

The Results of Data Collected from Surveys to Predict the Effectiveness of Undergraduate Research Experience in Ocean, Marine, and Polar Science Program (2009) and Virtual Conferences (2010)

Dalesha Cartman, Marvin Elder, and Dr. Yolanda McMillian (mentor)
Undergraduate Research Experience in Ocean, Marine, and Polar Sciences-Elizabeth City State University 1704 Weeksville Road, Box 672, Elizabeth City, North Carolina 27909

Keywords: Mean, Median, Mode, skew (positive or negative), normally distributed, bimodal, correlation, standard deviation, degrees of freedom, critical values, margin of errors, confidence intervals, variance, hypothesis (null or alternatives), right-tailed test, left-tailed test, and two-tailed test.

Abstract- The Undergraduate Research Experience in Ocean, Marine, and Polar Science (URE OMPS) program is set up to promote the professional development of undergraduate students through their participation in ocean, marine and polar science research. Each student was assigned to a specific research team, where they worked closely with the assigned faculty.

An additional component of the program was the opportunity for students to participate in virtual seminars. These seminars allowed the students to become more familiar with topics such as global warming and ice sheets, and it also allowed students to interact with the nation's most prominent scientists.

The primary focus of the research project was three-fold. First, research was conducted on the role of the Institutional Review Board (IRB). During the research of the

IRB, mock IRB approval applications were submitted for review. These actions had to be taken before any research could begin.

Second, the focus of the research project was to assess the hypothesized success of both the URE OMPS program and the virtual seminars through a comprehensive data analysis of questionnaire responses using experimental statistics.

Third, the design of experimental questionnaires was explored. Demographic, Likert Scale, open and close -ended survey questions were all used for questionnaires that were administered after the virtual seminars. Calculations of the statistical measures were done using the one sample and two sample tests for observational data using the statistical software packages Excel- StatPlus, and Minitab.

I. Introduction

The Undergraduate Research Experience in Ocean, Marine, and Polar

Science (URE OMPS) Program started 15 years ago in 1995. It had groups of about 20 students each year that worked with faculty members and getting hands-

on experience in the field of ocean, marine and polar science. The main goal of the research program was to investigate the trends in which global warming has on the environment. Some expert scientists in the program even received a chance to travel to Greenland and Antarctica to collect ice core samples and test them for any sudden changes over the years. Students obtain a chance to hear about the scientists' data and results through virtual conferences that the program offer.

Dr. Hayden is the principal investigator over the URE OMPS program and the director over the Center

Institutional Review Board (IRB)

Institutional Review Board (IRB), also known as the Independent Ethics Committee (IEC) or the Ethical Review Board (ERB), is a committee that has been organized to approve, monitor, and review biomedical and behavior research involving humans to protect their rights as research subjects. There are many research projects that can be done with human subjects but some cannot involve children. The different research projects are research involving normal educational practices; research involving the use of educational tests survey procedures, interview procedures, or observation of public behavior in which the subjects identity stays hidden and brings no harm to the subjects; research involving educational tests in which the subject identity stays hidden and no harm will come to the subject but this may include children; research involving the use of survey procedures or interview procedures or observation of public behavior for which subjects cannot be identified or release of the information would not be harmful

of Excellence in Remote Sensing Education and Research (CERSER). She has been a computer science professor at Elizabeth City State University 1980. She has contributed to many publications and received many awards.

This project consisted of collected surveys from the past URE OMPS program students, which recorded how effective the program was and it also consisted of collected surveys done this summer that tested the effectiveness of the virtual conferences viewed. The trends in the data explained how effective the program and the virtual conferences were.

to the subject (children can not be included in this); survey or interview of public or elected officials, research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified (children can not be included in this); research and demonstration projects that are conducted by or subject to the approval of Department or Agency heads, and which are designed to study or evaluate public benefits or services (children can not be included in this); and taste and food quality evaluation and consumer acceptance studies (children can not be included in this).

This project involved the use of research involving the use of educational tests survey procedures, interview procedures, or observation of public behavior in which the subjects identity stays hidden and brings no harm to the subjects and survey or interview of public or elected officials, and research involving the collection or study of

existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified. Before the research could start, there was a twenty-two page that was filled out. The packet asked

Surveys

The first surveys received were the surveys from the 2008 summer and the spring of 2010. The surveys were looked over and discussed seeing what the data could be used for and how to store the data.

The next surveys that were to be used were ones to collect the data from the virtual conferences and how effect they were to the students. The first step as to find a place to create surveys without a limitation in the amount of questions to be asked or answered. After searching the internet for a place to create surveys, Kwik Surveys came into view. Kwik Surveys was the questionnaire builder and also the storage device for the surveys created. Kwik Surveys was founded during January 2008 and has grown rapidly ever since. The company has attracted many clients internationally such as Disney, TATA, MySQL magazine, Glaxo Smith Kline, Pepsi as well as many government and educational organizations. This service is totally free of charge and has many advantages, such as an email facility, that allows users to send email notifications/ reminders about completing surveys in a timely manner. Also, the user's data is safe. The company has a secure server, which is backed up weekly and is located in a data center in Germany.

question so the IRB will know exactly what is being tested, how many test subjects, and if the subjects' identities would remain anonymous. If the subjects were to remain anonymous then there was a procedure to tell why they are remaining anonymous. After the checklist was filled out, the research began.

The only down-fault to this survey builder application is that a user must log into their account at least once every 18 months. Failure to comply with this requirement will result in the user's account and all data being removed from the servers.

Kwik Surveys was a very helpful tool in the research. It was used to create surveys about the overall rating of the virtual conferences. It contained questions such as, "would you attend this conference again?" or "how would you rate the presenter?" After the students listened to the speaker of a one of the virtual conferences, they went on the NIA website where the survey was located, read the information sheet, and answered the questions that were giving in the survey. After the students' answers were submitted, their results were taken and put into excel.

Data

The data that was collected from the past surveys and from the virtual conferences, were put into excel. In order to put the data into excel, the sentences had to be turned into phrases that could be understood. For example, "Family Socio-Economic Background" is FSEB, and "While growing up, did you have a parent, relative, or family friend who you knew was involved in computing science" is Involvement.

There were research questions created to separate the data and to compare certain parts of the data and to find trends in the data. For example, there would be a question asking, “Does the family socio-economic background effect the students’ decision to go to graduate school?” and there was a table created to see if the family’s socio-economic background effect the students decision for graduate school. There was also categorical data that was converted into numerical data in order to create graphs for the data and do calculations. After

turning the sentences into phrase and the categorical data into numerical data, each survey was put into alphabetical order and received an id number to tell the difference between them. The surveys from the summer, the surveys from the spring, and the virtual conference surveys where put into separate tables to compare the data from each one and to see if there were any trends. The data was also inputted in Mini-Tabs to create histogram for later calculation purposes.

Calculations

The first thing that was calculated were the URE program’s data and the virtual conference data. The mean, median, and mode were taken for every column of numerical data and put into the charts. These tests were done to find out if the data in the charts were balanced. The mean was the average of certain information that was highlighted to receive the results of the data. The median is the number that is in the middle of the group of numbers. The mode is the number that appears most out of the group of numbers. These three numbers should approximately be the same number or close to it. If the numbers had more then a .3 and above difference, there had to be observation made to explain why the averages were off.

- About 68% of the values lie within 1 standard deviation of the mean ($\mu \pm \sigma$)
- About 95% of the values lie within 2 standard deviations of the mean ($\mu \pm 2\sigma$)
- About 99% (99.7%) of the values lie within 3 standard deviations of the mean ($\mu \pm 3\sigma$)

These methods fell into place when the standard deviation equation is performed. The age column in the statistical data chart will be the example being used. The first step is to find the average of the data set.

$$\begin{aligned} \text{Ex:} \\ (19 \times 6) + (20 \times 4) + 21 + 23 + 18 &= 256 \\ 256/13 & \\ 19.69 & \end{aligned}$$

The next test that was taken was Standard Deviation. The Standard Deviation deals with many different steps to get to the results. This test is done to figure out if the information that is in the data is a normal distribution, which is the results lies within three standard deviations of the mean:

The second step is to take each number of the data set and subtract the average form it.

$$\begin{aligned} \text{Ex: } 19-19.69 &= -0.69 \text{ (6x)} & 20-19.69 &= \\ & 0.31 \text{ (4x)} & & \\ 23-19.69 &= 3.31 & 21-19.69 &= \\ & 1.31 & & \\ & & 18-19.69 &= -1.69 \end{aligned}$$

The third step is to square each difference of the difference

$$\begin{aligned} \text{Ex: } -0.69^2 &= 0.48 \text{ (6x)} \\ 0.31^2 &= 0.096 \text{ (4x)} \\ 3.31^2 &= 10.96 \\ 1.31^2 &= 1.72 \\ -1.69^2 &= 2.86 \end{aligned}$$

The fourth step is to add all the results from step 3.

$$\begin{aligned} \text{Ex:} \\ (0.48 \times 6) + (0.096 \times 4) + 10.96 + 1.72 + 2.86 &= 18.8 \end{aligned}$$

The fifth step is to divide the sum of squares by the number of numbers in the data set

$$\text{Ex: } \frac{13}{18.8} = 1.45$$

The sixth and final step is to take the square root of the results from step 5.

$$\sqrt{1.45} = 1.20$$

These were all the steps taken to find the standard deviation of the data in the charts but since excel had a formula that did all the work, the only things that had to be put in was the cells is the column that was being used.

After Standard Deviation, there was a test done called the Confidence Intervals. This test deals with a lot of different equation to develop the answer that is need for the results. It was used a range of value that could be certain the mean will fall between. The variables used for the test were as follows:

- μ - Population mean

- \bar{x} - Sample mean
- s- Sample Standard Deviation
- n- Sample Size

The problem that is presented as an example is “There was 13 surveys with the students age on them and measure the average age of students in this program. The sample mean is 19.69 with a sample standard deviation of 1.20. Find the 95% confidence interval for the mean temperature.”

The first step taken in this equation was to identify the variables that were used in the equation.

- n- 13
- s- 1.20
- \bar{x} -19.69
- 95% confidence interval

The second step is finding the standard error.

$$SE = 1.20 / \sqrt{13} = 0.33$$

The third step is to find alpha, the critical probability, and Degrees of Freedom (the number of values in the final calculation of a statistic that are free to vary).

$$\begin{aligned} \alpha &= 1 - \left(\frac{95}{100}\right) = 0.05 \\ \rho^* &= 1 - \frac{.05}{2} = 0.975 \\ df &= 13 - 1 = 12 \end{aligned}$$

The fourth step taken is finding the critical value (the value a test statistic must exceed for the null hypothesis to be rejected) based on the standard t Distribution table.

2.179

Results

Discussion

There were plenty of errors in this experiment. One of the errors was that there was not enough data from the surveys that was collected to make any results that shows some type of trend in the data. In the histogram graphs created the were not enough normal graphs trend to develop an accurate hypothesis...

The ratio of men to women in all the undergraduate programs tested was almost two to one every time. There are questions to see if it was because of personal preference of if the Secretary of Education had anything to do with it...

Instead of the students stating minors that they have or will receive in the future, they should state there major. This will make it simpler to find trends related to students major

