AMS 315, Spring 2023

Overall Format

You are responsible for the readings in your text. I will post study guides for each chapter that specify the sections that you are responsible for in each chapter and examples of potential examination questions. In general, the material discussed in class and the questions worked in lecture should be your highest priority for studying. I will also post on the class blackboard scientific papers for you to read. These papers will illustrate the application of the techniques that we are studying in this class to important biostatistical research. As you read these papers, work to understand the way that the researchers developed their research question and used statistical methods to address the question.

Opening reading assignments:

These two papers summarize the research that led to the development of the statin class of cholesterol reducing medicines and are on the class blackboard. The Tobert paper is an excellent survey of the broad range of studies that were used to identify cholesterol as a crucial agent in the development of heart disease. The discussion of the studies and their statistical analysis should be clear to you, especially at the end of the course. The discussion of the chemical properties of the medicines in the statin class are not essential to the work of this course. The Brown and Goldstein paper discusses the scientific basis of the statin class and is very heavy going. Read it quickly to get as much as possible from it. It is a summary of the authors' speech accepting their Nobel Prize.

Tobert, Jonathan A. (2003). Lovastatin and Beyond: The History of the HMG-COA Reductase Inhibitors. Nature Reviews, Drug Discovery, volume 2, 517-526.

Brown, Michael S., Goldstein, Joseph L. (1986). A Receptor-Mediated Pathway for Cholesterol Homeostasis. Science, volume 232, 34-47.

Chapter One, Statistics and the Scientific Method:

There will be no examination questions directly about this chapter. You should read and review it. It contains fundamental background material. I recommend that you focus on the schematic for the scientific method (Figure 1.1): 1. formulate research goal; 2. plan the study (specifically identify variables); 3. collect data; 4. inferences; 5. conclusions (decisions); 6. formulate new goals and return to step 2. You should learn the definition of populations and sample.

Concerning the reasons for studying statistics, your text does not mention my choices of the five major successes of statistics. These are: 1. the development of the randomized experiment (clinical trial); 2. the development of statistical procedures for identifying genes associated with diseases (R. A. Fisher was the first and most important of these researchers); 3. the development of quality control techniques; 4. the development of sample survey methodology; 5. the development of procedures to analyze data collected

in observational studies. The Tobert review paper on the class blackboard is an excellent case study of the application of genetic findings to develop the statin class of medicines and prove their effectiveness. The papers discuss the statistical analysis of observational data (e.g., data from the Framingham study) that documented the potential importance of cholesterol control.

Chapter Two, Using Surveys and Experimental Studies to Gather Data:

There will be no examinations questions directly about this material. It contains important definitions. The most important point is the distinction between a randomized experiment and an observational study. In a randomized experiment, the experimental units are randomly assigned to the settings of the factor. This has the effect of roughly balancing the distribution of each confounding factor for the experimental units assigned to each setting of the factors. Consequently, the only two explanations for an important difference are that the treatments caused the difference or there was a random error. In an observational study, one may only observe association not necessarily causation. You should know the distinctions between comparative and descriptive studies, between prospective and retrospective studies, and case-control studies. You should also know about confounding variables and have examples at hand. Other terms to master: are target population, samples population, sample, observation unit, sampling unit, sampling frame, simple random sample, and systematic sample. You should know that two problems associated with random sampling surveys are non-response and measurement problems.

Chapter Three, Data Description:

You may use a calculator but not a computer in examinations. The only way to perform the computational assignments is to use a computer and data analysis package. I recommend R, SAS, SPSS, or Minitab. For those of you who are going to seek a position as a computer oriented quantitative analyst, expertise in SAS should be helpful as a qualification for most of these positions.

Your text does not discuss scales of measurement. Wikipedia and your favorite search engine are effective tools to get a fuller definition of scales of measurement. The ***nominal scale*** of measurement is the simplest. The value of a nominal scale variable is actually a verbal characterization. For example, ethnicity or hair color are nominal variables. Traditionally, numbers are used in databases rather than the words. Pie charts, bar charts, and contingency tables are tools used to deal with a nominal scale variable. The ***ordinal scale*** of measurement has ranked values. For example, course grade with values A, B, …, is an example of an ordinal scale variable. Psychology statistics texts often recommend nonparametric statistical procedures. The common statistical procedures that you studied in AMS 310 can be viewed as an approximation to a permutation procedure. The permutation procedure is considered a gold standard statistical procedure in scientific research. In an ***interval scale***, differences have meaning. For example, the net balance in a bank account (positive for asset and negative for liability) is an example of a interval scale variable. The techniques that you studied in AMS 310 (such as t-tests) can be applied to interval scale variables. A ***ratio scale*** variable is one in which ratios have

meaning. For example, the weight of a study participant is a ratio scale variable. Often, a monotonic transformation of a ration scale variable is helpful.

I expect you to know about histograms and that they can be classified as: 1. unimodal, bimodal, or multimodal; 2. symmetric or skewed. I expect you to know the definitions of mean, mode, and median (see page 85). I also expect you to know the definition of standard deviation (which is the square root of the variance), range, percentiles, and interquartile range.

End of Guide