section 14: Hypothesis Testing (t-tests)

stuff to learn today:

1. hypothesis testing review
2. p-values
3. effect sizes
4. t-tests: 1 sample, 2 sample
5. type i & type ii errors

Statistical Testing Process

1. Set up hypotheses
2. Pick the statistical test based on your experiment
3. Pick your alpha (level of significance)
4. Calculate your test statistic
5. Find your p-value
6. Interpret

p-value

definition to remember:

a p-value is the **probability** of observing a test statistic at least as large as the one observed **by random chance**, assuming that the null hypothesis is true

effect size

effect size measures the difference between two groups

standardized effect size: Cohen's d

- measures the difference between groups in terms of the number of standard deviations
- 0-1, {small: 0.2, medium: 0.5, large: 0.8}

1-sample t-test

- very similar to our 1-sample z-test!
- information we need:
  - population mean, sample data

Sample question:

> *"Acme Ltd. wants to improve sales performance. Past sales data indicate that the average sale was 100 dollars per transaction. After training the sales force, recent sales data (from a random sample of 25 salesmen) is shown below:"*

```
[122.09, 100.64, 125.77, 120.32, 118.25,
  96.47, 111.4 ,  80.66, 110.77, 111.14,
  102.9, 114.54,  88.09,  98.59,  87.07,
 110.43, 101.9 , 123.89,  97.03, 116.23,
  108.3, 112.82, 119.57, 131.38, 128.39]
```

1-sample t-test

- Hypotheses:
  - $H_0: \mu = 100$
  - $H_A: \mu > 100$

- $\alpha = 0.05$

numbers to note/calculate:

- $\mu = 100$
- $n = 25$
- $x_{bar} = 109.5456$
- $s = 13.3388$ (sample std)
- $df = 24$

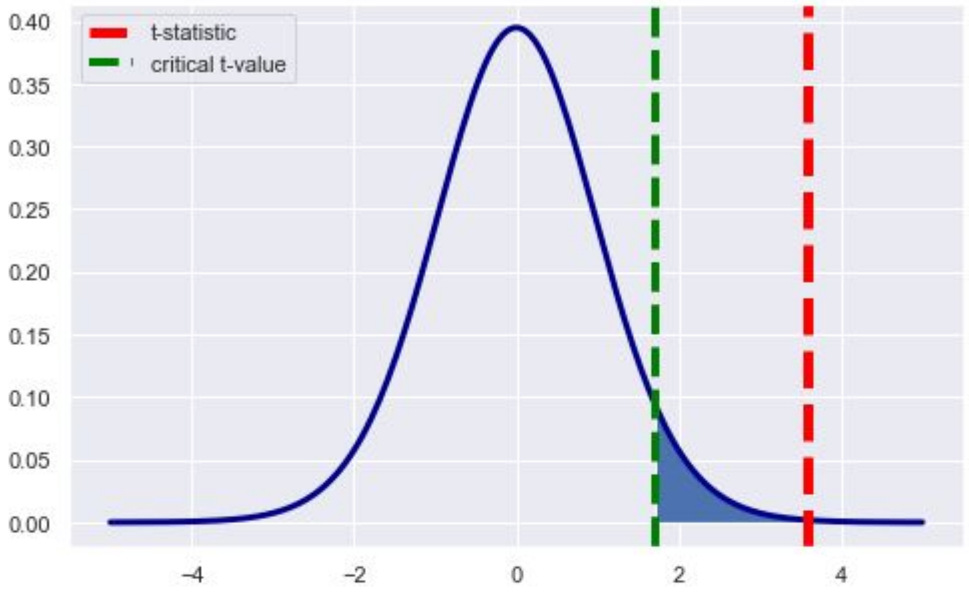
1-sample t-test

t-statistic

- (xbar - μ) / (s/√n)
- 3.5781

t distribution critical values

					Up
df	.25	.20	.15	.10	.05
1	1.000	1.376	1.963	3.078	6.314
2	0.816	1.061	1.386	1.886	2.920
3	0.765	0.978	1.250	1.638	2.353
4	0.741	0.941	1.190	1.533	2.132
5	0.727	0.920	1.156	1.476	2.015
6	0.718	0.906	1.134	1.440	1.943
7	0.711	0.896	1.119	1.415	1.895
8	0.706	0.889	1.108	1.397	1.860
9	0.703	0.883	1.100	1.383	1.833
10	0.700	0.879	1.093	1.372	1.812
11	0.697	0.876	1.088	1.363	1.796
12	0.695	0.873	1.083	1.356	1.782
13	0.694	0.870	1.079	1.350	1.771
14	0.692	0.868	1.076	1.345	1.761
15	0.691	0.866	1.074	1.341	1.753
16	0.690	0.865	1.071	1.337	1.746
17	0.689	0.863	1.069	1.333	1.740
18	0.688	0.862	1.067	1.330	1.734
19	0.688	0.861	1.066	1.328	1.729
20	0.687	0.860	1.064	1.325	1.725
21	0.686	0.859	1.063	1.323	1.721
22	0.686	0.858	1.061	1.321	1.717
23	0.685	0.858	1.060	1.319	1.714
24	0.685	0.857	1.059	1.318	1.711
25	0.684	0.856	1.058	1.316	1.708
26	0.684	0.856	1.058	1.315	1.706



3.57 > 1.711, reject H_0

2-sample t-test

- used to measure if two population means are equal

In the context of controlled experiments, you will often see talk about the "control" group and the "experimental" or "treatment" group. In a drug test example, the control group is the group given the placebo and the treatment group is given the actual drug. Researchers are interested in the average difference in blood pressure levels between the treatment and control groups.

> The 50 subjects in the control group have an average systolic blood pressure of 121.38 who have been given a placebo drug.

> The 50 subjects in the experimental / treatment group have an average systolic blood pressure of 111.56 after treatment with the drug being tested.

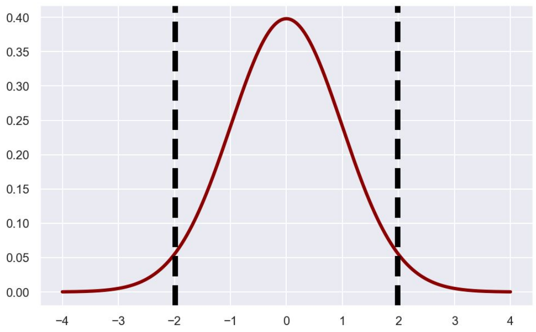
The apparent difference between experimental and control groups is -9.82 points. But with 50 subjects in each group, how confident can a researcher be that this measured difference is real? You can perform a two sample t-test to evaluate this.

2-sample t-test

t table

- Hypotheses:
  - $H_0: \mu_1 - \mu_2 = 0$
  - $H_A: \mu_1 - \mu_2 \neq 0$

- $\alpha = 0.05$ (two-tailed)



numbers to note/calculate:

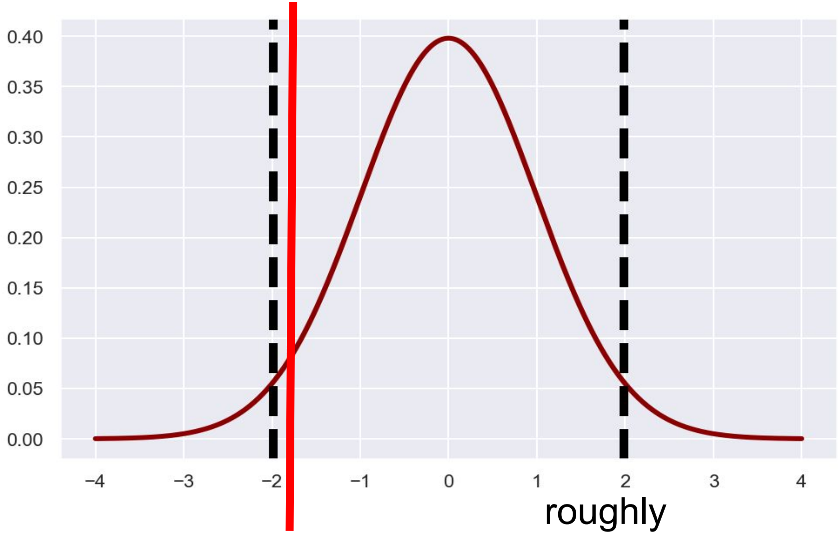
- $\mu_1 = 121.38$, $n_1 = 50$
- $\mu_2 = 111.56$, $n_2 = 50$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- $s_1^2 = 806.0$, $s_2^2 = 541.6$
- pooled sample var (s_p^2) = 673.8
- 2-sample t-statistic = -1.891

2-sample t-test

t table



roughly
±2.0

Conclusion: Fail to reject H_0, the drug does not make a difference in blood pressure levels

or...

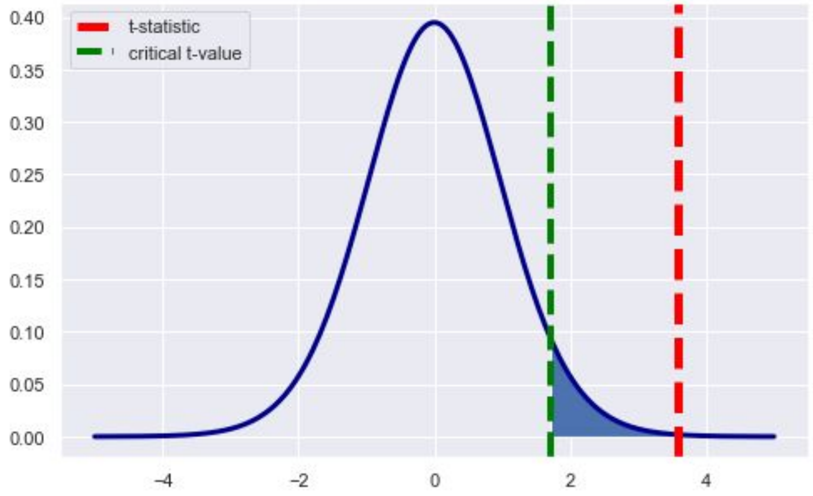
```
stats.ttest_ind(experimental, control)
```

```
Ttest_indResult(statistic=-1.8915462966190273, pvalue=0.061504240672530394)
```

Type I and Type II errors

- in statistical testing, everything is based on a probability or statistical significance
- there are situations where we reject the null hypothesis when we should not have (Type I error) and vice versa (Type II error)

- P(Type I error) = α
- False positive



Type II error / β

- Type II error is failing to reject the null hypothesis when it is actually false - **false negatives**
- P(Type II error) = β
- Power of the test = 1 - β
- Whether you have a higher α or β depends on context