**Announcements:**

- First midterm is a week from Friday (Feb 1), covering Chapters 1 to 6. Sample questions (and answers) have been on course website all along. Mixture of free-response and multiple choice. You are allowed one sheet of notes (both sides of page, typed or hand-written).
- Friday discussion *is* for credit this week.

**Today**: Section 4.4
Assessing the Statistical Significance of a 2×2 Table
**Homework (due Wed, Jan 30):**
Chapter 4: #49, 50 (count together as 1)
Chapter 15: #10, #12 (Use R Commander, counts double)

## Review from last time

What to do with two categorical variables:

- Create a "contingency table" with explanatory variable as rows, response variable as columns.
- Each combination of row and column is a *cell*.
- Use table to compute risk, relative risk, increased risk, odds, odds ratio.

Sometimes these measures don't make sense – just want to know if the two variables are *related*.

**Today**: How to determine if two categorical variables have a *statistically significant relationship*.

**<u>Example</u>** (Case Study 6.3, p. 199): 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 (baseline) | 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?

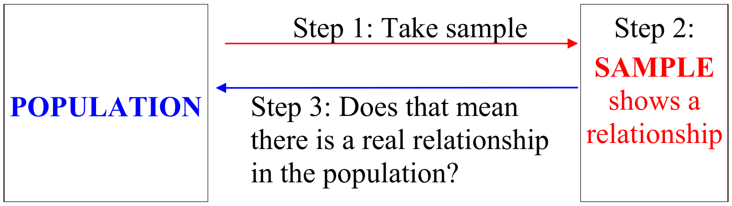
Definitions (from Chapters 1 and 5):

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** consists of the units in a subset of the population, for which measurements are available.

**Nicotine Patch Example:**
***Population***: <u>All</u> smokers with a desire to quit
***Sample***: <u>240 smokers</u> 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 4, then Chs 12, 13)
- Confidence intervals (Chapter 5, then Chs 10, 11)

*Confidence interval* – An interval of values that we are "confident" covers the truth about a population value. (Ch 5)

*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 associated with a *population*.
Example: Mean of a *population*, such as male heights for *all* college students. (Usually, value of parameter is unknown because we can't measure the whole population.)

A ***test statistic*** is a statistic that summarizes sample data into a number that can be 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 decide 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 between the explanatory and response variables in the population.

**Nicotine Patch Example: What to Expect if No Relationship**

| *Observed counts* | **Quit** | **Didn't** | **Total** | **% Quit** |
|---|---|---|---|---|
| Nicotine | 56 | 64 | 120 | 46% |
| Placebo (baseline) | 24 | 96 | 120 | 20% |
| **Total** | 80 | 160 | 240 | **33%** |

- Note that 80/240 = 1/3 (or 33%) quit smoking overall
- *If there is no difference* in the effect of patch type, we *expect* to see 1/3 of each type quit. So, we would expect:

| *Expected counts* | **Quit** | **Didn't** | **Total** | **% Quit** |
|---|---|---|---|---|
| Nicotine | 40 | 80 | 120 | **33%** |
| Placebo (baseline) | 40 | 80 | 120 | **33%** |

Five Steps to Determining Statistical Significance (**page 126**)

Here is how to do a hypothesis test:

**Step 1:** Determine the *null* and *alternative* hypotheses. (These are statements about the population.)

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

**Step 3:** Find the *p-value,* by assuming the null hypothesis is the truth about the population, and then computing how unlikely the (sample) test statistic would be in that case.

**Step 4:** Decide whether or not the result (relationship) 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. in the population.
- Alternative hypothesis is what researchers hope to show, that something interesting *is* going on in the population.

*For contingency tables*:

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

**Step 1 for the Nicotine Patch Example**:

**Population**: 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, there is no relationship between patch type and whether or not someone quits smoking.

Alternative hypothesis: In this population, there *is* a relationship between patch type and whether or not someone quits smoking.

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

For two categorical variables

- Data condition: All expected counts $\geq 1$, at least three $\geq 5$
- The test statistic is called the *chi-square statistic*.

**Logic of the chi-square statistic**:

- Compute *expected counts* under the assumption of *no relationship in population* , i.e when *null hypothesis* is true
- Compare these to *observed counts* in the cells of the table, using a summary measure (to be shown)
- If they are very different (far apart), conclude that there *is* a relationship between explanatory and response variables.

**O = Observed count** in each cell = actual sample data
**E = Expected count** (if null is true) in each cell =

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

| | Quit | Did not quit | Total |
|---|---|---|---|
| **Nicotine** | **56**<br>**(120)(80)/240 = 40** | **64**<br>**(120)(160)/240 = 80** | **120** |
| **Placebo** | **24**<br>**(120)(80)/240 = 40** | **96**<br>**(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 for the population?

- 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. These match "expected count" formula.

| | **Quit** | **Did not quit** | **Total** |
|---|---|---|---|
| **Nicotine** | **56 (Observed)**<br>**40 (Expected by chance)** | **64**<br>**80** | **120** |
| **Placebo** | **24**<br>**40** | **96**<br>**80** | **120** |
| **Total** | **80** | **160** | **240** |

**Continuing Step 2, Creating the test statistic**:

- For each cell, summarize difference between "observed" counts ($O$) and "expected" counts ($E$), using

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

- Sum these over all cells.

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

$$\chi^2 = \sum_{\text{all 4 cells}} \frac{(O-E)^2}{E}$$

Example: How far are *observed* numbers who quit from what we *expect* if there is no population 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:** Find the *p-value,* calculated by assuming the null hypothesis is true for the population.

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 128 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. Conclude there isn't enough sample evidence to convince us that there is a relationship in the population. Conclude IF p-value > .05.
(Use .05 or other "*level of significance*")

*Reject* the null hypothesis. Conclude there *is* a relationship in the population. Conclude this IF p-value ≤ .05.

Equivalent 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*."

Equivalent 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 of .00001177 is *much* less than .05, so relationship *is* statistically significant. We reject the null hypothesis. We accept the alternative hypothesis. The relationship *is statistically significant*.

**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. In other words, conclude that this is a real relationship in the population.*

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.

**Caution #1:** *p*-value depends on sample size. Easier to detect a real difference with *larger* sample. Therefore, *failure to detect a statistically significant relationship does not mean there is no relationship.*

Example: Aspirin and heart attacks (Case Study 1.6): Based on 22,071 men.

- $\chi^2 = 25.4$, p-value $\approx 0$, clearly there *is* a relationship between aspirin (yes/no) and heart attack (yes/no).
- Suppose the sample size is cut by a factor of 10 (to 2207) but same pattern, i.e. all observed counts are divided by 10 as well.

  Chi-square statistic = 2.54

  p-value = .111, *not* statistically significant.

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

Example: Suppose a drug is compared to a placebo:

| | Cured | Not Cured | Total | % Cured |
|---|---|---|---|---|
| Drug | 5100 | 4900 | 10000 | 51% |
| Placebo | 4900 | 5100 | 10000 | 49% |
| Total | 10000 | 10000 | 20000 | 50% |

Relative "risk" of cure = 51/49 = 1.04
Chi-square statistic = 8.0, *p*-value = .0047
Clearly reject the null hypothesis, conclude drug works!
But the difference is of *little practical importance*.

**New example: Question asked in Discussion 1 last year**

Explanatory: Sex (Male/Female)
Response: If no cops around, would you speed over 90?

| | Yes | No | Total |
|---|---|---|---|
| Male | 13 | 2 | 15 |
| Female | 15 | 18 | 33 |
| Total | 28 | 20 | 48 |

- Note that over half (58%) said yes, but most males (87%) and fewer than half of females (45.5%).
- Does M/F difference in *sample* reflect a true difference in the *population*, or is it just a chance difference?

Population: College students similar to UCI Stat 7 students.

**Step 1: Determine the null and alternative hypotheses**
*Two versions are shown here for each hypothesis.*

Two equivalent ways to state the Null hypothesis:

- For the *population* of students, proportion who would speed *does not differ* for males and females.
- For the *population* of students, speeding (yes/no) and sex (M/F) *are not related.*

Two equivalent ways to state the Alternative hypothesis:

- For the population of students, proportion who would speed differs for males and females.
- For the population of students, speeding (yes/no) and sex (M/F) *are* related.

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

| | Yes | No | Total |
|---|---|---|---|
| Male | 13 | 2 | 15 |
| Female | 15 | 18 | 33 |
| Total | 28 | 20 | 48 |

Results from R Commander:

| | |
|---|---|
| `X-squared = 7.2062, df = 1, p-value = 0.007265` | |
| **`Expected Counts`** | Compare these to observed counts in table. |
| `Yes No`<br>`Male 8.75 6.25`<br>`Female 19.25 13.75` | For example, for Male, Yes:<br>Expected = (28)(15)/48 = 8.75<br>Observed = 13 |
| Note: 13 (M, Yes) and 18 (F, No) larger than expected;<br>15 (F, Yes) and 2 (M, No) smaller than expected. | |
| `Chi-square Components`<br>`Yes No`<br>`Male 2.06 2.89`<br>`Female 0.94 1.31` | Interesting to look for cells with large values here.<br>In this case, males deviate more from expected. |

**Steps 4 and 5: Conclusion in statistical terms and context**

Step 4: Statistical conclusion:
*p*-value = .007 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 sex and whether someone would speed over 90 mph.

The relationship in the sample was strong enough that we can conclude that the relationship holds in the *population* represented by this sample.

**A Few More Details (for 2 × 2 tables only)**

1. If you have to compute the test statistic by hand, there is a "short-cut" formula; see page 593 (Chapter 15):

| | Column 1 | Column 2 | Total |
|---|---|---|---|
| Row 1 | A | B | $A+B = R_1$ |
| Row 2 | C | D | $C+D = R_2$ |
| Total | $A+C = C_1$ | $B+D = C_2$ | N |

$$\chi^2 = \frac{N(AD - BC)^2}{R_1 R_2 C_1 C_2}$$

2. For a 2 x 2 table only (2 rows and 2 columns):
*p-value* $\leq$ .05 if and only if chi-square value $\geq$ 3.84.
So *statistically significant relationship* if $\chi^2 \geq 3.84$.
In our example, $\chi^2 = 7.2052 > 3.84$.

**Homework (due Wed, Jan 30):**

Chapter 4: #49, 50 (count together as 1);
Chapter 15: #10, #12 (Use R Commander, counts double)
(Yes, really Chapter 15!)