Goals for the day:

1. Words: $P$-value, statistical significance, Type I and Type II errors
2. R: t.test
3. Big idea: Presenting the results of a hypothesis test

Decisions

There are two possible decisions:

- **reject** the null hypothesis we have enough evidence to say with confidence that the null hypothesis is false
- **do not reject** the null hypothesis we do not have enough evidence to say with confidence that the null hypothesis is false although it very well might be

Test statistics

A statistic is computed from the data. A good test statistic is one that for which more extreme values (in one or both directions) tend to favor the alternate hypothesis.

Examples:

- For testing hypotheses about $\mu$: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ ($\mu = \mu_0$ is the null hypothesis)
- For testing hypotheses about $\sigma$: 

$P$-values

The $P$ value is the number, expressed as a percentage, that we write in the following blank:

If $H_0$ is true, the probability that we would see a value of our test statistic at least as extreme as the one we got in our sample is

or

If $H_0$ is true, the value of the test statistic will be at least as extreme as the one we got in our sample in __________ of all random samples of this size.

We can compute the $P$-value if we know the distribution of the test statistic under the assumption that the null hypothesis is true.

$P$-values measure the strength of evidence against the null hypothesis (in favor of the alternative hypothesis). Smaller $P$-values are stronger evidence.
Statistical significance

We sometimes choose a fixed small number $\alpha$ to compare our $P$-value to. If our $P$-value is less than $\alpha$, we say that our evidence that the null hypothesis is false is statistically significant.

- $\alpha$ is often chosen to be 0.05 or 0.01.
- Never just say “statistically significant” but always say “statistically significant at the $\alpha$ level”. And it’s better to just report the $P$-value so the reader can make her own judgment.
- Never just say “significant.” A statistically significant finding might not be practically significant.

Errors

There are two possible errors that we could make as a result of a hypothesis test:

- Type I error: Reject the null hypothesis when it is true.
- Type II error: Not reject the null hypothesis when it is false.

If our decision rule is to reject $H_0$ whenever the $P$-value is less than $\alpha$, the probability of a type I error is $\alpha$.

The probability of a type II error is called $\beta$ but we cannot compute $\beta$ without knowing the true value of the parameter. (The alternative hypothesis could be true in many different ways.) $1 - \beta$ is called the power of our hypothesis test.

Confidence intervals and hypothesis tests

- Confidence intervals provide more information than a decision on the null hypothesis
- Confidence intervals give a direct measure of the precision of our estimate
- Any confidence interval can be turned into a hypothesis test