



Leveraging the power of Matlab, SPSS, EXCEL, and Minitab for Statistical Analysis and Inference

Dr. Mohammad Rafiq Muqri, DeVry University, Pomona

Dr. Mohammad R. Muqri is a Professor in College of Engineering and Information Sciences at DeVry University. He received his M.S.E.E. degree from University of Tennessee, Knoxville. His research interests include modeling and simulations, algorithmic computing, analog & digital signal processing, and statistical analysis and inference.

**Nikole Harper
Hasan Muqri
Mr. Brian Keith Wesr Sr.**

Leveraging the power of Matlab, SPSS, EXCEL and Minitab for Statistical analysis and inference

Abstract

For many undergraduate and graduate engineering technology students, data collection and data analysis—including methodology, statistical analysis, and data preparation—is the most daunting and frustrating aspect of working on capstone senior projects and master’s theses. This paper provides an introduction to a number of statistical considerations, specifically statistical hypotheses, statistical methods, appropriate analytic techniques, and sample size justifications. Statistical analysis of data utilizing statistical software packages, including MATLAB, SPSS, Minitab, EXCEL, and R, will be shown for scientific applications, quality assurance, corporate finance modeling and other purposes. Our goal is to prep our students so that they are adept in a variety of basic and advanced statistical methods and are able to use these tools judiciously to fully understand and interpret their analyses and results.

This paper will explain how this learning and teaching module is instrumental for progressive learning of students; the paper will also demonstrate how the numerical and integral algorithms are derived and computed through leverage of EXCEL, Minitab and MATLAB data structures. As a result, there will be a discussion concerning the comparison of SPSS, EXCEL, Minitab and MATLAB programming, as well as student feedback. The result of this new approach is expected to strengthen the capacity and quality of our undergraduate and graduate degree programs, in addition to enhancing overall student learning and satisfaction.

Objective

This teaching module was designed to enhance the knowledge and expertise of our student which enabled them to apply inferential methods to different configurations of data thereby enhancing statistical literacy and critical thinking. It includes both descriptive statistics and inferential concepts used to draw conclusions about a population. Research techniques, such as sampling and experiment design, are included for both single and multiple sample groups.

Choice of statistical software and Comparison

Most universities and institutions support a wide variety of Statistical software. The decision process related to the purchase of a statistical software is strongly depending on the context where we have to undertake the decision. Constraints like the budget, the surrounding IT infrastructure, the level of programming knowledge and skills of end user, the extent to which the statistical analysis has to be addressed, and many other context-based elements are much more than factors, they open to different scenario and require more than a reflection.

Table 1 is a partial comparative list of different statistical software used at our university by undergraduate and graduate students. It is not complete but gives a glimpse of its capabilities.

Table 1. Which Statistical Software to be used

Software	Interface	Learning Curve	Data Manipulation	Statistical Analysis	Graphics	Specialties
<i>SPSS</i>	Menus & Syntax	Gradual	Moderate	Moderate Scope Medium Versatility	Good	Custom Tables, ANOVA & Multivariate Analysis
Minitab	Menus & Syntax	Moderate	Strong	Broad Scope Medium Versatility	Very Good	Panel Data, Survey Data Analysis & Multiple Imputation
<i>SAS</i>	Syntax	Steep	Very Strong	Very Broad Scope High Versatility	Good	Large Datasets, Reporting, Password Encryption & Components for Specific Fields
<i>R</i>	Syntax	Steep	Very Strong	Very Broad Scope High Versatility	Very Good	Packages for Graphics, Web Scraping, Machine Learning & Predictive Modeling
<i>MATLAB</i>	Syntax	Moderate	Very Strong	Moderate Scope	Excellent	Simulations, Multidimensional Data, Image &

				High Versatility		Signal Processing
--	--	--	--	---------------------	--	----------------------

SAS (pronounced "sass") which stands for "statistical analysis system" initially began at North Carolina State University as a project to analyze agricultural research. Demand for such software capabilities began to grow, and SAS was founded in 1976 to help customers in all sorts of industries – from pharmaceutical companies and banks to academic and governmental entities. Microsoft Excel features calculation, graphing tools, pivot tables and a macro-programming language called Visual Basic for applications. It is user friendly, powerful and very popular among our undergraduate technology students.

Minitab is a statistical package distributed by Minitab Inc., a privately owned company headquartered in State College, Pennsylvania. It initially began as a light version of OMNITAB, a statistical analysis program by National Institute of Standards and Technology.

SAS (pronounced "sass") which stands for "statistical analysis system" initially began at North Carolina State University as a project to analyze agricultural research. Demand for such software capabilities began to grow, and SAS was founded in 1976 to help customers in all sorts of industries – from pharmaceutical companies and banks to academic and governmental entities. Microsoft Excel features calculation, graphing tools, pivot tables and a macro-programming language called Visual Basic for applications.

SPSS originally from SPSS Inc., stands for Statistical Package for the Social Sciences (SPSS). It is widely used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others. IBM acquired SPSS in 2009 and the current version is officially called IBM SPSS Statistics.

The following Statistics are included in the SPSS base software:

1. Descriptive statistics: Cross tabulation, Frequencies, Explore, Descriptive Ratio Statistics
2. Bivariate statistics: Means, t-test, ANOVA, Correlation (bivariate, partial, distances), Non parametric tests, Bayesian
3. Prediction for numerical outcomes: Linear Regression.
4. Prediction for identifying groups: Factor analysis, cluster analysis (two-step, K-means, hierarchal), Discriminant.
5. Geospatial analysis, simulation.
6. R extension(GUI).

Selected topics from the teaching module

The probabilities of events within the normal distribution are the same as finding the area under the curve. Excel provides a very convenient and easy way of finding areas under the standard normal curve. The Excel formula always calculates the probability at the given value or less. With this information, all the needed probabilities can be calculated. The syntax of the formula is =NORM.DIST(x_0 , μ , σ , TRUE) where, x_0 is the value of the continuous random variable, μ is the mean and σ is the standard deviation. The value TRUE request the cumulative probability for all x values in the range from $-\infty \leq x \leq x_0$.

Normal Distribution EXCEL Problem 1:

The mean incubation time for a type of fertilized egg kept at 100.1°F is 21 days. Suppose that the incubation times are approximately normally distributed with a standard deviation of 2 days.

- (a) What is the probability that a randomly selected fertilized egg hatches in less than 19 days?
- (b) What is the probability that a randomly selected fertilized egg hatches between 17 and 21 days?

$$P(x_0 < 19) = \text{NORM.DIST}(19, 21, 2, \text{true}) = 0.1587 \text{ or } 15.87 \%$$

$$P(17 < x_0 < 21) = \text{NORM.DIST}(21, 21, 2, \text{true}) - \text{NORM.DIST}(17, 21, 2, \text{true})$$

$$P(17 < x_0 < 21) = 0.472 = 47.2\%$$

Normal Distribution Inverse Problem 2:

The time spent (in days) waiting for a heart transplant for people ages 35-49 can be approximated by the normal distribution, as shown in the figure to the right. Given mean = 206 days, standard deviation = 24.3

- (a) What waiting time represents the 5th percentile?
- (b) What waiting time represents the third quartile?

$$5^{\text{th}} \text{ percentile waiting time} = \text{NORM.INV}(0.05, 206, 24.3) = 166 \text{ days}$$

$$\text{Third Quartile waiting time} = \text{NORM.INV}(0.75, 206, 24.3) = 222 \text{ days}$$

Binomial probability example

In a binomial experiment, the probability of exactly k successes in n trials ⁶ is:

$$P(k) = \frac{n!}{(n-k)!k!} p^k q^{n-k}$$

The result of a survey indicate that 67% of homemakers consider air conditioning a necessity. If 100 homemakers are selected at random. What is the probability that exactly 75 homemakers consider air conditioning a necessity?

Using EXCEL, $P(75) = \text{BINOM.DIST}(75, 100, 0.67, \text{FALSE}) = 0.0201004$

Similarly, the Poisson distribution is a discrete probability distribution of a random variable k , such that the probability of exactly x occurrences in an interval is:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

Poisson Distribution Example

The mean number of traffic accidents per month at a certain intersection is 3. What is the probability that in any given month five accidents will occur at this intersection?

Using MatLab, $P(5)$ is computed as follows

```
>> (3^5)*exp(3)/(5*4*3*2*1)
```

```
>> ans = 0.1008
```

Using EXCEL, $P(5) =$ Probability that in any given month five accidents will occur

$$= \text{POISSON.DIST}(5, 3, \text{FALSE}) = 0.100819$$

Confidence Interval Concepts

A confidence interval ⁸ (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI. The confidence level is the area in the center of the standard normal distribution. Imagine a normal distribution. For a 95% confidence interval, there would be 95% of the area in the middle of the distribution, there would be 2.5% at the low end, and there would be 2.5% at the high end. With a 95% confidence interval, one could be 95% sure that the true value of the population means is within that interval.

Since the population standard deviation is rarely known, we use sample standard deviation in its place

- $E \rightarrow$ Margin of Error
- $s \rightarrow$ sample standard deviation
- $\sigma \rightarrow$ population standard deviation
- $n \rightarrow$ sample size
- $z_c \rightarrow$ critical z-value

Typical critical values (z_c) are given below:

Confidence Level	Alpha α	Critical Value
80 %	0.20	1.282
90%	0.10	1.645
95%	0.05	1.960
98%	0.02	2.326
99%	0.01	2.5976

Note that the critical values and thus the margin of error become larger as the level of confidence increases.

We can also use EXCEL formula: critical z-value = NORM.S. INV () to find critical values. A random sample of 90 observations produced a mean of 25.9 and a standard deviation of 2.7. Find an approximate 95 % confidence interval for the population mean. For the confidence coefficient 0.95, $\alpha = .05$ this corresponds to $z_c = 1.96$.

Therefore the confidence interval is $25.9 + 1.96 * 2.7 / \text{sqrt}(90) = 25.34$ and $25.9 - 1.96 * 2.7 / \text{sqrt}(90) = 26.46$ which gives an interval of (25.34, 26.46).

Using EXCEL formula: CONFIDENCE.NORM(alpha, standard_dev, size) returns margin of error (E). To calculate the confidence interval for a population mean, the returned value 'E' must then be added to, and subtracted from, the sample mean.

$$\text{Confidence Interval} = (\bar{x} - E, \bar{x} + E)$$

$$E = \text{CONFIDENCE.NORM} (.05, 2.7, 90) = .5578$$

$$\text{Therefore Confidence Interval} = (25.9 - .5578, 25.9 + .5578) = (25.34, 26.46)$$

Confidence Intervals of Means with samples

From a small sample ($n < 30$), when the population standard deviation is not known, the steps are the same: Find E , then add and subtract E from mean. However, in this situation, the t-score must be used rather than the z-score. Therefore, a different formula in Excel is needed (=CONFIDENCE.T, rather than =CONFIDENCE.NORM), but the overall process is the same—find E and then add and subtract E from the mean.

For example, based on a sample of 20 bags, there is an average of 50 M&Ms in each bag with a standard deviation of 2. Construct the 95% confidence interval for this mean.

To do this in Excel, use formula =CONFIDENCE.T (alpha, standard deviation, sample size) which in this case would be =CONFIDENCE.T(0.05, 2, 20) = $E = 0.94$.

Then subtract and add ($50 - 0.94, 50 + 0.94$), which gives the 95% confidence interval of (49.06, 50.94).

Therefore, one can be 95% confident that the true population mean is between 49.06 and 50.94. Notice that with fewer in the sample, the confidence interval is wider. This makes sense because a smaller sample provides less information so the interval must be wider to maintain the same level of confidence.

Hypothesis Testing

In hypothesis testing we begin by making a tentative assumption about a population parameter. This tentative assumption is called the null hypothesis and is denoted by H_0 . We then define another hypothesis, called the alternative hypothesis, which is the opposite of what is stated in the null hypothesis and is denoted by H_a . The hypothesis testing procedure uses data from a sample to test the two competing statements indicated by H_0 and H_a .

Performing a One-Way ANOVA Test ⁷

A medical researcher wants to determine whether there is a difference in the mean lengths of time it takes three types of pain relievers to provide relief from headache pain. Several headache sufferers are randomly selected and given one of the three medications. Each headache sufferer records the time (in minutes) it takes medication to begin working. The results are shown in the table. At level of $\alpha = 0.01$, can you conclude that at least one mean time is different from others? Assume that each population of relief times is normally distributed and that population variances are equal.

$H_0: \mu_1 = \mu_2 = \mu_3$

H_a : At least one mean is different from others (Claim)

Because there are $k = 3$ samples, degree of freedom $df = k - 1 = 3 - 1 = 2$

The Minitab output and graphical plots depicted below were obtained using the following steps

- Step1. Select the Stat menu
2. Choose ANOVA
3. Choose One-Way
4. When the One-Way Analysis of Variance box appears.
Select Response data are in a separate column for each factor level.
5. Enter C2 C3 C4 in the response box
6. Click on Graph
7. Check – Individual value plot
8. Check boxplots of data
9. Residual Plots - Three in one

From Minitab output, $p = 0.26$. Given $\alpha = 0.01$

Interpretation: Since $p < \alpha$, There is not enough evidence at the 1% significance to conclude that there is a difference in the mean length of time it takes the three pain relievers to provide relief from headache pain.



Worksheet 1 ***

	C1-T	C2	C3	C4	C5	C6	C7	C8	C9
1	Med A	12	16	14					
2	Med B	15	14	17					
3	Med C	17	21	20					
4		12	15	15					
5			19						
6									

Session

Source	DF	SS	MS	F	P
Factor	2	21.92	10.96	1.50	0.269
Error	10	73.00	7.30		
Total	12	94.92			

S = 2.702 R-Sq = 23.10% R-Sq(adj) = 7.71%

Figure 1. One-Way ANOVA Test

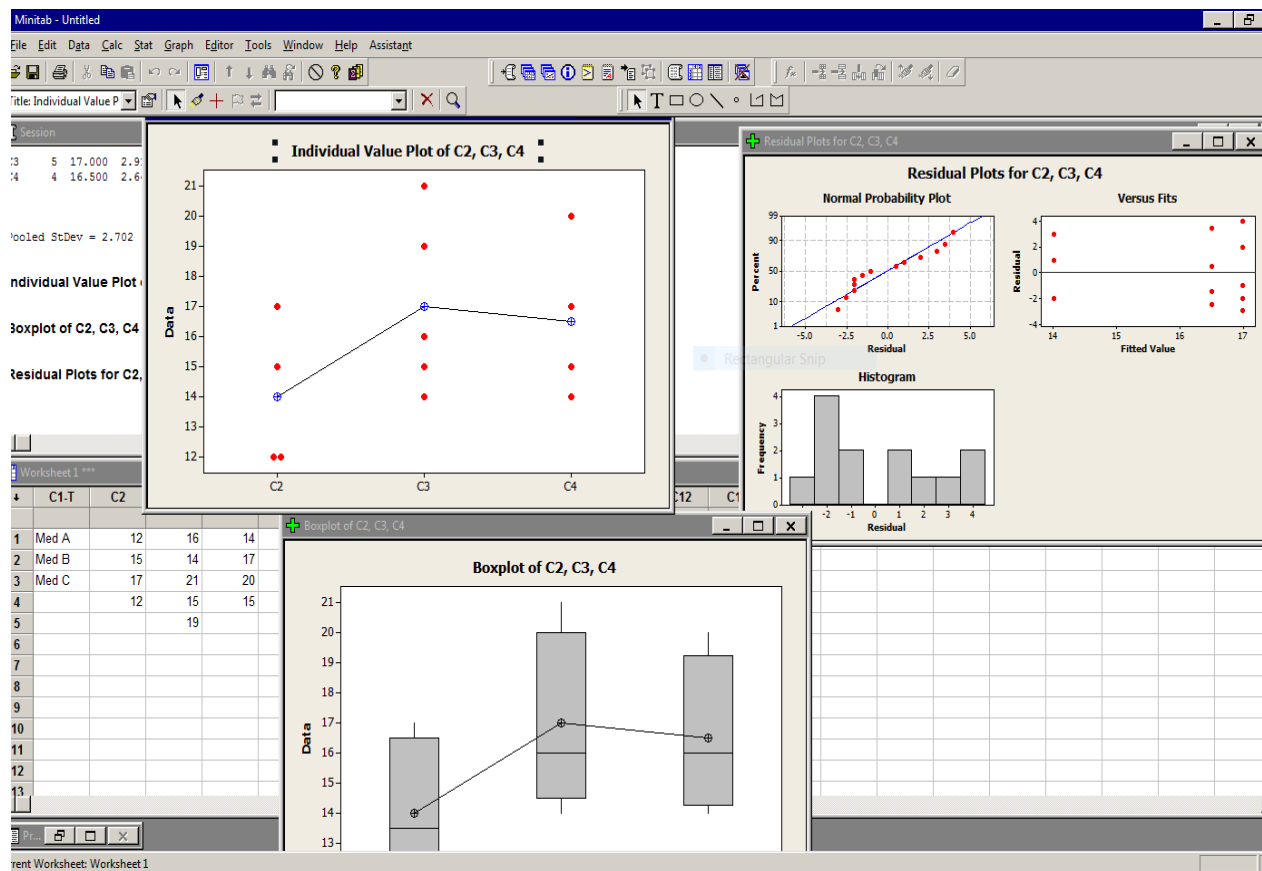


Figure 2. Individual Value, Box Plots and Residual Plots

R is primarily command-line driven program while S-Plus is more GUI-oriented. The user enters commands at the prompt (> by default) and each command is executed one at a time. The workspace is your current R working environment and includes any user-defined objects (vectors, matrices, data frames, lists, functions). At the end of an R session, the user can save an image of the current workspace that is automatically reloaded the next time R is started. For example, in either program you can type `t.test (Data)` to get detailed hypothesis test that by default, tests $H_0: \mu = 0$ and returns the test statistic, p-value, and other potentially useful quantities. But in S-Plus, you can simply click on a menu item instead of typing (and remembering) the `t.test` command. You may think of the analogy as R representing the command-line Linux OS and S-Plus representing the menu-driven Windows OS.

Typical R Example 1:

Suppose the manufacturer claims that the mean lifetime of a light bulb is more than 10,000 hours. In a sample of 30 light bulbs, it was found that they only last 9,900 hours on average. Assume the population standard deviation is 120 hours. At .05 significance level, can we reject the claim by the manufacturer?

Solution with R

The null hypothesis is that $\mu \geq 10000$. We begin with computing the test statistic.

```
> xbar = 9900      # sample mean
> mu0 = 10000     # hypothesized value
> sigma = 120     # population standard deviation
> n = 30          # sample size
> z = (xbar-mu0)/(sigma/sqrt(n))
> z              # test statistic
[1] -4.5644
```

We then compute the critical value at .05 significance level.

```
> alpha = .05
> z.alpha = qnorm(1-alpha)
> -z.alpha      # critical value
[1] -1.6449
```

The test statistic -4.5644 is less than the critical value of -1.6449. Hence, at .05 significance level, we reject the claim that mean lifetime of a light bulb is above 10,000 hours.

Alternative Solution using EXCEL

$$z_c = \text{NORM.S.INV}(0.05) = 1.644854$$

Not only engineers, scientists, and physicists, but also financial professionals worldwide use the interactive programming environment and prebuilt computational libraries of Matlab to develop quantitative applications in a fraction of time it would take them in C++ or Visual Basic. Matlab is not limited, but it is being successfully used for the following applications:

1. Chart historical and live
2. Market data
3. Model interest rates
4. Solve optimization problems
5. Develop quantitative models to optimize performance and minimize risk
6. Integrate with data sources and legacy software
7. Develop and deploy applications to production environments, desktops, servers, and the Web

Minitab basics:

This introduction to Minitab ⁵ is intended to provide one with enough information to get started using the basic functionality of Minitab. When you open Minitab, you will typically see something that looks like Figure 3.

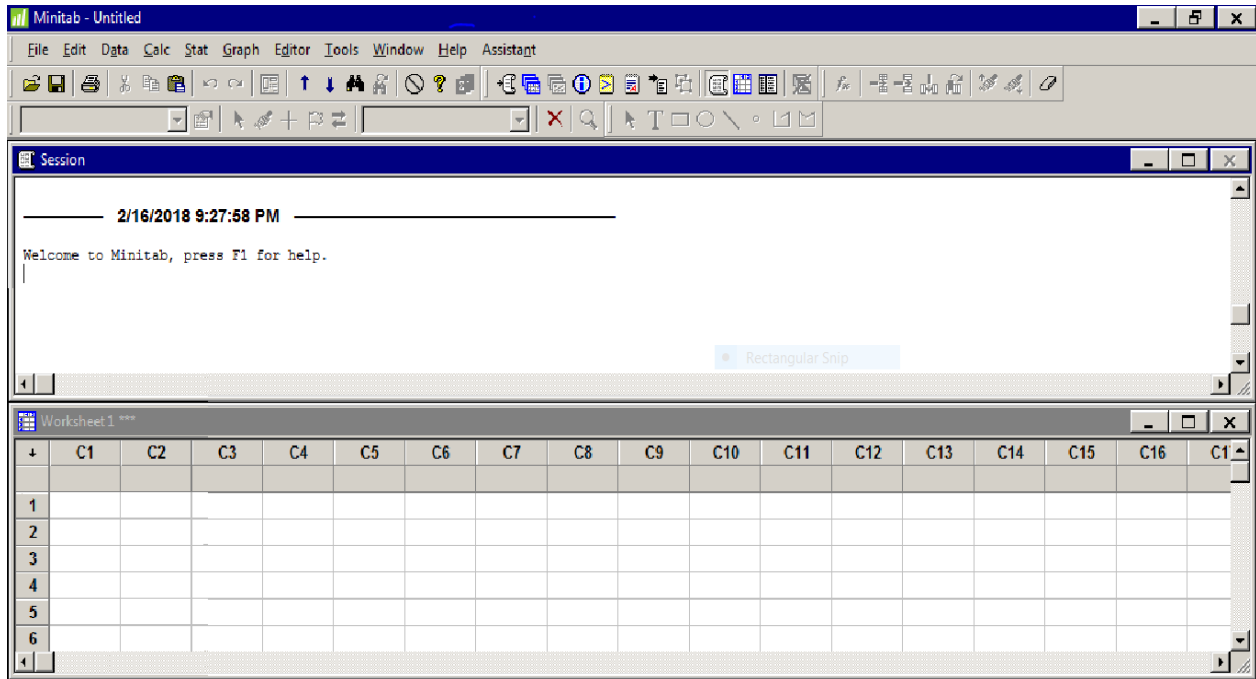


Figure 3. Session and Worksheet Windows with Minitab

The top window — called the "Session" window — is where Minitab will output the results of your requested statistical analyses. The bottom window — called the "Worksheet" window — is where you will copy and paste the data. The third type of window — called a "Graphics" window — only appears when you've asked Minitab to plot something.

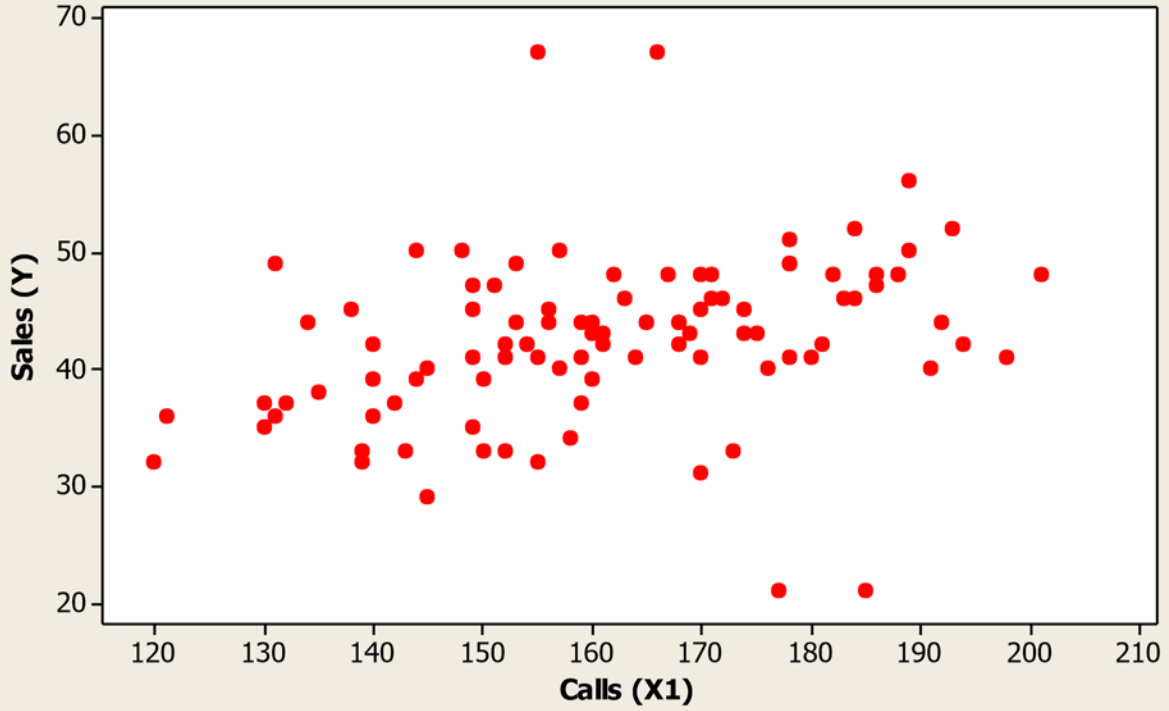
The active window is the window whose top bar appears as a darker blue. To make an inactive window active, simply click your mouse anywhere in the window. Convince yourself that you understand this behavior by clicking back and forth between the Session window and the Worksheet window — you should see the top bars of the two windows change back and forth from a lighter to a darker blue.

Here are few examples of a graphics window:

Generate a scatterplot for the specified dependent variable (Y) and the selected independent variable (X), including the graph of the "best fit" line. Due to sake of brevity, the data for the scatter plot and fitted line plots is not listed here. The course project gives students the opportunity to use Minitab data files and become proficient with different options and tools provided with Minitab.

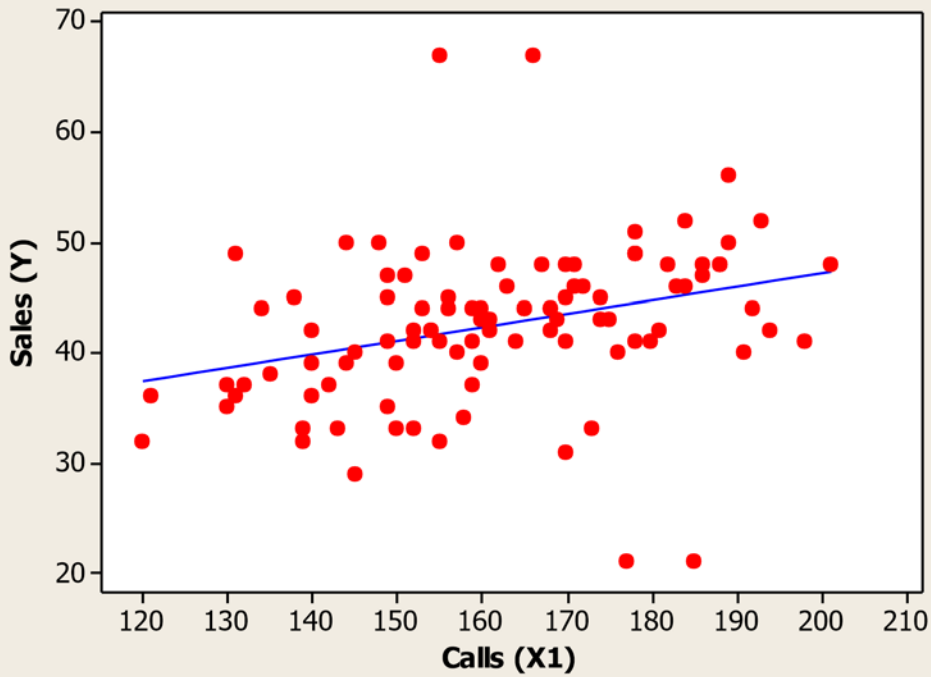
Figure 4. Scatter Plot and Fitted Line Plots for Sale (Y) vs. Calls (X1)

Scatterplot of Sales (Y) vs Calls (X1)



Fitted Line Plot

Sales (Y) = 22.52 + 0.1237 Calls (X1)



S	6.83308
R-Sq	10.1%
R-Sq(adj)	9.2%

Interpretation: The scatterplot of Sales (Y) and the Calls (X1) showed some clustering and outliers. By drawing the best fit line, a positive association is shown with linear association.

Determining the coefficient of correlation. Coefficient of correlation

$$(r) = \underline{0.318075}$$

Determining the coefficient of determination. Coefficient of determination

$$(r^2) = \underline{0.101172 \text{ or } 10\%}$$

Minitab can also be used to develop forecasts using different forecasting methods: exponential smoothing, moving averages, trend projection, and time series decomposition just to name a few.

Time series decomposition can be used to separate or decompose a time series into seasonal, trend, and irregular components. While this method can be used for forecasting, its primary applicability is to get a better understanding of the time series. To show how Minitab can be used to forecast a time series with trend and seasonality the following data ⁷ as shown in Table 2 was used to develop a forecast for the Television Set Sales using series decomposition.

Table 2. SEASONAL IRREGULAR VALUES FOR THE TELEVISION SET SALES TIME SERIES

Year	Quarter	Sales(1000s)
1	1	4.8
	2	4.1
	3	6
	4	6.5
2	1	5.8
	2	5.2
	3	6.8
	4	7.4
3	1	6
	2	5.6
	3	7.5
	4	7.8
4	1	6.3
	2	5.9
	3	8
	4	8.4

The following steps produce a forecast for the next quarter Television Set Sales:

1. Select the Stat menu
2. Choose Time Series
3. Choose Decomposition
4. When the Decomposition dialog box opens:

5. Enter C3 in variable box
6. Enter 4 in Season Length box
7. Select Multiplicative for method type
8. Select Trend plus Seasonal for Model components
9. Select Generate forecasts box
10. Enter 1 in the number of forecasts box
11. Enter 8 in the Starting from origin box
12. Click OK

The fitted trend equation, seasonal indexes, measures of forecast accuracy, and forecast for the next quarter are shown in session window.

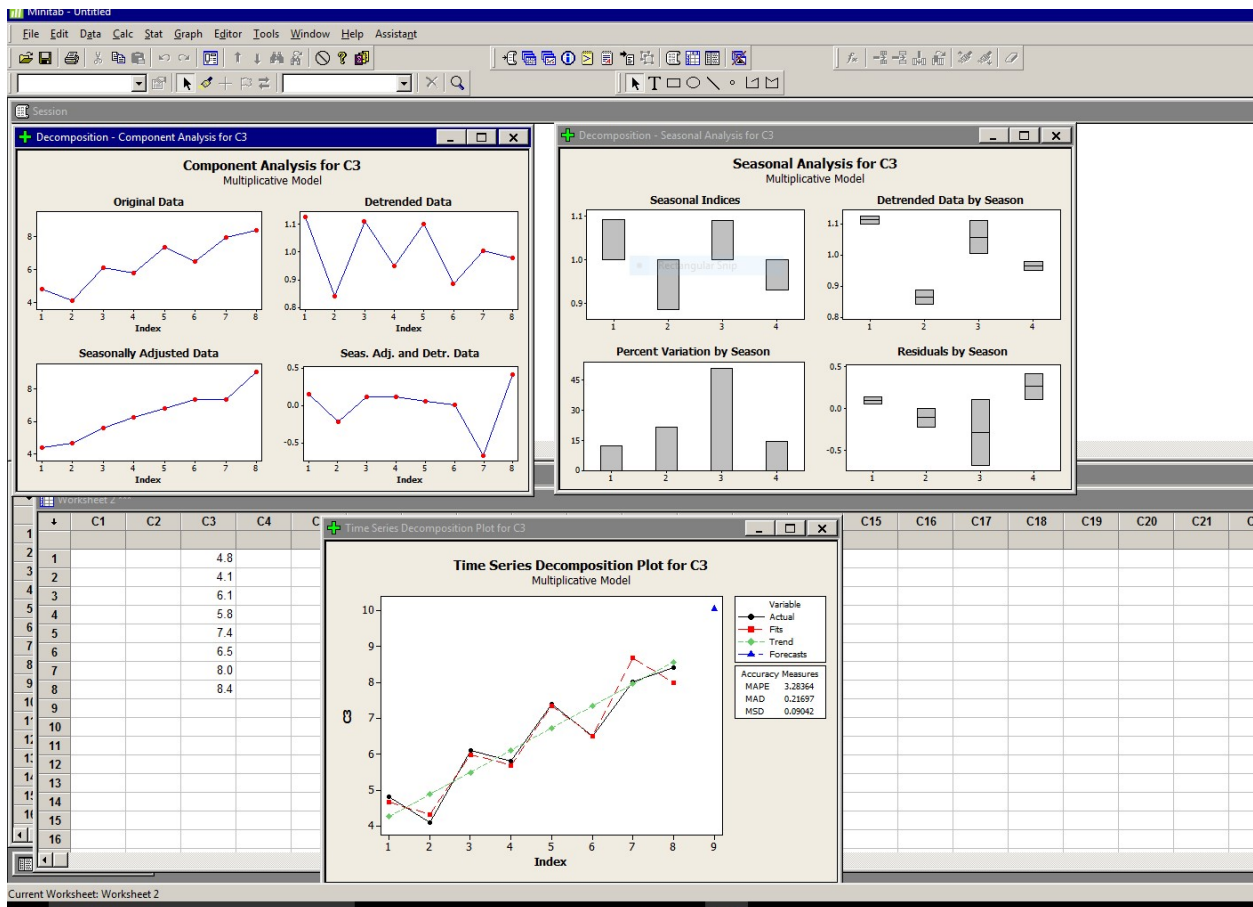


Figure 5. Deseasonalized Television Set Sales Time Series

In general, parametric tests⁹ have requirements about the nature or shape of the populations involved: nonparametric tests do not require that samples come from populations with normal distribution or any other particular distribution.

Given time, it can be proved that the group of students who are now actually taking the junior level courses, their understanding is far better than the group of students who were never exposed to this statistics-teaching module. This was witnessed about two years earlier as well when we developed a similar FE(EIT) ² certification exam prep module for our technology student.

Test Statistics and Analysis

Due to the nature of our diverse student population, a rank correlation method will be described. The rank correlation test, or Spearman's rank correlation, is a non-parametric test that uses ranks of sample data consisting of matched pairs. It is used to test for an association between two variables, so the null and alternate hypotheses are as follows (where p_s denotes the rank correlation coefficient for the entire population):

$H_0: p_s = 0$ (There is no correlation between the two variables.)

$H_0: p_s \neq 0$ (There is correlation between the two variables.)

The notation r_s will be used for the Spearman rank relation coefficient so as not to confuse it with the linear correlation coefficient r .

Rank Correlation Procedure and Notation

n = number of pairs of sampled data

d = difference between the ranks for the two values within a pair

= rank correlation coefficient for sample paired data (r_s is a sample statistic)

p_s = rank correlation coefficient for all the population data (p_s is a population parameter)

No ties: After converting the data in each sample to ranks, if there are no ties among the ranks for the first variable and there are no ties among the ranks for the second variable, the exact value of the test statistic can be calculated using this formula ⁹:

$$r_s = 1 - 6 \sum d^2 / n(n^2 - 1)$$

Ties: After converting the data in each sample to ranks, if either variable has ties among its ranks, the exact value of the test statistic can be found by using this formula:

$$rs = n \sum xy - \sum x \sum y / \text{sqrt} ((n \sum x^2 - (\sum x)^2) * (n \sum y^2 - (\sum y)^2))$$

Table 3 depicted below includes FE/EIT Subject Areas Grade Results Ranked by Pre-Test and Post-Test scores. It includes the difference d and the squares of the differences d^2 . The value of the rank correlation coefficient is computed in order to determine whether there is a correlation between the rankings of the Pre-test Subject Areas Grade Results and the rankings of the Post-Test Subject Areas Grade Results using a 0.05 significance level.

Table 3: FE/EIT Subject Areas Grade Results Ranked by Pre-Test and Post-Test

Subject Area	Student PreTest Grade Ranks	Student PostTest Grade Ranks	Difference	d ²
Linear Systems	1	1	0	0
Signal Processing	3	2	1	1
Electronics	2	3	1	1
Power Systems	5	7	2	4
Computer Systems	4	5	1	1
Control Systems	7	6	1	1
Communications	6	4	2	2
Ethics	8	6	2	4
Total				16

Following the procedure, the data in the Table 3, are in the form of ranks and the neither of the two variables has ties among ranks, so the exact value of the test statistic can be calculated as shown below using the equations above.

We use $n=8$ (for 8 pairs of data) and $\sum d^2=16$ (as shown in Table 2) to get $1 - (6 \sum d^2) / n(n^2 - 1) = 1 - [6(16) / 8(8^2 - 1)] = 1 - [96/504] = 0.8095$

If $n > 30$, critical values are found by the following formula, where the value of z corresponds to the significance level (for example if $\alpha = 0.05$, $z = 1.96$).

$$rs = \pm z / \sqrt{n - 1}$$

If $n \leq 30$, critical values are found by using⁹ the Table for Critical values of Spearman's Rank Correlation. For our case $n = 8$, so we determine that the critical values are ± 0.738 (based on $\alpha = 0.05$ and $n = 8$).

Because the absolute value of the test statistic $rs = 0.8095$ does exceed the positive critical value of 0.738, we reject the null hypothesis and conclude that there is a correlation. It has thus been demonstrated that there is significant correlation between the Pre-test Subject Areas Grade Results and the rankings of the Post-Test Subject Areas Grade results. The same analogy and similar analysis can be carried out and used to prove that Students appear to learn better the Statistical analysis and inference by studying more and going through rigorous practice and testing.

Applets and Statistics Calculators as Learning Tools

Students really liked the¹⁰ "Build a scatter plot" applet. This applet is a great learning tool to find out how two variables are related in the number of case (n) and observing the corresponding coefficient of correlation (r). Just Change the n (number of cases) or r (correlation coefficient) by sliding white triangles to see the effect on scatter plot. There were a good number of EXCEL based Statistic Calculators which were provided to

students for weekly topics. A wide range of other statistical applets and display tools⁴ are available on world wide web.

These are just few of the selective examples from the teaching module. Students were introduced early only to Matlab, Microsoft EXCEL, object based programming with C++, but also to basic concepts of descriptive statistics. Having being exposed to this basic programming, Regression Analysis: Model Building Time Series Analysis and Forecasting Learning Module has far-reaching implications. Advances in all fields of knowledge and the widespread availability of powerful computing devices have given a new meaning to the term statistical analysis and inference for real world applications.

Feedback and Assessment

We have offered the mini statistics crash course as a special topics course that is one way new courses are piloted locally at our university campus. This hands-on course is instrumental in the progressive learning of the students by reviewing, relating and applying fundamentals of Matlab, Minitab, EXCEL and SPSS for experimental design, using the estimated regression equation for estimation and prediction, trend projection and time series analysis, forecasting of variance, and statistical control. Although special topics are not evaluated in the same manner as standard session-long courses, feedback directly from students indicates that initial offering was well received.

A compilation of feedback was from eight junior students enrolled in the course for credit, and four sophomore students attended the lectures and discussions but did not receive the credit. The feedback from the student indicates that:

1. Students gained an appreciation of the statistical tools and their capabilities.. Using Minitab, EXCEL, SPSS and R puts many concepts in statistics analysis and inference to practice.
2. Students gained an appreciation for the difficulties involved in developing nonlinear models, detecting outliers and residual analysis.
3. Students spent a significant amount of time on quality assurance, statistical process control, and data analysis using Minitab, scripting with R and Matlab programming assignments (presumably relative to their other course work) but the results were satisfying. We did not receive any complaints about the level of effort required by, nor the time spent on the programming assignments.

Student's performance in the initial course offering and in the course of capstone projects was exceptionally high. This result was due to a biased sampling; the four juniors taking the special topic course initiated the effort, and the sophomores that attended regularly were invited by the instructor. We hope to see better understanding of basic principles and excellent performance in the future versions of the course,

Conclusion

Statistics Literacy and critical thinking is necessary in today's world that is fascinated with numbers and data. Even if one is not responsible for conducting statistical analysis, one needs the basic understanding to properly use the information for decision making. With proper guidance, monitoring, and diligent care, students were exposed early on scripting, discrete probability distributions, sampling distributions and statistical inference, design of experiments, analysis of variance. End of the course

survey and diagnostic quizzes demonstrated the marked increase in student understanding of statistics which is again attributed to early exposure of EXCEL, Matlab, R, and Minitab and proficiency acquired with statistical tools. It is hoped that the concepts covered in this paper will instigate future research and development in the field of statistical analysis and inference.

Bibliography

1. Muqri, M., Shakib, J., *A Taste of Java-Discrete and Fast Fourier Transforms*, American Society for Engineering Education, AC 2011-451.
2. Shakib, J., Saouli, M, Muqri, M., *An Electrical and Computer Startup Kit for Fundamentals of Engineering*, American Society for Engineering Education, AC 2016.
3. Mallat, S., Zhang, Z., Tutorial for Beginners – Quick R.
<https://www.statmethods.net/r-tutorial/index.html>
4. Statistics Simulations, Demonstrations - http://onlinestatbook.com/stat_sim/index.html
5. Statistics Resources for online courses – *Introduction to using Minitab*. Retrieved from <https://onlinecourses.science.psu.edu/statprogram/minitab>.
6. Larson, R, Farber, B., *Elementary Statistics Picturing the World*. Boston, MA: Pearson, 2015
7. Anderson, D. R., Sweeney, D. J., Cochran, J. J., *Statistics for Business and Economics*.: Cengage Learning.2016
8. Triola, T., F., *Essentials of Statistics*, Addison Wesley, 2007
9. Numbers Toolkit : *Spearman Correlation*. Retrieved from http://web.anglia.ac.uk/numbers/biostatistics/spearman/local_folder/critical_values.html
10. Huffman T., Russell M., *Exploring Correlation –Build a scatter plot*. Retrieved from <http://www.bc.edu/research/intasc/library/correlation.shtml>