Stat 411/511

Wilcoxon Rank Sum Test

Feb 6 2012

Charlotte Wickham stat511.cwick.co.nz

Midterm

Practice midterm posted online.

TRUE/FALSE, some multiple choice and a few short answer (one sentence) questions. 50 minutes long, 50 pts total.

Everything we cover up to the end of this week could be tested.

Review the conceptual exercises I've assigned.

Check out study guides also posted (Chap 4 at the end of this week).

Two independent samples

Two sample t-test

Randomization test

Wilcoxon Rank Sum

Today

Doesn't assume Normality and is resistant to outliers

Levene's test

test for equal population standard deviations

Wednesday

Welch's t-test

t-test without assumption of equal standard deviations

Two paired samples

Paired t-test

Sign test

Doesn't assume Normality and resistant to outliers, quick

Wilcoxon Signed Rank test

Doesn't assume Normality and resistant to outliers, more efficient

Friday

Wilcoxon Rank Sum

aka Rank Sum Test

aka Mann Whitney Test

Ranks

Instead of using the raw data, we convert to ranks.

Ranks are **resistant** to outliers since an outlying value will only ever be 1 unit away from next value.

Removes information about shape (no Normality assumption)

The **test statistic, T,** for the Wilcoxon **Rank Sum** test is the **sum of the ranks in the smaller group.**

Null hypothesis

Assuming the two groups have the same shape (same standard deviation).

Null: The population **medians** are equal (or the difference in population medians is zero)

OR

Null: The treatment effect is zero

If the two populations don't have the same shape, the test is still valid but it tests a different null hypothesis,

$P(Y_2 > Y_1) = P(Y_1 > Y_2)$

Under the Null

The test statistic, T, (the sum of the ranks in the smallest group) will:

- have mean $n_1\overline{R}$
- have standard deviation
  s_R sqrt($(n_1 n_2)/(n_1 + n_2)$)

where $\overline{R}$ = sample average of the ranks, and s_R = sample standard deviation of the ranks.

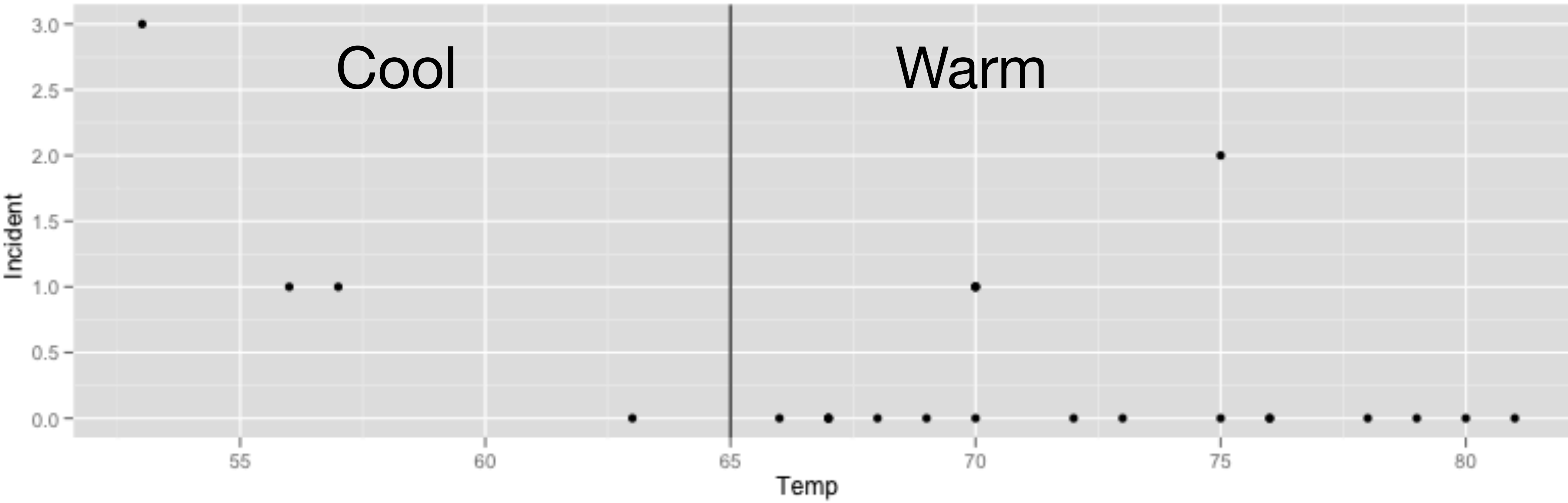
To get a p-value:

Either use a Normal to approximate the sampling distribution of T. Approximate p-value.

Or calculate the exact distribution of T. Exact p-value. Do this for small samples or lots of ties.

You are not responsible for knowing how to go from the Wilcoxon Rank Sum test-statistic to the p-value.

Space shuttle O-ring failures



Is there a higher risk of O-ring failure at lower temperatures?

Challenger launched on Jan 27 1986 at 29°F

Null: The population median number of O-ring failures at Cool temperatures is the same as at Warm temperatures.

One-sided alternative

Incidents	Launch Temp.
1	Cool
1	Cool
1	Cool
3	Cool
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
1	Warm
1	Warm
2	Warm

Your turn

Step 1:

Using data from both groups, order from largest to smallest.

Put the number of incidents in order from smallest to largest, and indicate whether each number is from a Warm or Cool launch

Incidents	Launch
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
0	Warm
1	Cool
1	Cool
1	Cool
1	Warm
1	Warm
2	Warm
3	Cool

Step 2:

Add a new column called "order", with the numbers from 1 to $n_1 + n_2$.

Incidents	Launch	order
0	Warm	1
0	Warm	2
0	Warm	3
0	Warm	4
0	Warm	5
0	Warm	6
0	Warm	7
0	Warm	8
0	Warm	9
0	Warm	10
0	Warm	11
0	Warm	12
0	Warm	13
0	Warm	14
0	Warm	15
0	Warm	16
0	Warm	17
1	Cool	18
1	Cool	19
1	Cool	20
1	Warm	21
1	Warm	22
2	Warm	23
3	Cool	24

Step 2:

Add a new column called "order", with the numbers from 1 to $n_1 + n_2$.

Step 3:

Search for ties, and replace the order, with the average order for the observations with the same value.

Your turn

Incidents	Launch	order
0	Warm	1
0	Warm	2
0	Warm	3
0	Warm	4
0	Warm	5
0	Warm	6
0	Warm	7
0	Warm	8
0	Warm	9
0	Warm	10
0	Warm	11
0	Warm	12
0	Warm	13
0	Warm	14
0	Warm	15
0	Warm	16
0	Warm	17
1	Cool	18
1	Cool	19
1	Cool	20
1	Warm	21
1	Warm	22
2	Warm	23
3	Cool	24

Replace the order of the observations that are tied, with their average order.

Hint: the average of the integers m, m + 1, ... , n is (m + n)/2.

Incidents	Launch	order	rank
0	Warm	1	9
0	Warm	2	9
0	Warm	3	9
0	Warm	4	9
0	Warm	5	9
0	Warm	6	9
0	Warm	7	9
0	Warm	8	9
0	Warm	9	9
0	Warm	10	9
0	Warm	11	9
0	Warm	12	9
0	Warm	13	9
0	Warm	14	9
0	Warm	15	9
0	Warm	16	9
0	Warm	17	9
1	Cool	18	20
1	Cool	19	20
1	Cool	20	20
1	Warm	21	20
1	Warm	22	20
2	Warm	23	23
3	Cool	24	24

Step 4:

Add the ranks of the smallest group

Test statistic,

$$T = 20 + 20 + 20 + 24$$
$$= 84$$

approximate p-value

The test statistic will:

have mean, $n_1\bar{R} = 4 * (1 + 24)/2 = 50$

have standard deviation,

$s_R\ \text{sqrt}((n_1 n_2)/ (n_1 + n_2)) = 10.3$

(in R `(SDT<- sd(case0401$rank) * sqrt(n1*n2/(n1 + n2)))`)

(T - 50) / 10.3 is approximately Normal:

```
1- pnorm((84 - 50) / 10.3) = 0.005
```

exact p-value

Elements of smaller group	No. of groupings	Value of T
0 0 0 0	2380	36
0 0 0 1	3400	47
0 0 0 2	680	50
0 0 1 1	1360	58
0 0 0 3	680	51
0 0 1 2	680	61
0 1 1 1	170	69
0 0 1 3	680	62
0 1 1 2	170	72
1 1 1 1	5	80
0 0 2 3	136	65
0 1 1 3	170	73
1 1 1 2	10	83
0 1 2 3	85	76
1 1 1 3	**10**	**84**
1 1 2 3	10	87
	10626	

List all the ways of reassigning the observed "Incidents" to the "Cool" group.

What proportion have as extreme values of T?

$T \geq 84$

(10 + 10) / 10626 = 0.0018

```
wilcox.test(Incidents ~ Launch, data  = case0401)
# OR
wilcox.test(subset(case0401, Launch == "Cool")$Incidents,
 subset(case0401, Launch == "Warm")$Incidents)
```

```
        Wilcoxon rank sum test with continuity correction

data:  subset(case0401, Launch == "Cool")$Incidents and
subset(case0401, Launch == "Warm")$Incidents
W = 74, p-value = 0.001144
alternative hypothesis: true location shift is not equal
to 0

Warning message:
In wilcox.test.default(subset(case0401, Launch == "Cool")
$Incidents,  :
  cannot compute exact p-value with ties
```

Read section 4.2.3

Two-sided

Not our definition