

# 3 HYPOTHESIS TESTING: ONE SAMPLE TESTS

---

## Objectives

After studying this chapter you should

- understand when, and be able, to carry out a one sample  $t$ -test;
- be able to carry out a test for a normal population variance;
- be able to test hypotheses for a binomial population proportion, either by calculating exact probabilities or by using an appropriate approximation;
- be able to test hypotheses for the mean of a Poisson population parameter, either by calculating exact probabilities or by using a normal approximation;
- be able to carry out a sign test for a population median, either by using a binomial distribution or an approximation.

## 3.0 Introduction

In Chapter 10 of *Statistics* you were introduced to some important basic concepts of hypothesis testing:

Null hypothesis ( $H_0$ ): an assertion that a parameter in a statistical model takes a particular value, and is assumed true until experimental evidence suggests otherwise.

Alternative hypothesis ( $H_1$ ): expresses the way in which the value of a parameter may deviate from that specified in the null hypothesis, and is assumed true when the experimental evidence suggests that the null hypothesis is false.

Type 1 error: rejecting the null hypothesis when it is, in fact, true.

- Type 2 error: accepting the null hypothesis when it is, in fact, false.
- Test statistic: a function of a sample of observations which provides a basis for testing the validity of the null hypothesis.
- Critical region: the null hypothesis is rejected when a calculated value of the test statistic lies within this region.
- Critical value: the value which determines the boundary of the critical region.
- Significance level ( $\alpha$ ): the size of the critical region; the probability of a Type 1 error.
- One-tailed test: the critical region is located wholly at one end of the sampling distribution of the test statistic;  $H_1$  involves  $<$  or  $>$  but not both.
- Two-tailed test: the critical region comprises areas at both ends of the sampling distribution of the test statistic;  $H_1$  involves  $\neq$ .

### Example

A consumer group, concerned about the mean fat content of a certain grade of steakburger submits to an independent laboratory a random sample of 12 steakburgers for analysis. The percentage of fat in each of the steakburgers is as follows.

21 18 19 16 18 24 22 19 24 14 18 15

The manufacturer claims that the mean fat content of this grade of steakburger is less than 20%. Assuming percentage fat content to be normally distributed with a standard deviation of 3, carry out an appropriate hypothesis test in order to advise the consumer group as to the validity of the manufacturer's claim.

### Solution

For this problem, denoting the percentage fat content by  $X$ , then

$X \sim N(\mu, 3^2)$  and it is required to test

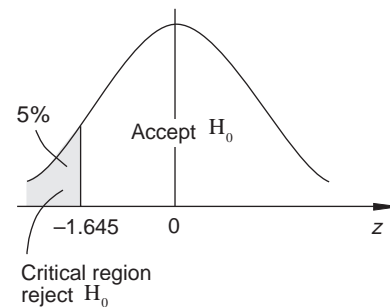
$$H_0: \mu = 20\%$$

$$H_1: \mu < 20\% \text{ (one-tailed)}$$

Significance level,  $\alpha = 0.05$  (say)

Critical region,  $z < -1.645$

$$\text{Under } H_0, \quad \bar{X} \sim N\left(20, \frac{3^2}{12}\right)$$



Now the test statistic is

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Calculation gives

$$\bar{x} = \frac{\sum x}{n} = \frac{228}{12} = 19$$

and thus 
$$z = \frac{19 - 20}{\frac{3}{\sqrt{12}}} = -1.15$$

This value does not lie in the critical region. Thus there is no evidence, at the 5% level of significance, to support the manufacturer's claim.

### 3.1 Normal population mean ( $n$ small, $\sigma$ unknown)

Consider again the previous example on steakhburgers and now suppose that it can be assumed that the percentage fat content is normally distributed but with an **unknown standard deviation**.

Now from Section 2.1 you have seen that if  $X \sim N(\mu, \sigma^2)$  then

$$\frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

is a  $t$  statistic with degrees of freedom  $\nu = n - 1$ , where

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (x - \bar{x})^2 = \frac{1}{n-1} \left[ \sum x^2 - \frac{(\sum x)^2}{n} \right].$$

So a  $z$ -test, used when  $\sigma$  is known, is replaced by a  $t$ -test when  $\sigma$  is unknown.

**What is meant by the term 'degrees of freedom' and why is its value important?**

A re-working of the steakburger problem, assuming that  $\sigma$  is now unknown, is thus as follows.

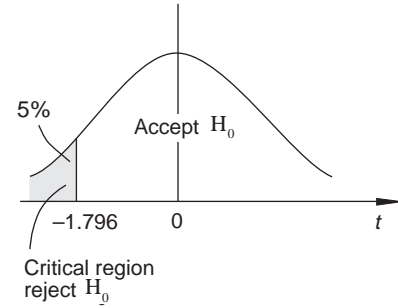
$$H_0: \mu = 20\%$$

$$H_1: \mu < 20\% \text{ (one-tailed)}$$

Significance level,  $\alpha = 0.05$  (say)

Degrees of freedom,  $\nu = n - 1 = 11$

Critical region,  $t < -1.796$



$$\text{Under } H_0, \quad \bar{X} \sim N\left(20, \frac{\sigma^2}{12}\right)$$

Now the test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

with  $\bar{x} = 19$  and  $\hat{\sigma}^2 = \frac{1}{11} \left( 4448 - \frac{228^2}{12} \right) = 10.545$  and so

$$\hat{\sigma} = 3.25.$$

$$\text{Hence } t = \frac{19 - 20}{\frac{3.25}{\sqrt{12}}} = -1.07$$

This value does not lie in the critical region. Thus there is no evidence, at the 5% level of significance, to support the manufacturer's claim.

### Activity 1 Weights and measures

Collect a random sample of between 5 and 10 'identical' items on which is stated the contents.

For example, boxes of matches, boxes of paper clips, packets of drawing pins, packets of crisps, chocolate biscuits, etc.

Count or measure, as appropriate, the contents of each item.

Investigate the hypothesis that your sample mean differs significantly from the value stated.

## Exercise 3A

- During a particular week, 13 babies were born in a maternity unit. Part of the standard procedure is to measure the length of the baby. Given below is a list of the lengths, in centimetres, of the babies born in this particular week.  
49 50 45 51 47 49 48 54 53 55 45 50 48  
Assuming that this sample came from an underlying normal population, test, at the 5% significance level, the hypothesis that the population mean length is 50 cm. (AEB)
- A random sample of 12 steel ingots was taken from a production line. The masses, in kilograms, of these ingots are given below.  
24.8 30.8 28.1 24.8 27.4 22.1  
24.7 27.3 27.5 27.8 23.9 23.2  
Assuming that this sample came from an underlying normal population, investigate the claim that its mean exceeds 25.0 kg. (AEB)
- A random sample of 14 cows was selected from a large dairy herd at Brookfield Farm. The milk yield in one week was recorded, in kilograms, for each cow. The results are given below.  
169.6 142.0 103.3 111.6 123.4 143.5 155.1  
101.7 170.7 113.2 130.9 146.1 169.3 155.5  
Stating clearly any distributional assumptions that you make, investigate the claim that the mean weekly milk yield for the herd is greater than 120 kg. (AEB)
- A random sample of 15 workers from a vacuum flask assembly line was selected from a large number of such workers. Ivor Stopwatch, a work-study engineer, asked each of these workers to assemble a one-litre vacuum flask at their normal working speed. The times taken, in seconds, to complete these tasks are given below.  
109.2 146.2 127.9 92.0 108.5  
91.1 109.8 114.9 115.3 99.0  
112.8 130.7 141.7 122.6 119.9  
Assuming that this sample came from an underlying normal population, investigate the claim that the population mean assembly time is less than 2 minutes. (AEB)
- In processing grain in the brewing industry, the percentage extract recovered is measured. A particular brewing introduces a new source of grain and the percentage extract on 11 separate days is as follows.  
95.2 93.1 93.5 95.9 94.0 92.0  
94.4 93.2 95.5 92.3 95.4  
Test the hypothesis that the true mean percentage extract recovered is 95.0. What assumptions have you made in carrying out your test? (AEB)
- A car manufacturer introduces a new method of assembling a particular component. The old method had a mean assembly time of 42 minutes. The manufacturer would like the assembly time to be as short as possible, and so he expects the new method to have a smaller mean. A random sample of assembly times (minutes) taken after the new method had become established was  
27 39 28 41 47 42 35 32 38  
Stating any necessary distributional assumptions, investigate the manufacturer's expectation. (AEB)

## 3.2 Normal population variance

In Section 2.2 you have seen that if  $\hat{\sigma}^2$  denotes the variance [divisor  $(n-1)$ ] of a random sample of size  $n$  from a normal population with variance  $\sigma^2$ , then

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2$$

with degrees of freedom,  $\nu = n - 1$ .

The statistic  $\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$  may thus be used to test hypotheses concerning a normal population variance or, by implication, a normal population standard deviation.

How does the  $\chi^2$  distribution differ in shape from the t-distribution and how does this affect the percentage points of the  $\chi^2$  distribution?

### Example

A user of a certain gauge of steel wire suspects that the standard deviation of its breaking strength, in newtons (N), is different from the value of 0.75 as specified by the manufacturer.

Consequently the user tests the breaking strength of each of a random sample of nine lengths of wire and obtains the following results.

72.1 74.5 72.8 75.0 73.4 75.4 76.1 73.5 74.1

Assuming breaking strength to be normally distributed, test, at the 10% level of significance, the manufacturer's specification.

### Solution

$$H_0: \sigma = 0.75 \text{ N or } H_0: \sigma^2 = 0.5625 \text{ N}^2$$

$$H_1: \sigma \neq 0.75 \text{ N or } H_1: \sigma^2 \neq 0.5625 \text{ N}^2 \text{ (two-tailed)}$$

Significance level,  $\alpha = 0.10$

Degrees of freedom,  $\nu = 8$

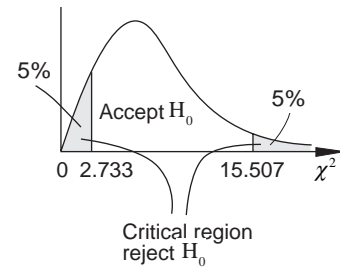
Critical region,  $\chi^2 < 2.733$  or  $\chi^2 > 15.507$

Calculation gives

$$\begin{aligned} (n-1)\hat{\sigma}^2 &= \sum x^2 - \frac{(\sum x)^2}{n} \\ &= 49430.49 - \frac{666.9^2}{9} = 13.2 \end{aligned}$$

$$\text{Under } H_0, \chi^2 = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{13.2}{0.5625} = 23.47$$

This value does lie in the critical region. Thus there is evidence, at the 10% level of significance, to dispute the manufacturer's specification as regards variability of breaking strength.



## Activity 2 Random standardised normal deviates

---

Obtain from tables, calculator, or computer a random sample of between 10 and 15 values of the standardised normal random variable.

Investigate the hypothesis that your sample provides significant evidence that the variance differs from unity.

---

## Exercise 3B

- Given the data in Question 1 of Exercise 3A, test, at the 5% level of significance, the hypothesis that the population standard deviation of length is 2.5 cm. (AEB)
- Referring to Question 2 of Exercise 3A, investigate the claim that the population standard deviation is greater than 2 kg. (AEB)
- Using the data given in Question 3 of Exercise 3A, test the hypothesis that the standard deviation of yield for the herd at Brookfield Farm is 20 kg. (AEB)
- The variability in unit mass is very critical in the pharmaceutical industry. The mass, in grams, of each of 15 randomly selected units of a specific drug is as follows.  
3.48 3.52 3.50 3.47 3.49 3.54 3.51 3.52  
3.46 3.45 3.55 3.48 3.51 3.52 3.50  
Assuming that this sample comes from an underlying normal population, test the hypothesis that the variance of the population exceeds  $0.002 \text{ gm}^2$ . (AEB)
- With reference to Question 5 of Exercise 3A, the variance of percentage extract recovered was 6.0 prior to introducing the new source of grain. Does the new source of grain result in a decrease in the variability of percentage extract recovered? (AEB)
- Referring to Question 6 of Exercise 3A, the manufacturer assumes that the new method of assembly will reduce the variability of assembly times. Given that the old assembly method had a standard deviation of 8 minutes, investigate the manufacturer's assumption. (AEB)

## 3.3 Binomial population proportion

In situations where the population parameter of interest in a hypothesis test is a proportion, or percentage, rather than a mean or variance, then a binomial distribution, or an appropriate approximation, provides the basis for the test.

Remember from Chapter 5 of *Statistics*, that if  $X$  denotes the number of successes in  $n$  repeated trials, each of which may result in a success with probability  $p$ ,

then  $X \sim B(n, p)$

and  $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$

where  $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

### Example

Until recently, an average of 60 out of every 100 patients have survived a particularly severe infection. When a new drug was administered to a random sample of 15 patients with the infection, 12 survived. Does this provide evidence that the new drug is effective?

### Solution

Let  $X$  denote the number of patients who survive.

Then  $X \sim B(15, p)$ , and it is required to test

$$H_0: p = 0.6 \quad (\text{not effective})$$

$$H_1: p > 0.6 \quad (\text{effective; one-tailed})$$

Now assuming  $H_0$  is true,  $X \sim B(15, 0.6)$ , and so

$$P(X = 12) = \binom{15}{12} 0.6^{12} 0.4^3 = 0.06339$$

$$P(X = 13) = \binom{15}{13} 0.6^{13} 0.4^2 = 0.02194$$

$$P(X = 14) = \binom{15}{14} 0.6^{14} 0.4^1 = 0.00470$$

$$P(X = 15) = \binom{15}{15} 0.6^{15} 0.4^0 = 0.00047$$

For a 5% level of significance, what is the critical region?

Hence, under  $H_0$ , the probability of

$$15 \text{ patients surviving} = 0.00047$$

$$14 \text{ or } 15 \text{ patients surviving} = 0.00517$$

$$13, 14 \text{ or } 15 \text{ patients surviving} = 0.02711$$

$$12, 13, 14 \text{ or } 15 \text{ patients surviving} = 0.09050$$



Note that if 14 surviving is adopted as a 'significant' result, the probability of 15 must be included as well (because 15 is actually a 'better' result than 14). Similarly with 13, the probabilities for 14 or 15 must be included, and so on.

From the above (cumulative) probabilities it can be seen that the critical region for say a 5% test is  $X \geq 13$

since  $P(X \geq 13) = 0.02711 < 0.05$

but  $P(X \geq 12) = 0.09050 > 0.05$

The number of patients who actually survived was 12. This value does not lie in the critical region. Thus there is no evidence, at the 5% level of significance, to suggest that the drug is effective.

Note that rather than actually determine the critical region, it is equally valid to simply argue that since

$P(X \geq 12) = 0.09050 > 0.05$ , then  $H_0$  cannot be rejected.

Remember also that tables of the Cumulative Binomial Distribution Function may often provide an easier and quicker alternative for calculating probabilities or finding critical regions.

### **Activity 3**      **Can you tell the difference?**

---

Can you tell HP Baked Beans from a supermarket brand? Can you tell Coca Cola from a supermarket brand? Can you tell Stork from butter?

You are going to set up an experiment to determine whether people really can tell the difference between two similar foods or drinks.

Each person taking part in the test is given 3 samples; 2 of one product and 1 of another (so they may have 2 beakers containing Coca Cola (say) and 1 beaker containing a supermarket brand or vice-versa).

Ask each person to try to identify the sample which is different from the other two.

Note that there are 6 possible groupings of the 3 samples,

ABB BAB BBA BAA ABA AAB

and a die may be used to decide which grouping to give each individual person taking part.

Plan the experiment carefully before you start. Write out a list showing the groupings for each of your participants (about 12, say).

Ensure that your participants take the test individually in quiet surroundings, free from odours. All 3 samples must be of the same size and temperature. If there are any differences in colour then you can blindfold your participants. Record each person's answer. Count the number of persons giving the correct answer. Participants who are unable to detect any difference at all in the 3 samples must be left out of the analysis.

Use your results to test the hypotheses

$$H_0: p = \frac{1}{3} \quad (\text{participants are simply guessing})$$

$$H_1: p > \frac{1}{3} \quad (\text{participants can distinguish})$$

---

Under what conditions does a Poisson distribution provide a suitable approximation to a binomial distribution?

### Example

A machine which manufactures black polythene dustbin bags, is known to produce 3% defective bags. Following a major breakdown of the machine, extensive repair work is carried out which may result in a change in the percentage of defective bags produced. To investigate this possibility, a random sample of 200 bags is taken from the machine's production and a count reveals 12 defective bags. What may be concluded?

### Solution

Here  $n = 200$ ,  $p =$  population proportion of defective bags produced, and it is required to test

$$H_0: p = 0.03 \quad (\text{no change})$$

$$H_1: p \neq 0.03 \quad (\text{change; two-tailed})$$

Significance level,  $\alpha = 0.05$ (say)

Again from Chapter 6 of *Statistics*, a Poisson distribution with  $\lambda = np$  provides an approximation to a binomial distribution when  $n \geq 50$  and  $p \leq 0.1$  (or  $p \geq 0.9$ ).

Thus if  $X$  denotes the number of defective bags in the sample, then under  $H_0$

$$X \sim Po(\lambda = 200 \times 0.03 = 6.0)$$

Using tables of the Cumulative Poisson Distribution Functions,

$$\begin{aligned} P(X \geq 12) &= 1 - P(X \leq 11) \\ &= 1 - 0.9799 = 0.0201 \end{aligned}$$

Since the test is two-tailed, this probability is compared with

$\frac{\alpha}{2} = 0.025$ , because a small number of defective bags could also lead to the rejection of the null hypothesis.

Here,  $0.0201 < 0.025$  so  $H_0$  is rejected. There is evidence, at the 5% level of significance, that the percentage of defective bags produced has changed following the repair work.

*Under what conditions does a normal distribution provide a suitable approximation to a binomial distribution?*

### Example

Company A proposes the take-over of Company B. The latter's Chief Executive claims that her Company's shareholders are equally divided for and against the take-over on the basis of the terms offered. However, the Chairman of Company A claims that more than half of Company B's shareholders are in favour of accepting his Company's offer.

To investigate these two rival claims, the view of each of 400 randomly selected shareholders of Company B is sought. A subsequent count reveals that 219 are in favour of the offer; the remainder are against.

Does this provide evidence, at the 1% significance level, that the claim made by Company B's Chairman is valid?

### Solution

Let  $p$  denote the actual proportion of Company B's shareholders who are in favour of the offer. Then

$$H_0: p = 0.50 \text{ (claim of Company B's Chief Executive)}$$

$$H_1: p > 0.50 \text{ (claim of Company A's Chairman; one-tailed)}$$

Significance level,  $\alpha = 0.01$

If  $X$  denotes the number of Company B's shareholders in the sample who are in favour of the offer, then under  $H_0$

$$\begin{aligned} X &\sim B(400, 0.50) \quad \text{with mean, } np = 200 \text{ and} \\ &\quad \text{variance, } np(1-p) = 100 \end{aligned}$$

From Chapter 8 of *Statistics*, a normal distribution provides an approximation to a binomial distribution when  $n \geq 50$  and  $0.1 \leq p \leq 0.9$ .

Thus under  $H_0$ ,  $X \sim N(200,100)$ , or denoting the sample

proportion by  $\hat{P} = \frac{X}{n}$  then

$$\hat{P} \sim N\left(\frac{200}{400}, \frac{100}{400^2}\right)$$

i.e.  $\hat{P} \sim N(0.50, 0.000625)$

So Critical region is  $z > 2.326$

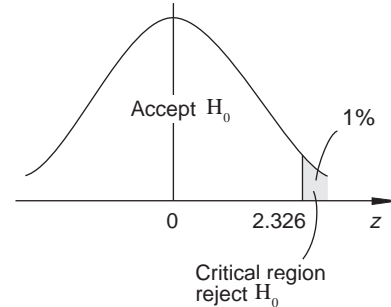
Now value of sample proportion,

$$\hat{p} = \frac{219}{400} = 0.5475$$

so test statistic,

$$z = \frac{0.5475 - 0.50}{\sqrt{0.000625}} = 1.90$$

This value does not lie in the critical region. Thus there is no evidence, at the 1% level of significance, to support the claim of Company A's Chairman.



### Exercise 3C

1. A pharmacist claims that more than 60% of all customers simply collect a prescription. One of her assistants notes that, in a random sample of 12 customers, 10 simply collected a prescription. Does this provide sufficient evidence, at the 5% level, to support the pharmacist's claim?
2. In a survey carried out in Funville, 14 children out of a random sample of 30 said that they bought the Bopper comic regularly. Test, at the 10% level of significance, the hypothesis that the true proportion of all children who buy this comic regularly is 0.35. (AEB)
3. A random sample of 30 coffee drinkers were each asked to taste-test a new brand of coffee. The responses are listed below with L representing 'like', I representing 'indifferent', and D representing 'dislike'.
 

L	D	L	L	D	L	L	L	L	I
L	L	L	I	L	D	L	I	L	L
I	L	L	L	D	L	L	L	L	I
4. Tins of baked beans are packed in boxes of 24. Results from a random sample of 25 boxes delivered to supermarkets show that a total of 8 tins were damaged. Assess the claim that less than 2% of tins are damaged during delivery. (AEB)
5. A survey of Aldervale revealed that 223 houses out of a random sample of 500 were fitted with some form of double glazing. Test the hypothesis that 40% of all houses in Aldervale have some form of double glazing. (AEB)
6. Prior to joining a coaching course, a netball player has scored with 45% of her shots at the basket, and so is considered to have a scoring ability of 0.45. In games following the course, she scores with 72 of her 135 shots at the basket. Stating clearly any assumptions made, investigate whether or not the course has improved her shooting ability.

Do these data support the claim that more than half of all coffee drinkers like this new brand of coffee?

## 3.4 Poisson population mean

In the previous Section it was seen how a Poisson distribution may be used as an approximation to a binomial situation when testing hypotheses concerning a population proportion. However, there are situations for which a Poisson distribution is appropriate in its own right when testing hypotheses concerning a population mean value.

*Under what conditions would you expect a Poisson distribution to provide a suitable model for a random variable?*

Recall from Chapter 6 of *Statistics*, that if  $X \sim Po(\lambda)$  then

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, 3, \dots,$$

with mean =  $\lambda$  and variance =  $\lambda$ .

*What do you know about the sum of two independent Poisson variables?*

### Example

The number of faults in one metre of a particular type of thread is known to have a Poisson distribution. It is claimed that the average number of faults is 0.02 per metre. A random sample of 100 one metre lengths of the thread reveals a total of 6 faults. Does this information support the claim?

### Solution

Remembering that if  $A \sim Po(a)$  independent of  $B \sim Po(b)$  then  $C = A + B \sim Po(a + b)$ , and denoting the total number of faults in the 100 one metre lengths by  $X$ , then if the claim is valid

$$X \sim Po(100 \times 0.02 = 2)$$

It is thus required to test

$$H_0: \lambda = 2$$

$$H_1: \lambda \neq 2 \quad (\text{two-tailed})$$

Significance level,  $\alpha = 0.05$  (say)

The observed number of faults is 6, and using tables of the Cumulative Poisson Distribution Function with  $\lambda = 2$

$$\begin{aligned} P(X \geq 6) &= 1 - P(X \leq 5) \\ &= 1 - 0.9834 = 0.0166 < 0.025 \quad (\text{two-tailed}) \end{aligned}$$

There is evidence, at the 5% level of significance, that the claim is not valid.

### Activity 4 Simulated traffic flow

---

The random variable  $X$  denotes the number of vehicles in a 30 second interval passing a traffic check point on a country road.

Use a three-digit random number,  $y$ , to generate an 'observed' value,  $x$ , of  $X$  using the following rules.

$000 \leq y \leq 548$	$x = 0$
$549 \leq y \leq 877$	$x = 1$
$878 \leq y \leq 976$	$x = 2$
$977 \leq y \leq 996$	$x = 3$
$997 \leq y \leq 999$	$x = 4$

Using the above rules, obtain 5 'observed' values of  $X$ .

Use the total of these values to test the hypothesis that the average number of vehicles passing the traffic check point in 30 seconds is 0.60.

Repeat the test for 10, 15 and 20 simulated 'observed' values of  $X$ .

Consider the conclusions of your four tests.

---

For a Poisson distribution with a mean greater than 15, a normal distribution with  $\mu = \sigma^2 = \lambda$  provides a good approximation (see Chapter 8 of *Statistics*).

### Example

Prior to extensive modernisation of its forecourt, the average number of vehicles calling for fuel at a small garage has been 5 per hour. Following the modernisation, a total of 543 vehicles called for fuel in a random sample of 100 one-hour intervals. Has the modernisation of the garage's forecourt significantly increased the average number of vehicles calling for fuel?

### Solution

If  $X$  denotes the total number of vehicles calling for fuel in the 100 one-hour intervals, then under the null hypothesis of no increase

$$X \sim Po(100 \times 5 = 500) \approx N(500, 500)$$

The test is thus as follows.

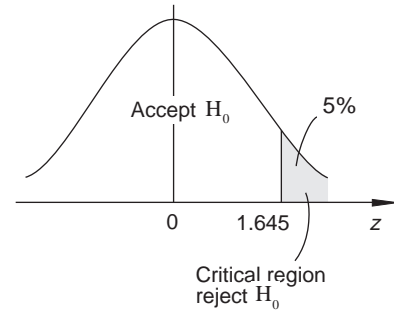
$$H_0: \lambda = 500$$

$$H_1: \lambda > 500 \quad (\text{one-tailed})$$

Significance level,  $\alpha = 0.05$  (say)

Critical region is  $z > 1.645$

$$\text{Test statistic, } z = \frac{543 - 500}{\sqrt{500}} = 1.92$$



This value does lie in the critical region. Thus there is evidence, at the 5% level of significance, to suggest that the forecourt modernisation has resulted in an increase in the average number of vehicles calling for fuel.

### Exercise 3D

- In the manufacture of commercial carpet on a particular machine, small faults occur at random in the carpet at an average rate of 0.925 per 25 m<sup>2</sup>. Following an overhaul of the machine, an inspection of a random sample of four 5 m × 5 m squares of carpet reveals only 2 small faults. Is there evidence, at the 5% level of significance, that, following the overhaul, the mean number of small faults has been reduced?
- It is known that the numbers of parasites on fish in a pond follow a Poisson distribution. The total number of parasites on a random sample of 3 fish taken from the pond is 8. Test the hypothesis that the mean number of parasites per fish is 4. (AEB)
- The daily number of letters of complaint from customers received by a department store follows a Poisson distribution. Over a 150-day period, a total of 407 letters of complaint were received. Investigate, at the 1% level of significance, the claim that an average of fewer than 3 letters of complaint per day are received. What assumption is needed regarding the 150-day period? (AEB)
- It has been established over a period of time that the weekly numbers of breakages of crockery and glassware in a hotel follow a Poisson distribution with mean 27. On appointment, a new manager announces to staff that unless this mean level is reduced, bonuses will suffer as a result. Following this announcement, the weekly numbers of breakages over an eight-week period were as follows.  
23 19 17 20 25 32 22 26  
Do these data provide evidence that the manager's announcement has been heeded by the staff?

## 3.5 Population median

In Chapter 3 of *Statistics*, the median,  $m$ , of  $n$  items of data is

defined as the  $\left\{ \left( \frac{n+1}{2} \right) \right\}$ th item of data; or, in other words, the middle item.

For a random variable,  $X$ , with median  $\eta$ , this leads to the result

$$P(X < \eta) = P(X > \eta) = 0.5$$

or 
$$P(X - \eta < 0) = P(X - \eta > 0) = 0.5$$

or 
$$P(X - \eta \text{ is negative}) = P(X - \eta \text{ is positive}) = 0.5$$

These equally probable positive and negative values lead to a test for a hypothesised value of a population median based upon a binomial distribution with  $p = P(+ \text{ sign}) = 0.5$  under  $H_0$ .

Note however that, contrary to all the other tests in this chapter, no assumption needs to be made about the distribution of  $X$  itself. As a result this sign test for a population median is an example of a **non-parametric** test.

Tests which require a knowledge of, or assumption about, the distribution of the parent population (e.g. normal, binomial, Poisson) are called **parametric** tests.

### Example

The values below are the scores (maximum 20) obtained in an aptitude test by a random sample of 11 graduates. It is known that for the non-graduate population the median score is 12. Is there evidence, at the 10% significance level, that graduates achieve a higher median score than the non-graduate population?

14 15 9 10 10 13 14 19 12 16 13

### Solution

$$H_0: \eta = 12$$

$$H_1: \eta > 12 \quad (\text{one-tailed})$$

$$\text{Significance level, } \alpha = 0.10$$

Signs of (score - 12) are

+ + - - - + + + 0 + +

Let  $X$  denote the number of + signs. Then, ignoring the one 0 in this case, under  $H_0$ ,

$$X \sim B(10, 0.50) \text{ with observed value of } X = 7.$$

Now 
$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.8281 = 0.1719 > 0.10$$

Thus there is no evidence, at the 10% level of significance, to suggest that graduates achieve a higher median score than the non-graduate population.



What additional information would be required in the above example in order to test  $H_0: \mu = 12$  versus  $H_1: \mu > 12$ ?

### Activity 5 Weights and measures re-visited

Using your data collected for Activity 1, test the hypothesis that the true median value for the contents is equal to the value stated as contents.

Explain the reasons for the similarity, or difference, between your conclusions for the two activities.

### Exercise 3E

1. The following values are the annual salaries, in £, of a random sample of 10 recent statistics' graduates.

13 250   7 485   15 136   12 258   11 019  
14 268   19 536   14 326   16 326   17 984

Investigate the claim, at the 10% significance level, that the median annual salary of all recent statistics' graduates exceeds £12 000.

2. Part of the assessment of a statistics course involves small groups of students giving a short presentation of their findings. The grades obtained by 13 such groups are as follows (highest = A, lowest = F).

C B A B B C D C F B A B B

Test the hypothesis that the median grade for all such group presentations is C.

3. 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.

42 35 24 30 32 42 56 35  
34 30 29 41 38 38 30 29  
34 39 43 72 38 40 30 62

Investigate the resident's claim.

4. During a working day a machine may require various adjustments. The machine's operator is asked to record the number of adjustments made to the machine each day for a period of 250 working days. She obtains the data displayed in the table below.

Number of adjustments	0	1	2	3	4	5	$\geq 6$
Number of days	61	54	65	18	12	9	31

Test the hypothesis that the true median daily number of adjustments is 1.

Explain why a test for the true mean daily number of adjustments would be difficult.

### 3.6 Miscellaneous Exercises

1. An investigation was conducted into the dust content in the flue gases of a particular type of solid-fuel boiler. Thirteen boilers were used under identical fuelling and extraction conditions. Over a similar period, the following quantities, in grams, of dust were deposited in traps inserted in each of the thirteen flues.

73.1 56.4 82.1 67.2 78.7 75.1 48.0  
53.3 55.5 61.5 60.6 55.2 63.1

Stating any necessary distributional assumptions, test, at the 5% level of significance, the hypothesis that

- (a) the population variance of dust deposit is  $35 \text{ g}^2$ ,  
(b) the population mean dust deposit is 60 g. (AEB)
2. As part of a research project, a random sample of 11 students sat a proposed national Physics examination and obtained the following percentage marks.
- 30 44 49 50 63 38 43 36 54 40 26
- (a) Use a  $t$ -test, at the 10% significance level, to test the hypothesis that the true mean examination mark is 40%.  
(b) Use a sign test, at the 10% level, to test the hypothesis that the median mark for all candidates is 40%.

What assumption is needed for (a) but not for (b), and do you think it is justified? (AEB)

3. Smallwoods Ltd run a weekly football pools competition. One part of this involves a fixed-odds contest where the entrant has to forecast correctly the result of each of five given matches. In the event of a fully correct forecast the entrant is paid out at odds of 100 to 1. During the last two years Miss Fortune has entered this fixed-odds contest 80 times. The table below summarises her results.

<b>Number of matches correctly forecast per entry</b> ( $x$ )	0	1	2	3	4	5
<b>Number of entries with <math>x</math> correct forecasts</b> ( $f$ )	4	13	27	22	13	1

Assuming that the number of matches correctly forecast per entry follows a binomial distribution, test the hypothesis that the probability of Miss Fortune forecasting correctly the result of a single match is 0.5.

On the evidence before you, and assuming that the point of entering is to win money, would you advise Miss Fortune to continue with this competition, and why? (AEB)

4. At a nuclear power station great care is taken to monitor employees' state of health. The table below gives the number of visits made by each of 10 employees from the reactor room to their general practitioners during one calendar year.

3 6 5 7 4 2 3 5 1 4

The number of such visits made by a member of the general public is known to have a Poisson distribution with mean 3. Assuming that the numbers of visits made by the nuclear power station's employees are also distributed as a Poisson random variable, test the hypothesis that the annual mean per employee is greater than 3. (AEB)

5. A random sample of 18 female computer operators each had their diastolic blood pressure measured to the nearest millimetre with the following results.

57 64 77 82 66 94 72 83 61  
72 65 89 54 73 55 67 71 68

- (a) Stating any necessary distributional assumptions, show that there is no significant evidence to reject the hypothesis that the standard deviation of female computer operators' diastolic blood pressures is equal to that of the female population, namely 10 mm.  
(b) Making appropriate use of the statement in (a), investigate the claim that the mean diastolic blood pressure of female computer operators differs from 75 mm. (AEB)
6. The number of misprints per page in each national daily newspaper is known to follow a Poisson distribution. The mean number per page for the tabloid newspapers is 4. The Daily Planet, a quality newspaper, claims that although its pages are much bigger and contain more text than those of the tabloid papers, its mean number of misprints per page is certainly fewer. A random sample of 16 pages from recent editions of the Daily Planet results in the following numbers of misprints.  
0, 3, 2, 3, 1, 4, 5, 2, 3, 4, 5, 2, 4, 3, 3, 1. Investigate, at the 1% significance level, the Daily Planet's claim. (AEB)
7. Explain what is meant by the following terms when used in the context of a hypothesis test.  
(a) Null and alternative hypotheses.  
(b) Type 1 and Type 2 errors.  
(c) One-tailed and two-tailed tests.
- A new dietary treatment for a severe allergy is claimed to have a better cure rate than the accepted value of 60% for the well-established standard drug treatment. A random sample of 20 patients, suffering from the allergy, are given the new dietary treatment and as a result 17 are cured. Is the claim valid? (AEB)

8. A large food processing firm was considering introducing a new recipe for its ice cream. In a preliminary trial, a panel of 15 tasters were each asked to score the ice cream on a scale from 0 (awful) to 20 (excellent). Their scores were as follows.

16 15 17 3 18 15 18 7  
12 14 6 4 19 14 17

The mean score in a similar trial for the firm's existing ice cream was 14. Test the hypothesis that the mean score for the new ice cream was no different.

Because of the erratic nature of the scores obtained, doubt was expressed as to the assumption that scores were normally distributed. As a result it was suggested that a sign test for a median was perhaps more appropriate. Given that the median score in the trial for their existing ice cream was 13, what can be concluded? (AEB)

9. The external diameter, in centimetres, of each of a random sample of 10 pistons manufactured on a particular machine was measured with the results below.

9.91, 9.89, 10.06, 9.98, 10.09,  
9.81, 10.01, 9.99, 9.87, 10.09.

Stating any necessary assumptions, test the two distinct claims that piston rings manufactured on this machine have a mean external diameter of 10 cm with a variance of 0.005 cm<sup>2</sup>. (AEB)

10. A sweet shop sells chocolates which appear, at first sight, to be identical. Of a random sample of 80 chocolates, 61 had hard centres and the rest soft centres. Test the hypothesis that 70% of chocolates have hard centres.

The chocolates are all in the shape of circular disks and the diameters, in millimetres, of the 19 soft centred chocolates were as follows.

279 263 284 277 281 269 266  
271 262 275 266 272 281  
274 279 277 267 269 275

Assuming that the diameters of the soft centred chocolates are normally distributed, test, at the 10% significance level, the hypothesis that their mean diameter is 275 mm.

What changes would you make to your test if it was known that the standard deviation of the diameters of soft centred chocolates was 5 mm? (AEB)

11. It is known that repeated weighings of the same object on a particular chemical balance give readings which are normally distributed. Past evidence, using experienced operators, suggests that the mean is equal to the mass of the object and that the standard deviation is 0.25 mg.

A trainee operator makes seven repeated weighings of the same object, which is known to have a mass of 19.5 mg, and obtains the following readings.

19.1 19.4 19.0 18.8 19.7 19.8 19.3

Is there any evidence that these results are more variable than those obtained by experienced operators?

Hence investigate whether or not the trainee operator's readings are biased. (AEB)

12. As part of a statistics project, students observed five private cars passing a college and counted the number which were carrying the driver only, with no passengers. This was repeated 80 times. The results for a particular student were as follows.

<b>Number of cars with driver only</b>	0	1	2	3	4	5
<b>Number of times observed</b>	0	3	12	27	26	12

Explain why the distribution of the number of cars with a driver only could be modelled by a binomial random variable. Assuming such a model is appropriate, state the value of the parameter  $n$  and estimate the value of the parameter  $p$ .

Hence investigate the claim that more than 60% of cars contain the driver only. (AEB)

13. Explain **each** of the terms *significance level*, *critical region* and *test statistic* as used in hypothesis testing.

The manager of a road haulage firm records the following times taken, in minutes, by a lorry to travel from the depot to a particular customer's factory.

43 35 47 180 39 58 40 39 51

The journey time of 3 hours was as a result of the driver being stopped by Customs & Excise Inspectors. The manager therefore removes this value before passing the data to you, as the firm's statistician, for analysis. Use the eight remaining values to test the hypothesis that the true mean journey time is 40 minutes.

Comment on the manager's decision to remove the value of 3 hours and state what assumption may have been violated if this value had been included.

What alternative test would you have considered if the manager had requested that all nine results should be analysed? (AEB)

14. A car insurance company found that the average amount it was paying on bodywork claims in 1992 was £435 with a standard deviation of £141. The first six bodywork payments, in £, in 1993 were  
 548 209 534 198 789 633.  
 Stating clearly all necessary assumptions, has there been a significant change in the average and in the variability of payments in 1993 as compared to 1992? (AEB)
15. (a) Before its annual overhaul, the mean operating time of an automatic machine was 100 seconds. After the overhaul, the following random sample of operating times, in seconds, was obtained.  
 90 97 101 92 101 95 95 98 96 95  
 Assuming that the time taken by the machine to perform the operation is a normally distributed random variable, test the hypothesis that the overhaul has improved the machine's operation.
- (b) The results of a survey showed that 3615 out of 10 000 families regularly purchased a specific weekly magazine. Test, at the 1% level of significance, the claim that more than 35% of families buy the magazine. (AEB)
16. A nurseryman decided to keep records of the first year's growth of his pine seedlings. On the first occasion he found a mean growth of 11.5 cm with a standard deviation of 2.5 cm. The following year he used an experimental soil preparation for all his seedlings and the first year's growth of a random sample of eight of the seedlings was  
 7, 23, 19, 25, 11, 18, 17 and 15 cm.
- (a) Assuming these data may be regarded as a random sample from a normal distribution, test at the 5% significance level, whether there has been a change in  
 (i) the standard deviation,  
 (ii) the mean.
- (b) Explain to the nurseryman why the conclusions in (a) cannot necessarily be attributed to the new soil preparation. (AEB)
17. The development engineer of a company making razors records the time it takes him to shave, on seven mornings, using a standard razor made by the company. The times, in seconds, were  
 217, 210, 254, 237, 232, 228, 243.  
 Assuming that this may be regarded as a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$  test, at the 5% level of significance, the hypothesis that  
 (a)  $\sigma = 10$  seconds,  
 (b)  $\mu = 240$  seconds. (AEB)
18. Packets of ground filter coffee have a nominal weight of 200 g. The distribution of weights may be assumed to be normal. A random sample of 30 packets had the following weights.  
 218 207 214 189 211 206 203 217 183 186  
 219 213 207 214 203 204 195 197 213 212  
 188 221 217 184 186 216 198 211 216 200  
 Investigate the assumption that the mean weight of all packets is 200 g.  
 Test the hypothesis that 15% of packets weigh less than 190 g. (AEB)
19. (a) The number of accidents per day on a stretch of motorway is known to have a Poisson distribution. Police claim that there is an average of more than one accident per day. Over a particular 7-day period there is a total of 13 accidents. Investigate the claim of the police.
- (b) The results of a survey to establish the attitude of individuals to a particular proposal showed that three quarters of those interviewed were house owners. Of the 200 interviewed, only 12 of the 70 in favour of the proposal were not house owners. Test the hypothesis that the percentage of house holders in favour of the proposal is 40. (AEB)
20. The resistances (in ohms) of a sample from a batch of resistors were as follows.  
 2314 2456 2389 2361 2360 2332 2402  
 Stating any necessary assumptions, test the hypothesis that the true standard deviation of the resistors is 30 ohms.  
 Hence test the hypothesis that the actual mean resistance of the resistors is 2400 ohms. (AEB)
21. Explain the difference between *parametric* and *non-parametric* hypothesis tests, and give an example of each type.  
 A local authority offers all of its employees regular health checks. As part of the check, several physiological measurements are taken on each person.  
 (a) The ordered scores for one of the measurements on nine employees were as follows.  
 9 11 50 54 58 69 76 91 95  
 Test the hypothesis that the median score for this measurement is 50. (AEB)
- (b) For another measurement, a random sample of 100 employees had 19 scores of exactly 50 and 30 scores above 50. Investigate the claim that the median score for this measurement is less than 50. (AEB)

22. Experimental components for use in aircraft engines were tested to destruction under extreme conditions. The survival times,  $X$  days, of ten components were as follows.

207 381 411 673 534 294 697 344 418 554

- (a) Assuming that the survival time, under these conditions, for all the experimental components is normally distributed, assess the validity of the claim that the mean survival time exceeds 400 days.
- (b) Given that the standard deviation of survival times is known to be 150 days, what changes would you make to your test procedure in (a)? (AEB)

23. The external diameters (measured in units of 0.01 mm above nominal value) of a sample of piston rings produced on the same machine were

11 9 32 18 29 11 21 19 6

- (a) Assuming a normal distribution, and given that the target value is 20 and a standard deviation of at most 10 is acceptable, carry out appropriate hypothesis tests so as to comment on the performance of the machine.
- (b) It was later discovered that an error had been made in zeroing the measuring device and that all the measurements in the sample should be increased by 12. How, if at all, does this affect your conclusions in (a)? (AEB)

24. Explain **each** of the terms *null hypothesis*, *critical region* and *test statistic* as used in hypothesis testing.

Employees of a firm carrying out motorway maintenance are issued with brightly coloured waterproof jackets. These come in different sizes numbered 1 to 5. The last 40 jackets issued were of the following sizes.

2	3	3	1	3	3	2	4	3	2
5	4	1	2	3	3	2	4	5	3
2	4	4	1	5	3	3	2	3	3
1	3	4	3	3	2	5	1	4	4

Assuming that the 40 employees may be regarded as a random sample of all employees, test the hypothesis, at the 5% significance level, that 40% of all employees require size 3.

Test the claim that size 3 is the median size. (AEB)

