Relationship between a CI and a (two-sided) hypothesis test:

• **A test of H_0: $\mu = m^*$ vs. H_a: $\mu \neq m^*$ will reject H_0 if and only if a corresponding CI for μ does not contain the number m^*.**

$1-\alpha = .95 \Rightarrow \alpha = .05$

Example: A 95% CI for μ is (2.7, 5.5).

(1) At $\alpha = 0.05$, would we reject H_0: $\mu = 3$ in favor of H_a: $\mu \neq 3$? No. Here, 3 is a reasonable value for μ.

(2) At $\alpha = 0.05$, would we reject H_0: $\mu = 2$ in favor of H_a: $\mu \neq 2$? Yes. 2 is not a reasonable value for μ (it falls outside the 95% CI)

(3) At $\alpha = 0.10$, would we reject H_0: $\mu = 2$ in favor of H_a: $\mu \neq 2$? Yes. 2 would certainly also be outside the (narrower) 90% CI for this sample.

(4) At $\alpha = 0.01$, would we reject H_0: $\mu = 3$ in favor of H_a: $\mu \neq 3$? No. 3 would certainly also be inside the (wider) 99% CI for this sample.

Power of a Hypothesis Test

• Recall the significance level α is our desired P(Type I error) = P(Reject H_0 | H_0 true)

The other type of error in hypothesis testing:
Type II error = "Fail to reject H_0 | H_0 false"

P(Type II error) = β = P(Fail to reject H_0 | H_0 false)

The power of a test is P(Reject H_0 | H_0 false)
$= 1 - \beta$

• High power is desirable, but we have little control over it (different from α)

Calculating Power: The power of a test about μ depends on several things: α, n, σ, and the true μ.

Example 1: Suppose we test whether the true mean nicotine contents in a population of cigarettes is greater than 1.5 mg, using $\alpha = 0.01$.

H_0: $\mu = 1.5$ H_a: $\mu > 1.5$

We take a random sample of 36 cigarettes. Suppose we know $\sigma = 0.20$ mg. Our test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 1.5}{0.20/\sqrt{36}}$$

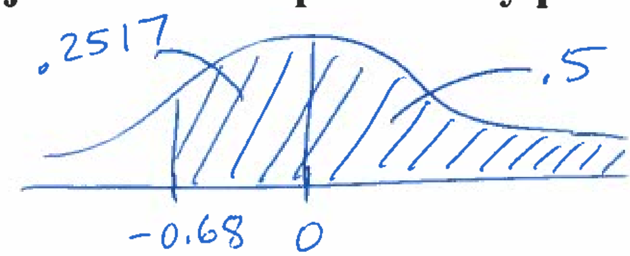
We reject H_0 if: $Z > Z_{.01} = 2.326$

$$\Rightarrow \frac{\bar{X} - 1.5}{0.20/\sqrt{36}} > 2.326 \Rightarrow \bar{X} - 1.5 > 0.0775$$
$$\Rightarrow \bar{X} > 1.5775$$

• Now, suppose μ is actually 1.6 (implying that H_0 is false). Let's calculate the power of our test if $\mu = 1.6$:

$$P(\bar{X} > 1.5775 \mid \mu = 1.6) = P\left(\frac{\bar{X} - 1.6}{0.20/\sqrt{36}} > \frac{1.5775 - 1.6}{0.2/\sqrt{36}}\right)$$
$$= P(Z > -0.68)$$

This is just a normal probability problem!



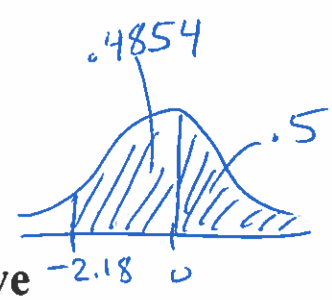
$P(Z > -0.68) = .7517$

$\Rightarrow P(\text{Reject } H_0 \mid \mu = 1.6) = .7517$

$\Rightarrow$ Power when $\mu = 1.6$ is $.7517$

• What if the true mean were 1.65?

Verify: $P(\bar{X} > 1.5775 \mid \mu = 1.65)$
$= P(Z > -2.18) = .9854$



• The farther the true mean is into the "alternative region," the more likely we are to correctly reject H_0.

Example 2: Testing H_0: $p = 0.9$ vs. H_a: $p < 0.9$ at $\alpha = 0.01$ using a sample of size 225.

$$\text{Reject } H_0 \text{ if } \quad Z = \frac{\hat{p} - 0.9}{\sqrt{\frac{(0.9)(0.1)}{225}}} < \overbrace{-2.326}^{z_{.01}}$$

$$\Rightarrow \text{ if } \quad \hat{p} - 0.9 < -0.0465$$

$$\Rightarrow \text{ if } \quad \hat{p} < 0.8535$$

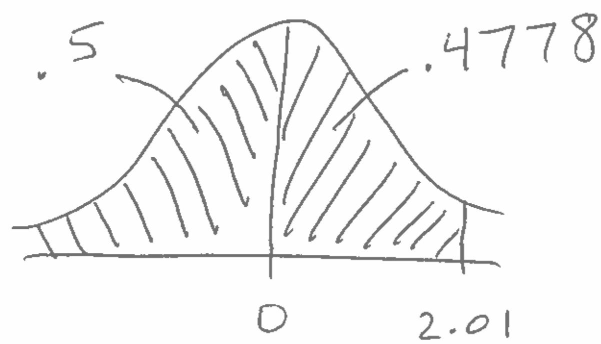
Suppose the true p is 0.8. Then our power is:

$$P(\hat{p} < 0.8535 \mid p = 0.8)$$

$$= P\left(\frac{\hat{p} - 0.8}{\sqrt{\frac{(0.8)(0.2)}{225}}} < \frac{0.8535 - 0.8}{\sqrt{\frac{(0.8)(0.2)}{225}}} \right)$$

$$= P(Z < 2.01)$$

$$= \boxed{.9778}$$



Power is 0.9778 when the true p is 0.8.

STAT 515 -- Chapter 9: Two-Sample Problems

Paired Differences (Section 9.3)

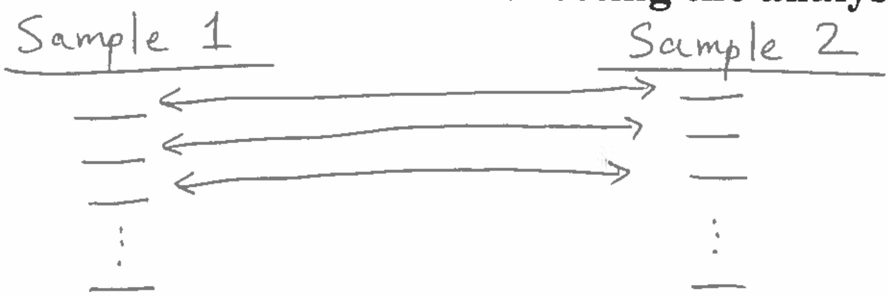
Examples of Paired Differences studies:

- **Similar subjects are paired off and one of two treatments is given to each subject in the pair.**

or

- **We could have two observations on the same subject.**

The key: With paired data, the pairings cannot be switched around without affecting the analysis.



We typically wish to perform inference about the mean of the differences, denoted μ_D.

Example 1: Six students are given two tests, one after being fed, and one on an empty stomach. Is there evidence that students perform better on a full stomach? (Assume normality of data, and use $\alpha = .05$.)

of differences

	Student					
Scores	**1**	**2**	**3**	**4**	**5**	**6**
X_1 (with food)	74	71	82	77	72	81
X_2 (without food)	68	71	86	70	67	80

$\mu_D = \mu_1 - \mu_2$

Calculate differences: D = $X_1 - X_2$

D: 6, 0, −4, 7, 5, 1

$H_0: \mu_D = 0$ vs. $H_a: \mu_D > 0$

$\bar{D} = 2.5$, $S_D = 4.231$

$$t = \frac{\bar{D} - 0}{S_D / \sqrt{n_D}} = \frac{2.5 - 0}{4.231 / \sqrt{6}} = 1.447$$

$(n_D - 1 = 5)$ ← d.f.

Rejection region: $t > t_{.05}$

(From t-table) $t > 2.015$

Since $1.447 \not> 2.015$, we fail to reject Ho. We cannot conclude the students' perform better on a full stomach.

Example 2: Find a 98% CI for the mean difference in arm strength for right-handed people (measured by the number of seconds a certain weight can be held extended).

	Person						
	1	**2**	**3**	**4**	**5**	**6**	**7**
X_1 (Right)	**26**	**35**	**17**	**47**	**22**	**16**	**32**
X_2 (Left)	**20**	**31**	**10**	**38**	**23**	**16**	**29**
D:	6	4	7	9	−1	0	3

↳ $X_1 - X_2$

Assume the population of differences is normally distributed.

$\bar{D} = 4.0 \qquad S_D = 3.65$

98 % CI for μ_D:

$1-\alpha = .98$

$\alpha = .02$

$\alpha/2 = .01$

$t_{.01}$ (6 d.f.) $= 3.143$

$$\bar{D} \pm t_{\alpha/2}\left(\frac{S_D}{\sqrt{n_D}}\right)$$

$$4.0 \pm (3.143)\left(\frac{3.65}{\sqrt{7}}\right)$$

$$\Rightarrow \boxed{(-0.336,\ 8.336)}$$

Interpretation: With 98% confidence, the mean right-arm strength is between 0.336 seconds <u>less</u> and 8.336 seconds <u>greater</u> than the mean left-arm strength. (We are 98% confident the mean difference is between -0.336 and 8.336 seconds.)

Note: With paired data, the two-sample problem really reduces to a one-sample problem on <u>the sample of differences</u>.

<u>Two Independent Samples</u> (Section 9.2)

Sometimes there's no natural pairing between samples.

Example 1: Collect sample of males and sample of females and ask their opinions on whether capital punishment should be legal.

Example 2: Collect sample of iron pans and sample of copper pans and measure their resiliency at high temperatures.

No attempt made to pair subjects – we have two <u>independent</u> samples.

We could rearrange the order of the data and it wouldn't affect the analysis at all.