Review Questions for Exam 1

1. Mammography is estimated to be about 85-90% accurate in detecting abnormalities that may be caused by breast cancer. If, in a given year, mammograms detected breast abnormalities in 1000 women, would it be correct to estimate that about 850-900 of these were true abnormalities? Explain your answer.

2. Identify three words or phrases that have substantially different meanings when used in a statistical context to when they are used in regular English. Explain the statistical meaning and how it may differ from the colloquial meaning.

3. Explain the terms “Sample” and “Population” with reference both to clinical trials and experimental lab science.

4. Define the term “confidence interval”. Use an example to help explain what it means.

5. Define the four types of variable: Nominal, Ordinal, Interval, and Ratio. Give examples of each.

6. Does it make sense to talk about a 10% increase in body mass (weight)? What about a 10% increase in body temperature? Explain carefully in both cases.

7. Explain the three measures of central tendency (“averages”), including interpretation and how they are calculated. Explain advantages and disadvantages of each and when each would be appropriate for use.

8. Explain the four measures of variability, including interpretation and how they are calculated. Explain advantages and disadvantages of each and when each would be appropriate for use.

9. Explain some general principles in presenting data graphically. Give examples of specific plot types and how well they meet these principles.

10. Explain what a column scatter plot is; use one or more examples to demonstrate. When is a column scatter plot useful?

11. Explain what a box-and-whisker plot is; use one or more examples to demonstrate. When is a box-and-whisker plot useful?

12. Explain what a bar chart is; use one or more examples to demonstrate. When is a bar chart plot useful?

13. Explain what a distribution is. Give your answer both in terms of a sample distribution and a population distribution.

14. Why is the normal distribution important in statistics?

15. Explain how the t-distribution is different to the normal distribution. Your answer should reference the central limit theorem.

16. Give three ways of calculating error bars when plotting bar charts. Explain the interpretation of what they represent and discuss pros and cons of using each.

17. What is the Standard Error of the Mean? Explain both how it is calculated and interpreted.
18. What is hypothesis testing from a statistical perspective?
19. What is a p-value?
20. Discuss the statement “The p-value was 0.04, so the result is significant”
21. Explain the terms “Type I Error” and “Type II Error” in the context of hypothesis testing.
22. Discuss the statement “The p-value was 0.04, so there is only a 4% chance the null hypothesis is true.”
23. Explain the difference between the False Discovery Rate and the p-value. You should include mention of the prior probability.
24. In investigating a potential new cancer therapy, which appears to have fewer side effects than existing treatments, as a first step an investigator performs a mouse model experiment. In this experiment, tumor growth is initiated in six mice. One group is treated with the standard treatment and one with the existing treatment. The investigator measures tumor size after one week of treatment and compares the two groups using a t-test. Because the p-value is large (p=0.27), the investigator concludes that there is no difference in the two treatments in their ability to restrict tumor growth. Is this interpretation correct? Explain.
25. In the scenario in question 24, what is the correct way to test whether the new treatment is at least as effective as the standard treatment?
26. What is statistical power?
27. When is it useful to compute statistical power? When is it not useful?
28. What is Multiple Hypothesis Testing? Why is it problematic?
29. What is the Bonferroni Correction (or equivalently, the Bonferroni method)? When is it used? What is the interpretation of statistically significant results when a Bonferroni correction is used?
30. What is a family-wise error rate?
31. Explain the Benjamini-Hochberg method. When is it used? How does it work? What is the interpretation of statistically significant results when it is used?
32. Explain why p-values are difficult to interpret in the context of microarray or other high-throughput experiments.
33. Explain the terms skewness and kurtosis.
34. Why do statisticians perform tests for normality on data? What are the limitations of these tests?
35. Is it acceptable to remove outliers from data before analyzing it? If so, how should it be done? If not, why is it not acceptable?
36. What is a one-class T-test? When is it useful (give an example)? How are the results interpreted?
37. What is an unpaired T-test? When is it useful (give an example)? How are the results interpreted?
38. What is a paired T-test? When is it useful (give an example)? How are the results interpreted?
39. What is the assumption of equal variances? Discuss how to deal with the possibility this assumption may not be true.
40. If the error bars on a bar chart overlap, what may be said about the statistical significance of testing the values in the bar chart are different?
41. If the error bars on a bar chart do not overlap, what may be said about the statistical significance of testing the values in the bar chart are different?

42. In a lab experiment, cells are cultured and then treated with either an agent or control. The experimental design used is that four plates of cells are grown independently, cells are then harvested from each plate and split into two, with one group of cells receiving the treatment and one the control. The data are analyzed using a paired t-test. Will a paired t-test always be more statistically powerful in this scenario? (Hint: think about how a paired t-test works, and think about sample sizes.)