Sections 7.5, 7.7, and 7.10

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

One-sided tests

- Often in scientific studies our hypothesis has a *direction* (larger or smaller) attached to it.

- If we come up with a new sleep aid, we may want to show that it works better than Lunesta, i.e. $H_A : \mu_1 > \mu_2$ where μ_1 is mean hours of sleep using our new drug and μ_2 is mean hours of sleep using Lunesta.

- We might want to show that eating fiber lowers cholesterol, i.e. $H_A : \mu_1 < \mu_2$ where μ_1 is mean cholesterol from those who eat at least 50 grams of fiber a day and μ_2 is mean cholesterol from those who eat under 50 grams per day.

One-sided tests

- In these examples, the alternative is $H_A : \mu_1 > \mu_2$ or $H_A : \mu_1 < \mu_2$.
- These are called *one-sided* (or one-tailed) t-tests.
- The null is still the same $H_0 : \mu_1 = \mu_2$, and so is the test statistic $t_s = \frac{\bar{Y}_1 - \bar{Y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$.
- The P-value is different though, because we're only looking in one direction $H_A : \mu_1 < \mu_2$ or $H_A : \mu_1 > \mu_2$ instead of both directions $H_A : \mu_1 \neq \mu_2$.

P-value for one-sided tests

- The P-value is the probability of seeing a sample mean difference $\bar{Y}_1 - \bar{Y}_2$ *further* in the direction of the alternative than the one we observe $\bar{y}_1 - \bar{y}_2$ if $H_0 : \mu_1 = \mu_2$ is true.

- For $H_A : \mu_1 < \mu_2$

  P-value $= \Pr\{\bar{Y}_1 - \bar{Y}_2 < \bar{y}_1 - \bar{y}_2 | H_0 \text{ true}\} = \Pr\{T_{df} < t_s\}$

  where T_{df} has Student's t distribution with df from the Welch-Satterthwaite formula.

- For $H_A : \mu_1 > \mu_2$

  P-value $= \Pr\{\bar{Y}_1 - \bar{Y}_2 > \bar{y}_1 - \bar{y}_2 | H_0 \text{ true}\} = \Pr\{T_{df} > t_s\}$.

- Use t.test(sample1,sample2,alternative="less") or t.test(sample1,sample2,alternative="greater") for one-sided alternatives.
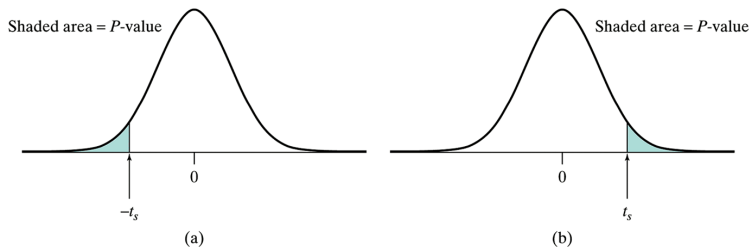
P-value for one-sided tests

Figure 7.5.1 One-tailed P-value for a t test, (a) if the alternative is H_A: $\mu_1 < \mu_2$ and t_s is negative; (b) if the alternative is H_A: $\mu_1 > \mu_2$ and t_s is positive

Example 6.7.1 Height of control & ancymidol plants

Ancymidol is hypothesized to slow growth. $n_1 = 8$ control (no ancymidol) and $n_2 = 7$ plants treated with ancymidol were measured (cm) after two weeks. Researchers want to show that ancymidol slows growth.

Table 6.7.1 Fourteen-day height of control and of ancy plants (cm)		
	Control (Group 1)	Ancy (Group 2)
	10.0	13.2
	13.2	19.5
	19.8	11.0
	19.3	5.8
	21.2	12.8
	13.9	7.1
	20.3	7.7
	9.6	
n	8	7
$\bar{y}$	15.9	11.0
s	4.8	4.7
SE	1.7	1.8

Example 6.7.1 Height of control & ancymidol plants

Let μ_1 be the mean two-week height of treatment (ancymidol) plants and μ_2 be the mean two-week height of controls. We want to test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 < \mu_2$.

```
> control=c(10.0,13.2,19.8,19.3,21.2,13.9,20.3,9.6)
> ancy=c(13.2,19.5,11.0,5.8,12.8,7.1,7.7)
> t.test(ancy,control,alternative="less")

        Welch Two Sample t-test

data:  ancy and control
t = -1.9939, df = 12.783, p-value = 0.03397
alternative hypothesis: true difference in means is less than 0
```

Since P-value < 0.05 we reject $H_0 : \mu_1 = \mu_2$ in favor of $H_A : \mu_1 < \mu_2$ at the $\alpha = 0.05$ level.

Sample size determination

Important question | How much data should I collect?

- The power of a test is the probability of rejecting

  $$\text{power} = \Pr\{\text{reject } H_0 | H_A \text{ true}\} = 1 - \beta.$$

- The larger the sample sizes n_1 and n_2, the higher the power.
- We can find sample sizes n_1 & n_2 to reach a fixed power, usually 80% or 90%.
- To make things easier, let's assume equal sample sizes in each group, $n_1 = n_2$.

Sample size determination

- We need estimates of μ_1, μ_2 and σ; the standard deviation is assumed to be the same in both populations $\sigma_1 = \sigma_2 = \sigma$.
- R has a sample size calculator that uses $\delta = |\mu_1 - \mu_2|$ and σ.
- The R command looks like
  power.t.test(delta=$|\mu_1 - \mu_2|$,sd=σ,sig.level=0.05,power=0.90, type="two.sample",alternative="two.sided")
- You can put different power levels, e.g. power=0.80 is common. You can also put in difference significance levels besides $\alpha = 0.05$.
- Need to indicate whether performing a two-sided test $H_A : \mu_1 \neq \mu_2$ or a one-sided test $H_A : \mu_1 < \mu_2$ or $H_A : \mu_1 > \mu_2$. In R, `alternative="two.sided"` or `alternative="one.sided"`.
- R returns a decimal for $n_1 = n_2$; round up to the next integer.

Example 7.7.3 Postpartum weight loss

- Want to test whether internet-based program helps women lose weight after giving birth.
- One group (treatment) women to be enrolled in internet weight loss program providing weekly exercise and dietary guidance, tracking progress, includes online forum for nutrition and exercise discussion with other recent mothers.
- Other group (control) given traditional written dietary and exercise guidelines by their doctors.

Example 7.7.3 Postpartum weight loss

- Response is weight lost at 12 months postpartum (kg).
- Let μ_1 be weight lost for controls and μ_2 weight lost for treatment, after 12 months postpartum.
- Previous studies show that at 12 months postpartum, mean weight loss is about 3.6 kg with a standard deviation of 4.0 kg, so $\mu_1 = 3.6$ kg and $\sigma = 4$ kg.
- Research team wants to show 50% improvement in weight loss for the treatment group, a 1.8 kg difference from controls. So $\mu_2 = 5.4$ kg.
- One-sided t-test at 5% significance level; want power $= 80\%$.

R output

We have $|\mu_1 - \mu_2| = 1.8$ kg, $\sigma = 4$ kg, $\alpha = 0.05$, power $= 0.8$.

```
> power.t.test(delta=1.8,sd=4,sig.level=0.05,power=0.80,
+ type="two.sample",alternative="one.sided")

     Two-sample t test power calculation

              n = 61.74997
          delta = 1.8
             sd = 4
      sig.level = 0.05
          power = 0.8
    alternative = one.sided

 NOTE: n is number in *each* group
```

Need $n_1 = n_2 = 62$ in each group to have an 80% chance of rejecting $H_0 : \mu_1 = \mu_2$ in favor of $H_A : \mu_1 > \mu_2$ where μ_1 is mean of control and μ_2 is mean of treatment.

Wilcoxon-Mann-Whitney test

- Often data are *not* normal.
- Example 7.8.1: Researchers took skin samples from $n_1 = 10$ patients who had breast implants and from a control group of $n_2 = 6$ patients & recorded level of interleukin-6 (in pg/ml/10 g of tissue), a measure of tissue inflammation.
- Data are highly skewed.
- One approach to handling data like this is the *permutation test* described in Section 7.1.
- Now we'll look at another approach called the Wilcoxon-Mann-Whitney test.

```
implant=c(231,308287,33291,124550,17075,22955,95102,5649,840585,58924)
control=c(35324,12457,8276,44,278,840)
```

Interleukin-6 data
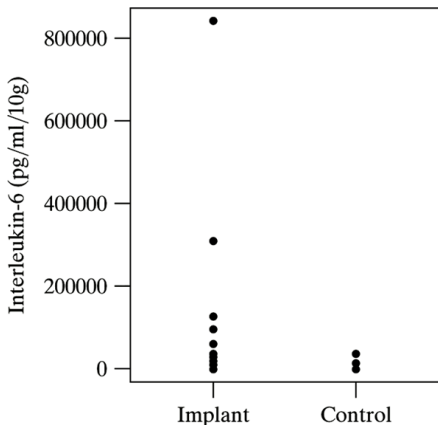
Side-by-side dotplots of data. Let's look at the normal probability plots in R...

Wilcoxon-Mann-Whitney test

- Denote the observations in the two samples by Y_1 and Y_2.
- Null and alternative hypotheses of Wilcoxon-Mann-Whitney test are

  H_0 : The population distributions of Y_1 and Y_2 are the same

  H_A : Y_1 tends to be either greater or less than Y_2

- One or two-sided alternatives are possible.
  - H_A : Y_1 tends to be less than Y_2
  - H_A : Y_1 tends to be greater than Y_2
- Let's see how to build the test statistic.

Wilcoxon-Mann-Whitney test statistic U_s

1. Arrange the observations in increasing order.

2. For each observation in sample 1, we count the number of observations in sample 2 that are smaller in value (that is, to the left); count $1/2$ for each tied observation. Add them up to get K_1.

3. For each observation in sample 2, we count the number of observations in sample 1 that are smaller in value (that is, to the left); count $1/2$ for each tied observation. Add them up to get K_2.

4. Check that $K_1 + K_2 = n_1 n_2$.

5. U_s is the larger of K_1 and K_2.

If H_0 is true then $U_s \approx \frac{n_1 n_2}{2}$; U_s near $n_1 n_2$ gives evidence to H_A. R computes the statistic and a P-value for us.

Interleukin-6 data

```
> implant=c(231,308287,33291,124550,17075,22955,95102,5649,840585,58924)
> control=c(35324,12457,8276,44,278,840)
> wilcox.test(implant,control)

        Wilcoxon rank sum test

data:  implant and control
W = 49, p-value = 0.04196
alternative hypothesis: true location shift is not equal to 0
```

What do we conclude?

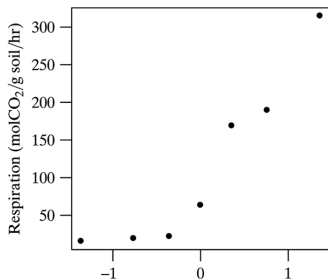## Example 7.10.1 Soil respiration

- Soil respiration affects plant growth.
- Soil cores taken from two locations in forest: (1) under opening in forest canopy (gap) and (2) under heavy tree growth (growth). Amount of carbon dioxide given off by each soil core was measured (mol CO_2/g soil/hr).
- H_0: The gap and growth areas do not differ with respect to soil respiration.
- Two-sided H_A: The distribution of soil respiration rates tends to be higher in one of the two populations.
- One-sided H_A: Soil respiration rates tend to be greater in growth vs. gap areas.
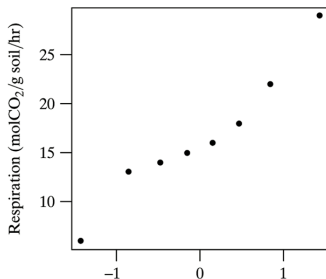
Soil respiration
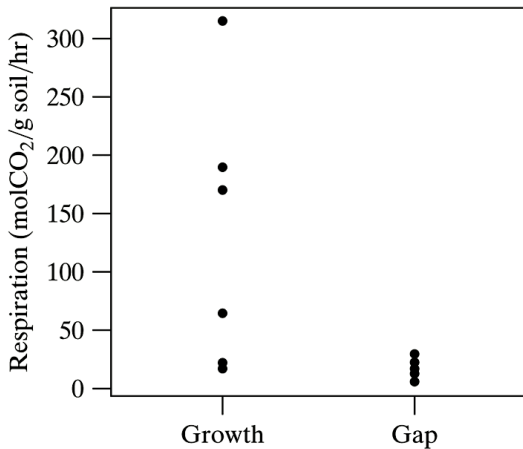
Side-by-side dotplots of data.

Soil respiration

(a) Normal scores

(b) Normal scores

Normality questionable for growth data (left).

Soil respiration

Number of gap observations that are smaller	Y_1 Growth data	Y_2 Gap data	Number of growth observations that are smaller
5	17	6	0
6	20	13	0
6.5	22	14	0
8	64	15	0
8	170	16	0
8	190	18	1
8	315	22	2.5
		29	3
$K_1 = 49.5$			$K_2 = 6.5$

Table 7.10.2 Wilcoxon-Mann-Whitney calculations for Example 7.10.2

$U_s = 49.5$ which is close to $n_1 n_2 = 56$ and far away from $n_1 n_2 / 2 = 28$. The P-value should be small...

Soil respiration, R code

```
> growth=c(17,20,22,64,170,190,315)
> gap=c(6,13,14,15,16,18,22,29)
> wilcox.test(growth,gap)

        Wilcoxon rank sum test with continuity correction

data:  growth and gap
W = 49.5, p-value = 0.015
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(growth, gap) :
  cannot compute exact p-value with ties
```

Ignore the warning. P-value is 0.015; at the $\alpha = 0.05$ level reject H_0 : carbon dioxide distributions are the same at growth and gap areas of the forest.

Ancymidol data with WMW test

Want to test:

H_0 : there is no difference between ancymidol and control plants
vs. H_A : ancymidol plants tend to be shorter than control plants.

```
> control=c(10.0,13.2,19.8,19.3,21.2,13.9,20.3,9.6)
> ancy=c(13.2,19.5,11.0,5.8,12.8,7.1,7.7)
> wilcox.test(ancy,control,alternative="less")

        Wilcoxon rank sum test with continuity correction

data:  ancy and control
W = 11.5, p-value = 0.03192
alternative hypothesis: true location shift is less than 0
```

Since P-value $= 0.032 < 0.05$, at $\alpha = 0.05$ we reject H_0 : *there is
no difference between ancymidol and control plants* and accept
H_A : *ancymidol plants tend to be shorter than control plants*.