Wilcoxon rank-sum test

Starter

1. **(Review of last lesson)** A sociologist is interested in comparing the ages of husbands and wives. He collected the data below, which shows the ages for the husband and wife in a random chosen sample of nine couples.

Couple	A	B	C	D	E	F	G	H	I
Husband's age (years)	79	39	55	71	37	39	48	63	54
Wife's age (years)	70	36	49	54	38	32	49	52	56

Use these data to test the hypothesis that most men are older than their wives using:
(a) a sign test,
(b) a Wilcoxon signed-rank test.
Carry out your tests at the $5\,\%$ significance level.
(c) What assumptions, if any, are required for each of the tests?
(d) Explain how it is possible for the two tests to give different results.

2. (a) How many different arrangements are there of the letters $FFFSSSSS$?

(b) The position of the letters is given a value so that if a letter appears in the first position, it has a value of 1. If it appears in the second position, it has a value of 2 etc. For example, the total for the F's in the arrangement $FSSFSFS$ is $1+4+6=11$.
Calculate the probability, that the total of the three F's will be:
(i) 6
(ii) 8
(iii) less than or equal to 9.
Give your answers as fractions and percentages to 4 s.f..

(c) Find x such that $P(\text{total} \leq x) \leq 5\,\%$.

Notes

What happens when the data is not paired i.e. there are more more values in one sample than the other? We can still compare the medians of the samples using the ***Wilcoxon rank-sum test***.

The ***Wilcoxon rank-sum test*** tests whether two samples come from populations with identical distributions. It ***does not require any assumptions*** about the two populations and, importantly, their medians do not even need to be known.

H_0 : The two distributions are the same
H_1 : The two distributions are different

Assume that the 2 distributions have the same shape, we can state that:

H_0 : The median scores are the same
H_1 : The median scores are different (or higher or lower)

E.g. 1 Over seven weeks, Jack visits his local supermarket on a Friday or Saturday and times how long, to the nearest minute, it takes to his shopping. The data is below.

Friday	38	56	60		
Saturday	74	58	61	50	64

(a) Rank the combined data by copying and completing this table.

Value								
Rank	1	2	3	4	5	6	7	8
Day								

(b) Find the sum of the Friday ranks, R_F.

(c) State the smallest value that R_F could be.

(d) State the largest value that R_F could be.

(e) H_0 : shopping time has the same distribution on Fridays and Saturdays
H_1 : shopping on Friday is likely to take less time than on Saturday
Is there evidence, at the $5\,\%$ level, to suggest that shopping on Friday takes less time than on Saturday? Make your decision using your answer to 2(c) from the starter.

It would be a pain to have to do all the arrangement calculations each time for a Wilcoxon rank-sum test, so it is no surprise to hear there are table of critical values (see p206 of textbook).

The tables contain the ***largest value*** which would lead to the ***rejection*** of the null hypothesis.

From tables, the critical value for two samples whose sizes are 3 and 5 is 7 i.e. the value found in question 2(c) of the starter.

E.g. 2 Let the sample sizes of two distributions be m and n where $m \leq n$.
Let R_m be the sum of the ranks of the sample with size m when the smallest ranked value is given the value of 1.
Let R'_m be the sum of the ranks of the sample with size m when the smallest ranked value is given the value of $m + n$ i.e. the ranks are reversed.
Let $r_1, r_2, r_3, \ldots, r_m$ be the individual ranks of the sample with size m in ascending order.

(a) If $m = 2$ and $n = 4$ with $r_1 = 2$ and $r_2 = 3$. State the value of R_m and find the value of R'_m

(b) For the general case, $R_m = r_1 + r_2 + r_3 + \ldots + r_m$. Find the value of R'_m in terms of m, n and R_m. Hint: consider the value $r_m + r'_m$ from (a)

Working: (a) $R_m = 2 + 3 = 5$
The original ranks are $\_, r_1, r_2, \_, \_, \_,$
The reversed ranks would be $\_, \_, \_, r_2, r_1, \_$
$r'_1 = 5$ and $r'_2 = 4 \quad \Rightarrow \quad R'_m = 5 + 4 = 9$

Success Criteria for carrying out a Wilcoxon rank-sum test

Let the sample sizes of the two distributions be m and n where $m \leq n$.

1. Merge the data sets and rank all the data, smallest to largest (smallest value is rank 1).
2. Add the ranks of the smallest sample, R_m.
3. Choose the test statistic W as the smaller value of R_m and $m(m+n+1) - R_m$.
4. If W is less than the critical value given by tables, reject H_0.

$m(m+n+1) - R_m$ would be the value of R_m if we ranked the data values from largest to smallest and solves the problem of opposite ranking and whether H_1 is "greater than" or "less than".

The formula booklet has a simplified method included.

E.g. 3 State the critical values for the following Wilcoxon rank-sum tests using the table on p206:

(a) a one-tailed test at the $5\,\%$ significance level where $m = 6, n = 10$

(b) a two-tailed test at the $2\,\%$ significance level where $m = 3, n = 9$

(c) a one-tailed test at the $2.5\,\%$ significance level where $m = 4, n = 8$

(d) a two-tailed test at the $10\,\%$ significance level where $m = 5, n = 7$

Working: (a) 35

E.g. 4 The lengths of the femur, in mm, in samples of a mouse from Britain and North Africa are given below:

Britain	12.3	12.7	13.1	10.8	11.3	11.8	12.4	13.2
North Africa	10.6	9.8	11.5	10.0	11.1			

Conduct a non-parametric test at the $5\,\%$ level to test whether the data are consistent with the assumption that the mice in Britain and North Africa are the same breed.

Working: H_0 : mice in Britain and North Africa are the same breed

H_1 : mice in Britain and North Africa are not the same breed

Here are the lengths ranked in order from smallest to largest:

Value	9.8	10.0	10.6	10.8	11.1	11.3	11.5	11.8	12.3	12.4	12.7	13.1	13.2
Rank	1	2	3	4	5	6	7	8	9	10	11	12	13
Area	NF	NF	NF	B	NF	B	NF	B	B	B	B	B	B

$R_{NF} = 18$

$m = 5, n = 8$: $\quad m(m+n+1) - R_m = 5(5+8+1) - 18 = 52$

$\therefore W = 18$

The $5\,\%$ critical value for a two-tailed test when $m = 5, n = 8$ is 21.

Since $W = 18 \leq 21$, we reject H_0.

There is evidence to suggest the mice in Britain and North Africa are not the same breed.

E.g. 5 An estate manager is wondering which trees to plant in a forest. She collects data from nearby forests on the heights of Stardust and Blue Gown trees after 10 years.

Stardust	1.9	1.5	1.7	2.4	2.3	2.0	3.4
Blue Gown	3.7	2.6	2.1	3.6			

Test at the $2.5\ \%$ level that the average height of Blue Gown trees after ten years is higher than the Stardust variety.

In the event of ***tied ranks***, the ***average rank*** is assigned.

E.g. 6 A study was undertaken to investigate the effect of vitamin C on the common cold. Fifteen students, each of whom had developed the symptoms of a common cold, were randomly assigned to two groups. Group A acted as the control group and received unknowingly only a daily sugar tablet, whereas Group B received one gram of vitamin C per day. The table shows the duration, in days, of cold symptoms for each student.

Group A	13	11	12	9	18	7	12
Group B	8	14	7	10	9	12	

Test at the $1\ \%$ level the suggestion that consumption of one gram of vitamin C each day improves the time to recover from a common cold

Video (password needed): Wilcoxon rank-sum test
Video: How to conduct the Wilcoxon rank-sum test

Solutions to Starter and E.g.s

Exercise
p56 4D Qu 1-4, (5 red)

Summary
H_0 : The two distributions are the same
H_1 : The two distributions are different
Success Criteria for carrying out a Wilcoxon rank-sum test:
Let the sample sizes of the two distributions be m and n where $m \leq n$.

1. Merge the data sets and rank all the data, smallest to largest (smallest value is rank 1).
2. Add the ranks of the smallest sample, R_m.
3. Choose the test statistic W as the smaller value of R_m and $m(m+n+1)-R_m$.
4. If W is less than the critical value given by tables, reject H_0

$m(m+n+1)-R_m$ would be the value of R_m if we ranked the data values from largest to smallest and solves the problem of opposite ranking and whether H_1 is "greater than" or "less than".

The formula booklet has a simplified method included.