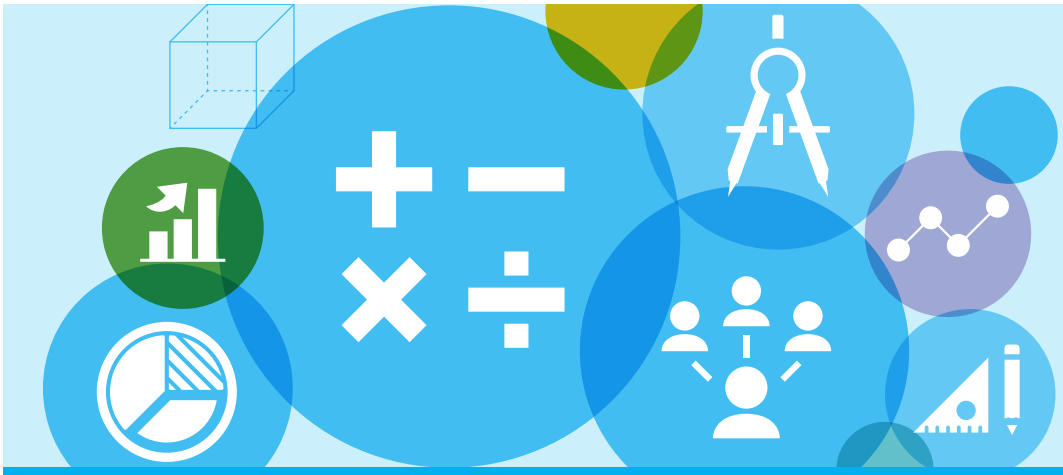


STUDENT VERSION

EDUCATIONAL ATTAINMENT AND MARRIAGE AGE — TESTING A CORRELATION COEFFICIENT'S SIGNIFICANCE

Activity Items

The following items are part of this activity and appear at the end of this student version.

- Item 1: Data Table
- Item 2: Optional Instructions for Calculating *r* on a TI-84 Plus
- Item 3: Critical Values of *r* at a 5 Percent Significance Level

Student Learning Objectives

- I will be able to predict and test the significance of the relationship between two quantitative variables.
- I will be able to write a line of best fit and interpret slope and *y*-intercept in the context of the data.
- I will be able to assess the strength and direction of a linear association based on a correlation coefficient.
- I will be able to compute a correlation coefficient and distinguish between correlation and causation.

NAME:_____ DATE:_____

Part 1 – Make Predictions

1. Between 2011 and 2019 in the United States, do you think the percentage of women aged 25–34 with a bachelor's degree or higher increased, decreased, or stayed the same? Explain your reasoning.

2. During the same period in the United States, do you think the median age of women when they were first married increased, decreased, or stayed the same? Explain your reasoning.

3. State the null and alternative hypotheses for whether there is a relationship between the variables in questions 1 and 2.

   Null hypothesis:

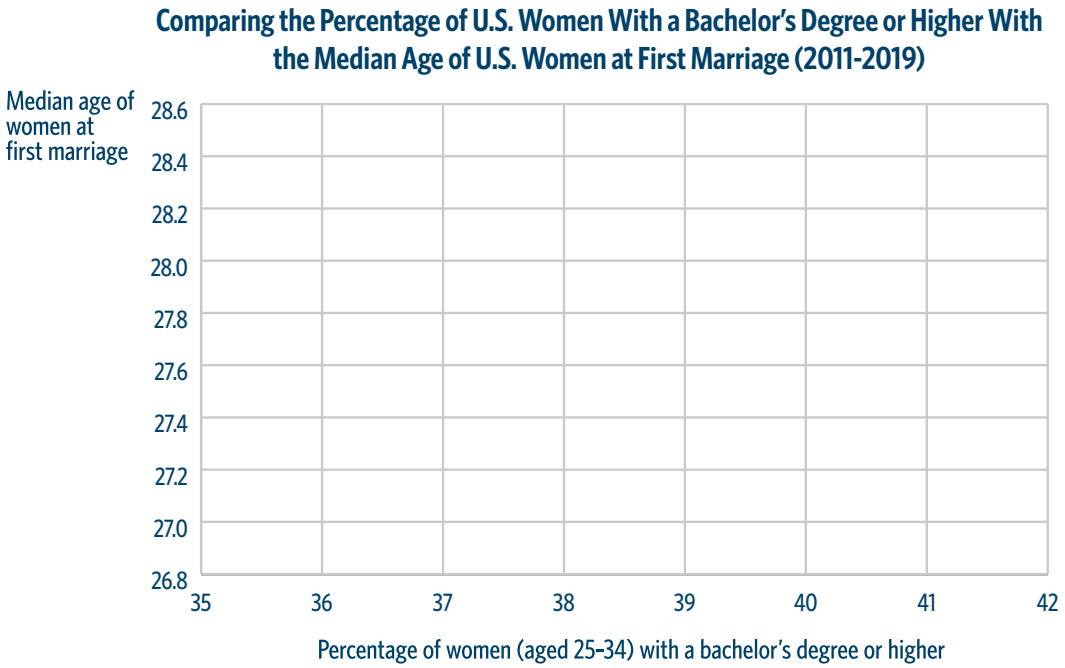
   Alternative hypothesis:

4. Make a conjecture predicting whether a relationship exists between the variables.

5. If you predicted a relationship in question 4, state its direction: Do you think there will be higher percentages of women with bachelor's degrees during years with *higher* or *lower* median ages of women when they were first married? Why?

Part 2 – Evaluate Data to Assess Predictions

1. Use **Item 1: Data Table** to create a scatter plot on the following grid. Use the education data as the independent variable and the marriage data as the dependent variable, keeping in mind that this particular choice is arbitrary.



2. Does a linear model appear to be a sufficient description of the relationship between the two variables in this sample? Explain your reasoning, while keeping in mind that the data are from sample estimates so they could include random error in their values.

3. Find a regression equation (by hand or using technology; feel free to reference **Item 2: Optional Instructions for Calculating *r* on a TI-84 Plus** for help) that best models the data in your scatter plot. Round your values to the nearest hundredth, and explain your equation's meaning.

4. Define the variables in the sample, and interpret the values of the regression coefficients in the context of the data.

   ***x* values:**

   ***y* values:**

   Slope:

   ***y*-intercept:**

5. How could you assess how accurately your regression equation represents the data?

6. Calculate the correlation coefficient (r) of your linear model from question 3 using graphing technology. (You can use **Item 2** for reference.) Round your answer to the nearest thousandth.

   a. Why would a person want to find a correlation coefficient?

   b. Based on the r value you calculated, how strong is the linear relationship between the variables?

7. Calculate the degrees of freedom (*df*) for your equation using the formula $n - 2$, where n represents the number of pairs of data points. Show your work.
   a. Find the corresponding *df* in **Item 3: Critical Values of *r* at a 5 Percent Significance Level**, and determine whether your *r* value is greater than or equal to it.

8. How does the critical value for the sample help us determine whether there is a significant relationship between the variables in the population?

9. Based on what you found in question 8, is there a significant relationship between the median age of U.S. women at first marriage and the percentage of U.S. women aged 25–34 with a bachelor's degree or higher in the years observed? If so, explain and state the direction.

10. Do the results support your initial conjecture of whether there would be a significant relationship? Explain.

11. Does a significant correlation between two variables also indicate a cause-and-effect relationship? Explain, thinking about the two variables in this case.

12. Explain three possible interpretations of a significant correlation for this data set.

13. Which of the possible interpretations that you identified in question 12 is most likely to explain the results? Justify your theory.

Item 1: Data Table

Year	Percentage of U.S. women (aged 25–34) with a bachelor's degree or higher	Median age of U.S. women (aged 15–54) at first marriage
2011	35.5	26.9
2012	36.3	27.1
2013	36.9	27.4
2014	37.5	27.6
2015	38.0	27.8
2016	38.9	27.9
2017	39.7	28.1
2018	40.3	28.3
2019	41.1	28.4

Source for education data: U.S. Census Bureau, Educational Attainment. 2011–2019. American Community Survey 1-Year Estimates.

https://data.census.gov/cedsci/table?q=S1501%3A%20EDUCATIONAL%20ATTAINMENT&tid=ACSST1Y2019.S1501

Copy and paste the link above into your browser to view the source data online.

Source for marriage data: U.S. Census Bureau, Median Age at First Marriage. 2011–2019. American Community Survey 1-Year Estimates.

https://data.census.gov/cedsci/table?q=B12007&tid=ACSDT1Y2019.B12007

Copy and paste the link above into your browser to view the source data online.

Item 2: Optional Instructions for Calculating *r* on a TI-84 Plus

Step 1: Turn on diagnostics to ensure that the *r* value will appear in the display when calculating a linear regression.

[2nd] [0] [x^{-1}] scroll down to DiagnosticOn [Enter] [Enter]

Should say "Done."

Step 2: Clear any previous data in L1 and L2, and then enter the *x* and *y* values from this activity in L1 and L2, respectively.

[Stat] [Enter] highlight L_1 [Clear] [Enter] highlight L_2 [Clear] [Enter]

Enter the X values in L_1 Enter the Y values in L_2

Step 3: Use the LinReg ($a + bx$) function to find the parameters for the line of best fit (linear regression), including *r*.

[Stat] Scroll to the right to Calc [8] [Enter] [Enter]

The last value listed is *r*.

Item 3: Critical Values of r at a 5 Percent Significance Level

Critical values of r for a = .05

df	*a*=.05
1	.997
2	.950
3	.878
4	.811
5	.754
6	.707
7	.666
8	.636
9	.602
10	.576
11	.553
12	.532
13	.514
14	.497
15	.482
16	.468
17	.456
18	.444
19	.433
20	.423

df	*a*=.05
21	.413
22	.404
23	.396
24	.388
25	.381
26	.374
27	.367
28	.361
29	.355
30	.349
35	.325
40	.304
45	.288
50	.273
60	.250
70	.232
80	.217
90	.205
∞	.195

Note: These critical values pertain to a two-sided t*-test.*