Objectives

We'll cover the chi-squared (χ^2) test for categorical data (goodness-of-fit test) and extend it to examine whether two categorical variables are related (contingency test). By the way, *chi* is pronounced *k*ai, not *ch*ai. Related supplementary material is presented for those who are interested.

Stuff with a solid edge, like this, is important.

But remember — you can totally ignore stuff with single / double wavy borders.

5.1 The chi-squared (χ^2) test

One categorical variable, two categories

The χ^2 test, sometimes called Pearson's χ^2 test, is all about analysing **categorical data.** Suppose we ask 100 people to choose between chocolate and garibaldi biscuits (so every person falls into one of two categories); 65 choose chocolate and 35 choose garibaldi. Does this differ from chance, i.e. a 50:50 split? The **expected values** based on the null hypothesis are 50 chocolate and 50 garibaldi. The **observed values** are 65 and 35. From this, we can calculate the **χ^2 statistic:**

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O is the observed frequency in each category, and E is the expected frequency. We sum over all the categories. **A big χ^2 means that the observed frequencies differ considerably from the expected frequencies. (Significant values of χ^2 are big. Non-significant values of χ^2 are close to zero.)** If we have c categories, we have **c – 1 degrees of freedom.**

This is called a **goodness-of-fit** test. It asks whether the data (observed values, O) are a good fit to some model (expected values, E).

So in this example, $\chi^2 = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} = 9$. We have two categories, but since we know n (100), then as soon as we know the frequency of one category (chocolate) we automatically know the frequency of the other (garibaldi). So we have only 1 *df*. All we need now is to know the **critical value** of χ^2_1 for our chosen value of α (say 0.05); our handy statistical tables will tell us that this is 3.84. Since our χ^2 value was 9, we can reject the null hypothesis and say that people's preferences differed from chance ($\chi^2_1 = 9.0$, $p < .05$). If we were using a computer, we could derive an exact p value for our χ^2 value of 9 — it's 0.0027 — so we could report our biscuit analysis like this: 'The group's preference differed from chance ($\chi^2_1 = 9.0$, $p = .0027$).'

Note that although the process of testing χ^2 involved a one-tailed test (was χ^2 *bigger* than a critical value?), the process of *obtaining* the value of χ^2 was inherently two-tailed (the way we calculate χ^2 detects observed values that are bigger *or* smaller than the expected value). So the α we use to obtain a critical value of χ^2 is effectively a two-tailed α. For more details on this, see Howell (1997, p. 144).

One categorical variable, more than two categories

This approach can be used for any number of categories, and any expected values. So if a furniture warehouse stocks a vast number of chair backs, chair seats, and chair legs, then we could take random samples of items, classify each item in the sample into one of these three categories (c = 3), and test the hypothesis that in the total stock (the population) these items were in the correct chair-building ratio 1:1:4 using a χ^2_2 test (note 2 degrees of freedom = $c - 1$).

More than one categorical variable (contingency tests)

We're often interested in data that's classified by more than one variable, and in asking whether these variables are *independent* of each other or in some way *contingent* upon each other. Here's an example (see Howell, 1997, p. 144), based on a 1983 study of jury decisions in rape cases. Decisions were classified on two variables: (1) guilty or not guilty; (2) whether the defence alleged that the victim was somehow partially at fault for the rape. The researcher analysed 358 cases:

Obtained values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	153 (*a*)	24 (*b*)	*177*
Victim portrayed as high-fault	105 (*c*)	76 (*d*)	*181*
Total	*258*	*100*	*358*

Now if these two variables (verdict and victim portrayal) are *independent*, then we would expect that $^a/_b = {^c/_d}$ and that $^a/_c = {^b/_d}$. But if they are not independent, we might expect a different picture. We can use a χ^2 test to answer this question. This is called a **contingency test,** because it asks whether one variable is in some way contingent upon the other. The null hypothesis is that the two variables are independent. We can calculate the expected value of each cell as follows:

$$E(row_i, column_j) = \frac{R_i C_j}{n}$$

where $E(row_i, column_j)$ is the expected value of the cell in row i and column j, R_i is the **row total** for row i, C_j is the **column total** for column j, and n is the overall total number of observations.

For our example, we can calculate that $E(1,1) = (177 \times 258) / 358 = 127.559$. We can fill in all the other expected values like this:

Expected values	Guilty verdict	Not guilty verdict	Total
Victim portrayed as low-fault	127.559	49.441	*177*
Victim portrayed as high-fault	130.441	50.559	*181*
Total	*258*	*100*	*358*

Then we can calculate χ^2 in the usual way:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

In general, if we have a table with R rows and C columns, we have **$(R - 1)(C - 1)$ degrees of freedom.** This method extends to any $R \times C$ table.

So in our example, there are four numbers to sum over (you should obtain the answer $\chi^2 = 35.93$), and we have (2–1)(2–1) = 1 *df*. This should make sense: once you know the row and column totals, you need to know only one cell frequency to be able to work out all the others. The critical value of χ^2_1 for $\alpha = 0.05$ is 3.84, so we reject the null hypothesis. When the victim was portrayed as low-fault, the defendant was found guilty 86% of the time, but when the victim was portrayed as high-fault, the defendant was convicted only 58% of the time, and this is a significant difference ($\chi^2_1 = 35.93$, $p < 0.001$).

Assumptions of the χ^2 test

All statistical tests have assumptions. If they are violated, using the test is pointless: the results of the test will not be the probabilities we're interested in, and therefore our conclusions will be meaningless. This is what the χ^2 test assumes:

- **Independence of observations.** In all the examples given so far, each observation has been independent. One person didn't affect another's biscuit choice, and one court case didn't affect another. If this is not the case, you can't use a χ^2 test. In particular, one thing you mustn't do is to analyse data from several subjects when there are multiple observations from one subject, because they won't be independent. (It's possible to analyse data from *only* one subject, because the observations are then *equally* independent, but your conclusion will only tell you something about that one subject.)

- **Normality.** There shouldn't be any very small expected frequencies **(none less than 5),** otherwise the data won't approximate a normal distribution.
(Actually, the "none <5" rule is a bit conservative; it's probably OK to use the test with even smaller expected frequencies if the row totals aren't too dissimilar and neither are the column totals; see Howell, 1997, p. 152 - but no expected value can be zero!)
- **Inclusion of non-occurrences.** To see what this means, let's take an example. Suppose that 17 out of 20 men supported the sale of alcohol in petrol stations, and 11 out of 20 women did. We want to know if significantly more men than women support this idea. This would be **wrong:**

Obtained values	Men	Women
Support booze	17	11

This would give us expected values of 14 and 14 under the null hypothesis of 'no difference', and therefore χ^2_1 = 1.29 (not significant). But this is wrong because we've *lost information* about the total number of responders. We should be doing this:

Obtained values	Men	Women
Yes to booze	17	11
No	3	9

This would give us χ^2_1 = 4.29 (p = 0.038). Including information on non-occurrences is vital — suppose we'd interviewed 2000 men and 17 said yes:

Obtained values	Men	Women
Yes to booze	17	11
No	1983	9

We'd have a totally different picture, which the first table would have missed completely.

5.2 Supplementary material: odds ratios and relative risk

Although a χ^2 test may tell you that two variables are associated, it won't tell you by how much. One way of doing this is by using the **odds ratio.** Here's some 1998 data in which 20,000 male physicians were given daily aspirin or placebo for some time, and the incidence of heart attacks monitored.

	Heart attack	No heart attack	*Total*
Aspirin	104 (*a*)	10,933 (*b*)	*11,037*
Placebo	189 (*c*)	10,845 (*d*)	*11,034*
Total	*293*	*21,778*	*22,071*

The probability of someone in the aspirin group having a heart attack was $^a/_{a+b}$ = 0.94%. The probability of someone in the placebo group having a heart attack was $^c/_{c+d}$ = 1.7%. The **probability ratio** or **relative risk** is therefore $^a/_{a+b} \div {}^c/_{c+d}$ = 0.55 (or, taking the reciprocal of this, 1.82). The **odds** of someone in the aspirin group having a heart attack were $^a/_b$ = 0.0095 (see Handout 1 for definition of odds). The odds of someone in the placebo group having a heart attack were $^c/_d$ = 0.0174. The **odds ratio** is $^a/_b \div {}^c/_d = {}^{ad}/_{bc}$ = 0.54 (and its reciprocal, $^{bc}/_{ad}$, is 1.83). So these men were about half as likely to have a heart attack if they were on aspirin.

Probability versus odds: be careful

Applying this technique to the rape jury data above might lead you to the conclusion that the jury were five times as likely to acquit if the defendant was portrayed as being at fault. The probability of conviction in the low-fault condition was 0.86, equivalent to odds of 6.40. The probability of conviction in the high-fault condition was 0.58, equivalent to odds of 1.38. The **odds ratio** is therefore 4.6 (or 0.22 depending on which way round you view it). However, the probability ratio (**relative risk**) is only 1.49 (or 0.67) and the **absolute risk** increased by 0.86 – 0.58 = 0.28. Were the jury 4.6 times as likely to convict if the defendant was portrayed as being at fault, or 1.5 times? This depends on what you mean by 'as likely'! Remember that **probability = odds / (1+odds).** The odds on them acquitting were increased 4.6 times; the probability was increased 1.5 times.

To get a feeling for these counter-intuitive numbers, consider a couple of examples. Take a 100-kg sack of potatoes that are 99% water. If you dried out the potatoes completely, they'd have a mass of 1 kg. What would their mass be if you dried them out partially, until they were 98% water? The answer is 50 kg. So consider a group of patients that has a 99% chance of dying from the disease. If you give them a drug that reduces their probability of dying to 98% (so relative risk of dying: 0.98 / 0.99 = .9899), you have halved their odds of dying from 100:1 to 50:1 (odds ratio 0.5). But beware another property of relative risk: it matters which way round you view things. The patients' chance of *survival* has increased from 1% to 2% (relative risk of surviving: 0.02 / 0.01 = 2, which is nothing like the reciprocal of the relative risk of dying) but their odds of survival have increased from 1:100 to 1:50 (odds ratio 2, which is exactly the reciprocal of the odds ratio of dying).

Be careful not to be misled by papers that report odds ratios. If the overall event rate is low, odds ratios and relative risk are very similar; if high, they can be very different. The mathematical properties of odds ratios encourage their use (you can't double a probability of 0.6, for example), and they can be used in studies where you do not know the absolute probabilities (risks) of something happening (e.g. clinical case–control studies). However, they don't reflect our intuitive view of probability very well. Perhaps the clearest way to report these things is to give absolute probabilities, if you can, and then readers can work out all the other measures.

5.3 Supplementary material: the binomial distribution

Where does the χ^2 test come from? Read on if you're interested…

Imagine you have a coin that you flip a number of times. Each time, there are only two possible outcomes (heads or tails). If it's a fair coin, the probability of a head on each trial, call it *p*, is 0.5. Let's call the probability of a tail *q*; also 0.5. If you flip the coin five times, what is the probability that you get five heads? There's only one way to do this — HHHHH. So the probability is $0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 = (0.5)^5 = p^5 = 0.03125$. Similarly, the probability of zero heads, i.e. five tails (TTTTT), is $q^5 = 0.03125$ as well. But if you flip the coin five times, what's the probability that you get *three* heads? This is trickier, because there are several ways to do it. You might throw HHHTT, or HHTTH, or TTHHH… the probability of each pattern is $(0.5)^5$, but we'd like an easy way to work out the number of ways of getting three heads.

Permutations and combinations

We might as well make this general. If we have *n* lottery balls in a lottery ball machine, and we draw out *r* of them in a particular order, the number of ways we could draw them is called the number of **permutations,** written P_r^n. For example, if there are 50 balls in the National Lottery and we draw out 6 of them, then one permutation is {1,2,3,4,5,6}; another is {6,5,4,3,2,1}; another is {17;42;22;5;38;9}. Since we don't care about the order of the balls in the lottery, we can also talk about the number of **combinations** of drawing *r* balls out of *n* balls, or C_r^n — combinations are the same as permutations except that they don't care about the *order*, so {1,2,3,4,5,6} and {6,5,4,3,2,1} count as two separate permutations but are just two ways of writing the same combination. We can calculate P_r^n and C_r^n very simply once we know what **factorial** means: 6 factorial, written **6!,** is $6 \times 5 \times 4 \times 3 \times 2 \times 1$. So, written mathematically, here's what we need to know:

$$x! = x(x-1)(x-2)\ldots1 \qquad \text{Note that } 0! = 1 \text{ (a special case)}$$

$$P_r^n = \frac{n!}{(n-r)!}\text{; } P_r^n \text{ is sometimes written } {}_nP_r$$

$$C_r^n = \frac{n!}{r!(n-r)!}\ ;\ C_r^n \text{ is sometimes written } {}_nC_r \text{ or } \binom{n}{r}.$$

We can use this to find out that there are $C_6^{50} = \frac{50!}{6!(50-6)!} = \frac{50!}{6!\times 44!} = 15{,}890{,}700$ possible outcomes in the National Lottery. But we can also use it to find out that there are $C_3^5 = \frac{5!}{3!\times 2!} = 10$ ways of flipping three heads in five coin flips.

The binomial distribution

Since we know that the probability of any particular sequence of five coin flips is $(0.5)^5 = 0.03125$, we now know that the probability of flipping three heads is $10 \times (0.5)^5 = 0.31$. In general, if we have ***n* independent trials,** each of which has **two outcomes,** one of which we'll call 'success' and one of which we'll call 'failure', where the probability of success is ***p*** and the probability of failure is ***q* = 1 – *p*,** and ***X*** is a discrete random variable representing the number of successes, then the probability of ***r* successes,** written $P(X = r)$, is given by the **binomial distribution:**

$$P(X = r) = C_r^n p^r q^{n-r}$$

We would call this distribution ***B*(*n*, *p*).** We can calculate the mean (the expected value) and the variance of $B(n, p)$:

$$E[B(n, p)] = np$$

$$Var[B(n, p)] = npq$$

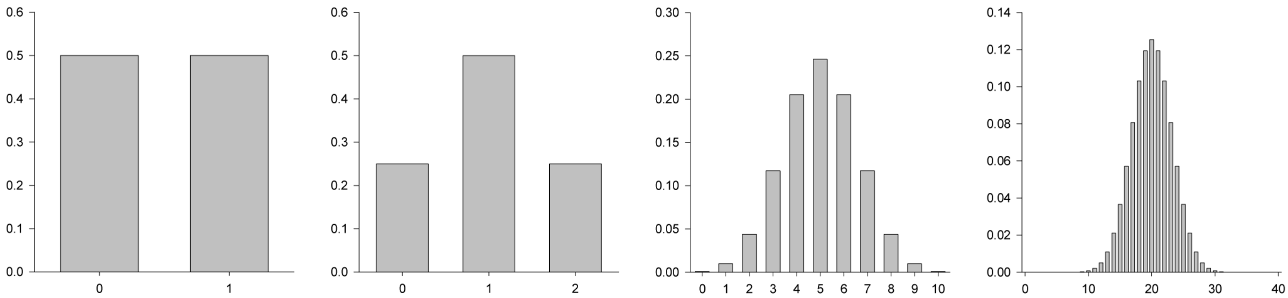
In other words, the mean number of heads in five coin flips is $5 \times 0.5 = 2.5$, and the variance of this is $5 \times 0.5 \times 0.5 = 1.25$ (so the standard deviation is $\sqrt{1.25} = 1.12$).

Using the binomial distribution as a statistical test

If a gambler inveigles us into a betting game, flips a coin 100 times and obtains 90 heads, is the coin fair? The null hypothesis is that the coin is fair ($p = q = 0.5$), and the observed number of heads was observed by chance. If the null hypothesis is true, then the number of heads in 100 flips should obey the binomial distribution $B(100, 0.5)$. The probability of obtaining 90 heads is therefore $P(X = 90) = C_{90}^{100} 0.5^{90} 0.5^{10}$. But we're actually interested in the probability of obtaining *90 or more* heads. We therefore want to know $P(X \geq 90) = P(X = 90) + P(X = 91) + P(X = 92) + \ldots + P(X = 100)$; a bit of calculation gives the answer $P(X \geq 90) = 1.53 \times 10^{-17}$. This is considerably less than our conventional α of 0.025 (we'd be using a two-tailed test here, since we'd want to detect a bias in either direction, so $\alpha = 0.025$ for each tail); we would therefore reject the null hypothesis and accuse the gambler of fraud. The clever fraudster would do better to use a very slightly biased coin: if he flipped 60 heads, $P(X \geq 60) = 0.028$, so a two-tailed test with overall $\alpha = 0.05$ wouldn't reject the null hypothesis of a fair coin. We'd need to observe the slightly biased coin for longer (more trials) to be able to detect it. This is a general principle of statistics: **more observations help you detect smaller effects.**

The normal distribution as an approximation to the binomial distribution

For large sample sizes (e.g. $np > 5$ and $nq > 5$), the binomial distribution $B(n, p)$ ap-



Probability (y axis) of all possible total numbers of heads observed (x axis) when you flip a coin 1, 2, 10, or 40 times (from left to right). The binomial distribution approximates a normal distribution as n increases.

proximates the normal distribution $N(np, npq)$ — that is, a normal distribution with mean np and variance npq (see figure).

5.4 Supplementary material: the sign test

The sign test (sometimes called the Fisher sign test) evolves from the binomial test and is very simple indeed. Using an example borrowed from Howell (1997, p. 127), suppose we want to test whether people that know each other are more tolerant of individual differences. We might ask a dozen male first-year students to rate the physical attractiveness of a dozen other first-years (of the same sex) at the start and the end of the year. Suppose the median ratings (high = attractive) are as follows:

Target	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Start	12	21	10	8	14	18	25	7	16	13	20	15
End	15	22	16	14	17	16	24	8	19	14	28	18
Gain	3	1	6	6	3	–2	–1	1	3	1	8	3

The sign test looks at the sign (direction), but not the magnitude (size) of each difference. The null hypothesis is that there is no change in rating. Ignoring gains of 0 (which we don't have here anyway), the null hypothesis would therefore predict that by chance, about half the ratings would improve and about half would worsen, i.e. $p(\text{higher}) = p(\text{lower}) = 0.5$. In our hypothetical data set, we have 10 improvements out of 12 targets. We want to calculate $P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12)$. Using the binomial distribution $B(12, 0.5)$, we know that $P(X = 10) = C_{10}^{12} 0.5^{10} 0.5^{2}$, and so on; the total $P(X \geq 10)$ is 0.0192. As this is less than our traditional $\alpha = 0.05$, we would reject the null hypothesis and say that there was a significant change in rating over the year.

The sign test using the normal approximation to the binomial distribution

For the null hypothesis, $p(\text{positive sign}) = p(\text{negative sign}) = 0.5$. So if the number of non-zero difference scores $n > 10$, and x is the number of difference scores of one sign (e.g. positive), we can use the normal approximation to the binomial distribution to get a quick answer. The mean of this distribution is $np = n/2$, and the variance is $npq = n/4$. So we can calculate a ***Z* score:**

$$z = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2\left(x - \frac{n}{2}\right)}{\sqrt{n}}$$

and test that Z score in the usual way (see Handout 1).

Comparing the sign test to the Wilcoxon matched-pairs signed-rank test

From our discussion of the Wilcoxon matched-pairs signed-rank test in Handout 4, you'll see that the sign test is pretty similar in overall logic — except that the sign test throws away *even more* information about the distribution (it doesn't care about the magnitudes of the difference scores at all, just their signs). You pay a price in power, but gain generality; the sign test is a nonparametric test that can be used with ordinal or even categorical data.

5.5 Supplementary material: the multinomial distribution

If we want to consider more than two alternatives for each trial, we need to use the **multinomial distribution.** Let there be n trials and k alternatives for each trial, numbered from 1 to k, each with the probabilities p_1, p_2, … p_k. Then the probability of obtaining exactly X_1 outcomes of event$_1$, X_2 outcomes of event$_2$, … and X_k outcomes of event$_k$ is given by

$$p(X_1, X_2, \ldots, X_k) = \frac{n!}{X_1! X_2! \cdots X_k!} p_1^{X_1} p_2^{X_2} \cdots p_k^{X_k}$$

An example: if we had a die with two black sides, three red sides, and one white side, then for each trial $p(\text{black}) = {}^2/_6$, $p(\text{red}) = {}^3/_6$, and $p(\text{white}) = {}^1/_6$. So if we roll the die 10 times, then the probability of obtaining exactly 4 blacks, 5 reds, and 1 white is

$$p(4,5,1) = \frac{10!}{4! \times 5! \times 1!}\left(\frac{2}{6}\right)^4\left(\frac{3}{6}\right)^5\left(\frac{1}{6}\right)^1 = 0.081$$

5.6 Supplementary material: the χ^2 distribution; an outline of deriving the χ^2 test

The χ^2 distribution

The χ^2 probability density functions are shown in the figure below; you can see that the shape of the distribution depends on the number of degrees of freedom, k. It is a *positively skewed* distribution, especially when k is small. The distribution is often written as χ^2_{df}, or sometimes $\chi^2(df)$. To obtain critical values of χ^2, we need to know the value of χ^2 above which (say) 5% of the area falls. In practice, we'll get this from tables or a computer.

Relationship between χ^2 and the normal distribution

If we have a normal random variable $N(\mu, \sigma^2)$, we can sample one value x from it, convert it to a standard normal variable z, and square it:

$$z^2 = \frac{(x-\mu)^2}{\sigma^2}$$

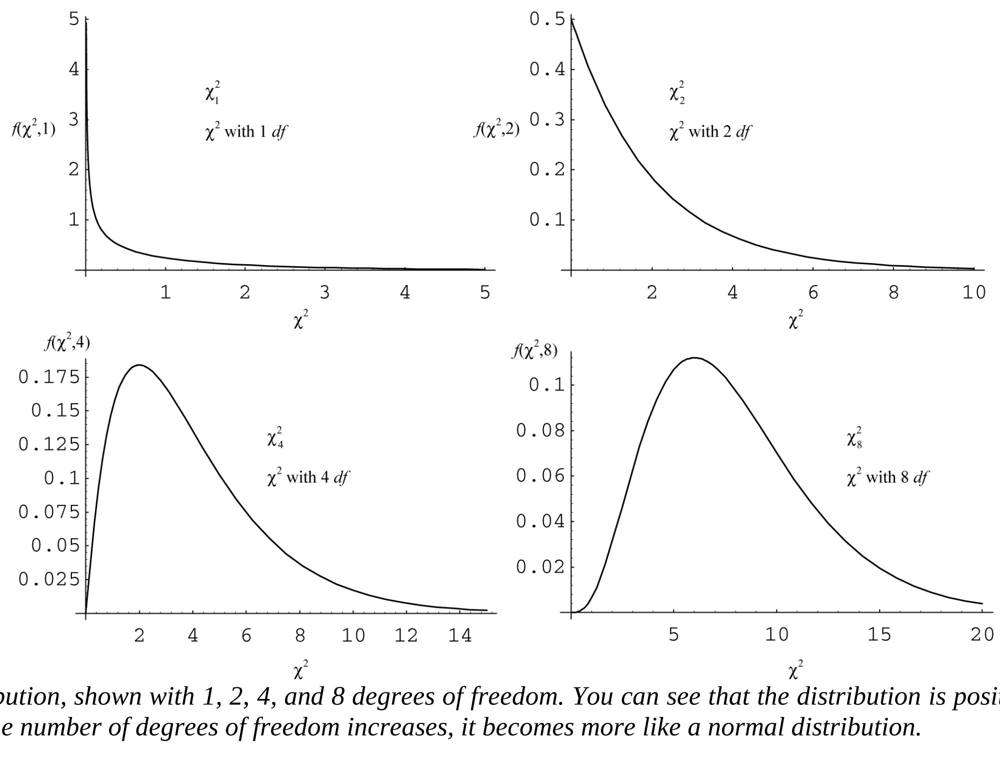
If we repeated this *ad infinitum*, sampling independently each time, we would have a great number of values of z^2. We could therefore plot the distribution of z^2. We would find that this distribution is the same as χ^2_1 (χ^2 with 1 *df*):

$$\chi^2_1 = z^2$$

Now suppose that instead of sampling one number at a time, we sample n numbers at a time. For each observation within each sample we calculate z^2; for each sample, we calculate Σz^2. So each sample produces one value of Σz^2. Now we plot the distribution of these values of Σz^2. We find that the distribution is the same as χ^2_n:

$$\chi^2_n = \sum_{i=1}^{n} z_i^2 = \Sigma \frac{(X_i - \mu)^2}{\sigma^2}$$

In other words, then if Y is the sum of squares of n independent standard normal



The χ^2 distribution, shown with 1, 2, 4, and 8 degrees of freedom. You can see that the distribution is positively skewed, but that as the number of degrees of freedom increases, it becomes more like a normal distribution.

variables, then Y is distributed as χ^2 with n degrees of freedom. (Since z_i^2 has a χ^2 distribution, this result also shows that the sum of a set of independent values of χ^2 itself has a χ^2 distribution, given the restrictions of independent sampling and an underlying population with a normal distribution.)

χ^2 tells us something about the distribution of sample variances

If we have a normal random variable $N(\mu, \sigma^2)$, we can draw an infinite number of samples from it. From each sample, we can calculate the sample variance s^2. We could then plot the distribution of these sample variances. We would find that it is related to the χ^2 distribution:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2} \text{ and therefore } s^2 = \frac{\chi^2\sigma^2}{n-1}$$

Since $\sigma^2/(n–1)$ is constant for a given σ^2 and sample size (n), the sampling distribution of the variance (the distribution of a set of sample variances) has a χ^2_{n-1} distribution. Since the χ^2 distribution is skewed, this tells us that the distribution of s^2 is too — although the average value of a lot of s^2 measurements will equal σ^2, more than half the time s^2 will be less than σ^2.

Deriving the χ^2 test from the binomial distribution (via the normal distribution)

Suppose we ask 100 people to choose between chocolate and garibaldi biscuits. Let's say that 65 choose chocolate and 35 choose garibaldi. Does this differ from chance, i.e. a 50:50 split? We could answer this with the binomial distribution, $B(100, 0.5)$, but there'd be a lot of adding up to find $P(X \geq 65)$. So let's do it a different way. **For large sample sizes** (e.g. $np > 5$ and $nq > 5$), **the binomial distribution $B(n, p)$ approximates the normal distribution $N(np, npq)$.** We've seen that

$$\chi^2_1 = z^2 = \frac{(x-\mu)^2}{\sigma^2}$$

where x is sampled from a normal distribution $N(\mu, \sigma^2)$. As we know the mean of a binomial distribution is np and the variance (σ^2) is npq, we can derive this approximation:

$$\chi^2_1 = z^2 = \frac{(x-np)^2}{npq}$$

To make things easier for later, we'll call the **observed frequencies** O_1 and O_2, and the **expected frequencies** E_1 and E_2. Specifically, $E_1 = np$ and $E_2 = nq$, and $O_1 + O_2 = E_1 + E_2 = n$. In our biscuit example, $O_1 = 65$, $O_2 = 35$, $E_1 = 50$, and $E_2 = 50$. Expanding and substituting these in to the previous formula, we would eventually get

$$\chi^2_1 = \frac{(x-np)^2}{np} + \frac{(n-x-nq)^2}{nq}$$

$$\chi^2_1 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2}$$

or, more generally,

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

which is the general formula for χ^2 that we've been using. This formula also extends to more than two categories, using the multinomial distribution.

Relevant functions in Excel (see Excel help for full details)

BINOMDIST()	Gives you the binomial p.d.f., $P(X = x)$, or c.d.f., $P(X \leq x)$, where X has a binomial distribution.
CHIDIST()	From χ^2 and the df, gives you the probability that $P(X > \chi^2)$, where X has a χ^2 distribution.
CHIINV()	From p and the df, gives you the critical value of χ^2 such that $P(X > \chi^2) = p$.
CHITEST()	Does a χ^2 test for you, working out the df automatically and returning the p value.

Bibliography

Howell, D. C. (1997). *Statistical Methods for Psychology*. Fourth edition, Wadsworth, Belmont, California.