THE UNIVERSITY OF SYDNEY

STAT2012 STATISTICAL TESTS

Semester 2	**Solution to Tutorial Week 12**	2015

Tutorial questions

1. The calculation is summarized in the following table:

Current Survey

User Segment	Obs. freq. y_i	Exp. prob. p_{i0}	Exp. freq. np_{i0}	$\frac{(y_i-np_{i0})^2}{np_{i0}}$
Business-Prof.	102	0.69	$150(0.69) = 103.5$	
Goverment	32	0.21	$150(0.21) = 31.5$	
Education	12	0.07	$150(0.07) = 10.5$	
Home	4	0.03	$150(0.03) = 4.5$	
Total	150	1	150	

Since $E_4 = 4.5 < 5$, we combine the last two classes. The revised table is

Current Survey

User Segment	Obs. freq. y_i	Exp. prob. p_{i0}	Exp. freq. np_{i0}	$\frac{(y_i-np_{i0})^2}{np_{i0}}$
Business-Prof.	102	0.69	$150(0.69) = 103.5$	$\frac{(102-103.5)^2}{103.5} = 0.0217$
Goverment	32	0.21	$150(0.21) = 31.5$	$\frac{(32-31.5)^2}{31.5} = 0.0079$
Education and home	16	0.10	$150(0.10) = 15$	$\frac{(16-15)^2}{15} = 0.0667$
Total	150	1	150	0.0963

The Chi-square goodness-of-fit test is

1. **Hypotheses:**

$$\begin{aligned} H_0 : &\quad p_1 = 0.69,\ p_2 = 0.21,\ p_3 = 0.07\ p_4 = 0.03 \quad \text{vs} \\ H_1 : &\quad \text{At least one of the equalities does not hold.} \end{aligned}$$

2. **Test statistic:** $\chi_0^2 = \displaystyle\sum_{i=1}^{3} \frac{(n - np_{i0})^2}{np_{i0}} = 0.0963.$

3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi_{k-1}^2$ approx.

4. ***P*-value:** p-value $= \Pr(\chi_2^2 \geq 0.0963) > 0.1$ ($\chi^2_{2,0.90} = 4.605,\ 0.95$ from R).

5. **Decision:** Since the p-value > 0.05, we accept H_0. The data is consistent with the percentages of computer uses in 1988.

2. The calculation is summarized in the following table:

Selection	Observed freq.	Expected freq.	Diff.	Test stat.
i	O_i	$E_i = np_i$	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
Nonsmokers	44	$100 \times 0.5 = 50$	-6	$\frac{(-6)^2}{50} = 0.72$
Current Smokers	24	$100 \times 0.2 = 20$	4	$\frac{4^2}{20} = 0.80$
Tobacco Chewers	13	$100 \times 0.1 = 10$	3	$\frac{3^2}{10} = 0.90$
Ex-smokers	19	$100 \times 0.2 = 20$	-1	$\frac{(-1)^2}{20} = 0.05$
Total	100	100	0	$\chi_0^2 = 2.47$

The Chi-square goodness-of-fit test is

1. **Hypothesis:** H_0: $p_1 = .5,\ p_2 = .2,\ p_3 = .1,\ p_4 = .2$ vs
   H_1: At least one equality do not hold.
2. **Test statistic:** $\chi_0^2 = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i} = 2.47$
3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi_{k-1}^2$ approx.
4. ***P*-value:** p-value $= \Pr(\chi_3^2 \geq 2.47) > 0.10$ $(\chi^2_{3,0.90} = 6.251,\ 0.480737$ from R)
5. **Decision:** Since p-value> 0.01, we accept H_0. The data is consistent with the proportions estimated by the researcher.

3. The calculation is summarized in the following table:

No.	Obs. f.	Exp. prob.	Exp. f.	Chi-square
x	y_i	$p_i = {}_5C_i p_0^i (1-p_0)^{5-i}$	$E_i = n_i p_i$	$\frac{(n_i - np_i)^2}{np_i}$
0	10	$C_4^0 0.5^0 0.5^4 = 0.0625$	$100(0.0625) = 6.25$	$\frac{(3.750)^2}{6.25} = 2.250$
1	18	$C_4^1 0.5^1 0.5^3 = 0.2500$	$100(0.2500) = 25.00$	$\frac{(-7.000)^2}{25.00} = 1.960$
2	35	$C_4^2 0.5^2 0.5^2 = 0.3750$	$100(0.3750) = 37.50$	$\frac{(-2.500)^2}{37.50} = 0.167$
3	28	$C_4^3 0.5^3 0.5^1 = 0.2500$	$100(0.2500) = 25.00$	$\frac{(3.000)^2}{25.00} = 0.360$
4	9	$C_4^4 0.5^4 0.5^0 = 0.0625$	$100(0.0625) = 6.25$	$\frac{(2.750)^2}{6.25} = 1.210$
Sum	100	1.0000	100.00	5.947

The Chi-square goodness-of-fit test for the binomial distribution is

1. **Hypothesis:** H_0: X_j follow a binomial distribution vs
   H_1: X_j does not follow a binomial distribution.
2. **Test statistic:** $\chi_0^2 = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i} = 5.947$
3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi_{k-1}^2$ approx.
4. *P*-**value:** p-value $= \Pr(\chi_4^2 \geq 5.947) > 0.1$ ($\chi^2_{4,0.90} = 7.779,\ 0.2031668$ from R)
5. **Decision:** Since p-value> 0.05, we accept H_0. The data is consistent with the null hypothesis that the data follow a binomial distribution with $p = 0.5$.

Extra problems

1. Calculation is summarized in the following table:

Location i	Observed freq. y_i	Expected freq. np_i	Diff. $(y_i - np_i)$	$\frac{(y_i-np_i)^2}{np_i}$
I	6	$30 \times 0.33 = 10$	-4	$\frac{(-4)^2}{10} = 1.6$
II	9	$30 \times 0.33 = 10$	-1	$\frac{(-1)^2}{10} = 0.1$
III	15	$30 \times 0.33 = 10$	5	$\frac{5^2}{10} = 2.5$
Total	30	30	0	$\chi_0^2 = 4.2$

The Chi-square goodness-of-fit test is

1. **Hypothesis:** H_0: $p_1 = 0.33,\ p_2 = 0.33,\ p_3 = 0.33$ vs
   H_1: At least one equality does not hold.
2. **Test statistic:** $\chi_0^2 = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i} = 4.2$
3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi_{k-1}^2$ approx.
4. **P-value:** p-value $= \Pr(\chi_2^2 > 4.2) > 0.1$ ($\chi^2_{2,0.90} = 4.605,\ 0.1225$ from R)
5. **Decision:** Since the p-value > 0.05, we accept H_0. The data is consistent with the null hypothesis of the proportions of preferences for the three proposals.