16
Wilcoxon Signed Rank Test

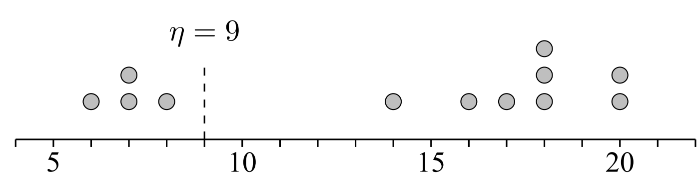
The z and t hypothesis tests from chapters 10 and 15 which explore the *locations* of distributions have all focused on the *population mean*, μ. Those *parametric* tests each rely on either normally distributed populations or sample sizes greater than 20 such that the sample mean is approximately normally distributed.

Given small samples ($n < 20$) from populations whose distributions may not be assumed to be normally distributed, a number of **non-parametric** tests may be used which instead explore the locations of distributions by considering the **population median**, η (*'eta'*).

16.1 Sign Test

The Sign Test is a simple non-parametric test that, whilst *not required for the Advanced Higher Statistics course*, helps to lay the groundwork for the tests covered later in this chapter. It uses a useful property of the population median, η, namely that 50% of observations sampled from the population can be expected to be greater than η and 50% can be expected to be less than η. This means that the number of observations on one side of the population median in a sample of size n should follow a $B(n, 0.5)$ distribution. For example, consider the following random sample of values ($n = 12$) and a claim that the population median is less than 9:

$$\left.\begin{array}{ll} \mathrm{H}_0: & \eta = 9 \\ \mathrm{H}_1: & \eta < 9 \end{array}\right] \begin{array}{l} \alpha = 0.05 \\ \text{One-tailed} \end{array}$$



With 4 out of the 12 observed values being less than 9, the p-value for this test can be calculated as the probability of observing 4 or fewer values less than $\eta = 9$. Letting random variable X represent the number of such values, which follows a $B(12, 0.5)$ distribution under the null hypothesis:

$$p\text{-value} = P(X \leqslant 4) = 0.1938$$

Since $0.1938 > 0.05$, do not reject H_0 at the 5% level of significance.

There is insufficient evidence to suggest that the population median is less than 9.

16.2 Wilcoxon Signed Rank Test

Whilst the Sign Test is only concerned with the number of values that fall on either side of the hypothesised population median, η, the Wilcoxon Signed Rank test also considers how far from the median these values lie. More specifically, it applies weighting to the values by assigning *rank* 1 to the value closest to η, *rank* 2 to the second closest, and so on.

Zeros: Any values that are exactly equal to the median, η, are ignored and not given a rank, with the sample size n reduced to reflect the number of *remaining values to be ranked*.

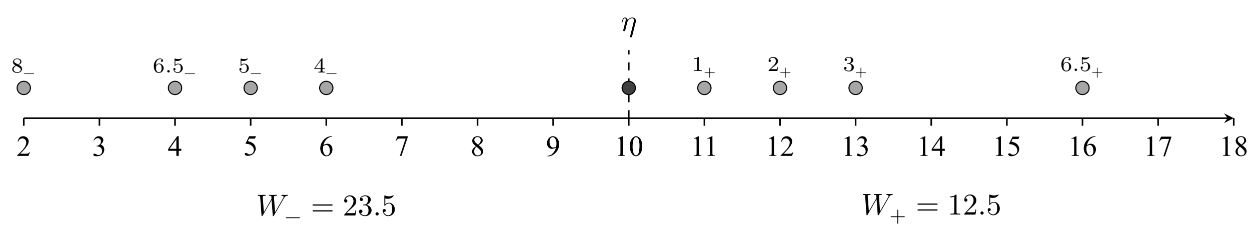
Tied ranks: Values that are equally distant from the median, η, share ranks. For example, if two values are equally far from η when looking to assign rank 4, then they are each ranked 4.5, since $\frac{4+5}{2} = 4.5$. If three values are equally distant from η when assigning rank 7, then they are each ranked 8 since $\frac{7+8+9}{3} = 8$.

- Ranks of values greater than the median are denoted as *positive ranks*, and their total as W_+.

- Ranks of values less than the median are denoted as *negative ranks*, and their total as W_-.

Consider a random sample of nine values from a distribution with median $\eta = 10$: $2, 4, 5, 6, 10, 11, 12, 13, 16$

$$\left.\begin{array}{ll} \mathrm{H}_0: & \eta = 10 \\ \mathrm{H}_1: & \eta \neq 10 \end{array}\right] \begin{array}{l} \alpha = 0.05 \\ \text{Two-tailed} \end{array}$$



Note that the sum of W_+ and W_- is $12.5 + 23.5 = 36$, which will always be the total when ranks are assigned to $n = 8$ values, since $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 = 36$. It is recommended to check that the sum of W_+ and W_- does give the appropriate total for the sample size, which can be calculated for a sample of size n as $\frac{1}{2}n(n+1)$.

For a Wilcoxon Signed Rank test:

- The **test statistic**, denoted W, is the *lower* of W_+ and W_-.
- The **critical value** can be obtained from Page 15 of the Data Booklet.
- H_0 is rejected if the test statistic W is **less than or equal to** the critical value.

Test statistic: $W = 12.5$

Critical value: $CV = 3$ ($\alpha = 0.05$, $n = 8$ and two-tailed.)

Since $12.5 > 3$, do not reject H_0 at the 5% level of significance.
There is insufficient evidence to suggest that the population **median** is not 10.

The Wilcoxon Signed Rank test will typically be used for small samples for which the Central Limit Theorem cannot be invoked, from populations which cannot be reasonably assumed to be normally distributed in order to use a t-test. The table of critical values only extends to samples up to size $n = 20$, once zeros have been discarded.

Conditions for valid use:

- The underlying distribution is **symmetrical**.
- The data was obtained through a **random sample**.

A suggested methodical approach to obtaining the signed ranks required without a diagram is:

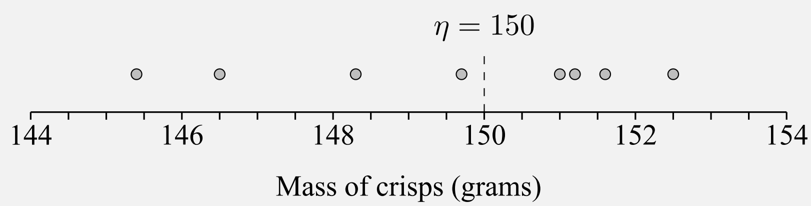
- **Subtract** η from every value in the sample.
- Ignoring any **zeros**, assign ranks to the remaining n values in ascending order by **absolute** value.

The following example should be compared to Example 1 from Section 10.7, Chapter 10, in which a t-test was used.

Example 1

Problem: A supermarket sells sharing-sized bags of a particular brand of crisps. A consumer watchdog is asked to investigate a claim that the median mass of crisps contained in a bag is less than the stated contents of 150 grams. A random sample of bags gives the following results, in grams:

148.3 151.2 149.7 152.5 145.4 151.0 146.5 151.6



Stating a necessary assumption, perform a non-parametric test to assess the claim at the 1% level of significance.

Solution:
Assume that the mass of crisps in a bag is symmetrically distributed.

$$\left.\begin{array}{ll} \mathrm{H_0}: & \eta = 150 \\ \mathrm{H_1}: & \eta < 150 \end{array}\right] \begin{array}{l} \alpha = 0.01 \\ \text{One-tailed} \end{array}$$

148.3	151.2	149.7	152.5	145.4	151.0	146.5	151.6	
-1.7	1.2	-0.3	2.5	-4.6	1.0	-3.5	1.6	(*subtract* $\eta = 150$)
5_-	3_+	1_-	6_+	8_-	2_+	7_-	4_+	(*assign signed ranks*)

$$W_+ = 15 \qquad \tfrac{1}{2} \times 8 \times 9 = 36$$
$$W_- = 21 \qquad \textit{Check: } 15 + 21 = 36$$

Test statistic: $W = 15$

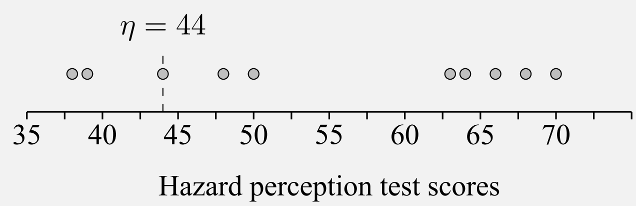
Critical value: $CV = 1$ ($\alpha = 0.01$, $n = 8$ and one-tailed.)

Since $15 > 1$, do not reject $\mathrm{H_0}$ at the 1% level of significance. There is insufficient evidence to suggest that the median mass of crisps in the bags is less than the stated value of 150g.

Example 2

Problem: Before learner drivers can sit their practical test, they must sit a theory test. Part of this is a hazard perception test, in which a maximum of 75 points are available. Prior to some changes to the hazard perception test, the median score was 44. The points scored by ten randomly selected learner drivers in the updated hazard perception test are recorded:

63 38 48 44 50 70 68 39 64 66



Test whether there is evidence, at the 10% level of significance, that the median score has changed from 44.

Solution:

$$\left.\begin{array}{ll} H_0: & \eta = 44 \\ H_1: & \eta \neq 44 \end{array}\right] \begin{array}{l} \alpha = 0.1 \\ \text{Two-tailed} \end{array}$$

63	38	48	44	50	70	68	39	64	66	
19	-6	4	0	6	26	24	-5	20	22	(*subtract* $\eta = 44$)
5_+	3.5_-	1_+		3.5_+	9_+	8_+	2_-	6_+	7_+	(*assign signed ranks*)

$$W_+ = 39.5 \qquad \tfrac{1}{2} \times 9 \times 10 = 45$$

$$W_- = 5.5 \qquad \textit{Check: } 39.5 + 5.5 = 45$$

Test statistic: $W = 5.5$

Critical value: $CV = 8$ ($\alpha = 0.1, n = 9$ and two-tailed.)

Since $5.5 < 8$, reject H_0 at the 10% level of significance. There is evidence to suggest that the median score achieved in the hazard perception tests has changed from 44 points.

Exercise 16.1

1. It is known that a certain tree produces apples with a median mass of 150g. Eight apples are chosen at random from a crate containing apples picked from the tree. The masses, in grams, are:

   147 138 171 142 152 145 141 143

   Use a non-parametric test, at the 5% significance level, to test whether the median mass of the apples is different from 150g.

2. For number of social media posts made in the last week are recorded for a random sample of eight UK teenagers, with the results:

   5 2 18 11 0 6 1 0

   Test at the 5% significance level the claim that the median number of social media posts made per week by a UK teenager is 10.

3. A recycling centre allows residents to deposit their metal waste into a skip set aside for it. The decision as to how often the skip will need to be emptied is based on an understanding that the median daily mass of metal waste deposited into the skip is 15kg, based on previous records. It has been suggested that this figure is no longer accurate. The recycling centre recorded the amount of metal waste deposited on eight randomly selected days. The results, in kg, were:

   11.6 14.7 15.1 14.3 16.2 9.7 15.2 10.7

   Perform a non-parametric test to assess, at the 10% level of significance, whether this sample provides evidence to support the suggestion that the median daily mass of metal waste deposited has changed.

4. The lifetimes of a random sample of a particular brand of candle, measured in minutes, are:

   372 352 335 364 345 360 354 358 348 341

   Determine if there is evidence at the 5% significance level that the median lifetime of this brand of candles is different from 6 hours.

5. A company's complaints department has a policy that states that the median length of time its customers should have to wait before receiving a response to any complaint should not exceed 14 days. A sample of ten customers who have received responses to complaints were asked how long they had to wait. The results, in days, were:

   15 10 17 13 18 12 23 20 19 21

   (a) State two assumptions required in order for a Wilcoxon Signed Rank Test to be performed.
   (b) Assess whether there is evidence to support a claim that the complaint department's policy is being breached, at the 5% level of significance.

6. A company is concerned about the length of time customers have to wait on a helpline before being connected to an agent. Previous records shows that the median waiting time was 34 minutes, and following additional training for the agents the company wishes to gauge the impact it may have had. Customer waiting times, in minutes, for a random sample of ten calls are recorded:

   55 20 31 12 18 32 28 16 14 33

   Stating an assumption required, perform a non-parametric test at the 5% level of significance to assess whether there is evidence the median waiting time has reduced from 34 minutes.

7. A medical centre schedules doctors appointments for patients on the basis of a median appointment duration of 8 minutes. Wishing to determine whether this figure is reasonable, the length of ten randomly chosen appointments are recorded, and displayed below (in minutes).

   7 8 6 5 9 5 7 5 10 4

   Assess at the 5% level whether the data suggests the median appointment duration is not 8 minutes.

16.3 Wilcoxon Signed Rank Test for Paired Data

In Chapter 15, the *paired t-test* was introduced. It is used for assessing the *mean of the differences between paired values* from small samples, and a key assumption for the test is that the *population of differences is normally distributed*. Where this assumption cannot reasonably be made, a **Wilcoxon Signed Rank Test for Paired Data** can instead be used to assesses the **median of the differences**, η_d.

Conditions for valid use of the Wilcoxon Signed Rank Test for Paired Data:

- The **population of differences are symmetrical**.
- The data was obtained through **random sampling**.

Just as with the paired t-test, the paired samples are considered as a *single sample of differences*, allowing a similar process to the one-sample Wilcoxon Signed Rank Test.

Example

Problem: An athletics coach wishes to assess the value to athletes of an intensive period of weight training. Twelve 400-metre runners are selected at random and their times to complete this distance, in seconds, are recorded. Following a programme of weight training they run the distance again, with their times in seconds again recorded. The table below summarises the results.

Athlete	A	B	C	D	E	F	G	H	I	J	K	L
Before	51.0	49.8	49.5	50.1	51.6	48.9	52.4	50.6	53.1	48.6	52.9	53.4
After	50.6	50.4	48.9	49.1	51.6	47.6	53.5	49.9	51.0	48.5	50.6	51.7

Stating an assumption required, perform a non-parametric test to assess the evidence that the training programme will improve athletes' times for the 400 metre distance.

Solution:
(Note that $d = \textit{Before} - \textit{After}$ will give differences such that a positive number represents an improvement in running time, and η_d represents the population median of differences).

$$\left.\begin{array}{ll} H_0: & \eta_d = 0 \\ H_1: & \eta_d > 0 \end{array}\right] \begin{array}{l} \alpha = 0.05 \\ \text{One-tailed} \end{array}$$

0.4	-0.6	0.6	1.0	0.0	1.3	-1.1	0.7	2.1	0.1	2.3	1.7	
2_+	3.5_-	3.5_+	6_+		8_+	7_-	5_+	10_+	1_+	11_+	9_+	(*assign signed ranks*)

$$\begin{array}{ll} W_+ = 55.5 & \tfrac{1}{2} \times 11 \times 12 = 66 \\ W_- = 10.5 & \textit{Check: } 55.5 + 10.5 = 66 \end{array}$$

Test statistic: $W = 10.5$
Critical value: $CV = 13$ ($\alpha = 0.05$, $n = 11$ and one-tailed.)

Since $10.5 < 13$, reject H_0 at the 5% level of significance. There is evidence to suggest that the median difference in 400 metres times is greater than zero, meaning times have improved after the weight training programme.

Exercise 16.2

1. To measure the effectiveness of a drug for asthmatic relief, twelve subjects, all susceptible to asthma, were each randomly administered either the drug or a placebo during two separate asthma attacks. After one hour an asthmatic

index was obtained on each subject with the following results:

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Drug	28	31	17	18	31	12	33	24	18	25	19	17
Placebo	32	33	23	26	34	17	30	24	19	23	21	24

Investigate the claim that the drug reduces the asthmatic index.

2. A group of 8 patients with a particular illness were given a special diet and it was desired to test the assess the impact at the end of a 2-week period. Their masses, measured before and after the diet, are shown below, in kg.

Patient	A	B	C	D	E	F	G	H
Before	82.27	78.18	86.36	85.00	95.45	75.45	83.18	83.64
After	82.87	79.54	87.36	86.10	94.99	75.48	83.54	82.15

Use a non-parametric test to determine if there is evidence at the 5% level that the diet increases patient mass, stating any assumptions made.

3. As part of her research into the behaviour of the human memory, a psychologist asked 15 randomly selected school pupils to talk for five minutes on ‘my day at school’. Each pupil was then asked to estimate how many times they thought that they had used the word “like” during the five minutes. The table below gives their estimates together with the true values.

Pupil	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
True value	12	20	1	8	0	12	12	17	6	5	24	23	10	18	16
Estimated value	9	19	3	14	4	12	16	14	5	9	20	16	11	17	19

Use a non-parametric test to investigate whether pupils can remember accurately the frequency with which they use a particular word in a verbal description.

4. Eight amateur runners are selected to study the effects of consuming energy gels before running on runners’ performances at middle-distances. Each has their finish time recorded for a 5000m run without consuming energy gels beforehand, then a month later their time is recorded again for a 5000m run before which each consumed two energy gels. The data is shown in the table below, with times recorded to the nearest minute.

Runners	A	B	C	D	E	F	G	H
Without gels	24	19	21	17	26	32	18	20
With gels	22	19	20	18	23	28	15	20

Perform a hypothesis test at the 5% level of significance to assess whether the data supports a conclusion that consuming energy gels improves performances for runners.

5. The delay in sound transmission through an audio interface is called *latency*. A music technology website wishes to investigate whether a company’s new audio interface, really offers low latency than its older one. A random sample of ten computers each have their audio latency measured, first with the older interface and then with the new interface. The results are as follows, measured in milliseconds:

Computer	1	2	3	4	5	6	7	8	9	10
Old interface	5.7	3.6	7.3	6.1	2.8	4.8	5.3	2.9	6.7	11.7
New interface	5.4	3.5	7.5	5.4	2.7	4.0	5.3	2.6	6.3	11.6

Stating an assumption required, perform a non-parametric test to assess whether the newer interface offers lower latency, at the 5% level of significance.

16.4 Normal Approximation to the Wilcoxon Signed Rank Test

The Data Booklet only provides Wilcoxon critical values for effective sample sizes where $n \leqslant 20$, due to the complexity of obtaining exact critical values as n increases. When dealing with a larger sample, where $n > 20$ after zeros have been removed, a Wilcoxon Signed Rank Test may still be performed using a **normal approximation**. Page 15 of the Data Booklet provides the required properties of the distribution of the test statistic, W:

$$E(W) = \frac{1}{4}n(n+1) \quad \text{and} \quad V(W) = \frac{1}{24}n(n+1)(2n+1)$$

Since here a *discrete* distribution is being approximated by the *continuous* normal distribution, a **continuity correction** is required. If the test statistic W is taken as the lower of W_+ and W_-, a continuity correction of $+0.5$ should be used.

Example

Problem: It is claimed by a local resident that more than 50% of all the vehicles on an urban road exceed the 30 mph speed limit. The speed of each of a random sample of 24 vehicles is recorded with the following results:

22	24	26	28	29	30	30	32	33	34	35	35
37	38	38	39	40	41	42	45	48	56	62	72

Use a non-parametric test to investigate the resident's claim.

Solution:

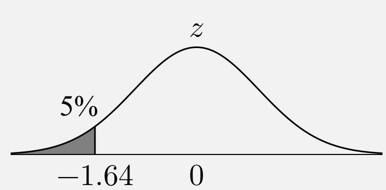
$$\left.\begin{array}{ll} H_0: & \eta = 30 \\ H_1: & \eta > 30 \end{array}\right] \begin{array}{l} \alpha = 0.05 \\ \text{One-tailed} \end{array}$$

22	24	26	28	29	30	30	32	33	34	35	35	
37	38	38	39	40	41	42	45	48	56	62	72	
-8	-6	-4	-2	-1	0	0	2	3	4	5	5	(*subtract* $\eta = 30$)
7	8	8	9	10	11	12	15	18	26	32	42	
12_-	9_-	5.5_-	2.5_-	1_-			2.5_+	4_+	5.5_+	7.5_+	7.5_+	(*assign signed ranks*)
10_+	12_+	12_+	13_+	14_+	15_+	16_+	17_+	18_+	19_+	20_+	21_+	22_+

$$\begin{array}{ll} W_+ = 223 & \frac{1}{2} \times 22 \times 23 = 253 \\ W_- = 30 & \textit{Check: } 223 + 30 = 253 \end{array}$$

$$E(W) = \frac{1}{4} \times 22 \times 23 = 126.5 \quad \text{and} \quad V(W) = \frac{1}{24} \times 22 \times 23 \times 45 = 948.75$$

Test statistic: $z = \dfrac{30.5 - 126.5}{\sqrt{948.75}} = -3.12$



Critical value: -1.64

Since $-3.12 < -1.64$, reject H_0 at the 5% level of significance. There is evidence to suggest that the median speed of the vehicles on the road is greater than 30 mph, and so more than 50% of the vehicles are speeding..

Exercise 16.3

Note: sums of signed ranks are often provided in questions that require a normal approximation to the Wilcoxon test. The data must still be inspected to check whether any values have been be discarded, to determine the effective sample size

1. Twenty six apples are chosen at random from a crate containing a large number of apples. The masses (to the nearest gram) are:

136	137	138	139	141	141	142	142	143	145	146	146	147
148	149	150	152	152	156	157	159	161	162	167	171	173

$$W_+ = 149.5 \qquad W_- = 175.5$$

   Use a Wilcoxon Signed Rank test, at the 5% significance level, to test whether the median mass of the apples is different than 150 grams, stating any assumptions made.

2. The lifetimes of a random sample of a particular type of candle, measured to the nearest minute, are:

268	335	339	341	341	345	346	348	352	354	355	356
357	358	359	361	362	363	364	367	371	373	374	374

$$W_+ = 91 \qquad W_- = 209$$

   The manufacturer claims that the average lifetime is at least 6 hours. Use a non-parametric test to assess whether the sample gives evidence to suggest the manufacturer's claim is not correct.

3. A local politician stated that the median hourly wage paid to young people aged 16-17 in the area is £9.20. Believing that the true median hourly wage paid is lower than this, a newspaper randomly selected 28 young people currently doing paid work and recorded their hourly wage (in £):

8.65	9.10	8.71	9.80	8.60	8.95	10.35	8.60	8.86	9.25	8.60	8.70	9.20	8.92
8.82	8.60	9.83	13.95	9.00	8.60	8.63	9.35	8.90	9.70	8.70	9.00	10.10	9.50

$$W_+ = 152 \qquad W_- = 173$$

   Use a non-parametric test to determine whether the sample supports the newspaper's suspicions.

4. A hillwalking website publishes details of popular hiking routes around Scotland. One particular route states that it takes 1 hour and 50 minutes to complete. Following upgrades to some of the paths which the route follows, the time taken to complete the route is recorded for a random sample of 32 hikers. The results of a non-parametric test conducted on the resulting data is shown below, with times recorded in minutes:

```
Wilcoxon test for the population median
data: hike times
alternative hypothesis: true location is less than 110
n = 32    sample median = 102
W = 148    p-value = 0.0154
```

   (a) State the null and alternative hypotheses for the test conducted.
   (b) Use the value of W provided in the computer output to perform a test of the above hypotheses at the 5% level of significance.
   (c) State an assumption required for the test.

Review Exercise

1. A random sample of nine primary school pupils are selected to test the number of times tables questions they can answer correctly in one minute. The results are are follows:

   24 33 36 19 37 21 16 47 20

   Perform a test to determine whether the data supports a claim that the median number of times tables questions a primary school pupil can answer in a minute is greater than 20.

2. To test whether displaying the speeds of cars travelling up a school drive on a digital display is effective in reducing driving speeds, ten cars that regularly drive up the school drive are randomly selected. For each, a speed is recorded on an occasion in which their speed is not displayed on a digital sign, and a speed is recorded when their speed is displayed. The results are as follows, in mph:

Vehicle	A	B	C	D	E	F	G	H	I	J
Speed not displayed	8	9	11	9	11	13	9	16	7	21
Speed is displayed	7	10	10	9	10	14	8	10	9	14

   Use a non-parametric tests to assess whether this suggests displaying the driving speeds of cars travelling up the school drive is effective in reducing speeds, stating an assumption required.

3. A national newspaper states that the median amount of fuel motorists put into their vehicles' tanks when stopping at a petrol station is 20 litres. To investigate this claim, a motoring magazine stops thirty people at random whilst leaving a petrol station and asks how much fuel they just bought. The results, to the nearest litre, are:

32	45	17	54	20	9	15	34	38	26
40	24	61	6	18	21	34	48	28	25
9	10	13	10	25	22	38	41	23	29

$$W_+ = 106 \qquad W_- = 329$$

   Perform a non-parametric test to assess whether there is evidence to suggest that the figure claimed by the national newspaper is incorrect, at the 5% level of significance.

4. A car company wishes to check that a new machine at their factory producing front bumpers is still maintaining the median mass of 2.8kg for the part as the old machine it replaced. Eight bumpers produced by the machine are randomly selected and weighed, in kilograms. Computer output from a test performed on the data is as follows:

```
Wilcoxon test for the population median
data: bumper mass
alternative hypothesis: true location is not equal to 2.8
W = 25    n = 7 ranks assigned    sample median = 2.95
```

   By first calculating the required test statistic, perform a non-parametric test at the 5% level of significance to determine whether there is evidence to suggest that the median mass of the part produced by the new machine has changed, stating an assumption required.

Chapter 16 Answers

For any questions for which a level of significance is not specified, a 5% level of significance has been used. Full answers should include, of course, conclusions in context

Exercise 16.1

1. $9 > 3 \implies$ do not reject H_0
2. $5.5 < 3 \implies$ reject H_0
3. $8 > 5 \implies$ do not reject H_0
4. $7.5 > 5 \implies$ do not reject H_0
5. $10 \leqslant 10 \implies$ reject H_0
6. $9 < 10 \implies$ reject H_0
7. $6.5 > 5 \implies$ do not reject H_0

Exercise 16.2

1. $8.5 < 13 \implies$ reject H_0
2. $11 > 5 \implies$ do not reject H_0
3. $46 > 21 \implies$ do not reject H_0
4. $1.5 < 2 \implies$ reject H_0
5. $4 < 8 \implies$ reject H_0

Exercise 16.3

1. $-0.34 > -1.96 \implies$ do not reject H_0
2. $-1.67 < -1.64 \implies$ reject H_0
3. $-0.27 > -1.64 \implies$ do not reject H_0
4. (a) $H_0 : \eta = 110, \quad H_1 : \eta < 110$
   (b) $-2.16 > -1.64 \implies$ reject H_0
   (c) Assume that time taken to complete the hike is symmetrically distributed.

Review Exercise

1. $5 \leqslant 5 \implies$ reject H_0
2. $14 > 8 \implies$ do not reject H_0
3. $-2.40 > -1.96 \implies$ reject H_0
4. $3 > 2 \implies$ do not reject H_0