Chapter 7: Chi-square Test

1 Example

1. H_0 : The die is fair $\left(p = \frac{1}{6}\right)$
2. H_1 : The die is unfair $\left(p \neq \frac{1}{6}\right)$
3. At $\alpha = 0.05$, critical value $= \chi^2_{\alpha;m-t-1}$.
4. $$\begin{aligned}\chi^2_{0.05;6-0-1} &= \chi^2_{0.05;5} \\ &= 11.070\end{aligned}$$
5. Rejection region: (remember Chi-square distribution does NOT have a negative region)
$$\chi^2 > 11.07$$
6.

Number	1	2	3	4	5	6	Total
O_i	89	113	98	104	117	79	600
E_i	100	100	100	100	100	100	600
$O_i - E_i$	-11	13	-2	-4	17	-21	
$(O_i - E_i)^2$	121	169	4	16	289	441	
$\frac{(O_i-E_i)^2}{E_i}$	1.21	1.69	0.04	0.16	2.89	4.41	10.40

7. Since $\chi^2 = 10.40 < 11.07$, we failed to reject H_0 at 5% significance level. Hence, we do not have enough evidence to conclude that the observed frequencies are significantly different from those expected of a fair die.

2 Example

Since every city have equal sales potential, therefore, each sale should be $\frac{\sum x}{n}$, which is the mean.

1. Hypothesis
   (a) Claim: Each of the seven cities have equal sales potential. (H_0)
   (b) Opposite: Each of the seven cities do not have equal sales potential. (H_1)
   (c) $H_0 : A = B = C = D = E = F = G$
   (d) $H_1 : A \neq B \neq C \neq D \neq E \neq F \neq G$
2. Find the critical value
   (a) $\alpha = 0.05$
   (b) $\chi^2_{0.05;7-1} = \chi^2_{0.05;6} = 12.592$
3. Find the rejection region
   (a) $\chi^2 > 12.592$
4.

City	A	B	C	D	E	F
O_i	120	185	260	190	210	175
E_i	200	200	200	200	200	200
$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(120-200)^2}{200} = 32$	$\frac{(185-200)^2}{200} = 1.125$	$\frac{(260-200)^2}{200} = 18$	$\frac{(190-200)^2}{200} = 0.5$	$\frac{(210-200)^2}{200} = 0.5$	$\frac{(175-200)^2}{200} = 3.125$

5. Since $\chi^2 = 73.25 > 12.592$, we reject $H_0$and do not have enough evidence to conclude that each of the seven cities have equal sales potential.

3 Example

1. Hypothesis
   (a) Claim: The local hospital follows the national pattern. (H_0)
   (b) Oppo: The local hospital do not follow the national pattern. (H_1)
2. Find the rejection region
   (a) $\alpha = 0.05$
   (b) $\chi^2_{0.05;7-1} = \chi^2_{0.05;6} = 12.592$
   (c) Rejection region: $\chi^2 > 12.592$
3. Find the Chi-square score
4.

Number of times admitted	1	2	3	4	5	6	7	Total
O_i	165	79	50	44	32	20	10	400
E_i	400*40%=160	80	56	40	32	24	8	
$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(165-160)^2}{160}$	$\frac{(79-80)^2}{80}$	$\frac{(50-56)^2}{56}$	$\frac{(44-40)^2}{40}$	$\frac{(32-32)^2}{32}$	$\frac{(20-24)^2}{24}$	$\frac{(10-8)^2}{8}$	2.378

5. Since $\chi^2 = 2.378 < 12.592$, we failed to reject H_0. Hence, we conclude that the local hospital follows the national pattern.

4 Example

1. Hypothesis
   (a) Claim: The results support the theory (H_0)
   (b) Opposite: The results do not support the theory (H_1)
2. Find the rejection region
   (a) $\chi^2_{0.05;3-1} = 5.991$
   (b) $\chi^2 > 5.991$
3. Calculate the Chi-score
   (a)

Color	O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
Red	84	83.25	0.00676
Blue	92	83.25	0.91967
Purple	157	166.5	0.54204
	333		1.46847

4. $\chi^2 = 1.46847 < 5.991$. We failed to reject H_0 at $\alpha = 0.05$ and conclude that the result support the genetic theory.

5 Example

1. Hypothesis
   (a) Claim: The random number generator is working incorrectly. H_1
   (b) Hypo: The random number generator is working correctly. H_0
2. The critical value
   (a) $\chi^2_{0.05;6-1} = 11.070$
   (b) Rejection region: $\chi^2 > 11.070$
3. Find the test statistic

(a)

Interval	O_i	Probability, P_i	$E_i = 500 * P_i$	$\frac{(O_i-E_i)^2}{E_i}$
$s < 4$	10	$P\left(Z < \frac{4-6}{1}\right) = P(Z < -2) = 0.02275$	11.375	0.16621
$4 \leq s < 5$	75	0.13595	67.975	0.74601
$5 \leq s < 6$	163	0.34130	170.65	0.34294
$6 \leq s < 7$	174	0.34130	170.65	0.06576
$7 \leq s < 8$	66	0.13595	67.975	0.05738
$s \geq 8$	12	0.02285	11.375	0.3434
Total		1.0000	500	**1.39264**

(b) Note: For the probability, use the Z-score

$$P(s > \bar{x}) = P\left(Z > \frac{\bar{x} - \mu}{\sigma}\right)$$

(c) Use Normal Distribution Calculator from OnlineStatBook to quickly calculate without converting to normal.

4. Conclusion

   (a) Since $\chi_2 = 1.39264 < 11.070$, we failed to reject H_0, and hence conclude that the random number generator is working correctly.

6 Example

1. Hypothesis

   (a) Claim: The distribution of X is $B(3, 0.25)$ (H_0)

   (b) Oppo: The distribution of X is not $B(3, 0.25)$ (H_1)

2. Find the critical value

$$\alpha = 0.01$$

   (a) Critical value

$$\begin{aligned}\chi^2_{\alpha;m-t-1} &= \chi^2_{0.01;3-0-1} \\ &= \chi^2_{0.01;2} \\ &= 9.210\end{aligned}$$

3. Rejection region

$$\chi^2 > 9.210$$

   (a) If this is true, the probability distribution of X is

$$P(X = x) =^3 C_x (0.25)^x \, 0.75^{3-x}$$

4.

X	O_i	Probability, p_i	$E_i = 64p_i$	$\frac{(O_i-E_i)^2}{E_i}$	Total
0	21	$\frac{21}{21+31+12+0} = 0.4219$	27	1.333	
1	31	0.4219	27	0.5926	
2,3	12	0.1563	10	0.4	2.3259

   (a) $p_0 =^3 C_0 * (0.25)^0 * (0.75)^3 = 0.4219$

   (b) $p_1 =^3 C_1 * (0.25)^1 * (0.75)^2 = 0.4219$

   (c) $p_{2+3} =^3 C_2 * (0.25)^2 * (0.75)^1 + {}^3C_3 * (0.25)^3 = 0.1563$

5. Since $\chi^2 = 2.3259 < 9.210$. We failed to reject H_0, and hence do not have enough evidence to show that the recorded data do not fit $B(3, 0.25)$.

7 Example

Note: We should not round it from 2.02 to 2.00, oterhwise it will be inaccurate during exams.

8 Example

8.1 Extra Notes

By following the probability multiplication rule, if they are independent, the expected frequency for a field in a two-way table will be

1. Proportion probability of something happening in the row, multiple by the proportion probability of something happening in a column (to get row * column proportion) and then multipled by the sample size.
2. Eg:

$$1000\,(sample) * \left(\frac{100}{1000}\,(row) * \frac{100}{1000}\,(column)\right) = 10\,(expected)$$

8.2 Example

1. Hypothesis
   (a) The number of accidents depend on the visits by the inspector. (H_0)
   (b) The number of accidents do not depend on the visits by the inspector. (H_1)
2. Critical value

$$\begin{aligned}\chi^2_{0.05;(4-1)(2-1)} &= \chi^2_{0.05;3} \\ &= 7.815\end{aligned}$$

   (a) Note : $D.O.F. = (Row - 1)\,(Column - 1)$
   (b)

	Number of accidents				
	0	1	2	3	Total row
Visit	33	8	5	4	**50**
Expected	27.77778	13.88889	5.555556	2.777778	
No visit	67	42	15	6	**130**
Expected	72.22222	36.11111	14.44444	7.222222	
Total column	**100**	**50**	**20**	**10**	
				Total acc.	**180**

3. Rejection region

$$\chi^2 > 7.815$$

4. Find the test-statistic

$$\begin{aligned}\chi^2 = &\frac{(33-27.778)^2}{27.778} + \frac{(8-13.889)^2}{13.889} + \frac{(5-5.556)^2}{5.556} + \frac{(4-2.778)^2}{5.556} + \frac{(67-72.222)^2}{72.222} + \frac{(42-36.111)^2}{36.111} \\ &+ \frac{(15-14.444)^2}{14.444} + \frac{(6-7.222)^2}{7.222} \\ = &\,5.3692\end{aligned}$$

5. Conclusion
   (a) Since $\chi^2 = 5.3692 < 7.815$. We failed to reject H_0, and hence do not have enough evidence to conclude that the number of accidents do not depend on the visits by the inspector.

9 Example

10 Example

1. Hypothesis
   (a) Hypo: The coin is fair (H_1)
   (b) Oppo: The coin is unfair (H_0)
2. Critical value at $\alpha = 0.05$
   (a) Critical value: $\chi^2_{0.05;2-0-1} = \chi^2_{0.05;1} = 3.841$
3. Rejection range: $\chi^2 > 3.841$
4. Test-statistics

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{m} \frac{(|O_i - E_i| - 0.5)^2}{E_i} \\ &= \frac{(|115 - 100| - 0.5)^2}{100} + \frac{(|85 - 100| - 0.5)^2}{100} \\ &= 4.205\end{aligned}$$

11 Example

1. Hypothesis
   (a) (H_0) There is no association between taking the new drug and attack by the disease
   (b) (H_1) There is an association between taking the new drug and attack by the disease
2. Critical value
   (a) $\chi^2_{0.05;(2-1)(2-1)} = \chi^2_{0.05;1} = 3.841$
3. Rejection range
   (a) $\chi^2 > 3.841$
4. Test-statistic (apply Yate's correction because $D.O.F. = 1$)
   (a)

	Drugged	Drugles
Attacked	24	32
Expected	(35.47)	(20.53
Not attacked	52	12
Expected	(40.53)	(23.47
Total column	76	44

   (b) $\chi^2 = \frac{(|24-35.47|-0.5)^2}{35.47} + \frac{(|32-20.53|-0.5)^2}{20.53} + \frac{(|52-40.53|-0.5)^2}{40.53} + \frac{(|12-23.47|-0.5)^2}{23.47} = 17.351$
5. Conclusion
   (a) Since $\chi^2 = 17.351 > 3.841$, we reject H_0. Therefore we have enough evidence to conclude that there is an association between taking the new drug and attack by the disease.