Today: Section 6.4
Assessing the Statistical Significance of a 2x2 Table

Homework (due Friday, October 16):
6.40 and 15.10 (yes, really Chapter 15!)

Example (Case Study 4.3, p. 129): Randomized experiment
Explanatory variable = wear nicotine patch or placebo
Response variable = Quit smoking after 8 weeks?   Yes/ No

Results:

	Quit	Didn't	**Total**	% Quit
Nicotine	56	64	120	46%
Placebo	24	96	120	20%
Total	80	160	240	33%

Relative "risk" of quitting with nicotine patch is .46/.20 = 2.3

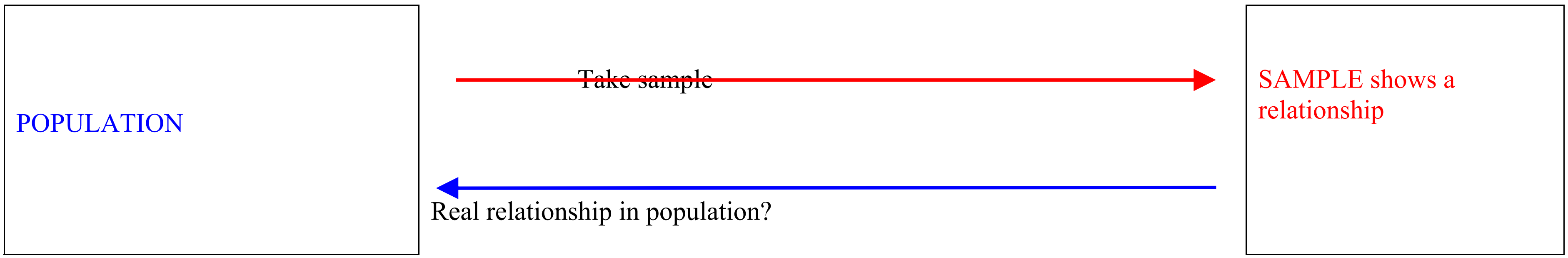
Question: Could the observed relationship be due to chance, or is there really a difference in proportions who would quit in the *population* from which this *sample* was taken?

Definitions (from Chapters 1 and 3):

A **population** is the entire group of units (college students, Old Faithful eruptions, babies, cities in US, smokers, …) about which information is desired.

A **sample** is the subset of the population for which measurements are available.

Nicotine Patch Example:
Population: Smokers with a desire to quit
Sample: 240 smokers at Mayo clinics in Minnesota, Florida and Arizona, who volunteered to participate.
Goal of **statistical inference**: Use the data from the sample to make conclusions (*inferences*) about the population.



So far, *descriptive statistics*

Now: *Inferential statistics*

- Hypothesis tests (Chapter 6, then Chs 12, 13)
- Confidence intervals (Chapter 3, then Chs 10, 11)

Confidence interval – An interval of values that we are "confident" covers the truth about a population value. (Ch 4)

Hypothesis test (also called a significance test) – based on *sample* determine if there is a relationship, difference, etc., in the *population*.

Definitions:
A ***statistic*** is a numerical summary of the data in a *sample*.
Ex: mean, median, correlation, etc, computed from *sample.*

A ***parameter*** is a number association with a *population*.
Ex: mean of a *population*, such as male heights for *all* college students.
A ***test statistic*** is a summary of the data from the sample, used in a hypothesis test.

A ***chi-square statistic***

- The test statistic we use to assess the strength of the relationship in a two-way table, and to determine if the relationship is "statistically significant".
- More complicated summary than seen so far, but still, just a numerical summary of sample data!

- Measures how far the *observed* numbers in the cells are from what we would *expect* if there is no relationship

Five Steps to Determining Statistical Significance (**page 208**)

Here is how to do a hypothesis test:

Step 1: Determine the *null* and *alternative* hypotheses.

Step 2: Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.

Step 3: Assuming the null hypothesis is true, find the p-value.

Step 4: Decide whether or not the result is statistically significant based on the p-value.

Step 5: Report the conclusion in the context of the situation.

For two categorical variables (two-way table):
Step 1: Determine the null and alternative hypotheses.

In general:
- Null hypothesis is "nothing going on," status quo, no difference, etc.
- Alternative hypothesis is what researchers hope to show, that something interesting *is* going on.

For contingency tables:
- Null hypothesis: The two variables are *not* related in the population.
- Alternative hypothesis: The two variables *are* related in the population.

Nicotine Patch Example:

Think of what happens in the population as the hypothetical behavior of *all* smokers with a desire to quit, *if* given nicotine patch compared with *if* given placebo patch.

Null hypothesis: In the population of smokers who want to quit, whether people quit or not is *not* related to patch type.

Alternative hypothesis: In this population, quitting or not *is* related to patch type.

Step 2: Verify necessary data conditions [skip for now], and if met, **summarize the data into an appropriate test statistic.**

For two categorical variables, remember, the appropriate test statistic is called a *chi-square statistic*.

Logic of the chi-square statistic:

- Compute *expected counts* under the assumption of *no relationship in population* , i.e, assuming *null hypothesis* is true (will see how to do this)
- Compare these to *observed counts* in the cells of the table, using a summary measure (to be shown)

E = Expected count in each cell =

$$\frac{(Row\ total)(Column\ total)}{Total\ n}$$

O = Observed count in each cell = actual sample data

	Quit	Did not quit	Total
Nicotine	**56** **(120)(80)/240** **= 40**	**64** **(120)(160)/240 = 80**	**120**
Placebo	**24** **(120)(80)/240** **= 40**	**96** **(120)(160)/240 = 80**	**120**
Total	**80**	**160**	**240**

NOTE: Only need compute E for one cell, others determined by totals

Why do these "expected counts" make sense if the null hypothesis is true?

- Overall, 80/240 = 1/3 quit (see "Total" row).
- If *no relationship*, we would expect (80/240) to have quit in *each* treatment (each row of the table).
- So, we expect 120 × 80/240 = 40 to have quit in each treatment (row) and 120 × 160/240 = 80 to have *not quit* in each treatment.

	Quit	Did not quit	Total
Nicotine	**56** **40**	**64** **80**	**120**
Placebo	**24** **40**	**96** **80**	**120**
Total	**80**	**160**	**240**

Continuing Step 2, Creating the test statistic:

- For each cell, summarize difference between "observed" and "expected" counts, using

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

- Sum these over all cells.

Chi-square statistic: ***Notation, Greek letter "chi"***

$$\chi^2 = \sum_{all\ cells} \frac{(O-E)^2}{E}$$

Example: How far are *observed* numbers who quit from what we *expect* if there is no difference for patch types?

$\frac{(56-40)^2}{40} = \frac{256}{40} = 6.4$	$\frac{(64-80)^2}{80} = \frac{256}{80} = 3.2$
$\frac{(24-40)^2}{40} = \frac{256}{40} = 6.4$	$\frac{(96-80)^2}{80} = \frac{256}{80} = 3.2$

So, $\chi^2 = 6.4 + 3.2 + 6.4 + 3.2 = 19.2$

Does that indicate a large difference or a small one??? A strong relationship or no relationship at all in the population?

Step 3: Assuming the null hypothesis is true, find the p-value.

Decide *how unlikely* such a big difference *in the sample* would be *if* there is no *real relationship in the population.* This is a black box to you for now! Called the **p-value**.

Using R Commander (See handout on website):

```
Statistics -> Contingency Table ->
Enter and analyze two-way table
```

Example:

`X-squared = 19.2, df = 1, p-value = 1.177e-05` [.00001177]

Note: You can use Excel, but you need to find the expected counts yourself first. See page 212 in book.

Step 4: Decide whether or not the result is statistically significant, based on the p-value.

Possible conclusions:

Do not reject the null hypothesis and conclude there isn't enough evidence to convince us that there is a relationship in the population. (IF p-value > .05)

(Use .05 or other "level of significance")

Reject the null hypothesis and conclude there *is* a relationship in the population (IF p-value < .05)

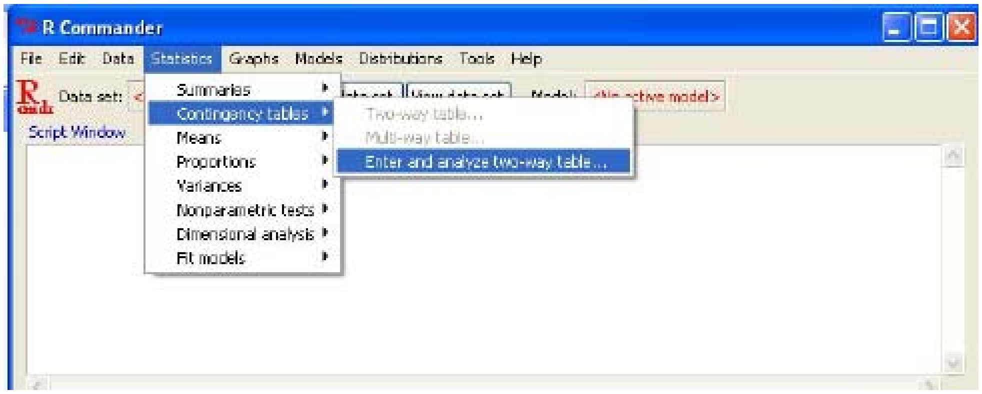
Other ways to say we *do not reject the null hypothesis*:

- There is *not enough evidence* to support the alternative hypothesis
- There is *not enough evidence* to reject the null hypothesis
- The relationship is *not statistically significant*

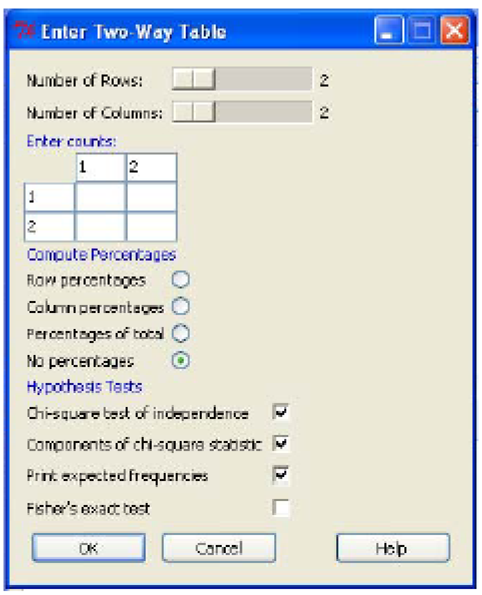
NOTE: It is *not okay* to "*accept the null hypothesis*."

How to use R Commander for Contingency Tables (Section 6.4)

Statistics → Contingency tables → Enter and analyze two-way table…
The screen shot below shows where to find these in the menu structure:



You should see a pop-up box like the one below. Check the first 3 boxes under "Hypothesis Tests" as shown in the screen shot. The default is to get the chi-square test statistic and p-value. Checking these boxes will give you the contributions from each cell, and the expected counts. If you want row, column, or total percentages, check those boxes as well.



Enter the data into the cells of the table, then click "OK". The results should look like this (these are the results for the nicotine patch example):

```
       Pearson's Chi-squared test

data:  .Table
X-squared = 19.2, df = 1, p-value = 1.177e-05

> .Test$expected # Expected Counts
   1  2
1 40 80
2 40 80

> round(.Test$residuals^2, 2) # Chi-square Components
    1   2
1 6.4 3.2
2 6.4 3.2
```

Other ways to say we *reject* the null hypothesis:

- We *accept* the alternative hypothesis
- There is a *statistically significant* relationship between the two variables.

Step 4 for the nicotine patch example:
The p-value is *much* less than .05, so relationship *is* statistically significant. We reject the null hypothesis. We accept the alternative hypothesis.

Step 5: Report the conclusion in the context of the situation.

Step 5 for the nicotine patch example:
There is a statistically significant relationship between type of patch worn and the ability to quit smoking.

And, because this was a randomized experiment, we can conclude that wearing nicotine patches would *cause* more people to quit smoking than wearing a placebo patch.

Cautions:

1. p-value depends on sample size. Easier to detect real difference with *larger* sample. Therefore, *failure to detect stat significance does not mean there is no relationship.*
   Example: Aspirin and heart attacks (Case Study 1.6):
   - $\chi^2 = 25.4$, p-value ≈ 0, clearly there is a relationship.
   - Suppose sample size cut by factor of 10 but same pattern.
     Chi-square statistic = 2.54, p-value = .111, *not* statistically significant.

2. *Statistical* significance is not the same thing as *practical* significance. With a *very large sample* even a *minor* relationship will be statistically significant.

New example: Question asked in Discussion 1

Explanatory: Are you employed or not?
Response: Would you leave a note if you hit a car? Yes/No

Employed?	Note	No Note	Total

Yes	20	6	26
No	5	8	13
Total	25	14	39

<u>Population</u>: All college students similar to those who take Statistics 7 at UCI.

Step 1: Determine the null and alternative hypotheses

<u>Null hypothesis</u>:
For the population of students, leaving a note or not is *not* related to employment status

<u>Alternative hypothesis</u>:
Leaving a note or not *is* related to employment status (in the population of students).

Steps 2 and 3: Compute the test statistic and p-value

Employed?	Note	No Note	Total
Yes	20	6	26
No	5	8	13
Total	25	14	39

Results from R Commander:

```
X-squared = 5.5714, df = 1, p-value = 0.01826
Expected Counts
          1        2
1 16.666667 9.333333
2  8.333333 4.666667
```

(For example, (26)(25)/39 = 50/3 = 16.67)

```
Chi-square Components
     1    2
1 0.67 1.19
2 1.33 2.38
```

(Interesting to look for cell with largest value here)

Steps 4 and 5: Conclusion in statistical terms and context

Step 4: Statistical conclusion:
p-value = .01826 is less than .05, so *reject* the null hypothesis. The relationship *is* statistically significant.

Step 5: Conclusion in context:

There is a statistically significant relationship between employment status and whether someone would leave a note if they hit a car.

We can conclude that the relationship holds in the *population* represented by this sample.