Week 1 Quiz

Quiz, 9 questions

✔ **Congratulations! You passed!**

Next Item

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

1.

A researcher is investigating a possible effect of Facebook use on exam grades and he wants to try a linear regression model. He has outside data, collected via a survey, showing that the correlation between the student's exam grade and the number of hours spent on Facebook in the days before the exam is about -0.12. If the researcher wanted to use linear regression to predict the exam score for a student who spent 6 hours on Facebook before his exam, which is of the following **best** describes the researcher's situation?

- He should use linear regression since the correlation coefficient is not 0 it is clear that there is a linear relationship between the two variables.
- He needs to plot the data (e.g. a scatterplot) before deciding whether to use linear regression.

Correct

This question refers to the following learning objective(s): When describing the association between two numerical variables, evaluate

- direction: positive ($x \uparrow, y \uparrow$), negative ($x \downarrow, y \uparrow$)
- form: linear or not
- strength: determined by the scatter around the underlying relationship

- He should use linear regression. He can predict the exam score for a student who spent 6 hours on Facebook before the exam using the equation score = 100 + 6 × (−0.12).
- He should not use linear regression because the correlation is not strong enough.
- He should use linear regression. Because of the negative correlation, spending more time on Facebook leads to lower exam scores..

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

2.

Suppose the correlation coefficient between an explanatory variable (the size of a person's home, measured in square feet) and a response variable (the amount they spend on alcohol per year, measured in US Dollars) is 0.34. If we change the units used to measure their home size from square feet to square meters, which of the following will be **true**? Assume 1 square foot is roughly 0.09 square meters.

- The new correlation coefficient should be approximately $R = 0.34 \times 0.09 = 0.03$.
- The new correlation coefficient should be the same: $R = 0.34$.

Correct

This question refers to the following learning objective(s): Note that correlation coefficient (R, also called Pearson's R) has the following properties:

- the magnitude (absolute value) of the correlation coefficient measures the strength of the linear association between two numerical variables
- the sign of the correlation coefficient indicates the direction of association
- the correlation coefficient is always between -1 and 1, -1 indicating perfect negative linear association, +1 indicating perfect positive linear association, and 0 indicating no linear relationship
- the correlation coefficient is unitless
- since the correlation coefficient is unitless, it is not affected by changes in the center or scale of either variable (such as unit conversions)
- the correlation of X with Y is the same as of Y with X
- the correlation coefficient is sensitive to outliers

The correlation coefficient is unitless and so it stays the same when the units of the involved variables change.

- The new correlation coefficient should be approximately $R = 0.34/0.09 = 3.78$.

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

3.

Fill in the blank: Residuals of linear models should be distributed nearly normally around _____.

- the mean of y
- the mean of x
- 0

Correct

This question refers to the following learning objective(s):

• Define residual (e) as the difference between the observed (y) and predicted ($\hat{y}$) values of the response variable.

$$e_i = y_i - \hat{y}_i$$

• Define the least squares line as the line that minimizes the sum of the squared residuals, and list conditions necessary for fitting such line:

1. linearity
2. nearly normal residuals
3. constant variability

- the mean of $\hat{y}$ (the predicted values)

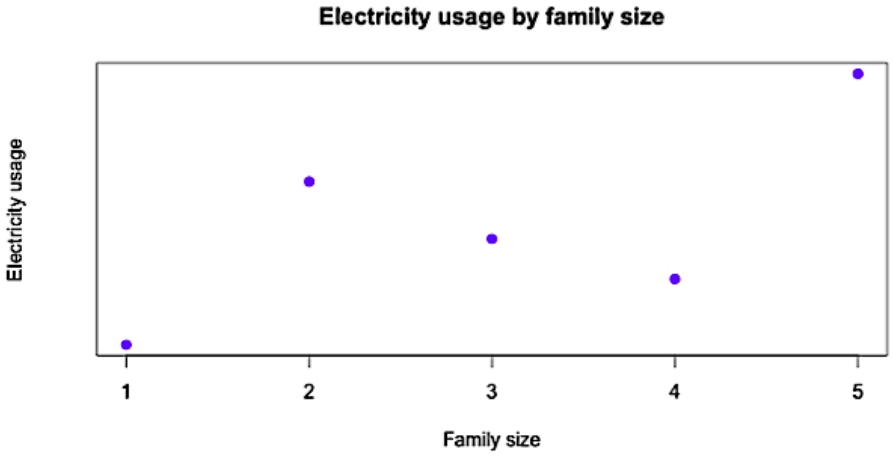
Week 1 Quiz

Quiz, 9 questions

1 / 1 point

4.

An ambitious young student collected data on household electricity usage for a few families. After she plotted the data (shown below), the student observed that there did not appear to be a strong, positive linear relationship between the two variables as she had expected. The student still suspects that such a relationship exists - which of the following is the **best** advice an experienced statistician could give to the girl in order to help her investigate whether there is a linear relationship?



- Collect one data point each for a family of size 6, 7, etc. in order to extend the plot off to the right.
- There is actually no practical strategy; whatever the strength of the association between these two variables, we cannot get a better idea of it just by collecting more data.
- Collect electricity usage data for more families of sizes 1 through 5.

Correct

This question refers to the following learning objective(s): When describing the association between two numerical variables, evaluate

- direction: positive ($x \uparrow, y \uparrow$), negative ($x \downarrow, y \uparrow$)
- form: linear or not

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

5.

For a certain professional basketball team, 32% of the variability in the team's points scored per game is explained by the total salary of the opposing team. For this particular team, which of the following **could be** the correlation between their points scored per game and the salary of the opposing team?

- $1 - \sqrt{0.32} = 0.434$
- $-0.32^2 = -0.102$
- $1 - 0.32^2 = 0.998$
- $-\sqrt{0.32} = -0.566$

Correct

This question refers to the following learning objective(s): Define R2 as the percentage of the variability in the response variable explained by the the explanatory variable.

- For a good model, we would like this number to be as close to 100% as possible.
- This value is calculated as the square of the correlation coefficient.

Correlation coefficient is the square root of R^2, both positive and negative values could be the correlation.

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

6.

A student is studying the relationship between how much money students spend on food and on entertainment per week. Based on a sample size of 270, he calculates a correlation coefficient of 0.013 for these two variables. Which of the following is the **most appropriate** interpretation?

- This correlation indicates a definite strong nonlinear relationship.
- There is no linear relationship but there may be a nonlinear relationship.

Correct

This question refers to the following learning objective(s):

- Define correlation as the **linear** association between two numerical variables.
- Note that a relationship that is nonlinear is simply called an association.
- Note that correlation coefficient (R, also called Pearson's R) has the following properties:

- the magnitude (absolute value) of the correlation coefficient measures the strength of the linear association between two numerical variables

- the sign of the correlation coefficient indicates the direction of association

- the correlation coefficient is always between -1 and 1, -1 indicating perfect negative linear association, +1 indicating perfect positive linear association, and 0 indicating no linear relationship

- the correlation coefficient is unitless

- since the correlation coefficient is unitless, it is not affected by changes in the center or scale of either variable (such as unit conversions)

- the correlation of X with Y is the same as of Y with X

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

7.

Fill in the blanks: A data point that has a negative residual is located _____ the regression line.

- on
- below

Correct

Residual is defined as observed minus predicted, therefore a negative residual means the observed is below the predicted (the regression line).

This question refers to the following learning objective(s): Define residual (e) as the difference between the observed (y) and predicted ($\hat{y}$) values of the response variable.

$e_i = y_i - \hat{y}_i$

- above

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

8.

The following ANOVA output is for the linear model predicting nicotine content (in mg) from tar content (in mg). Which of the following is R^2? Choose the **closest** answer.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
d$Tar	1	2.869	2.869	474.431	0.000
Residuals	23	0.139	0.006		
Total	24	3.008			

- 0.2%
- 4%
- 20%
- 5%
- 95%

Correct

This question refers to the following learning objective(s): Define R^2 as the percentage of the variability in the response variable explained by the the explanatory variable.

Week 1 Quiz

Quiz, 9 questions

1 / 1 point

9.

Based on a random sample of 170 married couples in Britain, a researcher finds that the relationship between the husbands' and wives' ages is described by the following equation:

$$\widehat{age}_{wife} = 1.57 + 0.91\ age_{husband}$$

Which of the following is the **best** interpretation of the slope estimate?

- For each additional year increase of husband's age, we would expect the wife's age to be 0.91 years higher, on average.

Correct

This question refers to the following learning objective(s): Interpret the slope as

- when x is numerical: "For each unit increase in x, we would expect y to be lower/higher on average by $|b_1|$ units"
- when x is categorical: "The value of the response variable is predicted to be $|b_1|$ units higher/lower between the baseline level and the other level of the explanatory variable."
- Note that whether the response variable increases or decreases is determined by the sign of b_1.

- Most wives in Britain are 0.91 years younger than their husbands.
- On average, when a husband in Britain gets 1 year older, his wife only gets 0.91 years older.
- For each additional year increase of wife's age, we would expect the husband's age to be 0.91 years higher, on average.

Week 1 Quiz

Quiz, 9 questions