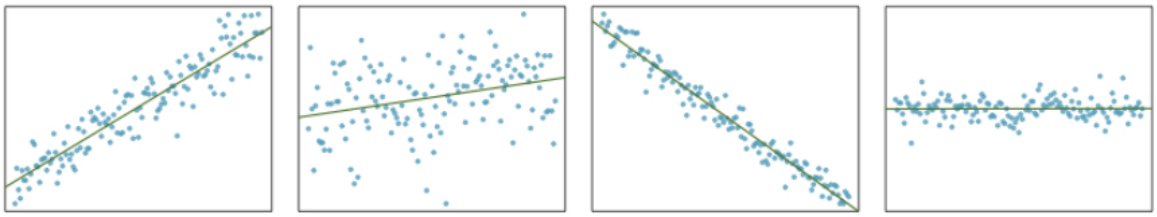
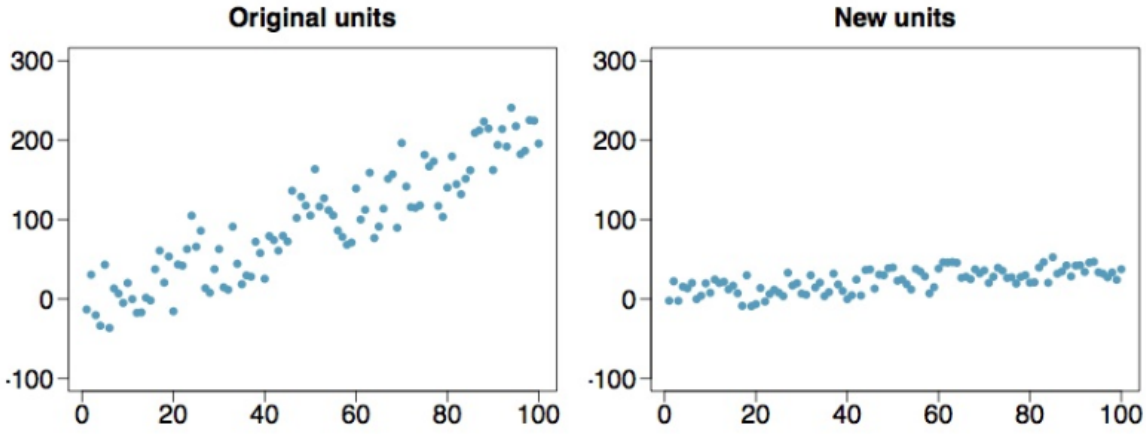
1. Of the four plots shown below, which one appears to show the weakest relationship between two variables?



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

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

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.
- ( ) 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 smaller 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.
- ( ) 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.

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

- (●) 0
- ( ) the mean of y
- ( ) the mean of x
- ( ) the mean of $\hat{y}$ (the predicted values)

✓ **Correct**

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

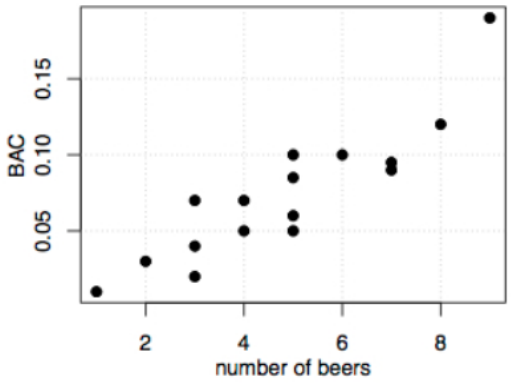
• Define residual (e) as the difference between the observed (y) and predicted ($\hat{y}$) values of the response variable.

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

• 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

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

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?

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

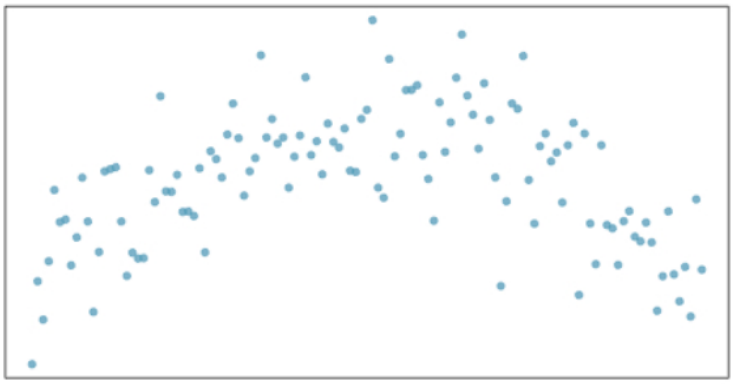
✓ **Correct**

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.

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?



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

7. The linear model below is used for predicting poverty rate from high school graduation rate in the 51 states in the US (including DC).

$$\widehat{poverty} = 64.68 - 0.62\ HS\ grad\ rate$$

High school graduation rate for North Carolina is 81.4% and the poverty rate is 13.1%. What is the residual for this observation? Choose the **closest** answer.

A snippet of the data matrix is provided below, pay attention to the scale of the data in solving this question:

	state	poverty rate	hs grad rate
1	Alabama	14.6	79.9
2	Alaska	8.3	90.6
...	...	...	...
51	Wyoming	9.5	90.9

- ○ 0
- ○ -24.8
- ◉ -1.1

8. Which of the following best describes SST (sum of squares total) in a regression?

- ○ Explained variability in the response variable.
- ○ Strength of the model fit.
- ○ Total variability in the explanatory variable.
- ◉ Total variability in the response variable.
- ○ Unexplained variability in the response variable.

✓ **Correct**

Sum of squares total is the total variability in the response variable.

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