

Practice Problems Section 4H

(#1-10) Use either the correlation coefficient or the T-test statistic and the corresponding critical values to fill out the table.

| | T-test statistic or Correlation Coefficient (r) | Sentence to explain T-test statistic or Correlation Coefficient (r) | Critical Value (T or r) | Does the T-test statistic or r-value fall in a tail determined by a critical value? (Yes or No) | Does sample data significantly disagree with H_0 ? |
|-----|---|---|-------------------------|---|--|
| 1. | T = -2.441 | | ± 1.775 | | |
| 2. | r = 0.183 | | 0.316 | | |
| 3. | T = +1.166 | | +2.003 | | |
| 4. | r = -0.799 | | ± 0.286 | | |
| 5. | T = +3.118 | | +2.714 | | |
| 6. | r = 0.921 | | 0.339 | | |
| 7. | T = -0.852 | | ± 2.322 | | |
| 8. | r = -0.026 | | -0.279 | | |
| 9. | T = +1.339 | | ± 1.997 | | |
| 10. | r = 0.483 | | +0.303 | | |

(#11-20) Use each of the following P-values and corresponding significance levels to fill out the table.

| | P-value Proportion | P-value % | Sentence to explain the P-value | Significance Level % | Significance level Proportion | If H_0 is true, could the sample data occur by random chance or is it unlikely? | Reject H_0 or Fail to reject H_0 ? |
|-----|-----------------------|-----------|---------------------------------|----------------------|-------------------------------|---|--|
| 11. | 0.521 | | | 10% | | | |
| 12. | 0.0426 | | | 1% | | | |
| 13. | 3.41×10^{-5} | | | 5% | | | |
| 14. | 0.0033 | | | 1% | | | |
| 15. | 0.768 | | | 5% | | | |
| 16. | 0 | | | 10% | | | |
| 17. | 0.0428 | | | 5% | | | |
| 18. | 0.277 | | | 10% | | | |
| 19. | 6.04×10^{-6} | | | 1% | | | |
| 20. | 0.0178 | | | 5% | | | |

21. List the assumptions that we need to check when performing a correlation hypothesis test.
22. How can we use the scatterplot and the correlation coefficient (r) to determine if the sample data follows a linear pattern?
23. Points in the scatterplot that are far from the regression line are considered outliers, but it is difficult to know if the outliers are influential or not. How can we use the scatterplot and the correlation coefficient (r) to determine if potential outliers are influential or not?
24. Explain the two assumptions that we check by using the histogram of the residuals.
25. Explain how to determine if the residual plot is evenly spread out or not.



(#26-29) Directions: For each of the following problems, use the Statcato printouts provided to answer the following questions.

- a) Write the null and alternative hypothesis for the correlation test. Address the quantitative relationship and label which is the claim.
- b) Write a sentence to explain the strength and direction based on the correlation coefficient (r).
- c) Write a sentence to explain the sample slope (b_1).
- d) Check all of the assumptions for the correlation test. Explain your answers.
- e) Write a sentence to explain the T-test statistic.
- f) Compare the T-test statistic to the critical value. Does the test statistic fall in a tail determined by the critical value?
- g) Does the sample data significantly disagree with the null hypothesis? Explain your answer.
- h) Is the sample slope significantly different from zero? Explain your answer.
- i) Write a sentence explaining the P-value.
- j) Compare the P-value to the significance level. Could the sample data or more extreme occur by sampling variability if the null hypothesis was true or is it unlikely? Explain your answer.
- k) Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.
- l) Write a conclusion for the test addressing evidence and the claim.

26. Use a 5% significance level and the Statcato printout below to test the claim that there is a linear relationship between the height (X) of a man and his weight (Y). This printout came from the random health data at www.matt-teachout.org.

Correlation and Regression: Significance level = 0.05

x = C16 Men Ht (in)

y = C17 Men Wt (Lbs)

Sample size n = 40

Degrees of freedom = 38

| | Test Statistic | Critical Value |
|---|----------------|----------------|
| r | 0.5222 | ± 0.3120 |
| t | 3.7750 | ± 2.0244 |

p-Value = 0.0005

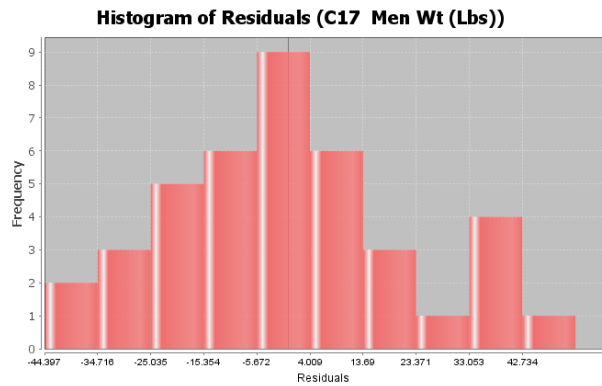
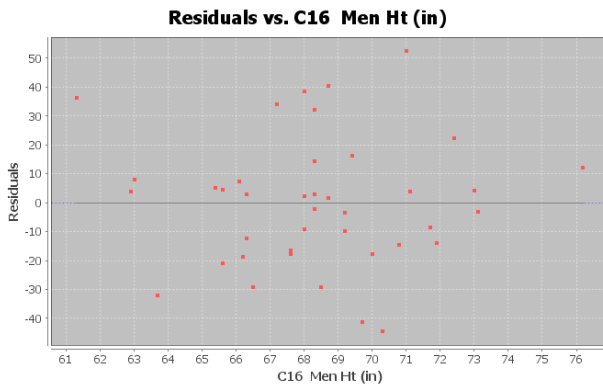
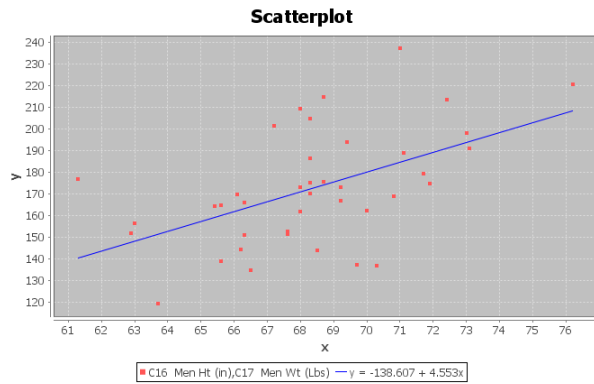
Regression:

Regression equation $Y = b_0 + b_1X$

$b_0 = -138.6070$

$b_1 = 4.5534$





27. Use a 5% significance level and the Statcato printout below to test the claim that there is NO linear relationship between the systolic blood pressure (X) of a woman and her diastolic blood pressure (Y). This printout came from the random health data at www.matt-teachout.org.

Correlation and Regression: Significance level = 0.05

x = C6 Women Syst BP
 y = C7 Women Diast BP
 Sample size n = 40
 Degrees of freedom = 38

| | Test Statistic | Critical Value |
|---|----------------|----------------|
| r | 0.7854 | ±0.3120 |
| t | 7.8209 | ±2.0244 |

p-Value = $1.9615 \cdot 10^{-9}$

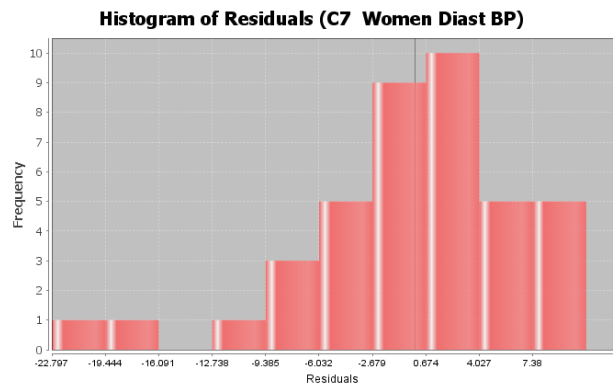
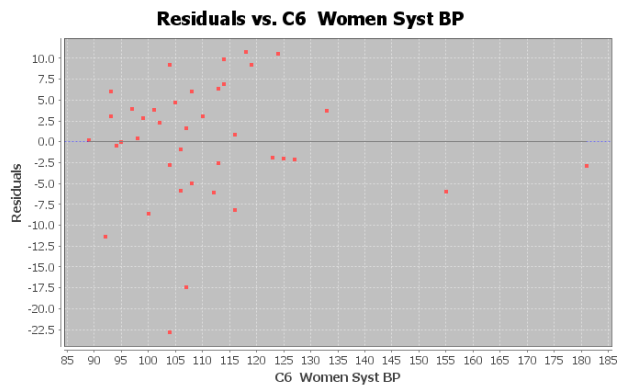
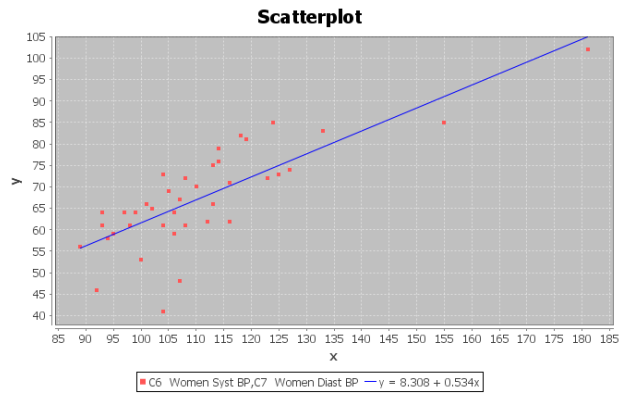
Regression:

Regression equation $Y = b_0 + b_1X$

$b_0 = 8.3079$

$b_1 = 0.5335$





28. Use a 5% significance level and the Statcato printout below to test the claim that there is a relationship between the head width (X) of a bear and its chest size (Y). This printout came from the random bear data at www.matt-teachout.org.

Correlation and Regression: Significance level = 0.05

x = C5 Head Width (In)

y = C8 Chest (in)

Sample size n = 54

Degrees of freedom = 52

| | Test Statistic | Critical Value |
|---|----------------|----------------|
| r | 0.7785 | ±0.2681 |
| t | 8.9451 | ±2.0067 |

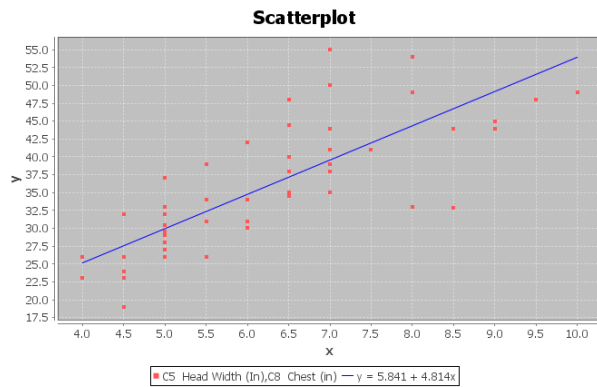
p-Value = $4.2208 \cdot 10^{-12}$

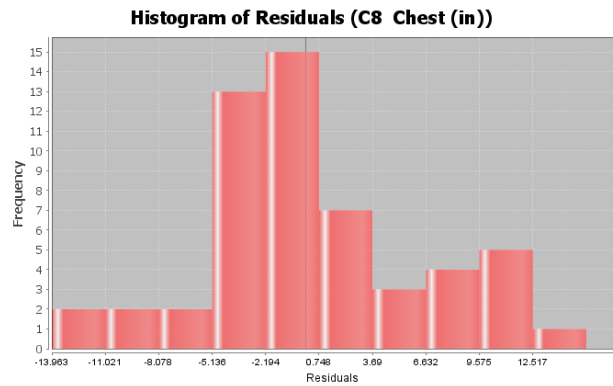
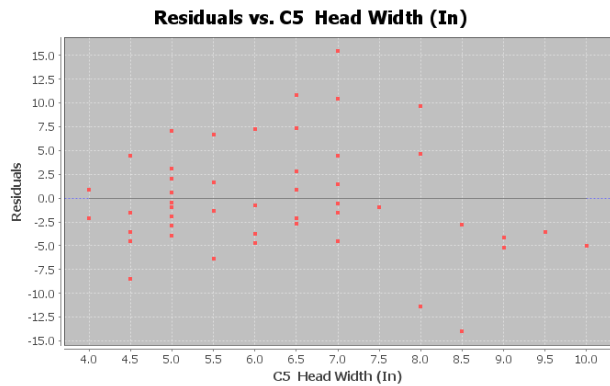
Regression:

Regression equation $Y = b_0 + b_1x$

$b_0 = 5.8408$

$b_1 = 4.8143$





29. Use a 5% significance level and the Statcato printout below to test the claim that there is NO relationship between the neck circumference (X) of a bear and its weight (Y). This printout came from the random bear data at www.matt-teachout.org.

Correlation and Regression: Significance level = 0.05

x = C6 Neck Circum (in)

y = C9 Weight (Lbs)

Sample size n = 54

Degrees of freedom = 52

| | Test Statistic | Critical Value |
|---|----------------|----------------|
| r | 0.9341 | ±0.2681 |
| t | 18.8612 | ±2.0067 |

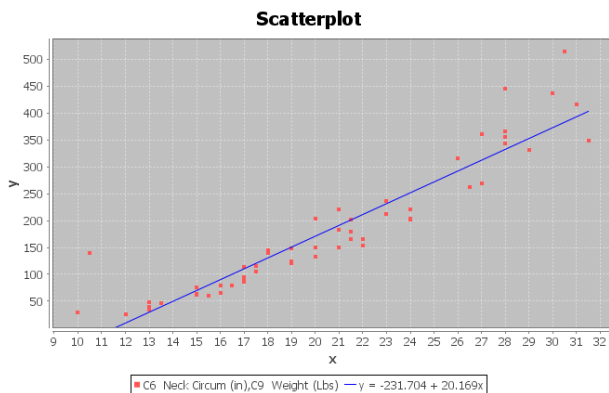
p-Value = 0

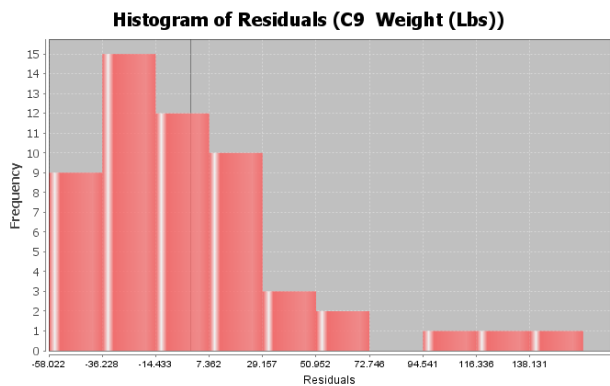
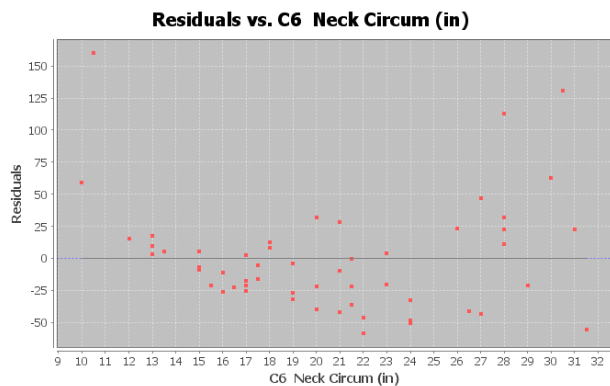
Regression:

Regression equation $Y = b_0 + b_1x$

$b_0 = -231.7044$

$b_1 = 20.1694$





(#30-32) Directions: Go to www.lock5stat.com and click on the "StatKey" button. Under "Randomization Hypothesis Tests", click the one that says, "Test for Slope, Correlation". Click "Generate 1000 Samples" a few times. Remember there are two ways of getting the P-value. If the top of the graph says "Randomization Dot plot of Correlation", then the null hypothesis is $\rho = 0$. Remember rho looks like a " ρ " but it is not a "P". If the top of the graph says

"Randomization Dot plot of Slope", then the null hypothesis is $\beta_1 = 0$. Remember when StatKey simulates correlation we will be comparing the original r-value to all the simulated r-values in the simulation. When StatKey simulates the slope, we will be comparing the original sample slope to the simulated slopes. You will get about the same P-value from either of these. Assume the assumptions are met. Use the simulation in StatKey to answer the following questions.

- Write the null and alternative hypothesis for the correlation test. Address the quantitative relationship and label which is the claim. Is this a right-tailed, left-tailed, or two-tailed test?
- Write a sentence to explain the strength and direction based on the "original sample" correlation coefficient (r).
- Does the original sample correlation coefficient fall in a tail of the correlation simulation?
- Write a sentence to explain the original sample slope (b_1).
- Does the original sample slope fall in the tail of the slope simulation?
- Is the sample slope significantly different from zero? Explain your answer.
- Does the sample data significantly disagree with the null hypothesis? Explain your answer.
- Put the original sample slope into the slope simulation to calculate the P-value. What is your estimated P-value? (Answers will vary.)
- Write a sentence explaining the P-value.
- Compare the P-value to the significance level. Could the sample data or more extreme occur by sampling variability if the null hypothesis was true or is it unlikely? Explain your answer.
- Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.
- Write a conclusion for the test addressing evidence and the claim.
- Use the original sample slope, the estimated standard error in the simulation, and the following formula to calculate the T-test statistic. (Answers will vary.) Write a sentence to explain the T-test statistic.

$$T\text{-test statistic} = \frac{(\text{Slope} - 0)}{\text{Standard Error}}$$



30. Open the “Car Data” in Excel from www.matt-teachout.org. Copy and paste the miles per gallon (mpg) and horsepower into two columns in new excel spreadsheet. The mpg should be on the left and the horsepower should be on the right. The mpg will be the explanatory variable (X) and the horsepower will be the response variable (Y). Now go to www.lock5stat.com and click on StatKey. Under “Randomization Hypothesis Tests” click on “Test for Slope, Correlation”. Under “Edit Data” paste the two columns into StatKey. Now click “Generate 1000 Samples” a few times. Use the randomized simulation in StatKey and a 1% significance level to test the claim that there is a negative (inverse) relationship between mpg and horsepower.

31. Open the “Car Data” in Excel from www.matt-teachout.org. Copy and paste the weight and horsepower into two columns in new excel spreadsheet. The weight should be on the left and the horsepower should be on the right. The weight in tons will be the explanatory variable (X) and the horsepower will be the response variable (Y). Now go to www.lock5stat.com and click on StatKey. Under “Randomization Hypothesis Tests” click on “Test for Slope, Correlation”. Under “Edit Data” paste the two columns into StatKey. Now click “Generate 1000 Samples” a few times. Use the randomized simulation in StatKey and a 10% significance level to test the claim that there is a positive (direct) relationship between the weight and horsepower of a car.

32. Open the “Health Data” in Excel from www.matt-teachout.org. Copy and paste the age of women and the height of women into two columns in new excel spreadsheet. The age of women should be on the left and the height of women should be on the right. The age of women in years will be the explanatory variable (X) and the height of women in inches will be the response variable (Y). Now go to www.lock5stat.com and click on StatKey. Under “Randomization Hypothesis Tests” click on “Test for Slope, Correlation”. Under “Edit Data” paste the two columns into StatKey. Now click “Generate 1000 Samples” a few times. Use the randomized simulation in StatKey and a 5% significance level to test the claim that there is NO relationship between the age and height of women.

