Chi-square test

Considering relationships

Cross-tabulation

Considering more than one variable at a time

- Cross tabulation of two or more variables may reveal additional information beyond simple descriptives
- Nominal measurements are counted and their frequencies are entered into appropriate cells

Cross-tabulation

Smokers by gender

		Variable 1 Smoker: Yes	No
Variable 2 Gender	Male	50	125
	Female	80	75

What can we conclude by looking at this table?

Cross-tabulation

More information

- Row and column totals as well as percentages add information

		Variable 1 Smoker: Yes	No	
Variable 2 Gender	Male	50	125	175
	Female	80	75	155
		130	200	330

Cross-tabulation

Question still remains

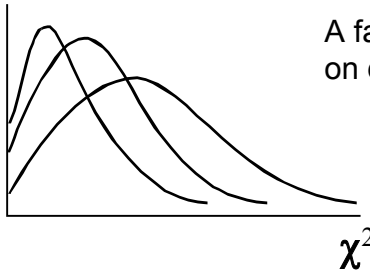
- Are smoking and gender related?
- The answer lies in Chi-square test
- The hypotheses are:
  - H_o: two variables are independent
  - H_a: two variables are related

    Testing for the presence of a relationship. Presence of relationship does not imply causality.

Chi-square test

Test of independence

- Chi-square (pronounce as Kai) is another probability distribution just like the normal distribution



A family of distributions based on degrees of freedom (df)

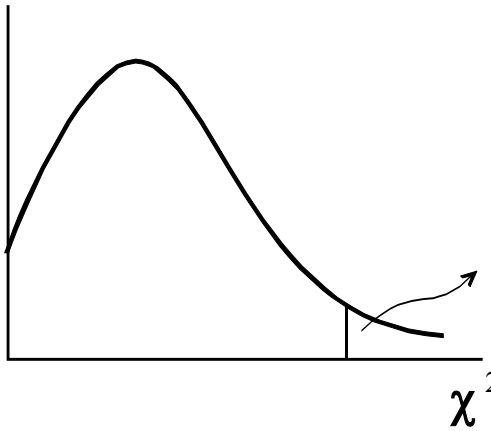
Chi-square test

Statistical reasoning

- With a given degrees of freedom and level of significance, there exists a theoretical χ^2 value
- If the calculated χ^2 statistic is greater than the theoretical value we will reject the null hypothesis

Chi-square test

A hypothetical curve



If:
degrees of freedom = 4
$\alpha = 0.05$
(level of significance)

Chi-square test

How to calculate the sample statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where

O_i – Observed frequencies

E_i – Expected frequencies

Chi-square test

Expected frequencies

- Think of a method to fill in the table so that the row and the column totals remain unchanged

Variable 2: Gender	Variable 1: Smoker – Yes	Smoker – No	
Male			175
Female			155
	130	200	330

Calculation

Calculate expected frequencies

- Expected frequency of male smokers:

Variable 2: Gender	Variable 1: Smoker – Yes	Smoker – No	
Male			175
Female			155
	130	200	330

Calculation

Use the observed and expected frequencies

	Observed: Smoker – Yes	Observed: Smoker – No	Expected: Smoker – Yes	Expected: Smoker – No	
Male	50	125			175
Female	80	75			155
			130	200	330

Calculation

Calculate the Chi-square statistic

$$\frac{(50-69)^2}{69} = 5.23$$

$$\frac{(125-106)^2}{106} = 3.41$$

$$\frac{(80-61)^2}{61} = 5.92$$

$$\frac{(75-94)^2}{94} = 3.84$$

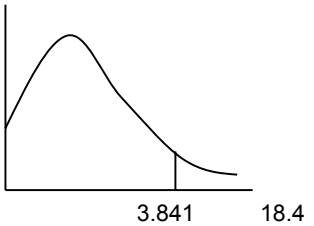
$$\chi^2 = 18.4$$

Interpretation

Compare it to a standard

- Look up the table value of Chi-square for 1 degree-of-freedom and 5% level of significance

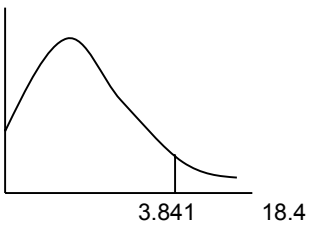
$$\chi^2_{\alpha = 0.05,\ 1\ \mathrm{df}} = 3.841$$



Interpretation

Compare it to a standard

- If the null hypothesis were true, we expect the calculated values to fall below 3.841 95% of the time.



Since our calculated value, 18.4, falls above this, we reject the hypothesis of independence and conclude that these variables are dependent on each other or related to each other.