STA4030: Categorical Data Analysis

Assignment 1

Due Date and Time: September 30, 2020 (Wednesday), 10:00PM

INSTRUCTION:

- Please scan your answers in **one single .pdf file** and submit via Blackboard System.
- **Late submissions** will receive a mark of zero.
- Students may discuss set problems with others, but your final submissions must be your own work.
- All these questions should be answered using a pen, paper, calculator (good practice for your midterm and final).
- You may use any software you like, e.g., R, Python, Excel, etc., to find the percentiles regarding relative distributions (for example, to find p-values).
- Show and write down your solutions in detail and clearly.

Problem Set 1:

1. (Exercise 1.1 from Agresti (2013)) Identify each variable as nominal, ordinal, or interval.

   (a). UK political party preference (Labour, Liberal Democrat, Conservative)

   (b). Anxiety rating (none, mild, moderate, severe, very severe)

   (c). Patient survival (in number of months)

   (d). Clinic location (London, Boston, Madison, Rochester, Montreal)

   (e). Response of tumor to chemotherapy (complete elimination, partial reduction, stable, growth progression)

   (f). Favorite grocery store for UK residents (Sainsbury, Tesco, Waitrose, other)

2. (Exercise 1.9 from Agresti (2013)) In an experiment on chlorophyll inheritance in maize, for 1103 seedlings of self-fertilized heterozygous green plants, 854 seedlings were green and 249 were yellow. Theory predicts the ratio of green to yellow is $3:1$. Test the hypothesis that $3:1$ is the true ratio using the Wald, Score and Likelihood Ratio tests. Report the P-values, and interpret.

3. Table 1 gives a random sample of size 150 of the random variable X. Do you think X follows the Poisson distribution? Use Pearson's chi-squared test and the likelihood ratio test with ($\alpha = 0.05$).

Values of X	0	1	2	3	4	5	6	7	8	9
Frequency	5	11	18	26	29	25	15	10	7	4

Table 1: Poisson Data.

4. For a given sample proportion p and standard normal percentile $z_{\alpha/2}$ show that the end points of the $100(1-\alpha)\%$ two-tailed Score confidence interval for binomial parameter π are given by the solutions of the following equation,

$$(1 + z^2_{\alpha/2}/n)\pi^2 + (-2p - z^2_{\alpha/2}/n)\pi + p^2 = 0.$$

Find these solutions and thus derive the so-called Wilson confidence interval for π.

5. The data in the Table 2 is obtained from a multinomial distribution.

Cell	1	2	3	4	5
Probability	π_1	π_2	π_3	π_4	π_5
Frequency	10	13	23	21	29

Table 2: Multinomial Data.

(a). Test with $\alpha = 0.05$ the null hypothesis $H_0 : \ \pi_1 = 0.1, \ \pi_2 = 0.1, \ \pi_3 = 0.25, \ \pi_4 = 0.25$ by using the Pearson chi-square test and the likelihood ratio test.

(b). Derive the maximum likelihood estimates of π_i, $i = 1, 2, \ldots, 5$ under the null hypothesis $H_0 : \ \pi_1 = \pi_2, \ \pi_3 = \pi_4$.

(c). Test with $\alpha = 0.05$ the null hypothesis $H_0 : \ \pi_1 = \pi_2, \ \pi_3 = \pi_4$ by using the Pearson chi-square test and the likelihood ratio test.

6. (Exercise 1.11 from Agresti (2007)) When a recent General Social Survey asked 1158 American adults, "Do you believe in Heaven?", the proportion who answered yes was 0.86. Treating this as a random sample, conduct statistical inference about the true proportion of American adults believing in heaven. Summarize your analysis and interpret the results in a short report of about 200 words.

THE END