# DATA 100: Vitamin 9 Solutions

October 25, 2019

## 1 Correlation

### 1.1 Definition

Fill in the blank:

Correlation coefficient measures the strength of the _____ between two quantitative variables.

- ☐ association
- ☐ relationship
- ☑ linear association
- ☐ causal relationship

**Explanation:** The correlation coefficient measures the strength of the *linear association* between two quantitative variables.

### 1.2 Perfect Linear Relationships

Suppose variables $x$ and $y$ have a perfect linear relationship. Select all possible values of the correlation coefficient between $x$ and $y$.

- ☑ 1
- ☐ 0
- ☑ $-1$
- ☐ Any value between 0% and 100%

**Explanation:** A perfect linear relationship between $x$ and $y$ implies that their correlation coefficient is either 1 (in the case of a perfect positive linear relationship) or $-1$ (in the case of a perfect negative linear relationship).

# 2 Residuals

## 2.1 Residual Plots

We fit a simple linear regression to our data $(x_i, y_i), i = 1, 2, 3$, where $x_i$ is the independent variable and $y_i$ is the dependent variable. Our regression line is of the form $\hat{y} = \hat{a} + \hat{b}x$. Suppose we plot the relationship between the residuals of the model and the $\hat{y}s$, and find that there is a curve. What does this tell us about our model?

- ☐ The relationship between our dependent and independent variables is well represented by a line.
- ☐ The accuracy of the regression line varies with the size of the dependent variable.
- ☑ The variables need to be transformed, or additional variables are needed.

**Explanation:** If we see a curve in our residual plot, then the relationship is not well represented by a line. Either more variables are needed, or transformations of the current variables are necessary.

## 2.2 Useful Properties

Which of the following are useful properties of residuals in an ordinary least squares regression (with an intercept) where the dependent and independent variables are vectors?

- ☑ The average of the residuals is 0.
- ☐ The inner product of the fitted values and the residuals is always positive.
- ☑ The residuals and the independent variable are orthogonal.

**Explanation:** See slides of lecture 15 for a list of these properties and an explanation.

# 3 Statistical Models

Which of the following statements are true regarding statistical models?

- ☐ They are always useful, even when they provide inaccurate descriptions of the data.
- ☑ They allow us to generalize beyond the subjects studied.
- ☑ We can use them to discover physical facts and social phenomena.

**Explanation:** Statistical models have many uses; this a non-exhaustive list taken from the lecture slides. Note that a statistical model is only useful it provides reasonably accurate descriptions of the data.