Week 1 Quiz

Quiz, 9 questions

9/9 points (100%)

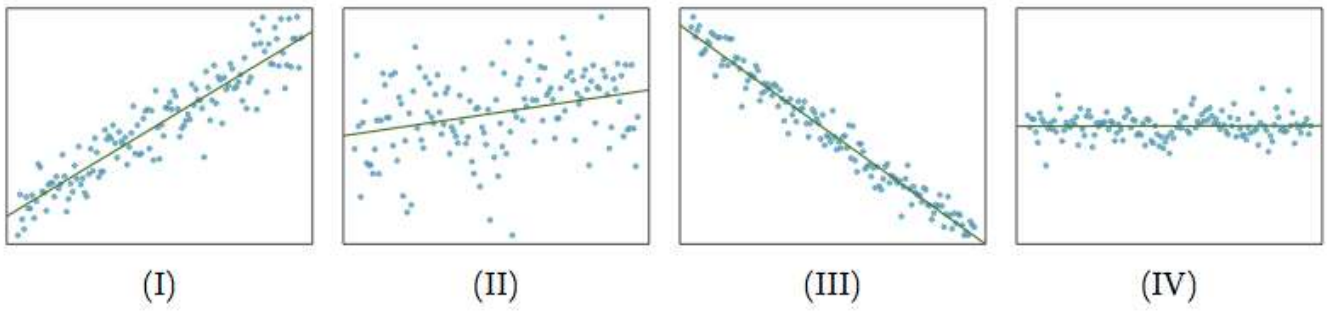
Congratulations! You passed!

Next Item

1 / 1 point

1.

Of the four plots shown below, which one appears to show the weakest relationship between two variables?



- (I)
- (II)
- (III)
- (IV)

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

Choices (I) and (III) show strong linear relationships. The relationship in (II) is weak but existent. However in (IV) there is no relationship.

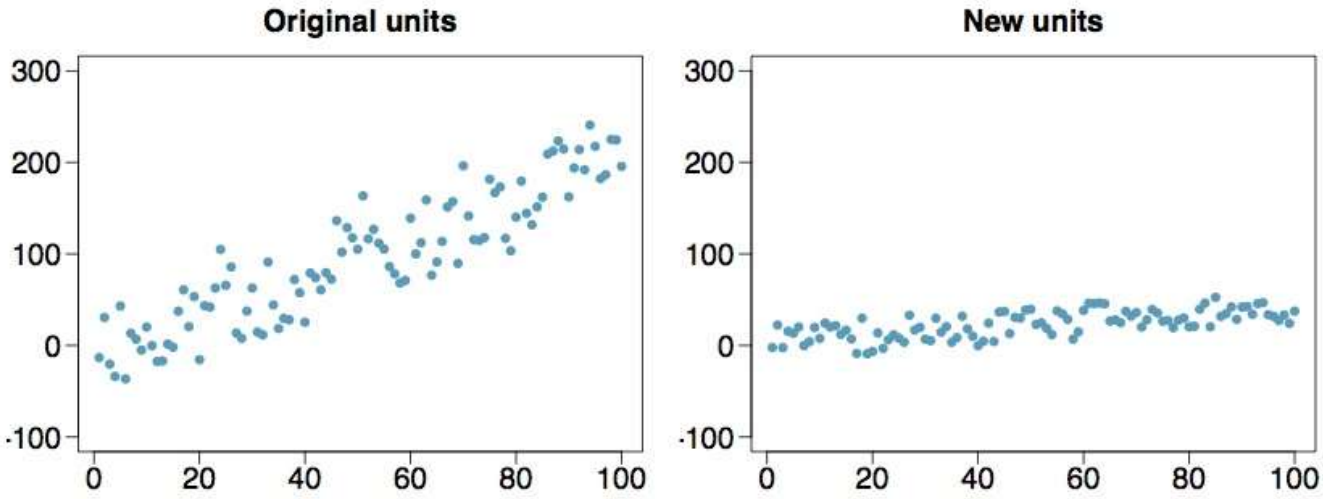
Week 1 Quiz
Quiz, 9 questions

1/ 1 point

9/9 points (100%)

2.

The first plot below was created by plotting data collected on two variables. Then, the second plot was created using the same data but with different units for the dependent variable. In the context of linear regression, which of the following best describes the differences between the two plots?



- The slope of the linear relationship in the first plot has a larger absolute value, but the correlation coefficients for the two plots are the same.

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 does not change if the units of the involved variables change. Magnitude of the slope is measured by the steepness of the relationship, the steeper the line, the larger the magnitude of the slope. Since the plots are plotted on the same x-axis and y-axis scales, we can say for sure that the regression slope has decreased in the 2nd plot compared to the 1st.

Week 1 Quiz

Quiz, 9 questions

9/9 points (100%)

The correlation coefficient for the second plot has a smaller absolute value, but the slopes of the linear relationships in the two plots are the same.

The slope of the linear relationship in the first plot has a smaller absolute value, but the correlation coefficients for the two plots are the same.

The correlation coefficient for the second plot has a larger absolute value, but the slopes of the linear relationships in the two plots are the same.

Both the slopes and the regression coefficients are the same for the two plots.

---

1 / 1 point

3.

Which of the following is **false**?

The variability of residuals should increase as x increases.

Correct

Constant variability in the residuals is a condition for fitting a least-squares 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.

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

• 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

A data point that has a negative residual is located below the regression line.

Residuals of linear models should be distributed nearly normally around 0.

The residuals plot (residuals vs. x) should show a random scatter around 0.

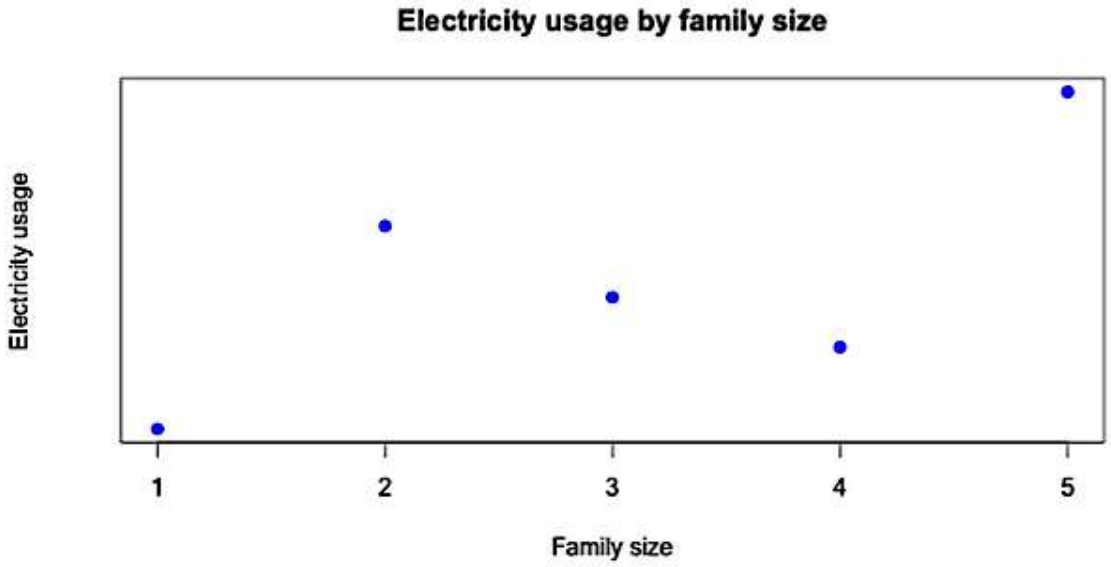
Week 1 Quiz
Quiz, 9 questions

1/ 1 point

9/9 points (100%)

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.
- 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
- strength: determined by the scatter around the underlying relationship

This is the correct answer because it would populate the current plot with more data; it would then be much easier to see whether there is a strong, positive linear relationship or not.

- 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.
- Plot the current data again, using a different scale for electricity usage. A poorly-chosen scale for this plot may be hiding a linear trend.

Week 1 Quiz

Quiz, 9 questions

9/9 points (100%)

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?

- [x] $-\sqrt{0.32} = -0.566$

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.

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

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.
- [x] 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

- the correlation coefficient is sensitive to outliers

Week 1 Quiz

Quiz, 9 questions

9/9 points (100%)

- This low correlation of 0.013 indicates there is no relationship.
- This correlation indicates there is a strong linear relationship.

✔ 1 / 1 point

7.

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

- above
- 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.

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

- on

Week 1 Quiz

Quiz, 9 questions

9/9 points (100%)

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			

- 4%
- 5%
- 20%
- 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.

- 0.2%

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?

- 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

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

9/9 points (100%)

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.