**Probability and Statistics**

**Unit 1: Exploratory Data Analysis**

**Module: Examining Distributions**

* Summarize and describe the distribution of a categorical variable in context.
* Generate and interpret several different graphical displays of the distribution of

a quantitative variable (histogram, stemplot, boxplot).

* Summarize and describe the distribution of a quantitative variable in context: a)

describe the overall pattern, b) describe striking deviations from the pattern.

* Relate measures of center and spread to the shape of the distribution and choose

the appropriate measures in different contexts

* Compare and contrast distributions (of quantitative data) from two or more

groups, and produce a brief summary, interpreting your findings in context.

* Apply the standard deviation rule to the special case of distributions having the

"normal" shape.

**Module: Examining Relationships**

* Classify a data analysis situation (involving two variables) according to the "role

type classification", and state the appropriate display and/or numerical

measures that should be used in order to summarize the data.

* Compare and contrast distributions (of quantitative data) from two or more

groups, and produce a brief summary, interpreting your findings in context.

* Produce a two-way table, and interpret the information stored in it about the

association between two cat. variables by comparing conditional percents.

* Graphically display the relationship between two quantitative variables and

describe: a) the overall pattern, b) striking deviations from the pattern.

* Interpret the value of the correlation coefficient, and be aware of its limitations

as a numerical measure of the association between two quantitative variables.

* In the special case of linear relationship, use the least squares regression line as

a summary of the overall pattern and use it to make predictions.

* Recognize the distinction between association and causation, and identify

potential lurking variables for explaining an observed relationship.

* Recognize and explain the phenomenon of Simpson's Paradox as it relates to

interpreting the relationship between two variables.

**Unit 2: Producing Data**

**Module: Sampling**

* Identify the sampling method used in a study and discuss its implication and

potential limitations.

* Critically evaluate the reliability and validity of results published in mainstream

media.

**Module: Designing Studies**

* Identify the design of a study (controlled experiment vs. observational study)

and other features of the study design (randomized, blind etc).

* Explain how the study design impacts the type of conclusions that can be drawn.
* Determine how features of a survey impact the collected data and its accuracy.
* Critically evaluate the reliability and validity of results published in mainstream media.

**Unit 3: Probability**

**Module: Introduction to Probability**

* Relate the probability of an event to the likelihood of this event occurring.
* Explain how relative frequency can be used to estimate the probability of an event.

**Module: Finding Probability of Events**

* Determine the sample space of a given random experiment.
* Find probability of events in the case when all outcomes are equally likely.
* Apply probability rules in order to find the likelihood of an event
* When appropriate, use tools such as Venn diagrams or probability tables as aids

for finding probabilities.

**Module: Conditional Probability and Independence**

* Explain the reasoning behind conditional probability, and how this reasoning is

expressed by the definition of conditional probability.

* Find conditional probabilities and interpret them.
* Determine whether two events are independent or not.
* Use the general multiplication rule to find the probability that two events occur

(P(A and B)).

* Use probability trees as a tool for finding probabilities.

**Module: Random Variables**

* Distinguish between discrete and continuous random variables.
* Find the probability distribution of discrete random variables, and use it to find

the probability of events of interest.

* Find the mean and variance of a discrete random variable, and apply these

concepts for solving real world problems.

* Apply rules of means and variances for finding the mean and variance of a linear

transformation of a random variable, as well as for the sum of two independent random variables.

* Fit the binomial model when appropriate, and use it to perform simple calculations.
* Explain how a density function is used to find probabilities involving continuous random variables.
* Find probabilities associated with the normal distribution.
* Use the normal distribution as an approximation to the binomial distribution, when appropriate.

**Module: Sampling Distributions**

* Identify and distinguish between a parameter and a statistic.
* Explain the concepts of sampling variability and sampling distribution.
* Apply the sampling distribution of the sample proportion (when appropriate).

In particular, be able to identify unusual samples from a given population.

* Apply the sampling distribution of the sample mean as summarized by the

Central Limit Theorem (when appropriate). In particular, be able to identify

unusual samples from a given population.

**Unit 4: Inference**

**(Inference for one variable)**

**Module: Estimation**

* Determine point estimates in simple cases, and make the connection between the sampling

distribution of a statistic, and its properties as a point estimator.

* Explain what a confidence interval represents and determine how changes in sample size and confidence level affect the precision of the confidence interval.
* Find confidence intervals for the population mean and the population proportion (when certain conditions are met), and perform sample size calculations.

**Module: Hypothesis Testing**

* Explain the logic behind and the process of hypotheses testing. In particular, explain what the

p-value is and how it is used to draw conclusions.

* In a given context, specify the null and alternative hypotheses for the population

proportion and mean

* Carry out hypotheses testing for the population proportion and mean (when appropriate),

and draw conclusions in context.

* Apply the concepts of: sample size, statistical significance vs. practical importance, and the

relationship between hypotheses testing and confidence intervals.

* Determine the likelihood of making type I and type II errors, and explain how to reduce them,

in context.

**Module: Inference for Relationships**

* Identify and distinguish among cases where independent samples, matched

pairs, and anova are appropriate.

* In a given context, carry out the inferential method for comparing groups and

draw the appropriate conclusions.

* Specify the null and alternative hypotheses for comparing groups.

**Module: Inference for Relationships Continued**

* Choose the appropriate inferential method for examining the relationship

between two variables and justify their choice.

* In a given context, carry out the appropriate inferential method for comparing

relationships and draw the appropriate conclusions.

* Specify the null and alternative hypotheses for comparing relationships.