Pearson's Chi-Square Test

Dr Azmi Mohd Tamil

$n \geq 20$

No expected value <5.

drtamil@gmail.com 2015

Concept

- Basically we are comparing the rate of the disease among the exposed against the rate of the disease among the non-exposed.
- If the exposure is related to the disease, we would expect those exposed to have a significantly higher ("causative") or lower ("protective") rate of the disease compared to those unexposed.

	Disease +	Disease -	TOTAL
Exposure +	a	b	a+b
Exposure -	c	d	c+d
TOTAL	a+c	b+d	n

drtamil@gmail.com 2015

e.g. Obesity & Diabetes Mellitus

Observed	Diabetes YES	Diabetes NO	TOTAL
Overweight	32 (32%)	68	100
Normal	7 (7%)	93	100
TOTAL	39 (19.5%)	161	200

drtamil@gmail.com 2015

Research Question

- Are overweight respondents more likely to develop diabetes mellitus (DM) compared to normal weight respondents?
- If it is true, then we expect a higher rate of DM among overweight respondents (32%) compared to among normal weight respondents (7%).
- Chi-Square test whether the rate difference (32% vs 7%) is significantly different than the expected values.

drtamil@gmail.com 2015

Null hypothesis

- Null hypothesis assumes that there is no association between overweight and diabetes mellitus. As though being overweight has no effect on the risk of developing diabetes.
- If being overweight has no effect on developing diabetes mellitus, then the rate of diabetes mellitus for both overweight and normal weight should be the same. This is the expected value as calculated in the expected table.
- The statistical test conducted is to decide whether or not to reject the null hypothesis. So if the observed table is sig. diff. than expected table, null hypothesis is rejected.

drtamil@gmail.com 2015

Null hypothesis

- There is no association between overweight and diabetes mellitus.
- Therefore there is no difference of DM (disease) rate among the overweight (exposed) and normal weight (non-exposed)

Expected	Diabetes	No Diabetes	Total
Overweight	19.5 (19.5%)	80.5 (80.5%)	100
Normal weight	19.5 (19.5%)	80.5 (80.5%)	100
Total	39 (19.5%)	161 (80.5%)	200

If there is no difference of diabetes mellitus rate between the two groups, then the rate of diabetes mellitus is similar between the two groups (overweight=19.5% & normal weight=19.5%).

drtamil@gmail.com 2015

Why do we need to calculate the expected value?

- Because we are testing whether to reject or not the null hypothesis. Null hypothesis stated that there is no difference of diabetes mellitus rate between overweight group and normal weight group.
- Expected value is the value if there is no difference of diabetes mellitus rate between the overweight and normal weight groups.

drtamil@gmail.com 2015

Why do we need to calculate expected value?

Observed	Diabetes	No Diabetes	Total
Overweight	32 (32%)	68 (68%)	100
Normal weight	7 (7%)	93 (93%)	100
Total	39 (19.5%)	161 (80.5%)	200

Expected	Diabetes	No Diabetes	Total
Overweight	19.5 (19.5%)	80.5 (80.5%)	100
Normal weight	19.5 (19.5%)	80.5 (80.5%)	100
Total	39 (19.5%)	161 (80.5%)	200

If there is no difference of diabetes mellitus rate between the two groups, then the diabetes mellitus rate is similar between the two groups (overweight=19.5% & normal weight=19.5%).

drtamil@gmail.com 2015

FORMULA

$$\chi^2 = \sum \left[\frac{(O-E)^2}{E} \right]$$

- How different is the observed value compared to the expected value? To get the difference, we deduct the expected value from the observed value. By squaring the difference, we nullify the negative value that may arises. Thus Chi-"Square"! Then that is compared against the expected value (as in the formula above).
- Therefore the larger the difference between the rate of the disease (DM) among the exposed (overweight) compared to the rate of the disease among the non-exposed (normal weight), the larger is the value of the chi-square. The larger the value of chi-square, the smaller is the value of p, therefore more significant.

drtamil@gmail.com 2015

Reject or Not the Null hypothesis

- Basically we are testing whether to reject or not the null hypothesis.
- Therefore if the observed table is similar to the expected table, then we will not reject the null hypothesis

Expected	Diabetes	No Diabetes	Total
Overweight	19.5 (19.5%)	80.5 (80.5%)	100
Normal weight	19.5 (19.5%)	80.5 (80.5%)	100
Total	39 (19.5%)	161 (80.5%)	200

If the observed table is significantly different than the expected table, then we will reject the null hypothesis.

drtamil@gmail.com 2015

Is the observed table similar of different than expected table?

Observed	Diabetes	No Diabetes	Total
Overweight	32 (32%)	68 (68%)	100
Normal weight	7 (7%)	93 (93%)	100
Total	39 (19.5%)	161 (80.5%)	200

Expected	Diabetes	No Diabetes	Total
Overweight	19.5 (19.5%)	80.5 (80.5%)	100
Normal weight	19.5 (19.5%)	80.5 (80.5%)	100
Total	39 (19.5%)	161 (80.5%)	200

We can clearly see that there is a huge difference between the two tables. Chances are that difference is statistically significant. For that test we use Pearson Chi-Square Test.

drtamil@gmail.com 2015

Example:

- $\chi^2 = \frac{(32-19.5)^2}{19.5} + \frac{(7-19.5)^2}{19.5} + \frac{(68-80.5)^2}{80.5} + \frac{(93-80.5)^2}{80.5}$
- $\chi^2 = 8.0128+8.0128+1.9410+1.9410=19.9076$
- df = (2-1)(2-1) = 1
- Critical value for df = 1 for p=0.05 is 3.84,
- The calculated χ^2 is larger than the critical value, therefore the null hypothesis is rejected.
- Conclusion: There is a sig. diff. of diabetes mellitus rate between overweight & normal weight. The overweight respondents has a significantly ($p<0.05$) higher rate of DM (32%) compared to normal weight respondents (7%).
- Hint: Memorise critical values of 3.84 for df=1 and 5.99 for df=2.

drtamil@gmail.com 2015

Table 3 : Percentage point of X^2

d.k.	P Value 0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	4.35	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	6.35	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	7.34	10.22	13.36	15.51	17.5	20.09	21.96	26.13
9	8.34	11.39	14.68	16.92	1	21.67	23.59	27.88
10	9.34	12.55	15.99	18.31	20.4	1	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	3.73	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.2	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74		29.69	34.53
14	13.34	17.12	21.06	23.68	26.12		31.32	36.12
15	14.34	18.25	22.31	25.00	27.49	3	32.80	37.70
16	15.34	19.37	23.54	26.30	28.85	32	4.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.4	72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.8	5	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19		43.82
20	19.34	23.83	28.41	31.41	34.17	37.57		45.32
21	20.34	24.93	29.62	32.67	35.48	38.93		46.80
22	21.34	26.04	30.81	33.92	36.78	40.29		48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	4	9.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45	18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	
27	26.34	31.53	36.74	40.11	43.19	46.96	49.64	
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	9
70	69.33	77.58	85.53	90.53	95.02	100.43	104.22	112.32
80	79.33	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	89.33	98.65	107.57	113.15	118.14	124.12	128.30	137.21
100	99.33	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Refer to Table 3.

Look at df = 1.

$X^2 = 19.91$, larger than 10.83 (p=0.001)

10.83(p=0.001) < 19.91

Therefore if $\chi^2 = 19.91$, p<0.001.

Since the calculated p < 0.05, null hypothesis rejected.

d.k.	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	14.86	18.47

Summary

- Hypothesis – Overweight higher risk of developing DM compared to normal weight. Therefore overweight has higher rate of DM (32%) compared to normal weight (7%).
- Null hypothesis
  - No difference of DM rate between overweight and normal weight; or
  - There is no association between being overweight and the risk of developing DM.
- Suitable statistical test – Pearson Chi-Square (n=200, no expected value less than 5).

drtamil@gmail.com 2015

Summary

- Calculation;
  - $\chi^2 = \frac{(32-19.5)^2}{19.5} + \frac{(7-19.5)^2}{19.5} + \frac{(68-80.5)^2}{80.5} + \frac{(93-80.5)^2}{80.5}$

    $= 8.0128+8.0128+1.9410+1.9410=19.9076$
  - df = (2-1)(2-1) = 1
  - $p < 0.001$
- Null hypothesis rejected because $p < 0.05$
- Conclusion: There is a sig. diff. of DM rate between overweight & normal weight respondents. The overweight respondents has a significantly ($p<0.05$) higher rate of DM (32%) compared to normal weight respondents (7%).

drtamil@gmail.com 2015