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.

Working: (a) H_0 : most husbands and wives are the same age
H_1 : most husbands are older than their wives
By H_0, X, the number of "—" signs is distributed such that
$X \sim \mathrm{B}(9, 0.5)$
The signs of the differences are:
$+ \quad + \quad + \quad + \quad - \quad + \quad - \quad + \quad -$
$p = P(X \geq 6) = P(X \leq 3) = 0.254$
Since $p = 0.254 \nleq 0.05$, we reject H_0.
There is no evidence to suggest that most husbands are older than their wives.

(b) H_0 : most husbands and wives are the same age
H_1 : most husbands are older than their wives

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
Differences	9	3	6	17	−1	7	−1	11	−2
|Differences|	9	3	6	17	1	7	1	11	2
Rank	7	4	5	9	1.5	6	1.5	8	3
Signed rank	7	4	5	9	−1.5	6	−1.5	8	−3

$W_+ = 39$ and $W_- = 6$
Check: when $n = 9$, $\frac{1}{2} \times 9 \times (9+1) = 45 = 39 + 6$ ✓
$T = 6$ (smallest value)
From tables, the critical value for a one-tail test at the 5% level with 9 values is 8.
Since $T = 6 \leq 8 = CV$, we reject H_0.
There is evidence to suggest that most husbands are older than their wives.

(c) The sign test in (a) requires no assumptions.
The signed-rank test in (b) requires that the distribution of the differences is symmetrical.

(d) The signed-rank test in (b) uses more information from the sample so it can discriminate better. It has a lower probability of giving a Type II error because it detects that the minor deviations are all much smaller than than the larger deviations.

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 $P(\text{total} \leq x) \leq 5\,\%$.

Working: (a) Since there are 8 items with 3 and 5 repetitions, there are $\dfrac{8!}{3!5!} = 56$ arrangements.

(b) (i) To get a total of 6 the F's must all appear at the start
i.e. $FFFSSSSS \quad 1+2+3=6$
$P(\text{total of } 6) = \dfrac{1}{56} = 1.786\,\%$

(ii) These are the possible arrangements for a total of 8:
$FSFFSSSSS$ and $FFSSFSSS$.
$P(\text{total of } 8) = \dfrac{2}{56} = 3.571\,\%$

(iii) Total of 7: $FFSFSSSS \quad \Rightarrow P(\text{total of } 7) = \dfrac{1}{56} = 1.786\,\%$
Total of 9: $SFFFSSSS$, $FSFSFSSSS$ & $FFSSSFSS$
$\Rightarrow \quad P(\text{total of } 9) = \dfrac{2}{56} = 3.571\,\%$
$P(\text{total} \leq 9) = \dfrac{7}{35} = 20\,\%$

(c) $P(\text{total of } 6) = \dfrac{1}{56} = 1.786\,\%$
$P(\text{total} \leq 7) = \dfrac{2}{56} = 3.571\,\%$
$P(\text{total} \leq 8) = \dfrac{4}{56} = 7.143\,\%$
So $x = 7$

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.

Working: (a)

Value	38	50	56	58	60	61	64	74
Rank	1	2	3	4	5	6	7	8
Day	F	S	F	S	F	S	S	S

(b) The sum of the Friday ranks, R_F, is 9.

(c) The smallest value that R_F could be is $1 + 2 + 3 = 6$

(d) The largest value that R_F could be is $6 + 7 + 8 = 21$
The average of the smallest and largest value is 13.5 so if R_F was larger than this value, there would be no need to do any further calculation.

(e) From the starter, $P(\text{total} \leq 7) \leq 5\%$.
Since $R_F = 9 \nleq 7$, we do not reject H_0.
There is no evidence to suggest that shopping on Friday is faster than on Saturday.

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 $r'_2 = 4 \quad \Rightarrow \quad R'_m = 5 + 4 = 9$

(b) When $m = 2$ and $n = 4$:

$r_1 = 2 \quad \Rightarrow \quad r'_1 = 5$

$r_2 = 3 \quad \Rightarrow \quad r'_2 = 4$

In each case, $r_m + r'_m = 7 \quad \Rightarrow \quad r_m + r'_m = m + n + 1$

$r'_m = m + n + 1 - r_m$

$R'_m = r'_1 + r'_2 + r'_3 + \ldots + r'_m$

$$R'_m = (m + n + 1 - r_1) + (m + n + 1 - r_2) + (m + n + 1 - r_3) + \ldots + (m + n + 1 - r_m)$$

$R'_m = m(m + n + 1) - r_1 - r_2 - r_3 - \ldots - r_m$

$R'_m = m(m + n + 1) - R_m$

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 (b) 7 (c) 14 (d) 21

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.

Working: H_0 : after 10 years the average height of the trees is equal
H_1 : after 10 years the average height of Blue Gown trees is higher than Stardust trees
Here are the weights ranked in order from smallest to largest:

Value	1.5	1.7	1.9	2.0	2.1	2.3	2.4	2.6	3.4	3.6	3.7
Rank	1	2	3	4	5	6	7	8	9	10	11
Variety	S	S	S	S	BG	S	S	BG	S	BG	BG

$R_{BG} = 5 + 8 + 10 + 11 = 34$
$m = 4, n = 7$: $\quad m(m+n+1) - R_m = 4(4+7+1) - 34 = 14$
$\therefore W = 14$
The 2.5% critical value for a one-tailed test when $m = 4, n = 7$ is 13.
Since $W = 14 \not\leq 13$, we do not reject H_0.
There is no evidence to suggest that the average height of Blue Gown trees after ten years is higher than the Stardust variety.

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

Working: H_0 : one gram of vitamin C does not improve recovery time from a cold
H_1 : one gram of vitamin C improves recovery time from a cold
Here are the recovery times ranked in order from smallest to largest:

Value	7	7	8	9	9	10	11	12	12	12	13	14	18
Rank	1.5	1.5	3	4.5	4.5	6	7	9	9	9	11	12	13
Group	A	B	B	A	B	B	A	A	A	B	A	B	A

$R_B = 1.5 + 3 + 4.5 + 6 + 9 + 12 = 36$
$m = 6, n = 7$: $\quad m(m + n + 1) - R_m = 6(6 + 7 + 1) - 19 = 48$
$\therefore W = 36$
The 10% critical value for a one-tailed test when $m = 6, n = 7$ is 25.
Since $W = 36 \nleq 25$, we do not reject H_0.
There is no evidence to suggest that consuming one gram of vitamin C per day improves recovery 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)