3X_1$ and $Y_2 = 2Y_1$. Now regress $Y_2$ on $X_2$. What is $R^2$? And what is $RSS$?

$R^2_{y_2|x_2} = r^2(x_2, y_2) = r^2(x_1, y_1) = R^2_{y_1|x_1} = 0.5$

$R^2_{y_2|x_2} = 1 - \frac{RSS_2}{TSS_2} = 0.5 = 1 - \frac{RSS_2}{40} \Rightarrow RSS_2 = 20$