THE CHINESE UNIVERSITY OF HONG KONG
Department of Statistics

STAT4006: Categorical Data Analysis
Problem Sheet 1

The deadline for this Problem Sheet is 5.30pm on Monday 5th October. Please submit your solutions via the link provided on the course Blackboard page - if you must submit your solutions in hard copy, please contact me at jawright@sta.cuhk.edu.hk in advance. **No late submissions will be accepted. A late submission will receive a mark of zero.** Students may discuss set problems with others, but their final submissions must be their own work.

Please answer the following problems. All these questions should be answered using a pen, paper, calculator (good practice for your midterm and final). That said, you may use any software you like to find percentiles (i.e. for finding p-values). Show your working.

1. **(Exercise 1.1 from Agresti (2007))** In the following examples, identify the response variable and the explanatory variables.
   (a) Attitude toward gun control (favor, oppose), Gender (female, male), Mother's education (high school, college).
   (b) Heart disease (yes, no), Blood pressure, Cholesterol level.
   (c) Race (white, nonwhite), Religion (Catholic, Jewish, Protestant), Vote for president (Democrat, Republican, Other), Annual income.
   (d) Marital status (married, single, divorced, widowed), Quality of life (excellent, good, fair, poor).
2. **(Adapted from 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 $2:1$. Test the hypothesis that $2:1$ is the true ratio using the Wald, Score and Likelihood Ratio tests. Report the p-values, and interpret.
3. **(Exercise 1.26 from Agresti (2013))** A binomial sample of size n has $y = 0$ successes.
   (a) Show that the confidence interval for π based on the likelihood function is $[0, 1-e^{-z^2_{\alpha/2}/2n}]$. For $\alpha = 0.05$, use the expansion of an exponential function to show that this is approximately $[0, 1.92/n]$.
   (b) For the score method, show that the confidence interval is $[0, z^2_{\alpha/2}/(n + z^2_{\alpha/2})]$.
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 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 1 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	21	23	29

Table 1: Multinomial Data

   (a) Test with $\alpha = 0.05$ the null hypothesis $H_0 : \pi_1 = 0.1, \pi_2 = 0.1, \pi_3 = 0.2, \pi_4 = 0.35$ by using the Pearson chi-square test and the likelihood ratio test.
   (b) Derive the maximum likelihood estimates of $\pi_i, i = 1, \ldots, 5$ 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. Table 2 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	29	26	25	15	10	7	4

Table 2: Poisson Data

THE END