STA4030 Assignment 1

Hong Haoyang 118010096

Due Date: September 30, 2020

Problem 1

Solution:

(a) Nominal variable.

(b) Ordinal variable.

(c) Interval variable.

(d) Nominal variable.

(e) Ordinal variable.

(f) Nominal variable.

Problem 2

Solution:

According to the problem, given a sample of 1103 i.i.d Bernoulli random variables with probability of success π. The two hypotheses are H_0: $\pi = 0.75$, H_1: $\pi \neq 0.75$.

I. Wald test:

Wald test statistic:

$$z = \frac{\hat{\pi} - \pi}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} = \frac{\frac{854}{1103} - 0.75}{\sqrt{\frac{854}{1103}(1 - \frac{854}{1103})/1103}} = 1.9266$$

Under H_0, $Z \sim N(0,1)$, then, p-value $= 2(1 - \Phi(1.9266)) = 0.0540 > 0.05$

Hence, fail to reject H_0 at 5% significant level.

II. Score test:

Score test statistic:

$$z = \frac{\mu(\pi_0)}{\sqrt{\iota(\pi_0)}} = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \frac{\frac{854}{1103} - 0.75}{\sqrt{0.75(1-0.75)/1103}} = 1.8601$$

p-value $= 2(1 - \Phi(1.8601)) = 0.0629 > 0.05$

Hence, fail to reject H_0 at 5% significant level.

III. Likelihood Ratio test:

The likelihood function is

$$L(X, y) = \pi^{\sum_{i=1}^{n} y_i} (1-\pi)^{1-\sum_{i=1}^{n} y_i}$$

Then, $\ell_0 = L(\pi_0; y)$. Because the maximum likelihood estimator of π is $\hat{\pi} = \sum_{i=1}^{n} \frac{y_i}{n}$, then $\ell_1 = L(\hat{\pi}; y)$. Hence, the likelihood ratio test statistic is given by:

$$\begin{aligned}
-2log(\frac{\ell_0}{\ell_1}) &= 2log(\frac{\hat{\pi}}{\pi_0})\sum_{i=1}^{n} y_i + 2(n - \sum_{i=1}^{n} y_i)log(\frac{1-\hat{\pi}}{1-\pi_0}) \\
&= 2log(\frac{854/1103}{0.75})854 + 2(1103 - 854)log(\frac{1 - \frac{854}{1103}}{1 - 0.75}) \\
&= 3.5390
\end{aligned}$$

Because the LR test statistic has the χ_1^2 distribution, then the p-value = 0.059940.05.

Hence, fail to reject H_0 at 5% significant level.

In short, all three methods reject H_0 at 5% significant level.

Problem 3

Solution:

According to the problem, given the test $H_0 : X$ follows the the Poisson distribution, $H_1 : X$ does not follow the Poisson distribution. Knowing that $\mu = \bar{x} = \frac{618}{150} = 4.12$.

For Poisson distribution, $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$. Then, $P(X = 0) = 0.0162$, $P(X = 1) = 0.0669$, $P(X = 2) = 0.1378 \cdots$. The summary table is as follows:

x value	0	1	2	3	4	5	6	7	8	9
P	0.0162	0.0669	0.1379	0.1893	0.1950	0.1607	0.1103	0.0649	0.0334	0.0254
O_j	5	11	18	26	29	25	15	10	7	4
E_j	2.430	10.035	20.685	28.395	29.250	24.105	16.545	9.735	5.010	3.810
$(O_j - E_j)^2/E_j$	2.7181	0.0928	0.3485	0.2020	0.0021	0.0332	0.1443	0.0072	0.7904	0.0095
$O_j ln(\frac{O_j}{E_j})$	3.6077	1.0100	-2.5027	-2.2910	-0.2489	0.9114	-1.4705	0.2686	2.3413	0.1947

Pearson Chi-square test:

$$X^2 = \sum_{i=0}^{9} \frac{(O_j - E_j)^2}{E_j} = 4.3482$$

Likelihood Ratio test:

$$G^2 = 2\sum_{i=0}^{9} O_j ln(\frac{O_j}{E_j}) = 3.6411$$

The degree of freedom $r = 10 - 1 - 1 = 8$, hence $\chi_{8,0.05} = 15.5073 > X^2$ and G^2. Hence we do not reject H_0, which is that X follows the Poisson distribution.

Problem 4

Solution: The score $100(1-\alpha)\%$ confidence interval for π is given by all π_0 that satisfies $|z| \leq z_{\frac{\alpha}{2}}$, where given p,

$$z = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Then,

$$\begin{aligned}
|z| \leq z_{\frac{\alpha}{2}} &\Rightarrow \left| \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \right| \leq z_{\frac{\alpha}{2}} \\
&\Rightarrow \frac{(p-\pi_0)^2}{\pi_0(1-\pi_0)/n} \leq z_{\frac{\alpha}{2}}^2 \\
&\Rightarrow p^2 + \pi_0^2 - 2\pi_0 p < z_{\frac{\alpha}{2}}^2 \frac{\pi_0(1-\pi_0)}{n} \\
&\Rightarrow p^2 + \pi_0^2 - 2\pi_0 p - \frac{z_{\frac{\alpha}{2}}^2}{n}\pi_0 + \frac{z_{\frac{\alpha}{2}}^2}{n}\pi_0^2 < 0 \\
&\Rightarrow (1 + \frac{z_{\frac{\alpha}{2}}^2}{n})\pi_0^2 + (-2p - \frac{z_{\frac{\alpha}{2}}^2}{n})\pi_0 + p^2 < 0
\end{aligned}$$

Then, according to the solution to the quadratic equation,

$$\pi_0 \in (\frac{2p + \frac{z_{\frac{\alpha}{2}}^2}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{z_{\frac{\alpha}{2}}{}^2}{n^2} + \frac{4p}{n} - \frac{4p^2}{n}}}{2(1 + \frac{z_{\frac{\alpha}{2}}}{n})}, \frac{2p + \frac{z_{\frac{\alpha}{2}}^2}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{z_{\frac{\alpha}{2}}{}^2}{n^2} + \frac{4p}{n} - \frac{4p^2}{n}}}{2(1 + \frac{z_{\frac{\alpha}{2}}}{n})})$$

Problem 5

Solution:

(a) According to H_0, $\pi_1 = 0.1$, $\pi_2 = 0.1$, $\pi_3 = 0.25$, $\pi_4 = 0.25$, then $\pi_5 = 1 - \sum_{i=1}^{4} \pi_i = 0.3$. The table is as follows:

x value	1	2	3	4	5
P	0.1	0.1	0.25	0.25	0.3
O_j	10	13	23	21	29
E_j	9.6	9.6	24	24	28.8
$(O_j - E_j)^2/E_j$	0.0167	1.2042	0.04167	0.3750	0.0014
$O_j ln(\frac{O_j}{E_j})$	0.4082	3.9414	-0.9789	-2.8042	0.2007

Then **Pearson chi-square test:** $X^2 = \sum_{i=1}^{5} \frac{(O_j-E_j)^2}{E_j} = 1.6389$. **Likelihood ratio test:** $G^2 = 2\sum_{i=1}^{5} O_j ln(\frac{O_j}{E_j}) = 1.5346$. $\chi^2_{4,0.05} = 9.4877 > X^2$ and G^2. Hence, we fail to reject H_0.

(b) Under the null hypothesis, the joint pdf

$$P(x_1 = n_1, x_2 = n_2, \cdots, x_5 = n_5) = \frac{n!}{n_1!n_2!\cdots n_5!}\pi_1^{n_1+n_2}\pi_3^{n3+n4}\pi_5^{n_5}$$

The likelihood function is the same as joint pdf. Then the log-likelihood function is as follows:

$$\begin{aligned}
\ell(\pi; X) &= ln(L(\pi; X)) = ln(P(x_1 = n_1, x_2 = n_2, \cdots, x_5 = n_5)) \\
&= ln(\frac{n!}{n_1!n_2!\cdots n_5!}) + (n_1+n_2)ln\pi_1 + (n_3+n_4)ln\pi_3 + n_5 ln\pi_5 \\
&= ln(\frac{n!}{n_1!n_2!\cdots n_5!}) + (n_1+n_2)ln\pi_1 + (n_3+n_4)ln\pi_3 + n_5 ln(1-2\pi_1-2\pi_3)
\end{aligned}$$

Then, take partial derivative with respect to π_1, π_3 and set to 0 respectively. We get:

$$\left\{\begin{aligned}
\frac{\partial\ell(\pi;X)}{\partial\pi_1} &= \frac{n_1+n_2}{\pi_1} - \frac{2n_5}{1-2\pi_1-2\pi_3} \overset{\text{set}}{=} 0 \\
\frac{\partial\ell(\pi;X)}{\partial\pi_3} &= \frac{n_3+n_4}{\pi_3} - \frac{2n_5}{1-2\pi_1-2\pi_3} \overset{\text{set}}{=} 0
\end{aligned}\right. \Rightarrow \left\{\begin{aligned}
\hat{\pi_1} &= \hat{\pi_2} = \frac{n_1+n_2}{2n} \\
\hat{\pi_3} &= \hat{\pi_4} = \frac{n_3+n_4}{2n} \\
\hat{\pi_5} &= \frac{n_5}{n}
\end{aligned}\right.$$

(c) According to problem (b), the estimated

$$\hat{\pi_1} = \hat{\pi_2} = \frac{n_1+n_2}{2n} = \frac{10+13}{2\times 96} = \frac{23}{192}$$

$$\hat{\pi_3} = \hat{\pi_4} = \frac{n_3+n_4}{2n} = \frac{23+21}{2\times 96} = \frac{44}{192} = \frac{11}{48}$$

$$\pi_5 = \frac{29}{96}$$

Then the table is as follows:

x value	1	2	3	4	5
P	23/192	23/192	11/48	11/48	29/96
O_j	10	13	23	21	29
E_j	11.5	11.5	22	22	29
$(O_j - E_j)^2/E_j$	0.1957	0.1957	0.0455	0.0455	0
$O_j ln(\frac{O_j}{E_j})$	-1.3976	1.5938	1.0224	-0.9769	0

The degree of freedom is $r = 5 - 1 - 2$(the number of estimated parameters).
Then **Pearson chi-square test:** $X^2 = \sum_{i=1}^{5} \frac{(O_j - E_j)^2}{E_j} = 0.4822$. **Likelihood ratio test:** $G^2 = 2\sum_{i=1}^{5} O_j ln(\frac{O_j}{E_j}) = 0.4834$. $\chi^2_{2,0.05} = 5.9915 > X^2$ and G^2. Hence, we fail to reject H_0.

Problem 6

Solution:
Use **Wald test**, the related $100(1-\alpha)\%$ C.I. for π is given by

$$\hat{\pi} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{\pi}(1-\hat{\pi})/n} = 0.86 \pm 1.96\sqrt{0.86 \times 0.14/1158} = (0.8400, 0.8800)$$

Consider the null hypothesis as H_0: $\pi_0 = 0.5$, then $z = \frac{\hat{\pi} - \pi_0}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} = 35.31 > z_{0.05} = 1.96$
Hence, we reject H_0. The result indicates that more than half of the American adults believe in heaven. And we also have at least 95% confidence that the proportion of American adults believing in heaven is between 0.84 and 0.88. The statistical reference process is based on the assumption that believing in heaven is an i.i.d sample in the American adults.