Steps in a hypothesis test (from page 475 of the book)

1. HYPOTHESES: Determine the null hypothesis and the alternative hypothesis.

2. TEST STATISTIC: Collect data and summarize them with a single number known as a test statistic.

3. P-VALUE: Determine how unlikely the test statistic is if the null hypothesis is true.

4. DECISION: Make a decision and explain it in plain English.

Assignment: Read Chapter 24
Take a look at the Exam #3 information

Nov. 18
Calcium and PMS

(Case study 21.1, pp. 462–463, and exercise 3, p. 490.) A study was conducted to see if a calcium supplement relieves the symptoms of premenstrual syndrome (PMS). Women were randomly assigned a placebo or a calcium supplement and a measure of severity of PMS was recorded.

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Calcium</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>.60</td>
<td>.43</td>
</tr>
<tr>
<td>SD</td>
<td>.52</td>
<td>.40</td>
</tr>
<tr>
<td>sample size</td>
<td>228</td>
<td>212</td>
</tr>
</tbody>
</table>

The p-value is [intentionally left blank; find it!]

What is your decision?
Hypotheses for calcium study?

Alternative (Research) Hypothesis: Calcium will reduce the severity of PMS, as compared with placebo.

Null Hypothesis (Skeptic): There is no effect due to calcium except the placebo effect.

The alternative in this case is one-sided.
The researchers concluded that calcium is effective. Which type of error could they have committed?

Type 1 error: claim calcium is effective when it is not.

Type 2 error: claim calcium is not effective when it is effective.

Since the researchers supported the Research Advocate, the only possible error is Type 1.
Can you understand this? (excerpt from a 2003 article)

**Correlates of the Hangover Symptoms Scale**

The HSS was significantly positively associated with the frequency of drinking \( r = 0.44, p < 0.001 \) and getting drunk \( r = 0.52, p < 0.001 \) and with the typical quantity of alcohol consumed when drinking \( r = 0.40, p < 0.001 \) in the past year. Thirty-two percent of the participants reported experiencing at least one of four alcohol-related problems in their lifetime. Those who reported an alcohol-related problem had significantly higher scores on the HSS.
Suppose you have a suspicious coin.

You flip this coin 100 times and observe 60 heads.

Set up a statistical hypothesis test to determine whether this is convincing evidence that the coin is weighted unfairly to come up heads.

OR

Set up a statistical hypothesis test to determine whether this is convincing evidence that the coin is weighted unfairly.

What's the difference?

(A) Alternative vs. Null  (C) Hypothesis test vs. CI
(B) 1-sided vs. 2-sided  (D) Observational study vs. experiment
The test statistic in this class is always a STANDARDIZED SCORE:

Proportion of heads: .60
Fair coin: .50

\[ \frac{\text{Estimate} - \text{Null value}}{\text{Standard deviation of estimate}} = 2.0 \]

SD of sample proportion:
\[ \sqrt{\frac{0.5 \times 0.5}{100}} = 0.05 \]
Two-sided alternative hypothesis

Calculate the test statistic,

\[
\frac{0.6 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{100}}} = 2.0
\]

then find the probability of seeing a test statistic more extreme on EITHER side.
Two-sided p-value cont'd

Test statistic = 2.0

From p. 175:

Probability larger than 2.0:  about 2.5%
Probability smaller than −2.0:  about 2.5%

Thus, p-value equals .025+.025 = .05
What decision do we make with a p-value of .05?

Good question! It's dangerous to proclaim strongly "reject the null!" or "fail to reject the null!"

Perhaps the best answer: "We have moderate but not strong statistical evidence (p=.05). Further study may be called for."

What if the p-value had been 0.005?

(A) Reject Alternative   (C) Reject Null
(B) Accept Null         (D) Fail to Reject Null
Remember these steps!

1. Hypotheses
2. Test statistic
3. p-value
4. Decision

Research question: Do _____ M&M's really occur in the proportion previously claimed on the M&M's website?

(Sadly, Mars Incorporated no longer appears to post this information. I'm devastated.)

<table>
<thead>
<tr>
<th>Data:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yellow</td>
<td>54/427</td>
<td>.14</td>
<td>35/427</td>
</tr>
<tr>
<td>Blue</td>
<td>123/427</td>
<td>.24</td>
<td>84/427</td>
</tr>
<tr>
<td>Green</td>
<td>51/427</td>
<td>.16</td>
<td>80/427</td>
</tr>
<tr>
<td>Brown</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(54/427) Yellow (123/427) Blue (51/427) Green (35/427) Brown (84/427) Red (80/427) Orange
**Hours of exercise per week, STAT 100, SP 2015**

Research Hypothesis: In 2015, PSU women spent a different amount of time exercising during the semester than men.

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.3</td>
<td>7.1</td>
</tr>
<tr>
<td>SD</td>
<td>3.5</td>
<td>5.2</td>
</tr>
<tr>
<td>sample size</td>
<td>48</td>
<td>22</td>
</tr>
</tbody>
</table>

A couple of outliers have been removed here.
Dotplots of exercise hours per week.

Note large outliers, probably not valid.

Also note non-normal distributions (but we have large samples, so the theoretical sample means will still be normal)
Dotplots of exercise hours per week, outliers removed

Female

Male
Which is true of the female exercise data?

(A) Mean > median
(B) Median = mean
(C) Mean < median
# Hours of exercise per week, STAT 100, SP 2015

Research Hypothesis: In 2015, PSU women spent a different amount of time exercising during the semester than men.

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>4.3</td>
<td>7.1</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>3.5</td>
<td>5.2</td>
</tr>
<tr>
<td><strong>sample size</strong></td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td><strong>p-value (2-sided)</strong></td>
<td></td>
<td>Let’s calculate it!</td>
</tr>
</tbody>
</table>

Conclusion?
“Bonferroni correction”? the presence or absence of the symptom. Men and women were equally likely to experience at least one of the hangover symptoms in the past year (men: 89%; women: 87%; $\chi^2 = 1.2$, df = 1, N = 1215, $p = 0.282$). Using a liberal criterion for statistical significance ($p < 0.05$), there were no symptoms that women experienced more often than men and there were only two symptoms that men experienced more often than women: vomiting (men: 50%; women: 44%; $\chi^2 = 4.7$, df = 1, N = 1215, $p = 0.031$) and sweating more than usual (men: 34%; women: 23%; $\chi^2 = 18.9$, df = 1, N = 1214, $p < 0.001$). After applying a Bonferroni correction, only the latter difference remained statistically significant; after controlling for the frequency of drinking and getting drunk and the typical quantity of alcohol consumed when drinking in the past year, neither sex difference remained.