Homework 2

Instructions: This homework is due in class on Friday Apr 19.

Please read the following guidelines for presenting your work and follow them diligently.

- Write your full name clearly on the top right of the first page. **Staple** pages on the left hand corner. Write neatly in complete sentences.
- You are required to work all the problems, however, only 5 will be graded. The page numbers below refer to the fifth edition of the text.
- Answer the questions in the order in which they are posed. Clearly number the questions as I have.
- You must work independently on the homework. Please post questions on the discussion board or come to office hours.
- Be sure to show/explain your work thoughtfully. How you write your answers is important.
- If you use R to make plots or as a calculator, it is enough to simply include the output (e.g., appropriately labeled plot) in the main part of your homework without the R code.

---

1. Wassermann, L. *All of statistics: inference* Problem 3 on page 200 in chapter 11.
2. Wassermann, L. *All of statistics: inference* Problem 10 on page 203 in chapter 11.
3. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(0, \sigma^2)$ where σ^2 is unknown. Give the form of the M.P test of size α for testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 = \sigma_1^2$ where $(\sigma_1^2 < \sigma_0^2)$. Is this test Uniformly Most Powerful for testing $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 < \sigma_0^2$? Why or why not?
4. (From the *Fundamentals of Biostatistics* by Bernard Rosner) The degree of clinical agreement among physicians on the presence or absence of generalized lymphadenopathy was assessed in 32 randomly selected participants from a prospective study of male sexual contacts with AIDS or an AIDS-related condition (ARC).

   The total number of palpable lymph nodes was assessed by each of two physicians on each patient. The data can be found in arc.csv.

   (a) Does there appear to be an association between the two assessments? Make a scatterplot with clearly labeled axes and comment on the picture. If there is a relationship, is it linear?

(b) How might you formally determine whether there is a systematic difference between the assessments of Doctor A vs. Doctor B? Clearly state the model assumption, test you are using, decision rule, etc. Should a one sided or a two sided test be performed?

(c) Perform the test at $\alpha = 0.05$ and also report a P value. (You may copy and paste output from R so long as you adequately preface it with a few sentences describing what you are doing.)

(d) Suppose the results show no significant difference between the assessments made by the two physicians. Does this necessarily mean that this type of assessment is highly reproducible? Why or why not?

5. Suppose $(X_1, X_2, \ldots, X_5)$ is a multinomial random variable with probabilities $p_1, p_2, \ldots, p_5$ respectively, where $\sum_{j=1}^{5} p_j = 1$ and $p_j \geq 0$ for each j. Let $\theta = (p_1, p_2, \ldots, p_5)$. The likelihood function is given by:

$$L(\theta|x_1, x_2, x_3, x_4, x_5) = \frac{n!}{x_1!x_2!\ldots x_5!} \prod_{j=1}^{5} p_j^{x_j},$$

where $n = x_1 + x_2 + \cdots + x_5$ denotes the number of trials.

Consider testing $H_0 : p_1 = p_2 = p_3$, and $p_4 = p_5$ versus $H_1 : H_0$ is not true.

(a) Show that

$$-2\ln(\lambda(\mathbf{x}) = 2\sum_{j=1}^{5} x_j \ln\left[\frac{x_j}{m_j}\right],$$

where $m_1 = m_2 = m_3 = (x_1 + x_2 + x_3)/3$ and $m_4 = m_5 = (x_4 + x_5)/2$.

(b) Under H_0 the statistic $-2\ln(\lambda(\mathbf{x}) \sim \chi^2_\nu$. What is ν? (*Hint:* there is only 1 free parameter unnder H_0. Why?)