Tests Concerning a Single Sample

41.2

Introduction

This Section introduces you to the basic ideas of hypothesis testing in a non-mathematical way by using a problem solving approach to highlight the concepts as they are needed. We only consider situations involving a single sample. In Section 41.3 we will introduce you to situations involving two samples and while the basic ideas will follow through, their practical application is a little more complex than that met here. However, once you have learned how to apply the basic ideas of hypothesis testing covered in this Workbook, you should be capable of applying hypothesis testing to a very wide range of practical problems and learning about methods of hypothesis testing which are not covered here.

Prerequisites

Before starting this Section you should . . .

① be familiar with the results and concepts met in the study of probability

② be familiar with a range of statistical distributions

③ be familiar with the contents of Section 41.1

Learning Outcomes

After completing this Section you should be able to . . .

✓ be able to apply the ideas of hypothesis testing to a range of problems underpinned by elementary statistical distributions and involving only a single sample.

1. Tests of Proportion

Problem 1

SwitchRight, a manufacturer of engine management systems requires its supplier of control modules to supply modules with at least 99% complying with their specification. The quality control operators at SwitchRight check a random sample of 1000 control modules delivered to SwitchRight and find that 985 match the specification. Does this result imply that less than 99% of the control modules supplied do not match SwitchRights specification?

Analysis

Firstly, we set up two hypotheses concerning the control modules. The first hypothesis, called the null hypothesis is denoted by

H_0 : 99% of the control modules match SwitchRights specification.

The second hypothesis, called the alternative hypothesis and is denoted by

H_1 : less than 99% of the control modules match SwitchRights specification.

The alternative hypothesis is essentially saying that in this case, that SwitchRight cannot rely on its supplier of control modules supplying delivering batches of modules where 99% match SwitchRights specification.

Secondly, we describe the random sample from a statistical point of view, that is we find a statistical distribution which describes the behaviour of the sample. Suppose that X is the number of control modules in a random sample of 1000 matching SwitchRights specification.

We assume that the control modules are independent and that for each module the specification is either matched or it isnt. Under these conditions, X has a Binomial distribution and the problem can be summarised as follows:

$$X \sim B(1000, p)$$

$$H_0: \quad p = 0.99 \qquad\qquad H_1: \quad p < 0.99$$

Thirdly, we set up a mechanism to enable us to make a decision between the two hypotheses.

This is done by assuming that H_0 **is correct** until we can show otherwise.

Given that H_0 is correct we can calculate the mean μ and the standard deviation σ of the distribution as follows:

$$\mu = np = 1000 \times 0.99 = 990$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1000 \times 0.99 \times 0.01} = 3.15$$

Notice that

(a) $np > 5$ and (b) $n(1-p) > 5$

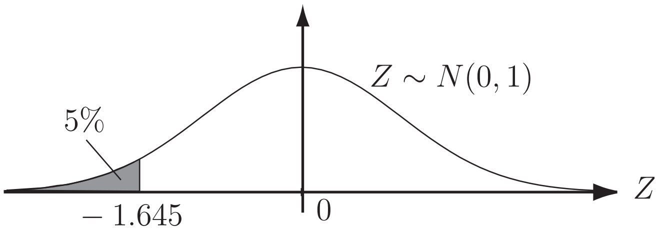
so that we can use the Normal approximation to the Binomial distribution, that is

$$B(1000, 0.99) \approx N(990, 3.15^2)$$

The sample value obtained is 985 and we now assess how close 985 is to the expected result of 990 by defining a remote left tail (in this case) of the Normal distribution and asking if the number 985 occurs in the left tail of the distribution or in the main body of the distribution.

In practice, we use the tail(s) of the standard Normal distribution and convert a problem involving the distribution $N(\mu, \sigma^2)$ into one involving the distribution $N(0, 1)$.

Diagrammatically the situation can be represented as shown below:



In general, the tails of a distribution can be defined to occupy any proportion of the distribution that we wish, the proportions chosen are usually taken as either 5% or 1%.

Given this information and a set of tables for the standard normal distribution we can assign values to the limits defining the tails.

Throughout this workbook we shall use the 5% proportion to define the tail(s) of a distribution unless otherwise stated.

In the case we have here, the alternative hypothesis states that p is **less** than 0.99. Because of this we use only one tail occupying a **total** of 5% of the distribution.

To discover where the number 985 lies within the distribution (tail or main body) we standardise 985 with respect to the normal distribution $N(990, 3.15^2)$ in the usual way (see Workbook 39).

The calculation is:

$$P(X \leq 985) = P\left(Z \leq \frac{985.5 - 990}{3.15}\right) = P(Z \leq -1.43)$$

Notice that 985.5 is used and not 985. This because we are using a *continuous* normal distribution to approximate a *discrete* binomial distribution and so

$$P(X = 985) \approx P(984.5 \leq X \leq 985.5)$$

the right hand side being calculated from the normal distribution.

The number $-1.43 > -1.645$ and so the number 985 occurs in the main body of the distribution not in the left tail. This suggests that the evidence does not support the claim that the number of control modules supplied meeting SwitchRights specification is different from 99%. Essentially, we accept the null hypothesis since we do not have the evidence necessary to reject it. Note that this result does not **prove** that the claim is true.

Before looking at similar problems, we will look at the possible ways of defining the tails of the standard normal distribution. As stated previously, we shall, in these notes, always use a total of 5% for the tail or tails of a distribution.

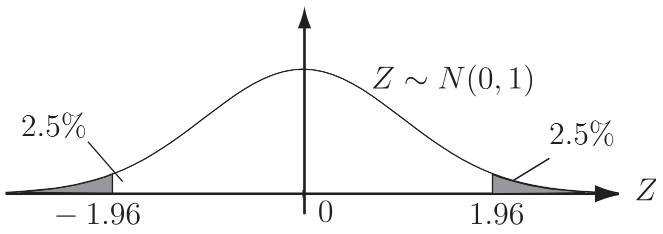
We say that we are making a decision at the **5% level of significance**.

The situation is represented by the following diagrams:

(1) **Hypotheses**:-

$H_0: \quad p = p_0$

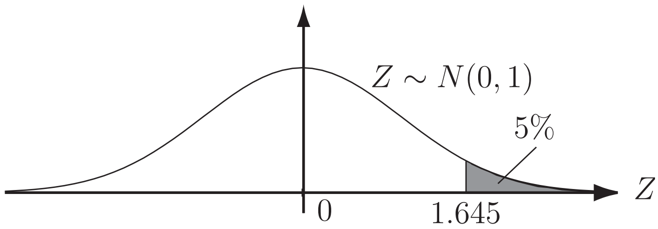
$H_1: \quad p \neq p_0$



(2) **Hypotheses**:-

$H_0: \quad p = p_0$

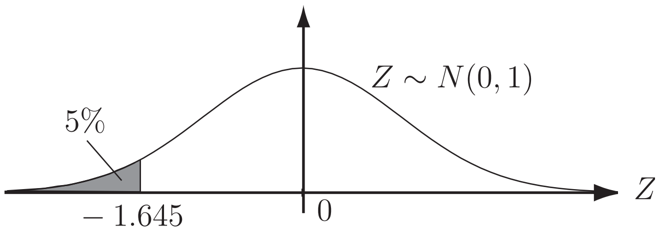
$H_1: \quad p > p_0$



(3) **Hypotheses**:-

$H_0: \quad p = p_0$

$H_1: \quad p < p_0$



The values ± 1.96, $+1.645$ and -1.645 are easily obtained from the standard normal tables given at the end of this workbook. The appropriate lines from the table are reproduced on the following page for ease of reference. Note that it is sometimes advisable to be 99% sure of either correctly accepting or rejecting a null hypothesis. In this case we say that we are working at the 1% level of significance. The situation diagrammatically is exactly the same as the one shown above except that the 5% tail areas become 1% and the 2.5% areas become 0.5%.

The corresponding values of Z are ± 2.58, $+2.33$ and -2.33 depending on whether a one-tailed or a two-tailed test is being performed.

Particular note must always be taken of the form of the hypotheses and the corresponding test, one-tailed or two-tailed.

Extracts from the Normal Probability Integral Table

Case 1 - the 5% Level of Significance

$Z = \frac{X-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.6	.4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1.9	.4713	4719	4726	4732	4738	4744	4750	4756	4762	4767

Case 2 - the 1% Level of Significance

$Z = \frac{X-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.3	.4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.5	.4938	4940	4941	4943	4945	4946	4948	4949	4951	4952

We shall now look at a problem which is similar in type to Problem 1 and solve it using the ideas discussed in the analysis of that problem.

Problem 2
The Head of Quality Control in a foundry claims that the castings produced in the foundry are 'better than average.' In support of this claim he points out that of a random sample of 60 castings inspected, 59 passed. It is known that the industry average percentage of castings passing quality control inspections is 90%. Do these results support the Heads claim?

Analysis

Let X denote the number of castings passing the quality control inspection from the sample of 60. Assuming that a casting either passes or fails the inspection process, we can assume that X follows the binomial distribution

$$X \sim B(60, p)$$

where p is the probability that a casting passes the inspection.

The null hypothesis H_0, is that the probability that a casting passes the inspection is the same as the industry average. The alternative hypothesis H_1, is that the Head of Quality Control is correct in his claim that castings produced in his foundry have a greater chance of passing the inspection. The problem can be summarised as:

$$X \sim B(60, p)$$

$$H_0: \quad p = 0.90 \qquad\qquad H_1: \quad p > 0.90$$

The form of the alternative hypothesis dictates that we do a one-tailed test.

If H_0 is correct we can calculate the mean and standard deviation of the binomial distribution above and, assuming that the appropriate condition are met, use the normal distribution with the same mean and standard deviation to solve the problem. The calculations are:

$$\mu = np = 60 \times 0.90 = 54$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{60 \times 0.90 \times 0.10} = 2.32$$

Notice that

$$\text{(a)} \;\; np > 5 \quad \text{and} \quad \text{(b)} \;\; n(1-p) > 5$$

so that we can use the normal approximation to the binomial distribution, that is

$$B(60, 0.90) \approx N(54, 2.32^2)$$

In order to make a decision, we need to know whether or not the value 59 is in the remote tails of the distribution or in the main body. Recall that the hypotheses are:

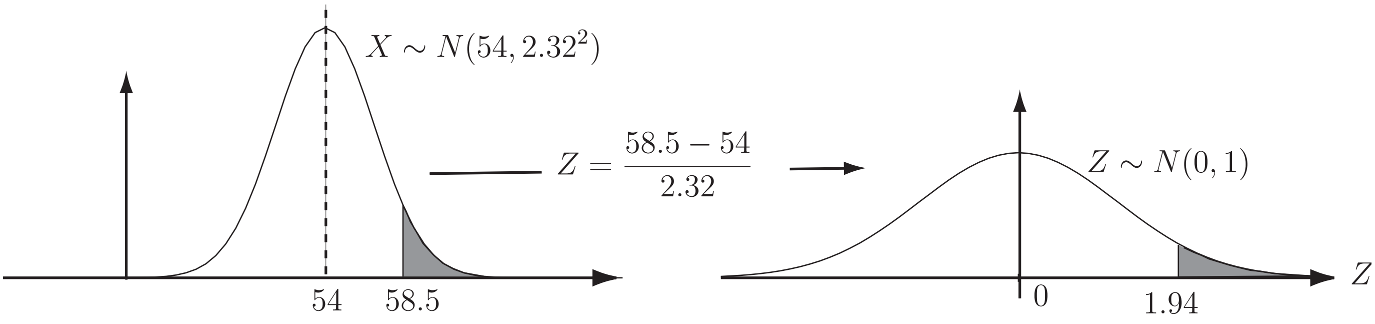
$$H_0: \quad p = 0.90 \qquad H_1: \quad p > 0.90$$

so that we must do a one-tailed test with a critical value of $Z = 1.645$.

The calculation is:-

$$P(X \geq 59) = P\left(Z \geq \frac{58.5 - 54}{2.32}\right) = P(Z \geq 1.94)$$

The situation is represented by the following diagram.



Since $1.94 > 1.645$, the result is significant at the 5% level and so we reject the null hypothesis. The evidence suggests that we accept the alternative hypothesis that, at the 5% level of significance, the Head of Quality Control is making a justified claim.

A firm manufactures heavy current switch units which depend for their correct operation on a relay. The relays are provided by an outside supplier and out of a random sample of 150 relays delivered, 140 are found to work correctly. Can the relay manufacturer justifiably claim that at least 90% of the relays provided will function correctly?

Your solution

Let X represent the number of relays working correctly. The required hypotheses are:

$$X \sim B(150, p) \qquad H_0: \; p = 0.90 \qquad H_1: \; p > 0.90$$

We perform a one-tailed test with critical value $Z = 1.645$. The necessary calculations are:

$$\mu = np = 150 \times 0.90 = 135$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{150 \times 0.90 \times 0.10} = 3.67$$

Since $np > 5$ and $n(1-p) > 5$, we can use the normal approximation to the binomial distribution. We approximate $B(150, 0.90) \approx N(135, 3.67^2)$. Hence:

$$P(X \geq 140) = P\left(Z = \frac{139.5 - 135}{3.67}\right) = P(Z \geq 1.23)$$

Since $1.23 < 1.645$ we cannot reject the null hypothesis at the 5% level of significance. There is insufficient evidence to support the manufacturers claim that at least 90% of the relays provided will function correctly.

2. Tests for Population Means

Tests concerning a Single Mean

Introduction

In cases where tests involving measurements are performed, it is often possible to statistically hypothesize about the results. Suppose that the boiling point of a particular coolant used in car engines is claimed by a manufacturer to be 110°C. Further suppose that a series of accurate measurements made in a laboratory using 8 random samples of the coolant are recorded as:

$$110.2^\circ,\ 110.3^\circ,\ 110.1^\circ,\ 109.8^\circ,\ 109.9^\circ,\ 110.0^\circ,\ 110.4^\circ,\ 110.1^\circ,$$

The mean of these results is 110.1°C.

It is reasonable to ask whether, on the basis of the results obtained, we may claim that the boiling point of the coolant is significantly higher than the assumed true boiling point of 110°C. We will return to this problem later in this workbook after looking at some general results.

General Results

In general terms, we need to make predictions, based on calculation, about the parameters of the population from which the random sample is drawn. As illustrated above we calculate the sample mean $\bar{x}$. The statistical tests used to answer the above question depend on whether the variance of the population is known or not.

Case (i) - Population Variance Known

Firstly, we form the null hypothesis that there is no difference between the sample mean $\bar{x}$ and the population mean μ, that is:

$$H_0: \quad \bar{x} = \mu$$

Secondly, we consider drawing samples of size n from a population of size n_p. If n is **large** ($n \geq 30$), then the Central Limit Theorem tells us that the sample means follow a normal distribution with mean and standard deviation (standard error of the means) s given by

$$s = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{n_p - n}{n_p - 1}}$$

In cases where the size of the population n_p is large in comparison to the sample size n, the quantity

$$\frac{n_p - n}{n_p - 1} \approx 1$$

so that the standard error of the means may be given as

$$s = \frac{\sigma}{\sqrt{n}}$$

Using the standard normal distribution $Z \sim N(0,1)$ leads us to calculate the value of Z as:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{n_p-n}{n_p-1}}}, \quad \text{for populations of size } n_p;$$

or

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}, \quad \text{for very large populations.}$$

We may now set up an alternative hypothesis which can take one of the three forms:

$$H_1: \quad \bar{x} \neq \mu$$

$$H_1: \quad \bar{x} > \mu$$

$$H_1: \quad \bar{x} < \mu$$

depending on the result we are looking for. A decision at the 5% level of significance may be made depending on whether:

$|Z| > 1.645$ for a two-tailed test

$Z > 1.96$, for a (right) one-tailed test;

$Z < -1.96$, for a (left) one-tailed test; In each case Z will lie in the remote tail(s) of the standard normal distribution and we would reject the null hypothesis in favour of the alternative hypothesis at the 5% level of significance.

In the case where n is **small**, ($n < 30$) the distribution of sample means in not normally distributed and we need to consider the following:

- If we know the value of *sigma* the quantity

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

  has a normal distribution

- If we have to estimate σ then the quantity

  $$\frac{\bar{x} - \mu}{s/\sqrt{n}}$$

  approximates to Students t-distribution (See Workbook 41).

Note that for large samples the two cases differ little but that the differences are important if the sample size is small. In this case we can still use the alternative hypotheses stated above but in conjunction with the t-values:

$$|t| = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Note that we must use these values of t in conjunction with tabulated values of t and take the number of degrees of freedom to be $n - 1$ where n is the sample size.

So that they are available when you need them, the percentage points of the distribution, along with a copy of the standard normal distribution tables are attached to the end of this workbook.

Example Dishwasher powder is poured into the cartons in which it is sold by an automatic dispensing machine which is set to dispense 3 kg of powder into each carton. In order to check that the dispensing machine is working to an acceptable standard (i.e. does not need adjustment), a production engineer takes a random samples of 40 cartons and weighs them. It is found that the mean weight of the sample is 3.005 kg. It is known that the dispensing machine operates with a variance of 0.015^2 kg^2 and that the manufacturer of the powder is willing to rely on a 5% level of significance. Does the sample provide the engineer with sufficient evidence that the machine is operating acceptably and so does not need adjustment?

Solution

Given that the dispensing machine can over-fill or under-fill the containers, the null and alternative hypotheses are:

$$H_0: \quad \bar{x} = 3 \qquad H_1: \quad \bar{x} \neq 3$$

Since the sample size is large (≥ 30) and we can regard the population as infinite but with a known variance, we can calculate the relevant value of the test statistic Z by using the formula:

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

Hence, in this case:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{3.005 - 3}{0.015/\sqrt{40}} = 2.108$$

and since we are performing a two-tailed test at the 5% level of significance and have found that $|Z| > 1.645$, that is, Z is outside the range $[-1.645, 1.645]$, we must reject the null hypothesis and conclude that the machine is not operating acceptably and needs adjustment.

Case (ii) - Population Variance Unknown

We have exactly the same situation as that described in Case (i) but do not know the value of the population variance σ^2. Fortunately, this turns out to be fairly straightforward to deal with since it may be shown that for a sample of size n with variance s^2 taken from a population with variance σ^2, the variances are related by the approximation:

$$s^2 \approx \frac{(n-1)}{n}\sigma^2$$

so that we can closely approximate the population standard deviation as

$$\hat{\sigma} = s\sqrt{\frac{n}{n-1}}$$

the $\hat{\sigma}$ notation indicating that we are effectively *estimating* the population standard deviation. The factor $\frac{n}{n-1}$ is called Bessels correction. Note that as the sample size increases

$$\sqrt{\frac{n}{n-1}} \to 1$$

and that for $n = 30$

$$\sqrt{\frac{n}{n-1}} \to 1.017$$

Essentially, for sample sizes larger than 30 we can neglect Bessels correction although we must take it into account for small samples or errors will be introduced into any calculations made which may have an effect on decisions made when performing hypothesis tests.

For large samples (σ unknown) the equation for Z quoted previously becomes

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

For small samples (σ unknown) we have

$$|t| = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

For small samples Bessels correction must be taken into account. If the population distribution is normal, then the means of small samples follow a normal distribution. If the population is not normal, there is nothing we can say in general about the distribution of the means of small samples. $\quad Z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

In summary we have the following.

Population	Variance	Sample size	Test
Normal	Known	Small	Normal (Z)
Normal	Known	Large	Normal (Z)
Normal	Unknown	Small	t
Normal	Unknown	Large	t but Z approximates
Not Normal	Either	Small	Non-parametric
Not Normal	Known	Large	Z approximates
Not Normal	Unknown	Large	Z and t approximate

Non-parametric testing is covered in Workbook 46.

Example The average useful life of a random sample of 33 similar calculator batteries taken from a batch of 100 is found to be 99.5 hours continuous use. The batch is classed as unsuitable for production purposes if the mean lifetime of the batch under conditions of continuous usage is significantly less than the population mean of 100 hours. If the variance of the sample lifetime is 18.49 hours^2, determine, at the 5% level of significance, whether the batch is suitable for production purposes.

Solution

The null and alternative hypotheses are:

$$H_0: \quad \bar{x} = 100 \qquad H_1: \quad \bar{x} < 100$$

Since the sample size may be taken as large (33) and the population small (100) and we have to estimate the population variance from the sample variance, we calculate Z using the formula previously used in conjunction with finite populations

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}\sqrt{\frac{n_p - n}{n_p - 1}}}}$$

In this case

$$Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}\sqrt{\frac{n_p - n}{n_p - 1}}}} = \frac{99.5 - 100}{\frac{\sqrt{18.49}}{\sqrt{33}}\sqrt{\frac{100-33}{100-1}}} = -0.812$$

and since we are performing a one-tailed test and have found that $-1.645 < -1.319$ we conclude that the value of Z is not in the remote tails of the $A \sim N(0, 1)$ and that, at the 5the batch is suitable for production purposes.

Solve the problem given at the start of this section. Note the sample is small and you will have to estimate the population variance from the sample variance. Use the tabulated values of the distribution given at the end of this workbook in conjunction with the appropriate number of degrees of freedom.

Your solution

The null and alternative hypotheses are:

$$H_0: \ \mu = 110 \qquad H_1: \ \mu > 110$$

The value of the sample variance is given by the formula

$$s^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{0.28}{8} = 0.035$$

We have a small sample and a large population so that the test statistic t is given by

$$|t| = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{110.1 - 110}{\sqrt{0.035}/\sqrt{7}} = \frac{0.1 \times \sqrt{7}}{0.187} = 1.415$$

At the 5% level of significance and using $8 - 1 = 7$ degrees of freedom, the value of $t_{\alpha,\nu}$ from tables is 1.895. Since $1.415 < 1.895$, we cannot reject the null hypothesis in favour of the alternative hypothesis. On the basis of the evidence available, we are not able to conclude that the boiling point of the coolant is greater than 110°C.

General Comments about Tests concerning a Population Mean

(a) The sample mean $\bar{x}$ is often used as a test statistic when testing a hypothesis concerning a population mean μ.

(b) Even if the population distribution cannot be assumed to be normal, the distribution of sample means can often be assumed to be normal. This depends on the sample size.

(c) The tests described above sometimes require us to assume that the population variance is known. This is often unrealistic and we turn to the t-test to deal with cases where the population standard deviation is unknown and must be estimated from the data available.

General Comment on the t-test

(a) The test only applies when the underlying distribution can be assumed to be normal.

(b) The test is used when the standard deviation of the parent population has to be estimated.

(c) As the sample size n get larger, the distribution approximates to the standard normal distribution.

(d) The distribution depends on the number of degrees of freedom, for a single sample or equal paired samples (see below), the number of degrees of freedom is always one less than the sample size.

Tests concerning Paired Data

Sometimes experimental data may be directly compared using an appropriate test. The following example looks at experimental data concerning the throttle reaction times of two turbochargers fitted to an internal combustion engine.

Example In order to test the hypothesis that two standard turbochargers A and B have the same throttle reaction times, a random sample of 7 cars were fitted with the turbochargers and the throttle reaction times measured. The results were as follows:

Car	1	2	3	4	5	6	7
Throttle Reaction time for A; $R1$	0.223	0.212	0.201	0.205	0.216	0.211	0.209
Throttle Reaction time for B; $R2$	0.208	0.207	0.203	0.204	0.205	0.202	0.206
$D = R1 - R2$	0.015	0.005	−0.002	0.001	0.011	0.009	0.003

Solution

Let D be the difference between the throttle reaction times of the two turbochargers. We assume that the distribution of D is normal. Let $\bar{d}$ be the sample mean. We must decide between the two hypotheses:

$$H_0: \quad \bar{d} = 0 \qquad H_1: \bar{d} \neq 0$$

$\bar{d}$ is the mean of the 7 differences and the alternative hypothesis dictates that we perform a two-tailed test. The sample size of 7 gives us 6 degrees of freedom so that the 5% critical values (from tables) are $+2.447$ and -2.447. The sample mean is calculated as

$$\bar{d} = \frac{\sum d}{7} = \frac{0.042}{7} = 0.006$$

The value of t is given by:

$$|t| = \frac{\bar{d} - 0}{s/\sqrt{n-1}} = \frac{0.006}{0.00552914/\sqrt{6}} = 2.658$$

Since $2.658 > 2.447$, we reject H_0 and conclude that, on the basis of the evidence available, there is a difference in the throttle reaction times of the two turbochargers at the 5% level of significance.

Two different methods of analysis were used to determine the levels of impurity present in a particular aircraft quality aluminium alloy. Eight specimens were analysed using both methods. Does the available evidence suggest that both methods lead to the same results

Alloy Specimen	1	2	3	4	5	6	7	8
Test 1	1.24	1.23	1.24	1.20	1.21	1.22	1.23	1.22
Test 2	1.23	1.20	1.20	1.21	1.20	1.20	1.21	1.25
$D =$ Test1 $-$ Test2	0.01	0.03	0.04	−0.01	0.01	0.02	0.02	−0.03

Your solution

Let D be the difference between the two methods of analysis. We assume that the distribution of D is normal. Let $\bar{d}$ be the sample mean. We must decide between the two hypotheses:

$$H_0: \ \bar{d} = 0 \qquad H_1: \ \bar{d} \neq 0$$

$\bar{d}$ is the mean of the 8 differences and the alternative hypothesis dictates that we perform a two-tailed test. The sample size of 8 gives us 7 degrees of freedom so that the 5% critical values (from tables) are +2.365 and 2.365. The sample mean is calculated as

$$\bar{d} = \frac{\sum d}{8} = 0.01125$$

The value of t is given by:

$$|t| = \frac{\bar{d} - 0}{s/\sqrt{n-1}} = \frac{0.01125}{0.02087912/\sqrt{7}} = 1.426$$

Since $-2.306 < 1.426 < 2.306$, we do not have the evidence to reject H_0 and conclude that, on the basis of the evidence available, there is no difference between the two methods of analysis at the 5% level of significance.