Overview (Activity #1)

1. Define the terms *probability, statistics (descriptive statistics),* and *Statistics (statistical inference)*
2. Make decisions based on reported probabilities
3. Explain the Law of Large Numbers and “Law of Small Numbers”
4. Explain why random assignment is important in designing studies
5. Calculate simple proportions, odds, relative rates, and odds ratios from a contingency table
6. Given the general definition of a null hypothesis, write the null hypothesis for a specific study
7. Use a simulation method (randomization) to estimate the p-value of a statistical test
8. Explain the difference between a probability and a likelihood

Probability Basics (Activities 2-3)

9. Define probability using either the relative frequency approach or the classical approach
10. Write out the sample space for simple and compound experiments
11. Apply the slots method (multiplication, factorial, permutation rules) or combinations to calculate the potential outcomes of an experiment
12. Explain the difference between a combination and a permutation using an example of each
13. Recognize when outcomes from an experiment are equally likely (or are not equally likely) to occur
14. Use the complement rule to calculate probabilities of disjoint (mutually exclusive) events
15. Give an example of two or more events that are disjoint; give an example of two events that are not disjoint
16. Draw Venn Diagrams to represent probability rules (such as the general addition rule or the complement rule)
17. Use Venn Diagrams or the general addition rule to calculate probabilities
18. Write out probability statements using correct set notation
19. Given a frequency table, determine probabilities
20. Given a list of probabilities, fill-in a frequency table
21. Use combinations and the definition of probability to solve simple probability problems (beginning hypergeometric probabilities)
22. Use a simulation to estimate probabilities from an experiment
23. Explain solutions to the birthday problem (probability of at least 2 people sharing a birthday) and the Let’s Make a Deal problem
24. Use Venn Diagrams and probability rules to solve probability applications
25. Explain the following problem (sample space comprehension): Amy has 2 children; the older child is a female. Barb has 2 children; one child is female. The probability that the youngest child is female is 1/2 for Amy and 1/3 for Barb.

Method of Randomization (Activity 3a)

26. Explain the difference between an experimental study and an observational study
27. Write out the null and alternate hypotheses for a given study
28. Write out all possible outcomes of a study under Fisher’s method of randomization
29. Use combinations to determine the number of ways of randomly assigning X people into G groups
30. Use Fisher’s method of randomization to calculate a p-value in a given study
31. Write the potential consequences of Type I and Type II errors in a given study
32. Write out an interpretation of a p-value in a given study
33. Given a simple data set, complete an analysis using the Method of Randomization

Discrete Random Variables & Conditional Probability (Activities 4, 5, 5a)

34. Given a list of probabilities, sketch a probability mass function
35. Given a list of probabilities, sketch a cumulative distribution function
36. Given a cumulative distribution function, calculate specific probabilities: What is $P(X<3)$?
37. Explain why a cumulative distribution function always starts at zero and ends at 1.0
38. Explain what is meant by a *conditional* probability
39. Write conditional probabilities using correct set notation: $P(A \mid B)$
40. Use Venn Diagrams or the conditional probability rule to calculate conditional probabilities
41. Use the general multiplication rule to calculate probabilities: $P(A \cap B)$
42. Given a frequency table, calculate conditional and joint probabilities (like the Alan & Beth movie example)
43. Explain what it means for two events to be independent
44. Use the conditional probability rule or the general multiplication rule to show that two events are independent
45. Use the law of total probability to calculate the probability of an event
46. Use Bayes’ Theorem to calculate posterior probabilities
47. Verify de Morgan’s Laws by sketching and shading Venn Diagrams

Discrete Random Variables, Expectation, and Variance (Activity #6)

48. Use the formula to calculate the expected value of a random variable
49. Explain what the expected value of a random variable represents
50. Explain the difference between an expected value and our expectation (of an outcome of an experiment)
51. Use the properties of expected values to determine what will happen if each value of X undergoes a linear transformation
52. Use the formula to calculate the variance of a random variable
53. Use the formula to calculate the standard deviation of a random variable
54. Use the properties of variances to determine what will happen if each value of X undergoes a linear transformation

Binomial Distribution (Activity #7 & Dog Experiment)

55. Define a discrete random variable
56. Explain the properties of a Bernoulli random variable
57. Derive the expected value and variance of a Bernoulli random variable
58. Evaluate a situation to see if the Binomial distribution applies (independent trials, constant probability of success)
59. Derive the pmf of a binomial distribution
60. Use the pmf to calculate binomial probabilities: $P(X<a)$ $P(X=a)$ $P(X>a)$
61. Derive the expected value of a binomial distribution
62. Calculate the expected value and variance of a binomial random variable
63. Use a calculator to calculate binomial probabilities
64. Use a binomial table to find binomial probabilities
65. Use the complement rule and a calculator to calculate $P(X>a)$ under a binomial distribution
66. Verify the results of a published research article

Binomial Test & Sign Test (Activity #7a)

67. Write null and alternate hypotheses (for the binomial or sign tests) using correct probability notation
68. Assuming the null hypothesis is true, calculate the p-value of an experiment (probability of observing something as or more extreme)
69. Draw an appropriate conclusion from a p-value
70. Conduct a complete analysis using the binomial test
71. Conduct a complete analysis using the sign test

If time permits...

72. Calculate the desired probability using information typically collected under the method of randomized responses (in surveys)
73. Calculate false positive, false negatives (sensitivity, specificity) rates for a given situation (drug testing, polygraph)

Discrete Distributions (Activities 8, 8a, 8b, 8c)

74. Define the term *random*
75. Derive pmf's for variables following geometric, negative binomial, hypergeometric, or poisson distributions
76. Derive formulas for the expected value and variance under geometric, negative binomial, hypergeometric, and poisson distributions
77. Use the formulas to calculate probabilities, expected values, and variances under these distributions
78. Sketch the pmf and cdf for a specific case under each distribution
79. Interpret the probabilities, expected values, and variances in an application of each distribution
80. Given an application, determine the appropriate discrete probability distribution to use
81. Explain how hypergeometric random variables differ from binomial random variables (sampling w/o replacement)
82. Explain the conditions under which each distribution can be used to model probabilities
83. Calculate probabilities for each distribution using a calculator (or computer)
84. Apply discrete distributions to solve realistic problems (hypergeometric probabilities, odds ratios, relative rates)
85. Use the multinomial distribution to calculate probabilities

Continuous Distributions (Activities 9, 9a, 9b)

86. Define a continuous random variable
87. Classify variables as either discrete or continuous
88. Given a histogram, calculate probabilities
89. Given a histogram, explain how probability is equivalent to the area under the histogram
90. Explain how to calculate probabilities (area under curves) or continuous random variables
91. Explain the conditions needed to demonstrate that a pdf is legitimate (area sums to 1.0; always positive)
92. Using either geometry or integration, calculate the area under a variety of curves
93. Explain why P(X=a) is always zero with continuous random variables
94. Derive the formulas for the expected value and variance of continuous random variables
95. Given a valid pdf, calculate expected values, variances, and percentiles of continuous distributions
96. Define the pth percentile of a distribution
97. Apply pdfs, area calculations, expected values, and percentiles to solve a realistic problem (industrial robot problem)

Special Continuous Distributions (Activity 10)

98. Calculate probabilities and expected values for variables following a uniform distribution
99. Explain the conditions under which exponential distributions may be appropriate
100. Given the pdf of an exponential distribution, derive the cdf
101. Model a situation using an exponential distribution and calculate probabilities
102. Provide an example of the memoryless property of exponential distributions
103. Use the gamma, beta, and weibull distributions to calculate probabilities
104. Determine the appropriate continuous distribution to model a given situation

Normal Distribution (Activity 11)

105. Describe the visual characteristics of a normal curve
106. Compare and contrast normal distributions for a variety of variables (location, spread)
107. Determine the appropriateness of using a normal distribution to model given random variables
108. Sketch symmetric, positively skewed, and negatively skewed distributions
109. Given the pdf for a standard normal distribution, prove the inflection points are at +/- 1
110. Sketch normal distributions given an expected value and variance
111. Explain why we cannot integrate to calculate normal curve probabilities
112. Use the empirical rule to state what percentage of observations lie within 1, 2, and 3 standard deviations of the mean
113. Interpret a z-score from a normal distribution
114. Calculate z-scores from a normal distribution
115. Explain what happens to the shape of a normal distribution when we transform it to a standard normal distribution
116. Calculate probabilities from a normal distribution
117. Use a calculator to calculate normal distribution probabilities
118. Calculate the pth percentile of a normal distribution
119. Standardize scales using z-scores and interpret the results (identify limitations)

Data Collection (Activity 12)

120. Classify variables into quantitative, qualitative, nominal, ordinal, interval, and ratio
121. Explain simple random, stratified, cluster, systematic, and convenience sampling methods
122. Determine whether a sampling method introduced bias (selection, response, measurement, nonresponse biases)
123. Define *independent* and *dependent* variables
124. Identify the sampling method, independent variable(s), and dependent variable(s) from a given research article

Exploratory Data Analysis (Activities 12a, 13)

125. Derive various methods to determine the "center" of a distribution

126. Demonstrate how the median minimizes the sum of absolute deviations
127. Prove (using calculus or graphical methods) how the sample mean minimizes the sum of squared deviations
128. Given a dataset, create a visual display (stemplot, histogram, and boxplot)
129. Given a visual display, draw conclusions about a dataset
130. Calculate the mean, median, and mode of a dataset
131. Determine which measure of central tendency is most appropriate for a given dataset
132. Determine the impact of an outlier (or linear transformation) on measures of central location
133. Calculate the pth percentile of a sample dataset
134. Calculate the IQR, range, variance, and standard deviation of a sample dataset
135. Calculate the population variance and unbiased estimate of the population variance
136. Determine the impact an outlier (or linear transformation) on measures of spread
137. Derive formulas to show the impact of a linear transformation on the mean and variance

Estimates and Estimators (Activities 14, 15)

138. Use correct notation to define population parameters (Greek letters) and point estimators
139. Use intuition to select the most appropriate point estimator of a parameter
140. Explain what is meant by a biased point estimate
141. Prove that the sample mean and sample proportion are unbiased estimators for the population mean and proportion
142. Explain why we want the variance of a point estimate to be minimized
143. Derive the variance of the sample mean and variance
144. Interpret the components of the mean square error of a point estimate
145. Explain why the unbiased estimate of the population variance has (n-1) in the denominator
146. Explain the term *degrees of freedom*
147. Explain the concept of *maximum likelihood*
148. Maximize the likelihood function in a simple (binomial) case
149. Use maximum likelihood to estimate item parameters in an Item Response Theory example

Sampling Distributions (Activity 16)

150. Given a small population, simulate the sampling distribution of the sample mean through repeated sampling
151. Given a population with a normal distribution, explain the shape and center of the sampling distribution for the sample mean
152. Define the term *standard error*
153. Explain what happens to the standard error as our sample size increases
154. Prove that the expected value of the sample mean is equal to the population mean
155. Prove that the variance of the sample means is equal to the population variance divided by the square root of the sample size
156. Given a population mean and standard deviation, sketch the sampling distribution of the sample mean for various sample sizes

Sampling Distributions (Activities 17, 17a, 17b)

157. Write out the Central Limit Theorem (explain it in your own words)
158. Explain the conditions under which the CLT “works”
159. Calculate probabilities from the sampling distribution: $P(\bar{X} < a)$
160. Run a computer simulation to verify the CLT

Confidence Intervals & t-distribution (Activities 18, 19)

161. Derive the formula for a confidence interval for the population mean
162. Interpret a confidence interval
163. Explain why we cannot say, “We are xx% confident that the population mean falls in our interval.”
164. Explain how our chosen level of confidence impacts the width of a confidence interval
165. Explain how the sample size impacts the width of a confidence interval
166. Given a desired confidence interval width, determine the necessary sample size
167. Calculate confidence intervals for given situations
168. Derive the formula for a confidence interval for the population proportion
169. Explain the formula for the standard error of the sampling distribution of the proportion
170. Calculate confidence intervals for the proportion in a given situation
171. Explain why a t-distribution must be used when the population standard deviation is unknown
172. Determine the degrees of freedom for a confidence interval for the population mean
173. Calculate confidence intervals using the t-distribution
174. Evaluate different interpretations of confidence intervals (determine which are correct/incorrect)
175. Explain why a confidence interval may be meaningless with a nonrepresentative sample

Hypothesis Testing (general concepts) (Activity 20)

176. Given a study, identify the research goal, target population, sample, sampling procedure, parameter of interest, observed estimator, dependent variable, and independent variable
177. Determine if a given study is observational or experimental
178. Write out the null and alternate hypotheses
179. Explain why hypotheses are written with respect to parameters (and not statistics)
180. Determine if a 1-tailed or 2-tailed test should be used in a given study
181. Define Type I (alpha) error and Type II (beta) error
182. Define power
183. Explain the potential consequences of alpha and beta errors in a given study
184. Set an appropriate level for alpha
185. Explain the logic of statistical inference (hypothesis testing)
186. Explain why we must assume the null hypothesis is true in order to conduct a hypothesis test
187. For a given study, sketch the sampling distribution and locate the critical value (z-score)
188. Convert a critical value in the z-score metric to the sample mean metric

189. Convert an observed sample mean into a z-score
190. Determine whether to retain or reject the null hypothesis by comparing observed and critical values
191. Explain why we never accept the null hypothesis
192. Calculate and interpret the p-value from a hypothesis test
193. Given distributions for both the null and alternate hypotheses, shade the areas corresponding to alpha, beta, and power
194. Calculate the probability of a Type II error and power for a given study

One-sample hypothesis tests for the mean and proportion (Activities 21, 21b, 21c, 22, 22a)

195. Explain the conditions under which we use a t-test rather than a z-test
196. Estimate the p-value under a t-test
197. Given a specified Type I error rate and a p-value, determine whether to retain or reject the null hypothesis
198. Explain the difference between *statistical significant* and *practically significant* results
199. Determine the impact of alpha error rate on power
200. Determine the impact of sample size on power
201. Given a set of data, complete a hypothesis test and write the conclusions
202. Explain the relationship between hypothesis testing and confidence intervals
203. Conduct hypothesis tests on a calculator or computer
204. Interpret results from computer output of a hypothesis test
205. Derive formulas in order to test hypotheses about a population proportion

Independent samples hypothesis tests for the mean and proportion (Activities 23, 24, 24a, 24b)

206. Write appropriate hypotheses regarding the means from two independent samples
207. Derive and sketch the sampling distribution of the difference in means
208. Derive the formula for s_{pooled} (weighted average standard deviation)
209. Determine when to pool variances (equal variance assumption) or to use the Welch-Satterthwaite Method
210. Derive the formula for the confidence interval of the difference in means
211. Conduct an independent samples hypothesis test (z-test and t-test)
212. Determine the appropriate degrees of freedom from a given study
213. Write appropriate conclusions from an independent samples t-test
214. Explain the assumptions needed to conduct an independent samples t-test (independence, equal variances, normal distributions)
215. Explain how to test if two populations have equal variances
216. Explain how to test the normality of a distribution (p-p plots, histograms, chi-square tests)

Dependent samples (matched-pairs) hypothesis tests for the mean (Activity 25)

217. Given a study, determine if the groups are independent or dependent (matched)
218. Derive the formulas and sampling distribution for a dependent samples t-test
219. Conduct a complete dependent samples t-test
220. Explain why dependent samples tests have higher power than independent samples tests

MATH 301 Outline

- I. Review
  - a. Experimental Design
  - b. Exploratory Data Analysis
  - c. Sampling distributions
    - i. Standard Error
    - ii. Mean
  - d. Confidence intervals
  - e. Hypothesis Tests
    - i. Error Rates
    - ii. Power

- II. Distributions of sample variance
  - a. Chi-squared distributions
    - i. Testing single variances
    - ii. Confidence Intervals
    - iii. Relationship to standard normal distribution
  - b. F- distributions
    - i. Testing two variances
    - ii. Confidence Intervals
    - iii. Relationship to chi-squared distribution
    - iv. Relationship to t-distribution

- III. Analysis of Variance
  - a. Advantage over multiple t-tests (alpha error)
  - b. Visual/conceptual comparison of means
  - c. Sums of squares
    - i. SS Total
    - ii. SS Error
    - iii. SS Between
    - iv. Partitioning SS
  - d. Degrees of freedom
  - e. Mean squares
    - i. MS Treatment
    - ii. MS Error
  - f. Mean square ratio
    - i. Critical F-value
  - g. Statistical model
    - i. Notation
      - 1. Treatment effect
      - 2. Error
      - 3. Group means
      - 4. Individual means
  - h. Expected value of mean squares
    - i. Under true null hypothesis
    - ii. Under false null hypothesis
  - i. Mean squares as estimates of variance
    - i. Use of F-distribution
  - j. ANOVA summary tables
  - k. Conclusions based on ANOVA
    - i. Effect Size
  - l. ANOVA Assumptions
    - i. Independence
    - ii. Normality
      - 1. Histograms
      - 2. Q-Q, P-P Plots
    - iii. Homogeneity of Variance
      - 1. Fmax test
      - 2. Bartlett's test
    - iv. Effect of violation of assumptions
      - 1. Robustness
      - 2. Modifying degrees of freedom (noncentral chi-squared)
  - m. Follow-up Tests
    - i. Bonferroni
      - 1. Experiment-wise (family-wise) error rate
      - 2. Adjustment
    - ii. Tukey's pairwise comparisons
      - 1. Critical difference
      - 2. Computations
    - iii. Scheffe
      - 1. Contrasts
      - 2. Computations

- IV. Experimental Design

a. AxB ANOVA
    i. Graphical Analysis (means plots)
    ii. Interaction
    iii. Formal statistical model
        1. Treatment effects
        2. Interaction
    iv. Partitioning SS
        1. SS Total
        2. SS A
        3. SS B
        4. SS AxB
        5. SS Error
    v. Summary table
    vi. Mean Squares
        1. Expectation
        2. Test Statistic
    vii. Conclusions
    viii. Follow-ups
        1. Simple effects
        2. Main effects
        3. Follow-up tests

b. AxS ANOVA
    i. Advantages of CRD
    ii. Statistical Model
        1. Treatment effect
        2. Error
            a. Subject effect
            b. Random error
    iii. Summary table
        1. Mean square ratio
        2. Critical value
        3. Conclusions
    iv. Follow-up tests

c. Other experimental designs
    i. Groups within treatments
    ii. Random replications
    iii. Latin Square
    iv. Split Plot

V. Bivariate Analysis
  a. Nominal/Ordinal Scale
    i. Chi-squared Goodness-of-Fit
    ii. Chi-squared tests for independence
        1. Limitations of chi-squared tests
        2. Applications of goodness-of-fit tests
    iii. Phi coefficient
    iv. Cramer's coefficient
  b. Interval/Ratio Scale
    i. Scatterplots
    ii. Pearson's Product Moment Correlation
        1. Inferences about correlation coefficient
    iii. Spearman's Rho
    iv. Kendall's Tau
    v. Characteristics of correlation coefficients
        1. Range reduction
        2. Maximum/minimum values
    vi. Visualization of linear regression
        1. Least squares regression line

VI. Linear Regression
  a. Conditional means approach
  b. Linear equations
  c. Visual estimation
  d. Least squares criteria
    i. Error
    ii. Squared error
    iii. Minimized squared error
  e. Formal model
    i. Full model
    ii. Reduced model

f. Beta weights
    i. Y-intercept
    ii. Slope
g. Indices of accuracy
    i. SSE
    ii. Standard error of estimate
    iii. 1 – R-squared
    iv. Coefficient of Determination
h. Assumptions
    i. Linearity
    ii. Homoscedasticity
    iii. Normality
    iv. Existence
    v. Independence
    vi. Violation of assumptions
i. Principle of parsimony
j. Significance of beta weights
    i. ANOVA summary table
    ii. Full vs. reduced models
    iii. Visual representation
    iv. Descriptive statistics
    v. Hypothesis tests
    vi. Conclusions

VII. Multiple Linear Regression
a. Modeling
    i. Full Model
    ii. Reduced Model
    iii. Model Selection Procedures
        1. Forward entry
        2. Backwards
        3. Stepwise
        4. Hierarchical
        5. Grouping independent variables (data reduction)
b. Sums of Squares
    i. SS Regression
    ii. Partial SS
    iii. SS Error
    iv. Visual displays
c. Omnibus F-test
    i. Using SS
    ii. Using R-squared

VIII. Nonparametric Tests
a. Specific Tests
    i. Sign Test
    ii. Mann-Whitney U
    iii. Fisher’s Exact Test
    iv. Independent \/Dependent/ANOVA variants
b. Bootstrap Method
c. Jackknife Method
d. Checking normality
    i. Q-Q Plots
    ii. Histograms
    iii. Chi-squared goodness-of-fit
    iv. Kolmogorov-Smirnoff Test
e. Nonparametric vs. Parametric
    i. Relative Efficiency

IX. Measurement
a. Constructs
b. Operationalization
c. Sources of error

MATH 400 Outline

X. In-Depth Exploratory Data Analysis
a. Numerical Summaries
    i. Selecting the best numerical summary
    ii. What’s lost in the summarization?
b. Graphical Analysis
    i. Charts
        1. Stemplots

            2. Dotplots
            3. Scatterplots
            4. Time series
            5. Boxplots
        ii. Evaluating visual displays
            1. Characteristics of effective visual displays
                a. Data density
                b. Chartjunk
                c. Clarity; not simplicity
                d. Case studies
        iii. Creating effective visual displays
            1. Bivariate displays
            2. Univariate displays
            3. Designs for specific purposes

XI. Applied Probability
    a. Conceptual review
    b. Applications
        i. Systems reliability
        ii. Relationship to significance testing
        iii. Compound probabilities
        iv. Tests for independence

XII. Significance Testing
    a. Applications
        i. One sample mean
        ii. Independent samples means
        iii. Dependent samples means
        iv. Proportions
        v. ANOVA
    b. Evaluating Significance Tests
        i. Understanding p-values
        ii. Effect of sample size
        iii. Effect size
        iv. False dichotomy of hypotheses
        v. Null hypothesis – never true?
    c. Nonparametric Applications

XIII. Applied Regression Analysis
    a. Review of linear regression
    b. Nonlinear regression
    c. Model-selection procedures
    d. Prediction models
    e. Explanation models

XIV. Statistical Applications
    a. SPSS
        i. Hypothesis tests
        ii. Regression analyses
        iii. Exploratory analysis
        iv. Data cleansing
        v. Data management
        vi. Data merging
        vii. Defining variables
    b. StatCrunch Software
    c. SAS
    d. STATA
    e. R

XV. Other Topics
    a. Multivariate Statistics
        i. Multivariate normal distribution
        ii. Matrix computations
    b. Principal Components Analysis
    c. Factor Analysis