I. The matched pairs $t$ test (Section 9.5)

- The procedures of Section 9.3 for comparing two means require independent samples from the two populations.

- Two important designs that don’t fit that framework are “before-after” studies and “matched pairs studies.”
Example: Before-after study

- Interested in whether a short summer course improves language skills of high-school foreign language teachers.
- Before the course, teachers were given a listening test.
- After the course, teachers were given a similar test.
- Want to see whether the course improved their listening skills.
- Clearly a teacher’s score on the second test is not independent of his score on the first test.
Example: Matched pairs study

- Want to determine whether a weight-training program increases bone density.

- Choose 20 people at random to participate in the weight-training program.

- For each of these people, choose another of the same age and gender who will not participate.

- Because we’ve matched the subjects based on age and gender, the scores from the two samples are likely not independent.

- The advantage of matching is that it hopefully eliminates the effect of the variables age and gender.
The matched pairs t test

- It’s easy to analyze data from such studies.
- Basic assumption of normality of data as usual.
- Let $d_1, d_2, \ldots, d_n$ represent the differences for the two samples.
- Let $\overline{d}$ and $S_d$ represent the sample mean and sample standard deviation of the differences.
- We can follow the one-sample t procedures from Chapter 8, using the differences as our sample.
Example: Before-after continued

- Recall that we’re interested in whether short course improved teachers’ language skills.

- Some of the data are in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>35</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>26</td>
<td>-6</td>
</tr>
</tbody>
</table>

- Altogether there are 20 pairs of scores.

- We’re interested in testing

  \[ H_0 : \mu_d \leq 0 \]
  \[ H_a : \mu_d > 0 \]

  where \( \mu_d \) is the population mean difference.
The sample mean difference is \( \bar{d} = 2.5 \).

The sample standard deviation of the differences is \( S_d = 2.893 \).

The test statistic is

\[
T = \frac{\bar{d}}{S_d/\sqrt{n}}.
\]

Plugging in the data values yields

\[
t_{\text{obs}} = \frac{2.5}{2.893/\sqrt{20}} \approx 3.86.
\]

We use the \( t \) density with \( n - 1 = 19 \) degrees of freedom to compute the \( p \)-value.

From Table B.3 this is less than 0.001.

Since the \( p \)-value is so small, we have strong evidence that the course increases listening skills test scores.
• For a confidence interval for $\mu_d$, we use

$$\bar{d} \pm c \frac{S_d}{\sqrt{n}}$$

where $c$ comes from Table B.3 with $n - 1 = 19$ degrees of freedom.

• For a 98% confidence interval this leads to

$$2.5 \pm 2.539 \frac{2.893}{\sqrt{20}} \approx 2.5 \pm 1.64.$$
Two methods of inference for the difference of means

- Goal is inference (test or CI) for $\mu_1 - \mu_2$.

- Independent samples:
  - Degrees of freedom are minimum of $n_1 - 1$ and $n_2 - 1$.
  - Test statistic is
    $T = \frac{X_1 - X_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$
  - Confidence interval is
    $(X_1 - X_2) \pm c \sqrt{S_1^2/n_1 + S_2^2/n_2}$.

- Paired samples (usually before-after study):
  - Degrees of freedom are $n - 1$.
  - Test statistic is
    $\frac{\bar{d}}{S_d/\sqrt{n}}$.
  - Confidence interval is
    $\bar{d} \pm c \frac{S_d}{\sqrt{n}}$. 


II. Chapter 10: Categorical Data

- In Chapter 10 we’ll learn how to perform hypothesis tests for categorical data in three settings.

- Deciding whether a probability distribution fits a categorical population. (Section 10.1, Goodness-of-fit).

- Deciding whether two categorical variables are independent. (Section 10.2, Test of Independence).

- Deciding whether two or more categorical populations have the same distribution. (Section 10.3, Test of Homogeneity).

- In all three cases we use the same idea to construct the test statistic, and use the same distribution (called chi-squared) to compute p-values.
Goodness-of-fit tests (Section 10.1)

- The goal is to determine whether a probability distribution fits a discrete population well.
- The basic idea is simple: We compare the data to what we would expect to see if the probability distribution is correct.
  - If the data are *far* from what we’d expect to see, we have evidence against the model.
  - If the data are *close* to what we’d expect to see, we have evidence in favor of the model.
Example: Mendel’s genetic theories

• Gregor Mendel postulated models for inheritance that form the basis for much of modern genetics.

• How could a scientist test Mendel’s theories?

• For example, a type of pea plant has 4 kinds of seeds: smooth yellow, smooth green, wrinkled yellow, wrinkled green.

• Mendel’s model predicts that these should occur in proportions 9/16, 3/16, 3/16, and 1/16.

• We’ll compare data on these peas to what Mendel’s model predicts.
• The null hypothesis is that Mendel’s model is correct. Symbolically,
\[ H_0: P(SY) = \frac{9}{16}; \quad P(SG) = \frac{3}{16}; \quad P(WY) = \frac{3}{16}; \quad P(WG) = \frac{1}{16}. \]
• The book would write this as
\[ H_0: \pi_1 = \frac{9}{16}; \quad \pi_2 = \frac{3}{16}; \quad \pi_3 = \frac{3}{16}; \quad \pi_4 = \frac{1}{16}. \]
Here are the observed data with what we’d expect if $H_0$ were true:

<table>
<thead>
<tr>
<th>Category</th>
<th>SY</th>
<th>SG</th>
<th>WY</th>
<th>WG</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>275</td>
<td>90</td>
<td>98</td>
<td>37</td>
<td>500</td>
</tr>
<tr>
<td>Expected</td>
<td>281.25</td>
<td>93.75</td>
<td>93.75</td>
<td>31.25</td>
<td>500</td>
</tr>
</tbody>
</table>

- We get the expected counts by multiplying the total (500) by the hypothesized proportions.
- For example, the expected number of SY seeds is $(500)(9/16) = 281.25$. The expected number of WG seeds is $(500)(1/16) = 31.25$.
- The expected counts are not exactly the same as the observed counts. The question is whether the difference is due to $H_0$ being untrue, or due to the fact that we’re only seeing a sample, not the whole population.
• First technical question: How should we measure how “far” the observed counts are from what we’d expect?

• Answer: We’ll call the answer $\chi^2$ (read “chi-squared”). Let $o_j$ and $e_j$ represent the observed and expected counts for category $j$. Then

$$\chi^2 = \sum \frac{(o_j - e_j)^2}{e_j}$$

• For the Mendel data this becomes

$$\chi^2 = \frac{(275 - 281.25)^2}{281.25} + \frac{(90 - 93.75)^2}{93.75} + \frac{(98 - 93.75)^2}{93.75} + \frac{(37 - 31.25)^2}{31.25}$$

$$= .139 + .150 + .193 + 1.058$$

$$= 1.54.$$
• Second technical question: How large does $\chi^2$ have to be for us to reject $H_0$?

• Answer: We’ll use a “chi-squared” table to compute p-values.
  – We’ll always compute areas to the right, since large values of $\chi^2$ give evidence against $H_0$.
  – The “degrees of freedom” will be the number of categories minus 1.
  – In our case, we have $4 - 1 = 3$ degrees of freedom.

• From the chi-squared table we see that the area to the right of 0.584 is 0.900 and the area to the right of 4.64 is 0.200, so the p-value is between 0.200 and 0.900. Clearly there’s not much evidence against $H_0$. 
Example:

- Are births equally likely to take place on any day of the week?

Data:

<table>
<thead>
<tr>
<th>Category</th>
<th>Weekend</th>
<th>Weekday</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>300</td>
<td>600</td>
<td>900</td>
</tr>
</tbody>
</table>

- Since the categories are “Weekend” with 2 days and “Weekday” with 5 days, the null hypothesis is

\[ H_0: \pi_1 = \frac{2}{7}; \quad \pi_2 = \frac{5}{7}. \]

- The expected counts are

\[ (900)\left(\frac{2}{7}\right) \approx 257.14 \text{ and } (900)\left(\frac{5}{7}\right) \approx 642.86. \]
Data with expected counts:

<table>
<thead>
<tr>
<th>Category</th>
<th>Weekend</th>
<th>Weekday</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>300</td>
<td>600</td>
<td>900</td>
</tr>
<tr>
<td>Predicted</td>
<td>257.14</td>
<td>642.86</td>
<td>900</td>
</tr>
</tbody>
</table>

The chi-squared statistic is

$$\chi^2 = \frac{(300 - 257.14)^2}{257.14} + \frac{(600 - 642.86)^2}{642.86}$$

$$= 7.14 + 2.86 = 10.00.$$

- The p-value is the area to the right of 10 under a chi-squared density with $2 - 1 = 1$ degrees of freedom.
- From the chi-squared table this is between 0.001 and 0.0025.
- Since the p-value is so small, we have strong evidence against the equally likely theory.
- From the data it seems that there are more weekend births than expected.
III. Chi-squared tests: basics

- Goal is to decide between two hypotheses $H_0$ and $H_a$.
- Collect data with observed count $o_j$ in category $j$.
- Compute the expected count $e_j$ in category $j$, assuming $H_0$ is true.
- Compute the “chi-squared statistic”

$$\chi^2 = \sum \frac{(o_j - e_j)^2}{e_j}$$

which measures how far the observed data are from what we expect if $H_0$ is true.
- The p-value is the area to the right of $\chi^2$ from the “chi-squared” table, Table B.4.
- Interpreting $\chi^2$ and the p-value:
  - Larger $\chi^2$ gives stronger evidence against $H_0$
  - Smaller p-value gives stronger evidence against $H_0$
Application 1: Goodness of fit, Section 10.1

- Expected count in category $j$ is $n\pi_j$.
- Degrees of freedom is number of categories minus 1.
IV. Application 2: Test of Independence, Section 10.2

- Are smoking and Socioeconomic status independent?

- A sample of $n = 356$ adults were classified on their SES (High, Middle, Low), and Smoking status (Current, Former, Never).

- Here are the data:

<table>
<thead>
<tr>
<th>SES</th>
<th>Smoking</th>
<th>High</th>
<th>Middle</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>51</td>
<td>22</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Former</td>
<td>92</td>
<td>21</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>68</td>
<td>9</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>
Let’s compute the probability of being a current smoker for the three groups:

- For High SES: $\frac{51}{211} \approx 0.24$
- For Middle SES: $\frac{22}{52} \approx 0.42$
- For Low SES: $\frac{43}{93} \approx 0.46$

If the variables were independent, we’d expect these to be equal.

We can do the same for former smokers and those who never smoked.

But the question will still remain: Are the differences due to the fact that we’re only looking at a sample?

To answer this, we use a chi-squared test!
• For the chi-squared test we need to know how to
  – Compute the degrees of freedom to use Table B.4.
  – Compute the expected category counts $e_{ij}$.
• The degrees of freedom is
  \[(\# \text{ rows } - 1) \times (\# \text{ columns } - 1)\]
• For our data this is \((3 - 1) \times (3 - 1) = 4\).
• For a category in the table, the expected count is
  \[
  \frac{(\text{row total}) \times (\text{column total})}{n}
  \]
• This formula comes from the fact that for independent events $A$ and $B$,
  \[
P(A \text{ and } B) = P(A)P(B).
  \]
  Details are in Section 10.2 of the text.
**Original Data with totals:**

<table>
<thead>
<tr>
<th>Smoking</th>
<th>SES</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>High</td>
<td>51</td>
<td>22</td>
<td>43</td>
<td>116</td>
</tr>
<tr>
<td>Former</td>
<td>High</td>
<td>92</td>
<td>21</td>
<td>28</td>
<td>141</td>
</tr>
<tr>
<td>Never</td>
<td>High</td>
<td>68</td>
<td>9</td>
<td>22</td>
<td>99</td>
</tr>
<tr>
<td>Total</td>
<td>High</td>
<td>211</td>
<td>52</td>
<td>93</td>
<td>356</td>
</tr>
</tbody>
</table>

**Expected counts:**

<table>
<thead>
<tr>
<th>Smoking</th>
<th>SES</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>High</td>
<td>68.75</td>
<td>16.95</td>
<td>30.30</td>
</tr>
<tr>
<td>Former</td>
<td>High</td>
<td>83.57</td>
<td>20.60</td>
<td>36.83</td>
</tr>
<tr>
<td>Never</td>
<td>High</td>
<td>58.68</td>
<td>14.46</td>
<td>25.86</td>
</tr>
</tbody>
</table>

- For example, \((116)(211)/356 \approx 68.75\).
- For example, \((141)(93)/356 \approx 36.83\).
Computing the $\chi^2$ statistic
We have to compute $(o_j - e_j)^2/e_j$ for each category, then add:

$$\chi^2 = \frac{(51 - 68.75)^2}{68.75} + \frac{(22 - 16.95)^2}{16.95} + \ldots$$

• For this table, we get $\chi^2 \approx 18.51$.
• Using the 4 degree of freedom row of Table B.4, we find that the p-value is about 0.001.
Example: Are older workers more likely to be laid off than younger workers?

- The null hypothesis states that age is unrelated to being laid off.
- The alternative hypothesis states that the variables are related.

Data:

<table>
<thead>
<tr>
<th></th>
<th>Under 40</th>
<th>Over 40</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laid off</td>
<td>7</td>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td>Retained</td>
<td>504</td>
<td>765</td>
<td>1269</td>
</tr>
<tr>
<td>Total</td>
<td>511</td>
<td>806</td>
<td>1317</td>
</tr>
</tbody>
</table>

Expected counts:

<table>
<thead>
<tr>
<th></th>
<th>Under 40</th>
<th>Over 40</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laid off</td>
<td>18.624</td>
<td>29.376</td>
<td>48</td>
</tr>
<tr>
<td>Retained</td>
<td>492.376</td>
<td>776.624</td>
<td>1269</td>
</tr>
<tr>
<td>Total</td>
<td>511</td>
<td>806</td>
<td>1317</td>
</tr>
</tbody>
</table>
**Test statistic:**

\[
\chi^2 = \frac{(7 - 18.624)^2}{18.624} + \frac{(41 - 29.376)^2}{29.376} \\
+ \frac{(504 - 492.376)^2}{492.376} - \frac{(765 - 776.624)^2}{776.624} \\
\approx 12.303.
\]

**p-value:**

Area to the right of 12.203 from Table B.4 using \((2 - 1)(2 - 1) = 1\) degree of freedom.

This is less than 0.001.

**Conclusion:**

The data strongly support the conclusion that age is related to job status, with older workers more likely to be laid off.
V. Announcements

- We will *not* cover Section 10.3.
- Next week will be devoted to review.
- Next Wednesday we will fill out course evaluations and take a quiz.
- If you have a conflict with the exam (more than two exams that day) let me know (preferably by email) as soon as possible.