Lecture 2 Review Quiz

Mallick Hossain

University of Pennsylvania

Class Logistics

1. Have you turned in your homework?
    - The correct answer is “Yes”
2. Have you chosen a grading scheme?
    - The correct answer is “Yes”

Real Questions

1. What is the relationship between variance and covariance?
    - Covariance is a general concept that includes variance. Variance is simply a random variable's covariance with itself. Recall that the formula for covariance is

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

2. What is the sum of deviations from the average for a dataset?
    - Zero
3. Why do we prefer standard deviation over variance?
    - The units of standard deviation are the same as the variable. Variance is the units squared

Real Questions

4. Why do we prefer correlation over covariance?
    - Correlation is unitless and normalized between -1 and 1. Covariance has units of X times units of Y.
5. What problems do outliers cause?
    - They affect the mean, variance, and range of our data
6. What does the z-score measure?
    - The number of standard deviations the data is away from the mean

Real Questions

7. How do we fit a line to data using Ordinary Least Squares (OLS)?
    - Specify a linear relationship: $y_i = a + bx_i$
    - Find a and b such that they minimize the sum of squared deviations from the line: $\min_{a,b}(y_i - a - bx_i)^2$
    - $a = \bar{y} - b\bar{x}$
    - $b = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}$
8. True or False? If two variables are unrelated, their correlation is zero.
    - True. The other direction is false though.
9. Ready for the next lecture?

Real Questions