

Applied Stats I

Exam One Review

Disclaimer: The following questions are meant to serve as preparation, and are examples of what may appear on the first exam. Do not rely on this content as your sole means of preparation, this is meant to guide your studying.

1 Terms

Be able to define these terms and describe why they're important:

- Describing data:
 - Parameter vs. statistic
 - Data
 - Observations
 - Population
 - Sample
 - Variable
 - Qualitative/quantitative data
 - Nominal/ordinal
 - Granularity (continuous/discrete)
 - Skew
 - Outliers
- Measures of Central Tendency and Dispersion:
 - Mean
 - Median
 - Variance
 - Standard deviation
- Distributions
 - Normal Distribution
 - T Distribution
 - Degrees of Freedom
 - Standard deviation/variance
 - Sampling distribution
 - Central Limit Theorem
 - Standard error
- Estimation
 - Point estimate
 - Confidence interval
 - Confidence level
 - Bias
 - Efficiency

- Hypothesis testing
 - Hypothesis
 - Null/alternative hypotheses
 - Test statistic
 - P-value
 - Significance level (α -level)
 - Type I & Type II error
- Regression
 - Linear regression
 - Regression analysis
 - Y-intercept
 - Slope
 - Least squares line
 - Sum of squared error
 - Residual sum of squares
 - Total sum of squares
 - Correlation coefficient (r/R)

2 Calculate/execute these concepts:

- Measures of central tendency:
 - Mean
 - Median
- Measures of dispersion:
 - Variance
 - Standard deviation
 - Standard error
- Distributions and probability:
 - Define a sampling distribution of a mean
 - Calculate a Z-score
 - Use the Z-score to find p-value
- Regression
 - Interpret y-intercept and slope for a linear function
 - Write a prediction equation
 - Calculate Sum of Squared Errors
 - Calculate Total Sum of Squares
 - Interpret a scatter plot
 - Construct a CI around β
 - Conduct a hypothesis test for β
 - Calculate standard error for β
 - Calculate S_x and S_y
 - Calculate r and interpret its meaning
 - Calculate r^2/R^2
 - Interpret a regression analysis table
 - Interpret regression coefficients

3 Practice problem set:

- The following table contains the GDP per capita (in thousands of international dollars) for four European countries.

Belgium	Germany	France	Luxembourg
38	38	35	90

- Find the variance.
- Find the standard deviation.
- Would you say that one of these observations is an outlier?

- The "Freshman Fifteen" is an expression that commonly refers to an amount (somewhat arbitrarily set at fifteen pounds) of weight often gained during a student's first year at university. You decide to test whether this expression holds true for Trinity College students. You randomly select 16 second years and gather data on how much weight (in pounds) they gained the previous year. The mean change of your data is 14.5 lbs. and the sample standard deviation is 0.8 lbs.

1. Identify the population for this study.
2. Describe the sample distribution for this study.
3. Describe the sampling distribution for this study as precisely as possible.
4. Calculate the point estimate and a 95% confidence interval for the population mean. Explain what your confidence interval means.

- Imagine you are interested in the different patterns of support for the Spanish government among citizens of Catalan population. You decide to conduct a survey asking people "Do you have confidence in the national government?" Possible answers include Yes or No. You were able to poll 243 Catalans. Of these 243 respondents, 86 said, "Yes."

1. Provide a point estimate for the percent of Catalans that have confidence in the government.
2. Identify the sampling distribution of this study. Be precise.
3. Construct a 92% confidence interval of the percent of Catalans that have confidence in the government.
4. Test the theory that less than 40% of Greeks support the government using a 0.05 significance level.

- Suppose a random sample is taken of 200 rat-hunting dogs in New York City. The mean number of rats killed by a dog is 19, with a standard deviation of 2. Construct and interpret a 92% confidence interval for the mean number of rats killed.

- The distribution of Quantitative GRE scores for graduate students who drop out of graduate school has a mean of 550 and standard deviation of 30. These scores are distributed normally.
 1. Assume that you do not know the true population mean and standard deviation. You take a random sample of 24 graduate school dropouts and calculate a mean of 570 and a standard deviation of 16. Construct a 95% interval. Does this contain the true population mean?
 2. The average GRE scores for all graduate students is 582.5. Using only the sample described in part 1, test the research hypothesis that graduate school dropouts have lower GRE scores than average. Specify all five steps for conducting a hypothesis test. Use a 0.05 level of significance.