Stat 218 - Day 18
Transformations, More on Interpreting Tests and Intervals

We have studied *t*-intervals and *t*-tests for comparing means between two groups. The technical conditions required for the *t*-procedures to be valid are that:

1) the data can be regarded as independent random samples
2) the underlying populations follow normal distributions, or the sample sizes are large.

What can we do when this first condition is not met? Conduct a new study with a better design for the data collection phase.

What can we do when this second condition is not met? Two options:

1) Transform the data so that the distribution looks roughly normal on the transformed scale.
2) Conduct a non-parametric Mann-Whitney-Wilcoxon test

Example: Cloud seeding (cont.)
Recall that we examined data on the amount of rain produced by clouds that had been seeded and by clouds that had not been seeded (`CloudSeeding.mtw`).

(a) Do the technical conditions for a two-sample *t*-test appear to be met here? Explain.

(b) Try a log transformation of the rainfall amounts (`let c5 = logt(c3)`). Do the transformed data appear to be more normally distributed than the raw data?

(c) Conduct a two-sample *t*-test on the transformed data. What conclusion would you draw?

Now we will discuss four issues related to interpreting tests and intervals correctly and applying them properly.

- **Relationship between tests and intervals**

  - Hypothesis tests and confidence intervals have different goals but are closely related.
    - A test assesses the degree of evidence against the hypothesis that there is no difference/effect between two groups
    - An interval estimates the magnitude of the difference/effect between two groups
  - When a two-sided test results in rejecting H_0 at significance level α, then a CI with confidence level $100(1-\alpha)$ will not include the value zero.

Example: Fish oil diet (cont.)
We've seen that the test statistic equals 3.06, so the *P*-value for testing H_0: $\mu_F = \mu_R$ vs. H_A: $\mu_F \neq \mu_R$ is about .014.

(a) Based on this *P*-value alone, what can you say about a 95% CI for $\mu_F - \mu_R$? Explain.

(b) Based on this *P*-value alone, what can you say about a 99% CI for $\mu_F - \mu_R$?

(c) Calculate these CI's (Minitab: `Stat> Basic statistics> 2-sample t` with `FishOil.mtw`) to see if your conjecture is correct.

- **Significance vs. importance**
    - A test reveals whether an observed difference is "statistically significant," meaning that it is unlikely to have occurred by chance
        - A statistically significant result may or may not be practically important
        - Especially with large sample sizes, even a very small difference can be statistically significant
    - A CI estimates the magnitude of the difference and so can estimate the practical importance

Example: SAT coaching (hypothetical)
Suppose that 5000 students are randomly assigned to either take an SAT coaching course or not, with the following results in their improvements in SAT scores:

	Sample size	Sample mean	Sample std dev
Coaching group:	2500	46.2	14.4
Control group:	2500	44.4	15.3

(a) Use Minitab (`Stat> Basic statistics> 2-sample t, summarized data`) to conduct a test of whether the sample data provide evidence that SAT coaching is helpful. State the hypotheses, and report the *P*-value. Draw a conclusion in the context of this study.

(b) Use Minitab to produce a 99% CI for the difference in population mean improvements between the two groups. (Note: To produce the CI, Minitab requires that the alternative be set to "not equal.") Interpret this interval.

(c) Do the sample data provide *very* strong evidence that SAT coaching is helpful? Explain whether the *P*-value or the CI helps you to decide.

(d) Do the sample data provide strong evidence that SAT coaching is *very* helpful? Explain whether the *P*-value or the CI helps you to decide.

- **Non-sacredness of common α levels**

While it is common to report whether H_0 is rejected at common α levels such as .05 and ,01, it is much more informative to report the *P*-value. There is not a hard-and-fast cut-off point, only increasing strong evidence against H_0 as the P-value gets smaller and smaller.

- **Type I and Type II Errors**

In any hypothesis testing situation, there are two errors (mistakes) that could be made:

  - Reject H_0 when H_0 is actually true (called Type I error)
    - The significance level α represents the probability of making a Type I error
  - Fail to reject H_0 when H_0 is actually false (called Type II error)

Example: Drug development

Suppose that a pharmaceutical company is testing whether a new drug reduces patients' pain more than a standard drug.

(a) State the appropriate null and alternative hypotheses in symbols.

(b) Describe what the consequences of Type I error would be in this situation.

(c) Describe what the consequences of Type II error would be in this situation.