Biostatistics
BIOM 559
4 Blocks (6 Credit Hours)
July 6, 2020– February 26, 2021
Master of Science in Clinical Research

Course Description: This course is an introduction to the basic concepts and methods of probability and statistics designed specifically for clinical and translational research scientists. Clinical and translational research examples are obtained from novel, textbook, and published sources. We will use the statistical software package Stata 16 to analyze real clinical and translational data sets.

Topics include numerical and graphical summaries of data, basic probability concepts, sampling distributions, basic concepts for estimation, estimation for population means and proportions, basic concepts for hypothesis testing, hypothesis tests for population means and proportions, analysis of variance, power and sample size calculations, simple linear regression, multivariable linear regression, additional regression techniques, chi-square distribution and analysis of frequencies, statistical evaluation of diagnostic tests, survival analysis, and nonparametric and distribution-free statistics.

Instructors:

Huining Kang, PhD, Associate Professor
Division of Epidemiology, Biostatistics, and Preventive Medicine
Department of Internal Medicine
Cancer Research Facility (CRF) Building, Rm G17
HuKang@salud.unm.edu
(505) 272-4195

Vernon Shane Pankratz, PhD, Professor
Division of Epidemiology, Biostatistics, and Preventive Medicine
Department of Internal Medicine
Cancer Research Facility (CRF) Building, Rm G17
vpankratz@salud.unm.edu
(505) 272-3718

Teaching Assistant:

Jonathan Odumegwu
Graduate Student
Department of Mathematics & Statistics
jodumegwu@unm.edu

Lectures & Labs: Lectures and labs will be conducted virtually through Zoom, Mondays, Wednesdays, and Fridays @ 2:45 – 3:45 pm.

Office hours: The TA's office hours will be immediately following the lab.
Drs. Kang and Pankratz's office hours are by appointment.

Course Materials:

Required: Daniel WW, Biostatistics: A Foundation for Analysis in the Health Sciences, 10th/11th Ed. Wiley, Hoboken NJ

Stata 15/16 IC or SE, StataCorp, College Station, TX, 2019

Optional Resources: Bland. An Introduction to Medical Statistics, 4th Ed., Oxford University Press, Oxford, UK, 2015

Biostatistics
BIOM 559
4 Blocks (6 Credit Hours)
July 6, 2020– February 26, 2021
Master of Science in Clinical Research

Van Belle, Fisher, Heagerty, and Lumley. Biostatistics, A methodology for the health sciences. 2nd Ed. Wiley, Hoboken NJ, 2004

SAS 9.4, SAS Institute, Cary, NC 2012

R. http://www.**R**-project.org/

Course Grading:

Exam	50%
Homework	35%
Class participation (includes attendance, team work assessment as applicable, and contribution to in-class discussions)	15%

Course grades are as follows:

Letter Grade	Numeric Grade
A+	97-100%
A	93-96%
A-	90-92%
B+	87-89%
B	83-86%
B-	80-82%
C+	77-79%
C	73-76%
C-	70-72%
D+	67-69%
D	63-66%
D-	60-62%

Notes to class:

- Many students find the homework for this class to be time-consuming. Not only might you be learning new concepts for the first time, you are also likely learning a new software package which alone can be challenging. Plan accordingly and seek help either from course instructors and the TA, and/or your classmates before you feel too lost.
- All homework assignments are expected to be typed with a word processor, with answers and explanations clearly identified. Sloppy and/or disorganized work will be returned ungraded or penalized.

Biostatistics
BIOM 559
4 Blocks (6 Credit Hours)
July 6, 2020– February 26, 2021
Master of Science in Clinical Research

- Homework will be graded by the teaching assistant. Any concerns regarding homework grades should initially be addressed with the TA. The instructors will grade all exams at the end the blocks and will assign final grades.
- Students may work together on homework assignments, but *each student must turn in his or her own work.* Please refer to UNM policies regarding academic dishonesty.
- The exams must be completed individually.
- Unexcused absences and frequent tardiness will lower your participation grade. Please send the instructors an email notifying the instructors of any scheduled absences.
- In accordance with University Policy 2310 and the Americans with Disabilities Act (ADA), academic accommodations may be made for any student who notifies the instructor of the need for an accommodation. It is imperative that you take the initiative to bring such needs to the instructor's attention, as he/she are not legally permitted to inquire. Students who may require assistance in emergency evacuations should contact the instructor as to the most appropriate procedures to follow. Contact Accessibility Resource Center at 505-277-3506 for additional information.
- A Note about Sexual Violence and Sexual Misconduct: As a UNM faculty member, we are required to inform the Title IX Coordinator at the Office of Equal Opportunity (oeo.unm.edu) of any report we receive of gender discrimination which includes sexual harassment, sexual misconduct, and/or sexual violence. You can read the full campus policy regarding sexual misconduct at https://policy.unm.edu/universitypolicies/2000/2740.html. If you have experienced sexual violence or sexual misconduct, please ask a faculty or staff member for help or contact the LoboRESPECT Advocacy Center.

Each student is expected to maintain the highest standards of honesty and integrity in academic and professional matters. The University reserves the right to take disciplinary action, up to and including dismissal, against any student who is found guilty of academic dishonesty or otherwise fails to meet the standards. Any student judged to have engaged in academic dishonesty in course work may receive a reduced or failing grade for the work in question and/or for the course.

Course Outline and Learning Objectives: We strive to cover all of these topics, but some topics may be omitted or others may be included. The order of topics may change slightly.

Biostatistics
BIOM 559
4 Blocks (6 Credit Hours)
July 6, 2020– February 26, 2021
Master of Science in Clinical Research

Block 1 (July 6 – August 28, 2020)

1. Basic concepts and descriptive statistics

a. Introduction to biostatistics
b. Basic concepts: types of data and variables, population, sample, types of measurement scales
c. Numerical summaries: proportion, mean, standard deviation, median and percentiles
d. Graphical summaries: bar graphs, histograms, dot plots, stem-and leaf plot, box plots, scatter plots

Objective: Understand different types of data and variables. Produce and understand numerical and graphical summaries that are commonly used in clinical and translational research.

2. Basic probability concepts

a. Random variable and event
b. Definition of probability
c. Rules of obtaining probabilities: conditional probability, joint probability multiplication and addition rules, independent events, complementary events, marginal probability
d. Baye's theorem and statistical evaluation of biomedical tests

Objective: Understand classical, relative frequency probabilities. Understand the properties of probability and selected probability rules. Be able to calculate the probability of an event. Understand the measures for evaluating a biomedical test and be able to use Bayes' theorem when calculating screening test results.

3. Probability distribution

a. Probability distributions of discrete variables
b. Binominal distribution
c. Poisson distribution
d. Probability distributions of continuous variables
e. Normal distribution

Objective: Understand the basic concepts and properties probability distribution. Articulate the types of data that might be modeled by a binary variable, by a continuous variable, or neither. Be able to calculate the probability of commonly used distributions using statistical software.

4. Sampling distribution

a. Construction and important characteristics of sampling distributions
b. Central limit theorem
c. Sampling with/without Replacement
d. Distributions of the sample mean
e. Distributions of the difference between two sample means
f. Distributions of the sample proportion
g. Distributions of the difference between two sample proportions

Objective: Understand the concept of sampling distributions applied to sample proportions and sample means. Understand the central limit theorem and when to apply it. Be able to calculate basic probabilities using a sampling distribution.

Biostatistics
BIOM 559
4 Blocks (6 Credit Hours)
July 6, 2020– February 26, 2021
Master of Science in Clinical Research

Block 2 (August 31 – October 23, 2020)
5. Estimation

a. Basic concepts of point and interval estimates for population parameters
b. Confidence intervals for the population mean and for the difference between two population means
c. t-distribution
d. Confidence intervals for the population proportion and for the difference between two population proportions
e. Determination of sample size for estimating means and proportions
f. Confidence intervals for variance of a normally distributed population and Chi-square distribution
g. Confidence intervals for the ratio of the variance of two normally distributed populations, and *F* distribution

Objective: Understand the importance and basic principles of estimation. Be able to calculate interval estimates for a variety of parameters. Be able to interpret a confidence interval from both a practical and a probabilistic viewpoint. Understand the basic properties and uses of the t distribution, chi-square distribution, and F distribution.

6. Hypothesis testing

a. Hypothesis testing for the population mean: null and alternative hypotheses, rejection rule, significance levels and p-values
b. Relationship between confidence intervals and hypothesis testing
c. Hypothesis tests for a population proportion
d. Two sample problems: tests for the difference between the means or the proportions of two populations
e. Hypothesis tests for a single population variance
f. Hypothesis tests for a ratio of two population variances

Objective: Understand the basic concepts and principles of statistical hypothesis tests. Articulate the general procedure of the hypothesis test. Understand the relationship between confidence intervals and hypothesis tests and be able to generate the decision rule of the hypothesis tests based on the confidence intervals. Understand the assumptions underlying the tests for population means, proportions, and variances. Be able to perform such tests using statistical software.

7. Power and sample size

a. Meaning of insignificance in statistics, Type I and Type II errors
b. Definition of power of a significance test
c. Power and sample size calculation of the tests for population means and proportions

Objective: Understand the two types of error in a hypothesis test and the concept of power of the statistical test. Understand the relationship between power and sample size, as well as the relationship between the power and certain characteristics of the population. Be able to perform power and sample size calculations for tests of population means and proportions using statistical software.

Block 3 (October 26 – December 18, 2020)

8. ANOVA and multiple comparisons

a. Assumptions and related statistical design
b. One-way ANOVA model
c. Various sums of squares
d. ANOVA table and F-test
e. Multiple comparisons between means
f. Repeated measures ANOVA
g. Two-way ANOVA

Objective: Understand the assumptions and basic principles underlying ANOVA; understand the completely random design. Know how to perform ANOVA using statistical software and be able to explain the ANOVA table. Understand why p-values need to be adjusted in multiple comparisons and know how to calculate them using software.

9. Simple linear regression and correlation

a. Scatter plots
b. Model: assumptions, interpretation
c. Least squares estimate
d. Evaluation of regression equation: inference concerning the coefficients, coefficient of determination
e. Prediction
f. Correlation model and Pearson correlation coefficient

Objective: Understand the population model for simple linear regression, the least squares method, regression parameter estimates, the ANOVA table, RMSE value and R-squared, and be able to produce and interpret such an analysis and make predictions using the model.

Block 4 (January 4 – February 26, 2021)

10. Multiple linear regression and correlation

a. Model
b. Least squares estimate
c. Evaluation of regression equation; inference concerning the coefficients, coefficient of determination
d. Multiple correlation model (optional)

Objective: Understand how to include more than one independent variable in a regression equation. Be able to obtain a multiple regression model and use it to make predictions. Understand how to calculate and interpret correlation coefficients.

11. Additional regression techniques

a. Qualitative independent variables
b. Variable selection procedures
c. Logistic regression

Objective: Understand how to include qualitative variables in a regression analysis. Understand how to use automated variable selection procedures to develop regression models. Be able to perform logistic regression analysis for dichotomous dependent variables; output and interpret odds ratios.

12. Chi-square distribution and analysis of frequencies

a. Chi-square distribution: definition, calculation of the probability and critical points
b. Observed versus expected frequencies and Chi-square test statistic.
c. Contingency table showing the independence of two categorical variables and test of independence
d. Contingency table showing homogeneity in proportions of two or more populations and test of homogeneity
e. Fisher's exact test
f. Relative risk, odds ratio, and the Mantel-Haenszel (optional) statistic

Objective: Understand the definition of Chi-square distribution, the procedure of constructing the Chi-square statistic using observed and expected frequencies and general steps to conduct hypothesis testing using the Chi-square statistic. Understand the concepts of independence of two categorical variables and homogeneity in proportions of two or more populations, and be able to convert either a practical association problem or a problem for examining the homogeneity in proportion of two or more populations to a statistical problem of a hypothesis test for a contingency table. Be able to perform a Chi-square test using Stata. Understand the limitations of Chi-square tests and be able to apply Fisher's exact test to the contingency table where the Chi-square test is not appropriate.

13. Survival analysis

a. Descriptive statistics for survival data
b. Log-rank test for examining the difference between two survival curves
c. Proportional hazards or Cox regression model

Objective: Understand the concepts of survival data. Understand the Kaplan-Meier procedure for estimating the survival function and log-rank test for testing the difference between two survival functions. Be able to use Stata to estimate and plot the survival function and perform the log-rank test. Learn the concepts of proportional hazards and the Cox regression model.