In this lecture we'll begin by addressing some final issue concerning chi-squared tests, including it's assumptions and limitations. Then we'll discus s an alternative approach known as 'exact tests'. Following this we'll conduct a classroom exercise.

1. Assumptions and Limitations of Chi-Squared Tests

Degrees of Freedom

Before proceeding to the assumptions and limitations of chi-squared tests, let's revisit the issue of degrees of freedom. In the last lecture we learned that for a chi-squared independence test of two variables (i.e., with data organized as a contingency table), the *df* are $(R - 1)(C - 1)$, where R and C are the number of rows and columns. Let's examine more closely why this is so.

Consider the 2 x 2 contingency table below. Assume that the row and column marginal frequencies (grey cells) are known. If we then know that the frequency of cell (1, 1) is, say *a*, this one number determines the frequencies of the other three cells (shaded blue).

	Variable 2		
Variable 1	**Level 1**	**Level 2**	**Total**
Level 1	*a*	(30 – *a*)	*30*
Level 2	(40 – *a*)	(30 + *a*)	*70*
Total	*40*	*60*	*100*

Thus we see that, if the observed marginal frequencies are fixed (which is an assumption of the chi-squared test of independence), then if we stipulate a frequency for *a*, the other three frequencies follow automatically. The same is true if we stipulate cell (1, 2), cell (2, 1), or cell (2, 2) – i.e., if we fix any one cell, all the other three follow. Thus in a 2 x 2 table, we have only 1 *df*. This is what our formula predicts, because (R – 1) (C – 1) here is (2 – 1) (2 – 1) = 1.

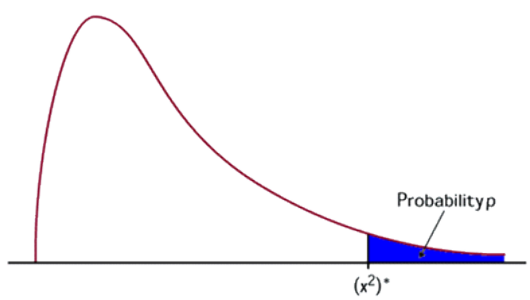
Now consider a larger table, say a 3 x 3 one.

	Variable 2			
Variable 1	**Level 1**	**Level 2**	**Level 3**	**Total**
Level 1	*a*	*b*	60 – (*a* + *b*)	60
Level 2	*c*	*d*	30 – (*c* + *d*)	*30*
Level 3	50 – (*a* + *c*)	40 – (*b* + *d*)	(*a* + *b* + *c* + *d*) – 90	*30*
Total	*50*	*40*	*30*	*120*

Here we find that if we fix the marginal frequencies (gray cells) and stipulate values *a*, *b*, *c* and *d* for the four unshaded cells, then the frequencies in the remaining blue cells are determined. So here we have (*R* – 1) (*C* – 1) = (*3* – 1) (3 – 1) = 4 degrees of freedom. By the same reasoning we may see that for any two-way frequency table, the *df* are (*R* – 1) (*C* – 1).

Discontinuity

Now we'll look at some possible problems with the chi-squared test of independence. The first is the *discontinuity* issue. Recall that the χ^2 distribution for any given *df* has a continuous shape:



And that to evaluate a test statistic like the Pearson X^2 we consider areas in the upper tail of this distribution.

The continuous distribution implies that *any* value (≥ 0) of χ^2 (*x*-axis) is possible. However in a contingency table, especially with a small N, there are only a finite number of possible ways to arrange the N observations amongst the cells. For example, given an extremely small N in a 2 x 2 table, only a few arrangements are possible.

	Variable 2		
Variable 1	**Level 1**	**Level 2**	**Total**
Level 1	1	0	*1*
Level 2	0	1	*1*
Total	*1*	*1*	*2*

That is, in this case where *N* = 2, if we consider the marginal frequencies fixed there are actually only two possible arrangements (1's on the main diagonal, or 1's on the off-diagonal cells). This is an extreme case but the same principle applies generally: in a two-way frequency table, there are only a finite number of arrangements of N observations, and therefore a finite number of possible values of the Pearson X^2. That means that χ^2 distribution, which is continuous and considers every value as possible, cannot correspond perfectly to our Pearson X^2 test statistic, and that p-value we produce by integrating the χ^2 distribution are here only approximate at best.

Small Frequencies

For this and possibly other reasons, a common rule-of-thumb suggests that the Pearson X^2 test should not be used when more than 20% of the *expected* frequencies of a table are less than 5. And not used for a 2 x 2 table if any expected frequency is less than 5.

Discontinuity Correction

To account for this discontinuity issue, some statisticians add a correction term in the formula to compute the Pearson X^2 statistic:

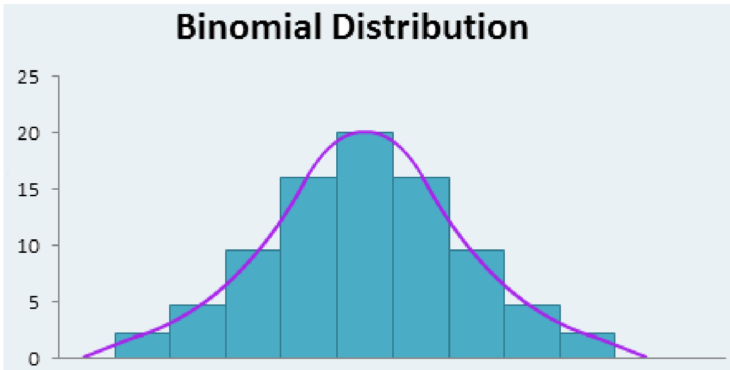
$$X^2_{corrected} = \sum_i \sum_j \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$$

The problem here is that not everyone agrees that the corrected version is better in every case. So now we have two formulas, neither of which is exactly correct – almost a worst state than before ('*A man with two watches never knows the time*.')

2. Exact Tests

It turns out, however, that there is a completely different approach to testing independence of two nominal variables that is exact. These tests are in fact called *exact tests*, and are becoming increasingly popular.

You may recall reading in the book, or mentioned briefly in the lectures, that there is an alternative to the binomial distribution called the 'normal approximation to the binomial.' This normal approximation seeks to cumulative binomial probabilities by making use of the fact that, for large N, and *p* not to close to 0 or 1, the discrete binomial distribution has somewhat the same shape as a normal distribution with mean *Np and variance N p* (1 – *p*).



However you may also recall that I didn't teach the normal approximation to the binomial, suggesting that, since computers now let us calculate cumulative binomial probabilities exactly, even with very large N, the normal approximation – something developed in pre-computer days – was basically no longer necessary.

It turns out that we have the same situation when it comes to chi-squared tests of independence. The approach we've considered in the last two lectures, based on calculating a test statistic and comparing it to a continuous theoretical χ^2 distribution, is just an approximation – yet today we have the computer power to compute exact values, just as we can compute exact binomial probabilities.

Consider again a 2 x 2 table. Suppose the following are our expected probabilities of observations falling in each cell.

	Variable 2		
Variable 1	**Level 1**	**Level 2**	**Total**
Level 1	0.25	0.25	0.5
Level 2	0.25	0.25	0.5
Total	0.5	0.5	1.0

As already stated, if we cross-classify N observations on these two variables, there are only a finite number of possible outcomes. And, given the above probabilities (and using the hypergeometric distribution, which is a generalization of the binomial distribution) we can compute the probability of each possible observed distribution occurring. This means that for some observed outcome – say 100 cases with frequencies of (10, 20, 30, 40), we can compute exactly compute the probability of this outcome or a rarer one (rare here meaning having a lower probability of occurring) out of all possible outcomes. This is a p-value – an exact one – just as we can compute an exact binomial probability for, say, getting 3 or fewer 'heads' out of 10 coin flips.

The exact test for 2 x 2 tables was actually developed around the 1920s by Ronald Fisher (and hence is called Fisher's *exact test*. Being more computationally complex than the chi-squared test, however, it wasn't used much. Now advances both in computing power and algorithm sophistication have made exact tests widely available – not just for 2 x 2 tables, but even for large contingency tables with large *N*'s.

Many people, therefore, would consider these exact tests the state of the art, and would avoid using chi-squared tests with contingency tables altogether.

Example

Yesterday we considered data on voting preferences of male and female voters:

	Outcome	
Treatment	**Failure**	**Success**
Drug 1	10	20
Drug 2	40	30

We'll now analyze the same data using a online exact-test calculator here:
http://vassarstats.net/tab2x2.html

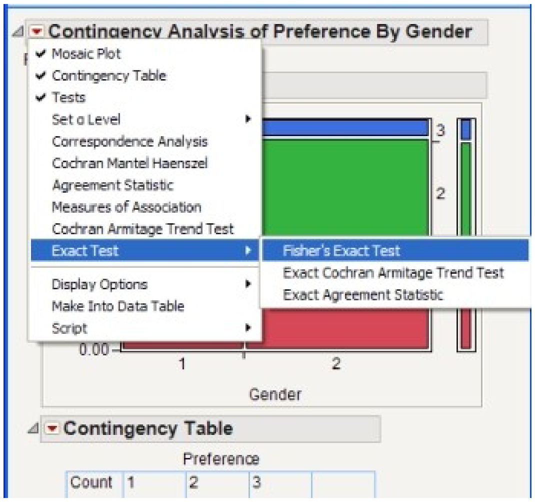
> Note: Online calculators for a wide range of statistical tests are an increasingly common and viable alternative to purchased computer software.

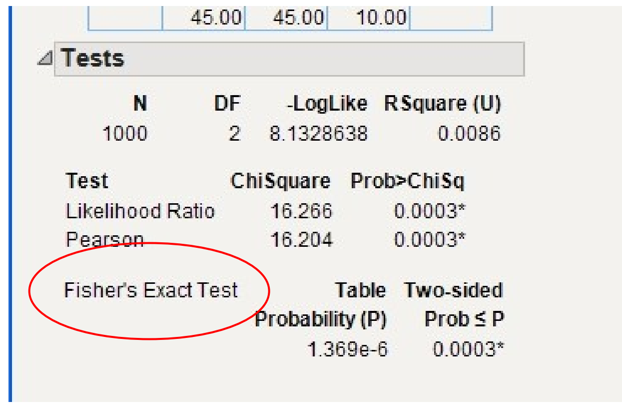
Chi-squared test:	0.0291
Corrected chi-squared test:	0.0495
Exact test:	0.0486

JMP

JMP will also perform an exact test for an R x C table. We first perform a chi-squared test, as shown yesterday. Then, clicking the red arrow in the Contingency Analysis section (top) of the report, we select **Exact Test > Fisher's Exact Test.**

The results will appear in the Tests section.





And now for something completely different....

Previously we've seen how to construct confidence intervals for various values, such as a 95% confidence interval for the mean, or a difference of means. We did this based on the sampling distribution of a statistic assuming samples of some fixed size, N.

In practice, however, and especially in quality applications, a somewhat different problem is encountered. Before a study is conducted, you, as the experimenter or quality engineer, can decide in advance how large N should be. The larger N is, the stronger your conclusions will be from a study, but also the more expensive the study will be.

One way to approach this problem is to ask what sample size N would be necessary to obtain some pre-specified level of precision.

Suppose, for example, we wish to estimate the performance of some manufactured electronic device, measured on some variable, like output voltage. We believe the population standard deviation is 0.8 volts, and we wish to estimate from a sample the mean output voltage in the population.

Question; how many units would we need to sample so that our 95% confidence limits are +/- 1 volt of the population mean?

Hint: to begin, recall the formula for the confidence limits of a sample mean:

$$LL = \overline{X} - z_{crit} \times s_{\overline{X}}$$
$$UL = \overline{X} + z_{crit} \times s_{\overline{X}}$$

The critical z value for a 95% confidence interval is 1.95.

As a class, derive the formula to estimate the required N so that the UL and CL of a 95% credible/confidence interval for the sample mean is exactly +/- E wide, i.e., $CI_{95\%}$ = [mean – E, mean + E].