DSO530 Applied Statistical Learning

Lecture 2a: Simple linear Regression addition exercises

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

Q0 (not a regression problem)

- Collins and his wife were accused of robbery.
- Collins was a black man with a beard and his wife was a blond white woman.
- At Trial, the prosecutors had difficulty establishing a positive identification, so they resorted to probabilistic evidence.
- An *expert witness*:
  - black man with beard: prob 1/10
  - man with monstache: prob 1/4
  - White woman with pony tail: prob 1/10
  - White woman with blonde hair: prob 1/3
  - Yellow motor car: prob 1/10
  - Interracial couple in car: prob 1/1000
- Q: what is the probability of seeing such a couple with these characteristics?
- The Trial Court found Collins guilty. He appealed.
- The California Supreme Court reversed the decision.

Q0 (not a regression problem)

- Collins and his wife were accused of robbery.
- Collins was a black man with a beard and his wife was a blond white woman.
- At Trial, the prosecutors had difficulty establishing a positive identification, so they resorted to probabilistic evidence.
- An *expert witness*:
  - black man with beard: prob 1/10
  - man with monstache: prob 1/4
  - White woman with pony tail: prob 1/10
  - White woman with blonde hair: prob 1/3
  - Yellow motor car: prob 1/10
  - Interracial couple in car: prob 1/1000
- Q: what is the probability of seeing such a couple with these characteristics?
- The Trial Court found Collins guilty. He appealed.
- The California Supreme Court reversed the decision.

Q1

- Given the regression equation

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

  How is change in x associated with change in y?
- How about $\log y = 1 + x$?
- And $y = 1 + 2 \log x$?

Q2

- 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
  - regress Y on X_2
  - regress Y on X_1 and X_2

Q3

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

Q4

- What is the meaning of $1 - R^2$?

Q5

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