Welcome to Biostatistics for Clinical Research. My hope is that your experience in this workshop will be pleasurable as well as enlightening. I realize that Latin phrases are not commonly exchanged among scholars these days, but I want to share the above with you. Here is the translation:
Finally, brethren, whatever is true, whatever is honorable, whatever is just, whatever is pure, whatever is lovely, whatever is gracious, if there is any excellence, if there is anything worthy of praise, think about these things. (RSV)

This is an excerpt from St. Paul's Letter to the Phillipians (4:8) and I am sure that people will be somewhat perplexed by my weaving Holy Scripture into a statistics workshop. However, I sincerely hope that your experience in this educational offering will help you recognize that statistical analysis possesses beauty and is a thing that is honorable, excellent, and worthy of praise. Given that our lives are immersed in a politically dense, data oriented, technical world that distracts us from thinking about whatever is true, honorable, just, pure, lovely, and gracious, I think that it is good for scientists to recapture the belief that scientific endeavors are meant to edify us. Ultimately, that means bringing us closer to God and deepening our appreciation of His creation. I hope you will be able to sense this attitude in the workshop and will find it a refreshing change from what you normally experience in academia and the culture.

Although I hope to transmit a sense of beauty, I am patently aware of the need to be pragmatic. There is, of course, no reason to believe that we can not strike a balance between theology and pragmatism. Hence, it is my aim to help participants acquire skills and knowledge that will enable them to manage common statistical tasks that confront researchers in the health and biological domains. From this, participants will be better prepared to interpret research journal articles and to conduct their own data analysis. Given this, the objectives for this workshop are as follows.

**General Objectives**

By the end of this workshop, participants should be able to ...

- classify statistical analysis as a collection of techniques that are intended to uncover information in data.
- discuss how statistical methodology is used as a research tool to aid thinking.
- describe the purpose and proper use of the statistical techniques that are presented in the workshop.
- define basic statistical concepts.
- conduct statistical tests of research hypotheses using methods presented in the workshop.
- interpret the meaning of tests of hypotheses and other statistical computations.
- identify the information that one needs to conduct a sample size estimation.

**Participant Activity**

**Practice Exercises.**
Participants will be given a set of practice exercises that will be distributed during the course of the workshop. These exercises are designed to help learners apply their knowledge to a practical problem. Solutions to the exercises will be provided so that
participants have timely feedback on their work. At the beginning of each session I will address questions about peoples’ computations and their interpretation. Everyone is highly encouraged to complete the practice exercise *before* analyzing data from their own research projects.

**Readings.**
Throughout the workshop, chapters from the Motulsky text will be given. In addition, special articles and other readings will be emailed or will be made available on the web.

**Computer Experience.**
Participants will learn to operate a current version of SPSS during the workshop. SPSS is a powerful statistical program that works well with clinical research. My hope is that everyone will come to value this program as an effective research tool. For most exercises, participants will be asked to use SPSS to replicate the work they have completed by hand. This will teach people how to use the program and will help nurture a sense of confidence in computerized statistical computations. Workshop participants will be able to purchase a copy of SPSS through the Chief of Staff’s office.

**Examinations & Grades.**
There are no examinations in this workshop. No grades or progress reports will be issued. This workshop should be viewed as continuing education.

**Computing Aids**

It goes without saying that learners are highly encouraged to use hand held calculators for this workshop. In fact, this may be the opportunity where some people will finally learn to decipher those cryptic symbols on their calculators! Bring your calculators to our sessions. You also are welcome to bring a laptop computer, if you have one.

**Your Responsibility**

I approach continuing education with the expectation that learners manage their own needs. Hence, do not be afraid to confront me or to ask questions. Be assertive. I expect that you will view Biostatistics for Clinical Research as your workshop.
**WORKSHOP TOPICS**

**Apr 13** Exploratory Data Analysis: Histograms and Boxplots.

The growing awareness of statistical methods and related probability values have had a major impact on the way scientists conduct data analysis. Unfortunately, many people have placed more attention on the technical aspects of statistical computations than on procedures for scrutinizing their data. The precipitous growth of high speed digital computers has only made this problem worse.

In today's session we will discuss the importance of exploring one’s data by looking at distributions of data by way of histograms. We also will cover the basic tenets of exploratory data analysis (EDA) that was promulgated by John W. Tukey. Participants will learn how to construct boxplots and how to generate histograms and boxplots on SPSS.

The primary goal of today's session is to help people realize that the purpose of data analysis is to uncover information in data, not to dictate the importance of a scientific finding. It is always and forever the scientist who must judge the value of empirical results. The methods of statistical analysis are simply tools to aid our thinking.

*In preparation for this session read:*

Motulsky; chapters 1, 7, 8, & 14.


**Apr 20** Probability: Quantifying Uncertainty.

One of the commonly believed falsehoods is that tests of statistical significance reveal truth – what really “is” and what really “isn’t.” Oddly enough, investigators won’t admit to believing this but, at the same time, they don’t want to make any statements about their research findings without a $p$ value. Hence, their behavior reveals what they are thinking.

In this class we will discuss the basic concept of probability and the law of large numbers. Standardized scores will be presented and we will learn how they aid us in computing probabilities from distributions of continuous variables. The purpose of this session is to lay the foundation for the proper understanding of hypothesis testing, which is based upon *probability*, not
certainty. The normal or Gaussian distribution will be presented as one type of continuous probability distribution.

In preparation for this session read:
Motulsky; chapters 2, 9, & 10.

Apr 27 The Logic of Hypothesis Testing: The One-Sample z and t Tests

Scientists are in the business of making educated guesses about the world and our hypotheses are usually made in the context of a scientific study. In today's session we will discuss how scientists rely upon statistical methods to help them test their hypotheses. The primary focus of this session will be upon the logic or steps of hypothesis testing. We will begin to learn about this process by discussing the z test. This is a very important session because the logic of hypothesis testing underlies the work for the remainder of the workshop. We will extend our discussion of hypothesis testing by learning about sampling distributions. An extension of the z test and Student's t test for a single sample will be presented.

In preparation for this session read:
Motulsky; chapters 3, 15, & 16.

May 4 Student’s t-Tests for Independent and Related Samples

The simplest form of an experiment compares two groups -- one that has received a treatment and one that has not. In today's session we will learn how scientists test hypotheses within this type of design. Two forms of Student's t test will be presented. The first assumes that the two groups are comprised of different individuals. The second assumes that the same set of individuals have participated in two experimental conditions or that two different groups of subjects are linked in some way.

In preparation for this session read:
Motulsky; chapters 12, 30, & 31 (also 18 & 19).

May 11 One-way Analysis of Variance.

Student's t-test is helpful for studies that compare two groups of subjects. What do we do when there are three or more groups to compare? In today's session we will learn the principle behind the analysis of variance, derived from the pioneering work of Sir Ronald Fisher. Following this, we will learn about a priori or planned comparisons in which we use multiple groups of subjects to test specific
hypotheses. We also will cover the less powerful (and over used) *a posteriori* or post hoc tests that allow us to make complete pair-wise comparisons of multiple groups.

*In preparation for this session read:*

Motulsky; chapters 22, 39 (pp. 369 – 374), & 40.

**May 18**  
Factorial Analysis of Variance.

The beauty of analysis of variance is that it allows us to analyze data from multiple groups and it gives us a mechanism to study the *interaction* of two or more factors. In today's session we will learn how to use ANOVA to study the influence of two independent variables on a dependent variable.

*In preparation for this session read:*

Motulsky; Chapter 39 (p. 374-375).

**May 25**  
Memorial Day – No session

**Jun 1**  
Bivariate Linear Correlation & Regression.

The fundamental activity of a scientist is to look for relationships. This is exactly what we have done with tests of statistical significance. However, in today's session participants will learn how relationships are measured with two scaled (continuous) variables. The Pearson correlation coefficient will be covered and participants will learn why it cannot range beyond ± 1.00 (you might have wondered about this if you had a previous statistics class). In addition, regression analysis, the method by which one uncovers the slope and y-intercept of a linear relationship, also will be presented.

*In preparation for this session read:*

Motulsky; chapters 32 & 33.

Jun 8 Multiple Linear Correlation & Regression.

In the previous session, we learned how to assess the relationship between two scaled variables. However, there will be times when an investigator wants to know how one variable (also called a target) is related to a set of other variables (known as predictors). This is what multiple correlation and regression (MCR) analysis allows one to accomplish. In today’s session participants will learn how to run MCR and how to interpret the major components of an SPSS output. They will learn how to evaluate the meaningfulness of a MCR analysis and how to look for and deal with multicollinearity.

In preparation for this session read:

Motulsky; chapter 37.


Jun 15 Chi Square Test For Goodness of Fit & Association.

The tests of significance that we have discussed to this point have all required data that are numeric measurements, in other words data that express an amount. There are times, however, when our data are categorical in nature and express type rather than amount. The chi-square test is uniquely designed to help scientists test hypotheses with data that are of this nature. In today’s session, participants will learn how “fit” a set of categorical frequencies against expected values and will learn how to find associations among two categorical variables.

In preparation for this session read:

Motulsky; chapter 27 (pp. 237-240).

Jun 22 Multiple Logistic Regression.

In a previous session, we saw how multiple linear correlation and regression analysis is used when an investigator wants to explain the variance of a target variable by a set of predictors. When the target is dichotomous (i.e., alive or dead, pass or fail, etc.) a different approach is needed. In this session, multiple logistic regression analysis will be presented as a method for classifying a dichotomous outcome with multiple predictors. Participants will learn how to evaluate the meaningfulness of a logistic regression analysis and how to interpret adjusted odds ratios.

(Continued on next page)
Jun 29  How to Think About Sample Sizes: Decoding the Mystery.

Our final session will be devoted to the topic of sample size estimation. In the past, this aspect of data analysis was sorely neglected. It has, however, recently been receiving proper attention among scientists. In this session we will discuss the concept of effect size and Type I and Type II errors. The focus of this session is to help participants learn about the type of information they need to bring a statistician so that they can avoid the blind question “How many subjects do I need?”

In preparation for this session read:

Motulsky; chapters 20 & 26.