TOPIC: The χ^2 (chi-square) distribution

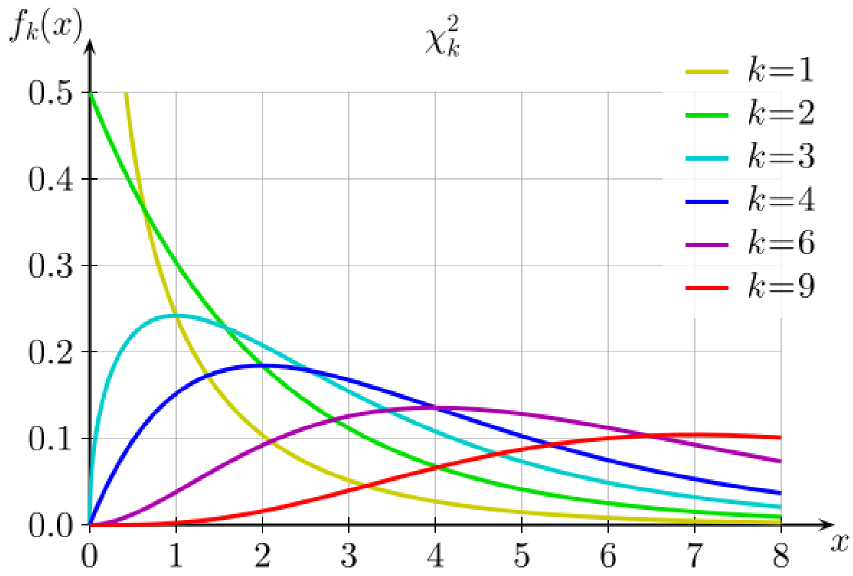
The χ^2 distribution allows us to test different hypotheses about frequencies (or proportions).

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

f_o = observed frequency
f_e = expected frequency

Since χ^2 is made up of a sum of squares it can never be negative, and the minimum value it can take on is 0. When f_o=f_e, the value of the χ^2 statistic is 0.

As with the Student's t distribution, χ^2 is a series of distributions, a family of curves. There is a different x distribution for each value of degrees of freedom.



Source: http://en.wikipedia.org/wiki/Chi-squared_distribution

There are 3 types of x tests:

χ^2 Test for Differences among the Proportions of C Populations
χ^2 Test for Independence – to determine whether two nominal variables are related
χ^2 Goodness of Fit Test – to determine whether a sample may be regarded as a random sample drawn from a population with a specified distribution (e.g., normal, binomial, etc.)

χ^2 Test for Differences among the Proportions of 2 Populations

In a sample of 200 individuals, researchers are interested in the proportions of males and females who use a particular brand of shampoo.

	MALE	FEMALE	
USER	20	50	70
NON-USER	80	50	130
	100	100	200

Hypotheses:

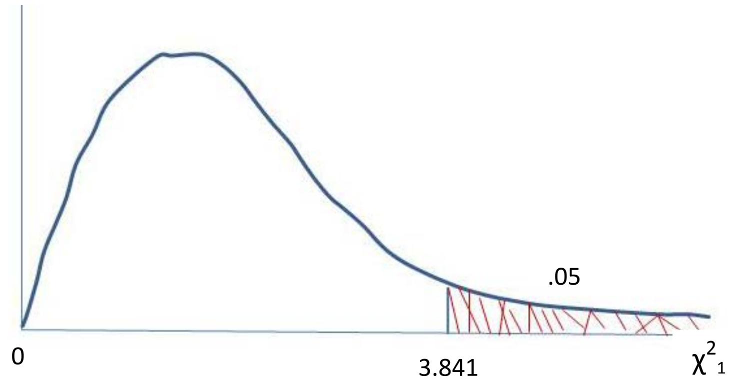
H_0: $P_1 = P_2$ (in other words, sex and usage are independent)
H_1: $P_1 \neq P_2$ (in other words, sex and usage are related)

Test at α=.05

Critical value of the χ^2 test statistic:

We need to know the degrees of freedom for the χ^2 test statistic.
Degrees of freedom = (r-1)(c-1)

So in this example, df = (2-1)(2-1) = 1. From the table, the critical value of the χ^2 test statistic, with 1 degree of freedom and a tail probability (α) of .05, is: 3.841



Calculated value of the χ^2 test statistic:

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

We can compute the expected frequencies (f_e) for each cell by:

$$f_e = \frac{row\ total * column\ total}{n}$$

Expected frequencies (f_e) are sometimes called theoretical frequencies (f_t).
Expected Frequencies (in red):

	MALE	FEMALE	
USER	35	35	70
NON-USER	80	50	130
	100	100	200

f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
---	---	-----	--------	------------
20	35	-15	225	6.43
80	65	+15	225	3.46
50	35	+15	225	6.43
50	65	-15	225	3.46
		0		19.78

[Make sure all the f_e are ≥ 5 or you may not have a valid χ^2 test.]

Conclusion:

The calculated value of the test statistic (19.78) is greater than the critical value at the .05 level of significance (3.841). Hence,

Reject H_0
Women are more likely to use the product than men.

NOTE: When testing for differences between proportions of two populations, we can use either the Z test or the χ^2 test with exactly the same results. In this case, we could have done the problem using the Z test for the difference between 2 proportions:

$$Z = \frac{.20-.50}{\sqrt{(.35)(.65)\left(\frac{1}{100}+\frac{1}{100}\right)}} = \frac{-.30}{.06745} = -4.448$$

At α=.05 and a two-tail test, the critical values of Z are ±1.96. Hence, Reject H_0.

[NOTE that $\chi^2=Z^2$.
$3.841 = 1.96^2$ and $19.78 = 4.448^2$]

χ^2 Test for Differences among the Proportions of C Populations

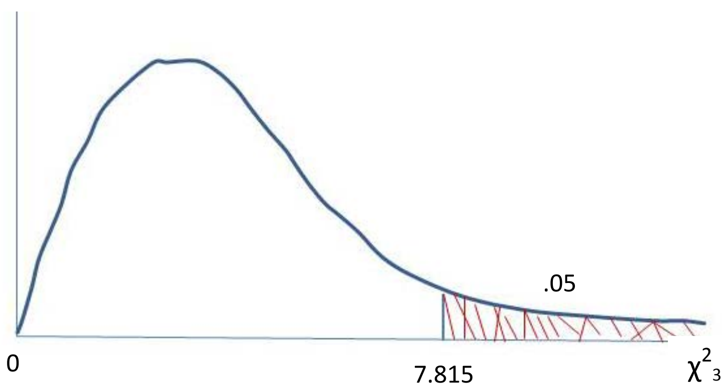
The advantage of the χ^2 test over the Z test for proportions is that the x test statistic can be used to test for differences among the proportions of more than 2 populations.

EXAMPLE: A research studies 4 ways of selling business insurance. Is there a difference in success rate? Test at α = .05.

	Cold call	Referred lead	Mass email	Personal email	Totals
Appointment Made	6	29	11	16	62
No Appointment Made	94	71	89	84	338
Totals	100	100	100	100	400

H_0: $P_1 = P_2 = P_3 = P_4$
H_1: At least one Proportion is different from the others

Degrees of freedom = (r-1)(c-1) = (2-1)(4-1) = 3
From the table, the critical value of the χ^2 test statistic, with 3 degrees of freedom and a tail probability (α) of .05, is: 7.815.



Expected frequencies:

	Cold call	Referred lead	Mass email	Personal email	Totals
Appointment Made	15.5	15.5	15.5	15.5	62
No Appointment Made	84.5	84.5	84.5	84.5	338
Totals	100	100	100	100	400

The calculated value of the test statistic is computed as $\chi^2 = \frac{(f_o - f_e)^2}{f_e}$ over all cells.

f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
---	---	-----	--------	------------
6	15.5	-9.5	90.25	5.82
29	15.5	13.5	182.25	11.76
11	15.5	-4.5	20.25	1.31
16	15.5	0.5	0.25	0.02
94	84.5	9.5	90.25	1.07
71	84.5	-13.5	182.25	2.16
89	84.5	4.5	20.25	.24
84	84.5	-0.5	0.25	0.00
		0		**22.38**

$\chi^2_{(2-1)(4-1)} = 22.38$

Reject H_0

χ^2 Test for Independence – OPTIONAL TOPIC

H_0: Rows and Columns are independent
H_1: Rows and Columns are related

EXAMPLE: Is there a relationship between social class and brand preference? Test at α = .05.

Observed frequencies (f_o):

BRAND PREFERENCE	UPPER CLASS	MIDDLE CLASS	LOWER CLASS	
Brand A	130	100	70	300
Brand B	30	400	70	500
Brand C	20	60	20	100
Brand D	20	40	40	100
	200	600	200	1,000

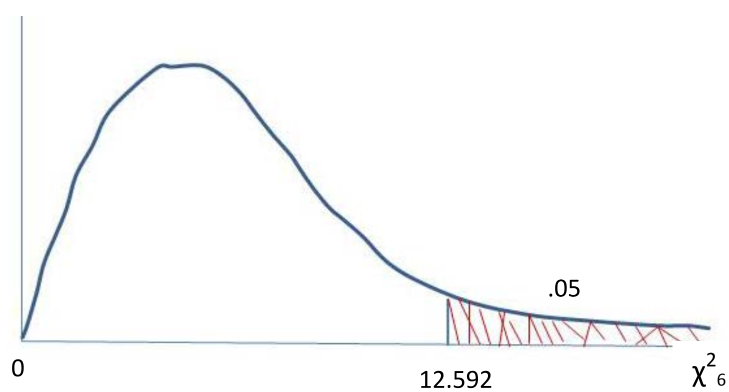
Expected frequencies (f_e):

BRAND PREFERENCE	UPPER CLASS	MIDDLE CLASS	LOWER CLASS	
Brand A	60	180	60	300
Brand B	100	300	100	500
Brand C	20	60	20	100
Brand D	20	60	20	100
	200	600	200	1,000

[Note that all the fe's are ≥ 5.]

H_0: Social Class and Brand Preference are independent
H_1: Social Class and Brand Preference are related

Degrees of freedom = (r-1)(c-1) = (4-1)(3-1) = 6
From the table, the critical value of the χ^2 test statistic, with 6 degrees of freedom and a tail probability (α) of .05, is: 12.592.



f_o	f_e	f_o-f_e	$(f_o-f_e)^2$	$(f_o-f_e)^2/f_e$
---	---	-----	--------	------------
130	60	70	4900	81.67
30	100	-70	4900	49.00
20	20	0	0	0
20	20	0	0	0
100	180	-80	6400	35.56
400	300	100	10000	33.33
60	60	0	0	0
40	60	-20	400	6.67
70	60	10	100	1.67
70	100	-30	900	9.00
20	20	0	0	0
40	20	20	400	20.00
		0		**236.90**

Already larger than the critical value.

The calculated value of $\chi^2_{(4-1)(3-1)=6} = 236.90$

Reject H_0

χ^2 Goodness of Fit Test – OPTIONAL TOPIC