

Lab 8. Two Sample Tests and Power

One Sample Tests for the Mean		
	Independent Data	Dependent/Paired Data
H0	$\mu = \mu_0$	$\mu_D = 0$
Test	One-sample t-test	Paired t-test
Test Statistic	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$	$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$
Distribution of Test Statistic	t distribution with $n - 1$ d.o.f.	t distribution with $n-1$ d.o.f.

Multi-sample Tests for the Mean of Normally Distributed Variables			
	Two sample		K samples
	Equal Variance	Unequal Variance	Coming soon ☺
H0	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 = 0$	
Test	Two-sample t-test with equal variances	Two-sample t-test with unequal variances	
Test Statistic	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 [(1/n_1) + (1/n_2)]}}$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{[(s_1^2/n_1) + (s_2^2/n_2)]}}$	
Distribution of Test Statistic	t distribution with $n_1 + n_2 - 2$ d.o.f.	t distribution with ν d.o.f.	

DO NOT MEMORIZE THE FOLLOWING:

- Equation for pooled variance: $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
- Equation for degrees of freedom: $\nu = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)]}$

1. In the general population, the mean gestational age is 40 weeks. Perform a one-sample t-test to see if the mean gestational age for the population from which this sample was drawn is the same as that of the general population.

- a. What are your null and alternative hypotheses?

$$H_0: \mu = 40$$

$$H_A: \mu \neq 40$$

- b. What is your p-value?

`ttest gestage=40`

$p < 0.0001$

- c. What do you conclude?

We reject the null hypothesis. There is sufficient statistical evidence to suggest the mean gestational age in this sample is not equal to 40 weeks. This sample had an average gestational age of 28.9 weeks (95% CI: 28.4 to 29.4 weeks).

- d. Construct a 95% confidence interval for the true mean gestational age of this population.

True population means between 28.4 and 29.4 weeks are consistent with our observed sample.

2. Now we will compare the birth weights of babies whose mothers had toxemia to those babies whose mothers did not.

- a. What is the sample mean birth weight of toxemia babies and non-toxemia babies in this dataset? Do they appear similar?

`ttest birthwt, by(toxemia) unequal`

1097g for those whose mothers did not have toxemia and 1105g for those whose mothers did have toxemia.

- b. What is the correct statistical test to answer whether babies whose mothers had toxemia had the same gestational age of babies whose mothers did not?

Unpaired t-test with unequal variance

- c. How did you decide between a t-test with equal variance and a t-test with equal variance?

While there are formal tests to conclude whether variances are equal, these tests can behave quite poorly when the sample size is small. It is always safest just to assume the variances are not equal when testing two means.

- d. What are your null and alternative hypotheses?

$$H_0: \mu_{\text{tox}} - \mu_{\text{not tox}} = 0$$

$$H_A: \mu_{\text{tox}} - \mu_{\text{not tox}} \neq 0$$

- e. Perform the test. What is your p-value?

$p = 0.912$

- f. What do you conclude?

We fail to reject the null hypothesis. There is not sufficient statistical evidence to suggest the mean birth weight of babies whose mothers had toxemia is different from the birth weight of babies whose mother did not have toxemia. The difference between the average birth weight in these two groups was -7.8g (95% CI: -151 to 135g) with the babies of toxemic mothers having slightly higher birth weight.

- g. How is the confidence interval related to the test?

Because the interval includes 0, which is the value of our null hypothesis, we will not be able to reject the null hypothesis.

3. Suppose we were interested in whether a baby's length is the same as their head circumference.

- a. What is the sample mean head circumference and length? Do they appear similar?

`ttest headcirc=length`

The mean head circumference is 26.5cm and the mean length is 36.8 cm.

- b. What is the correct statistical test to answer whether babies whose mothers had toxemia had the same gestational age of babies whose mothers did not?

paired t-test

- c. What are your null and alternative hypotheses?

$H_0: \mu_{\text{head}} - \mu_{\text{length}} = 0$

$H_A: \mu_{\text{head}} - \mu_{\text{length}} \neq 0$

- d. Perform the test. What is your p-value?

$P < 0.0001$

- e. What do you conclude?

We reject the null hypothesis. There is sufficient statistical evidence to suggest babies' head circumference is shorter than their length. The average difference between head circumference and length was -10.4cm (95% CI: -10.9 to -9.9cm).

- f. What conclusion could you draw about a difference between head circumference and length of -10 cm? Also state the hypothesis.

$H_0: \mu_{\text{head}} - \mu_{\text{length}} = -10$

$H_A: \mu_{\text{head}} - \mu_{\text{length}} \neq -10$

We would fail to reject the null hypothesis. Our data are consistent with a null hypothesis of the difference being -10cm.

Power

Suppose you were interested in studying the births of toxemic mothers more closely. Suppose the dataset used in the previous questions was from a pilot study.

1. How many patients do you need to have in your study to test the null hypothesis that the mean gestational age for babies born to toxemic mothers is 30 weeks with 80% power? Use the pilot data to estimate the key quantities needed in a power calculation. These estimates include the estimated mean and standard deviation.

```
ttest gestage, by(toxemia) unequal
```

28.4 weeks for those whose mothers did not have toxemia and 30.9 weeks for those whose mothers did have toxemia.

sd=2.3 for babies whose mothers had toxemia

```
sampsi 30 30.9, sd(2.3) onesam power(0.8)
```

n=52

2. Suppose there is a new toxemia treatment you are testing. Your study will have two groups, one treated and one not treated. The treatment will be deemed successful if it adds 3 weeks to gestational age. Also, the standard deviation in the treatment group is expected to be twice that of the untreated group. How many patients will you need to detect a 3 week difference with 80% power?

Again, use the pilot data to estimate the key quantities needed in a power calculation. These estimates include the estimated mean and standard deviations.

```
ttest gestage, by(toxemia) unequal
```

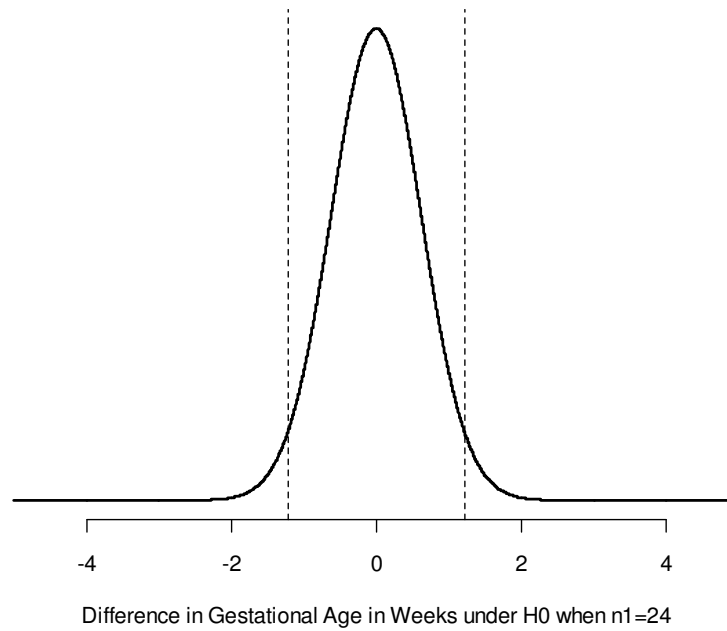
```
sampsi 33.9 30.9, sd1(4.6) sd2(2.3) power(0.8)
```

n1=n2=24, or 48 total

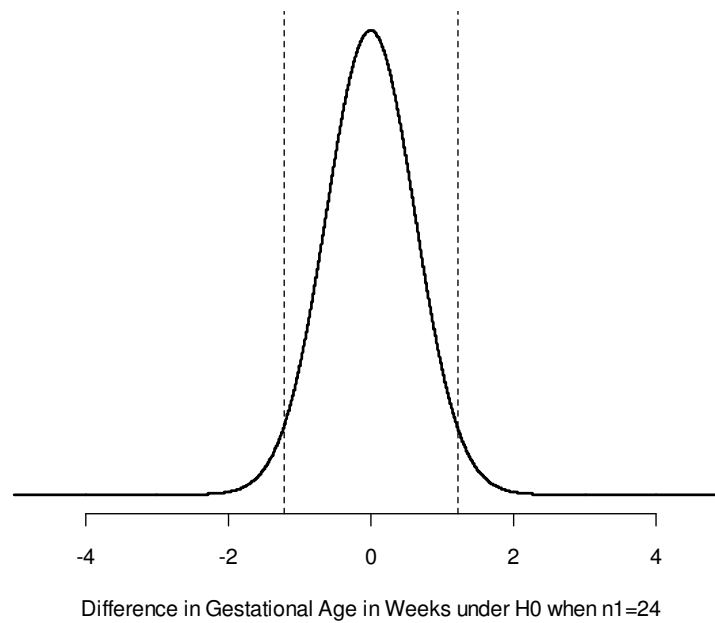
3. How does n change as the clinically meaningful difference, sd and power changes? What change has the most dramatic effect on n ?

Clinically Meaningful Difference	Power	sd	N1
4 weeks	0.8	2 times	13
2 weeks	0.8	2 times	52
1 week	0.8	2 times	208
4 weeks	0.9	2 times	18
2 weeks	0.9	2 times	70
1 week	0.9	2 times	278
4 weeks	0.8	equal	6
2 weeks	0.8	equal	21
1 week	0.8	equal	84
4 weeks	0.9	equal	7
2 weeks	0.9	equal	28
1 week	0.9	equal	112

4. The figure below gives the null distribution of no difference between the treated and the untreated group when sd is 2.3 in both groups and the total n is 48. Draw a picture of the alternative distribution when the observed mean difference is 3 weeks. Shade the region that corresponds to the alpha level of 0.05. The dashed lines are at $1.96 \cdot se$ of the difference.



5. The figure below gives the null distribution of no difference between the treated and the untreated group. Draw a picture of the alternative distribution when the observed mean difference is 3 weeks. Shade the region that corresponds to power of 0.80.



6. How will the picture change if $n_1=100$? What happens to the power?

