Stat 218 - Day 20
Analyzing Paired Data: Paired *t*-test

Example: Marriage ages
A student of mine wanted to test whether husbands tend to be older than their wives on average. He went to the county courthouse and took a sample of 24 couples who had applied for marriage licenses, recording the ages of the man and woman in each case (`MarriageAges.mtw`). Some summary statistics are:

	Sample size	Sample mean	Sample std dev
Husbands	24	35.71	14.56
Wives	24	33.83	13.56

a) Perform a two-sample *t*-test of whether these sample data provide evidence that the population mean age of husbands exceeds that of wives.

b) Use a two-sample *t*-procedure to estimate the difference in population mean ages with a 90% confidence interval.

This two-sample t-procedure is *not valid* because the samples are *not independent*. The age of a wife is related to the age of a husband, because older people tend to marry older people and younger people tend to marry younger people. The above analysis is completely inappropriate here. This analysis only would have been appropriate if the student had gathered ages for one sample of 24 husbands and then independently gathered ages for a different sample of 24 wives.

c) Look at a scatterplot to see the relationship between the ages of a husband and a wife (`MTB> plot c1*c2`).

These data are *paired*, because the observational units are *couples*, not independent individuals. The appropriate analysis is to calculate the *differences* in ages for each couple, and then apply *one*-sample *t*-procedures to those differences.

Some more summary statistics, with differences calculated as husband's age minus wife's age:

	Sample size	Sample mean	Sample std dev
Differences	24	1.875	4.812

d) Determine a 90% confidence interval for the population mean difference in ages between husbands and wives.

A paired *t*-test procedure tests a null hypothesis of H_0: $\mu_d = 0$ and uses a test statistic of $t_s = \frac{\bar{y}_d}{s_d / \sqrt{n}}$ with a *P*-value based on the *t*-distribution with (*n*-1) degrees of freedom.

e) Conduct a paired *t*-test of whether the sample data provide strong evidence that the population mean difference exceeds zero.

f) Explain why the paired analysis produces such a different conclusion than the independent-samples analysis.

g) Was the student wise for gathering paired data rather than independent-samples data to investigate his research question? Explain.

The pairing is effective here because there is a lot of variation in the ages of people who apply for marriage licenses, but there is a strong correlation between the ages of the husband and wife within a couple. Thus, there is much less variation in the differences in ages than there is in the ages themselves.

The technical conditions for the paired *t*-procedures are:

0) that the data come from a matched-pairs design
1) that the data can be regarded as a random sample from the population of differences
2) that the population of differences is normal or the sample size is large

h) Check these technical conditions for the marriage age study.

Paired Design

Note that a paired analysis follows from the data having been collected according to a paired design in the first place.

a) Explain how you could conduct a paired design to gather data for studying whether a baseball player gets from second base to home plate more quickly by taking a wide angle or a narrow angle around third base. Also explain why pairing would be useful in this study.

b) Would it be appropriate to analyze the "cloud seeding" data with a paired *t*-test? Explain.

c) Would it be appropriate to analyze the "bone length of sparrows" data with a paired *t*-test? Explain why it would not even be possible.

Example: Melting morsels (cont.)

Analyze the data from our class experiment. Start with graphical and numerical summaries and then apply the appropriate test and interval. Be sure to check the technical conditions, and write a paragraph summarizing your conclusions (in context).