STAT120C Homework 7

Assigned Wednesday May 30th, 2019
Due Friday June 7th, 2019 by 5pm in the Dropbox in DBH

Problem 1. *Weighted Least Squares.* Problem 14.9.7 in Rice.

Problem 2. Let $Y = (Y_1, Y_2, Y_3)$ follows a multinomial distribution with n trials and probabilities $p = (p_1, p_2, p_3)$. Noet that the sum of p_i's is 1, i.e., $p_1 + p_2 + p_3 = 1$. The probability mass function (pmf) is

$$Pr(Y = (y_1, y_2, y_3)) = \frac{n!}{y_1!y_2!y_3!} p_1^{y_1} p_2^{y_2} p_3^{y_3}$$

Here y_1, y_2, y_3 are nonnegative integers that satisfy $y_1 + y_2 + y_3 = n$.

(a) Show that Y_1 follows $Binomial(n, p_1)$ by showing that

$$Pr(Y_1 = y_1) = \sum_{y_2, y_3} P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = \frac{n!}{y_1!(n - y_1)!} p_1^{y_1} (1 - p_1)^{n - y_1}$$

Hint: the Binomial theorem is useful:

$$(a + b)^n = \sum_{x=0}^{n} \frac{n!}{x!(n - x)!} a^x b^{n-x}$$

(b) Prove that $Cov(Y_1, Y_2) = -np_1p_2, Cov(Y_1, Y_3) = -np_1p_3, Cov(Y_2, Y_3) = -np_2p_3$.

Hint: $E[Y_1Y_2] = \sum_{y_1,y_2,y_3} y_1y_2Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$. Show that it equals $n(n - 1)p_1p_2$. The trinomial theorem is useful: $(a + b + c)^n = \sum_{x+y+z=n} \frac{n!}{x!y!z!} a^x b^y c^z$.

Problem 3 (**Modified from 13.8 of Rice with cell values changed**) Adult-onset diabetes is known to be highly genetically determined. A study was done comparing frequencies of a particular allele in a sample of such diabetics and a sample of nondiabetics. The data are shown in the following table:

	Diabetic	Normal
Bb or bb	3	1
BB	5	6

(a) Are the relative frequencies of the allele significantly different in the two groups? State your hypotheses, test statistic, significance level and whether you should reject your null based on Fisher's exact test.

(b) Why is the Pearson Chi-squared test not recommended in this situation?

Problem 4 Suppose that 300 persons are selected at random from a large population, and each person in the sample is classified according to blood type: O, A, B, or AB, also according to Rh: positive or negative. The observed numbers are given below.

	O	A	B	AB
Rh+	82	89	54	19
Rh-	13	27	7	9

(a) Conduct a Pearson's chi-square test (at level $\alpha = 0.05$) to test the hypothesis that the two classifications of blood types are independent.

(b) Confirm your calculation in (a) using R.

```
> rhp = c(82, 89, 54, 19)
> rhn = c(13, 27, 7, 9)
> chisq.test(rbind(rhp, rhn), correct=F)
```

(c) Calculate the likelihood ratio statistic for testing independence. To do so, first calculate the maximized likelihood under the full model, i.e., the model with no constraint. Denote it by L_1. Second, calculate the maximized likelihood under the reduced model, i.e., the model assumes independence. Denote it by L_0. Third, calculate $2(log(L_1) - log(L_0))$.

(d) Compare the test statistic in (c) to Pearson's chi-square statistic. Under the null hypothesis of independence, the likelihood ratio statistic follows a chi-squared distribution with three degrees of freedom. Based upon the likelihood ratio statistic, would you reject the null the hypothesis at level $\alpha = 0.05$?