Math 741 Assignment 21 (Quiz)

Arnold Jiadong Yu

May 16, 2019

10.3.1 solution:

$$\sum_{i=1}^{t} \frac{(X_i - np_u)^2}{np_i} = \sum_{i=1}^{t} \frac{(X_i^2 - 2np_iX_i + n^2p_i^2)}{np_i}$$

$$= \sum_{i=1}^{t} \frac{X_i^2}{np_i} - 2\sum_{i=1}^{t} \frac{np_iX_i}{np_i} + \sum_{i=1}^{t} \frac{n^2p_i^2}{np_i} = \sum_{i=1}^{t} \frac{X_i^2}{np_i} - 2\sum_{i=1}^{t} X_i + n\sum_{i=1}^{t} p_i$$

$$= \sum_{i=1}^{t} \frac{X_i^2}{np_i} - 2n + n = \sum_{i=1}^{t} \frac{X_i^2}{np_i} - n$$

10.3.2 solution: Let i denote the outcomes of getting a white chips, this is clearly a hyper-geometric where $N = 10, w = 4, r = 6, n = 2$. From the information, a table can be formulated.

i	k_i	p_i	np_i
0	35	1/3	33.33
1	55	8/15	53.33
2	10	2/15	13.33

where k_i is the observed frequency, p_i is the probability to i outcomes of white chips with $\frac{\binom{w}{i}\binom{N-w}{n-i}}{\binom{N}{n}}$, and np_i is expected frequency.
Therefore, a test of whether it is the true probability also can be formulated.

$$H_0 : p_1 = 1/3, p_2 = 8/15, p_3 = 2/15$$

$$H_1 : \text{At least one } p_i \text{ is different}$$

with $\alpha = 0.05$ and $t = 3$.

$$d = \sum_{i=0}^{2} \frac{[k_i - np_i]^2}{np_i} = \frac{[35 - 100 \cdot (1/3)]^2}{100 \cdot (1/3)} + \frac{[55 - 100 \cdot (8/15)]^2}{100 \cdot (8/15)} + \frac{[10 - 100 \cdot (2/15)]^2}{100 \cdot (2/15)}$$

$$= 1/12 + 5/96 + 5/6 = 31/32$$

$$p - value = 1 - P(0 \leq \chi^2_{t-1} \leq 31/32) = 1 - 0.383918 = 0.616082$$

since $p - value = 0.616082 > \alpha = 0.05 \implies$ Fail to reject H_0. Hence, there is enough evidence to show that these probability is the true probability.

10.3.3 solution: Let i denote the outcomes of getting a white chips, this is clearly binomial where $n = 10, w = 4$ and $r = 6$. From the information, a table can be formulated.

i	k_i	p_i	np_i
0	35	0.36	36
1	55	0.48	48
2	10	0.16	16

where k_i is the observed frequency, p_i is the probability to i outcomes of white chips with $\binom{n}{i}(w/n)^i(r/n)^{n-i}$, and np_i is expected frequency.
Therefore, a test of whether it is the true probability also can be formulated.

$$H_0 : p_1 = 0.36, p_2 = 0.48, p_3 = 0.16$$

$$H_1 : \text{At least one } p_i \text{ is different}$$

with $\alpha = 0.05$ and $t = 3$.

$$d = \sum_{i=0}^{2} \frac{[k_i - np_i]^2}{np_i} = \frac{[35 - 36]^2}{36} + \frac{[55 - 48]^2}{48} + \frac{[10 - 16]^2}{16}$$

$$= 1/36 + 49/48 + 36/16 = 475/144$$

$$p - value = 1 - P(0 \leq \chi^2_{t-1} \leq 475/144) = 0.19218$$

since $p - value = 0.19218 > \alpha = 0.05 \implies$ Fail to reject H_0. Hence, there is enough evidence to show that these probability is the true probability.

10.3.4 solution: From the information, a test can be formulated. The births occurs uniformly in all time periods vs. the birth doesn't occur uniformly.

There are 4 hours between midnight and 4 a.m. Let p_1 denote the probability of birth occurs in these 4 hours and p_2 denote the probability of birth occurs otherwise.

$$H_0 : p_1 = 1/6, p_2 = 5/6$$

$$H_1 : \text{At least one } p_i \text{ is different}$$

with $\alpha = 0.05$ and $t = 2$. A table also can be formulated,

i	k_i	np_i
1	494	441.67
2	2156	2208.33

where k_i is the observed frequency, and np_i is expected frequency.

$$d = \sum_{i=1}^{2} \frac{[k_i - np_i]^2}{np_i} = \frac{[494 - 441.67]^2}{441.67} + \frac{[2156 - 2208.33]^2}{2208.33} = 7.440215$$

$$p - value = 1 - P(0 \leq \chi^2_{t-1} \leq 7.440215) = 0.006378$$

since $p - value = 0.006378 < \alpha = 0.05 \implies$ Reject H_0. Hence, there is enough evidence to say that the births doesn't occur uniformly in all time periods.

10.3.5 solution: 10.3.4 can also be done using one-sample stat test.

$$H_0 : p = 1/6$$

$$H_1 : p \neq 1/6$$

with $\alpha = 0.05, x = 494, n = 2650$. Therefore,

$$z_0 = \frac{x - np_0}{\sqrt{np_0(1 - p_0}} = 2.72786$$

$$p - value = 1 - P(-2.727886 \leq Z \leq 2.72786) = 0.006375$$

since $p - value = 0.006375 < \alpha = 0.05 \implies$ Reject H_0. Hence, there is enough evidence to say that the births doesn't occur uniformly in all time periods.

10.3.7 solution: Let p_i denote the outcomes of the color i M&M, where i is brown, yellow, red, orange, blue and green. A test can be formulated,

$$H_0 : p_1 = 0.30, p_2 = 0.20, p_3 = 0.20, p_4 = 0.10, p_5 = 0.10, p_6 = 0.10$$

$$H_1: \text{At least one } p_i \text{ is different}$$

with $\alpha = 0.05, t = 6$ and $n = 1527$. A table also can be formulated,

i	color	k_i	p_i	np_i
1	brown	455	0.30	458.1
2	yellow	343	0.20	305.4
3	red	318	0.20	305.4
4	orange	152	0.10	152.7
5	blue	130	0.10	152.7
6	green	129	0.10	152.7

Therefore,

$$d = \sum_{i=1}^{6} \frac{[k_i - np_i]^2}{np_i} = \frac{[455 - 458.1]^2}{458.1} + \frac{[343 - 305.4]^2}{305.}$$

$$+\frac{[318 - 305.4]^2}{305.4} + \frac{[152 - 152.7]^2}{152.7} + \frac{[130 - 152.7]^2}{152.7} + \frac{[129 - 152.7]^2}{152.7}$$
$$= 12.2261515$$

$$p - value = 1 - P(0 \leq \chi^2_{t-1} \leq 12.2261515) = 0.031817$$

since $p - value = 0.031817 < \alpha = 0.05 \implies$ Reject H_0. Hence, there is enough evidence to say that that the data are not consistent with the company's stated intentions.

10.3.8(H) solution: From the information given, a test can be formulated

$$H_0: \text{ These data are compatible witht he model that}$$

$$\text{each World Series game is an independent Bernoulli trial with}$$

$$p = P(\text{AL wins}) = P(\text{NL wins}) = \frac{1}{2}$$

$$H_1: \text{ These data are not compatible witht he model that}$$

$$\text{each World Series game is an independent Bernoulli trial with}$$

$$p = P(\text{AL wins}) = P(\text{NL wins}) = \frac{1}{2}$$

with $\alpha = 0.10, t = 4, n = 50$. A table also can be formulated,

i	Number of Games	Number of Years(k_i)	p_i	np_i
1	4	9	1/8	152.7
2	5	11	1/4	152.7
3	6	8	5/16	152.7
4	7	22	5/16	152.7

where

$$p_1 = P(X = 4) = (1/2)^4 + (1/2)^4 = 1/8$$

$$p_2 = P(X = 5) = \binom{4}{1}(1/2)^5 + \binom{4}{1}(1/2)^5 = 1/4$$

$$p_3 = P(X = 6) = \binom{5}{2}(1/2)^6 + \binom{4}{1}(1/2)^6 = 5/16$$

$$p_4 = P(X = 7) = \binom{6}{3}(1/2)^7 + \binom{4}{1}(1/2)^7 = 5/16$$

Therefore,

$$\chi_0^2 = \sum_{i=1}^{4} \frac{[k_i - np_i]^2}{np_i} = 7.712$$

$$p - value = 1 - P(0 \leq \chi_{t-1}^2 \leq 7.712) = 0.05235$$

since $p-value = 0.05235 < \alpha = 0.10 \implies$ Reject H_0. Hence, there is enough evidence to say that these data are not compatible witht he model that each World Series game is an independent Bernoulli trial with $p = P(\text{AL wins}) = P(\text{NL wins}) = \frac{1}{2}$.

10.3.10 solution: A test can be formulated by given information,

$$H_0: \text{ The data follows normal distribution}$$

$$H_0: \text{ The data does not follow normal distribution}$$

with $\alpha = 0.10$ and $n = 70$. A table can be formulated,

i	class	k_i	p_i	np_i
1	$220 \le y \le 230$	1	0.0102	0.714
2	$230 \le y \le 240$	5	0.0394	2.758
3	$240 \le y \le 250$	10	0.1071	7.497
4	$250 \le y \le 260$	16	0.1933	13.531
5	$260 \le y \le 270$	23	0.2467	17.269
6	$270 \le y \le 280$	7	0.2119	14.833
7	$280 \le y \le 290$	6	0.1226	8.582
8	$290 \le y \le 300$	2	0.0668	4.676

where k_i is the observed frequency, p_i is the probability with $P(\frac{x_1-\mu}{\sigma} \le Z \le \frac{x_2-\mu}{\sigma})$, and np_i is expected frequency. And $t = 8$

$$d = \sum_{i=1}^{8} \frac{[k_i - np_i]^2}{np_i} =$$

$$p - value = 1 - P(0 \le \chi^2_{t-1} \le) =$$

10.3.11(H) solution: Given,

$$f_Y(y) = \begin{cases} \frac{1}{9}y^2 & 0 < y \le 3 \\ 0 & o.w. \end{cases}$$

Then

$$F_Y(y) = \begin{cases} \frac{1}{27}y^3 & 0 < y \le 3 \\ 0 & o.w. \end{cases}$$

A table can be formualted,

i	class	k_i	p_i	np_i
1	$0 < y \le 1$	8	1/27	50/27
2	$1 < y \le 2$	16	7/27	350/27
3	$2 < y \le 3$	26	19/27	950/27

where k_i is the observed frequency, p_i is the probability with $P(y_1 < Y \le y_2) = F(y_2) - F(y_1)$, and np_i is expected frequency. And $t = 3$
Since $np_1 \le 5$, we need to pool the table. A new table need to be formulated,

i	class	k_i	p_i	np_i
1	$0 < y \le 2$	24	8/27	400/27
2	$2 < y \le 3$	26	19/27	950/27

A test can be formualted,

$$H_0: \text{ The data is consistent with } f_Y(y)$$

$$H_1: \text{ The data is not consistent with } f_Y(y)$$

with $\alpha = 0.05$ and $t = 2$. Then

$$\chi_0^2 = \sum_{i=1}^{2} \frac{[k_i - np_i]^2}{np_i} = 8.09263$$

$$p - value = 1 - P(0 \leq \chi_{t-1}^2 \leq 8.09263) = 0.0044446$$

since $p - value = 0.0044446 < \alpha = 0.05 \implies$ Reject H_0. Hence, there is enough evidence to say that the data is not consistent with $f_Y(y)$.