-6-

Testing of hypotheses

6.1 Introduction

Consider a factory producing PE sacks. The manufacturer claims that the mean mass of produced sacks exceeds 10 kg. This claim is known as the **null hypothesis** and is written: H_0: $\mu > 10$.

The hypothesis that does not support the claim is known as the **alternate hypothesis** and is written as follows: H_a: $\mu \leq 10$.

This defines two regions on a line indicating the mean mass of population: An **acceptance region** if $\mu > 10$ and a **rejection region** if $\mu \leq 10$. This situation in shown in Figure (6.1).

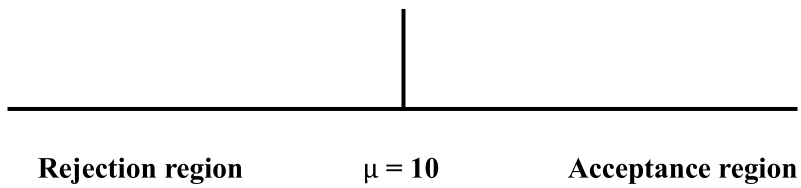

Fig (6.1): Acceptance and rejection regions for H_0: $\boldsymbol{\mu > 10}$

In practice it is common to allow for a certain deviation from the theoretical case discussed above. This means accepting the null hypothesis for values of mean $= \mu - a$. The determination of the allowed error (a) will be explained later.

A similar case may show up if a manufacturer of a certain chemical claims that, on the average, the percent impurities in his product is less than 0.5%. In that case: The null hypothesis is H_0: $\mu < 0.5$ and the alternate hypothesis: H_a: $\mu \geq 0.5$. In that case the acceptance and rejection regions show as in Figure (6.2).

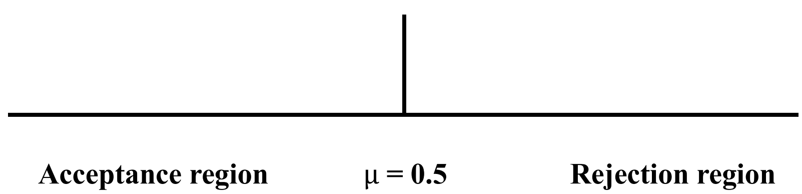

Fig (6.2): Acceptance and rejection regions for H_0: $\boldsymbol{\mu < 0.5}$

Here also, we may accept the null hypothesis for values of mean $= \mu + a$.

In the two previous examples, the acceptance or rejection of the null hypothesis depends on whether the mean of population is higher or lower than a certain control value. Such cases represent a **one – tailed hypothesis.**

Consider now the production of several tons per day of distilled water of mean pH = 7. Any deviation from that figure would be considered in principle inadequate so that the null hypothesis takes the form: H_0: $\mu = 7$ while the alternate hypothesis is H_a: $\mu \neq 7$. This way, there will appear two rejection regions from both sides of the hypothetical value 7 (Figure 6.3). This is called a **two – tailed hypothesis**.

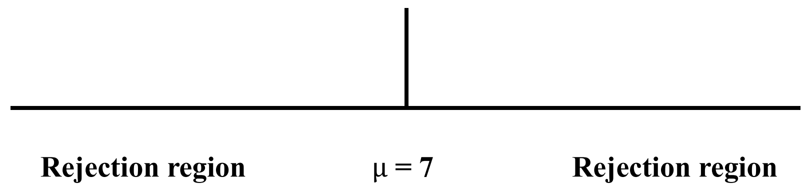

Fig (6.3): Rejection regions for H_0: $\mu = 7$

In that case, it is common practice to allow for deviations from the mean from both sides of its hypothetical values, that is in the range $]\mu - a\ ;\ \mu + a[$.

6.2 Hypotheses concerning the mean of a population:

6.2.1 The case of large samples $(n > 30)$

To check the veracity of the null hypothesis we choose a sample form the population of size n and determine the mean value of the parameter of interest $\bar{x}$ and its standard deviation s. In that case, the null hypothesis to be tested will consist of one of two possibilities, assuming that the studied parameter is normally distributed along the population.

(a) One – tailed hypotheses

This is the case when the null hypothesis takes the form: H_0: $\mu > k$ or H_0: $\mu < k$. We first calculate the **test statistic *z*** form the expression:

$$|z| = \frac{|\bar{x} - k|}{\frac{s}{\sqrt{n}}} \tag{6.1}$$

The next step is to compare the value of the calculated statistic to a critical z – value (z_{crit}) obtained from the function NORM.S.INV (L). The null hypothesis will be accepted if $|z| < |z_{crit}|$. Figure (6.4) explains the rationale of this criterion for the two cases where the null hypothesis is H_0: $\mu > k$ or H_0: $\mu < k$. The figures show that the alternate hypothesis H_a is more probable if $|z| > |z_{crit}|$.

Example 6.1

In the preparation of concrete slabs, concrete cubes are tested for compressive strength and the average strength should exceed 30MPa. If 35 cubes are tested and the mean strength was found to be 28.7MPa with a standard deviation of 2.66MPa, would you accept the null hypothesis H_0: $\mu > 30$ at 0.05 significance level?

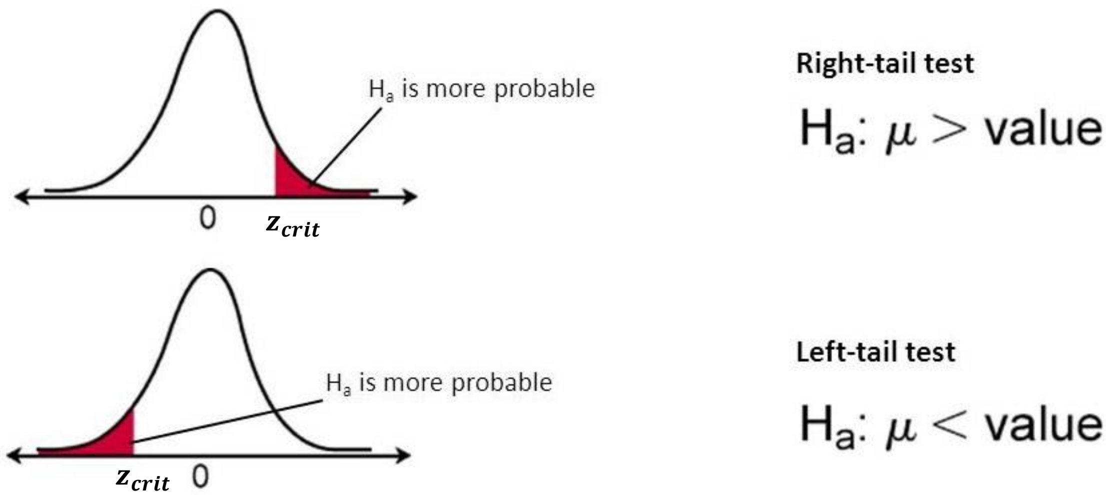

Fig (6.4): Alternate hypotheses for one – tailed tests

Solution:

The mean value of sample $\bar{x}$ = 28.7 and the standard deviation s = 2.66. The null hypothesis H$_0$: $\mu > 30$ corresponds to a value of $k = 30$.

$|z|$ = $\frac{|28.7-30|}{\frac{2.66}{\sqrt{35}}}$ = 2.89

At $\alpha = 0.05, L = 0.95, z_{crit} = 1.645$

Since $2.89 > 1.645$, the null hypothesis cannot be accepted.

Example 6.2

The daily amount of rejected items from a production line is normally distributed. The company's policy requires that this should not exceed 7% of the production. Data gathered over a one-month period (30 days) show that the daily mean percentage of rejected items = 7.62% with a standard deviation of 1.41%. What minimum significance level should be taken in order not to reject the null hypothesis: H$_0$: $\mu \leq 7$?

Solution:

The mean value of sample $\bar{x}$ = 7.62 and the standard deviation s = 1.41 while $n =$ 30 and $k = 7$.

$|z|$ = $\frac{|7.62-7|}{\frac{1.41}{\sqrt{30}}}$ = 2.408

The L value corresponding to 2.408 is obtained from NORM.S.DIST $(2.408, \text{TRUE})$. This gives $L = 0.992$. This means that the null hypothesis cannot be rejected if $L > 0.992$ or $\alpha < 1 - 0.992 \rightarrow \alpha < 0.008$.

Hence the maximum significance level that would not reject the null hypothesis is $\boldsymbol{\alpha = 0.008}$

(b) Two – tailed hypotheses

When the null hypothesis takes the form H_0: $\mu = k$, the z – statistic is also determined from equation (6.1). However, the two rejection regions where the alternate hypothesis is accepted are now form both sides of the mean as revealed from Figure (6.5). In that case, the critical vlaue of z is obtained from the function NORM.S.INV $\left(\frac{1+L}{2}\right)$

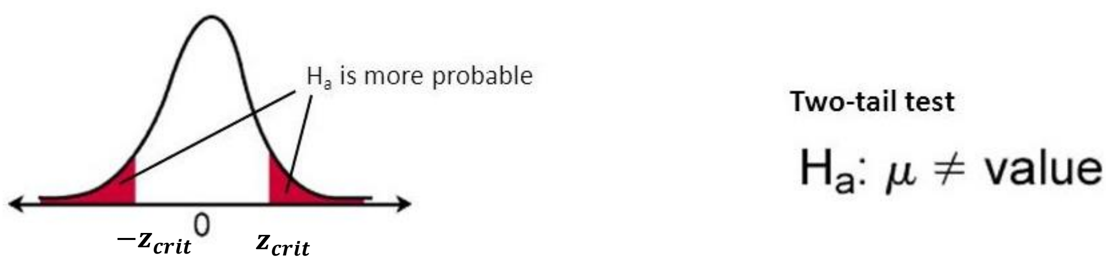

Fig (6.5): Alternate hypotheses for two – tailed tests

Example 6.3

For proper operation, the length of a particular lever in an engine should equal 645mm. When 40 pieces from the daily production were chosen, the mean length was 645.5mm with a standard deviation of 0.84mm. At 0.05 significance level, would you consider that the manufacturing process of these parts needs some adjustment?

Solution:

This is a two – tailed hypothesis with H_0: $\mu = 645$ and $\bar{x}$ = 645.5, s = 0.84, n = 40 and $L = 0.95$. The test statistic is:

$$|z| = \frac{|645.5 - 645|}{\frac{0.84}{\sqrt{40}}} = 3.7646$$

The critical value of z is obtained from NORM.S.INV $\left(\frac{1 + 0.95}{2}\right)$ = 1.96

Since $3.7646 > 1.96$, then the null hypothesis H_0: $\mu = 645$ cannot be accepted and the manufacturing process effectively needs some adjustment.

6.2.2 The case of small samples $(n < 30)$

More often than not, it is not possible to test large number of specimens and the need for a modification of the last criterion arises. In case of $n < 30$, the t – distribution is used whereby the test statistic takes the form:

$$|t| = \frac{|\bar{x}-k|}{\frac{s}{\sqrt{n-1}}} \qquad (6.2)$$

The null hypothesis is accepted if $|t| < |t_{crit}|$. The critical value of t is obtained as follows:

- **For one – tailed tests**: Use the function T.INV $(\alpha\ , n-1)$
- **For two – tailed tests**: Use the function T.INV.2T $(\alpha\ , n-1)$

Example 6.4

Six specimens were chosen from the production line of HDPE bottles and their density was determined. The values were as follows (g.cm^{-3}): $0.962\ , 0.949\ , 0.958\ , 0.961\ , 0.960\ , 0.944$. The factory claims that the mean density does not exceed 0.95 g.cm^{-3}. At a significance level = 0.05, would you accept that claim?

Solution:

The null hypothesis to be tested is: H_0: $\mu \leq 0.95$.

First the mean value and standard deviation of sample density is determined: $\bar{x}$ = 0.9548, s = 0.0079 with n = 6.

$$|t| = \frac{|0.9548-0.95|}{\frac{0.0079}{\sqrt{6-1}}} = 1.355$$

The critical value of $|t|$ = $|\mathrm{T.INV}\ (0.05\ , 6-1)|$= 2.015

Since $1.355 < 2.015$, then the null hypothesis cannot be rejected.

Example 6.5

The pH in a certain chemical reaction must be fixed at pH = 8.3 for proper results. Specimens from the reaction mixture were drawn at regular hourly intervals for 8 hours and their pH determined. The results obtained were as follows: $8.29\ , 8.18\ , 8.09\ , 8.12\ , 8.31\ , 8.34\ , 8.22\ , 8.19$. Show that the alternate hypothesis H_a: $\mu \neq 8.3$ cannot be rejected and determine the maximum significance level necessary for the null hypothesis H_0: $\mu = 8.3$ not to be rejected.

Solution:

The mean value and standard deviation of sample density is determined: $\bar{x}$ = 8.2175, s = 0.09 with n = 8.

$|t| = \frac{|8.2175-8.3|}{\frac{0.09}{\sqrt{8-1}}} = 2.424$

The critical value of $|t| = |\mathrm{T.INV.2T}\ (0.05\ ,8-1)| = 2.364 < 2.424$. Hence the null hypothesis cannot be accepted and consequently we do not reject the alternate hypothesis.

Using Goal – Seek, we can get the maximum value of α which would not reject the null hypothesis. This is found to equal $\boldsymbol{\alpha = 0.0458}$.

6.3 Sample size

When it is required for the sample mean to be as close as possible to the population mean, one must choose a large sample. If D is the difference between sample mean and population mean ($|\bar{x} - \mu|$) then from equation (6.1):

$$z_{crit} = \frac{D.\sqrt{n}}{\sigma}$$

Or $n = \left(\frac{z_{crit}\sigma}{D}\right)^2$ (5.3)

6.3 Hypotheses concerning the mean of a proportion

Sometimes we are more interested in testing a proportion rather than a parameter. The null hypothesis in that case takes the form:

- H_0: $\pi < k$ or $\pi > k$ for **one tailed hypotheses**
- H_0: $\pi = k$ for **two tailed hypotheses**

In case of large samples ($n > 30$), the test statistic is:

$|z| = \frac{|p-k|}{\sqrt{\frac{\pi.\tau}{n-1}}}$ (6.4)

Here, p is the proportion of sample while π is that of population ($\pi = k$) and $\tau = 1 - \pi$

The critical value of z is determined either from NORM.S.INV (L) or NORM.S.INV $\left(\frac{1+L}{2}\right)$ depending on whether the test is one – tailed or two – tailed respectively.

In case of small samples ($n < 30$), the test statistic is also obtained from equation (6.1) and the critical value of t obtained from $|\mathrm{T.INV}\ (\alpha\ , n-1)|$or T.INV.2T ($\alpha\ , n-1$) for one- tailed and two – tailed tests respectively.

Example 6.6

An instructor claims that at least 60% of his students have passed the final exam. When a sample of 30 students was chosen, 15 out of them turned out passing. At 0.05 significance level, would you accept the instructor's claim?

Solution:

H_0: $\pi \geq 0.6$ Hence $\pi = 0.6, \tau = 0.4$

The sample proportion $p = \frac{15}{30}$ = 0.5

$|z| = \frac{|0.5 - 0.6|}{\sqrt{\frac{0.6\times0.4}{30-1}}} = 1.1$ At $\alpha = 0.05, z_{crit} = 1.65 > 1.1$ Hence H_0 can be accepted.

Example 6.7

A coin was tested for uniformity (That is the probabilities of both heads or tails are equal). This coin was thrown 26 times and the number of tails obtained was 9. At a 95% confidence level, test the null hypothesis H_0: $\pi = 0.5$

Solution:

H_0: $\pi = 0.5$ Hence $\pi = \tau = 0.5$

The sample proportion $p = \frac{9}{26}$ = 0.346

$$|t| = \frac{|0.346 - 0.5|}{\sqrt{\frac{0.5\times0.5}{26-1}}} = 1.54$$

At $\alpha = 0.05, t_{crit} = 2.06 > 1.54$ Hence H_0 can be accepted.

6.4 Hypotheses concerning the difference between two means

6.4.1 Introduction

Suppose that two catalysts are tested to enhance the rate of a certain reaction. The reaction is repeated several times using each time the two catalysts individually and conversion is determined each time. The null hypothesis to be tested is whether the mean values of conversion using the two catalysts are comparable. That is:

H_0: $\mu_1 = \mu_2$

6.4.2 The case of large samples

Let the mean values of the two set of samples be $\bar{x}_1$ and $\bar{x}_2$ and their standard deviations s_1 and s_2 respectively. The statistic used takes the form:

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \qquad (6.5)$$

Example 6.8

Two different mixers were proposed to blend plastic powders. The uniformity of the blend was assessed by the time taken for the standard deviation of the mix density to stabilize. When 30 runs were carried out on the first mixer, the average time required to reach homogeneity was 23.5 min. with a standard deviation of 2.5 min. The corresponding figures for the second mixer were 25.2 and 3.8 min. respectively for a 35-size sample. Determine at 0.05 significance level, whether the performance of the two mixers can be considered comparable.

Solution:

H_0: $\mu_1 = \mu_2$

$\bar{x}_1 = 23.5 \quad s_1 = 2.5 \quad n_1 = 30$

$\bar{x}_2 = 25.2 \quad s_2 = 3.8 \quad n_2 = 35$

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|23.5 - 25.2|}{\sqrt{\frac{2.5^2}{30} + \frac{3.8^2}{35}}} = 2.16$$

At $\alpha = 0.05 \rightarrow L = 0.95, z_{crit} = 1.96 < 2.16$ (The function NORM.INV $\left(\frac{1+0.95}{2}\right)$ was used.

Hence the null hypothesis cannot be accepted and the performance of the two mixers cannot be considered comparable.

6.4.3 The case of small samples

In that case, the null hypothesis H_0: $\mu_1 = \mu_2$ is tested in a different way:

First a “pooled” standard deviation is used using the expression:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{6.6}$$

The statistic of the test is:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{6.7}$$

This is then compared to the critical value of t for two – tailed test at the required significance level for a number of degrees of freedom = $n_1 + n_2 - 2$

Example 6.9

Two different methods were used to determine the nickel content in a steel alloy using 4 specimens each time. The results were as follows:

$\bar{x}_1 = 3.285 \quad s_1 = 0.00774 \quad n_1 = 4$

$\bar{x}_2 = 3.258 \quad s_2 = 0.00960 \quad n_2 = 4$

Is the difference between the two methods significant at 0.05 significance level?

Solution:

$$s_p = \sqrt{\frac{(4-1)\times 0.0074^2 + (4-1)\times 0.0096^2}{4+4-2}} = 0.00857$$

$$t = \frac{|3.285 - 3.258|}{0.00857\times\sqrt{\frac{1}{4}+\frac{1}{4}}} = 4.45$$

The critical value of t is obtained from T.INV.2T $(0.05, 4+4-2) = 2.447 < 4.45$. Hence, the difference between the two methods of testing is significant.

6.5 Hypotheses concerning paired samples

In paired samples, we compare two sets of matched specimens. For example, consider a group of students having failed an exam that are to attend extra sessions. To test whether these sessions have had any effect on their performance, we compare the scores of these students before and after the sessions. The null hypothesis concerns the mean difference in scores $\overline{D}$:

Null hypothesis: $\overline{D} = 0$ (The extra sessions did not improve the students' status)

Alternate hypothesis: $\overline{D} > 0$ (The extra sessions improved the students' status).

Whether the sample size is large or small, such problem is dealt with one – tailed statistics. The test statistic is:

$$t = \frac{\overline{D}}{s_D}\sqrt{n-1} \qquad (6.8)$$

Example 6.10

A nano-oxide is claimed to stabilize the viscosity index (V.I.) of lube oils when added at 0.5% level. Seven specimens of the same oil were tested for V.I. before and after the addition. Determine at 0.05 significance level whether the addition has had any significant effect in improving the viscosity index.

V.I. Before	101.3	111	103.5	107.6	105.5	108.2	104.1
V.I. After	103.2	111.5	106.5	108.2	109.3	111	106.3

Solution:

H_0: $\overline{D} = 0$

H_a: $\overline{D} > 0$

We first determine the difference:

V.I. Before	101.3	111	103.5	107.6	105.5	108.2	104.1
V.I. After	103.2	111.5	106.5	108.2	109.3	111	106.3
D	1.9	0.5	3	0.6	3.8	2.8	2.2

The mean value and standard deviation of differences are: $\bar{D} = 2.114, s_D = 1.228$

$t = \frac{\bar{D}\sqrt{7-1}}{s_D} = 4.217$

The critical value of t as calculated from $|\text{T.INV}\,(0.05\,,7-1)| = 1.943$

Since $4.21 > 1.94$, then the null hypothesis cannot be accepted, and the alternate hypothesis cannot be rejected. Hence, we conclude that the addition significantly improved the viscosity index.

6.6 Type I and Type II errors

Consider the following case: When a sample of 6 foamed polystyrene slabs were tested for density, the mean value of density was 0.78g.cm^{-3} with a standard deviation of 0.066g.cm^{-3}. To check the null hypothesis:H_0: Mean density < 0.7 at 0.05 significance level, we calculate the statistic from equation (6.2) to obtain:

$|t| = \frac{|0.78-0.7|}{\frac{0.066}{\sqrt{6-1}}} = 2.71$ while $|t_{crit}| = 2.015$.

This means that the null hypothesis cannot be accepted. What if the null hypothesis was correct but the choice of the sample resulted in its rejection? In that case, we are in the presence of a **type I error in which a hypothesis is rejected while it is correct.** The error here is simply the significance level α (0.05). Decreasing the error means choosing a lower significance level. Effectively, if the significance level is reduced to 0.021, then $|t_{crit}| = 2.015$ and the null hypothesis can no more be rejected. Obviously, this means that we need to ensure to a probability of $1 - 0.021 = 0.979$ that our judgment is correct, which justifies not rejecting the null hypothesis in that case.

On the other hand, if in the previous example, the mean density of the sample turned out to be 0. 75g.cm^{-3} with the same standard deviation, then we would had obtained $|t| = 1.69 < 2.015$ and the null hypothesis would have been accepted. Again, if another sample was chosen, what if its average density or standard deviation would have resulted in a value of $|t_{crit}| > 2.015$? here we are in presence of **a type II error where a hypothesis has been accepted while it is false.** This type of error is denoted by β and can be obtained using the Goal – Seek technique by determining the significance level at which $|t_{crit}| = 1.69$.This results in $\beta = 0.076$.

6.7 Exercise problems

In all forthcoming problems, assume the properties to be normally distributed along population.

1. The presence of nitrogen oxides in the ambient atmosphere in the premises of a chemical plant is limited to a maximum of 600 mg.L^{-1}. On 30 consecutive days in a certain month, samples were drawn that resulted in a mean value of 624 mg.L^{-1} and a standard deviation of 42.5 mg.L^{-1}. At 0.05 significance level, would you accept the null hypothesis that the mean nitrogen oxides concentration does not exceed 600 mg.L^{-1}? What maximum significance level is required so as not to reject this hypothesis?

2. High sulfur content in fuel oil is undesirable, and a refinery claims that its production of fuel oil contains no more than 2% sulfur. The standard deviation of sulfur content is known to be about 0.22%. What sample size would ensure that a mean sulfur content of 2.1% in that sample would not result in rejecting the refinery's claim at 0.05 significance level?

3. Design calculations have shown that the compressive strength of concrete slabs should exceed 32MPa. When 35 core concrete cubes were tested, they resulted in a mean strength of 30.6MPa with a standard deviation of 1.8MPa. At 0.05 significance level, would you accept or reject the null hypothesis: H_0: Strength > 32?

4. The diameters of flywheels produced by a factory should equal 280mm. When a 40 size sample was chosen, the distribution of diameters was as shown below.

***D* mm**	277	278	279	280	281	282	283
f	3	11	7	8	5	4	2

At 0.05 significance level, would you consider the null hypothesis H_0: $D = 280$ valid?

5. The yield strength of steel bars should exceed 125MPa. When 6 bars were tested they produced the following results for yield strength: $117, 121, 122, 119, 117, 118$. What maximum significance level would not reject the null hypothesis: H_0: $\sigma > 125$?

6. According to a research center, a polymerization reaction reached a yield of 90% after 2 hours under certain experimental conditions. To confirm these results, the reaction was repeated five times under the same conditions and the yield after one hour determined. The results obtained for yield were as follows:

87.6 , 88.8 , 90.0 , 86.5 , 87.9. At 0.05 significance level, would you consider the null hypothesis H_0: Yield $= 90$ valid?

7. The company that manufactures a certain type of mixers stated that the time required reaching a homogeneous mixture on agitating two specific immiscible liquids does not exceed 10 min. Mixing runs were carried out twelve times and the time required to homogenize the mixture was assessed by reaching a steady mean density. The results for mixing time were as follows (min.):

   10.6 , 11.2 , 9.8 , 10.7 , 11.0 , 10.5 , 9.7 , 11.5 , 11.2 , 10.0 , 9.5 , 10.2.

   Test the null hypothesis: H_0: $t \leq 10$ at significance level = 0.05.

8. Two sets of students enrolled in two different programs (A and B) were examined in the same subject. The results are summarized as follows:

	$\boldsymbol{n}$	$\boldsymbol{\bar{x}}$	$\boldsymbol{s}$
A	41	64.3	15.6
B	51	59.5	17.2

   Use $\alpha = 0.05$ to decide whether the differences in their scores are significant.

9. The yields of a chemical reaction carried out in two different reactors produced the following results:

   - Reactor I: Nº of tests = 5, average yield = 96.3%, standard deviation = 2.75
   - Reactor II: Nº of tests = 6, average yield = 93.3%, standard deviation = 3.35

   Using 0.05 significance level, determine whether the difference in yield between the two reactors is significant.

10. A manufacturer of PE bottles claims that the percent defective production does not exceed 4%. 80 bottles were chosen at random and 6 of them were found to be defective. What maximum significance level would not reject the claim?

11. Two different techniques were used to prepare a certain catalyst and the specific surface area ($cm^2.g^{-1}$) is determined by nitrogen adsorption (BET analysis). The results of samples obtained on using the two techniques were as follows:

A	115.6	127.8	100.9	134.6	122.7	119.8	123.2	116.9	109.8
B	103.8	110.6	118.2	101.4	104.7	110.0	103.2		

   Determine at a significance level = 0.05 whether there is any significant difference in applying the two techniques.

12. A researcher wants to confirm whether the use of his own method in the preparation of an elastomer (A) yields better yield than a conventional method

(B). To that aim, he carries out two sets of tests and obtains the following results for the elastomer yield:

A	0.91	0.88	0.85	0.86	0.83	0.91
B	0.92	0.89	0.82	0.81	0.8	0.86

Prove that at significance level = 0.05, the difference between the two methods is statistically insignificant.

13. Seven students who failed to pass an exam were given extra tutorials and re-examined. Their scores (before and after the sessions) were as shown. Determine whether these sessions have significantly improved their performance at significance level = 0.05.

Before	9	6	4	11	10	2
After	11	6	7	12	10	5

14. A researcher wishes to prove that a certain doping additive has improved the viscosity index of a specific lubricating oil. He performs 5 tests on five specimens before and after the addition and obtains the following results:

V.I. Before	122	119	120	127	122
V.I. After	129	121	118	129	127

Prove that at significance level = 0.05, the proposed addition has not significantly affected the V.I.

Now, the researcher, who is unscrupulous, decides to doctor his data to prove his point of view by changing the third entry in the second row. What figure should he put instead of (118) to prove his point?