

**Introduction to Statistics**  
**Dr. Lee**  
**Review for Midterm Exam**

The midterm exam will consist of 15 “matching” items covering definitions of key concepts, 25 multiple choice questions, and 4 problems. Each of the 40 “choice” items will be worth 2 points. The problems will be worth 5 points each. The exam will cover chapters 1 – 5 in Asquith, Pyrczak sections 1 – 6, 8 – 11, and 18 – 21, and class lectures and discussions up through the introduction to “Significance Tests.” The key concepts in the readings are a “must study.” I will limit the “matching” questions to those concepts that are in bold in Asquith. These concepts will also be covered in the rest of the exam.

Below is a list of study comments. If you develop the knowledge and skills these comments focus on, you should do well on the exam. I encourage you to answer these on your own, and then to share your answers in study groups. Beyond studying the topics below, I recommend that you puzzle through some of the exercises at the end of each chapter of Asquith’s book.

Do not study the binomial distributions discussed in Asquith chapter 3.

1. Be familiar with how sociologists approach research. What things do good researchers try to do? What is the point of doing research? What role does theory play in research? How important are theories to sociology and to conducting good research? Of course, you should also be aware of the role that statistics play in sociological research.
2. Why not conduct a census every time you collect data?
3. Be sure you understand the different levels of measurement. Why is it important to know the levels of measurement?
4. You should be able to draw (sketch really) and interpret the various types of graphs.
5. You should see the direct, logical connection between histograms and probability distributions.
6. Make sure you understand the meanings and qualities of the measures of central tendency (mean, median, and mode). **Be sure you can calculate them using data.** How effective is each at accurately describing central tendencies of distributions? How do outliers (or even a skew in the data) affect these? Can you draw a skewed distribution? Definitely study the inflating or deflating influence that a skew (or even an outlier) has on the mean and standard deviation?
7. Yes. You need to be able to calculate the measures of variation, and **know the formulas for variance and standard deviation (know the five steps).** You should concentrate on the conceptual formulas rather than the computational formulas in the book (computational formulas will not be on the exam). You are expected to know the components of the standard deviation, the logic behind its formula and its use, and the different formulas for populations versus samples. What are the properties of the standard deviation? How does one interpret its magnitude? Be able to say what standard deviation means in English.
8. What is a normal curve, and what is a z-score? **Be able to calculate the numbers that will be used on a flippy ruler (z-scores).  $z = Y - M/s.d.$**
9. You should have the empirical rule committed to memory, understand how it can be used to provide proportions of cases at each scale value for a variable, and that it is used only for

normal distributions. Also, make a distinction between the empirical rule for normal curves and the actual precise measurements of normal curves.

10. You should understand that the mean and standard deviation are concepts that can be used to describe normal distributions in lieu of drawing them out.
11. The normal distribution is amazing in how it helps us understand the world. Be glad we have it! You should see the many ways it helps us. This means seeing that all normal curves, regardless of steepness or width, share the same properties such as relationship of standard deviation to area under the curve. Work through a few exercises where you have to find areas under the curve that correspond to particular z-scores. Find z-scores for areas under the curve. Also, just what is a z-score? If the data for a case on two different variables have the same z-score, what does that mean in terms of the case in relation to the population (e.g., percentile rank) and in terms of comparing the scores across the two variables (e.g., comparing SAT and ACT scores)?
12. The sampling distribution is an extremely important concept to master. We don't ever do it in practice, but what would happen if we recorded the mean for a variable when sampling with a particular sample size from a population and repeatedly re-sampling from that population until we exhausted all possibilities, and created a frequency polygon of the measures? What relationship would that polygon have to the mean of the variable in the population? What shape would the polygon take?
13. Study how the properties of sampling distributions can give us a way to gain insight into population parameters by merely using sample data.
14. Be sure that you can say what the standard error is in English. **Be able to calculate the standard error:  $se = s/\sqrt{n}$**
15. What does the central limit theorem say? Why are thirty cases usually enough with which to make inferences? In general, how would sampling distributions be affected for larger versus smaller sample sizes in repeated sampling (e.g., 500 cases versus 30 cases)?
16. Do you know what unbiased and efficient estimates are?
17. How does the shape of the underlying population distribution affect the shape of a sampling distribution when sample sizes are above 30? Under 30?
18. **You should be proficient at creating and interpreting confidence intervals.  $CI = M \pm z(s.e.)$ .** What determines the width of confidence intervals? Be able to explain in simple English what a confidence interval is?
19. How do you create and interpret confidence intervals for proportions?
20. Do you understand the logic of significance tests? What are the assumptions of significance tests?
21. What is a research hypothesis and what is a null hypothesis? Do you see how null hypotheses represent guesses? Do you understand that the guess represents the parameter of a hypothetical population?
22. What is a P-Value and how does it relate to sampling distributions? Be sure that you can determine a P-Value for significance tests.
23. Who determines the "significance" of a test? Where does the  $\alpha$ -level come from?
24. What is a null hypothesis? When do you reject the null?

*Please contact me if you have any problems understanding the material.*

*Best of luck on the exam. I hope that you do very well.*