# Stat 411/511

## Sign Test & Wilcoxon Signed Rank Test

Feb 10 2012

Charlotte Wickham

stat511.cwick.co.nz

| Levene's test | Welch's t-test |
| --- | --- |
| The difference in populations standard deviations is zero. | The difference in population means is zero.<br>OR<br>The treatment effect is zero. |
| •Normal populations of deviations<br>•Equal population standard deviations of deviations.<br>•Independence of subjects within and between groups. | •Normal populations<br>•Independence of subjects within and between groups. |
| Sleuth says it is robust. | •Robust to non-Normal populations with large samples. |
| Not resistant | Not resistant |
| two sample t-statistic on<br>$Z_1 = (Y_1 - \bar{Y}_1)^2$ &<br>$Z_2 = (Y_2 - \bar{Y}_2)^2$ | $( (\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1) ) /$<br>$SE_{\bar{Y}2 - \bar{Y}1}$<br>with different SE to two-sample t-test |

# Back to paired data

That finishes our two **independent group tests.**

We will now talk about two resistant **paired sample tests.**

**Either** a two independent sample test is appropriate **or** a paired sample test is appropriate, **never both**.

But, within the two independent sample tests, more than one test may be appropriate, same with paired sample tests.

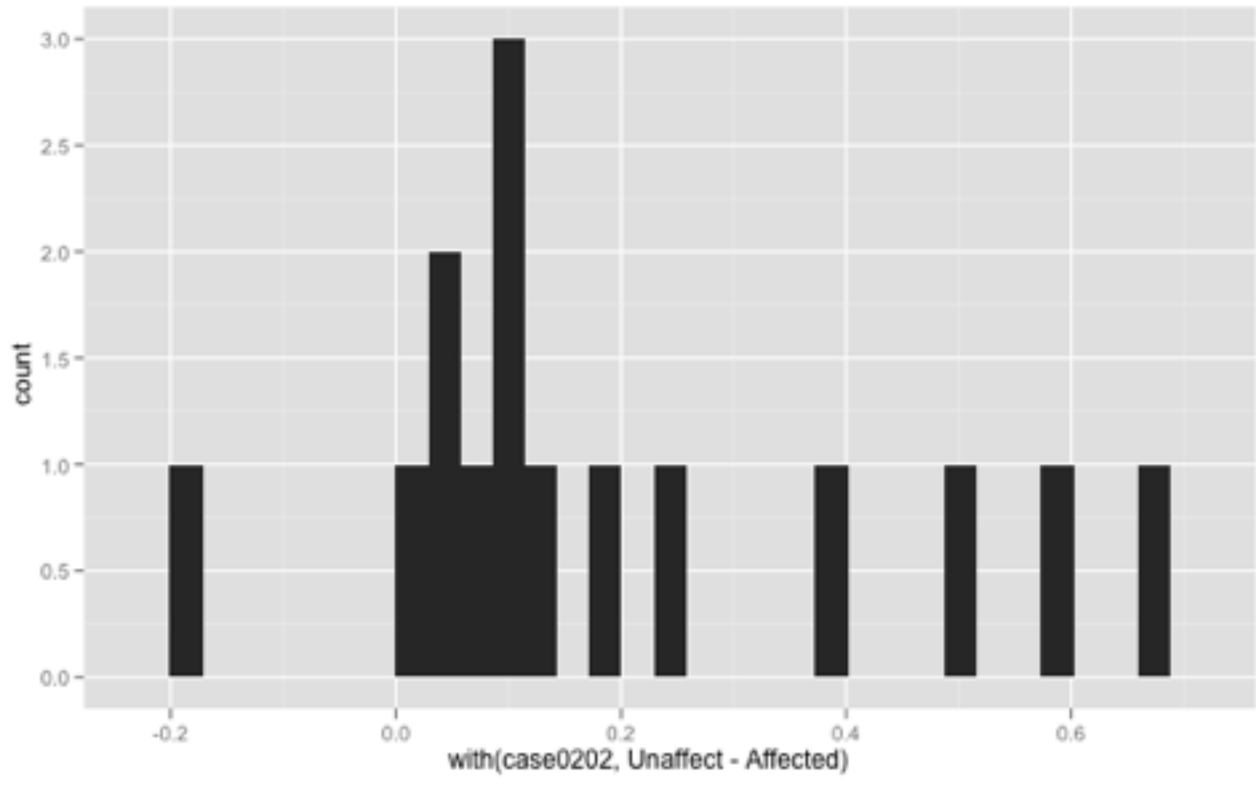
More about choosing a test on Monday

# Schizophrenia Case Study

One sample of differences.

The difference in brain volume between a non-schizophrenic and their schizophrenic twin.

```
with(case0202,
  Unaffect - Affected)

0.67 -0.19  0.09  0.19
0.13  0.40  0.04  0.10
0.50  0.07  0.23  0.59
0.02  0.03  0.11
```



# Sign Test

# Your turn

Imagine we just have one pair of twins.

If schizophrenia has no relationship to brain volume, what is the probability the difference (schizophrenic volume - non-schizophrenic volume) is positive?

If we have two sets of twins, what is the probability both differences are positive?

# Sign test

**Null:** Median difference is zero. (Or no treatment effect). (Sleuth says mean, it is wrong)

**Test statistic:** the number of positive differences.

Under the null hypothesis, seeing positive differences should be like seeing heads in a fair coin flip.

# Schizophrenia Case Study

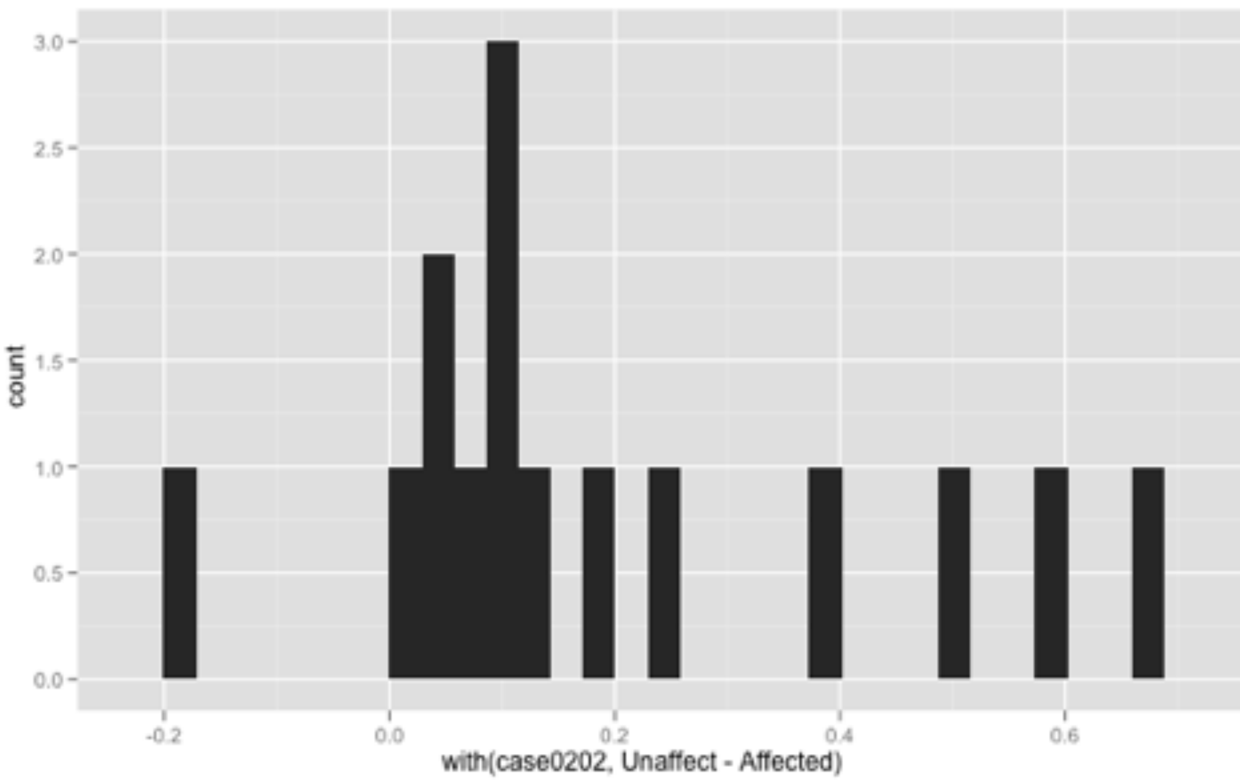
**Null:** Median difference in brain volume between a non-schizophrenic and their schizophrenic twin is zero.

Under the null, the probability of seeing 14 positive differences is the same as the probability of seeing 14 heads in 15 fair coin flips. the p-value

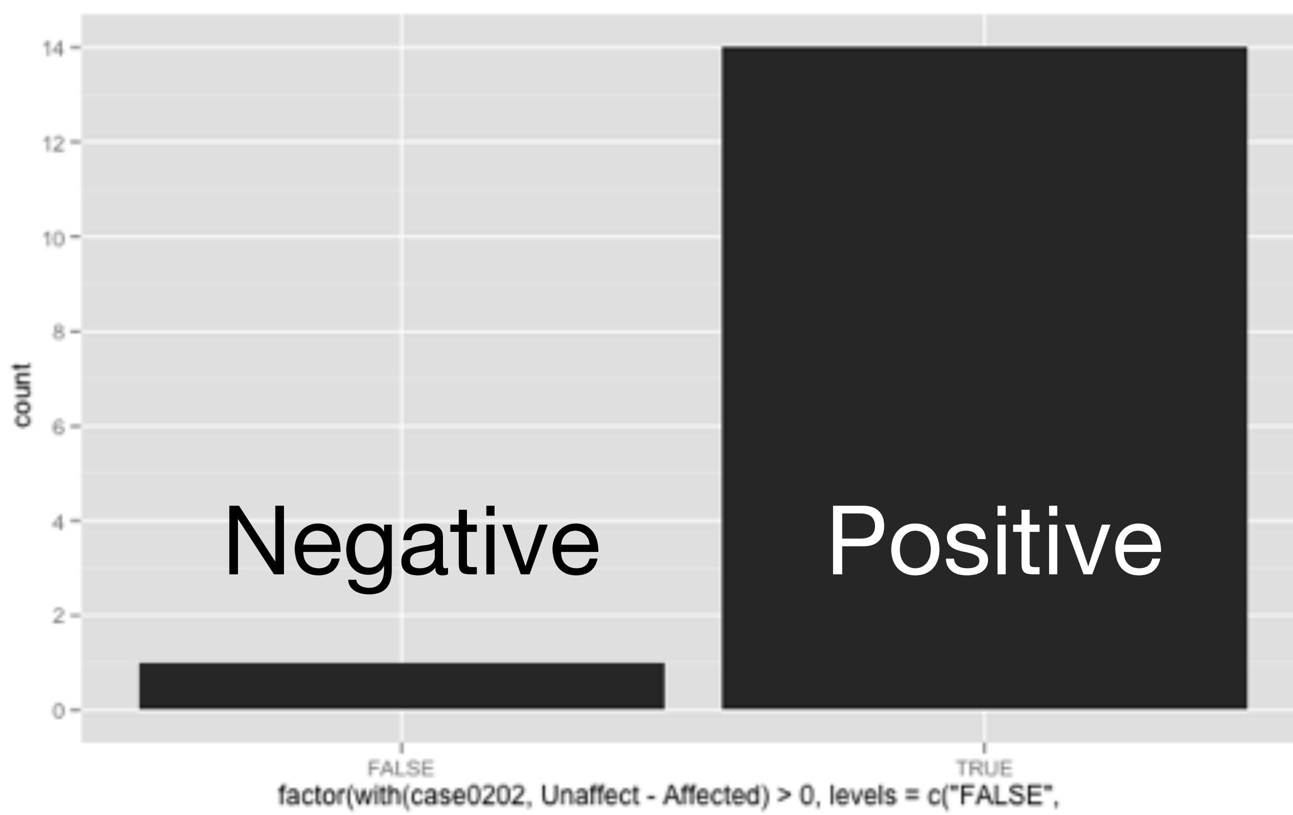
```
> binom.test(14, 15)

   Exact binomial test

data:  14 and 15
number of successes = 14, number of trials = 15,
p-value = 0.0009766
alternative hypothesis: true probability of
success is not equal to 0.5
95 percent confidence interval:
 0.6805154 0.9983136
sample estimates:
probability of success
             0.9333333
```

Sleuth does a Normal approximation
(useful but you don't need to know it)

# Takes this:




# And simplifies to this:




No Normality assumption

Very resistant to outliers

But throws a lot of information away

# Your turn

# Wilcoxon Signed Rank Test

Uses ranks and signs

# Wilcoxon Signed Rank Test

**Null:** Median difference is zero. (Or no treatment effect).

**Test statistic:** the sum of the ranks of the **positive** differences.

| | | |
|---|---|---|
| 0.67 | | 0.02 |
| -0.19 | | 0.03 |
| 0.09 | | 0.04 |
| 0.19 | | 0.07 |
| 0.13 | | 0.09 |
| 0.40 | | 0.10 |
| 0.04 | | 0.11 |
| 0.10 | ⟶ | 0.13 |
| 0.50 | | -0.19 |
| 0.07 | | 0.19 |
| 0.23 | | 0.23 |
| 0.59 | | 0.40 |
| 0.02 | | 0.50 |
| 0.03 | | 0.59 |
| 0.11 | | 0.67 |

# Step 1:

Using absolute values of the differences order the differences from smallest to largest.

0.02
0.03
0.04
0.07
0.09
0.10
0.11
0.13
-0.19
0.19
0.23
0.40
0.50
0.59
0.67

# Your turn

## Step 1.5:

Drop any zeros from the list.

## Step 2:

Rank the absolute values from smallest to largest (just like in Wilcoxon rank sum test, gives ties the average rank).

| Diffs | Rank |
|---|---|
| 0.02 | 1.0 |
| 0.03 | 2.0 |
| 0.04 | 3.0 |
| 0.07 | 4.0 |
| 0.09 | 5.0 |
| 0.10 | 6.0 |
| 0.11 | 7.0 |
| 0.13 | 8.0 |
| -0.19 | 9.5 |
| 0.19 | 9.5 |
| 0.23 | 11.0 |
| 0.40 | 12.0 |
| 0.50 | 13.0 |
| 0.59 | 14.0 |
| 0.67 | 15.0 |

# Step 3:

Add up the ranks of the positive differences.

= 110.5

**The Wilcoxon Signed Rank test-statistic**

Under the null, we expect the test statistic to be about:
$n(n+1)/4 = 15*14 /4 = 52.5$

```
wst <- wilcoxsign_test(Unaffect ~ Affected,
  data = case0202, alternative = "greater",
  distribution = exact())

> wst

   Exact Wilcoxon-Signed-Rank Test

data:  y by x (neg, pos)
    stratified by block
Z = 2.8966, p-value = 0.001007
alternative hypothesis: true mu is greater than 0

> statistic(wst, "linear")

neg 111
```

Exact p-value

Weird rounding issue with this data

# Your turn

Using the worksheet with test summaries, think about how you might choose a test.

What is the largest set of assumptions you need to check?

How do you proceed if any of them are violated?

How do outliers affect your choice?

How does sample size affect your choice?