

# Stat 201 – Fall 2024

## Topics to study for Exam 2

### 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-Square 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

# Stat 201 – Fall 2024

## Topics to study for Exam 2

### **Chapter 10: Sample surveys**

- Populations and samples
- Parameters and statistics
- Representative samples
- Sampling frame
- Nonrandom (Bad) samples
  - Voluntary samples
  - Convenience samples
- Randomized samples
  - Simple random samples
  - Stratified random samples
  - Cluster random samples
  - Systematic samples
  - Multistage samples
- Sampling biases
  - Nonresponse
  - Undercoverage
  - Response bias

### **Chapter 12: From Randomness to Probability**

- Formal definition of probabilities
  - Probabilities have to be between 0 and 1 inclusive
  - What it means to have a probability of 0 or 1
  - Recognize valid and invalid assignments of probabilities
- Law of large numbers
- Know that “disjoint” is the same as “mutually exclusive” and means that if one event happens, the other can’t happen
- Know what it means for two events to be independent
- Know that disjoint events are never independent (and, independent events are never disjoint)
- Understand the addition rule and necessary conditions for using this rule
- Understand the multiplication rule and necessary conditions for using this rule
- Understand complement rule and be able to apply it [ $P(\text{not } A) = 1 - P(A)$ ]
  - Recognize that when the words “probability of at least one....” are used, this typically involves using this rule
- State whether or not it would be reasonable to assume a set of events is independent or disjoint

# Stat 201 – Fall 2024

## Topics to study for Exam 2

### **Chapter 13: Sampling Distributions for proportions**

- Know the difference between a population and a sample
- Know what  $\hat{p}$  and  $p$  represent
- Knowing how to identify a “success”
- Have a general understanding of sampling distributions for a proportion
- Know how to find the mean and standard deviation of the sampling distribution of  $\hat{p}$ , given the known population proportion,  $p$ , and the sample size,  $n$ .
- Know when the shape of a sampling distribution of  $\hat{p}$  is approximately Normal
- Independence Assumption or as a practical approximation
  - Randomization
  - 10% condition
  - Success/Failure condition
- Be able to use the 68-95-99.7 rule to find the approximate probability of a specific  $\hat{p}$  (or less) from a normally distributed sampling distribution of  $\hat{p}$ .

### **Chapter 14: Sampling Distributions for Means**

- Know the difference between a population and a sample
- Have a general understanding of sampling distributions for a mean
- Know what  $\bar{y}$  and  $\mu$  represent
- Have a general understanding of the Central Limit Theorem
- Know assumptions for CLT when the parameters ( $\mu$  and  $\sigma$ ) are known
  - Independence condition or as a practical approximation
    - Randomization condition
    - 10% condition
  - Large enough sample condition
- Know the difference between two distributions
  - Real-world distribution of the sample that one might see in a histogram
  - Math-world distribution of a sample statistic such as the sample mean
- Given the mean and standard deviation of a population, be able to calculate the mean and standard deviation of the sampling distribution of  $\bar{y}$  (given the sample size  $n$ )