CIEE Global Institute - Yucatan

Course name: Statistics
Course number: (GI) MATH 2002 MEME
Programs offering course: Yucatan Open Campus Block: STEM and Society
Open Campus track: STEM and Society
Language of instruction: English
U.S. semester credits: 3
Contact hours: 45
Term: Spring 2020

Course Description

This course is a one-semester introduction to statistics course that follows the content and sequence required for STEM majors, focusing on real world applications of statistics and their impact on culture and society. It requires some knowledge of intermediate algebra. Students will work with data, including its summary and appropriate presentation using Excel and other widely available resources. They will perform comparative statistical tests in a variety of ways as well: by hand, using online calculators, freeware statistical packages, including R, Excel and other statistical software. Throughout, students will explore how statistics inform society.

Learning Objectives

Upon completion, students taking this course will be able to:

- Determine if data are qualitative or quantitative, whether they are discrete (nominal and ordinal) or continuous, and whether they are parametric or nonparametric.

- Apply basic criteria for sound experimental design and use of statistics.

- Describe groups of observations, including frequency distributions, central tendency and variability, for both parametric and nonparametric data.

- Manage large data sets in Excel or similar spreadsheets, including pivot tables, graphing capabilities and basic statistical functions.

- Present common summary statistics as graphs and tables using Excel and other software as they would appear in STEM peer-reviewed professional journals.

- Use concepts of the probability to define statistical significance.

- Evaluate statistical differences between groups using Goodness of Fit (Chi Square), t-tests, Analysis of Variance and Mixed Models, as well as nonparametric equivalents of parametric tests.

- Explore statistical relationships between sets of variables using correlation, regression

and multiple regression analyses.

- Choose the appropriate statistical test given data type and experimental design.

- Mathematically transform data to construct linear or parametric models.

- Apply statistics to everyday phenomena.

- Articulate intuitively how statistics works and its importance to local and global society.

Course Prerequisites

None

Methods of Instruction

Students will attend lectures, problem-solving workshops, discussions and excursions. Lectures will emphasize theory and applications. Considerable time will be spent solving problems individually and in groups with instructor oversight. In addition, students will be given data sets to manipulate, present and compare. They will discuss the relevance of statistics to a host of real-life situations involving physics, engineering, architecture, conservation, public health and human physiology. Excursions will explore the use of statistics in real life, including in the local culture.

Assessment and Grading

Participation	20%
Weekly Quizzes (5)	30%
Problem Sets	20%
Lecture Activities	20%
Statistics and Culture Essays (2)	10%
Total	100%

Course Requirements

Weekly Quizzes

Each week, students will take a quiz on the previous week's course material, including lectures, activities and readings. Quizzes will have true/false, multiple choice, calculations, filling in blanks and short answer questions. Quizzes will cover only new material from that week but will build on previous concepts.

Problem Sets

Each session will include a set of problems for students to solve individually and in groups. Certain of these will be handed in a graded. Grading will include both the ultimate solution, the student's reasoning in solving the problem, and the student's ability to clearly and intuitively explain the problem and its solution.

Lecture Activities

After each lecture, students will have a series of tasks and demonstrations related to the lecture material. They will work in groups to complete the tasks, handing in answers to a series of questions before leaving the class.

Statistics and Culture Essays

Students will use statistics to explore facets of culture. Students will write two 500-word essays: one on a cultural feature of the host culture and another comparing two cultures. In each case, statistical methods will be used in the analysis and to make major points.

Participation

Participation is valued as meaningful contribution in the digital and tangible classroom, utilizing the resources and materials presented to students as part of the course. Meaningful contribution requires students to be prepared in advance of each class session and to have regular attendance. Students must clearly demonstrate they have engaged with the materials as directed, for example, through classroom discussions, online discussion boards, peer-to- peer feedback (after presentations), interaction with guest speakers, and attentiveness on co-curricular and outside-of-classroom activities.

Class Attendance

Regular class attendance is required throughout the program, and all unexcused absences will result in a lower participation grade for any affected CIEE course. Due to the intensive schedules for Open Campus programs, unexcused absences that constitute more than 10% of the total course will result in a written warning.

Students who transfer from one CIEE class to another during the add/drop period will not be considered absent from the first session(s) of their new class, provided they were marked present for the first session(s) of their original class. Otherwise, the absence(s) from the original class carry over to the new class and count against the grade in that class.

For CIEE classes, excessively tardy (over 15 minutes late) students must be marked absent. Attendance policies also apply to any required co-curricular class excursion or event, as well as to Internship, Service Learning, or required field placement. Students who miss class for personal travel, including unforeseen delays that arise as a result of personal travel, will be marked as absent and unexcused. No make-up or re-sit opportunity will be provided.

Attendance policies also apply to any required class excursion, with the exception that some class excursions cannot accommodate any tardiness, and students risk being marked as absent if they fail to be present at the appointed time.

Unexcused absences will lead to the following penalties:

Percentage of Total Course Hours Missed	Equivalent Number of Open Campus Semester classes	Minimum Penalty
Up to 10%	1 content classes, or up to 2 language classes	Participation graded as per class requirements
10 – 20%	2 content classes, or 3-4 language classes	Participation graded as per class requirements; **written warning**
More than 20%	3 content classes, or 5 language classes	Automatic **course failure**, and possible expulsion

Weekly Schedule

NOTE: this schedule is subject to change at the discretion of the instructor to take advantage of current experiential learning opportunities.

Week 1   Sampling and Describing Data

Session 1.1   Definitions, Sampling, Experimental Design and Descriptive Statistics

Students will use key statistical terms appropriately, recognizing and differentiating important statistical concepts concerning the collection, analysis, interpretation and presentation of data. Students will differentiate groups of observations into qualitative vs. quantitative, and discrete vs. continuous. Students will place different data sets into four levels of measurement: nominal, ordinal, interval and ratio. Next, they will group data by observation frequency, relative frequency and cumulative relative frequency. Students will use concepts of explanatory vs. response variables and experimental units to explore solid experimental design. Students will discuss experimental ethics, including the importance of Institutional Review Boards when dealing with human subjects. Students will apply these concepts to a study they construct, going into the city center to gather observational data to demonstrate systematic sampling, construction of frequency distributions and interpretation of results. Students will work in groups to generate original discrete data, solve related problems, and put lecture concepts into practice. Students will define basic descriptive statistics and use software in Excel and R to generate them. Students will work with provided and original data sets to display data graphically, including stemplots, line graphs, histograms, scatterplots, box plots and others. They will also interpret graphs of varying formats. In addition, students will present data in tables.

Readings: Chapter 1 Sampling and Data , Chapter 2 Descriptive Statistics, and assigned section of Richardson, E., 2016. *Applying Statistics and Probability to Everyday Life*. Cavendish Square Publishing, LLC., Previte, J. 2018. The 2018 Digest of the Most Valuable Company Culture Statistics. Bluleadz at

Watch: SAS Software. 2014. How Statistics Touch our Everyday Lives. https://www.youtube.com/watch?v=yI1NH_XZ2_g

Due: Problem Set 1

Week 2   Probability and Discrete Random Variables

Quiz 1

Session 2.1   Probability Topics.

Students will use probability terminology in practice, including defining what is an experiment, outcome, sample space, events and relative frequency. They will calculate probability of an event A when all outcomes in the sample space are equally likely. They will learn the law of large numbers as well as biased outcomes and conditional probability. Students will consider independent and mutually exclusive events and the impacts of sampling with and without replacement. They will differentiate and calculate probabilities using Addition and Multiplication Rules. Students will also construct and interpret Contingency Tables, Venn Diagrams and Tree Diagrams. Students will collect original data for Contingency Tables, as well as use data provided to calculate probabilities in different contexts. Students will use R to explore these distributions further.

Reading: Chapter 3 Probability Topics and assigned problems

Watch: How to Predict the Odds of Everything. 2014. SciShow. https://www.youtube.com/watch?v=jFzahAm3qFQ

Due: Problem Set 2

Session 2.2   Discrete Random Variables.

Students will use random variable notation. They will recognize and understand the two characteristics of discrete probability distribution functions. They will calculate and interpret expected values or long-term average. Students will recognize the three characteristics of a binomial experiment, the resulting probability distribution, associated notation and applications, including Bernoulli trials. They will also understand the underlying principles and conditions of the Poisson distribution and apply it appropriately. Students will also consider geometric and hypergeometric probability distributions and applications. Students will then apply discrete distributions to playing cards, dice and other systems.

Readings: Chapter 4 Discrete Random Variables and assigned problems.

Due: Problem Set 3

Week 3   Continuous Random Variables, the Normal Distribution and the Central Limit Theorem

Session 3.1   Continuous Random Variables

Students define properties of continuous variables and continuous probability distributions. They will use cumulative distribution functions to calculate area under the curve for uniform, exponential and normal distributions. They will explore general characteristics of uniform, exponential and normal distributions, including the memoryless property of exponential distributions and its importance. Students will generate a data set using a random number generator and compare the resulting distribution with other distribution types. Students will work in groups to practice recognizing and working with different frequency distributions.

Quiz 2

Readings: Chapter 5 Continuous Random Variables and assigned problems.

Due: Problem Set 4

Session 3.2   The Normal Distribution

Students delve more deeply into the unique characteristics and properties of the normal distribution. They will use measures of student height in class to generate a frequency distribution and compare it to an idealized normal distribution. They will calculate common descriptive statistics for their data. They will define and calculate z scores for this and other normal distribution, relating z-score to standard deviation. Students will explore the Empirical Rule and its properties using the normal distribution. They will use the normal distribution to calculate probabilities. Students will work in groups to generate other normal curves using continuous data they gather. Students will discuss which continuous variables are likely to be normally-distributed and why.

Readings: Chapter 6 The Normal Distribution and assigned problems.

Watch: Crash Course Statistics. 2018. Henrietta Lacks, the Tuskegee Experiment, & Ethical Data Collection, Crash Course Statistics #12. https://www.youtube.com/watch?v=CzNANZnoiRs

Due: Problem Set 5

Session 3.3   The Central Limit Theorem.

In this session, students will explore central tendency in normal distributions, including the mean and the Central Limit Theorem. They will use sample data to

calculate minimum sample size. They will apply the Central Limit Theorem to distributions of means from samples of a large population and how this leads to the standard error of the mean. Likewise, students will use the Central Limit Theorem for Sums. Students will work in groups, using the Central Limit Theorem in a variety of applied contexts.

Readings: Chapter 7 The Central Limit Theorem and assigned problems.

Due: Problem Set 6, Statistics and Culture Essay 1

Week 4   Confidence Intervals and Hypothesis Testing

Quiz 3

Session 4.1   Confidence Intervals

Students will define, calculate and interpret confidence intervals for estimating a population mean and a population proportion, calculating the Error Bound Mean (EBM) as well as finding the z-score for a stated confidence level. Students will calculate sample size for a desired margin of error. They go on to explore the Student's t distribution and consider how its probability distribution changes with sample size. They interpret when to apply the Student's t distribution vs. the normal distribution. Students use the concept of population proportion p and sample size n to different real life scenarios.

Readings: Chapter 8 Confidence Intervals and assigned problems.

Due: Problem Set 7

Session 4.2   Visit to Statistics Institute.

During this visit, students will speak with professional statisticians about current projects connecting Statistics to society. Students will have a tour, speak with several statisticians and discuss how statistics, society and culture interact. Students will then use online resources to further explore demographic data between two cultures to explain similarities and differences between them.

Readings: Jasilionis, D. and Shkolnikov, V.M., 2016. Longevity and education: a demographic perspective. *Gerontology*, *62*(3), pp.253-262.

Watch: Milgram, A. 2014. Why Smart Statistics are the Key to Fighting Crime
https://www.youtube.com/watch?v=ZJNESMhIxQ0

Due: Problem Set 8

Session 4.3   Hypothesis Testing with One Variable.

Students will use formal logic to construct pairs of mutually exclusive hypotheses

for a variety of observations, as null and alternative hypothese. They will consider how to isolate a single variable, the importance of repeat observations to test pairs of hypotheses, how avoid pseudo-replication, best practices for isolating variables, and appropriate controls when testing hypotheses. Students will consider possible errors in hypothesis testing, as Type I and Type II errors. They will then work in small groups with data sets, both those provided and created by students, to conduct and interpret hypothesis tests for a single population. They will do this when the standard deviation is known, when standard deviation is unknown and for single population proportion. Along the way, students will define and calculate the actual probability of getting the test result, called the p-value.

Reading: Chapter 9 Hypothesis Testing with One Variable and assigned problems.

Due: Essay 2 from excursion to Statistics Institute: Statistical comparison of two cultures.

Week 5   Hypothesis Testing and Comparative Statistics

Quiz 4

Session 5.1   Hypothesis Testing with Two Samples

Students extend hypothesis testing to two sample experiments or comparisons. They will test if two samples are drawn from the same population. They consider the importance of underlying distribution with hypothesis testing when sample sizes are small or large. Students will use the Aspin-Welch t-test to compare two samples with unknown standard deviations. They will use the t-test to compare two samples with known standard deviations. Students will also compare two independent population proportions and consider the conditions necessary to conduct this comparison. They move onto paired samples, where two measurements are drawn from the same subject. Students will generate data, recognize the proper two sample test to use in a variety of situations, compare samples statistically with software and present their results graphically.

Readings: Chapters 10 Hypothesis Testing with Two Variables and assigned problems.

Due: Problem Set 9

Session 5.2   The Chi-Square Distribution

Students will explore underlying conditions of the Chi-Square Distribution and be able to describe it intuitively and mathematically. They will list the most common applications of the Chi-Distribution in statistics, including Goodness of Fit, the Test of Independence, Test of Homogeneity, and the Test of a Single Variance. Students will learn alternatives to the Chi-Square test when certain necessary

conditions are unmet. Students will generate data, recognize when to use the Chi-Square test in a variety of situations, compare samples statistically by hand, using online calculators and with software, presenting their results graphically.

Readings: Chapters 11 The Chi-Square Distribution and assigned problems

Beatty, B. 2016. The Bell Curve, in Plucker, J.A. and A. Esping (eds) Human Intelligence: Historical influences, current controversies, teaching resources. Retrieved [7 January 2019] from https://www.intelltheory.com/bellcurve.shtml#part1

Due: Problem Set 10

Session 5.3   Correlation and Regression

Students define components of the equation for a linear relationship between pairs of variables, identifying independent and dependent variables when present and how these are related to cause and effect. They will use scatterplots of linear or near-linear relationships and will use software to determine the best straight line, the linear regression equation associated, and interpret the least square criteria for best fit. They will describe the slope, intercept, the Regression Coefficient and the Coefficient of Determination. They compare statistical tests and their results for regressions and correlations. Students use software to test the significance of the Correlation Coefficient. Students will use linear regression to predict future outcomes and how to work with outliers. They will then apply correlation and regression analyses to a variety of data sets, including using a correlation matrix for many pairwise comparisons. Finally, students will use mathematical transformations of non-linear data to allow their use with linear tests.

Readings: Chapters 12 Linear Regression and Correlation, and assigned problems

Due: Problem Set 11

Week 6   F-test, Analysis of Variance and Non-Parametric Tests

Quiz 5

Session 6.1   F-Distribution and the One-Way Analysis of Variance

Students will learn tests comparing more than two samples. They begin by considering the conditions necessary to use a One Way Analysis of Variance to compare means of more than two groups of data. Students describe the F distribution and how it is used to test variance between samples by comparing it to variance within samples. To do this, they will define and calculate the sum of squares, mean square, the F-ratio, between and within sources of variation and

use proper notation to report these values. They will then use data and software to compare multiple means and multiple variances using the One Way ANOVA and the F-test. Students will then use post hoc or multiple comparisons tests to tease apart which pairs of groups statistically differ from one another.

Readings: Chapter 13 F-Distribution and One-Way ANOVA

Due: Problem Set 12

Session 6.2   Non-Parametric Statistics

Students will define parametric and non-parametric statistics, based on whether samples follow a normal distribution. They will consider what to do with small sample sizes, where directly testing if data are normally distributed is not possible. They will learn differences in statistical power between parametric and non-parametric statistics and discuss why non-parametric statistics are not always preferred. Students will then learn non-parametric analogs to the t-test, paired t-test, regression/correlation and ANOVA. Students will work in groups with different data sets and statistical software to test data for normality. They will compare statistical outcomes for specific comparisons using both parametric and non-parametric statistics.

Reading:

Watch:

Due: Essay 2 Contemporary Applications of Calculus to Mayan Life

Session 6.3   Multiple Regressions, 2 and 3 Way ANOVA, Next Steps

Students will extend their knowledge of statistics to more complex comparisons, including linear regressions with more than one independent variable and tests for means comparing multiple groups for more than one factor. Students will also discuss limitations of statistics, the difference between statistical significance and what is likely important in real life situations, how statistics can be misused and the ethics of statistical comparisons of human subpopulations.

Readings: and assigned problems.

Watch: Liddell, M. 2016. How Statistics Can Be Misleading
https://www.youtube.com/watch?v=sxYrzzy3cq8

Due: Final Problem Set, Final Quiz

Course Materials

Course Textbook

Illowski, B. and S. Dean (eds). 2017. Introductory Statistics. Houston, TX: Openstax CNX

Readings

Beatty, B. 2016. The Bell Curve, in Plucker, J.A. and A. Esping (eds) Human Intelligence: Historical influences, current controversies, teaching resources. Retrieved [7 January 2019] from https://www.intelltheory.com/bellcurve.shtml#part1

Jasilionis, D. and Shkolnikov, V.M., 2016. Longevity and education: a demographic perspective. *Gerontology*, *62*(3), pp.253-262.

Previte, J. 2018. The 2018 Digest of the Most Valuable Company Culture Statistics. Bluleadz at *https://www.bluleadz.com/blog/annual-digest-of-company-culture-statistics*

Richardson, E., 2016. *Applying Statistics and Probability to Everyday Life*. Cavendish Square Publishing, LLC.

Online Resources

Crash Course Statistics. 2018. Henrietta Lacks, the Tuskegee Experiment, & Ethical Data Collection, Crash Course Statistics #12. https://www.youtube.com/watch?v=CzNANZnoiRs

How to Predict the Odds of Everything. 2014. SciShow. https://www.youtube.com/watch?v=jFzahAm3qFQ

Liddell, M. 2016. How Statistics Can Be Misleading https://www.youtube.com/watch?v=sxYrzzy3cq8

Milgram, A. 2014. Why Smart Statistics are the Key to Fighting Crime https://www.youtube.com/watch?v=ZJNESMhIxQ0

SAS Software. 2014. How Statistics Touch our Everyday Lives. https://www.youtube.com/watch?v=yI1NH_XZ2_g