Sample Standard Deviation

_____ (y units) is a typical deviation of _____ (y-variable) from the mean _____ (y-variable) in the sample.

"2.83 inches is a typical deviation of a girl's height from the mean height of the girls in the sample."

Describing & Comparing Distributions

SOCS
- Shape
- Outliers
- Center
- Spread (clusters and gaps also)

"The center of the distribution of the boys (around 71 inches) is greater than the center of the girls' distribution (around 67 inches). The range for the boys (8 inches) is greater than the range for the girls (6 inches). The distribution of the boys' heights is positively skewed while the distribution of girls' heights is roughly symmetric."

Describing the Distribution of Bivariate Data

DOFS
- Direction
- Outliers and influential points
- Form
- Strength

"There is a strong positive linear association between lean body mass and metabolic rate. The data point at (45, 1200) appears to be an outlier."

Slope of Regression Line

For every 1 _____ (unit of x) increase in _____ (x-variable), the value of _____ (y-variable) is predicted to increase (decrease) by _____ (units of y).

"For each one mile per hour increase in running speed, the predicted pulse rate increases by 15.953 beats per minute."

Coefficient of Determination R^2

___ % of the variation in _____ (y-variable) is explained by the least-squares regression line relating _____ (x-variable) to _____ (y-variable).

"78.1% of the variation in running speed is explained by the least-squares regression line relating speed and age."

Correlation r

r = _____: There is a _____ (weak, moderate, or strong) _____ (positive or negative) linear relationship between _____ (x-variable) and _____ (y-variable).

"r = 0.769: There is a moderately strong positive linear relationship between the height of the plant and the amount of fertilizer applied."

p-Value

_____ is the probability, given that _____ (null hypothesis) is true, that you would get a value of _____ (statistic) as extreme as or more extreme as _____ (value of statistic).

"0.021 is the probability, given that the mean height of all women in the town is 63.5 inches, that we would get a t-statistic as extreme as or more extreme than 2.180."

Confidence Interval

We are _____ % confident that the interval from _____ (units) to _____ (units) captures the true population _____ (parameter).

"We are 95% confident that the interval from 62.4 to 67.5 inches captures the true population mean height of women in the town."

Confidence Level

In _____ % of all samples, the resulting confidence interval will contain the true _____ (parameter).

"In 95% of all samples, the resulting confidence interval will contain the true mean height of women in the town."

Hypothesis Test Conclusion

Since the p-value = _____ (p-value) is _____ (less than or greater than) α = (significance level), _____ (reject or fail to reject) H_0; therefore there _____ (is sufficient evidence or is not sufficient evidence) that _____ (conclusion with reference to parameter in context).

"Since the p-value = 0.001 is less than alpha = 0.05, reject the null hypothesis. Therefore there is sufficient evidence that the mean height of women in the town is less than 61 inches."

Standard Errors

Standard Error of the Residuals s – *This measures the amount by which the points in a scatterplot vary vertically from the least-squares regression line.*

"On average, the predicted student exam scores varied from the line of best fit by 4.775 points."

Standard Error of the Slopes of the Regression Line s_b – *Given a large number of samples, this measures the variation in the set of slopes of the regression lines from these samples.*

"0.237 is an estimate of the standard deviation of the slopes of the regression lines linking exam score and quarter score over all possible samples of 20 students."

Sample Selection & Random Assignment

Label the members of the population and then use slips of paper, random digits, or technology to select the sample. When using random digits use leading zeros to make numbers the same length; when using technology do not.

"Label the 500 students with numbers from 001 to 500. Using a random number table, record 3-digit numbers, skipping any that are not between 001 and 500 and any repeated numbers, until you have 25 unique numbers between 001 and 500. Survey the 25 students whose numbers were chosen."

Simulations

Describe how to use a chance device to imitate one repetition of the process being simulated. Then perform many repetitions.

What is the probability that, in a random selection of 10 passengers, none from first class are chosen?
"Number first class passengers 01-12 and the other passengers 13-76. Choose ten 2-digit numbers from the random number table, skipping any that are not between 01 and 76 and any repeated numbers. Count the number between 01 and 12. Divide the result by 10 to compute the simulated probability."