Statistics II : Tutorial 7 - Chi-Square Test

July 23, 2019

1.

(a) Hypothesis:

i. The die is unbiased at the 5% significance level (claim) (H_0)

ii. The die is NOT unbiased at the 5% is significance level (opposite) (H_1)

(b) Significance level, critical point

i. $\alpha = 0.05$, Degree of Freedom $= 6 - 1 = 5$

ii. $\chi^2_{0.05;5} = 11.070$

(c) Rejection region

i. $\chi^2 > 11.070$

(d) Test statistic

i. If the die is unbiased, then each of them should have uniform distributed probability:

$$P\left(X = x\right) = \frac{1}{6}$$

ii.

Number	O_i	P_i	$E_i = P_i * n$	$\frac{(O_i - E_i)^2}{E_i}$	Total
1	16	$\frac{1}{6}$	22	$\frac{(16-22)^2}{22}$	1.6363
2	20	$\frac{1}{6}$	22	$\frac{(20-22)^2}{22}$	0.1818
3	25	$\frac{1}{6}$	22	$\frac{(25-22)^2}{22}$	0.4091
4	14	$\frac{1}{6}$	22	$\frac{(14-22)^2}{22}$	2.9091
5	29	$\frac{1}{6}$	22	$\frac{(29-22)^2}{22}$	2.2273
6	28	$\frac{1}{6}$	22	$\frac{(28-22)^2}{22}$	1.6364

iii. Chi-Square Test statistic

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{6} \frac{\left(O_i - E_i\right)^2}{E_i} \\ &= 1.6363 + 0.1818 + 0.4091 + 2.9091 + 2.2273 + 1.6364 \\ \chi^2 &= 9\end{aligned}$$

(e) Conclusion

i. Since $\chi^2 = 9 < 11.070$, we failed to reject H_0, and hence do not have enough evidence to conclude that the die is NOT unbiased at the significance level of 5%

2.

(a) Hypothesis

i. Claim: The eight machines have equal performance (H_0)

ii. Oppo: The eight machines DO NOT have equal performance (H_1)

(b) Critical value

i. $\alpha = 0.01, D.O.F\,(8-1) = 7$

ii. $\chi^2_{0.01;8-0-1} = \chi^2_{0.01;7} = 18.475$

(c) Rejection region

i. $\chi^2 > 18.475$

(d) Test statistic

i. If equal performance, all machines should produce the equal number of items per hour. $E_i = \frac{\sum O_i}{n} = \frac{8+7+6+9+10+8+6+10}{8} = 8$

ii.

Machine	O_i	$E_i = P_i * n$	$\frac{(O_i-E_i)^2}{E_i}$	Total
1	8	8	$\frac{(8-8)^2}{8}$	0
2	7	8	$\frac{(7-8)^2}{8}$	0.125
3	6	8	$\frac{(6-8)^2}{8}$	0.5
4	9	8	$\frac{(9-8)^2}{8}$	0.125
5	10	8	$\frac{(10-8)^2}{8}$	0.5
6	8	8	$\frac{(8-8)^2}{8}$	0
7	6	8	$\frac{(6-8)^2}{8}$	0.5
8	10	8	$\frac{(10-8)^2}{8}$	0.5

iii. Chi-square test statistic

$$\chi^2 = \sum_{i=1}^{8} \frac{(O_i - E_i)^2}{E_i}$$
$$= 0 + 0.125 + 0.5 + 0.125 + 0.5 + 0 + 0.5 + 0.5$$
$$\chi^2 = 2.25$$

(e) Conclusion

i. Since $\chi^2 = 2.25 < 18.475$, we failed to reject H_0, and hence do not have enough evidence to say that the eight machines DO NOT have equal performance at 1% significance level.

3.

(a) Hypothesis

i. Claim: the true percentages of the colours produced differ from the manufacturer's stated percentages (H_1)

ii. Oppo: the true percentages of the colours produced do not differ from the manufacturer's stated percentages (H_0)

(b) Critical value

i. $\alpha = 0.02, D.O.F. = 6 - 0 - 1$ (Note: -0 is because we do not "predict" any expected outcome)

ii. $\chi^2_{0.02;5} = 13.388$

(c) Rejection region

i. $\chi^2 > 13.388$

(d) Test statistics

i. $\sum n = 84 + 79 + 75 + 49 + 36 + 47 = 370$

ii.

Colour	O_i	P_i	$E_i = P_i * \sum n$	$\frac{(O_i-E_i)^2}{E_i}$	Total
Brown	84	0.3	111	$\frac{(84-111)^2}{111}$	6.5676
Yellow	79	0.2	74	$\frac{(79-74)^2}{74}$	0.3378
Red	75	0.2	74	$\frac{(75-74)^2}{74}$	0.0135
Orange	49	0.1	37	$\frac{(49-37)^2}{37}$	3.8919
Green	36	0.1	37	$\frac{(36-37)^2}{37}$	0.0270
Tan	47	0.1	37	$\frac{(47-37)^2}{37}$	2.7027

iii. $\chi^2 = 6.5676+0.3378+0.0135+3.8919+0.0270+2.7027 = 13.5405$

(e) Conclusion, since $\chi^2 = 13.5405 > 13.388$, we reject H_0 and hence have enough evidence to conclude, at 2% significance level, the true percentages of the colours produced differ from the manufacturer's stated percentages.

4.

(a) Hypothesis

i. Claim: The recent sales follow the past pattern of sales (H_0)

ii. Opposite: The recent sales DO NOT follow the past pattern of sales (H_1)

(b) Critical value

i. $\alpha = 0.025, D.O.F. = 3 - 0 - 1 = 2$

ii. $\chi^2_{0.025;2} = 7.378$

(c) Rejection range:

i. $\chi^2 > 7.378$

(d) Test statistics:

i.

Television	O_i	P_i	$E_i = P_i * \Sigma n$	$\frac{(O_i - E_i)^2}{E_i}$
Small-screen	24	0.3	30	1.2
Medium-screen	37	0.4	40	0.225
Large-screen	39	0.3	30	2.7
Σn	100			$\chi^2 = 4.125$

ii. $\chi^2 = 4.125 < 7.378$

(e) Conclusion, since $\chi^2 = 4.125 < 7.378$, we failed to reject $H_0$and hence do not have enough evidence to say that the recent sales DO NOT follow the past pattern of sales (H_1) at 2.5% significance level.

5.

(a) Hypothesis

i. Claim: All four coins are fair (H_0)

ii. Opposite: Not all four coins are fair. (H_1) (Note: "All four coins are not fair" is NOT acceptable, because it can be 3 fair,1 not fair)

(b) Critical value

i. $\alpha = 0.025, D.O.F. = 5 - 1 = 4$

ii. $\chi^2_{0.025;4} = 11.143$

(c) Rejection region:

i. $\chi^2 > 11.143$

(d) Test statistics:

i. If the coins are fair, then it should follow a binomial distribution of $B(4, 0.5)$

ii. Each of the probability are: $P_i = \binom{4}{i} 0.5^i * 0.5^{4-i}$

iii.

Numbr of heads	O_i	P_i	$E_i = P_i * \Sigma n$	$\frac{(O_i - E_i)^2}{E_i}$
0	5	0.0625	6.25	0.25
1	23	0.25	25	0.16
2	39	0.375	37.5	0.06
3	19	0.25	25	1.44
4	14	0.0625	6.25	9.61
Σn	100			$\chi^2 = 11.52$

(e) Conclusion, since $\chi^2 = 11.52 > 11.143$, we reject H_0 and hence can say that we have sufficient evidence to say that not all four coins are fair at 2.5% significance level.

6.

(a) Table

i.

Q6				
Intensity of bookings	Type of flights			
	Internal	Regional	International	TOTAL
Fully booked	154	171	275	600
Expected	150	150	300	
Not fully booked	96	79	225	400
Expected	100	100	200	
TOTAL	250	250	500	1000

(b) Hypothesis

i. Claim: There is no evidence of significant association between the type of flights and the intensity of bookings (H_0)

ii. Opposite: There is evidence of significant association between the type of flights and the intensity of bookings (H_1)

(c) Find the critical value, at $\alpha = 0.01$

$$\begin{aligned}\chi^2_{0.01;(2-1)(3-1)} &= \chi^2_{0.01;2} \\ &= 9.210\end{aligned}$$

(d) Find the rejection region

i. $\chi^2 > 9.210$

(e) Find the test statistics

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \\ &= \frac{(154-150)^2}{150}+\frac{(171-150)^2}{150}+\frac{(275-300)^2}{300}+\frac{(96-100)^2}{100}+\frac{(79-100)^2}{100}+\frac{(225-200)^2}{200} \\ &= 12.825\end{aligned}$$

(f) Conclusion

i. Since $\chi^2 = 12.825 > 9.210$, we reject H_0 . Hence, we have sufficient evidence to conclude that there is evidence of significant association between the type of flights and the intensity of bookings at 1% significance level.

7.

(a) Table

	Type of Machines			
% defective, d	Cutting Machine	Grinding machine	Milling machine	TOTAL
d <= 1	22	74	102	198
Expected	19.8	79.2	99	
1 < d < 2	31	102	143	276
Expected	27.6	110.4	138	
d >= 2	7	64	55	126
Expected	12.6	50.4	63	
TOTAL	60	240	300	600

(b) Hypothesis

i. Claim: the degree of defectiveness is independent from the types of machine. (H_0)

ii. Oppo: the degree of defectiveness is NOT independent from the types of machine (H_1)

(c) Critical value $\chi^2_{0.02;(3-1)(4-1)} = \chi^2_{0.02;6} = 15.033$

(d) Rejection range:

$$\chi^2 > 15.033$$

(e) Test statistics:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$
$$= \frac{(22-19.8)^2}{19.8} + \frac{(74-79.2)^2}{79.2} + \frac{(102-99)^2}{99} + \frac{(31-27.6)^2}{27.6} + \frac{(102-110.4)^2}{110.4} + \frac{(143-138)^2}{138} +$$
$$= 9.0905$$

(f) Conclusion

i. Since $\chi^2 = 9.0905 < 15.033$,we failed to reject H_0 and hence do not have sufficient evidence to conclude thatthe degree of defectiveness is NOT independent from the types of machine at 2% significance level.

8.

(a) Table

Q8					
			Productivity		
Training method	High	Good	Average	Low	TOTAL
In-house	26	43	62	11	142
Expected	27.61111111	39.44444444	55.22222222	19.72222222	
Outside agency	28	40	59	20	147
Expected	28.58333333	40.83333333	57.16666667	20.41666667	
Previous employment	16	17	19	19	71
Expected	13.80555556	19.72222222	27.61111111	9.861111111	
TOTAL	70	100	140	50	360

(b) Hypothesis

i. Claim: the level of productivity DO NOT depends on the method of training (H_0)

ii. Opposite: the level of productivity depends on the method of training (H_1)

(c) Critical value at $\alpha = 0.05$

$$\begin{aligned}\chi^2_{0.05;(3-1)(4-1)} &= \chi^2_{0.05;2\cdot 3} \\ &= \chi^2_{0.05;6} \\ &= 12.592\end{aligned}$$

(d) Rejection region

$$\chi^2 > 12.592$$

(e) Test statistic

i.

Oi	Ei	(Oi-Ei)^2/Ei
26.0000	27.6111	0.0940
43.0000	39.4444	0.3205
62.0000	55.2222	0.8319
11.0000	19.7222	3.8574
28.0000	28.5833	0.0119
40.0000	40.8333	0.0170
59.0000	57.1667	0.0588
20.0000	20.4167	0.0085
16.0000	13.8056	0.3488
17.0000	19.7222	0.3757
19.0000	27.6111	2.6856
19.0000	9.8611	8.4696
	Total	**17.0797**

$$\begin{aligned}\chi^2 &= \sum_{i=0}^{r}\sum_{j=0}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \\ &= 17.080\end{aligned}$$

(f) Conclusion

i. Since $\chi^2 = 17.080 > 12.592$. We reject $H_0$and hence have sufficienct evidence to conclude that the level of productivity depends on the method of training at 5% significance level.

(g) NOTE:

i. Slight difference in answer (at 3th decimal place onwards) due to possible rounding errors

9.

(a) Hypothesis

i. Claim: The coin is fair (H_0)

ii. Opposite: The coin is not fair (H_1)

(b) Find the critical point at $\alpha = 0.02$

i. $\chi^2_{0.02;2-0-1} = \chi^2_{0.02;1} = 5.412$

(c) Find the rejection region:

i. $\chi^2 > 5.412$

(d) Find the test statistic

i.

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{m} \frac{(|O_i - E_i| - 0.5)^2}{E_i} \\ &= \frac{(|454 - 500| - 0.5)^2}{500} + \frac{(|546 - 500| - 0.5)^2}{500} \\ &= 8.281\end{aligned}$$

(e) Conclusion

i. Since $\chi^2 = 8.281 > 5.412$, we reject H_0 and hence conclude that the coin is not fair at 2% significance level.

10.

(a) Hypothesis

i. Claim: The vaccine do not have relationship with recovery from the disease (H_0)

ii. Opposite: The vaccine have relationship with recovery from the disease (H_1)

(b) Critical value, $\alpha = 0.05$

i. $\chi^2_{0.05;m-t-1} = \chi^2_{0.05;2-0-1} = \chi^2_{0.05;1} = 3.841$

(c) Rejection region

i. $\chi^2 > 3.841$

(d) Test-statistic (since degree of freedom is 1, we need to apply Yate's correction)

i.

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{r} \frac{(|O_i - E_i| - 0.5)^2}{E_i} \\ &= \frac{(|75 - 65| - 0.5)^2}{65} + \frac{(|25 - 35| - 0.5)^2}{35} \\ &= 3.967\end{aligned}$$

(e) Conclusion

i. Since $\chi^2 = 3.967 > 3.841$, we reject $H_0$and hence have sufficient evidence to conclude that the vaccine have relationship with recovery from the disease at 5% significance level.