Part 3

The Wilcoxon rank-sum test

The theory behind the Wilcoxon rank-sum test

Wilcoxon rank-sum test

Step 1: Order the dependent variable from smallest to largest

Group	Frequency
Intervention	0
Intervention	3
Intervention	1
Intervention	0
Intervention	2
Control	2
Control	31
Control	29
Control	30
Control	5

Group	Frequency
Intervention	0
Intervention	0
Intervention	1
Intervention	2
Control	2
Intervention	3
Control	5
Control	29
Control	30
Control	31

Wilcoxon rank-sum test

Step 2: Rank the dependent variable from smallest to largest

- Assign the lowest value of your dependent variable a rank of 1, the second lowest a rank of 2, etc.

- If two are equal, they are ‘tied ranks’. Take the average of the two ranks and make both values equal that.

- E.g. Lowest two values are 0:
  - Rank 1 and Rank 2.
  - Take the average of these (1+2/2) = 1.5

Wilcoxon rank-sum test

Step 2: Rank the dependent variable from smallest to largest

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	1	3
Intervention	2	4.5
Control	2	4.5
Intervention	3	6
Control	5	7
Control	29	8
Control	30	9
Control	31	10

Rank 1 + Rank 2 (1+2/2) = Both equal 1.5

Rank 4 + Rank 5 (4+5/2) = Both equal 4.5

Wilcoxon rank-sum test

Step 3: Add up the ranks per group

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	1	3
Intervention	2	4.5
Control	2	4.5
Intervention	3	6
Control	5	7
Control	29	8
Control	30	9
Control	31	10

- Intervention: 1.5+1.5+3+4.5+6 = 16.5
- Control: 4.5+7+8+9+10 = 38.5
- This is referred to as the **sum of ranks**

Wilcoxon rank-sum test

Step 4: Correct for the number of people in the group

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	1	3
Intervention	2	4.5
Control	2	4.5
Intervention	3	6
Control	5	7
Control	29	8
Control	30	9
Control	31	10

- This is necessary, as otherwise larger groups would have larger ranks.
- Firstly, calculate the mean rank for each group:

$$\textbf{Mean rank} = \frac{N*(N+1)}{2}$$

$$\textbf{Intervention mean rank} = \frac{5*(5+1)}{2} = \frac{5*6}{2} = \frac{30}{2} = 15$$

$$\textbf{Control mean rank} = \frac{5*(5+1)}{2} = \frac{5*6}{2} = \frac{30}{2} = 15$$

Wilcoxon rank-sum test

Step 4: Correct for the number of people in the group

Then minus the mean rank from each group's sum of ranks:

Sum of ranks – mean rank

- Intervention group: 16.5 - 15 = 1.5
- Control group: 38.5 – 15 = 23.5
- The lowest value is the test statistic (*W*):
  – Intervention group = **1.5**

Running the Wilcoxon rank-sum test in R

Running a Wilcoxon rank-sum test in R
Import the data

	Group	Frequency
1	Intervention	0
2	Intervention	3
3	Intervention	1
4	Intervention	0
5	Intervention	2
6	Control	2
7	Control	31
8	Control	29
9	Control	30
10	Control	5

frequency_data dataframe

You do not need to rank the data – R does this for you. Just enter the raw information (i.e. the IV and DV values)

Basic code to run the Wilcoxon rank-sum test

Keep output

DV

Dataframe

```r
model <- wilcox.test(Frequency ~ Group, data = frequency_data, paired = FALSE)
model
```

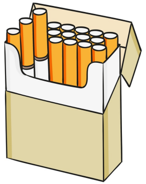
Function to run test

IV

Tells R data is not paired (i.e. independent groups design)

Displays the model output

Output for our cigarette example



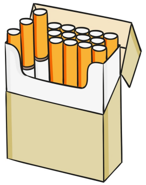
Lancaster University

```
> model <- wilcox.test(Frequency ~ Group, data = frequency_data, paired = FALSE)
Warning message:
In wilcox.test.default(x = c(0L, 3L, 1L, 0L, 2L), y = c(2L, 31L,   :
  cannot compute exact p-value with ties
> model

        Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 1.5, p-value = 0.02733
alternative hypothesis: true location shift is not equal to 0
```

Same as when we manually calculated W

- The lowest value is the test statistic:
  - Intervention group = **1.5**

Output for our cigarette example



```
> model <- wilcox.test(Frequency ~ Group, data = frequency_data, paired = FALSE)
Warning message:
In wilcox.test.default(x = c(0L, 3L, 1L, 0L, 2L), y = c(2L, 31L,  :
  cannot compute exact p-value with ties
> model

	Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 1.5, p-value = 0.02733
alternative hypothesis: true location shift is not equal to 0
```

There is a significant difference in the number of cigarettes smoked between the intervention and control groups

Is there a significant difference between the groups?

p ≤ .05 = there is a significant difference between the groups

p > .05 = there is no significant difference between the groups

In what direction is our result?

With parametric tests, we typically report the mean. With non-parametric tests, the median is typically preferred.

Produces the descriptive statistics for each group

Dataframe

Name of the independent variable

```
frequency_data %>%
  group_by(Group) %>%
  summarise(med = median(Frequency), n = n())
```

Instructs R to summarise

Calculate the median frequency and call this 'med'

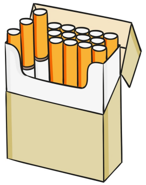
How many participants per group? Call this 'n'

In what direction is our result?

```
> frequency_data %>%
+   group_by(Group) %>%
+   summarise(med = median(Frequency), n = n())
# A tibble: 2 x 3
  Group          med     n
  <fct>        <int> <int>
1 Intervention     1     5
2 Control         29     5
```

Median number of cigarettes consumed is lower in the intervention group than the control group

Output for our cigarette example



```
> model <- wilcox.test(Frequency ~ Group, data = frequency_data, paired = FALSE)
Warning message:
In wilcox.test.default(x = c(0L, 3L, 1L, 0L, 2L), y = c(2L, 31L,   :
  cannot compute exact p-value with ties
> model

	Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 1.5, p-value = 0.02733
alternative hypothesis: true location shift is not equal to 0
```

What does this warning message mean? Do I need to worry about it?!

Different ways R calculates the p-value we will cover

1. The exact method

2. The normal approximation with continuity correction

Lancaster University

1. The exact method

- Uses a ”Monte Carlo” method

- Creates lots of datasets that are the same as the sample – but instead of assigning the correct group, assigns group randomly

Group	Frequency	Rank
Control	0	1.5
Intervention	0	1.5
Intervention		
Control		
Intervention		
Intervention		
Control		
Control		
Control		
Intervention		

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Interventio		
Interventio		
Control		
Interventio		
Control		
Control		
Control		
Control		

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	2	3.5
Control	2	3.5
Control	4	5
Control	7	6
Control	14	7.5
Control	14	7.5
Intervention	15	9
Intervention	16	10

x1000s of times

1. The exact method

- In these datasets, we know the null hypothesis is true (there is no difference between groups because we have randomly assigned group)

- How often is the difference that appears in the generated datasets (blue) as large as the difference in the true data (red and green)?

Group	Frequency	Rank
Control	0	1.5
Intervention	0	1.5
Interven		
Contr		
Interven		
Interven		
Contr		
Contr		
Contr		
Interven		

Group	Frequency	Rank
Intervention	0	1.5
Interve		
Interve		
Interve		
Con		
Interve		
Con		
Con		
Con		
Con		

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	2	3.5
Control	2	3.5
Control	4	5
Control	7	6
Control	14	7.5
Control	14	7.5
Intervention	15	9
Intervention	16	10

x1000s of times

Group	Frequency	Rank
Intervention	0	1.5
Intervention	0	1.5
Intervention	1	3
Intervention	2	4.5
Control	2	4.5
Intervention	3	6
Control	5	7
Control	29	8.5
Control	30	8.5
Control	31	10

2. The normal approximation with continuity correction

- Assumes that the sampling distribution of the W statistic is normal

  → This produced a standard error

  → This can be used to calculate z (and then a p-value)

- By default, it applies a continuity correction

  → This corrects for the fact we are assuming the sampling distribution of W is normal, but a person can change in ranks only in set amounts (e.g. 1, or 0.5 if tied ranks)

Don't worry if you don't fully understand this. The key is to understand there are two different methods for how the p-value is calculated

Which is the default method?
It depends…

- Sample size < 50 in all group **and** no tied ranks = exact method

- Sample size ≥ 50 in any group **OR** tied ranks = normal approximation with continuity correction

Note: you can also specify which method should be run, but we won’t cover that in this set of lectures

How do I know which method R has used?

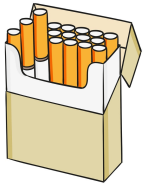
```
> model

	Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 5, p-value = 0.14
alternative hypothesis: true location shift is not equal to 0
```

Normal approximation with continuity correction

```
> model

	Wilcoxon rank sum test

data:  DV by IV
W = 1300, p-value = 0.7314
alternative hypothesis: true location shift is not equal to 0
```

The exact method

Output for our smoking example



Lancaster University

```
> model <- wilcox.test(Frequency ~ Group, data = frequency_data, paired = FALSE)
Warning message:
In wilcox.test.default(x = c(0L, 3L, 1L, 0L, 2L), y = c(2L, 31L,   :
  cannot compute exact p-value with ties
> model

	Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 1.5, p-value = 0.02733
alternative hypothesis: true location shift is not equal to 0
```

This warning message just tells you that it tried to do the exact method (because <50 participants per group), but couldn't be completed due to tied ranks

The p-value in our smoking example was calculated using the normal approximation with continuity correction

What about an effect size?

Not provided, but it can be estimated using the R output

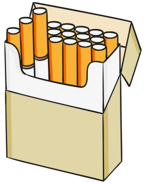
$$r = \frac{z}{\sqrt{N}}$$

Can be calculated from the p-value by running the following code:

```
qnorm(model$p.value/2)
```

Total number of observations in the study

Our smoking example



```
        Wilcoxon rank sum test with continuity correction

data:  Frequency by Group
W = 1.5, p-value = 0.02733
alternative hypothesis: true location shift is not equal to 0
```

```
> qnorm(model$p.value/2)
[1] -2.206794
```

$$r = \frac{z}{\sqrt{N}} = \frac{-2.206794}{\sqrt{10}} = \frac{-2.206794}{3.16} = -0.70$$

How do we interpret the effect size?

Interpretation		
	Positive	Negative
Small	0.1 to 0.3	-0.1 to -0.3
Medium	0.3 to 0.5	-0.3 to -0.5
Large	0.5 to 1.00	-0.5 to -1.00

In our example, r = -0.70.

Large effect size

Reporting the results in APA format

A Wilcoxon rank-sum test revealed that the number of cigarettes smoked was significantly lower in the intervention group (Median = 1; Range = 0-3) relative to the Control group (Median = 29; Range = 2-31), $W = 1.5$, $p = .027$, $r = -.70$.

Wait…
Why is my W different when calculated manually and in R?

- If you were to calculate W manually, W should be the smallest value (sum of ranks – mean rank)
- In R, W is reported for the first factor level
- Doesn't affect significance, so not something you need to worry about
  → You can just report what R outputs

- Intervention group: 16.5 - 15 = 1.5
- Control group: 38.5 – 15 = 23.5
- The lowest value is the test statistic (*W*):
  – Intervention group = **1.5**

Calculating manually – W = 1.5

Reported by R – W = 1.5 OR 23.5