Lecture 20 (Ch 8)

When we compare 2 props \((\pi_1, \pi_2)\) (e.g. \(H_0: \pi_1 - \pi_2 < 0.1\)) we have two populations, each with 2 groups/cats. (e.g. Boys/Girls)

In that case, one proportion (e.g. prop of boys, \(\pi_{\text{boys}}\)) is enough to describe each pop., because the other prop. (e.g. \(\pi_{\text{girls}}\)) is fixed by \(1 - \pi_{\text{boys}}\). In other words, our 2-sample CIs/tests involve two proportions, one from each of two populations.

So, an example would be \(\begin{cases} \pi_1 = \pi_{\text{boys}} \text{ in Northern hemisphere.} \\ \pi_2 = \pi_{\text{boys}} \text{ in Southern hemisphere.} \end{cases}\)

Note that both \(\pi_1\) and \(\pi_2\) refer to boys, but in two different populations (e.g. Northern and Southern hemispheres).

Then, we can test, e.g. \(H_0: \pi_1 - \pi_2 = 0\) vs. \(H_1: \pi_1 - \pi_2 \neq 0\).

But there are situations where we have one population, with more than 2 categories, and we want to test some claim about the proportions of each category.

If we have one pop., with \(k\) categories, we can test \(H_0: \pi_1 = \pi_{\text{category 1}}, \pi_2 = \pi_{\text{category 2}}, \ldots, \pi_k = \pi_{\text{category k}}, \text{ prop. of } k^{th} \text{ cate. in pop.}\)

\(H_1: \text{At least one of } \pi_i \text{ is wrong}\) \(\sum_{i=1}^{k} \pi_i = 1\)

I'll explain this later.

Of course, given that there is only one pop., we must have \(\pi_1 + \pi_2 + \cdots + \pi_k = 1\).

Below, we will see how to do this test.

There will be a new distribution: Chi-squared.

Also note that a pop. with 2 groups can be thought of as being described by one random variable with 2 levels. Similarly, a pop. with \(k\) groups can be described with one r.v. with \(k\) levels.
Does data provide sufficient evidence to support an association between climate and tornado activity?

<table>
<thead>
<tr>
<th>El Nino</th>
<th>La Nina</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_1 = 14 )</td>
<td>( \eta_2 = 28 )</td>
<td>( \eta_3 = 44 )</td>
</tr>
</tbody>
</table>

Number of years classified as:

- \( \eta_1 = 12 \)
- \( \eta_2 = 17 \)
- \( \eta_3 = 25 \)

Proportion:

\[
\frac{\eta_1}{\eta_3} = \frac{14}{86} = 0.16 \\
\frac{\eta_2}{\eta_3} = \frac{28}{86} = 0.33 \\
\frac{\eta_3}{\eta_3} = \frac{44}{86} = 0.51
\]

Hypotheses:

- \( H_0 \): There is no association, i.e., true prop of tornado days in El Nino years. Etc.
- \( H_1 \): At least one of these assignments is wrong.

If \( H_0 \) = True, how many tornadoes do you expect in each of the \( k^3 \) categories?

\[
\begin{align*}
\text{Expected counts:} & \quad 0.22(86) \quad 0.32(86) \quad 0.46(86) \\
\text{Observed counts:} & \quad 14 \quad 28 \quad 44
\end{align*}
\]

\[
\begin{align*}
(\text{Exp.}-\text{obs})^2 : & \quad (4.9)^2 \quad (-0.5)^2 \quad (-4.4)^2 \\
(\text{Exp.}-\text{obs})^2 \quad \text{Exp} : & \quad 1.27 \quad 0.009 \quad 0.49
\end{align*}
\]

\[
\chi^2 \quad \text{obs.} = \sum_{i=1}^{3} \frac{(\text{exp.}-\text{obs})^2}{\text{exp.}} = 1.77
\]
If there were really no difference at all in the # of tornadoes between the 3 categories, then this would be near zero.

So, is this $X_{obs}^2$ far away from 0 to reject $H_0$ (in favor of $H_1$)?

Note: $X^2$ is non-negative, unlike $z$, $t$

We need to know the sample dist. of $X^2$, when $H_0 = T$.

Theorem: Under the null hypothesis, $X^2$ has a chi-squared dist. with df = $k - 1$ ($= 3 - 1 = 2$)

What's a chi-squared dist.? It's just another Table (VII). But FYI.

$p$-value = $\text{prob}(X^2 > X_{obs}^2) = \text{prob}(X^2 > 1.77) > 0.1$

$s^2$ df = $3 - 1 = 2$

R: $\text{pchisq}(1.77, \text{df} = 3 - 1) = 0.41$

For the chi-squared test, p-value is always right area.

Conclusion (at $\alpha = .01$): $p$-value > $\alpha$

\(\Rightarrow\) at least 1 is wrong.

In words: Cannot reject $H_0$ in favor of $H_1$.

\(\Rightarrow\) ($\pi_1 = .22, \pi_2 = .32, \pi_3 = .46$)

In English: There is no evidence from data to suggest that the 3 props are not .22, .32, .46, i.e.

I.e. There is no evidence from data that there is an association between tornado activity and climate.

The chi-squared density function is

\[ f(x) = \frac{X^2 - 1}{e} \left( \frac{X^2}{2} \right)^{df - 1} \frac{e}{\Gamma(\frac{df}{2})} \frac{1}{2^{df/2}} \]
Here is a generalization of the above to any \( k \) (above \( k=3 \)). I propose that you do not use the formula at the bottom, but instead do things like I did above. After, you are completely comfortable with the steps, then you can use this page.

Let \( \pi_i \) = proportion of cases in category \( i \) (where \( i = 1, 2, \ldots, k \)).

\[
\begin{array}{c|c|c}
\hline
\text{Null params} & \text{Example} \\
\hline
\pi_1 & \pi_{o1} & 0.22 \\
\pi_2 & \pi_{o2} & 0.32 \\
\pi_3 & \pi_{o3} & 0.46 \\
\hline
\end{array}
\]

If \( H_0 = \text{True} \), then \( \pi_i = \pi_{o1}, \pi_2 = \pi_{o2}, \ldots \)

Thus in a sample of size \( n \), how many would we expect in category 1: \( n \pi_{o1} = 18.9 \),
2: \( n \pi_{o2} = 27.5 \),
3: \( n \pi_{o3} = 39.6 \).

But according to data, we observe this many:
\[
\begin{align*}
\sum_{i=1}^{3} n_i &= n \\
n_1 &= 14 \\
n_2 &= 28 \\
n_3 &= 44
\end{align*}
\]

**Punch line:**
Thus the theorem tells us that counts, not proportions!

\[
\chi^2_{\text{obs}} = \sum_{i=1}^{k} \frac{(n\pi_i - n_i)^2}{n\pi_i}
\]

has a chi-sq. distr with \( df = k-1 \).

\( p \)-value \( = \text{prob}( \chi^2 > \chi^2_{\text{obs}} ) \)

In fact, even though we wrote a lot of things in terms of props, the data are all in counts. This chi-sq test models count data!
The chi-squared test shows up in many places. Here is one more:

Independence of 2 categorical variables, \( X \) and \( Y \), one with \( k \) levels, the other with \( r \) levels.

- **Ho:** 2 categorical vars. are indep.
- **H1:** \( \cdots \) not indep.

Analog of scatter plot for cat var.

In such problems, the data are shown as a **Contingency Table**:

![Contingency Table](image)

The test of \( Ho \) (i.e., independence) turns out to be a chi-squared test, but with \( df = (k-1)(r-1) \).

Recall: counts, not p-value

\[ \chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \]

I.e. compute

\[ X^2_{\text{obs}} = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \]

and p-value = \( P(X^2 > X^2_{\text{obs}}) \)

The only question is what are the expected counts (under Ho)?

It can be shown (see page 19, below):

**Expected counts**

Assuming Ho = True

\[
\begin{pmatrix}
\frac{(a+b+c)(a+b)}{n} & \frac{(a+b+c)(b+c)}{n} & \cdots \\
\frac{(a+b+c)(a+c)}{n} & \frac{(a+b+c)(c+d)}{n} & \cdots \\
\end{pmatrix}
\]

Remember, this result a "row x col. marginals".
Derivation of “row x col marginal” method for computing The C-table of expected counts. For simplicity, let v=2, k=2. Start with The observed counts:

\[
\begin{array}{c|cc}
\text{category} & 1 & 2 \\
\hline
\text{population A} & a & b \\
\text{population B} & c & d \\
\hline
\text{a+b} & \text{c+d} & \text{a+c, b+d, \overline{1/n}}
\end{array}
\]

Even though we talk about population A, B, ... The counts a,b,c,d, are The observed Sample (data).

Switch to proportions:

\[
\begin{pmatrix}
\frac{1}{n} \\
\frac{2}{n}
\end{pmatrix}
\]

\[\bar{P}_A = \text{true prop. of catg. 1 in pop. A} \]

Etc.

If \(H_0=\text{true}\), i.e. if X and Y are indp., Then \(\overline{P}_A = \overline{P}_B\), and \(\overline{P}_A = \overline{P}_B\).

But if \(\bar{P}_A = \bar{P}_B\), Then it must be That \(\overline{P}_A = \overline{P}_B = \overline{P}_1\),
where \(\overline{P}_1\) = prop. of category 1, period!

So, if \(H_0 = \text{true}\), how many do we expect to get in pop.A?

\[\text{Answer: } \frac{(a+b) \, \overline{P}_1}{\overline{1/n}}\]

\# in pop.A

We can estimate \(\overline{P}_1\) with The sample prop. of 1's : \(\frac{a+c}{n}\).

\[\therefore \text{Answer: } \frac{(a+b) \, (a+c)}{n}\]

And that is The 1st element of The expected C-Table. Etc.
Example: Does the email filter actually work? I.e., is the variable “True State of an email” (safe/unseen) independent of the variable “predicted/classified State” (safe/unseen)? (This data is actually a “famous” data set on whether Belief in afterlife is independent of Gender! I’ve updated the variables.)

Here are the data in the form of a contingency table:

<table>
<thead>
<tr>
<th></th>
<th>Safe</th>
<th>Uns1afe</th>
<th>Unclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truly Safe</td>
<td>435</td>
<td>58</td>
<td>89</td>
</tr>
<tr>
<td>Truly Unsafe</td>
<td>375</td>
<td>50</td>
<td>84</td>
</tr>
<tr>
<td>Unclassified</td>
<td>810</td>
<td>108</td>
<td>173</td>
</tr>
</tbody>
</table>

\[
\text{Expected Counts: } \begin{pmatrix} \frac{582 \times 509}{1091} & \cdots & \frac{582 \times 1091}{509} \\ \cdots & \cdots & \cdots \\ \frac{509 \times 1091}{582} & \cdots & \frac{509 \times 582}{1091} \end{pmatrix} = \begin{pmatrix} 432.1 & 57.6 & 92.3 \\ 377.9 & 50.4 & 80.7 \end{pmatrix}
\]

\[
\chi^2 = \frac{(435 - 432.1)^2}{432.1} + \frac{(58 - 57.6)^2}{57.6} + \cdots + \frac{(92.3 - 92.3)^2}{92.3}
\]

\[
= 0.019 + 0.0028 + 0.118 + 0.022 + 0.0035 + 0.135
\]

\[
= 0.3
\]

Table VII: \( \chi^2 \text{ p-value} = 0.86 \)

\[
df = (2-1)(3-1) = 2 \Rightarrow \text{p-value} > 0.1 \text{ (huge)}
\]

\[\text{Cannot reject } H_0 \text{ in favor of } H_1 \text{ at } \alpha = 0.01 \text{ (or } 0.05)\]

I.e., there is no evidence the 2 variables are not independent.

I.e., there is no evidence that the email filter actually works! (I.e., it may not be working.)
Based on this data, we cannot say that the filter works! Mathematically, the reason is that $X_{obs}^2$ was too small.

But suppose, $X_{obs}^2$ had turned out to be huge. Then we could conclude that the filter is working, i.e. that there is some kind of difference between truly safe and unsafe emails - a difference that the classifier has identified. Then, we can look at the relative size of the various terms in $X_{obs}^2$ to see which ones make $X_{obs}^2$ large.

In this example, the big terms are 0.118, 0.135, which correspond to the "predicted/classified as unsafe" category.

In short, if the $X_{obs}^2$ had turned out to be huge (i.e. p-value = small, i.e. statistically significant), then we would conclude that 1) filter works, and additionally 2) whatever the difference between truly safe and unsafe emails is, that difference is biggest when the filter classifies things as unsafe.

The signs of the terms in $X_{obs}^2$ can also be interpreted, but we'll skip that, too.
how to use Table VII:

Table VII gives the area to the right of some value of $x^2_{\text{obs}}$, i.e., it gives a p-value. However, it does not give all p-values; the only ones it provides are listed in the left-most column. E.g.,

$X^2_{\text{obs}} = 8.49$, df = 4 $\Rightarrow$ p-value = 0.075

$X^2_{\text{obs}} = 8.66$, df = 4 $\Rightarrow$ p-value = 0.070

One might think that putting bounds on p-value is not enough for hypothesis testing, but it often is.

For example, suppose we get $X^2_{\text{obs}} = 8.55$ with df = 4. Then we can say $0.070 < \text{p-value} < 0.075$. That is good enough if $\alpha = 0.05$, because p-value > $\alpha$, and so we cannot reject $H_0$ in favor of $H_1$. 
Additional comments:

Note that the first $H_0$, $H_1$ above are just a generalization of:

$H_0: \pi = \pi_0$ (z-test).

$H_1: \pi \neq \pi_0$

to more than 2 categories in the population.

However, there are no 1-sided/2-sided varieties of chi-squared.

When $X^2_{obs}$ is small (say no), then the observed counts are consistent with the expected counts if $H_0$ is true (i.e. $\pi_1 = \pi_{01}$, $\pi_2 = \pi_{02}$, ..., $\pi_k = \pi_{0k}$). So, if $X^2_{obs}$ is large, then at least one of the $\pi_{0i}$ must be wrong.

In other words, the appropriate hypotheses are:

$H_0: \pi_1 = \pi_{01}, \pi_2 = \pi_{02}, ..., \pi_k = \pi_{0k}$

$H_1: \text{At least one of these specifications is wrong.}$

And it is the "At least" which gives us

$p\text{-value} = \text{prob}(X^2 \geq X^2_{obs})$ (Table VII)

i.e. We are always interested in the upper tail area only.

Said differently, for the chi-squared test of the above $H_0/H_1$, the $p\text{-value}$ is only the right area, because violation of each part of $H_0$, increases $X^2$. 
Summary

In Ch. 7, we learned how to build CIs for either a prop, \( \pi_x \), or the difference between 2 props, \( \pi_1 - \pi_2 \), where \( \pi_i \) = prop of something (e.g., boys) in population \( i \), and \( \pi_2 = \) "Same Thing."

In Ch. 8, we did hyp. tests on \( \pi_x \) and \( \pi_1 - \pi_2 \) (note: \( \pi_1 + \pi_2 \neq 1 \)).

But in all of these situations, the 2 pops have 2 categories (boy/girl), and \( \pi_i \) is the prop. of 1 of the categories in the \( i^{th} \) population.

The Tornado/Climate eg. deals with the situation where we have one population with 3 categories. For \( k \) categories:

We learned that the relevant dist. is \( \chi^2 \)-squared with \( df = k - 1 \). And the quantity that follows that dist. is:

\[ x^2 = \sum_{i=1}^{k} \left( \frac{\text{obs}_i - \text{exp}_i}{\sqrt{\text{exp}_i}} \right)^2 \]

where \( \text{obs}_i \) and \( \text{exp}_i \) are observed and expected counts in the \( i^{th} \) category (still of 1 population). The latter are computed assuming \( H_0 \) is true, where

\[ H_0 : \pi_1 = \pi_{01} \quad \pi_2 = \pi_{02} \quad \ldots \quad \pi_k = \pi_{0k} \]

\[ H_1 : \text{At least one of these is wrong.} \]

Finally, the chi-sq test tests indep. of \( X, Y \), both categorical:

\[ H_0 : X \& Y \text{ are indep.} \quad H_1 : \text{They are not indep.} \]
By hand
Consider the data from an example in a past lecture where a survey of students in 390 yielded the following data:
17 students like lab
48 do not like lab
15 have no opinion.
Suppose I believed that the proportion of students in each of the 3 categories (like, no-like, no-opinion) was equal. Does this data contradict that belief? Let alpha=0.05.

By hand
A sample of 210 Bell computers has 56 defectives. Theory suggests that a third of all Bell computers should be defective. Does this data contradict the theory (at alpha=0.05)? Specifically,
a) Do a z-test,
b) Do a chi-squared test with k=2 categories. Hint: The pi's (and pi_0's) of the k categories must sum to 1.
c) Are the conclusions in a and b consistent?

By hand
Have you ever wondered whether soccer players suffer adverse effects from hitting "headers"? The authors of the article "No Evidence of Impaired Neurocognitive Performance in Collegiate Soccer Players" (The Amer. J. of Sports Medicine, 2002: 157-162) investigated this issue. The paper reported that 45 of the 91 soccer players in their sample had suffered concussion, 28 of 96 nonsoccer athletes had suffered concussion, and only 8 of 53 student controls had suffered concussion. Suppose we want to apply the chi-squared test to this problem. I hope it's clear that only the second test (i.e. test of independence) is possible.
a) What are the categorical variables whose independence can be tested? When you identify them, make sure you also state the number of levels for each.
b) State the hypotheses.
c) Write the data in the form of a contingency table.
d) Compute the expected counts.
e) Compute the p-value (or specify a range for it).
f) State the conclusion "in English."