Observed frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	938	119	1057
	Did not vote	Count	99	42	141
Total		Count	1037	161	1198
Total (%)			0.87%	0.13%	100%

How likely is it that we obtained these numbers by chance?

That is: if there was no relationship between ownership and voting in the population, how likely is it we get numbers which are so far away from what we would expect (or more extreme)?

Expected frequencies

			Owner or tenant		Total	Total (%)
			Owner	Tenant		
Vote in 2007 election	Did vote	Count	A	C	1057	88%
	Did not vote	Count	B	D	141	12%
Total		Count	1037	161	1198	100%
Total (%)			0.87%	0.13%	100%	

- If ownership and vote were not related, how many respondents should we expect in cell A?

Expected frequencies

			Owner or tenant		Total	Total (%)
			Owner	Tenant		
Vote in 2007 election	Did vote	Count	A = 88%*0.87%	C = 88%*0.13%	1057	88%
	Did not vote	Count	B = 12%*0.87%	D = 12%*0.13%	141	12%
Total		Count	1037	161	1198	100%
Total (%)			0.87%	0.13%	100%	

- If ownership and vote were not related, how many respondents should we expect in cell A?
  - If among all voters, 88% did vote, we would expect that among owners, also 88% would vote.
  - If among all owners, 87% did vote, we would expect that among voters, also 87% would vote.

Expected frequencies

			Owner or tenant		Total	Total (%)
			Owner	Tenant		
Vote in 2007 election	Did vote	Count	A	C	1057	88%
	Did not vote	Count	B	D	141	12%
Total		Count	1037	161	1198	100%
Total (%)			0.87%	0.13%	100%	

- If ownership and vote were not related, how many respondents should we expect in cell A?
- Expected frequency (f_e)= row margin$*\frac{\text{column margin}}{\text{total}}$
- $f_e = 1037 * \frac{1057}{1198}$
- f_e = 1037 * 0.88 = 914.9

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	A	C	1057
	Did not vote	Count	B	D	141
Total		Count	1037	161	1198

$$\text{Expected frequency } (f_e) = \frac{\text{row margin} * \text{colum margin}}{\text{total}}$$

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	A	C	1057
	Did not vote	Count	B	D	141
Total		Count	1037	161	1198

- B: f_e = 1037 * 141 / 1198 = 122.1
- C: f_e = 1057* 161 / 1198 = 142.1
- D: f_e = 141 * 161 / 1198 = 18.9

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	914.9	142.1	1057
	Did not vote	Count	122.1	18.9	141
Total		Count	1037	161	1198

Observed frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	938	119	1057
	Did not vote	Count	99	42	141
Total		Count	1037	161	1198

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	914.9	142.1	1057
	Did not vote	Count	122.1	18.9	141
Total		Count	1037	161	1198

Observed frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	938	119	1057
	Did not vote	Count	99	42	141
Total		Count	1037	161	1198

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	914.9	142.1	1057
	Did not vote	Count	122.1	18.9	141
Total		Count	1037	161	1198

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Cell A: $\frac{(f_o - f_e)^2}{f_e} = \frac{(938-914.9)^2}{914.9} = \frac{533.61}{914.9} = 0.58$

Observed frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	938	119	1057
	Did not vote	Count	99	42	141
Total		Count	1037	161	1198

Expected frequencies

			Owner or tenant		Total
			Owner	Tenant	
Vote in 2007 election	Did vote	Count	914.9	142.1	1057
	Did not vote	Count	122.1	18.9	141
Total		Count	1037	161	1198

Cell B: $\frac{(f_o - f_e)^2}{f_e} = \frac{(99 - 122.1)^2}{122.1} = \frac{533.61}{122.1} = 4.37$

Cell C: $\frac{(f_o - f_e)^2}{f_e} = \frac{(119 - 142.1)^2}{142.1} = \frac{533.61}{142.1} = 3.76$

Cell D: $\frac{(f_o - f_e)^2}{f_e} = \frac{(42 - 18.9)^2}{18.9} = \frac{533.61}{18.9} = 28.23$

Chi squared

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = .58 + 4.37 + 3.76 + 28.23 = 36.94$$

Interesting, but what does that mean?

Chi squared

- We need to compare the chi squared we **obtained** with the **critical** value for chi squared.

- If $\chi^2_{obtained} > \chi^2_{critical}$ we can conclude that it is unlikely that the relationship we found is just due to sampling error.

The cricical value

- First, we need to set a **confidence level**, normally 95%
- This corresponds to a ***p*** **value** of 0.05 (1 – 95/100).
- Second, we need to know the **degrees of freedom**: df = (c-1)(r-1)

The critical value

In our example:

- The degrees of freedom:
  - 2 rows
  - 2 columns
  - df = (2 -1) * (2 – 1) = 1 * 1 = 1
- The critical value corresponding df = 1 and p = 0.05 is found in Field, appendix A.4:

A.4. Critical values of the chi-square distribution

	p			
df	0.05	0.01	df	0.05
1	3.84	6.63	25	37.65
2	5.99	9.21	26	38.89
3	7.81	11.34	27	40.11
4	9.49	13.28	28	41.34
5	11.07	15.09	29	42.56
6	12.59	16.81	30	43.77
7	14.07	18.48	35	49.80

Comparing obtained and critical value

- $\chi^2_{obtained} = 36.94$
- $\chi^2_{critical} = 3.84$

- As $\chi^2_{obtained} > \chi^2_{critical}$ we conclude that there is a statistically significant relationship.