

This review sheet is intended to cover everything that could be on the exam; however, it is possible that I will have accidentally left something off. You are still responsible for everything in the chapters covered except anything that I explicitly say you are not responsible for. Therefore, if I left something off of this sheet, it can still be on the exam. There will be no multiple-choice questions. Most of the questions will be like the ones in the homework assignments, and possibly a few definition questions, but I am more likely to ask questions that make you use the definitions rather than recite them.

The review session will be at a time to be determined in class, probably Sunday 3/1.

For all chapters of the textbook, I will not ask for a list of items like, “What are the research ethics for the participants?” Instead, I will ask a question like, “The **research ethics** for the **participants**, requires that they be informed about their **rights** and **protections**. Why is this important?” *Therefore, reading in the book about what I list is more advantageous than memorizing the list.* In addition to the questions from the homework, the questions at the end of the chapters are helpful.

Chapters 1: Why must good research: have the **purpose of the research** be clearly defined, have the **process** clearly defined, have the **design** thoroughly planned, have **ethical standards** applied, frankly reveal its **limitations**, do an adequate analysis for the **decision maker**'s needs, have the **findings** presented unambiguously, and have justified **conclusions**? What are the advantages and disadvantages of the following types of research: **reporting, descriptive, explanatory, and predictive**?

Chapter 2: (The paragraph before Chapter 1 belongs here.) What is meant by **informed consent**? (See Page 31.) Why do some people feel **deception** is never ethical, while others say it can be done to **protect confidentiality**? What are the following parts of the **debriefing**: explain any deception, describe **hypothesis**, give **post-study results**, get post-study follow on medical or psychological conditions? How can you keep **privacy**? Why can **data mining** be bad? Why should the **sponsor**: keep confidentiality, not falsify or change data, not falsify or change the presentation, not reinterpret the data from a **biased** view, not omit data, not make recommendations beyond the **scope of the study**? Why should the **researchers** be concerned about **safety** and the **code of ethics** of their field? What does the **Institutional Review Board (IRB)** do at Bethany College and at other institutions?

Chapter 3: How is a **construct** used in research? What does it mean that they are **hypothetical** and **conceptual**? Why is a good **operational definition** so important? What is meant by the following types of **variables**: **independent** variable, **dependent** variable, **moderating** variable, **extraneous** variable, **control** variable, **confounding** variable, and **intervening** variable? What are the following types of hypotheses and when would you use them: **descriptive hypothesis**, and the two types of **relational hypotheses (correlation and causation)**? Why must a strong hypothesis be adequate for the purpose, testable, and better than rival hypotheses? What is meant by a **theory**? *This is something most people will get wrong. It is a detailed explanation of a hypothesis – it is not the opposite of fact.* Know what each of the six tenets of the **scientific method** on Page 66 mean and why they are important. Know how to do both **deductive** and **inductive reasoning**.

Chapter 4: The stages of research: Stage 1, clarification. Know how to clarify the question from a **management question**, to **research question, investigative question, and measurement question**. *Basically, you start with the question somebody asks you, and you refine it until you have a question you*

can get numbers for. Stage 2, proposing. What are the strengths and weakness of the following types research: **ex post facto evaluation, prior or interim evaluation, option analysis, and decision theory?** Basically, they are, in order, after the fact, prior to the experiment, prior but with an eye as to which way to go, similar but making rules to use in the decision making. Stage 3 designing. Here is where you decide how you are going to answer the question. You figure out what type of **sampling** you will use, what is the **target population?** Later chapters go into details on this Stage 4, data collection. What are the advantages and disadvantages of **primary and secondary data?** Stage 5, analysis, and Stage 6 reporting will be gone over in later chapters. Know why the following could cause problems with your research, **avored technique syndrome, database strip mining, unsearchable questions, and politically motivative research.**

Chapter 6: Understand what each of the steps on the **critical path method** on Page 125 mean and why you do that step. Do not worry about the numbers on the diagram. Know the following descriptors of the research and why it is important to know about each for the research. **Degree of crystallization** (is it **exploratory** or **formal**), **control of variables** (is it an **experiment** or **ex post facto?**), **purpose** (is it **reporting data, describing data, causal to explain or predict?**), **time frame** (is it **cross-sectional or longitudinal?**), the **scope** (**statistical studies** have **breadth**, but **case studies** have **depth.**), **research environment** (is it in the **field, lab, or simulation?** Note that the order listed are from the most realistic setting to the least realistic, but also from the least control over other variables to the most control.), and **participant awareness** (is the participant unaware and acting normal, aware but does not believe the research is related to what is different so may or may not change behavior or are there researcher induced changes in behavior?

Still Chapter 6, but all of this is one big topic. For explorative studies, understand why they are predominately qualitative (This is not about the quality of the research. It is about what qualities or aspects the people or products have.), **conversations** rather than **structured questions**, and often **observational**. Why is it good to start with secondary data? Why does it often have open-ended questions about what happened, what was tried, and what were the results of the trials? Why would you want to ask **newcomers, marginal people, well established people**, and people with different skills? Why is it a **two-stage process?** For descriptive studies, it is similar to explorative studies but you are trying to find relationships. For causal studies, you want to find **covariation** between the variables, a **time relationship** between them, and you want to eliminate other possible causes. The two biggest types of logic we discuss are **Mills Method of Agreement** and **Mills Method of Difference**. The former finds the only common (i.e., item in agreement) link and the latter finds the only thing that changes (i.e., is different.) Understand the differences between **symmetric, reciprocal, and asymmetric relationships between the variables.**

Chapter 7: What is meant by a **qualitative study?** Why might it be **subjective** and **easily biased?** How can it give insights not gotten otherwise? How does the list on the bottom of Page 145 help ensure the weaknesses are minimized? Note that several of the items on the checklist include words like “carefully” or “thoroughly.” Know how to clarify the question (See Chapter 4). Why may that mean participants may do **pre-tasking?** Why is the sampling not **random**, rather it is “**non-probability sampling**” like **purposeful sampling, snowball sampling, or convenience sampling?** What are the advantages and disadvantages of **interviews** which are **unstructured, semi-structured, and structured?** Why is the interviewer often a guide? Why do **pre-interview screening?** How do the **projective techniques** on Pages 155 - 156 work and why would you want to do them? Why do **in depth interviews** pay \$1/minute for consumers and more for skilled professionals? What are the advantages and disadvantages of **homogeneous vs. heterogeneous** and **experts vs. non-expert** group interviews?

Chapter 14: What do these terms from **sampling** refer to: **population, census, population element, and sample frame**? How does sampling mean low cost, may be more accurate, increases speed, and may use available elements, when compared to a census? Why should we avoid **systematic variance**? Why should we calculate the **sampling error**? (*You will not have enough information to do the calculation from this course alone.*) Know what the following steps of designing a sample mean? 1) Find the **target population**. 2) What are the **parameters of interest** (*what data to collect and what part of the population is relevant*)? 3) When finding the sample frame, why does desired the **sample size** increase when the **population variance, desired precision, and the number of subgroups** increase. 4) Why might you do each of the following types of probability sampling? **simple random, systematic sample, stratified random sample (proportionate and disproportionate), or cluster sampling**? Non-probability sampling was discussed in Chapter 7, but is mentioned here again. 5) Size determination is beyond the scope of this course.

Chapter 15: Why must you **edit data**? What is **field editing**? Why do you **pre-code** for **closed questions** but **post-code** for **open-ended questions**? With pre-coding, why do you need a code for missing? For post-coding, why must the codes be **appropriate, exhaustive, and mutually exclusive**? A **proximity plot** like on Page 388 will help with this. Why might a **don't know** answer be a problem and why might it not be? Why might missing data in a longitudinal study mean you eliminate that **element**? Why might missing data mean you eliminate the variable from all elements? Why might you replace missing data with an estimated value? Know how to enter data. Basically, it is using a **keyboard, OCR software, or voice recognition software** to put it into a **spreadsheet**.

Chapter 16: Much of this is covered in the early labs. Know how and why you would use **frequency tables, bar charts, pie charts, histograms, and stem & leaf tables**. Know when to do the following with tables **round data, sort, averages** (but not if it is a **bimodal distribution**), use percentages, **common scales**, use blank spaces, summarize, and always use labels. How do **Pareto diagrams** help information to be relayed? Understand how to read a **box plot** and **cross tabulation**.

Labs: I cannot ask you to do the computer work. Read each lab for the types of questions I could ask.

Lab 1: For this lab, I will ask you the advantages of **line charts, pie charts, and bar charts**.

Lab 2: Know what each of the terms in descriptive statistics mean. I will give you the a printout of the **descriptive statistics** and you have to tell me what the data looks like. The key terms are **mean, standard error** (the 65% probability the real mean is within that distance of the mean), **median, mode, standard deviation** (65% probability the data will be within that distance of the mean), **kurtosis** (<0 is more peaked, >0 is flat), **skewness** (the sign tells you which way the tail points), **range, maximum, minimum, count, largest (K), smallest (K), and confidence level**.

Lab 3: What do **scatter diagrams** and **histograms** tell us? I will give you one of them.

Lab 4: What do the following terms mean **mutually exclusive, collectively exhausted, independent, and conditional probability**? I will give you the following formulas $\text{permut}(\text{number}, \text{number_chosen})$, $\text{combin}(\text{number}, \text{number_chosen})$ $\text{binom.dist}(\text{number_s}, \text{trials}, \text{probability_s}, \text{cumulative})$, and $\text{hypgeom.dist}(\text{sample_s}, \text{number_sample}, \text{population_s}, \text{number_pop})$. I will then give you a scenario and ask you which formula to use and what numbers would go in each spot. (Just like in the lab, except you do not type it in.) Use **permutations** for when the order matters and **combinations** for when the

order does not matter. Use **binomial** for independent events and **hypergeometric** for dependent events. Cumulative tells you up to and including that number. Therefore, if you want $> x$, you do $1 - \text{distrib}$ where distrib is either hypergeometric or binomial, have cumulative be true, and the number you put in is one less than x because you want to eliminate all less than x .

Lab 5: I will give you the following formulas: `normdist(x, mean, standard_dev, cumulative)` and `norminv(probability, mean, standard_dev)`. Use the former to find out the percentage of values in a **continuous normal distribution** are below x . (Should always have `cumulative = true`.) The latter is used to find at which value, the percentage (probability) are below that value. I will ask questions like I did in the lab and for Lab 4.

Non-graded Homework Assignment #4A to be reviewed with Assignment #4.

- 1) (10 points) What is meant by “editing data”? And why would you do it?
- 2) (5 points) Why must the coding for a post-code be mutually exclusive?
- 3) (10 points) When would you decide to eliminate a variable from all elements? Why did you choose that?
- 4) (15 points) What is meant by a “stem and leaf table”? What are the advantages and disadvantages of it over a histogram?
- 5) (10 points) Draw a box plot for data set which has a negative skew. State how your plot shows that.
- 6) (10 points) What is the advantage of doing cross tabulation?