Stat Support Activity: Correlation Hypothesis Test

Notes: Correlation Hypothesis Test Null and Alternative Hypothesis.

- Population Correlation Coefficient "rho" = ρ
- Population Slope "beta-one" = β_1
- If there is no correlation, ρ and β_1 will be zero.
- Two-tailed Correlation Test (Quantitative Relationship Test)
  $H_0: \rho = 0 \ or \ \beta_1 = 0$ (No correlation between the two quantitative variables.)
  $H_A: \rho \neq 0 \ or \ \beta_1 \neq 0$ (Is correlation between the two quantitative variables.)

- Right-tailed Positive Correlation Test (Direct Relationship)
  $H_0: \rho \leq 0 \ or \ \beta_1 \leq 0$ (No positive correlation between the two quantitative variables.)
  $H_A: \rho > 0 \ or \ \beta_1 > 0$ (Is positive correlation or direct relationship between two
  quantitative variables)
- Left-tailed Negative Correlation Test (Inverse Relationship)
  $H_0: \rho \geq 0 \ or \ \beta_1 \geq 0$ (No negative correlation between the two quantitative variables.)
  $H_A: \rho < 0 \ or \ \beta_1 < 0$ (Is negative correlation or inverse relationship between two
  quantitative variables)

Problem Directions: Write the null and alternative hypothesis in context for the following claims. Is the alternative hypothesis or the null hypothesis the claim?

1. "We claim that the age of a person (years) and the amount of money ($) in their savings accounts are directly related (positive correlation)."
2. "We claim there is no relationship between the resting heart rate (beats per minute) of a person and their cholesterol (mg per deciliter)."
3. "We claim there is an inverse relationship (negative correlation) between the costs and profits of a company in thousands of dollars."
4. "We claim that the gas mileage of a car (miles per gallon) is related to the time (seconds) it takes for the car to go from 0 mph to 60 mph.
5. "We claim there is not a direct relationship (positive correlation) between the time (months) this company has been in business and their revenue (thousands of dollars)."

Notes: Using the Correlation Coefficient (r)

- If r is close to +1 (like r = +0.893) then there is a Strong, Positive Correlation (line going up from left to right (positive slope) and the points in scatterplot are close to line) ,
  (r ≈ +0.6, +0.7, +0.8 , +0.9 usually indicate pretty strong positive correlation)

- If r is close to −1 (like r = −0.916) then there is a Strong Negative Correlation (line going down from left to right (negative slope) and the points in the scatterplot are close to the line)
  (r ≈ −0.6, −0.7, −0.8 , −0.9 usually indicate pretty strong negative correlation)

- If r close to zero (like +0.037 or −0.009) then there is a No linear correlation. Points in the scatterplot do not follow any linear pattern. There still could be a nonlinear curved pattern though. (r ≈ ±0.1, ±0.0 usually indicate no linear correlation)

Stat Support Activity: Correlation Hypothesis Test

- If r ≈ ±0.4, ±0.5 usually indicate moderate linear correlation. There is a linear pattern and points are moderately close to the regression line.

- If r ≈ ±0.2, ±0.3 usually indicate weak linear correlation. There is some linear pattern but the points are not very close the regression line.

- Correlation Coefficients (r) may be used to determine if the sample data significantly disagrees with the null hypothesis (no correlation). If r is close to zero, it indicates that data does not significantly disagree with the null hypothesis (no correlation). It can be difficult to know where no correlation ends and weak correlation begins. Computer programs often give critical values to determine the right and left tails of the normal r distribution. Think of the right tail as where "some positive correlation" begins. Think of the left tail as where "some negative correlation" begins.

Problem Directions: For each correlation coefficient critical value(s), draw a picture of the normal r distribution. Does the sample correlation coefficient fall in a tail? Describe the strength (weak, moderate, strong, or no correlation) and the direction (negative, positive, or none) of the sample correlation.

	Correlation Coefficient (r)	Tail of Test	Critical Value or Values	Does r fall in a tail determined by a critical value? (In tail or not in tail)	Strength of Correlation (Weak, Moderate, Strong, or No Correlation)	Direction of Correlation (Positive, Negative, or None
6.	r = 0.123	Right	0.316			
7.	r = −0.799	Two	±0.286			
8.	r = 0.921	Right	0.339			
9.	r = −0.026	Left	−0.279			
10.	r = 0.483	Two	±0.303			
11.	r = −0.834	Left	−0.316			
12.	r = 0.112	Two	±0.245			
13.	r = −0.461	Left	−0.292			

Notes: Using T-test Statistic

- Most statisticians prefer to use the T-test statistic to see if the sample data significantly disagrees with the null hypothesis.
- Sentence: T-test statistic for correlation counts the number of standard errors that the sample slope is above (+) or below (−) zero.
- Calculate the critical values and P-value with the T-distribution and n − 2 degrees of freedom.

- Calculate the T-test statistic using the following formula:

T-test statistic for correlation $= \dfrac{(Sample\ Slope - 0)}{Standard\ Error}$

Stat Support Activity: Correlation Hypothesis Test

- Calculate the degrees of freedom (df = n – 2) where n is the number of ordered pairs of quantitative data.

- Go to www.lock5stat.com and click on "StatKey". Under "Theoretical Distributions" click on "T". Use the given significance level and tail of the test to look up the critical value (where the tail or tails begin). Remember to put the significance level proportion in the top box of the tail. If it is a two tail test, put half of the significance level in the each of two top boxes in StatKey. The bottom box or boxes will be the critical value or values. Does your T-test statistic fall in a tail? If it does, then the sample data significantly disagrees with the null hypothesis.

- Go to www.lock5stat.com and click on "StatKey". Under "Theoretical Distributions" click on "T". Use the given significance level, tail, and test statistic to look up the P-value. Put the T-test statistic in the bottom box in StatKey. The top box will be the P-value. If it is two-tailed, you will need to add the top boxes to get the P-value. Is the P-value lower than the significance level? If it is, then it is unlikely to have occurred just because of sampling variability.

Problems Directions: Fill out the following table using a calculator and StatKey.

	Sam-ple Slope (b_1)	Stan-dard Error	T test stat	Sam-ple Size (n)	De-grees of Free-dom n – 2	Signif-icance Level (α)	Tail	Crit-ical value or values	T-test Stat in tail?	P Value	P Value less than α?
14.	−32.45	10.167		30		0.01	Left				
15.	+6.318	5.922		50		0.05	Right				
16.	+1.416	1.662		40		0.1	Two				
17.	−2.507	0.943		45		0.05	Left				
18.	+5.373	1.918		35		0.01	Right				
19.	−4.48	3.661		55		0.05	Two				
20.	−14.29	7.724		60		0.1	Left				
21.	+3.301	2.894		36		0.05	Right				