# Thought Questions for Midterm

**The questions listed below are designed for discussion and preparation before the midterm. When reviewing these questions:**
1. **Illustrate your points with specific examples**
2. **Be as complete as you can with your definitions, in other words, be as specific as you can and do not leave anything up to the reader (be exclusive and inclusive).**

1. Why do we make distinctions between samples and populations in statistics?
2. Discuss the differences between the descriptive and the inferential approaches in statistics. Is one "better" than the other? Are they competitive or complementary? Illustrate the kind of situation in which each approach is appropriate.
3. Discuss the use of exploratory data analysis. Illustrate with an example.
4. What is statistical visualization? Why do we use it? What is its limitation?
5. Discuss histograms, frequency plots, box plots, stem-and leaf plots. What are their strengths and weaknesses? What do we use each for?
6. Discuss the phrases "Being a Statistician is not having to be certain"; "Uncertainty - Something you can always count on", and "Statistics: Stability from variability."
7. What is the point of constructing frequency distribution tables and graphs?
8. What are the properties of a random sample?
9. Know the differences between each type of measurement scale (nominal, ordinal, interval, ratio). Illustrate each type with an example. What measures of central tendency and variability are appropriate for each type?
10. Know how to make a cumulative frequency distribution table. Be able to describe its contents and applications.
11. What is the relation between the median, percentiles, and the box plot?
12. What is skew? Kurtosis? Define and illustrate each with an example.
13. Compare and contrast the type of information you get with relative frequencies versus cumulative frequencies.
14. Why to statisticians care about the distribution of their data?
15. Is it enough to know general descriptive information about a distribution (i.e., the range, skewness, mean, etc.)? Why or why not?
16. What are outliers? What impact do they have on how you describe your data?
17. What are some of the special qualities of the mean?
18. What is variability? Discuss measures of variability. What are their strengths and weaknesses? What is its relation to quality of the product/service?
19. What is the sum of squares and sum of products short-cut formulas?
20. Why do we standardize data?
21. What does a Z-score tell you (i.e., how do you interpret one)? How do you convert a raw score to a Z-score?
22. What is the normal distribution? What is the standard-normal distribution?
23. How do you use the normal distribution table to find the percentage of the population that is expected to fall between two points or beyond or below one point in the distribution?
24. What are some of the benefits of using the normal distribution as a model for data at the population level?
25. What does the central limit theorem tell us?
26. What does "independent events" mean?
27. What is the purpose of inferential statistics? How is it different than descriptive statistics? Are they complementary or competitive?
28. Why do we care about probability at all when we are talking about inferential statistics?
29. What are the four steps involved in hypothesis testing?
30. What does specifying the null hypothesis mean? What about the alternative hypothesis? What is the benefit of being so specific about the hypotheses?
31. As discussed in class, what are the differences schools of probability concepts?

32. With inferential statistics, the goal is to reject the null hypothesis. What does this mean? Do we conclude that the alternative hypothesis is correct? Why or why not?
33. Why is the standard error of the mean (based on many samples) going to be smaller than the standard deviation of a single sample? In explaining your answer, be sure to describe the interpretation of a standard error of the mean.
34. What types of error can occur when making decisions based on test of hypothesis? Be specific.
35. Why are points beyond + or - 2 standard deviations often considered outliers by some researchers?
36. What does it mean if a researcher sets his/her alpha at say .01, and rejects the null hypothesis? How does this differ from setting the alpha at .05 and rejecting the null? In which case is the researcher going to be most likely to reject the null hypothesis?
37. What are degrees of freedom, and when do you use them?
38. What is the difference between a one-tailed and a two-tailed test? When would you use each (if at all)?
39. What is meant (and not meant) when a researcher says a finding is "statistically significant"?
40. What is the generic formula for the T-Test? Explain it conceptually (not mathematically).
41. What is the relationship between T-Tests and F-Tests?
42. What is the generic formula for the F-Test? Explain it conceptually (not mathematically). How does this relate to the generic formula for the T-Test?