Homework 7

Instructions: This homework is due in class on Monday June 3. This is your LAST homework.

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 first work independently on the homework. Please post questions on the discussion board or come to office hours once you have tried the problems.
- 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. The data calcium.txt gives the measurements in milligrams of daily calcium intake for 38 women between the ages of 18 and 24 years. Suppose we preume a normal model for these data with unknown mean θ and known $\sigma^2 = 500$. Assess the goodness of fit of this model by examining
   (a) an index plot of the standardized residuals
      - state formula for standardized residuals
      - give index plot
      - comment

   (b) QQ plot of the standardized residuals
      - state explicit formula that show how the values on the x and y axis can be calculated
      - give QQ plot
      - comment

(c) a discrepancy statistic

- give the formula for the statistic
- its distribution under the presumed model
- its observed value and conclusion

(d) χ^2 goodness of fit using the intervals $(-\infty, 600], (600, 1200], (1200, 1800]$ and $[1800, \infty)$.

- give expressions for multinomial probabilities based on assumed model
- give a graph of the multinomial likelihood as a function of θ.
- calculate MLE based on the multinomial likelihood
- χ^2 test statistic calculation and p-value based on the appropriate reference χ^2 distribution.

Your Rcode can be included in an appendix, but plots should be in the main body of your homework. The main body of your homework must also clearly demonstrate your understanding of what is being calculated step by step as indicated above and also your interpretations.

2. It would be more realistic in problem 1 to assume that σ^2 is unknown. Re-do the χ^2 goodness of fit test and report your conclusion. (Please show expressions for the multinomial probabilites under the presumed model, MLEs based on the multinomial likelihood, etc. to demonstrate your understanding of the various steps involved. Rcode can be in an appendix. Plots should be in the main body of the homework.)

3. Let $X_1, X_2, \ldots, X_k$ follow a Multinomial distribution with parameters n and probabilities $p_1, p_2, \ldots, p_k$. Suppose that $k = 5$ and we want to investigate if the following model fits our data: $p_1 = p_2 = p$, and $p_3 = p_4 = p_5 = q$.

(Note: since the multinomial probabilities must add to one, $2 \times p + 3 \times q = 1$ and therefore there is only one parameter to be estimated in the presumed model.)

(a) Show the MLE of p is

$$p \;=\; \frac{X_1 + X_2}{2\,n}.$$

(b) Conduct a χ^2 goodness of fit test for the data: $x_1 = 10, x_2 = 20, x_3 = 20, x_4 = 20,\ and\ x_5 = 30$. Use an $\alpha = 0.05$.

4. Suppose we are given independent P-values $P_1, P_2, \ldots, P_m$ for testing m different hypotheses H_{0i}.

(a) Show that the distribution of the $V = \min_i P_i$ when all the H_{0i} are true is a $Beta(a = 1, b = m)$. That is,

$$f(v) \;=\; m(1-v)^{m-1}, \quad 0 < v < 1.$$

(b) Suppose we reject all H_{0i} for which $P_i < t$. Define the event

$$A_i = \text{ith hypothesis is rejected.}$$

Show that the probability of at least one false rejection when the null hypothesis is true is equal to $1 - (1 - t)^m$. More specifically:

$$P(\cup A_i) = 1 - (1 - t)^m,$$

where the probability is calculated assuming H_{0i} holds for each i.

(c) Find the t that makes the probability in part (b) exactly equal to α. How does this compare to Bonferroni's rule?

5. In a study that tested associations of 25 dietary variables with mammographic density, an important risk factor for breast cancer, the p-values in *problem4.txt* were obtained.

(a) Give the Bonferroni rule for ensuring a family wise error rate $\alpha = 0.05$ and use it to identify significant associations.

(b) Give the FDR rule for ensuring a false discovery rate $\alpha = 0.05$ and use it to identify significant associations.