Week 1 Quiz

Back to Week 1

9/9 points earned
(100%)

Quiz passed!

1 / 1
points

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. Because of the negative correlation, spending more time on Facebook leads to lower exam scores..
- 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.

---

1 / 1 points

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 $R = 0.34 \times (0.09)^2$ = 0.003.
- The new correlation coefficient should be approximately $R = 0.34/0.09 = 3.78$.

---

1 / 1 points

3.

Which of the following is **false**?

- The residuals plot (residuals vs. x) should show a random scatter around 0.
- Residuals of linear models should be distributed nearly normally around 0.
- A data point that has a negative residual is located below the regression line.
- 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:

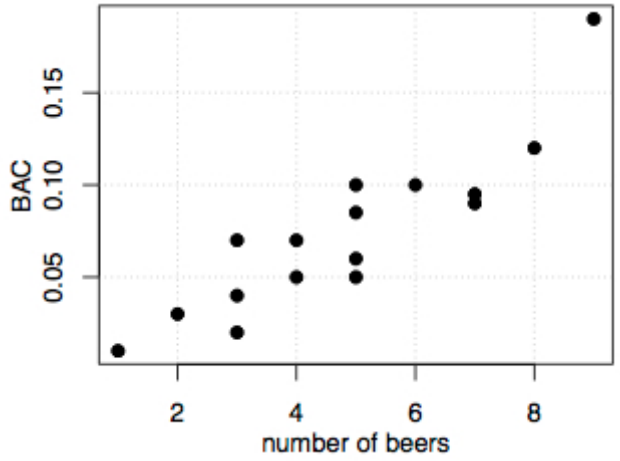
1. linearity
2. nearly normal residuals
3. constant variability

---

1 / 1 points

4.

Sixteen student volunteers at Ohio State University drank a randomly assigned number beers. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood. The scatterplot displays the relationship between BAC and number of beers consumed. Suppose **a mistake was found** in the data: the student who supposedly drank the highest number of beers (9 beers) actually only drank 6. His BAC was recorded correctly. In a new scatterplot, how would the strength of the association appear - compared to the strength of the association shown here?



- Roughly the same as the strength of the association shown in the above scatterplot.
- Weaker than the strength of the association shown in the above scatterplot.

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

Moving the rightmost point back to x=6 would result in a scatterplot where there is **more** scatter around the main trend of the data. Therefore we would say the new strength of association would be weaker than that shown above.

- Stronger than the strength of the association shown in the above scatterplot.
- It's impossible to tell.

---

1 / 1 points

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?

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

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.

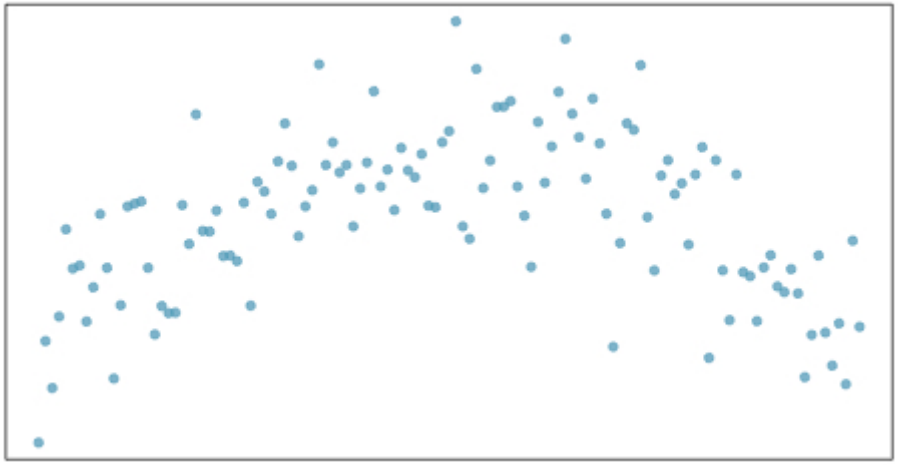
Correlation coefficient is the square root of R^2, both positive and negative values could be the correlation.

---

1 / 1 points

6.

A colleague needs some help with a statistics problem: He brings you the plot shown below, along with a correlation coefficient of 0.03 which he calculated himself. The plot shows two numerical variables which are obviously strongly related, and as a result your colleague is afraid he made a mistake calculating the correlation coefficient: that is, he was surprised to get an answer so close to 0. Given only this information, which of the following responses is the **best** to give your colleague?



- Your colleague must have made a mistake in his calculations. A much higher correlation coefficient is expected for variables that show a clear association.

The correlation coefficient measures the strength of the linear relationship, therefore two variables that have a strong non-linear association might still have a low correlation coefficient.

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

---

1 / 1 points

7.

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

- on
- 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.

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

---

1 / 1 points

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			

- 20%

- 0.2%
- 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.

- 5%
- 4%

---

1 / 1 points

9.
Based on an observational study, a clinical psychologist finds that the relationship between the number of weeks spent in a therapy hospital and number of seizures per week is described by the following equation:

$$\widehat{seizures} = 14.09 - 0.91\ weeks$$

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

- Each additional week spent in a therapy hospital decreases the number of seizures per week by 0.91.
- All patients start their treatment with at least 14.09 seizures per week.
- For each additional seizure per week, we would expect the average number of additional weeks spent in the therapy hospital to be higher by 0.91 seizures.

For each additional week spent in the therapy hospital, we would expect the average number of seizures per week to lower by 0.91 seizures.

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.