INTRO TO STATISTICS COURSE

Week	Module Number	Topics/Concepts	Learning Outcome(s)	Bloom's Taxonomy Level*	Activities and/or Assessments
1	1	Intro to stats	• Identify variables as numerical and categorical. • Define associated variables as variables that show some relationship with one another. Further categorize this relationship as positive or negative association, when possible. • Define variables that are not associated as independent. • Identify the explanatory variable in a pair of variables as the variable suspected of affecting the other. • Classify a study as observational or experimental and determine and explain why whether the study's results can be generalized to the population and whether they suggest correlation or causation between the variables studied. • Question confounding variables and sources of bias in a given study.	Knowledge Comprehension Application Analysis	Intro discussion Reading Videos Practice Quiz
2-3	2	Study designs	• Distinguish between simple random, stratified, cluster, and multistage sampling, and recognize the benefits and drawbacks of choosing one sampling scheme over another. • Identify the four principles of experimental design and recognize their purposes: control any possible confounders, randomize into treatment and control groups, replicate by using a sufficiently large sample or repeating the experiment, and block any variables that might influence the response. • Identify if single or double blinding has been used in a study.	Knowledge Comprehension Application	Q&A discussion Reading Videos Quiz Assignment
4-5	3	Representing Data	• Use scatterplots for describing the relationship between two numerical variables making sure to note the direction (positive or negative), form (linear or non -linear) and the strength of the relationship as well as any unusual observations that stand out.	Comprehension Application Analysis	Q&A discussion Reading Videos Quiz Assignment

Contributed with permission by Marie Lippens, University of Waterloo; OpenIntro: Statistics

			• When describing the distribution of a numerical variable, mention its shape, center, and spread, as well as any unusual observations. • Identify the shape of a distribution as symmetric, right skewed, or left skewed, and unimodal, bimodal, multimodal, or uniform. • Use histograms and box plots to visualize the shape, center, and spread of numerical distributions, and intensity maps for visualizing the spatial distribution of the data. • Define a robust statistic (e.g. median, IQR) as measures that are not heavily affected by skewness and extreme outliers, and determine when they are more appropriate measured of center and spread compared to other similar statistics. • Recognize when transformations (e.g. log) can make the distribution of data more symmetric, and hence easier to model.		
6-8	4	Using Variability to Understand Probability	• Define trial, outcome, and sample space. • Explain why the long-run relative frequency of repeated independent events settle down to the true probability as the number of trials increases, i.e. why the law of large numbers holds • Distinguish disjoint (also called mutually exclusive) and independent events. • Draw Venn diagrams representing events and their probabilities. • Define a probability distribution as a list of the possible outcomes with corresponding probabilities that satisfies the three rules. • Define complementary outcomes as mutually exclusive outcomes of the same random process whose probabilities add up to 1.	Knowledge Comprehension Application	Q&A discussion Reading Videos Quiz Midterm
9-10	5	Foundations for Inference	• Define sample statistic as a point estimate for a population parameter. • Recognize that point estimates (such as the sample proportion) will vary from	Knowledge Comprehension Application	Q&A discussion Reading Videos

			one sample to another and define this variability as sampling variation. • Calculate the sampling variability of the proportion, (the standard error). • Recognize that when the sample size increases we would expect the sampling variability to decrease. • Notice that sampling distributions of point estimates coming from samples that don't meet the required conditions about sample size and independence will not be normal. • Define a confidence interval as the plausible range of values for a population parameter. • Calculate an approximate 95% confidence interval. • Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.		Quiz Assignment
11	6	Inference for Categorical Data	• Define population proportion (parameter) and sample proportion (point estimate). • Note that if the CLT doesn't apply and the sample proportion is low (close to 0) the sampling distribution will likely be right skewed, if the sample proportion is high (close to 1) the sampling distribution will likely be left skewed. • Note that the standard error calculation for the confidence interval and the hypothesis test are different when dealing with proportions, since in the hypothesis test we need to assume that the null hypothesis is true. • Calculate the required minimum sample size for a given margin of error at a given confidence level. • Note that the calculation of the standard error of the distribution of the difference in two independent sample proportions is different for a confidence interval and a hypothesis test. • Use a chi-square test of goodness of fit to evaluate if the distribution of levels of	Knowledge Comprehension Application Analysis	Q&A discussion Reading Videos Quiz

			a single categorical variable follows a hypothesized distribution.		
12	7	Inference for Numerical Data	• Use the t-distribution for inference on a single mean, difference of paired (dependent) means, and difference of independent means. • Explain why the t-distribution helps make up for the additional variability introduced by using s (sample standard deviation) in calculation of the standard error, in place of σ (population standard deviation). • Describe how the t-distribution is different from the normal distribution, and what 'heavy tail' means in this context. • Note that the t-distribution has a single parameter, degrees of freedom, and as the degrees of freedom increases this distribution approaches the normal distribution. • Describe how to obtain a pvalue for a t-test and a critical tscore for a confidence interval. • Define observations as paired if each observation in one dataset has a special correspondence or connection with exactly one observation in the other data set. • Recognize that a good interpretation of a confidence interval for the difference between two parameters includes a comparative statement (mentioning which group has the larger parameter). • Recognize that a confidence interval for the difference between two parameters that doesn't include 0 is in agreement with a hypothesis test where the null hypothesis that sets the two parameters equal to each other is rejected.	Knowledge Comprehension Application Analysis	Q&A Discussion Reading Videos Final Exam