

Stat 201 – Fall 2019

Topics to study for Exam 1

Chapter 1: Type of Variables

- Quantitative versus categorical variable
- Recognizing what is not a variable at all
- Being able to recognize an “Identifier Variable”

Chapter 2: Quantitative Variables

- Calculating a Range and the IQR (given Q1 and Q3)
- When to use mean and s versus median and IQR
- Interpreting median, mean, quartiles, IQR, range and standard deviation
- Interpreting shape, center, spread and unusual features from a histogram
- Comparing histograms versus stem and leaf displays
- Identifying symmetry – skewness – outliers – gaps (from histograms or stem and leaf)
- Calculating a median by hand and interpreting it
- Know what the 5 number summary is
- Calculating a mean by hand and interpreting it
- Given the variance, calculate the standard deviation

Chapter 3: Categorical Variables

- Reading pie charts, bar charts, comparing the two
- Frequency table versus relative frequency table
- Being able to calculate a relative frequency
- Being able to calculate a marginal distribution
- Being able to calculate a conditional distribution
- Interpreting segmented bar charts, mosaic plots
- Recognizing independence or dependence from mosaic plots

Chapter 4:

- Interpreting side by side box plots
- Comparing distributions (using histograms)
- Identifying and interpreting box plot components
- Use of common scaling on histograms when making comparisons
- Recognizing a trend in a time plot

Pivot Tables & Pivot Charts:

- Reading Pivot Tables in Excel
- Being able to recognize unusual information displayed in Pivot Tables and/or Charts
- Being able to make recommendations to make presented information better

Chapter 5:

- Calculating Z-scores, comparing two values
- 68-95-99.7 rule
- Understanding Z-scores, s as a "ruler"
- Interpreting normal probability plots and goodness of fit tests
- Identifying outliers
- Understanding the difference between positive and negative Z-scores
- Recognizing the normal model, relationship between normal model and Z-scores
- "Nearly Normal Condition"
- Be comfortable with output from the "David Lane" Normal Curve calculator
- Note: no use of the "Z-table"

Chapter 6

- Know which type of variables we use to create a scatter plot
- Interpret a scatter plot
- Direction (distinguish positive versus negative relationship)
- Form (distinguish linear versus curvy relationship)
- Strength (distinguish strong versus weak relationship)
- Unusual Features (be able to spot an outlier or groups)
- Know the difference between explanatory and response variables
- Know the bounds for r
- Be able to match r to example scatter plots
- Know the 3 necessary conditions for correlation analysis
 - Quantitative variables
 - Straight-enough condition
 - No outlier condition
- Describe what it means for a correlation to be -1 , 0 , and 1
- Understand the difference between correlation and causation
- Have a general understanding of what lurking variables are

Chapter 7: Linear Regression

- Four conditions for valid regression
 - Quantitative variables
 - Straight-enough condition
 - No outlier condition
 - "Does the plot thicken?" condition
- What is special about the regression ("least squares") line compared to any other line drawn through the data
- Given JMP output, write out the regression model with actual variable names
- Interpret regression coefficients (b_0 and b_1)
- Know when b_0 has no logical interpretation
- Know the difference between y and \hat{y}
- Be able to use a regression equation to make an estimate of y for a given value of x
- Be able to calculate and interpret a residual
- Be able to interpret a residuals plot and spot problems
- Be able to find r from r^2 or vice versa
 - Taking care of the direction of the slope
- Interpretation of R-square as the percentage of the variation in y associated with the variation in x
- Don't interpret the relationship to imply cause and effect

Chapter 8: Regression Wisdom

- Understand extrapolation
 - Dangers
 - Valid only when one can safely assume that the model will hold outside the range of the data into the region where the extrapolation is desired
- Groups in residual plots
- Correlation is not causation
- Lurking variable and how they explain a correlation that is not due to causation
- Identify outliers from a scatter plot