EXAMINING THE STATISTICAL KNOWLEDGE OF PROSPECTIVE AND PRACTICING SECONDARY MATHEMATICS TEACHERS GRADUATED FROM UPRM

By

Kevin Molina Serrano

A project submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In

MATHEMATICS STATISTICS

UNIVERSITY OF PUERTO RICO

MAYAGUEZ CAMPUS

2018

Approved by:

Marggie González Toledo, Ph.D. Date
President, Graduate Committee

Dámaris Santana Morant**,** Ph.D. Date
Member, Graduate Committee

Edgardo Lorenzo González, Ph.D Date
Member, Graduate Committee

Victor Ocasio González, Ph.D. Date
Interim Chairperson of the Department

José R. Ferrer López, Ed.D. Date
Representative of Graduate School

Abstract of Project to Graduate School
of the Univeristy of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master in Science

EXAMINING THE STATISTICAL KNOWLEDGE OF PROSPECTIVE AND PRACTICING SECONDARY MATHEMATICS TEACHERS GRADUATED FROM UPRM

By: Kevin Molina Serrano

Chair: Dra. Marggie González Toledo
Department: Mathematical Sciences

Research have shown that prospective and practicing mathematics teachers lack of statistical knowledge and as result are not well prepared to teach statistics in K-12. In this study we evaluated the statistical knowledge of prospective and practicing mathematics teachers, all from the mathematics education undergraduate program at University of Puerto Rico at Mayaguez Campus years using the CAOS 4 test instrument. In addition, we provided feedbacks and recommendations to the program to improve the effectiveness of preparing teachers to teach statistics at secondary level. We used a binomial regression to model the proportion of correct answers and to identify which factors were significant. Results demonstrated subjects lack the statistical knowledge required to teach statistics, with a sample CAOS average score of 0.41. Finally, the majority agreed the program did not prepared them to teach statistics and that ESMA 3016 course is mainly focused on programming.

Resumen de Proyecto Presentado a Estudios Graduados
de la Universidad de Puerto Rico como Requisito Parcial de los
Requerimientos para el Grado de Maestría en Ciencias

EXAMINING THE STATISTICAL KNOWLEDGE OF PROSPECTIVE AND PRACTICING SECONDARY MATHEMATICS TEACHERS GRADUATED FROM UPRM

Por: Kevin Molina Serrano

Consejera: Dra. Marggie González Toledo
Departamento: Ciencias Matemáticas

Estudios investigativos han demonstrado que futuros maestros y maestros en servicio carecen de conocimiento estadístico y como consecuencia, no están bien preparados para enseñar estadística a nivel K-12. En este estudio se evaluó el conocimiento estadístico requerido para enseñar estadística a nivel secundaria de futuros maestros y maestros en servicio que se graduaron del programa de bachillerato de educación matemática de la Universidad de Puerto Rico Recinto Mayagüez usando el instrumento CAOS 4. Además, se le proveyó recomendaciones al programa para mejorar la preparación de maestros para enseñar estadísticas. Se utilizó una regresión binomial para modelar la proporción de respuestas correctas en el CAOS 4 para identificar factores significativos. Los resultados demuestran un pobre conocimiento estadísticos en los sujetos con una puntuación promedio muestral de 0.41. Además, la mayoría concordaron que el programa no los prepara bien para enseñar estadística y que el curso ESMA 3016 está mayormente enfocado en la programación.

Copyright © 2018

por

Kevin Lee Molina Serrano

Acknowledgement

I want to extend my gratitudes to the following people because each of them played a significant role in my academic preparation and personal life:

To God because he is the center of my life and everything I achieved is thanks to Him.

To my family, girlfriend and friends for their support, unconditional love and for being there for me in the good and bad times.

To my mentor, Dra. Marggie González Toledo for her coaching, patience and being a great mentor. You are a blessing from the sky and thanks in large part to you, I will continue the doctoral degree in mathematics education.

To the members of my committee, Dr. Edgardo Lorenzo and Dra. Dámaris Santana Morant for their support and coaching in statistics.

To Dr. Pedro Torres Saavedra for his availability and his coaching in statistics even though he was not part of my graduate committee.

And lastly to Dra. Mayra Mendez from Industrial Engineering because she is like an academic mom for me and helped me a lot in my undergraduate program.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Statistics and Probability was fully incorporated into the mathematics curriculum of USA in recent years. Its progression has been a journey from relative insignificance, reserved for the most able students in high school, to prominence as a fundamental component recommended for all students at all grade levels (Jones & Tarr, 2010). Figure 1 shows how the incorporation of Statistics and Probability has progressed through the years into the mathematics K-12 school curriculum in the USA (Jones & Tarr, 2010).

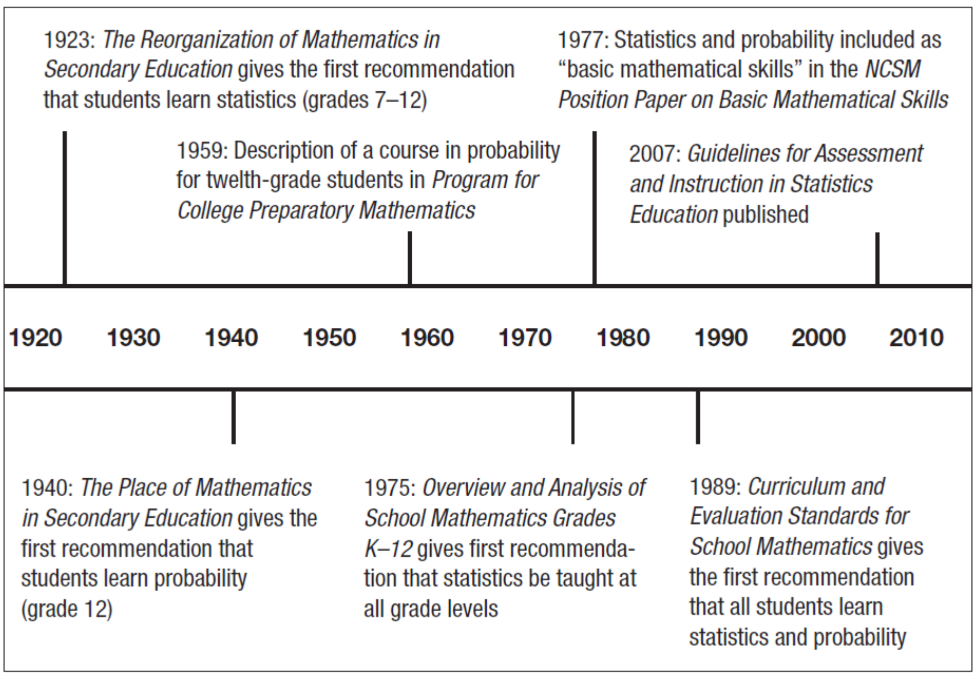


Figure 1: A timeline of shifting emphases on statistics and probability in K-12 school curriculum (Jones & Tarr, 2010).

Despite Statistics and Probability being relatively new to the mathematics curriculum, it is not the reason why students are not receiving statistics education in schools. Rather, maybe this is happening because Statistics is often taught by mathematics teachers that have not been properly trained to teach statistics. As a consequence, they do not feel confident nor prepare to teach statistics (Burrill & Biehler, 2011; Gattuso, 2008; Makar & Fielding-Wells, 2011; Franklin et al., 2007).

The *K-12 Guidelines for Assessment and Instruction in Statistics Education* report encourages teachers to teach statistics emphasizing the understanding of the concepts of the statistical ideas rather than focusing on computation and procedures (Franklin et al., 2007). Moreover, it is mentioned that a change of habits is necessary when teaching statistics (Burrill & Biehler, 2011). Adopting habits like using real data, build intuitions, begin with a graph, explore alternate representations of data, investigate and explore before introducing formulas and conduct projects helps students to develop conceptual and fundamental understanding in statistics (Burrill & Biehler, 2011). Nevertheless, when teachers teach statistics, it is often teach with a mechanical and computational approach focusing in the computational procedures of statistical measures (Makar & Confrey, 2004) and creating graphical representations (Sorto, 2006).

This situation arise a cause for concern. To begin with, we are living in a world where the importance of data is increasingly exponentially. In every industry, all kind of data is being collected. Not just to analyze and store it but to extract useful information and to be later translated into knowledge and that leads to better decisions making and finding solutions. Medical doctors benefits from Statistics to develop and prescribe better treatments to their patients. Social Networks benefits from Statistics by identifying topics that are most popular among their clients and providing more content from that topics. Sports benefits from Statistics to identify which athletes are better or how much a player will increase their revenues. Schools and Colleges are using Statistics to compare how well their students are performing against other schools students. These are some of the examples of how Statistics can be beneficial to society. Unfortunately, Statistics can also lead to false information or conclusions if it is not used with the proper statistical knowledge. A very common example is when a political poll is conducted and when the elections arrived, the results are the opposite of the polls results. There is a possibility the problem is not the computation of the statistical measures, rather it was a conceptual problem because all the data that it is gather for a poll is dictated by the sampling method used. If the sampling method is not correct, the data will be biased and the results will not be accurate with reality. This is one of many examples where conceptual understanding of the statistical ideas is equally important as the computational procedures.

The world is rapidly moving toward data analysis. Thus, it became imperative to have K-12 mathematics teachers with the best preparation, especially when it comes to teaching probability and statistics. Unfortunately, recent research (e.g.,Lovett et al. (2016)), have shown that prospective and practicing teachers struggles with understanding statistical concepts such as sampling distribution, variability, covariance, p values, confidence intervals, probability and conceptual understanding of measures of centers such as median, mean and mode. This leads to the foundation of this project.

As one of the main university in PR that prepares secondary mathematics teachers, it is neccesary to evaluate how is the program preparing our future teachers to teach such topics. The study aims to answer the following research questions:

1. To what extended is the curriculum at UPRM preparing future mathematics teachers to teach statistics and probability at the secondary level?

2. What recommendations can be offered to the Department of Mathematical Sciences in order to improve the mathematics education program to prepare teachers to teach statistics at the secondary level?

In the Literature Review, we will present the framework, the definition of statistical knowledge used in our study, studies evaluating the statistical knowledge of prospective and practicing mathematics teachers and key differences between teaching statistics and teaching mathematics. Following that chapter, we have the Methodology chapter in which an overview about the mathematics education program at UPRM and the statistics course ESMA 3016 is given. In addition, a description of the instrument that is used in this study for evaluating the statistical knoweldge will be provided. The Results chapter follows in which different quantitative and qualitative analysis will be conducted . After that, in the Discussion chapter a summary of the analysis conducted and answers to the research questions are provided. At the end, a conclusion will be given in addition with some limitations and future work for the study.

Chapter 2 Literature Review

In this chapter, we will begin by presenting the framework used in this study, followed by recent studies conducted on prospective and practicing mathematics teachers to evaluate their statistical knowledge. To conclude this chapter, key differences between teaching statistics and mathematics will be mentioned.

Framework

Statistical knowledge for teaching refers to the type of knowledge a teacher needs in order to teach statistics. It encompasses many different areas of knowledge for teaching. Groth (2013) expanded a framework for statistical knowledge for teaching by combining Groth (2007) and Hill, Ball, and Schilling (2008) statistical knowledge frameworks. This statistical knowledge is divided into two parts: Subject Matter Knowledge and Pedagogical Content Knowledge. The Subject Matter Knowledge is divided in three components: common content knowledge, specialized content knowledge and horizon knowledge. The Pedagogical Content Knowledge is divided in three components: knowledge of content and students, knowledge of content and teaching and curriculum knowledge. The way Groth (2013) defined these components is illustrated in Table 1.

Table 1: Statistical knowledge for teaching framework (Groth, 2013)

Subject Matter Knowledge	
Components	Definition
Common content knowledge	Knowledge that is used in the work of teaching in ways in common with how it is used in many other professions that are also use mathematics.
Specialized content knowledge	Mathematical knowledge that allows teachers to engage in particular teaching tasks; including how to accurately represent mathematical ideas and provide mathematical for common rules and procedures and examine and understand unusual solutions to problems.
Horizon knowledge	Entails knowing statistics beyond the prescribed curriculum
Pedagogical Content Knowledge	
Components	Definition
Knowledge of content and students	Content knowledge intertwined with knowledge of how students think about, know, or learn this particular content.
Knowledge of content and teaching	Provides teachers with content- specific teaching strategies.
Curriculum knowledge	Allows teachers to perform tasks such as appropriately sequencing the introduction of statistical ideas

Common content knowledge, according with Groth (2013), is based on accurately reading graphs, constructing survey questions, computing descriptive statistics and choosing appropriate descriptive statistics for a given context. The common content knowledge that is required to teachers to teach statistics is illustrated in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) K-12 report (Franklin et al., 2007). For the current study, we will use the four phases on the Statistics Investigation Cycle Wild and Phannkuch (1999) proposed a Statistics Investigation Cycle later used in the GAISE framework (Franklin et al., 2007).

The statistical investigative cycle has four phases: (1) formulate questions, (2) collect data, (3) analyze data and (4) interpret results. Table 2 provides a description of each phase of the statistical investigative cycle described in Franklin et al. (2007). Thus, when referring to statistical knowledge, it means the knowledge in statistics in each one of the four phases of the Statistical Investigative Cycle and the Puerto Rico Core Standars for data analysis and probability topics.

Table 2: Description of the four phase of the statistical investigative cycle (Franklin et al., 2007)

Phases of Statistical Investigative Cycle	**Description**
Formulate Questions	Clarify the problem at hand and formulate one or more questions that can be answered with data.
Collect Data	Design a plan to collect appropriate data and employ the plan to collect data.
Analyze Data	Select appropriate graphics and numerical methods and use these methods to analyze the data.
Interpret Results	Interpret the analysis and relate the interpretation to the original question.

Studies have shown lack of prepardeness and conceptual statisical knowledge in the prospective and practicing mathematics teachers. In the following parts, reviews about studies of prospective and practicing mathematics teachers evaluating their statistical knowledge will be presented.

Research on Prospective Math Teachers and Stat Knowledge

Although there has been limited studies investigating the statistical knowledge of prospective secondary mathematics teachers (Batanero, Burrill, & Reading, 2011), there has been some studies that reflect the lack of statistical knowledge this population tend to have. Several studies have demonstrated the difficulties that prospective teachers have on statistical knowledge. Groth and Bergner (2006) conducted a study where 46 prospective teachers were ask to complete a written content knowledge assessment to evaluate their statistical conceptual and procedure thinking about the mean, mode and median and how they defined each one of them. Two instruments were used in the evaluation of their knowledge: Structure of the Observed Learning Outcome (SOLO) taxonomy and Profound Understanding of Fundamental Mathematics (PUFM). Results showed teachers had difficulties identifying the differences between median, mode and mean and identifying when to use each one of them. In fact, the statistical thinking exhibited by the prospective teachers in the study strongly resembles the thinking of elementary and middle school children. Browning, Goss, and Smith (2014), were interested in improving the statistical knowledge of measures of variablity such as mean absolute deviation and standard deviation of 13 prospective elementary teachers. The task consisted of using TinkerPlots' software to analyze two data sets with equally measure of center but differences in variability. The results showed that using ThinkerPlots helped prospective elementary teachers develop a better understanding of mean absolute deviation and standard deviation. Leavy (2010) evaluated the obstacles 26 prospective teachers faced when designing and teaching informal inference, and investigated the development of the content and pedagogical knowledge related to teaching informal statistical inference. Overall, participants demonstrated profiency reasoning about many of the elements fundamental to informal inferential although some of them showed some content knowledge difficulties related to descriptive statistics, the median and graph knowledge that were addressed during their effort to prepare Lesson Study. Where these teachers encountered most difficulties were in transforming their subject matter content knowledge to pedagogical content knowledge. In a different study, Leavy (2006) evaluated 23 prospective teachers conceptual understanding of distribution, expressed in the measures and representations used to compare distributions. The participants focused on summary statistics such as measures of central tendency, neglecting important concepts such as variability and exploring graphical displays of datasets. González and Pinto (2008) evaluated the conception and knowledge of four prospective teachers regarding the use of statistical graphs. Results showed limited knowledge of graphical representation, especially in stem and leaf graphs. Also a focused on procedures to contruct graphs rather than conceptual understanding of them.

The studies mentioned above are small-scale studies; however, the following three are of bigger scale. Lovett et al. (2016) conducted a study to evaluate the statistical knowledge of 217 prospective secondary mathematics teachers across 18 institutions using the Levels of Conceptual Understanding of Statistics (LOCUS) assessment. LOCUS is a collection of assessments items designed to measure conceptual understanding of statistics at the levels hypothesized by the GAISE framework (Jacobbe, Whitaker, Case, & Foti, 2014). The results revealed a proficient knowledge identifying appropriate measures of center, but difficulties with sampling distributions, variability, p values, and confidence intervals. Leavy (2006) conducted a study to gain insight into the understanding of the mean of 263 prospective elementary teachers. The results showed that 57% of them correctly utilized the mean to compare two data set, but only 25% of the participants showed conceptual understanding of the mean. Lastly, Hannigan, Gill, and Leavy (2013) investigate the conceptual understanding of statistics of 134 prospective secondary mathematics teachers. The instrument used in this study was the Comprehensive Assessment of Outcomes in Statistics (CAOS). It is focused on reasoning with statistical information with an increased emphasis on reasoning about variability and conceptual understanding in statistics. A mean of 50% in the CAOS test indicated a lack of statistical reasoning and conceptual undertanding in the prospective teachers, struggling on items related to data production, in particular randomization, sampling and populations, and extrapolating from a regression model. Similar studies(Fabrizio, López, and Plencovich (2011) and Tintle, Topliff, VanderStoep, Holmes, and Swanson (2012)) have also used CAOS test to evaluated the conceptual statistical knowledge of its participants. Both of the studies agree with Hannigan et al. (2013), obtaining low means scores in CAOS of 42% and 44% respectively.

Research on Practicing Mathematics Teachers and Statistical Knowledge

In terms of practicing teachers, there are fewer studies conducted to evaluate their statistical knowledge. In Wessels and Nieuwoudt (2011), an adapted SOLO framework was used to determine levels of statistical thinking in 90 teachers understanding sample and average. In the topic of sampling, the participants showed strong theoretical knowledge but inability to apply it to real life context. In addition, teachers showed more familiarity and higher confidence levels when working with the topic of average. Jacobbe and Horton (2010) conducted a study to analyze data displays comprehension of three practicing elementary mathematics teachers. Participants were evaluated in the following five categories: Reading Data, Computations, Comparisons, Trend and Selection and Construction of Data Displays. Results demonstrated proficiency in the first three categories, but struggled with identyfing trends and selection and construction of data displays. Even though there were only three participants in this study, these three teachers were recommended by their district supervisor as exemplary mathematics teachers.

Bansilal (2014) performed a study to evaluate the knowledge of the normal probability distribution of 290 teachers from a teacher development program. Responses to the task were analyzed using the Action, Process, Object, Schema (APOS) framework that specified a standardization and a probability layer of understanding. The standarization layer evaluates the transformation from a normal distribution variable to a standardized normal distribution variable. Meanwhile, the probability layer evaluates the association between the z score and the p value. Only 27% of the participants have successful rate in the standarizartion layer and only 14 % have successful rate in the probability layer. Casey and Wasserman (2015) explored the knowledge about the line of best fit of nineteen prospective and practicing mathematics teachers. The teachers showed a strong knowledge of calculating the line of best fit, but their conceptions and criteria were sometimes inaccurate. Kataoka, da Silva, and Cazorla (2014b) evaluated the reasoning on variation of 23 secondary mathematics teachers using the (SOLO) taxonomy. Participants demostrated difficulties with median and quartiles on dotplots but overall, the reasoning on variation improved using dotplot. Jacobbe (2012) evauated the knowledge of the mean and median of three elementary mathematics teachers. The teachers focused on procedures definitions and showed lack of conceptual knowledge between the mean and the median. Similar to Jacobbe (2012), Hobden (2014) evaluated the conceptual understanding of the median of 316 practicing mathematics teachers. Participants demostrated low levels of statistical literacy and consequently many of them fail to correctly interpret the median. In Kataoka, da Silva, and Cazorla (2014a) the understanding of covariance of 24 high school practicing teachers was examined. At first teachers struggled with the undertanding of covariance, but through the task their understanding improved. Lastly, the reasoning of variation from nine practicing mathematics teachers was evaluated in Borim and Coutinho (2008), reflecting inability to relate the deviation from the mean, the mean and the standard deviation.

As mentioned above, prospective and practicing mathematics teachers are having trouble demonstrating the statistical knowledge required to teach in elementary, middle and secondary school. Topics that they commonly struggle with are:

- conceptual understanding of measures of center such as mean, media and mode; variability; interpreting graphs as a whole
- distributions
- p values
- confidence intervals
- sampling

Differences between Teaching Mathematics and Teaching Statistics

Most of the statistical knowledge of prospective and practicing mathematics teachers is focused on procedures and calculations but lack of conceptual statistical knowledge. Statistics is often seen as a branch of mathematics. However, there are sufficent evidence that shows how teaching and learning statistics is different from teaching and learning mathematics. Franklin et al. (2007) in the K-12 Report claims there are two beliefs that distinguish statistics and mathematics. The first one is that in statistics you have to take into consideration variability in the data as opposed to the deterministic nature of mathematics. The second one is the role of context, where in statistics context provides meaning and in mathematics context provides the opportunity for application. In addition, key comparisons between statistics and mathematics are presented below (Rossman, Chance, Medina, & Obispo, 2006, p. 2-10).

1. *Crucial Role of Context*: Mathematics is an abstract field of study; it exists independently of context. But in statistics, the context can not be ignore when analyzing data.

2. *Issues of Measurement*: In mathematics, measurement includes getting students to learn about appropriate units to measure attributes of an object such as length, area, and volume

and to use formulas to measure those attributes. In statistics, drawing conclusions from data depends critically on taking valid measurements of the properties being studied.

3. *Importance of Data Collection*: When mathematicians examine data they typically focus on detecting and analyzing patterns in the data. How the data were collected is not relevant to purely mathematical analyses. However in statistics, data collection is crucial. The design of the data collection strategy determines the scope of conclusions that can be drawn.

4. *Lack of Definitive Conclusions*: Mathematics involves rigorous deductive reasoning, proving results that follow logically from axioms and definitions. The quality of a solution is determined by its correctness and succinctness, and there is often an irrefutable correct answer. In contrast, statistics involves inductive reasoning and uncertain conclusions. All of statistical inference requires one to use inductive reasoning, as informed inferences are made from observed results to defensible, but ultimately uncertain, conclusions.

5. *Communicating Statistical Knowledge*: Terminology is essential in mathematics as well as statistics, but one difference is that many common terms from everyday language have technical meanings in statistics. Examples include words such as bias, sample, statistic, accuracy, precision, confound, correlation, random, normal, confident, and significant.

Similar to Rossman et al. (2006), Burrill and Biehler (2011, p. 62-63), presented how the following seven topics are taught in mathematics and statistics:

1. *Data*: Data are typically used in mathematics classrooms in the context of visualisation of numbers and the study of functions, but the work rarely reaches the level of context-related reading between and beyond the data.

2. *Variation*: Variation has a different nature in the two disciplines. Mathematics is often taught in school as being exact and precise. Statistics is about noise, that is, how to measure and control variability. Real data in statistics are contextual, containing uncertainty and error

while data in many school mathematics classrooms are typically assumed to perfectly fit a mathematical model.

3. *Distribution*: Distributions are developed in the context of teaching statistics only and do not evoke specific tensions with concepts taught in mathematics.

4. *Representation*: Statistics and mathematics differ in approaches to representations of data in several ways including the following: most statisticians begin with a graph; many mathematics students and teachers crunch numbers without paying attention to a visual representation of the data; and while in statistics different graphs or representations are used to identify different aspects of the same data (transnumeration), graphs in mathematics are often used in showing the same relationship in different representations (tables, graphs, and symbols).

5. *Association and modelling relations between two variables*: Cartesian coordinate plots are typically used in mathematics classrooms only to draw graphs of functions and not as scatter plots for bivariate data.

6. *Probability models for data-generating processes*: teaching probability has to be enriched by broad phenomenological experiences, in which simulation can play a prominent role. However, this modelling applies only to those samples randomly drawn from a population or from a random allocation and assignment. This fundamental aspect is often neglected in a typical mathematical treatment of probability, which then undermines statistical understanding. This modelling will depend on assumptions, such as independence or equiprobability, which do not always hold, and are often taken as given and not to be considered or checked.

7. *Sampling and inference*: what is important for statistics is sample-to-sample variation and how this variation decreases as the sample size increases. The mathematical approach to proportional reasoning, however, often undermines the statistical approach for reasoning from samples. Percentages in mathematics are often applied in simple contexts, where the reference is set and the units are clear and constant. Careful statistical statements made

about margin of error and confidence intervals are replaced by simplistic inferences from sample to population, assuming a perfect proportional relationship. Ignoring uncertainty and variability, sample results are reported in point estimates rather than interval estimates in many media reports.

As literatures shows, there are key differences between teaching statistics and teaching mathematics which can lead to a lack of preparedness in mathematics teachers to teach statistics. This is why it becomes important to train future mathematics teachers not only to teach math concepts and ideas. It is necessary to provoke a change of mind so those teachers get prepared to teach statistics and probability.

In the following chapter, information about the mathematics education program at UPRM and ESMA 3016 statistics course will be presented.In addition, information about the instrument used to evaluate the statistical knowledge to teach statistics at secondary level of prospective and practicing mathematics teachers from the mathematics education program at University of Puerto Rico Mayaguez will be presented.

Chapter 3 Methodology

Since the purpose of this study is to evaluate the statistical knowledge of the prospective and practicing secondary mathematics teachers from the mathematics education program in the University of Puerto Rico Mayaguez Campus (UPRM), it is important to provide some background information about its mathematics education program.

Mathematics Education Program at UPRM

The Mathematics Education Program is one of the three bachelors degrees offered by the Mathematical Science Department at UPRM. Its focus is to train specialists in mathematics with a secondary mathematics teacher certification from the Puerto Rico Department of Education . It is a four-year program with one hundred and thirty-nine courses credits. Table 3 shows the distribution of the courses credits. The program provides a preparation in the following mathematics topics: calculus, geometry, lineal algebra, number theory, discrete mathematics and introduction to abstract algebra. In the last semester of the program, the students are required to teach mathematics in a school under the supevision of a school teacher and a university supervisor which is usually a professor from the Department of Mathematical Science. It is the only mathematics education program in PR inside a Mathematical Science Department. For complete details of the mathematics education curriculum at UPRM, see Appendix A.

Table 3: Mathematics Education Program Course Distibution

Type of Courses	Number of Credits
Mathematics courses	42
Statistics courses	3
Education courses	21
Teaching practice course	6
Others	67

As you can see in Table 3, a single statistics course is incorporated into the curriculum. Details of this statistics course will be provided in the following section.

Statistical Analysis of Data (ESMA 3016) Course

The name of the statistics course is Statistical Analysis of Data and is a three-credit course which is included in the curriculum of the mathematics education program and the computer science program. The course is focused on descriptive and inferential statistics and the exploratory analysis of data. Statistics topics discussed in this course includes: sampling, source of error, quantitative and qualitative variables, descriptive statistics, probability (adding and multiplication rule, independence, conditional, mutually exclusive and empiric probability), Binomial and Poisson distribution, discrete and continuous random variable, sampling distribution, central limit theorem and inferential statistics. In addition, a fifty-minute laboratory section is incorporated in the course to perform statistical analysis of data using the R statistical software. For a complete information about the courses syllabus, see Appendix B. This statistics course meets the required statistical topics establish in Franklin et al. (2007) and the Puerto Rico Core Standards (Department of Education of Puerto Rico, 2014) for the preparation of mathematics teachers to teach statistics at secondary level. In the following section, we will introduce and described the instrument used in this study .

CAOS Test

The instrument used in this study is the Comprehensive Assessment of Outcomes in Statistics (CAOS 4) test. This test is an important component of the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project funded by the National Science Foundation (NSF) to develop reliable, valid, practical, and accessible assessment items and instruments for statistics educators (DelMas, Joan, Ooms, & Chance, 2007). The purpose of the CAOS test was to develop a set of items that students, completing any introductory statistics course in college, would be expected to understand. The CAOS test consists of forty multiple choice items and a student should need 30-45 minutes to complete the test. All items were revised to ensure they involved real or realistic contexts and data. Each item on the CAOS test was designed to measure a specific learning outcome. Table 4 illustrates the measured learning outcome for each of the forty items.

Table 4: Measured Learning Outcome for each CAOS item

Item	Measured Learning Outcome
1	Ability to describe and interpret the overall distribution of a variable as displayed in a histogram, including referring to the context of the data.
2	Ability to recognize two different graphical representations of the same data (boxplot and histogram).
3	Ability to visualize and match a histogram to a description of a variable (negatively skewed distribution for scores on an easy quiz).
4	Ability to visualize and match a histogram to a description of a variable (bell-shaped distribution for wrist circumferences of newborn female infants).
5	Ability to visualize and match a histogram to a description of a variable (uniform distribution for the last digit of phone numbers sampled from a phone book).
6	Understanding to properly describe the distribution of a quantitative variable (shape, center, and spread), need a graph like a histogram which places the variable along the horizontal axis and frequency along the vertical axis.
7	Understanding of the purpose of randomization in an experiment.
8	Ability to determine which of two boxplots represents a larger standard deviation.
9	Understanding that boxplots do not provide accurate estimates for percentages of data above or below values except for the quartiles.
10	Understanding of the interpretation of a median in the context of boxplots.
11	Ability to compare groups by considering where most of the data are and focusing on distributions as single entities.
12	Ability to compare groups and comparing differences in averages

Table 4: Measured Learning Outcome for each CAOS item

Item	**Measured Learning Outcome**
13	Understanding that comparing two groups does not require equal sample sizes in each group, especially if both sets of data are large.
14	Ability to correctly estimate and compare standard deviations for different histograms. Understands lowest standard deviation would be for a graph with the least spread (typically) away from the center.
15	Ability to correctly estimate standard deviations for different histograms. Understands highest standard deviation would be for a graph with the most spread (typically) away from the center.
16	Understanding that statistics from small samples vary more than statistics from large samples.
17	Understanding of expected patterns in sampling variability.
18	Understanding of the meaning of variability in the context of repeated measurements and in a context where small variability is desired.
19	Understanding that low p-values are desirable in research studies.
20	Ability to match a scatterplot to a verbal description of a bivariate relationship.
21	Ability to correctly describe a bivariate relationship shown in a scatterplot when there is an outlier (influential point).
22 23	Understanding that correlation does not imply causation Understanding that no statistical significance does not guarantee that there is no effect.
24	Understanding that an experimental design with random assignment supports causal inference.
25	Ability to recognize a correct interpretation of a p-value.
26	Ability to recognize an incorrect interpretation of a p-value (probability that a treatment is not effective).

Table 4: Measured Learning Outcome for each CAOS item

Item	**Measured Learning Outcome**
27	Ability to recognize an incorrect interpretation of a p-value (prob. treatment is effective).
28	Ability to detect a misinterpretation of a confidence level (the percent of sample data between confidence limits)
29	Ability to detect a misinterpretation of a confidence level (percent of population data values between confidence limits).
30	Ability to detect a misinterpretation of a confidence level (percent of all possible sample means between confidence limits)
31	Ability to correctly interpret a confidence interval.
32	Understanding of how sampling error is used to make an informal inference about a sample mean.
33	Understanding that a distribution with the median larger than mean is most likely skewed to the left.
34	Understanding of the law of large numbers for a large sample by selecting an appropriate sample from a population given the sample size
35	Understanding of how to select an appropriate sampling distribution for a particular population and sample size.
36	Understanding of how to calculate appropriate ratios to find conditional probabilities using a table of data.
37	Understanding of how to simulate data to find the probability of an observed value.
38	Understanding of the factors that allow a sample of data to be generalized to the population.
39	Understanding of when it is not wise to extrapolate using a regression model.
40	Understanding of the logic of a significance test when the null hypothesis is rejected

In order to summarize the measured learning outcomes, we classify each one of them into one of these statistics topics: data collection and design, descriptive statistics, graphical representations, probability, bivariate data, sampling distribution and variability, confidence intervals and test of significance (see Table 5). For this classification, the study of DelMas et al. (2007) was used as reference. In this study, they focused their analysis in these categories and information in their discusion section along with Table 4, provided an insight on which items fall in each category. In addition, in terms of the four phases of the statistical investigation cycle described by Franklin et al. (2007), each item is categorized into these four phases: formulate a question, collect data, analyze data and interpret results (see Table 6). For his classification, we used the definition of each phase presented in Franklin et al. (2007) along with samples questions of the LOCUS test provided in their website. Since the CAOS test does not have items that evaluates the phase of formulate a question, only the phases of collect data, analyze data and interpret results were taking into consideration.

Table 5: Statistics Topics associated to each CAOS item

Topics	**CAOS items**	**Total Number of CAOS Items**
Bivariate data	20, 21, 39	3
Confidence Interval	28, 29, 30, 31	4
Data Collection and Design	7, 22, 24, 38	4
Descriptive Statistics	8, 9, 10, 11, 12, 13, 14, 15, 18	9
Graphical Representation	1, 2, 3, 4, 5, 6, 33	7
Probability	36, 37	2
Sampling Distribution and Variability	16, 17, 32, 34, 35	5
Test of Significance	19, 23, 25, 26, 27, 40	6

Table 6: Statistical Investigation Cycle Phases associated to each CAOS item

Topics	CAOS items	Total Number of CAOS Items
Collect data	7, 38	2
Analyze Data	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 33, 34, 35, 36, 37, 39	22
Interpret Results	18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 40	16

According to DelMas et al. (2007), the CAOS test was developed through a three-year process of acquiring and writing items, revisions, feedback from advisors and class testers, and two large content validity assessments. Through out that three years, four version of the CAOS test were developed, being CAOS 4 the final version. On March 2006, a final analysis of content validation was conducted to the CAOS 4 and a group of 18 members of the advisory and editorial boards of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) were used as expert raters. These members were considered experts and leaders in the national statistics education community and they all agreed that CAOS 4 measures important basic learning outcomes and 94% of the members agreed that it measures important learning outcomes. The authorization to used the CAOS test in this study was provided via email by one of the authors.

As mentioned earlier, all the topics evaluated in the CAOS 4 test are covered in the introductory statistics course (ESMA3016) and each one of the forty-multiple choice evaluated the different standards and expectations in statistics topics mentioned in the Puerto Rico Core Standards (Department of Education of Puerto Rico, 2014). Table 7 shows how standards mentioned in the Puerto Rico Core Standards are associated with each question of the CAOS 4.

Table 7: Puerto Rico Core Standards topic associated with each CAOS item

Standard Code	CAOS Item
8.E.12.1 and 8.E.12.2	1,2,3,4,5,6,8,9,10,11,12,13,14,15,18,33,34
8.E.15.2	7
ES.E.42.2	16
9.E.17.2	17,37
ES.E.45.1	19,23,24,25,26,27,28,29,30,31,32,40
8E.14.1	20,21
ES.E.41.3	22
ES.E.41.2	35
ES.E.47.1	36
7.E.17.2	38
ES.E.44.1	39

Subjects

In this study, there are two different populations. The first population is the students graduated from the mathematics education program in the last five years (from 2012-2013 to 2017-2018). The second one is the students who are currently in the mathematics education program and have approved the course ESMA 3016 with a grade of C or more. The first population is referred as practicing mathematics teachers regardless of their current status as mathematics teachers. The second population is referred as prospective mathematics teachers. From the practicing teachers, all fifteen potential subjects were contacted. From those fifteen, only four responded and three of them decided to participate in the study. In terms of the prospective teachers, all six potential subjects were contacted in which four of them responded and decided to participate in the study. Thus, in total there are seven subjects in which three are practicing and four are prospective teachers. Table 8 shows the summary described above.

Table 8: Subjects in the Study

Population	Contacted	Responded	Participated
Prospective	6	4	4
Practicing	15	4	3
Total	**21**	**8**	**7**

Table 9 shows the year when the practicing teachers completed their program. In terms of having mathematics teaching experience, Figure 2 shows that two of the three practicing teachers have had experiences teaching mathematics. Of those, one of them has taught statistics (Figure 3). In terms of additional statistics courses taken through out their academic preparation, besides ESMA 3016, Figures 4 and 5 shows how many of them took additional statistical courses during or after the bachelor's degree.

Table 9: Year Program Completed for each Participant

Subjects	Year Program was Completed
S1	In progress
S2	In progress
S3	In progress
S4	2015-2016
S5	2012-2013
S6	2012-2013
S7	In progress

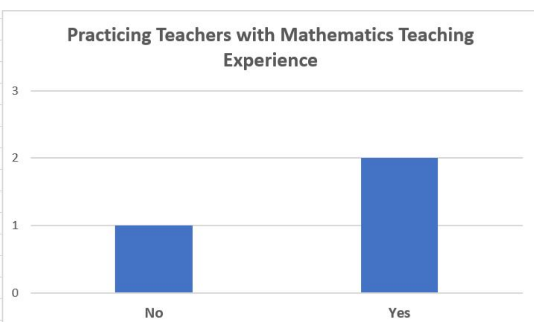


Figure 2: Practicing teachers with experience in teaching Mathematics

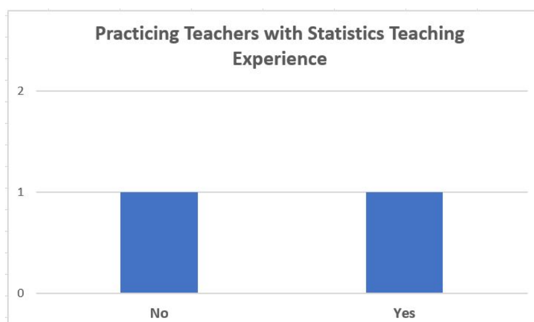


Figure 3: Practicing teachers with experience in teaching Statistics

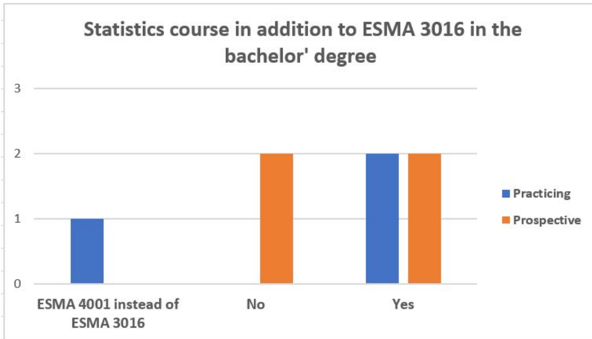


Figure 4: Subjects with additonal Statistics Course in the bachelor degree

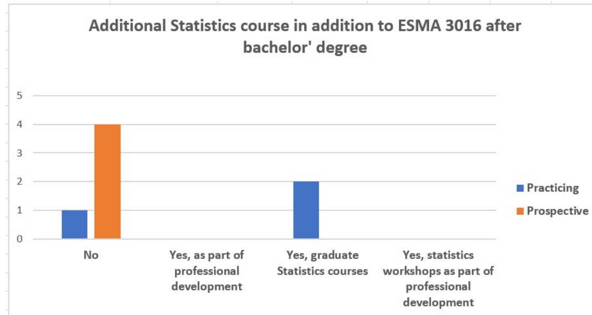


Figure 5: Subjects with additonal Statistics Course after the bachelor degree

Data Sources

The CAOS 4 intsrument was uploaded as a Google Form and sent to subjects via their institutional UPR email and through Facebook (when an account was available). In order to contact the subjects, a protocol was first approved through the IRB at UPRM (see Appendix C). Through the Google form, subjects must first read and electronically accept the concent form before proceeding to the test. In addtion to the CAOS test, 3 questions were conducted to gather subjects opinions and feedbacks on how effective the mathematics education program from UPRM is in preparing mathematics teachers to teach statistics.

Data Analysis

To answer the first research question established in the last part of the Introduction chapter, quantitative and qualitative analysis will be conducted in the following chapter. Quantitative analysis includes confidence intervals for the overall mean CAOS score and binomial regressions to test if different factors contributes to the proportion of correct answers in the CAOS test. For the factors that were significant, contrast tests were conducted between the high levels and the lower levels of those factors. For the goodness of fit of the binomial regressions, we used the quotient of the statistic chi-squared and the degree of freedom of the error. If this quotient is between 0.5 and 1.5, it is an indication of a good fit (Dobson & Barnett, 2008). If it is below 0.5, it may be a signal of underdispersion. Meanwhile, if it is over 1.5, it may be a signal of overdispersion (Dobson & Barnett, 2008). In addtion, for the quantitative analysis, we identify in which CAOS items the subjects struggle the most and in which CAOS items they were proficient. Most of the analysis was conducted using SAS statistical software. In terms of answering the second question, based on the feedbacks and recommendations given by each subjects, a series of recommendations and feedbacks will be giving to the Mathematical Science Departments to improve the mathematics education program.

Chapter 4 Results

In this chapter, we evaluated how effective prospective and practicing secondary mathematics teachers are being prepared to teach statistics by the UPRM mathematics education program. This chapter is divided into four quantitative analysis sections and one qualitative analysis section. The first section will be evaluating the overall performance of the seven subjects in the CAOS test. Next, we will take into consideration how factors like *Population*, *Statistics Topic* and *Statistics Phase* influence the subject's CAOS Score. After that, the same analysis will be conducted, but with *Statistics Course* factor instead of the *Population* factor. In addition, an analysis will be conducted for each item in the CAOS test to identify those items the subjects struggle the most. The last part will summarize the results of the feedbacks and recommendations provided by the subjects about the effectiveness of the UPRM mathematics education program to prepare future mathematics teachers to teach statistics at the secondary level.

Overall CAOS Scores

Each item in the CAOS test was worth one point; giving the CAOS test a total score of forty points. The number of corrected answer and the score for each of the seven subjects are shown in Table 10 As we can see, none of them had a score greather than 60% in the CAOS test. In Figure 6, a dotplot of the overall CAOS Scores is shown.

Table 10: CAOS test scores of each participant

Subject	**Statistics Course**	**Teaching Experience**	**Number of Correct Answers (out of 40)**	**CAOS Score**
Subject 1	Additional Course	No	15	37.50%
Subject 2	Only ESMA 3016	No	19	47.50%
Subject 3	Only ESMA 3016	No	23	57.50%
Subject 4	Additional Course	No	17	42.50%
Subject 5	Additional Course	Yes	13	32.50%
Subject 6	Additional Course	Yes	14	35.00%
Subject 7	Additional Course	No	15	37.50%

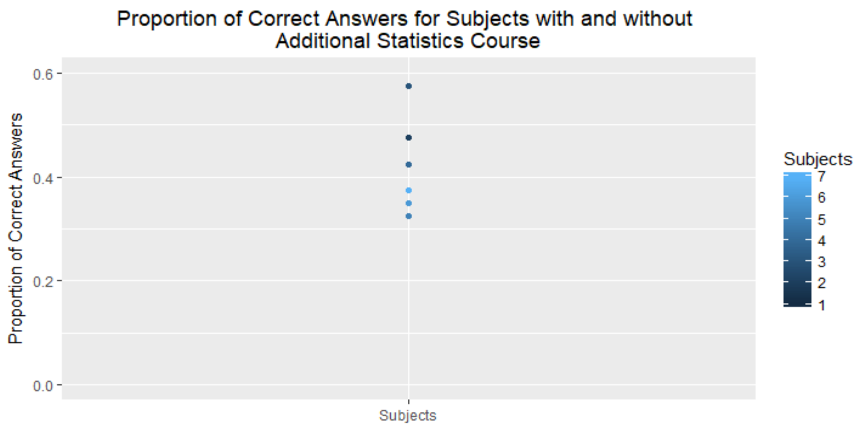


Figure 6: Dotplot of CAOS Scores

In order to get an estimation of the CAOS Score mean from the prospective and practicing teachers in general from the UPRM mathematics education program, we will calculate a 95% Confidence Interval for the mean of the CAOS score. First, we need to validate some assumptions. Since each subject is a different person, each subject is independent. In addition, we have to verify if the population follow a normal distribution by testing the hypothesis established in Equation (1). In Figures 7 and 8, the Anderson Darling test and the Kolmogorov Smirnov test are performed to test the following statistical hypothesis with a significance level of 0.05.

$$
\begin{aligned}
&H_0 : The\ population\ follows\ a\ normal\ distribution \\
&H_1 : The\ population\ does\ not\ follows\ has\ a\ normal\ distribution
\end{aligned} \tag{1}
$$

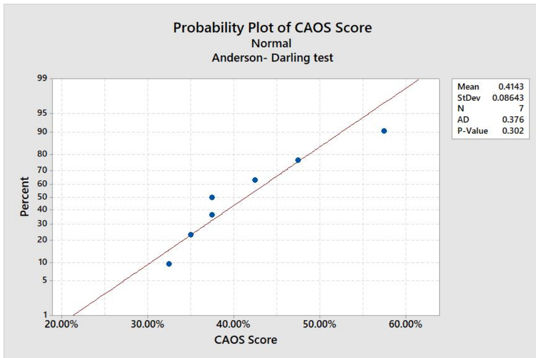


Figure 7: Anderson Darling Normality test

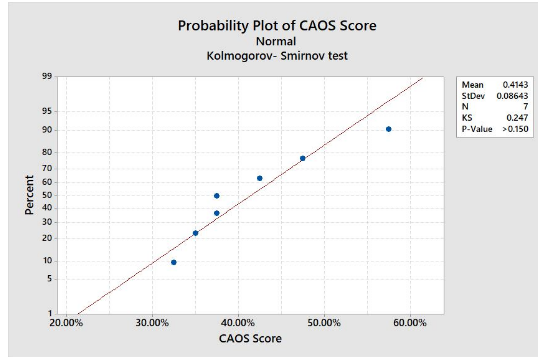


Figure 8: Kolmogorov Smirnov test

Since in both tests, the p value is greather than 0.05, there are not sufficient evidence to reject H_0 with a significance level of 0.05. As a result, from Equation (1), we will assume the population follows a normal distribution. Since the population variance is unknown and the sample size is small (less than 30), the student t distribution will be used to calculated the 95% Confidence Interval for the mean. Table 11 shows the value and the meaning of the parameters used in Equation (2), and the 95% Confidence Interval for the mean of the CAOS score for all the subjects is calculated.

Table 11: Parameters for the 95% Confidence Interval for the mean of the CAOS Score

Parameter	**Meaning**	**Value**
$\bar{x}$	sample average CAOS Scores	0.4143
s	sample standard deviation of CAOS Scores	0.0864
n	sample size	7
$t_{\frac{\alpha}{2},n-1}$	Critical Value	2.4469
α	significance level	0.05
μ	population mean of the CAOS Score of the propective and practicing secondary mathematics teachers graduated from the mathematics education program at UPRM	to be estimated

$$\begin{gathered}\bar{x} - t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right) \\ 0.4143 - 2.4469\left(\frac{0.0864}{\sqrt{7}}\right) < \mu < 0.4143 + 2.4469\left(\frac{0.0864}{\sqrt{7}}\right) \\ 0.3344 < \mu < 0.4942\end{gathered} \tag{2}$$

With 95% confidence, the real value of the mean of the CAOS score of the students graduated in the last five years or currently enrolled in the mathematics education program with ESMA 3016 course approved with C plus falls approximately between 0.33 and 0.49. Technically, we can say with 95% confidence that CAOS score average for the our population is an F grade.

In the next section, we will analyze if the factor of *Population* in combination with *Statistics Topic* and *Statistics Phase* factors, respectively, influence the CAOS Score.

Prospective versus Practicing Teachers

Population and Statistics Topic

For this section, we conducted an analysis to evaluate if such factor as *Population* and *Statistics Topic* influenced the performance in the CAOS test. From hereafter, we will refer to the CAOS Score as the proportion of correct answer in the CAOS test. The seven subjects were divided into two groups: prospective and practicing teachers; and the proportions of correct answers was divided for each statistics topic mentioned in Table 5 (Bivariate data, Confidence Interval, Data Collection and Design, Descriptive Statistics, Graphical Representation, Probability, Sampling Distribution and Variability, and Test of Significance). Figure 9 shows dotplots of the proportion of correct answers for each population, using the data in Table 10. From Figure 9, we can informally mentioned that prospective subjects performed a little better than practicing teachers, but seems there is not a significance difference in their CAOS Score means.

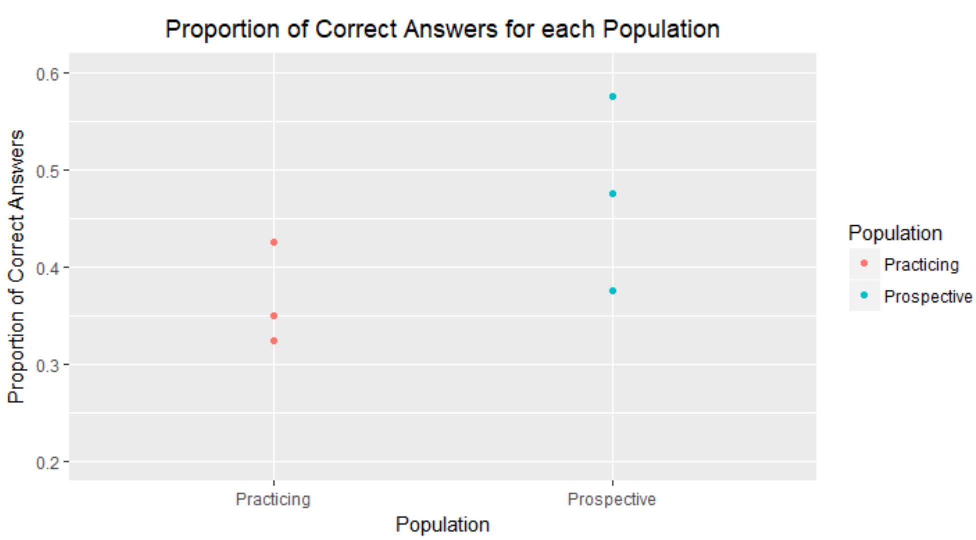


Figure 9: Dotplot of Proportion of Correct Answers for each Population

For the *Statistics Topic* factor, the number of correct answers for each subject in each statistics topic is shown in Table 12. Figure 10 shows dotplots of the proportion of correct answers in each statistics topic, using the data in Table 12. In Figure 10, is not clear that a significance difference exists in the proportion of correct answers in each statistics topic since there is a large variability in them. We will refer to Figure 11 later in the analysis.

Table 12: Population and Statistics Topic Factors

		Population						
		Prospective				Practicing		
Statistics Topic	**Item**	S_1	S_2	S_3	S_7	S_4	S_5	S_6
Bivariate data	Correct	3	2	3	1	2	1	2
	Total	3	3	3	3	3	3	3
Confidence Interval	Correct	2	1	2	3	1	1	3
	Total	4	4	4	4	4	4	4
Data Collection and Design	Correct	0	1	0	2	1	2	0
	Total	4	4	4	4	4	4	4
Descriptive Statistics	Correct	4	5	6	2	5	2	4
	Total	9	9	9	9	9	9	9
Graphical Representation	Correct	1	3	5	2	5	3	1
	Total	7	7	7	7	7	7	7
Probability	Correct	2	1	2	0	1	2	0
	Total	2	2	2	2	2	2	2
Sampling Distribution and Variability	Correct	1	3	2	1	1	1	1
	Total	5	5	5	5	5	5	5
Test of Significance	Correct	2	3	3	4	1	1	3
	Total	6	6	6	6	6	6	6

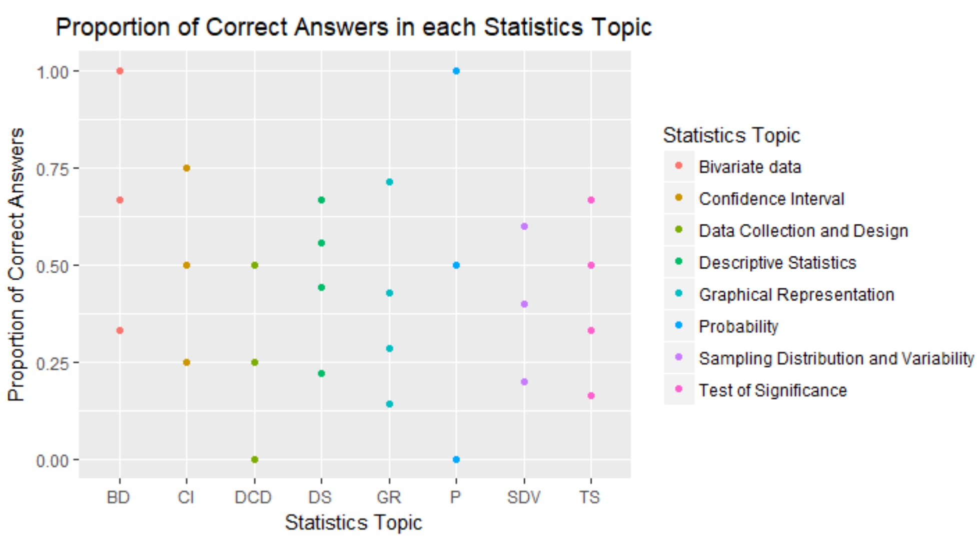


Figure 10: Dotplot of Proportion of Correct Answers for each Statistics Topic

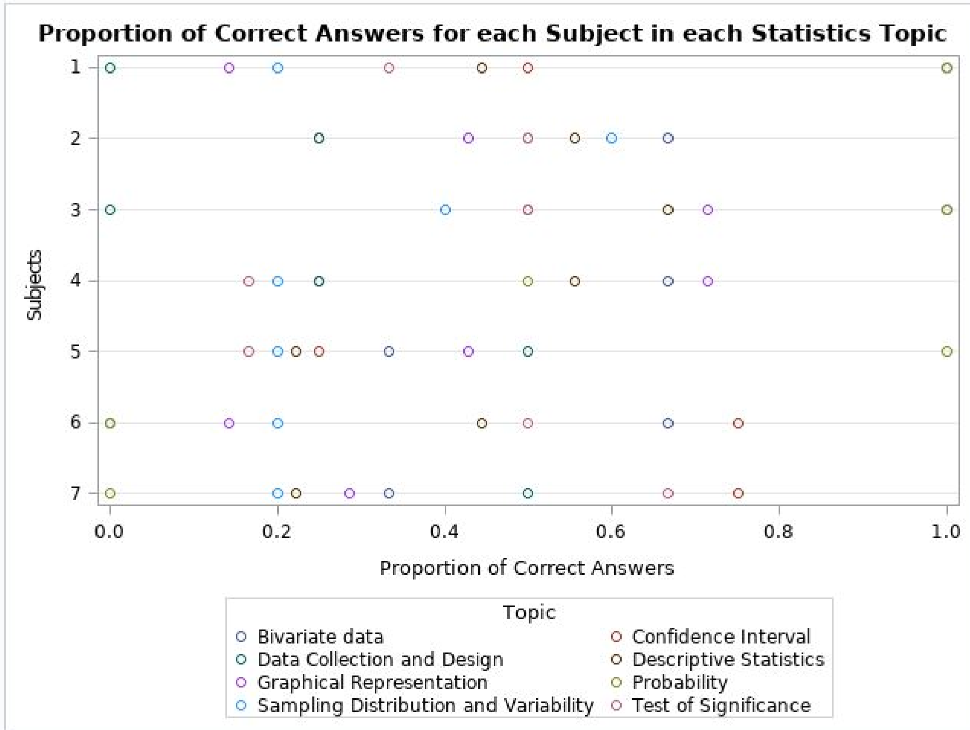


Figure 11: Proportion of Correct Answer for each Subject in each Statistics Topic

Looking at both factors at the same time, Figures 12 and 13 show the proportion mean of correct answers for each Population for each Statistics Topic. From Figure 12, we can observe that on average both Prospective and Practicing subjects performed better in the Bivariate Data and Probability topics. On the other hand, Data Collection and Design and Sampling Distribution and Variability were the topics that both Prospective and Practicing subjects struggle the most with. From this figures, we can think that an interaction between the two factors may exist.

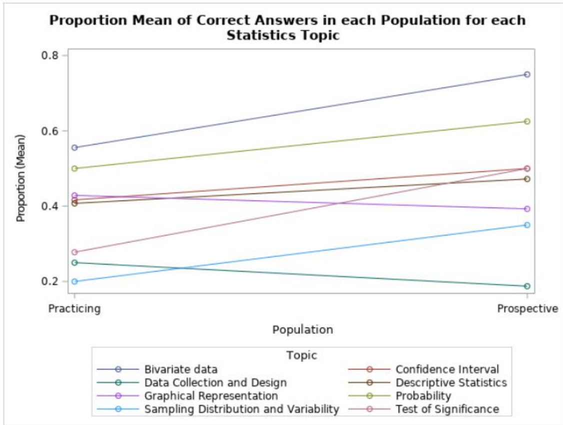


Figure 12: Proportion Mean of Correct Answers in each Population for each Statistics Topic

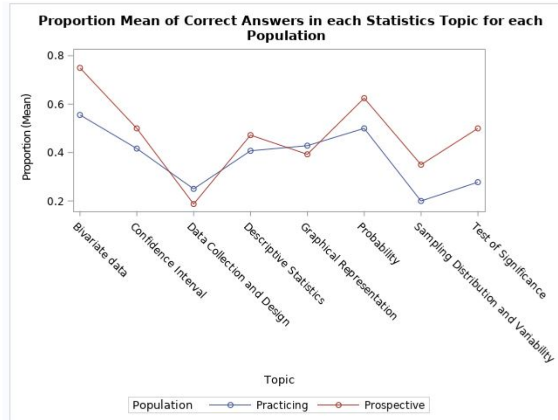


Figure 13: Proportion Mean of Correct Answers in each Statistics Topic for each Population

In order to test for significance of these two factors formally, we conducted a binomial regression analysis with the logit function as the link function. The reason we used this regression is because the number of items in each statistics topic differs from each one of them and it is more accurate to model the proportion of correct answers taking into consideration the sample size of each statistics topic. In addition, our random variable is the number of correct answers from diferent number of question in each Topic. For this model, we have two fixed effects (*Population* and *Statistics Topic*) and their interaction. Also, we add the subjects as a random effects since we have correlated data. The correlation is assumed since the number of correct answers is calculated for each subject in each statistics topic. In Equation (3), the binomial regression model used is shown with its assumptions, and Table 13 explains each parameter in the model.

$$
\begin{aligned}
logit(p_{ijk}) &= \mu_0 + \alpha_i + \beta_j + b_k + (\alpha\beta)_{ij} \\
Y_{ijk} &\sim Binomial(n_i, p_{ijk}) \\
b_k &\sim N(0, \sigma_s^2) \\
Y_{ijk} &: number\ of\ correct\ answers\ in\ the\ Statistics\ Topic\ i\ of\ the\ subject\ k \\
&\quad from\ Population\ j\ in\ the\ CAOS\ test \\
n_i &: number\ of\ items\ in\ the\ Statistics\ Topic\ i\ in\ the\ CAOS\ test \\
p_{ijk} &: proportion\ of\ correct\ answers\ in\ the\ Statistics\ Topic\ i\ of\ the\ Subject\ k \\
&\quad from\ Population\ j\ in\ the\ CAOS\ test
\end{aligned} \tag{3}
$$

Table 13: Description of Parameters and Response Variable for Model 1

Term	**Description**
p_{ijk}	proportion of correct answers in the CAOS test in the Statistics Topic level i of subject k in Population j
μ_0	Intercept
α_i	effect of Statistics Topic i in the proportion of correct answers in the CAOS test
β_j	effect of Population j in the proportion of correct answers in the CAOS test
b_k	effect of Subject k in the proportion of correct answers in the CAOS test
$(\alpha\beta)_{ij}$	interaction effect between Statistics Topic i and Population j in the proportion of correct answers in the CAOS test

For this analysis, the three hypothesis (Equation (4)) will be tested at a significance level of 0.15. The reason we chose a significance level of 0.15 is because this study does not need a very conservative type one error since we are trying to identify which factors influence the proportion of correct answers in the CAOS test. In addition, if this study was perfomed with a small sample of subjects and a significance level of 0.05, it would be very difficult for the test to detect significance in the factors.

$$
\begin{aligned}
&H_{0a} : \alpha_i = 0 && H_{0b} : \beta_j = 0 && H_{0c} : (\alpha\beta)_{ij} = 0 \\
&H_{1a} : at\ least\ one\ \alpha_i \neq 0 && H_{1b} : at\ least\ one\ \beta_j \neq 0 && H_{1c} : at\ least\ one\ (\alpha\beta)_{ij} \neq 0
\end{aligned} \tag{4}
$$

The analysis was conducted using procedure glimmix in SAS software. In the Figures 14-19, the outputs of SAS are shown.

Table 14: Description of the Index of Model 1

Index	**Value**	**Description**
i	1	Bivariate Data
	2	Confidence Interval
	3	Data Collection and Design
	4	Descriptive Statistics
	5	Graphical Representation
	6	Probability
	7	Sampling Distribution and Variability
	8	Test of Significance
j	1	Prospective
	2	Practicing
k	1	S1
	2	S2
	1	S3
	2	S7
	3	S4
	4	S5
	5	S6

Model Information	
Data Set	WORK.STAT_TOPIC_LOGIC
Response Variable (Events)	Correct
Response Variable (Trials)	Trials
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Not blocked
Estimation Technique	Residual PL
Degrees of Freedom Method	Satterthwaite

Figure 14: Information about the model

Fit Statistics	
-2 Res Log Pseudo-Likelihood	142.85
Generalized Chi-Square	48.70
Gener. Chi-Square / DF	1.22

Figure 15: Fit Model Criteria

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
Subjects	0.01773	0.08037

Figure 16: Covariance Parameter Estimator of the Subjects effect

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Topic	7	35	1.87	0.1047
Population	1	35	1.73	0.1969
Topic*Population	7	35	0.40	0.8974

Figure 17: F test for the model

Figure 14 provides information about the fitted model. As we can see, a binomial regression was fitted with the logit function as the link function. Figure 15 provides information about the goodness of fit of the model. The quotient X^2/DF = 48.70/40 = 1.22 indicates that the model fit the data well since it is close to 1. Moving on to Figure 16, the covariance estimate for the subject

Solutions for Fixed Effects							
Effect	Topic	Population	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			5.309E-6	0.4138	5	0.00	1.0000
Topic	Bivariate data		1.0992	0.7819	35	1.41	0.1686
Topic	Confidence Interval		1.93E-15	0.6457	35	0.00	1.0000
Topic	Data Collection and Design		-1.4671	0.7597	35	-1.93	0.0616
Topic	Descriptive Statistics		-0.1113	0.5275	35	-0.21	0.8341
Topic	Graphical Representation		-0.4356	0.5627	35	-0.77	0.4440
Topic	Probability		0.5111	0.8369	35	0.61	0.5453
Topic	Sampling Distribution and Variability		-0.6194	0.6218	35	-1.00	0.3260
Topic	Test of Significance		0	.	.	.	.
Population		Practicing	-0.9557	0.6739	35	-1.42	0.1650
Population		Prospective	0	.	.	.	.
Topic*Population	Bivariate data	Practicing	0.07960	1.1569	35	0.07	0.9455
Topic*Population	Bivariate data	Prospective	0	.	.	.	.
Topic*Population	Confidence Interval	Practicing	0.6191	1.0182	35	0.61	0.5471
Topic*Population	Confidence Interval	Prospective	0	.	.	.	.
Topic*Population	Data Collection and Design	Practicing	1.3240	1.1396	35	1.16	0.2532
Topic*Population	Data Collection and Design	Prospective	0	.	.	.	.
Topic*Population	Descriptive Statistics	Practicing	0.6922	0.8418	35	0.82	0.4165
Topic*Population	Descriptive Statistics	Prospective	0	.	.	.	.
Topic*Population	Graphical Representation	Practicing	1.1035	0.8877	35	1.24	0.2221
Topic*Population	Graphical Representation	Prospective	0	.	.	.	.
Topic*Population	Probability	Practicing	0.4445	1.2823	35	0.35	0.7309
Topic*Population	Probability	Prospective	0	.	.	.	.
Topic*Population	Sampling Distribution and Variability	Practicing	0.1886	1.0394	35	0.18	0.8571
Topic*Population	Sampling Distribution and Variability	Prospective	0	.	.	.	.
Topic*Population	Test of Significance	Practicing	0	.	.	.	.
Topic*Population	Test of Significance	Prospective	0	.	.	.	.

Figure 18: Fixed Effects Parameter Estimated

Solution for Random Effects						
Effect	Subjects	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Subjects	1	-0.04574	0.1260	35	-0.36	0.7188
Subjects	2	0.01526	0.1260	35	0.12	0.9043
Subjects	3	0.07621	0.1259	35	0.61	0.5490
Subjects	4	0.03575	0.1270	35	0.28	0.7799
Subjects	5	-0.02554	0.1270	35	-0.20	0.8418
Subjects	6	-0.01021	0.1270	35	-0.08	0.9364
Subjects	7	-0.04574	0.1260	35	-0.36	0.7188

Figure 19: Random Effects Parameter Estimated

effect ($\sigma_s^2 = 0.0177$) is very close to 0, indicating that maybe there is no correlation induced by the subjects. If we look at Figure 11, we can see the proportion of correct answers by each subject for each statistics topic. Since the range of the proportion of correct answers through all the subjects is almost the same, this explains why the covariance estimator is almost 0. As a result, the random effect in the model may not be needed. A formal hypothesis test will be conducted later on to test if the random effect is necessary.

We will focused now in the last three figures of the output. In Figure 17, the three hypothesis from Equation (4) has been tested. Since the p value of the hypothesis associated to the Statistics Topic factor is less than 0.15, we can conclude that at a significance level of 0.15, the Statistics Topic factor affects the proportion of correct answers in the CAOS test.

Next, we proceed to conduct a Contrast test to see if there are significance differences between the statistics topics of Bivariate Data and Data Collection and Design, Bivariate Data and Sampling Distribution and Variability and Bivariate Data and Probability since they are the topics in which the subjects obtained the highest and the lowest proportion of correct answers in the CAOS test. Figure 20 shows the results of the Contrast tests. As we can see, there is significance diferences between Bivariate Data with both Data Collection and Design and Sampling Distribution and Variability at a significance level of 0.15. However between Bivariate Data and Probability there is not a significance difference at a significance level of 0.15. This results validates what we saw in Figure 12. In conclusion, the overall test of the factor of Statistics Topic was significant at a significance level of 0.15 with the Statistics Topics of Bivariate Data and Probability being the ones with higher proportion of correct answers and Data Collection and Design and Sampling Variability and Distribution being the ones with lower proportion of correct answers.

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
Bivariate Data - Data Collection and Design	1	40	8.64	0.0054
Bivariate Data - Probability	1	40	0.31	0.5784
Bivariate Data - Sampling Distribution	1	40	7.23	0.0104

Figure 20: Contrast test for Statistics Topic

Recalling from Figure 16, $\sigma_s^2 = 0.01773$ which is very close to 0. Is in our interested to test formally this hypothesis to see if the random effect is really necessary or if the subjects are not inducing correlation in the data. The hypothesis to be tested is presented in Equation (5). Since the variance is a non negative value, we are testing a null hypothesis with a value in the border of the domain of the parameter. As a result, the test is a mixure chi square with an adjusted p value. Figure 21 presented the result for test using SAS statistical software. Since the p value is greater than 0.15, there is not sufficient evidence to reject the hypothesis in Equation (5) at a significance level of 0.15. Based on that, we can say that $\sigma_s^2 = 0$ and as a result, the random effect is not necessary in the model.

$$H_{0s} : \sigma_s^2 = 0$$
$$H_{1s} : \sigma_s^2 > 0 \tag{5}$$

Tests of Covariance Parameters
Based on the Residual Pseudo-Likelihood

Label	DF	-2 Res Log P-Like	ChiSq	Pr > ChiSq	Note
random effect variance	1	142.91	0.06	0.4037	MI

MI: P-value based on a mixture of chi-squares.

Figure 21: Covariance test for $\sigma_s^2 = 0$

Population and Statistics Phase

In this subsection, we will perform the same analysis described above, but using the *Statistics Phase* factor mentioned in Table 6, instead of the *Statistics Topic* factor. The statistics phase are based on the Statistical Investigation Cycle Phases mentioned established in the GAISE Report (Franklin et al. (2007)). It is important to clarify that we are going to take into consideration only the statistics phases of Collect Data, Analyze Data and Interpret Results since the CAOS test does not have items for the Formulate Question phase. The data used for this analysis in presented in Table 15 and Figure 22 presents dotplots of the proportion of correct answers for each statistics phase. From Figure 22, we can see that the subjects had most difficulties in the Collect Data phase.

Table 15: Population and Statistics Phase Factors

		Population						
		Prospective				Practicing		
Statistics Phase	**Item**	S_1	S_2	S_3	S_7	S_4	S_5	S_6
Collect Data	Correct	0	0	0	1	0	0	0
	Total	2	2	2	2	2	2	2
Analyze Data	Correct	8	11	15	4	12	8	5
	Total	22	22	22	22	22	22	22
Interpret Results	Correct	7	8	8	10	5	5	9
	Total	16	16	16	16	16	16	16

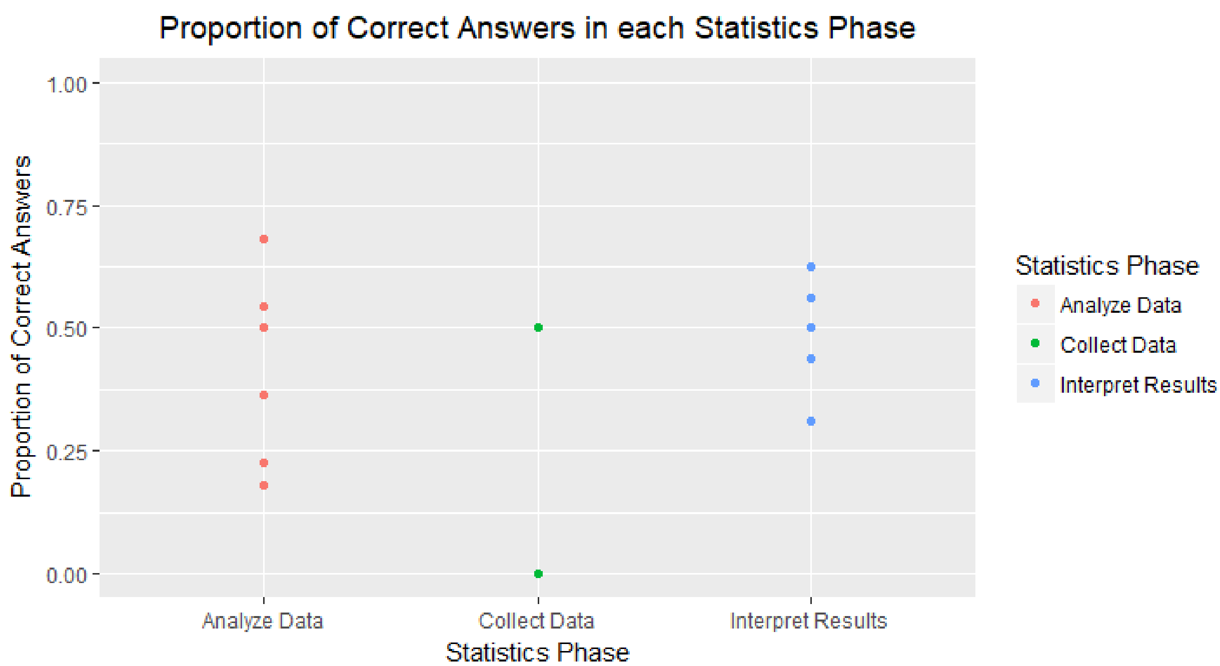


Figure 22: Dotplot of Proportion of Correct Answers in each Statistics Phase

Figure 23 and Figure 24 show the proportion mean of correct answers in each Population for each Statistics Phase and viceversa, respectively. From Figure 23, we can see that on average that Collect Data was the phase that both Prospective and Practicing subjects struggle more with and that in Analyze Data and Interpret Results, there may not be a significance difference. From these figures, we can think that an interaction between the two factors does not exist.

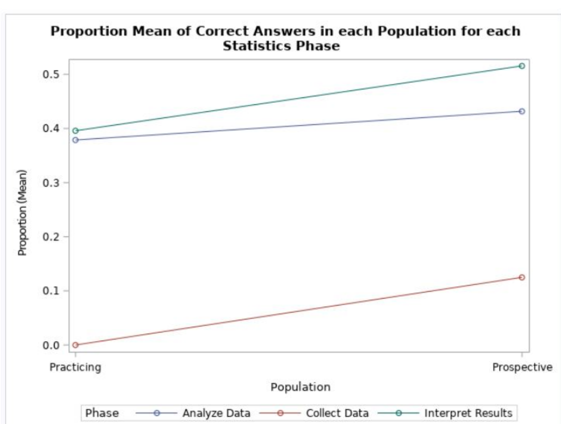


Figure 23: Proportion Mean of Correct Answers in each Population for each Statistics Phase

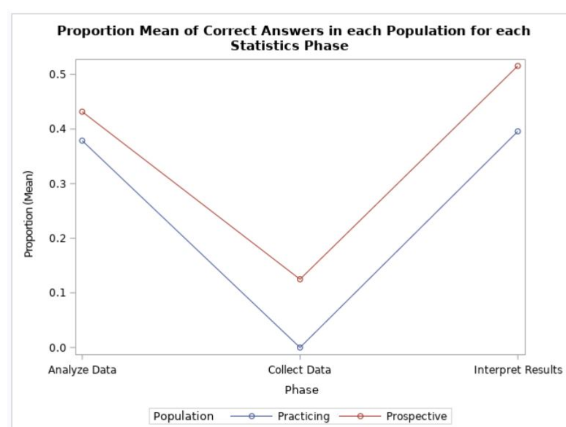


Figure 24: Proportion Mean of Correct Answers in each Statistics Phase for each Population

For this analysis, we will used the same binomial regression analysis with the logit function as the function link. For this model, we initially added the random effect of the subjects but since there was a problem of convergence in the analysis, we decided to remove it from the model. The model along with the asumptions used for this analysis is shown in Equation (6) and Table 16. The parameters in the model are explained in Table 17. The three hypothesis to be tested are shown in Equation (7).

$$\begin{aligned}
logit(p_{ijl}) &= \eta_0 + \gamma_i + \beta_j + (\alpha\beta)_{ij} \\
Y_{ijl} &\sim Binomial(n_i, p_{ijl}) \\
Y_{ijl} &: number\ of\ correct\ answers\ in\ the\ Statistics\ Phase\ i\ in\ the\ replicate\ l \\
&\quad from\ Population\ j\ in\ the\ CAOS\ test \\
n_i &: number\ of\ items\ in\ the\ Statistics\ Phase\ i\ in\ the\ CAOS\ test \\
p_{ijl} &: proportion\ of\ correct\ answers\ in\ the\ Statistics\ Phase\ i\ in\ the\ replicate\ l \\
&\quad from\ Population\ j\ in\ the\ CAOS\ test
\end{aligned} \tag{6}$$

Table 16: Description of the Index of Model 2

Index	**Value**	**Description**
i	1	Collect Data
	2	Analyze Data
	3	Interpret Results
j	1	Prospective
	2	Practicing
l	1	S1 if j=1
	2	S2 if j=1
	3	S3 if j=1
	4	S7 if j=1
	1	S4 if j=2
	2	S5 if j=2
	3	S6 if j=2

$$\begin{array}{lll}
H_{0d}: \gamma_i = 0 & H_{0e}: \beta_j = 0 & H_{0f}: (\gamma\beta)_{ij} = 0 \\
H_{1d}: at\ least\ one\ \gamma_i \neq 0 & H_{1e}: at\ least\ one\ \beta_j \neq 0 & H_{1f}: at\ least\ one\ (\gamma\beta)_{ij} \neq 0
\end{array} \tag{7}$$

Table 17: Description of Parameters and Response Variable

Term	**Description**
p_{ijk}	proportion of correct answer in the Statistics Phase level i in Population j
η_0	Intercept
γ_i	effect of Statistics Phase i in the proportion of correct answers
β_j	effect of Population j in the proportion of correct answers
$(\gamma\beta)_{ij}$	interaction effect between Statistics Phase i and Population j in the proportion of correct answers

Model Information	
Data Set	WORK.STAT_PHASE_LOGIC
Response Variable (Events)	Correct
Response Variable (Trials)	Trials
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Figure 25: Information about the model

Fit Statistics	
-2 Log Likelihood	72.30
AIC (smaller is better)	84.30
AICC (smaller is better)	90.30
BIC (smaller is better)	90.57
CAIC (smaller is better)	96.57
HQIC (smaller is better)	85.66
Pearson Chi-Square	24.21
Pearson Chi-Square / DF	1.61

Figure 26: Fit Statistics Model 2

The analysis was conducted using procedure glimmix in SAS software. On Figure 25, the information for the fitted model is presented. In Figure 26, the quotient $X^2/DF = 1.61$ indicated that there may be a warning of overdispersion since it is greather than 1.50. However, a negative impact of overdispersion is that produces falses positives and since, none of the factors were significant in the analysis, the overdispersion did not had a negative effect in the analysis. Finally, Figure 27 shows the results from the three hypothesis tested from Equation (7). Since the three p values are greather than 0.15, we can say that at a significance level of 0.15, there is not sufficient evidence to reject the three hypothesis in Equation (7). As a result, we can conclude that none of the factors influence in the proportion of correct answers in the CAOS test.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Phase	2	15	0.32	0.7299
Population	1	15	0.00	0.9767
Phase*Population	2	15	0.14	0.8745

Figure 27: F test for Model 2

Parameter Estimates							
Effect	Phase	Population	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			0.06252	0.2509	15	0.25	0.8066
Phase	Analyze Data		-0.3370	0.3300	15	-1.02	0.3234
Phase	Collect Data		-2.0084	1.0979	15	-1.83	0.0873
Phase	Interpret Results		0	.	.	.	.
Population		Practicing	-0.4854	0.3869	15	-1.25	0.2288
Population		Prospective	0	.	.	.	.
Phase*Population	Analyze Data	Practicing	0.2651	0.5103	15	0.52	0.6110
Phase*Population	Analyze Data	Prospective	0	.	.	.	.
Phase*Population	Collect Data	Practicing	-11.4711	426.38	15	-0.03	0.9789
Phase*Population	Collect Data	Prospective	0	.	.	.	.
Phase*Population	Interpret Results	Practicing	0	.	.	.	.
Phase*Population	Interpret Results	Prospective	0	.	.	.	.

Figure 28: Fixed Effects Parameter Estimates

In the following section, the subjects are divided into subjects who had taken additional statistics courses besides ESMA 3016 and subjects with only ESMA 3016 as statistics course and analyze how this factor in combination with *Statistics Course* and *Statistics Phase* factors influence in the proportion of correct answers in the CAOS test.

Subjects with Additional Statistics courses versus Only ESMA 3016 course

Statistics Course and Statistics Topic

For this analysis, the subjects are divided into two groups: those who had taken additional statistics courses besides ESMA 3016 and those who only had only taken ESMA 3016 as a statistics course. Figure 29 shows the dotplots of the proportion of correct answers for each level in Statistics Course. From this figure, we notice that subjects with only ESMA 3016 course performed better than subjects with addtional statistics courses and it appears to be a significant difference between them.

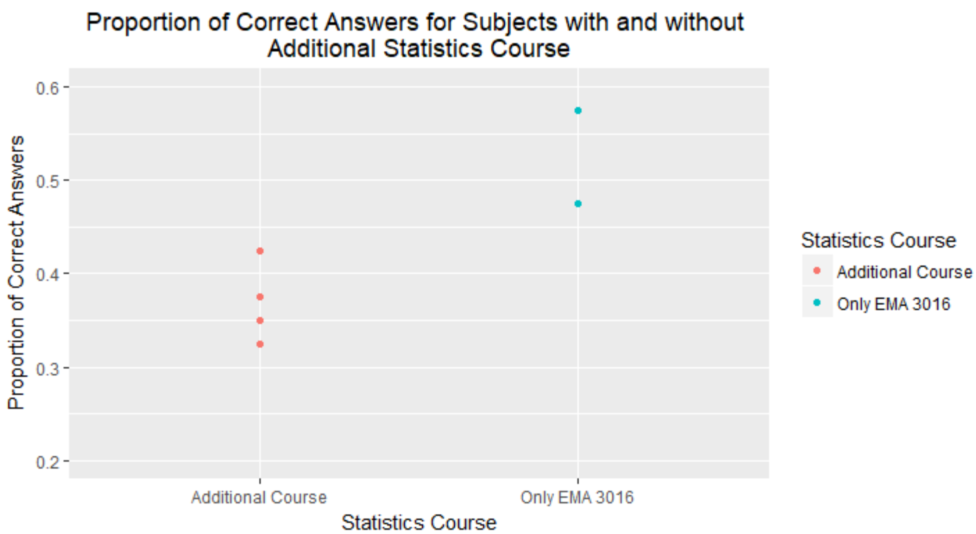


Figure 29: Dotplot of Proportion of Correct Answers for Subjects with and without Additional Statistics Course

Observing both factors at the same time, the next figures show the proportion mean of correct answers in each Population for each Statistics Phase (Figure 30) and viceversa (Figure 31). From Figure 31, we can see that on average subjects with only ESMA 3016 performed better in all Statistics Topics except in Confidence Interval and Data Collection and Design. As result, an interaction between the two factors may exists.

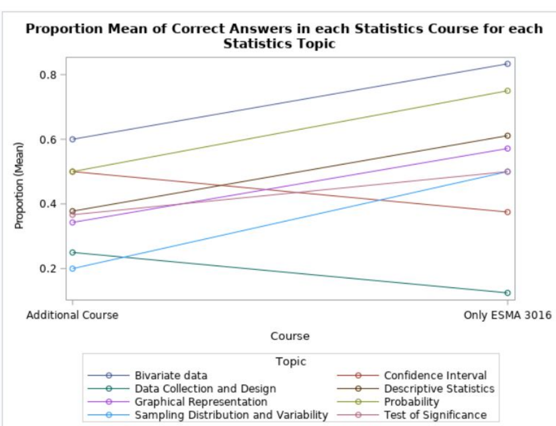


Figure 30: Proportion Mean of Correct Answers in each Statistics Course for each Statistics Topic

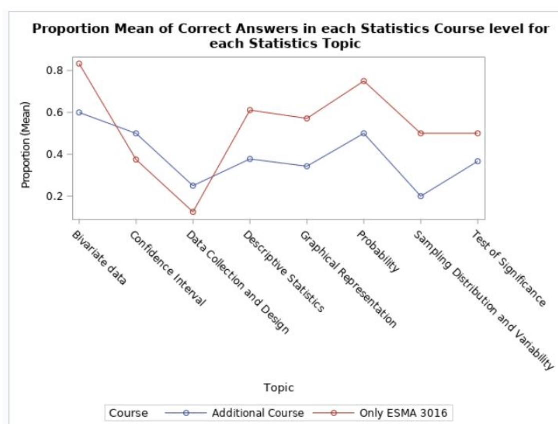


Figure 31: Proportion Mean of Correct Answers in each Statistics Topic for each Statistics Course

Just like in the section of Prospective versus Practicing, we want to test how the *Statistics Course* factor influenced the proportion of correct answers in the CAOS test in combination with the *Statistics Topic* factor. The random effect of the subjects was not included in this model due to the problem of convergence. The data for this analysis is shown in Table 18. The binomial regression analysis with the logit function as the link function is shown in Equation (8) with its assumptions. Also, Table 20 and Table 19 explain each of the parameter used in the model and their index respectively. Lastly, Equation (9) shown the three hypothesis to be tested in this analysis.

Table 18: Statistics Course and Statistics Topic Factors

		Statistics Course						
		Only ESMA 3016		Additional Course				
Statistics Topic	**Item**	S_2	S_3	S_7	S_4	S_5	S_6	S_1
Bivariate data	Correct	2	3	1	2	1	2	3
	Total	3	3	3	3	3	3	3
Confidence Interval	Correct	1	2	3	1	1	3	2
	Total	4	4	4	4	4	4	4
Data Collection and Design	Correct	1	0	2	1	2	0	0
	Total	4	4	4	4	4	4	4
Descriptive Statistics	Correct	5	6	2	5	2	4	4
	Total	9	9	9	9	9	9	9
Graphical Representation	Correct	3	5	2	5	3	1	1
	Total	7	7	7	7	7	7	7
Probability	Correct	1	2	0	1	2	0	2
	Total	2	2	2	2	2	2	2
Sampling Distribution and Variability	Correct	3	2	1	1	1	1	1
	Total	5	5	5	5	5	5	5
Test of Significance	Correct	3	3	4	1	1	3	2
	Total	6	6	6	6	6	6	6

$$
\begin{aligned}
logit(p_{ijl}) &= \theta_0 + \alpha_i + \delta_j + (\alpha\delta)_{ij} \\
Y_{ijl} &\sim Binomial(n_i, p_{ijl}) \\
Y_{ijl} &: number\ of\ correct\ answers\ in\ the\ Statistics\ Topic\ i\ in\ the\ replicate\ l \\
&\quad from\ Statistics\ Course\ j\ in\ the\ CAOS\ test \\
n_i &: number\ of\ items\ in\ the\ Statistics\ Topic\ i\ in\ the\ CAOS\ test \\
p_{ijl} &: proportion\ of\ correct\ answers\ in\ the\ Statistics\ Topic\ i\ in\ the\ replicate\ l \\
&\quad from\ Statistics\ Course\ j\ in\ the\ CAOS\ test
\end{aligned}
\tag{8}
$$

Table 19: Description of the Index of Model 3

Index	Value	Description
i	1	Bivariate Data
	2	Confidence Interval
	3	Data Collection and Design
	4	Descriptive Statistics
	5	Graphical Representation
	6	Probability
	7	Sampling Distribution and Variability
	8	Test of Significance
j	1	Additional Statistics Course
	2	Only ESMA 3016
l	1	S2 if j=1
	2	S3 if j=1
	1	S7 if j=2
	2	S4 if j=2
	3	S5 if j=2
	4	S6 if j=2
	5	S1 if j=2

Table 20: Description of Parameters and Response Variable Model 3

Term	Description
p_{ijk}	proportion of correct answer in the Statistics Topic level i of subject k in Population j
θ_0	Intercept
α_i	effect of Statistics Topic i in the proportion of correct answers
δ_j	effect of Statistics Course j in the proportion of correct answers
$(\alpha\delta)_{ij}$	interaction effect between Statistics Topic i and Statistics Course j in the proportion of correct answers

$$\begin{array}{lll} H_{0g} : \alpha_i = 0 & H_{0h} : \delta_j = 0 & H_{0i} : (\alpha\delta)_{ij} = 0 \\ H_{1g} : at\ least\ one\ \alpha_i \neq 0 & H_{1h} : at\ least\ one\ \delta_j \neq 0 & H_{1i} : at\ least\ one\ (\alpha\delta)_{ij} \neq 0 \end{array} \quad (9)$$

The analysis was conducted using SAS statistical sofware. Looking at Figure 33, the quotient X^2/DF = 1.07 indicated the model fits the data well since it tis close to 1. Moving on to Figure 34, the three hypothesis from Equation (9) has been tested. From the three hypothesis, the hypothesis associated with the Statistics Topic and Statistics Course factors were significant at a significance level of 0.15. As a result, we can conclude that the both Statistics Topic and Statistics Course factors influenced in the proportion of correct answers in the CAOS test. The result from the Statistics Topic was the same as in the analysis of Model 1. In terms of the Statistics Course factor, if we look at Figure 35, the β parameter estimated for the subjects with additional course is -0.5465 and a p value of 0.0865. This means that subjects with addtional statistics course performed worst than subjects wih only ESMA 3016.

Model Information	
Data Set	WORK.STAT_TOPIC_LOGIC
Response Variable (Events)	Correct
Response Variable (Trials)	Trials
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Figure 32: Information of Model 3

Fit Statistics	
-2 Log Likelihood	142.65
AIC (smaller is better)	174.65
AICC (smaller is better)	188.60
BIC (smaller is better)	207.06
CAIC (smaller is better)	223.06
HQIC (smaller is better)	187.21
Pearson Chi-Square	42.62
Pearson Chi-Square / DF	1.07

Figure 33: Fit Statistics Model 3

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Topic	7	40	1.73	0.1285
Course	1	40	3.14	0.0838
Topic*Course	7	40	0.72	0.6531

Figure 34: F test for Model 3

Parameter Estimates							
Effect	Topic	Course	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			-583E-18	0.5774	40	-0.00	1.0000
Topic	Bivariate data		1.6094	1.2383	40	1.30	0.2011
Topic	Confidence Interval		-0.5108	0.9309	40	-0.55	0.5863
Topic	Data Collection and Design		-1.9459	1.2150	40	-1.60	0.1171
Topic	Descriptive Statistics		0.4520	0.7531	40	0.60	0.5518
Topic	Graphical Representation		0.2877	0.7906	40	0.36	0.7179
Topic	Probability		1.0986	1.2910	40	0.85	0.3998
Topic	Sampling Distribution and Variability		8.94E-16	0.8563	40	0.00	1.0000
Topic	Test of Significance		0	.	.	.	.
Course		Additional Course	-0.5465	0.6906	40	-0.79	0.4334
Course		Only ESMA 3016	0	.	.	.	.
Topic*Course	Bivariate data	Additional Course	-0.6574	1.3981	40	-0.47	0.6407
Topic*Course	Bivariate data	Only ESMA 3016	0	.	.	.	.
Topic*Course	Confidence Interval	Additional Course	1.0574	1.1001	40	0.96	0.3422
Topic*Course	Confidence Interval	Only ESMA 3016	0	.	.	.	.
Topic*Course	Data Collection and Design	Additional Course	1.3938	1.3735	40	1.01	0.3163
Topic*Course	Data Collection and Design	Only ESMA 3016	0	.	.	.	.
Topic*Course	Descriptive Statistics	Additional Course	-0.4044	0.8973	40	-0.45	0.6546
Topic*Course	Descriptive Statistics	Only ESMA 3016	0	.	.	.	.
Topic*Course	Graphical Representation	Additional Course	-0.3917	0.9462	40	-0.41	0.6811
Topic*Course	Graphical Representation	Only ESMA 3016	0	.	.	.	.
Topic*Course	Probability	Additional Course	-0.5521	1.4867	40	-0.37	0.7123
Topic*Course	Probability	Only ESMA 3016	0	.	.	.	.
Topic*Course	Sampling Distribution and Variability	Additional Course	-0.8398	1.0615	40	-0.79	0.4336
Topic*Course	Sampling Distribution and Variability	Only ESMA 3016	0	.	.	.	.
Topic*Course	Test of Significance	Additional Course	0	.	.	.	.
Topic*Course	Test of Significance	Only ESMA 3016	0	.	.	.	.

Figure 35: Fixed Effects Parameter Estimated

Statistics Course and Statistics Phase

Lastly, we will perform the same analysis described above but with the *Statistics Phase* factor mentioned in Table 6 instead of the *Statistics Topic* factor. Since we have examinated these factors isolated in the previous analysis, we will go straight to analyzing both factors at the same time. Figures 36 and 37 show the proportion mean of correct answers in each Statistics Course for each Statistics Phase and viceversa, respectively. Again we see, both level of Statistics Course struggle the most in the phase of Collect Data and a little of interaction in the Analyze Data and Interpret Results phases and Statistics Course.

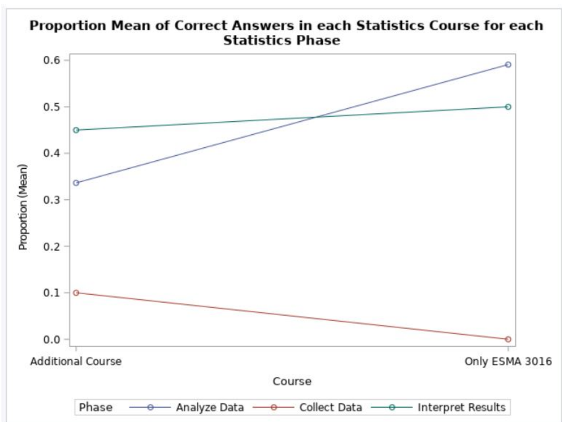


Figure 36: Proportion Mean of Correct Answers in each Statistics Course for each Statistics Phase

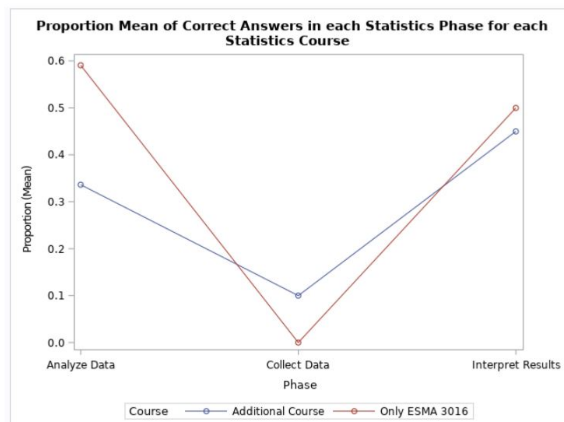


Figure 37: Proportion Mean of Correct Answers in each Statistics Phase for each Statistics Course

The data, model, parameter explanation and hypothesis to be tested are shown in Table 21, Equation (10), Table 23 and Equation (11) respectively. The random effect of the subjects was not included in the model due to the problem of convergence.

Table 21: Statistics Course and Statistics Phase Factors

		Population						
		Prospective				Practicing		
Statistics Phase	**Item**	S_1	S_2	S_3	S_7	S_4	S_5	S_6
Collect Data	Correct	0	0	0	1	0	0	0
	Total	2	2	2	2	2	2	2
Analyze Data	Correct	8	11	15	4	12	8	5
	Total	22	22	22	22	22	22	22
Interpret Results	Correct	7	8	8	10	5	5	9
	Total	16	16	16	16	16	16	16

$$
\begin{aligned}
&logit(p_{ijl}) = \zeta_0 + \gamma_i + \delta_j + (\gamma\delta)_{ij} \\
&\quad Y_{ijl} \sim Binomial(n_i, p_{ijl}) \\
&\quad Y_{ijl} : number\ of\ correct\ answers\ in\ the\ Statistics\ Phase\ i\ in\ the\ replicate\ l \\
&\qquad from\ Statistics\ Course\ j\ in\ the\ CAOS\ test \\
&\qquad n_i : number\ of\ items\ in\ the\ Statistics\ Phase\ i\ in\ the\ CAOS\ test \\
&\quad p_{ijl} : proportion\ of\ correct\ answers\ in\ the\ Statistics\ Phase\ i\ in\ the\ replicate\ l \\
&\qquad from\ Statistics\ Course\ j\ in\ the\ CAOS\ test
\end{aligned} \tag{10}
$$

Table 22: Description of the Index of Model 4

Index	**Value**	**Description**
i	1	Collect Data
	2	Analyze Data
	3	Interpret Results
j	1	Additional Statistics Course
	2	Only ESMA 3016
l	1	S2 if j=1
	2	S3 if j=1
	1	S7 if j=2
	2	S4 if j=2
	3	S5 if j=2
	4	S6 if j=2
	5	S1 if j=2

$$
\begin{array}{lll}
H_{0j} : \gamma_i = 0 & H_{0k} : \delta_j = 0 & H_{0t} : (\gamma\delta)_{ij} = 0 \\
H_{1j} : at\ least\ one\ \gamma_i \neq 0 & H_{1k} : at\ least\ one\ \delta_j \neq 0 & H_{1t} : at\ least\ one\ (\gamma\delta)_{ij} \neq 0
\end{array} \tag{11}
$$

Table 23: Description of Parameters and Response Variable

Term	Description
p_{ijk}	proportion of correct answer in the Statistics Phase i of observation k in Statistics Course j
ζ_0	Intercept
γ_i	effect of Statistics Phase i in the proportion of correct answers in the CAOS test
δ_j	effect of Statistics Course j in the proportion of correct answers in the CAOS test
$(\gamma\delta)_{ij}$	interaction effect between Statistics Phase i and Statistics Course j in the proportion of correct answers in the CAOS test

The analysis was conducted using procedure glimmix in SAS software. Looking at Figure 39 the quotient X^2/DF =1.28 indicated the model fit the data well since it is almost 1. Finally in Figure 40, the three hypothesis from Equation (11) has been tested. Since the three p values are greather than 0.15, we can say that at a significance level of 0.15, there is not sufficient evidence to reject the three H_0. As a result, we can conclude that neither of the factors nor the interaction influence in the proportion of correct answers in the CAOS test.

Model Information	
Data Set	WORK.STAT_PHASE_LOGIC
Response Variable (Events)	Correct
Response Variable (Trials)	Trials
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Figure 38: Information Model 4

Fit Statistics	
-2 Log Likelihood	66.23
AIC (smaller is better)	78.23
AICC (smaller is better)	84.23
BIC (smaller is better)	84.50
CAIC (smaller is better)	90.50
HQIC (smaller is better)	79.59
Pearson Chi-Square	19.18
Pearson Chi-Square / DF	1.28

Figure 39: Fits Statistics Model 4

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Phase	2	15	0.02	0.9792
Course	1	15	0.00	0.9766
Phase*Course	2	15	1.16	0.3413

Figure 40: F test for Model 4

Parameter Estimates							
Effect	Phase	Course	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			0	0.3542	15	0.00	1.0000
Phase	Analyze Data		0.3677	0.4680	15	0.79	0.4442
Phase	Collect Data		-12.9024	316.73	15	-0.04	0.9680
Phase	Interpret Results		0	.	.	.	.
Course		Additional Course	-0.2007	0.4189	15	-0.48	0.6388
Course		Only ESMA 3016	0	.	.	.	.
Phase*Course	Analyze Data	Additional Course	-0.8466	0.5570	15	-1.52	0.1493
Phase*Course	Analyze Data	Only ESMA 3016	0	.	.	.	.
Phase*Course	Collect Data	Additional Course	10.9058	316.73	15	0.03	0.9730
Phase*Course	Collect Data	Only ESMA 3016	0	.	.	.	.
Phase*Course	Interpret Results	Additional Course	0	.	.	.	.
Phase*Course	Interpret Results	Only ESMA 3016	0	.	.	.	.

Figure 41: Fixed Effects Parameter Estimated Model 4

In the following section, we will examine the CAOS test items individually and identify which items the subjects struggle most.

CAOS Test Items

In this section, we want to identify which items in the CAOS Test, the subjects had the most struggles with. Since there are seven subjects, the maximum number of times each item was answer correctly is seven. For the purpose of this study, we will define an item to be diffcult or that subjects struggled with if less than 50% of the subjects answered correctly. In other words, if number of subjects that answered correctly a specific item is less than 3.5, then that item is categorize as an item the subjects struggle with. Table 24 shows the number of times each CAOS test item was answered correctly.

Table 24: Number of Times each Caos test was Answered Correctly

Item	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
Number of times Answered Correctly	5	5	2	3	3	0	1	3
Item	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16
Number of times Answered Correctly	0	0	6	6	4	1	4	1
Item	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Number of times Answered Correctly	3	4	2	6	6	3	5	2
Item	Item 25	Item 26	Item 27	Item 28	Item 29	Item 30	Item 31	Item 32
Number of times Answered Correctly	1	4	1	4	1	3	5	1
Item	Item 33	Item 34	Item 35	Item 36	Item 37	Item 38	Item 39	Item 40
Number of times Answered Correctly	2	3	2	5	3	0	2	4

In Table 25, each CAOS item is categorized into Dominate or Struggle. There are 15 items in the Dominate category and 25 items in the Difficult category. There was not an item in which all the subjects answered correctly. However, in items 6, 9, 10 and 38, none of the subjects answered correctly.

Table 25: Number of Times each Caos test Item was Answered Correctly

Dominate		Struggle	
Item	**Number of Times Answered Correctly**	**Item**	**Number of Times Answered Correctly**
11	6	4	3
12	6	5	3
20	6	8	3
21	6	17	3
1	5	22	3
2	5	30	3
23	5	34	3
31	5	37	3
36	5	3	2
13	4	19	2
15	4	24	2
18	4	33	2
26	4	35	2
28	4	39	2
40	4	7	1
		14	1
		16	1
		25	1
		27	1
		29	1
		32	1
		6	0
		9	0
		10	0
		38	0

Now, is in out best interest to identify which of the items were classified as Dominate and Struggle for each Statistics Topic. Figures 42 and 43 shows for each Statistics Topic, which item was classify as Dominate and Struggle. As we can see, all the items in Data Collection and Design and Sampling Distribution and Variability were classify as Struggle. This validates the results from the Population and Statistics Topic analysis that indicated that Data Collection and Design and Sampling Distribution and Variability were the topics the subjects struggle the most with. On the other hand, topics like Bivariate Data and Descriptive Statistics were the topics that subjects dominated most.

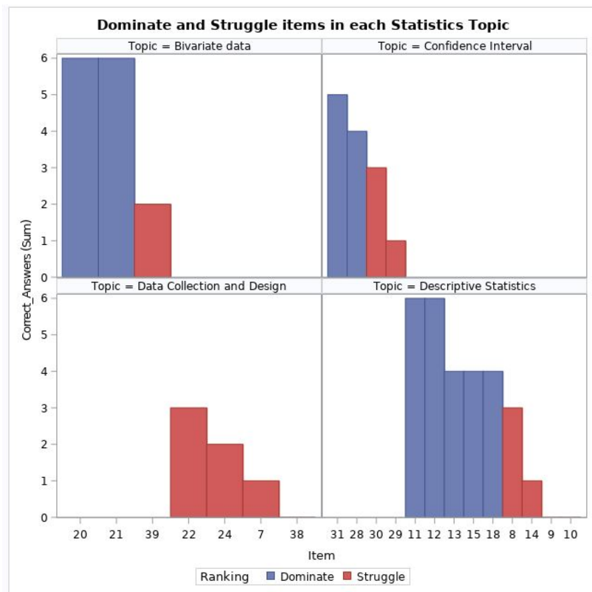


Figure 42: CAOS Items Classification in each Statistics Topic 1

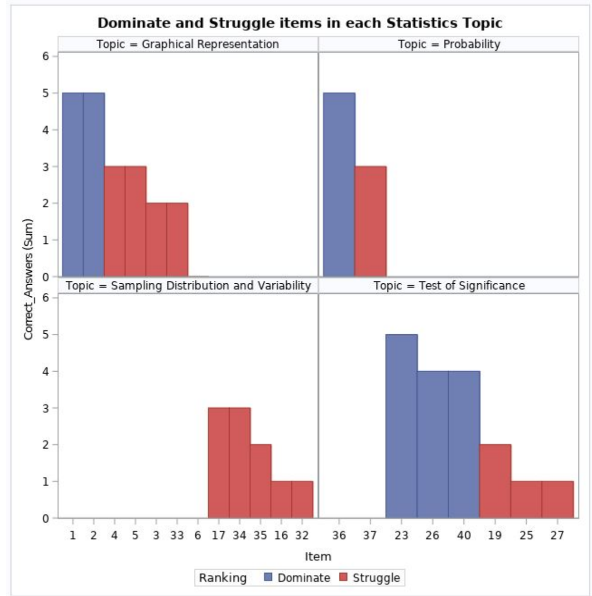


Figure 43: CAOS Items Classification in each Statistics Topic 2

Conducting the same analysis but with the Statistics Phase, we obtained the results in Figures 44 and 45. Analyze Data and Interpret Results have very similar results. However, in Collect Data, there are only two items and in item 7, only one subject answered correctly and in item 38, none of the subjects answered correctly.

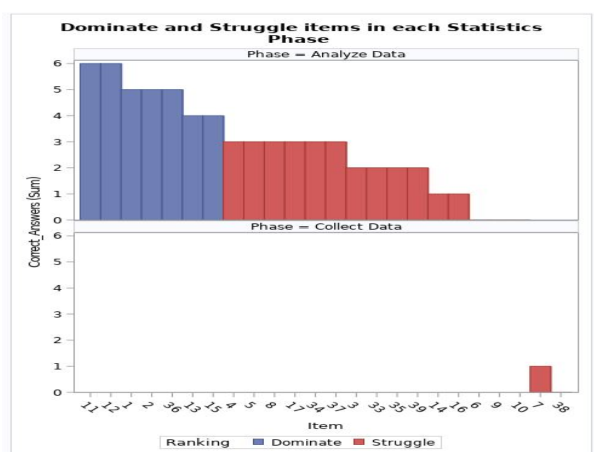


Figure 44: CAOS Items Classification in each Statistics Phase 1

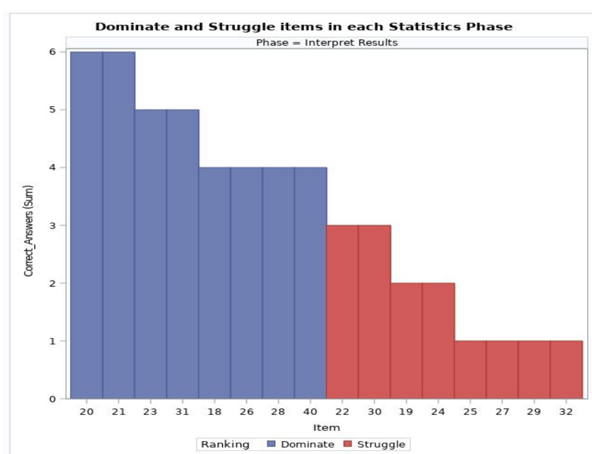


Figure 45: CAOS Items Classification in each Statistics Phase 2

In the following subsection, we will show the feedbacks and recommendations each suject gave to improve the mathematics education program in the preparation of mathematics teachers to teach statistics at secondary level.

Feedbacks and Recomendations

In addition to taking the CAOS test, each subject also provided feedbacks and recommendations on how to improve the effectiveness of the mathematics education program at UPRM. Specifically on how to prepare better future mathematics teachers to teach statistics at the secondary level. The first question asked was if they believed the statistics courses taken in the bachelors degree prepared them to teach statistics using the following scale: Completely Disagree, Disagree, Neutral, Agree and Completely Agree. A summary of the answer of each subject is provided in Figure 46. Observing, five of the seven subjects disagreed on the preparedness provided by the mathematics education program to teach statistics at the secondary level.

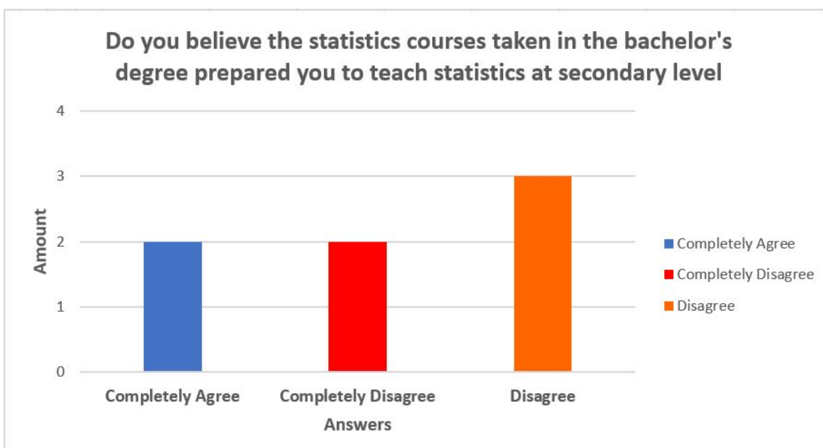


Figure 46: Question #1

The second question was for the subjects to provide recommendations on how to improve the effectiveness of the mathematics education program to prepare future mathematics teacher to teach statistics at the secondary level. Table 26 specify the pre determined recommendations available for the subjects to choose. Each subject had the opportunity to select more than one option. Figure 47 shows the recomendations provided by each subject and Table 27 provides a summary of the recommendations by the subjects. It can be observed that most of the subjects (71%) recommended to add a pedagogical course focused on the teaching of statistics and modifying the focus or the topics that are being taught ESMA 3016 is being taught and 57% of the subjects recommended adding more advanced statistics courses to the program.

Table 26: Recommendations for the Mathematics Education Porgram

Question	Options
Recommendations	Add a pedagogical course focused on the teaching of statistics
	Add a supervised practice in teaching statistics
	Add more advanced statistics courses in addition to ESMA 3016
	Add one hour of laboratory to the ESMA 3016 course
	Modify the topics or the focus in which ESMA 3016 is being taught
	The program does not require improvement

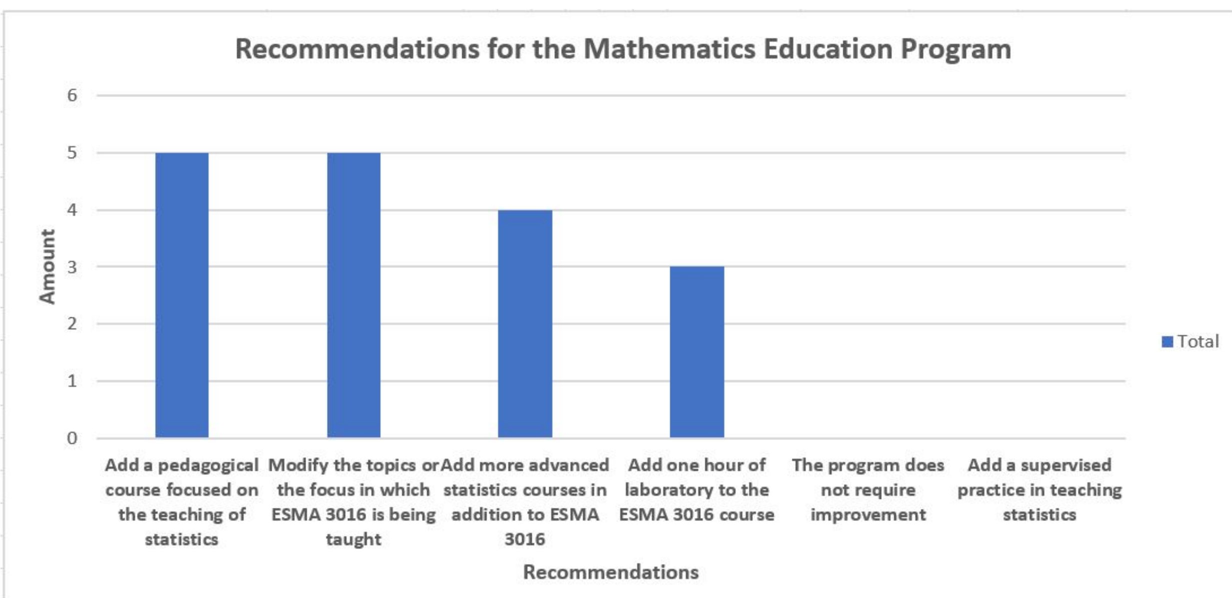


Figure 47: Recommendations for the Mathematics Education Program

Table 27: Summary of Recommendations for the Mathematics Education Porgram

Recommendations	**Number of times selected**
Add a pedagogical course focused on the teaching of statistics	5
Modify the topics or the focus in which ESMA 3016 is being taught	5
Add more advanced statistics courses in addition to ESMA 3016	4
Add one hour of laboratory to the ESMA 3016 course	3
Add a supervised practice in teaching statistics	0
The program does not require improvement	0

Lastly, each subject had the opportunity to provide a free written feedbacks and opinions about the statistics course they took as part of the program, ESMA 3016 course. Table 28 shows each subject complete feedback and opinion. There responses were provided in spanish by each subject and were translated in english. From all the responses, we can see that there is a strong argument that the primary focus of ESMA 3016 is programming instead of developing conceptual understanding of the statistical topics they have to learn. It gives the impression that ESMA 3016 is being taught to future statistician rather than to future mathematics teachers. It is important to clarify that the subjects did not provided a complete detail of what they meant for the word programming and we did not ask the professors that taught this course how they are incorporating programming in the course.

Table 28: Response for Open Question of each Subject

Subjects	**Answer**
Subject 1	The focus of ESMA 3016 is to create programs to solve statistics problems but that is not the focus of a teacher. The teacher wants to predict and study the results to make decisions to improve the learning process of the students. Also the teacher has to take a theoretical statistics course to learn all the topics that are required to teach in high school. The course of ESMA 3016 is not enough to prepare teachers to teach statistics at secondary level.
Subject 2	I understand that the course is much more focused on programming than on the application of what has been learned.
Subject 3	Certainly, we need more courses in statistics. The course is not complicated, but it is easy to forget if you do not practice it frequently. Sometimes, I think ESMA 3016 course focuses more on programming than statistics.
Subject 4	Many times elementary courses such as ESMA 3016 are taken at the beginning of the bachelor's degree and by the end of the program, the topics are forgotten. That is why I propose an advanced statistics course at the end of the program.
Subject 5	Additional statistics course is needed. The laboratory is practical and useful if you are interested in being a statistician but if you are going to be a teacher, you need a different focus.
Subject 6	With a laboratory hour, the student can practices and creates exercises helping to create a deep understanding of the topics.
Subject 7	By taking the course I actually felt that I did not learn much. Mostly because of the mix between trying to learn statistics and programming at the same time. Having difficulties with programming made it difficult for me to work and understand the statistics course topics.

Chapter 5 Discussion

The purpose of this study was to answer the following questions:

1. To what extended is the curriculum at UPRM preparing future teachers to teach statistics and probability at the secondary level?
2. What recommendations can be offered to the Department of Mathematical Sciences in order to improve the mathematics education program to prepare teachers to teach statistics at the secondary level?

To answer the first question, the CAOS 4 test was used to measure the conceptual statistical knowledge of the seven subjects. In the overall score in the CAOS test, the seven subjects had an average score of 0.41 with a 95% Confidence Interval for the population CAOS score mean between 0.33 and 0.50, indicating a poor performance in the CAOS test. This result agrees with the results from Hannigan et al. (2013), Fabrizio et al. (2011) and Tintle et al. (2012) in which the average scores for the subjects in their study were 0.50, 0.42 and 0.44 respectively.

Looking in the Statistics Topics, this factor was significant in the Type lll test at a 0.15 significance level with a p value of 0.1047. Futher analysis showed that topics like Sampling Distribution and Variability, and Data Collection and Design were the topics in which the subjects struggled the most and Bivariate data was the topic that the subjects dominated the most. The struggles in Data Collection and Design aligns with results shown in Hannigan et al. (2013); while in Sampling Distribution and Variability aligns with results shown in Hannigan et al. (2013), Fabrizio et al. (2011) and Lovett et al. (2016). Probability was a topic that subjects struggled with in Hannigan et al. (2013) but in this study, it was the second highest proportion of correct answers. In terms of the Statistics Phase mentioned in Franklin et al. (2007), the factor was not significant at a significance level of 0.15.

Futhermore, there was not a significance difference between the prospective and practicing teachers although there was significance difference between subjects who took addtional statistics courses and subjects who only had taken ESMA 3016 as a statistics course. Surprisingly, the subjects with additonal statistics course performed worse than subjects with only ESMA 3016.

In terms of the CAOS items individually, in 15 of the items more than half of the subjects answered correctly and in 25 items less than half of the subjects answered correctly. None of the items were answered correctly by all seven subjects. The items in which none of the subjects answered correctly were item 6,9,10 and 38 and the items in which they performed really well were item 11, 12, 20, 21,1,2,23,31 and 36.

On the other hand, answering the second question, five of the seven subjects agreed the mathematics education program at UPRM does not prepares future mathematics teachers effectively to teach statistics at the secondary level. In addition, 71% of the subjects recommended to add a pedagogical course focused on the teaching of statistics to the program and to add more advanced statistics courses to the program and 57% of them recommended to modify the topics or the focus in which ESMA 3016 is being taught. Finally, in the open write question, most of the subjects agreed that the focus of ESMA 3016 is more oriented to programming instead of developing conceptual understanding of the statistics topics.

Chapter 6 Limitations

In this study, although we gained important information regarding the mathematics education program, it had some limitations. First of all, we had very few subjects in the study (n=7) and this affected the power of the overall tests in the binomial regression analysis. With a very small sample, it is difficult to detect significance differences in the factors and interactions. Also due to the small sample, we could not evaluate the factor of subjects with teaching experience statistics since only one of the subjects have experience teaching statistics. Moreover, in the analysis of the CAOS item individually, it did not include the use of the Item Response Model since this model works with large sample size. The purpose of this model is to estimates the ability of the subject to answer the items correctly and the difficulty of the items.

In terms of the instrument used to evaluate the statistical knowledge, the CAOS test is a very good instrument, especially if we want to examine how the subjects performs by statistics topics. However if we are interested in the statistical phases, there is a big discrepency in the number of items in each statistical phase especially in the Collect Data phase compare to the rest of the phases. In addition, only three statistical phase were evaluated since the CAOS test does not have items to evaluate the phase of Formulate a Question.

Chapter 7 Future Work

For future work, data can be collected on more subjects to improve the power of the analysis. In addition, more factors could be incorporated into the analysis such as the factor of statistics teaching experience. Futhermore, we could incorporate an analysis using the Item Response Model. Also, we can try differents instruments such as Level Of Conceptual Understanding in Statstics Assesment (LOCUS) when working with the Statistical Investigation Cycle Phases. In addition, we can evaluate more pedagogical aspects such as statistics teaching efficacy using instruments like Self Efficacy to Teach Statistics (SETS) used in Lovett et al. (2016). Finally, this study can be expanded by incorporating all the mathematics education programs of Puerto Rico to evaluate how prepared the prospective and practicing mathematics teachers from Puerto Rico are to teach statistics at secondary level.

Chapter 8 Conclusion

Statistics has been treated as a branch of mathematics for many years. However, the reality is that statistics differs from mathematics in many aspects such as role of context, incorporating variability, reasoning, precision and the importance of data collection (Rossman et al., 2006; Scheaffer, 2006; Gattuso, 2008). Studies have shown that prospective and practicing mathematics teachers lack of statistical knowledge and as result are not well prepared to teach statistics in school (Burrill & Biehler, 2011; Gattuso, 2008; Makar & Fielding-Wells, 2011; Franklin et al., 2007). All seven subjects in this study took mathematics topics such as Arithmetic, Geometry and Algebra in school. Five of them took Pre Calculus and only three of them took statistics in school. This ilustrates that students in schools are not receiving the statistics education required. In evaluating the statistical knowledge of the prospective and practicing mathematics teachers from the mathematics education program at UPRM, we noticed a poor performance from them in the CAOS 4 test; giving the impression they lack of the statistical knowledge required to teach statistics at secondary level.

The recommendations and feedbacks from the subjects are a great place to start. To begin with, the course ESMA 3016 should be a course focused on developing conceptual knowledge in statistics instead of programming. The excessive focus in programming is affecting the learning experiences of the subjects since they have to learn the statistics topics and how to program at the same time. It seems ESMA 3016 is focusing on students who are pursuing a mathematics or statistics degree, but for mathematics education, the focus should be modify. Since this course is in both, the mathematics education program and in the computer science program, the students of the computer science program are benefiting from the programming but the students from the mathematics education program are struggling with the programming part. Another useful recommendation was to add a pedagogical course focusing on the teaching and learning of statistics since it is importance not only to have statistical knowledge but to know how to teach it. Lastly, perhaps adding additional statistics course is not as necessary since ESMA 3016 covers all the topics listed in the PR content standards. However it might be a good idea to split ESMA 3016 into two different courses so each topic can be taught deeper and better and to separate the students from the mathematics education program and the computer science program in different course sections.

Statistics education is an area that is growing every day since the analysis of data is playing an important role in our society. Each day data is being collected and used to extract information and finding solutions to real problems. As result it is important to have mathematics teachers with proper statistics knowledge and pedagogical statistical knowledge. This was a preliminary study to examine how the mathematics education program at UPRM is preparing future mathematics teachers to teach statistics. From the results in this study, the program has a lot of room to grow. It is important to recognize it needs to be improved and to identify those areas that definitively needs improvements.

References

Bansilal, S. (2014). Using an apos framework to understand teachers responses to questions on the normal distribution. *Statistical Education Research Journal*, *13*(2), 42-57.

Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint icmi/iase study: the 18th icmi study* (Vol. 14). Springer Science & Business Media.

Borim, C., & Coutinho, C. (2008). Reasoning about variation of a univariate distribution: a study with secondary mathematics teachers. *Joint ICMI/IASE study: teaching statistics in school mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study*, *18*.

Browning, C., Goss, J., & Smith, D. (2014). Statistical knowledge for teaching: Elementary preservice teachers. In *Sustainability in statistics education. proceedings of the ninth international conference on teaching statistics (icots9*.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 57–69). Springer.

Casey, S. A., & Wasserman, N. H. (2015). Teachers knowledge about informal line of best fit. *Statistical Education Research Journal*, *14*(1), 8-35.

DelMas, G., Joan, G., Ooms, A., & Chance, B. (2007). Assessing students'conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, *6*(2).

Department of Education of Puerto Rico. (2014). *Puerto rico core standards for mathematics programs.*

Dobson, A., & Barnett, A. (2008). *An introduction to generalized linear models* (3rd ed.). Chapman; Hall/CRC Press.

Fabrizio, M., López, M. V., & Plencovich, M. C. (2011). Statistics in teacher training colleges in buenos aires, argentina: Assessment and challenges. *Proceedings of the 56th Session of* *the International Statistics Institute. Lisbon: Portugal. http://www.stat.auckland.ac.nz/* iase/publications/isi56/CPM80_Fabrizio.pdf. Accessed*, *30*.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer, R. (2007). Guidelines for assessment and instruction in statistics education (gaise) report: A pre-k-12 curriculum framework. *American Statistical Association*.

Gattuso, L. (2008). Mathematics in a statistical context. *C. Batanero, G. Burrill, C. Reading, & A. Rossman*.

González, M., & Pinto, J. (2008). Conceptions of four pre-service teachers on graphical representation. *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study*, *18*.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, *38*(5), 427-437.

Groth, R. E. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning*, *15*(2), 121-145.

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median and mode. *Mathematical Thinking and Learning*, *8*(1), 37-63.

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education*, *16*(6), 427–449.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372-400.

Hobden, S. (2014). When statistical literacy really matters: Understanding published information about the hiv/aids epidemic in south africa. *Statistical Education Research Journal*, *13*(2), 72-82.

Jacobbe, T. (2012). Elementary school teachers understanding of the mean and median. *Interna-*

tional Journal of Science and Mathematics Education, *10*(5), 1143-1161.

Jacobbe, T., & Horton, R. M. (2010). Elementary school teachers comprehension of data displays. *Statistical Education Research Journal*, *9*(1), 27-45.

Jacobbe, T., Whitaker, D., Case, C., & Foti, S. (2014). The locus assessment at the college level: conceptual understanding in introductory statistics. In *Sustainability in statistics education. proceedings of the ninth international conference on teaching statistics (icots9), flagstaff, arizona, usa. voorburg: International association of statistics education.*

Jones, D., & Tarr, J. (2010). Recommendations for statistics and probability in school mathematics over the past century. *Mathematics curriculum issues, trends, and future directions: 72nd yearbook*, 65-75.

Kataoka, V., da Silva, C., & Cazorla, I. (2014a). Analysis of teachers understanding of covariation in the vitruvian man context. In *Sustainability in statistics education. proceedings of the ninth international conference on teaching statistics (icots9).*

Kataoka, V., da Silva, C., & Cazorla, I. (2014b). Analysis of teachers understanding of variation in the dot-boxplot context. In *Sustainability in statistics education. proceedings of the ninth international conference on teaching statistics (icots9).*

Leavy, A. (2006). Using data comparison to support a focus on distribution: examining preservice teachers understandings of distribution when engaged in statistical inquiry. *Statistical Education Research Journal*, *5*(2), 89-114.

Leavy, A. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistical Education Research Journal*, *9*(1), 46-67.

Lovett, J. N., et al. (2016). The preparedness of preservice secondary mathematics teachers to teach statistics: A cross-institutional mixed method study.

Makar, K., & Confrey, J. (2004). Secondary teachers statistical reasoning in comparing two groups. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353–373). Springer.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In

Teaching statistics in school mathematics-challenges for teaching and teacher education (pp. 347–358). Springer.

Rossman, A., Chance, B., Medina, E., & Obispo, C. (2006). Some key comparisons between statistics and mathematics, and why teachers should care. *Thinking and reasoning with data and chance: Sixty-eighth annual yearbook of the National Council of Teachers of Mathematics*, 323–333.

Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. *Thinking and reasoning with data and chance*, *68*, 309–321.

Sorto, M. A. (2006). Identifying content knowledge for teaching statistics. In *Working cooperatively in statistics education: Proceedings of the seventh international conference on teaching statistics, salvador, brazil.[cdrom]. voorburg, the netherlands: International statistical institute.[online: http://www. stat. auckland. ac. nz/˜ iase/publications/17/c130. pdf].*

Tintle, N. L., Topliff, K., VanderStoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, *11*(1), 21.

Wessels, H., & Nieuwoudt, H. (2011). Teachers' professional development needs in data handling and probability. *Pythagoras*, *32*(1), 1–9.

Wild, C. J., & Phannkuch, M. (1999). Statistical thinking in empirical inquiry. *International Statistical Review*, *67*(3), 223-248.

Appendix

Chapter A Mathematics Education Program Curricula

Universidad de Puerto Rico
Recinto de Mayagüez
Departamento de Ciencias Matemáticas

CURRÍCULO EN EDUCACIÓN MATEMÁTICA (1222)

PRIMER AÑO

PRIMER SEMESTRE	Pre-Requisito	SEGUNDO SEMESTRE
MATE 3005 (5) Precálculo I y II (730 College Board Aprovechamiento Matemático)	*MATE 3005 o MATE 3172*	**MATE 3031* (4)** Cálculo I
EDFU 3001 (3) Crecimiento y Desarrollo Humano I	*EDFU 3001*	**EDFU 3002 (3)** Crecimiento y Desarrollo Humano II
CIBI 3031 (3) Ciencias Biológicas I	*CIBI 3031*	**CIBI 3032 (3)** Ciencias Biológicas II
ESPA 3101 (3) Español Básico I	*ESPA 3101*	**ESPA 3102 (3)** Español Básico II
INGL 3--- (3) Inglés		**INGL 3--- (3)** Inglés
Total de créditos: 17		**Total de créditos: 16**

SEGUNDO AÑO

Pre-Requisito	PRIMER SEMESTRE	Pre-Requisito	SEGUNDO SEMESTRE
*MATE 3031**	**MATE 3032* (4)** Cálculo II	*MATE 3032**	**MATE 3063* (3)** Cálculo III
*MATE 3031**	**MATE 3020* (3)** Int. Fundamentos Matemática	*MATE 3032**	**MATE 3030* (3)** Int. Geométrica
	HIST 3241 O 3242* (3) Historia de Puerto Rico	*MATE 3171 o MATE 3005*	**COMP 3010* (3)** Int. Programación Computadora I
	ESPA 3--- (3) Español		**ESPA 3--- (3)** Español
	ELECTIVAS EN CIENCIAS (4)		**ELECTIVAS EN CIENCIAS (4)**
	EDFI ---- (1) Educación Física		**EDFI ---- (1)** Educación Física
	Total de créditos: 18		**Total de créditos: 17**

TERCER AÑO

Pre-Requisito	PRIMER SEMESTRE	Pre-Requisito	SEGUNDO SEMESTRE
*MATE 3032**	**MATE 4031* (3)** Álgebra Lineal	*MATE 3020**	**MATE 4008* (3)** Int. Estructuras Algebraicas
*MATE 3032**	**MATE 3040* (3)** Teoría de Números	*MATE 3031**	**MATE 3181* (3)** Matemática Discreta
MATE 3031 y COMP 3010**	**ESMA 3016* (3)** Análisis Estadístico de Datos		**EDFU 4019 (3)** Fund. Filosóficos de la Educación
	EDFU 3007 (3) Fund. Sociales de la Educación		**EDPE 3129* (3)** Uso Microcomputadoras Salón Clase
	HIST 3111 O 3112 (3) Historia de los Estados Unidos		**INGL 3--- (3)** Inglés
	INGL 3--- (3) Inglés		**ELECTIVA LIBRE (3)**
	Total de créditos: 18		**Total de créditos: 18**

CUARTO AÑO

Pre-Requisito	PRIMER SEMESTRE	Pre-Requisito	SEGUNDO SEMESTRE
DIR	**EDPE 4145 (3)** Teoría y Metodología Ense. Mate	*DIR y EDPE 4145*	**EDPE 4146 (6)** Práctica de la Enseñanza Mate
*MATE 3020**	**MATE 4023* (3)** Educación Matemática I	*MATE 4023* y EDPE 3129**	**MATE 4039* (2)** Uso de Tecnología Enseñanza Mate
	MATE 4120* (3) Historia Matemáticas	*HUMA 3111*	**HUMA 3112 (3)** Int. Cult. Occidental II
	HUMA 3111 (3) Int. Cultura Occidental I		**ELECTIVA LIBRE (3)**
	EDES 4006 (3) Natu. y Nec. Nino Excepcional		**ELECTIVA LIBRE (3)**
	ELECTIVA LIBRE (3)		
	Total de créditos: 18		**Total de créditos: 16**

Cursos con * son cursos de especialidad

Pre-requisito de EDPE 4145: EDFU3001-3002, EDFU3007, EDFU4019, EDPE3129 Y 18 créditos aprobados en matemática.

IMPORTANTE: Para obtener la Certificación de maestro de matemáticas de nivel secundario del Departamento de Educación de Puerto Rico, este requiere un promedio de **3:00** en **especialidad** y **general**. Se recomienda tomar como **electiva libre** uno de los siguientes cursos en fundamentos de la educación para cumplir los requisitos de la Certificación:

- EDFU 3017 Medición y Evaluación del Aprendizaje
- EDFU 4006 El niño y su ambiente social
- EDFU 3055 Aspectos legales

****** ELECTIVAS EN CIENCIAS ******

*****FÍSICA**

FISI 3151 / 3153 Física Moderna de Colegio I y Laboratorio I
FISI 3152 / 3154 Física Moderna de Colegio II y Laboratorio II
FISI 3171 / 3173 Física I y Laboratorio I
FISI 3172 / 3174 Física II y Laboratorio II

*****QUÍMICA**

QUIM 3131 / 3133 Química I y Laboratorio I
QUIM 3132 / 3134 Química II y Laboratorio II

*****GEOLOGÍA**

GEOL 3025 Ciencias de la Tierra
GEOL 3026 La Vida en el Pasado
GEOL 3027 Aspectos geológicos de las Ciencias Ambientales
GEOL 3045 Geología Planetaria
GEOL 3046 Recursos de la Tierra
GEOL 3055 Cristalografía Morfológica y Óptica
GEOL 3056 Química Cristalina y Geoquímica de Sistemas Minerales

Cursos con * son cursos de especialidad
Pre-requisito de EDPE 4145: EDFU3001-3002, EDFU3007, EDFU4019, EDPE3129 Y 18 créditos aprobados en matemática.

Chapter B ESMA 3016 Course Syllabus

Universidad de Puerto Rico
Recinto de Mayagüez
Facultad de Artes y Ciencias
DEPARTAMENTO DE CIENCIAS MATEMATICAS

Curso: Análisis Estadístico de Datos

Codificación: ESMA 3016

Número de horas/crédito: 3

Prerrequisitos, correquisitos y otros requerimientos: (MATE 3031 ó MATE 3144) y COMP 3010

Información del profesor:

Nombre	
Horas de Oficina	
Oficina	
Ext.	
Dirección Electrónica	

Texto	Statistics Informed Decisions Using Data Tercera Edición (ISBN-13: 978-0-321-57527-2)
Autor	Michael Sullivan III

Descripción del Curso:

Análisis Estadístico de Datos incluyendo estadística descriptiva e inferencial y análisis exploratorio de datos.

Objetivos del Curso: Al final del curso se espera que el estudiante pueda

- Resumir un conjunto de datos y presentarlos en tablas y gráficas
- Aplicar técnicas de Análisis Exploratorio de Datos para analizar un conjunto de datos.
- Calcular medidas estadísticas de centralidad y variabilidad basadas en la muestra tomada
- Establecer la relación entre dos variables cualitativas
- Entender el significado del concepto de correlación para relacionar dos variables cuantitativas
- Establecer una línea de regresión para representar la tendencia de la relación lineal de dos variables cualitativas.
- Determinar las probabilidades de eventos de experimentos aleatorios
- Aplicar herramientas de cálculo diferencial e integral a través del concepto de variables aleatorias para calcular probabilidades de eventos.
- Modelar experimentos aleatorios de acuerdo a los modelos de distribuciones conocidas como la Binomial, la Poisson y la Normal.

- Simular datos que siguen una distribución conocida, haciendo uso de un programa estadístico de computadoras
- Entender el significado del Teorema del Límite Central y la distribución de la media muestral.
- Aplicar los métodos de inferencia estadística tales como prueba de hipótesis e intervalos de confianza que le permitan para sacar conclusiones de la población usando la muestra extraída de ella

Bosquejo de contenido y distribución del tiempo:

LECCION	ARTICULO	TEMAS	EJERCICIOS
1	1.1-1.2	Introducción	P. 11: 1-8, 10-12, 13-36, 45, 47, 55, 56 P. 20: 1-4, 8, 9-16
2	1.3	Muestreo	P. 27: 1-4 P. 36: 1-24
2	1.4,1.5	Fuentes de Error y Diseño de Experimentos	P. 43: 13-24 P. 53: Impares 13 al 23
3		Laboratorio: Muestreo	
4	2.1	Organizando datos Cualitativos	P. 73: 1-8, 10,12,14,15,17,18,26
4	2.2-2.3	Organizando datos Cuantitativos	P. 94: 13,14,18,19,31,35,41,42 P. 106: 16,19,23
5	3.1	Medidas de tendencia central	P. 137: 15,21,27,31,47
5	3.2	Medidas de Dispersión	P. 153: 19,23,31,32,37
5-6	3.4-3.5	Medidas de localización, BoxPlot y Valores Atípicos	P. 173: 11,17,19,21,23,25,28 P. 181: 9,13,15,16,17
7		Laboratorio: Estadística Descriptiva	
8		Laboratorio: Estadística Descriptiva	
9	5.1	Leyes de probabilidad Laboratorio: Probabilidad Empírica	P. 269: 1-3,5-9,22,27-30,33,35,36,47
10		**EXAMEN I** **Jueves, 17 de septiembre**	
9	5.2	Regla de Adición y Complemento	P. 281: 1-4,13-20,26,29,30,39,41
11	5.3	Independencia y regla de Multiplicación	P. 290: 1-6,9-10,13,15,17,22,27
11	5.4	Probabilidad Condicional y Regla de Multiplicación General	P. 298: 1-2,3-8,9,19,22,26,28, 29
12	6.2	La distribución Binomial	P. 354: 36,37,38,46,47

13	6.3	La distribución Poisson	P. 362: 12,13,14,17
14		Laboratorio: Probabilidad	
15		Laboratorio: Probabilidad	
16	7.1- 7.2	Distribuciones Continuas Propiedades de la Distribución Normal y La Distribución Normal Estándar	P. 381: 19-21,23,24,32,33 P. 394: 2-4, 5,7,9,11,13,15,21,25,33,39,43
16	7.3-7.4	Aplicaciones de la Distribución Normal Avaluando normalidad	P. 402: 13,15,17,18,21,25,29 P. 409: 3-8
18		**EXAMEN II** **Jueves, 15 de octubre**	
17	7.5	Usando aproximación normal para la distribución binomial Laboratorio: Continuas, Normal y Binomial	P. 416: 21-23,26
19	8.1-8.2	Distribución de la media de la muestra y de la proporción de la muestra Laboratorio: TLC	P. 438: 3,6,9,10,17,19,21,24,26,30 P. 448: 11,16,17,20
20	9.1	La lógica en construir intervalos de confianza para la media poblacional (desviación estándar conocida)	P. 468: 7-12,23,24,43,47
20-21	9.2	Intervalo de confianza para la media poblacional (desviación estándar desconocida)	P. 483: 9,13,25
21	9.3	Intervalo de confianza para proporciones	P. 493: 19,20,23,28
22		Laboratorio: IC	
23	10.1-10.2	El lenguaje de las Pruebas de Hipótesis , Prueba de hipótesis para la Media poblacional (Desviación estándar conocida)	P. 521: 15-18,20-21 (solo escriba la nula y la alterna) P. 536: 11,13,21,25,26
23-24	10.3	Prueba de hipótesis para la Media poblacional (en la práctica)	P. 547: 5,6,7,19,21
24	10.4	Prueba de hipótesis para una proporción	P. 558: 10,11,13,15
25		Laboratorio: PH	
26		**EXAMEN III** **Martes, 17 de noviembre**	
27	11.1	Inferencia sobre dos medias de muestras dependientes	1,2,5-10,11,13,15

27	11.2	Inferencia sobre dos muestras. Muestras independientes	3,5,9,11,14
27	11.3	Inferencia sobre proporción de dos poblaciones	5,7,9,11,15,19,27,31,33
28	4.1	Diagrama de Dispersión y Correlación	P. 202: 9-12,13,14,23,27,29,31,40
28-29	4.2	Regresión Mínimos Cuadrados	P. 218: 10,21,24,26
29	4.3	Diagnostico en la recta de regresión(Mínimos cuadrados)	P. 232: 15,21,23,25
30		Laboratorio: Regresión	

Estrategias instruccionales:

Conferencias en donde se presentan: los conceptos y métodos fundamentales de la estadística, ejemplos, ejercicios y aplicaciones.

Laboratorios: Se discute el uso de la tecnología y aplicaciones del programa R.

El uso de otras estrategias (tales como uso aprendizaje cooperativo, trabajo en clase, discusión abierta, sesiones abiertas a preguntas, proyectos, etc.) se deja a discreción del profesor.

Recursos de aprendizaje o instalaciones mínimos disponibles o requeridos:

Las clases de este curso se reunirán en el salón M-202A donde se presentará el uso del programa estadístico R.

Política Universitaria

Según se establece en el Catálogo Sub-graduado 2007-2008, (pág. 65): "los estudiantes deben asistir a todos los exámenes. Los estudiantes que se ausenten a un examen por una razón justificada aceptable para el profesor, tomarán reposición del mismo. Si el estudiante no se presenta a la reposición, obtendrá F en dicho examen." Si un estudiante falta a un examen debe comunicarse lo más pronto posible con la profesora y deberá presentar evidencia que justifique su ausencia para poder tomar la reposición que será al final del semestre en una fecha que se le anunciará al estudiante.

El Catálogo Sub-graduado 2007-2008 (pág. 64) indica que: "la asistencia a clases es obligatoria. Las ausencias frecuentes afectan la nota final y la responsabilidad de reponer las tareas o trabajos recae en el estudiante."

Chapter C IRB Protocol Approval

Institutional Review Board
University of Puerto Rico Mayagüez Campus
Dean of Academic Affairs
Call Box 9000
Mayagüez, PR 00681-9000

April 09, 2018

Greetings Kevin Molina,

As a member of the Institutional Review Board of the University of Puerto Rico - Mayagüez Campus, I have considered the Review Application for your project titled "Evaluating the statistical knowledge of the prospective and practice mathematical teachers graduaded" (Protocol num. 20180308001). After an evaluation of your protocol, I have determined that your research qualifies for an Expedited approval.

Remember that any modifications or amendments to the approved protocol or its methodology must be reviewed and approved by the IRB before they are implemented. The IRB must be informed immediately if an adverse event or unexpected problem arises related to the risk to human subjects. The IRB must likewise be notified immediately if any breach of confidentiality occurs.

We appreciate your commitment to uphold the highest standards of human research protections and remain.

Sincerely,

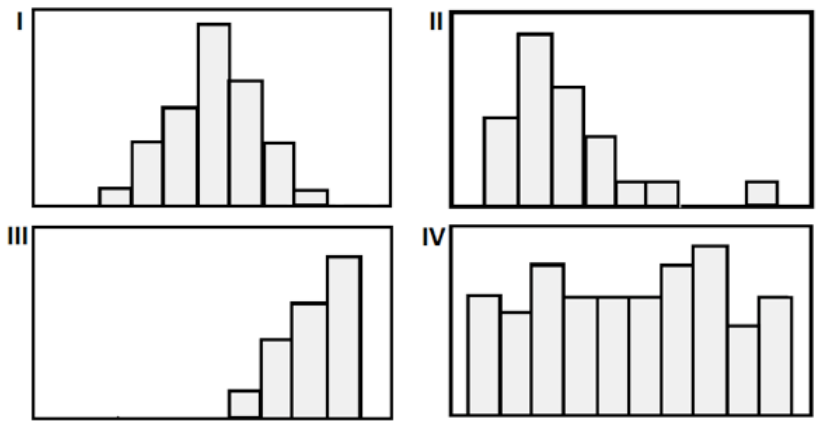
Institutional Review Board (IRB)
University of Puerto Rico,
Mayagüez Campus
Office: Celis 108
Tel.: (787) 832-4040 Ext. 6277
Web Page: http://www.uprm.edu/cpshi/

Chapter D Sample CAOS 4 Test Items

Evaluación Comprensiva de Resultados para un primer curso de Estadística (CAOS 4)

Las preguntas de la 3 a la 5 se refieren a la siguiente situación:

A continuación se muestran 4 histogramas. Para cada pregunta elige el histograma que mejor coincida con la descripción dada.



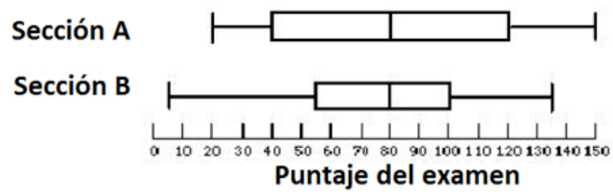
3. La distribución para un conjunto de resultados de un examen, en donde el examen fue muy fácil está representada por:
   a. Histograma I
   b. Histograma II
   c. Histograma III
   d. Histograma IV
4. La distribución para el conjunto de circunferencias de la muñeca de la mano derecha (medidas en centímetros), de una muestra aleatoria de niñas recién nacidas, está representada por:
   a. Histograma I
   b. Histograma II
   c. Histograma III
   d. Histograma IV
5. La distribución para el último dígito de los números telefónicos de una muestra del directorio telefónico (por ejemplo, para el número de teléfono 968-96-67, el último digito 7 sería seleccionado), es representado por:
   a. Histograma I
   b. Histograma II
   c. Histograma III
   d. Histograma IV

7. En un reciente estudio de investigación, se dividieron aleatoriamente a los participantes para recibir diariamente diferentes niveles de vitamina E. Un grupo recibió sólo una pastilla de placebo. El estudio dio seguimiento a los participantes por 8 años para ver cuántos desarrollaron un tipo particular de cáncer durante ese período. ¿Cuál de las siguientes respuestas da la mejor explicación en cuanto al propósito de la asignación al azar en este estudio?
    a. Para incrementar la exactitud de los resultados de investigación.
    b. Para estar seguros que todos los pacientes potenciales de cáncer tenían igualdad de oportunidades de ser seleccionados para el estudio.
    c. Para reducir la cantidad del error de la muestra.
    d. Para producir grupos de tratamiento con características similares.
    e. Para evitar sesgo en los resultados.

Evaluación Comprensiva de Resultados para un primer curso de Estadística (CAOS 4)

Las preguntas de 8 a 10 se refieren a la siguiente situación:

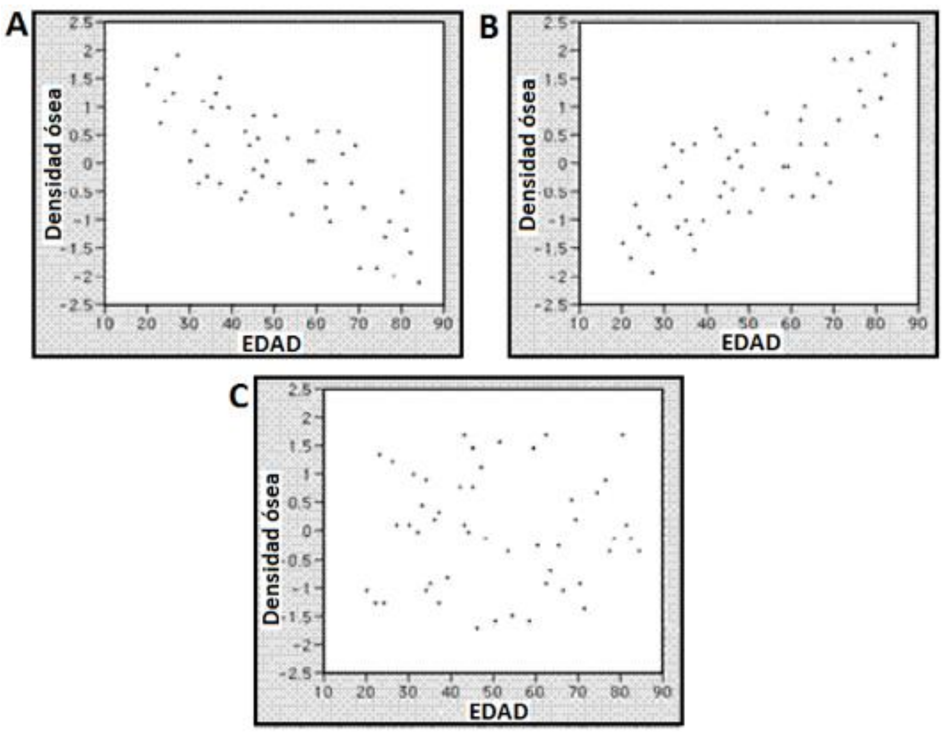
Los dos diagramas de caja siguientes muestran los resultados finales de un examen para todos los estudiantes en dos diferentes secciones del mismo curso.



8. ¿En cuál sección esperarías que se presentara una mayor desviación estándar en los puntajes de los exámenes?
   a. Sección A
   b. Sección B
   c. Ambas secciones son mas o menos igual
   d. Esto es imposible decirlo
9. ¿Cuál conjunto de datos tiene un mayor porcentaje de estudiantes con puntajes iguales o menores a 30?
   a. Sección A
   b. Sección B
   c. Ambas secciones son mas o menos igual
   d. Esto es imposible decirlo
10. ¿Cuál sección tiene un mayor porcentaje de estudiantes con puntajes iguales o mayores que 80?
   a. Sección A
   b. Sección B
   c. Ambas secciones son mas o menos igual

19. Una estudiante está diseñando un estudio de investigación. Ella espera demostrar que los resultados de un experimento son estadísticamente significativos. ¿Qué tipo de valor p (p value) podría obtener?
    a. Un valor p (p value) grande
    b. Un valor p (p value) pequeño
    c. La magnitud de un valor p (p value) no tiene impacto sobre la significancia estadística.

20. La densidad ósea se mide típicamente como un puntaje estandarizado con una media de 0 y una desviación estándar de 1. Las puntuaciones más bajas corresponden a una densidad ósea menor. ¿Cuál de las siguientes gráficas muestran que a medida que envejecen las mujeres tienden a tener menor densidad ósea?



    a. Gráfica A
    b. Gráfica B
    c. Gráfica C

Evaluación Comprensiva de Resultados para un primer curso de Estadística (CAOS 4)

Las preguntas 28 a 31 se refieren a la siguiente situación:

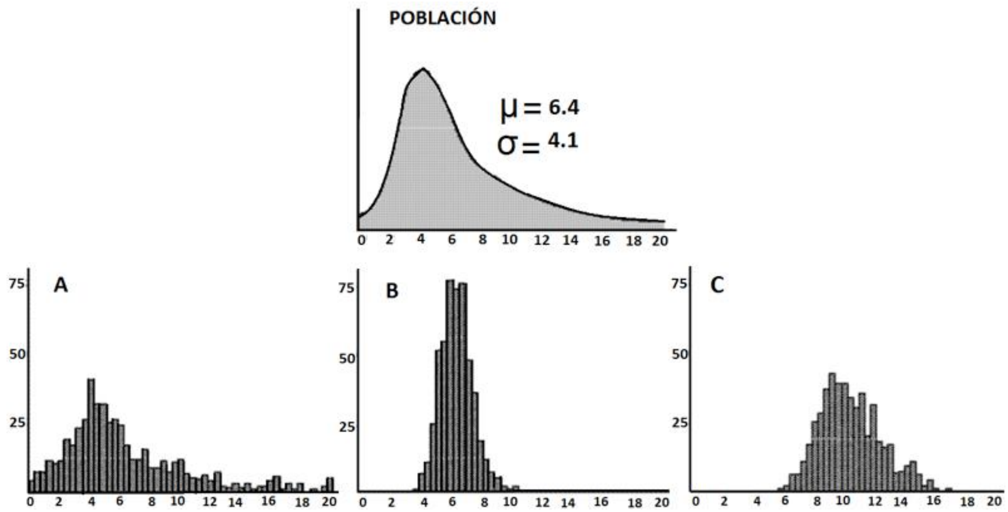
En la clase de estadística de una preparatoria, se quiere calcular el número promedio de chispas de chocolate en una marca genérica de galletas de chispas de chocolate. Obtienen una muestra aleatoria de las galletas, cuentan las chispas de chocolate en cada galleta y calculan un intervalo de confianza al 95% para el número promedio de chispas de chocolate por galleta (18.6 a 21.3). Las preguntas 28, 29, 30 y 31, presentan cuatro diferentes interpretaciones de estos resultados. Indica si cada interpretación es válida o inválida.

28. Estamos 95% seguros que cada galleta para esta marca tiene aproximadamente 18.6 a 21.3 chispas de chocolate.
    a. Válido
    b. Inválido
29. Esperamos que el 95% de las galletas tengan entre 18.6 y 21.3 chispas de chocolate.
    a. Válido
    b. Inválido
30. Podríamos esperar que alrededor del 95% de todas las posibles medias de las muestras para esta población sean entre 18.6 y 21.3 chispas de chocolate.
    a. Válido
    b. Inválido

31 Estamos 95% seguros que el intervalo de confianza de 18.6 a 21.3 incluye el numero real promedio de las chispas de chocolate por galleta.
   a. Válido
   b. Inválido

Evaluación Comprensiva de Resultados para un primer curso de Estadística (CAOS 4)

Las preguntas 34 y 35 se refieren a la siguiente situación.
A continuación, se presentan cuatro gráficas. La gráfica en la parte superior es una distribución de los resultados de una prueba para cierta población. La puntuación media es de 6.4 y la desviación estándar es de 4.1.



34. ¿Cuál gráfica (A, B, o C) piensas que representa una muestra aleatoria de 500 valores para esta población?
    a. Gráfica A
    b. Gráfica B
    c. Gráfica C

35. ¿Cuál gráfica (A, B, o C) piensas que representa la distribución de las medias muestrales de 500 muestras aleatorias cada una de tamaño 9?
    a. Gráfica A
    b. Gráfica B
    c. Gráfica C

37. Una estudiante participa en una prueba de sabor de *Coca – Cola* vs *Pepsi*. Ella identifica correctamente el refresco en 4 de 6 intentos. Ella afirma que esto demuestra que puede con confianza diferenciar entre los 2 tipos de refrescos. Tú has estudiado estadística y quieres determinar la probabilidad de que alguien tenga mínimo 4 aciertos de los 6 intentos sólo por azar. ¿Cuál de las siguientes proporcionaría una estimación acertada de esta probabilidad?
    a. Tener un estudiante que repita el experimento muchas veces y calcular el porcentaje de tiempo en que distingue correctamente entre las marcas.
    b. Hacer una simulación en la computadora, con un 50% de probabilidad de adivinar la bebida correcta en cada intento y calcular el porcentaje de veces en el que hay 4 o más aciertos en un total de 6 intentos.
    c. Repetir este experimento con una muestra muy grande de gente y calcular el porcentaje de gente quien hizo 4 aciertos correctos de 6 intentos.
    d. Todos los métodos mencionados anteriormente proporcionaría una estimación acertada de la probabilidad.

.

Chapter E SAS Models Codes

```
/* Model Population and Statistics Topic factors with random effect (Subjects) */
proc glimmix data=Stat_Topic_Logic;
class Topic Population Subjects;
model Correct/Trials= Topic Population Topic*Population/ ddfm=satterthwaite solution;
random Subjects/ solution;
contrast'Bivariate Data - Data Collection and Design' Topic 1 0 -1 0 0 0 0 0 ;
contrast'Bivariate Data - Probability' Topic 1 0 0 0 0 -1 0 0;
contrast'Bivariate Data - Sampling Distribution' Topic 1 0 0 0 0 0 -1 0;
covtest 'random effect variance' ZEROG;
run;

/* Model Population and Statistics Phase factors */
proc glimmix data=Stat_Phase_Logic ;
class Phase Population Subjects Course;
model Correct/Trials= Phase Population Phase*Population Course Phase*Course/
ddfm=satterthwaite solution;
contrast'Collect Data - Analyze Data' Phase -1 1 0  ;
contrast'Collect Data - Interpret Results' Phase 0 1 -1;
contrast'Analyze Data - Interpret Results' Phase 1 0 -1;
run;

/* Model Statistics Course and Statistics Topic factors */
proc glimmix data=Stat_Topic_Logic;
class Topic Course Subjects;
model Correct/Trials= Topic Course Topic*Course/ ddfm=satterthwaite solution;
contrast'Additional Course - Only ESMA 3016' Course 1 -1 ;
run;
```

/* Model Statistics Course and Statistics Phase factors */

```
proc glimmix data=Stat_Phase_Logic;
class Phase Course;
model Correct/Trials= Phase Course Phase*Course/ ddfm=satterthwaite solution;
run;
```