Chi-Square Test

Make sure you read about degrees of freedom and probability values in the description of Student's T-Test.

The T-Test is used for continuous variables, for things like numbers and percents. Chi-Squared test is used for categorical variables: colors, flavors, presence-absence. A Chi-Squared test will compare your actual values (**Observed**) to values based on your values (**Expected**). The easiest way to think of a Chi-Square Test is to put your values into a table.

Say you want to compare types of snacks eaten by kindergarteners compared to preschoolers, when they are allowed to choose one snack from three types: cookies, crackers, and apple slices. Then your hypotheses are:

Hypothesis: Kindergarten kids select different snacks than preschool kids.
Null Hypothesis: There are no differences in the types of snacks selected.

Count the number of kids that eat each type of snack. Then you can arrange your data like this:

	Cookies	Crackers	Apple Slices
Preschoolers	67	11	23
Kindergarteners	83	5	19

There are two important assumptions for Chi-Square: **sample size** and **independence**. Independence means the data are independent. If you are comparing the food selected by kids, you should not compare the food a kid selects today to the selections that kid makes tomorrow. Kids often have favorite snacks. So, information from the same individual over time is considered a repeated measure, not an independent measure.

The other problem with Chi-Square is sample size. Already, from the snack example above, we know we have very few kids eating crackers. As we start to calculate **Expected** frequencies, we may get values smaller than 5. **If we have any expected values <5, then: 1) we cannot use Chi-Square or 2) we must combine crackers with another snack (probably apple slices).**

The table above shows the actual number of kids that chose each type of snack. It's easy to see that more kids picked cookies. What's hard to see is whether kindergarteners picked cookies more often than preschoolers.

The next step is to calculate totals for your **Observed** values:

	Cookies	Crackers	Apple Slices	Row Totals
Preschoolers	67	11	23	101
Kindergarteners	83	5	19	107
Column Totals	150	16	42	208

Now you have to calculate **Expected** values:

Expected = (row total x column total)/(grand total)

	Cookies	Crackers	Apple Slices	Row Totals
Preschoolers	(101x150)/208=72.8	(101x16)/208=7.8	(101x42)/208=20.4	101
Kindergarteners	(107x150)/208=77.2	(107x16)/208=8.2	(107x54)/208=27.8	107
Column Totals	150	16	42	208

	Observed Values:			**Expected** Values:		
	Cookies	Crackers	Apple Slices	Cookies	Crackers	Apple Slices
Preschoolers	67	11	23	72.8	7.8	20.4
Kindergarteners	83	5	19	77.2	8.2	27.8

We are in luck! All expected values are ≥5. This means our sample size is big enough. It means we can keep all three types of snacks. We don't have to combine any snacks to meet the requirements of Chi-Square. Now, we get to do the test and calculate chi-square:

First do **(Observed – Expected)** or **(Expected – Observed)**. Just avoid negative numbers.

	Cookies	Crackers	Apple Slices
Preschoolers	72.8-67=5.8	11-7.8=3.2	23-20.4=2.6
Kindergarteners	83-77.2=5.8	8.2-5=3.2	27.8-19=8.8

Now, subtract 0.5 from each cell, but only if you have 2 columns and 2 rows. This square has 2 rows and 3 columns, so do not subtract 0.5:

	Cookies	Crackers	Apple Slices
Preschoolers	5.8~~-0.5=5.3~~	3.2~~-0.5=2.7~~	2.6~~-0.5=2.1~~
Kindergarteners	5.8~~-0.5=5.3~~	3.2~~-0.5=2.7~~	8.8~~-0.5=8.3~~

Now square each cell:

	Cookies	Crackers	Apple Slices
Preschoolers	5.8^2=33.64	3.2^2=10.24	2.6^2=6.76
Kindergarteners	5.8^2=33.64	3.2^2=10.24	8.8^2=77.44

Now divide each cell by that cell's **Expected** value:

	Cookies	Crackers	Apple Slices
Preschoolers	33.64/72.8=0.46	10.24/7.8=1.21	6.76/20.4=0.33
Kindergarteners	33.64/77.2=0.44	10.24/8.2=1.25	77.44/27.8=2.78

Add it all up to get your chi-square value: **0.46+0.44+1.21+1.25+0.33+2.78=6.47=Chi-square**

Now, what is your **probability value** and what are your **degrees of freedom**?

In the **Student's T-Test**, we discussed how Biologists often use **probability values** of 0.05 and what that means. In medicine, **probability values** are often 0.0001. What does that mean?

For Chi-square, the way to calculate **degrees of freedom (df)** is simple:

Degrees of freedom = (number of rows – 1) x (number of columns – 1)
For us, it is: (2 rows –1) x (3 columns –1) = 2

Now, we can find out if there are any differences in snack selection between kindergarteners and preschoolers. Look on this table of Chi-squared statistics or find one on the internet:

P-value	df=1	df=2	df=3	df=4
0.1	2.706	4.605	6.251	7.779
0.05	3.841	5.991	7.815	9.488
0.01	6.635	9.210	11.345	13.277
0.005	7.879	10.597	12.838	14.86
0.001	10.83	13.82	16.27	18.47

Use your degrees of freedom to find out if your chi-square value is bigger than the table's value. If it is, then there are significant differences between kindergarteners and preschoolers. For us, the chi-square is 6.47. Our probability value is 0.05. Our degrees of freedom is 2. The table's chi-square value is 5.99. If our value is smaller than the table's value, we cannot reject our null hypothesis. In this case, our value is larger than the table's value, so we can reject our null hypothesis!

Preschoolers and kindergarteners all like cookies, but preschoolers also like apple slices (chi-square=6.47, df=2, $p<0.05$).