RAFFLES INSTITUTION
H2 Mathematics (9758)
2025 Year 6

Additional Practice Questions for Chapter S5: Hypothesis Testing (Solutions)

1 **9758/2017/02/Q7**

The production manager of a food manufacturing company wishes to take a random sample of a certain type of biscuit bar from the thousands produced one day at his factory, for quality control purposes. He wishes to check that the mean mass of the bars is 32 grams, as stated on the packets.

(i) State what it means for a sample to be random in this context. **[1]**

The masses, x grams of a random sample of 40 biscuit bars are summarised as follows.

$$n=40 \quad \sum(x-32)=-7.7 \quad \sum(x-32)^2=11.05$$

(ii) Calculate unbiased estimates of the population mean and variance of the mass of biscuit bars. **[2]**

(iii) Test at 1% level of significance, the claim that the mean mass of biscuit bars is 32 grams. You should state your hypotheses and define any symbols you use. **[5]**

(iv) Explain why there is no need for the production manager to know anything about the population distribution of the masses of the biscuit bars. **[2]**

Solution

(i) For a sample to be random, every biscuit bar has equal chance of being selected into the sample, and the selection of one biscuit bar is independent of the selection of another biscuit bar.

(ii) Let $y=x-32$

Then $\overline{y}=\dfrac{-7.7}{40}=-0.1925$ and $\overline{x}=\overline{y}+32=31.8075$ (note that this is an exact answer)

Therefore an unbiased estimate of the population mean is 31.8075 grams.

$$s_y^2=\frac{1}{n-1}\left[\sum y^2-\frac{\left(\sum y\right)^2}{n}\right]=\frac{1}{39}\left[11.05-\frac{(-7.7)^2}{40}\right]\approx 0.24533=s_x^2$$

Therefore an unbiased estimate of the population variance is 0.245 grams^2 (3.sf)

(iii) Let μ denote the population mean mass (in grams) of a biscuit bar

$H_0: \mu=32$

$H_1: \mu\neq 32$

Perform a 2-tail test at 1% significance level.

Under H_0, since $n=40$ is large $\overline{X}\sim N\left(32,\dfrac{0.24533}{40}\right)$ approximately by Central Limit Theorem.

Using a z-test, p-value $= 2\mathrm{P}\left(\overline{X} \le 31.8075\right) \approx 0.0140 > 0.01$

$\therefore$ We do not reject H_0 and conclude that there is insufficient evidence at 1% level of significance, that the **mean** mass of a biscuit bar is not 32 grams.

(iv) No assumption about the population is needed, as sample size of 40 is large and hence $\overline{X}$ follows a normal distribution approximately by Central Limit Theorem.

2 SRJC/2006/02/Q31OR(a)

To investigate the length of time spent by students in the school canteen during lunch break, a random sample of 90 students was taken. The time, x minutes, spent by each student was measured and it was found that

$$\sum (x-40) = 270 \qquad \text{and} \qquad \sum x^2 = 167095\,.$$

A previous survey indicated that students spent an average of 42.5 minutes in the canteen during lunch break. Test, at the 5% level, whether the average time spent has increased. **[6]**

State what you understand by the expression 'at the 5% significance level' in the context of this question. **[1]**

Solution

From sample, $\bar{x} = \dfrac{270}{90} + 40 = 43$, $\quad s^2 = \dfrac{1}{89}\left(167095 - \dfrac{(90\times 43)^2}{90}\right) = \dfrac{685}{89} \approx 7.6966$

Let μ denote the population mean time spent (in min) by a student in the school canteen during lunch break.

H_0: $\mu = 42.5$ vs H_1: $\mu > 42.5$

Perform a 1-tail test at 5% significance level.

Under H_0, since $n = 90$ is large, $\overline{X} \sim \mathrm{N}\left(42.5, \dfrac{7.6966}{90}\right)$ approximately by Central Limit Theorem.

Using a z-test, p-value $= \mathrm{P}\left(\overline{X} \ge 43\right) = 0.0437$ (3sf)

Since p-value $\approx 0.0437 \le 0.05$, we reject H_0 and conclude that there is sufficient evidence at 5% significance level that the average time spent has increased.

5% significance level means that there is a probability of 0.05 to conclude that the **average** time spent by the students in the canteen is more than 42.5 mins when it is 42.5 mins.

3 **9758/2019/02/Q9**

A company produces resistors rated at 750 ohms for use in electronic circuits. The production manager wishes to test whether the mean resistance of these resistors is in fact 750 ohms. He knows that the resistances are normally distributed with variance 100 ohms^2.

(i) Explain whether the manager should carry out a 1-tail test or a 2-tail test. State hypotheses for the test, defining any symbols you use. **[2]**

The production manager takes a random sample of 8 of these resistors. He finds that the resistances, in ohms, are as follows.

742 771 768 738 769 752 742 766

(ii) Find the mean of the sample of 8 resistors. Carry out the test, at the 5% level of significance, for the production manager. Give your conclusion in context. **[5]**

The company also produces resistors rated at 1250 ohms. Nothing is known about the distribution of the resistances of these resistors.

(iii) Describe how, and why, a test of the mean resistance of the 1250 ohms resistors would need to differ from that for the 750 ohms resistors. **[2]**

Solution:

(i) The manager should carry out a 2-tail test because he wishes to know whether the mean resistance of these resistors is in fact 750 ohms or different from 750 ohms.

Let X denote the resistance of a resistor rated at 750 ohms and μ the population mean resistance of resistors.

Null hypothesis, H_0: $\mu = 750$

Alternative hypothesis, H_1: $\mu \neq 750$

(ii) Sample mean resistance, $\bar{x} = \frac{\sum x}{n} = \frac{6048}{8} = 756$.

Perform a 2-tail test at 5% significance level.

Under H_0, $\bar{X} \sim N\left(750, \frac{100}{8}\right)$.

Using a z-test, p-value $= 2P\left(\bar{X} \geq 756\right) = 0.0897 \text{ (3 sf)} > 0.05$.

$\therefore$ We do not reject H_0.

There is insufficient evidence, at 5% level of significance, to conclude that the mean resistance of these resistors is not 750 ohms.

(iii) As nothing is known about the distribution of the resistances of the resistors rated at 1250 ohms, a random sample of large enough size (say, 30) must be taken so that Central Limit Theorem can be applied to approximate the distribution of the sample mean resistance of resistors rated at 1250 ohms to a normal distribution.

4 **9233/2004/02/Q30 OR**

The mean of a random variable X is denoted by μ. A sample of 50 random observations of X is taken and the results are summarized by $\sum x = 527.1$.

(i) It is given that the population variance is 15. Carry out a 2-tail test of the null hypothesis $\mu = 9.5$, at the 5% significance level. State, giving a reason, whether any assumptions about the population are needed in order for the test to be valid. **[4]**

(ii) It is given instead that $\sum x^2 = 6172.31$. In a 1-tail test of the null hypothesis $\mu = 11.5$, the alternative hypothesis is accepted. State the alternative hypothesis, and find an inequality satisfied by the significance level of the test. **[8]**

Solution

(i) $\bar{x} = \dfrac{\sum x}{n} = \dfrac{527.1}{50} = 10.542$

To test $H_0 : \mu = 9.5$ vs $H_1 : \mu \neq 9.5$

Perform a 2-tail test at the 5% significance level.

Under H_0, since $n = 50$ is large,

$\bar{X} \sim N\left(9.5, \dfrac{15}{50}\right)$ approximately, by Central Limit Theorem

p-value $= 2P\left(\bar{X} \geq 10.542\right) \approx 0.0571 > 0.05$

$\therefore$ We do not reject H_0 and conclude that there is insufficient evidence, at 5% level of significance, that the population mean is not 9.5.

No assumption about the population is needed, as sample size 50 is large and hence $\bar{X}$ follows a normal distribution approximately by Central Limit Theorem.

(ii) $H_0 : \mu = 11.5$ vs $H_1 : \mu < 11.5$

Perform a 1-tail test at the α% significance level.

From sample, $s^2 = \dfrac{1}{n-1}\left[\sum x^2 - \dfrac{\left(\sum x\right)^2}{n}\right] = \dfrac{1}{49}\left[6172.31 - \dfrac{(527.1)^2}{50}\right] \approx 12.564$

Under H_0, since $n = 50$ is large, $\bar{X} \sim N\left(11.5, \dfrac{12.564}{50}\right)$ approximately by Central Limit Theorem.

Using a $z-$test, $p-$value $= P\left(\bar{X} \leq 10.542\right) \approx 0.027995$.

For H_1 to be accepted, H_0 is rejected. Thus $p-$value $\approx 0.027995 \leq \dfrac{\alpha}{100}$

i.e. significance level $\geq 2.80\%$

5 RI/9758/2021/02/Q6

The recruitment manager of the private car hire company, I-ber, claims that the mean weekly earnings of a full-time driver is $980. The managing director suspects that the mean weekly earnings is less than $980 and he instructs the recruitment manager to carry out a hypothesis test on a sample of drivers. It is given that the population standard deviation of the weekly earnings is $88.

(i) State suitable hypotheses for the test, defining any symbols that you use. **[2]**

The recruitment manager takes a random sample of 10 drivers. He finds that the weekly earnings in dollars, are as follows.

942 950 905 1003 883 987 924 920 913 968

(ii) Find the mean weekly earnings of the sample of these 10 drivers. Carry out the test, at 5% level of significance, for the recruitment manager. Give your conclusion in context and state a necessary assumption for the test to be valid. **[5]**

(iii) Find the smallest level of significance at which the test would result in rejection of the null hypothesis, giving your answer correct to 1 decimal place. **[1]**

Solution

(i) Let μ be the population mean weekly earnings of a full-time driver in dollars.

Null hypothesis, $H_0 : \mu = 980$

Alternative hypothesis, $H_1 : \mu < 980$

(ii) Let X be the weekly earnings of a driver (in \$).

Using GC, $\bar{x} = 939.5$

Perform a 1-tailed test at 5% significance level.

Under H_0, $\bar{X} \sim \mathrm{N}\left(980, \dfrac{88^2}{10}\right)$

Since $p\text{-value} = 0.0728 > 0.05$, we do not reject H_0 and conclude that there is insufficient evidence for the managing director to conclude, at 5% level of significance, that the mean weekly earnings of a driver is less than $980.

Assumption:

Assume that the weekly earnings of the drivers are normally distributed.

(iii) To reject H_0, p-value = $0.0728 \le \dfrac{\alpha}{100} \Rightarrow \alpha \ge 7.28$

Thus the smallest level of significance = 7.3% (1 d.p)

6 **CJC/9758/2021/02/Q10**

The Simozer vaccine developed by a pharmaceutical company is marketed to have a 91% efficacy rate against Covid-19, measured from seven days to six months after the completion of doses. The government of the country Elbonia is interested in purchasing this vaccine for its people. It decides to test a random sample of 100 citizens and measure the efficacy rate, x%. The results obtained are summarised as follows.

$$\sum x = 9052 \qquad \sum x^2 = 962373$$

(i) Explain the meaning of 'a random sample' in the context of the question. **[1]**

(ii) Calculate unbiased estimates of the population mean and variance of the efficacy rate. **[2]**

(iii) Test, at the 5% level of significance, if the company has overstated the vaccine's efficacy rate. You should state your hypotheses and define any symbols you use. **[5]**

Three months later, the government decides to check the efficacy rate for the same 100 citizens. It found that the mean efficacy rate is 90.2% and the standard deviation was m%. A test is carried out at the 1% level of significance.

(iv) Find the range of values of m if there is sufficient evidence supporting the claim that the efficacy rate was overstated by the pharmaceutical company. **[4]**

Explain why there is no need for the government of Elbonia to know anything about the population distribution of the vaccine's efficacy rate. **[2]**

Solution

(i) A random sample means that

- each of the citizens is selected **independently** from each other and
- each citizen has an **equal** chance of being selected.

(ii) An unbiased estimate of the population mean is $\bar{x} = \frac{\sum x}{n} = \frac{9052}{100} = 90.52$

An unbiased estimate of the population variance is

$s^2 = \frac{1}{99}\left(962373 - \frac{9052^2}{100}\right) = 1444.3 = 1440 \text{ (3s.f.)}$

(iii) Let X be the random variable denoting the efficacy rate of the Simozer vaccine.

$\mathrm{H}_0 : \mu = 91$

$\mathrm{H}_1 : \mu < 91$,

where μ is the population mean efficacy rate of the Simozer vaccine.

Perform a 1-tail test at 5% level of significance.

Under H_0, Since n = 100 is large,

$\bar{X} \sim \mathrm{N}\left(91, \frac{1444.3}{100}\right)$ approximately. by Central Limit Theorem.

p-value $= \mathrm{P}\left(\bar{X} \le 90.52\right) = 0.44975 > 0.05$

Do not reject H_0 and conclude that there is insufficient evidence at 5% level of significance to claim that the mean efficacy rate of the Simozer vaccine has been overstated.

(iv) $\bar{x} = 90.2,\ s^2 = \frac{100}{99}m^2$

$H_0 : \mu = 91$

$H_1 : \mu < 91$

Under H_0, Since n = 100 is large,

$\bar{X} \sim \mathrm{N}\left(91,\ \frac{\frac{100}{99}m^2}{100}\right)$ approximately by Central Limit Theorem

At 1% significance level, reject H_0 if $p - \text{value} \le 0.01$

$$\mathrm{P}\left(\bar{X} \le 90.2\right) \le 0.01$$

$$\mathrm{P}\left(Z \le \frac{90.2 - 91}{\sqrt{\frac{m^2}{99}}}\right) \le 0.01$$

$$\frac{-0.8}{\frac{m}{\sqrt{99}}} \le -2.3263$$

$$0.8 \ge 2.3263\left(\tfrac{m}{\sqrt{99}}\right)$$

$$m \le 3.4217$$

$0 < m \le 3.42$ (since the variance of a random variable is always positive)

Since the sample size of 100 is large, by Central Limit Theorem, the distribution of the sample mean efficacy rate follows a normal distribution approximately. Hence, the hypothesis test could still be carried out without knowing the distribution of the vaccine's efficacy rate.

7 DHS/9758/2021/02/Q11

In this question you should state clearly the variables and the values of the parameters of any distribution you use.

A pharmaceutical company has 300 employees from the research department and 330 employees from the manufacturing department.

(i) Billy wishes to find out about the stress level of employees in the research department, so he sends a survey to the 300 employees in the research department. Explain whether these 300 employees form a sample or a population. **[1]**

(ii) Charlie wishes to conduct a survey regarding the employees' experiences with a new facility. 100 of them are randomly selected from the research department and the other 100 are randomly selected from the manufacturing department. Explain whether the method will form a random sample. **[2]**

The pharmaceutical company produces pills against a particular bacteria. The manager claims that the average time taken to manufacture a batch of pills is less than 12 hours. The time taken in hours, x was measured on 80 occasions and the results are summarised as follows:

$$\sum(x-12)=-12.8,\quad \sum(x-12)^2=16.19.$$

(iii) Find the unbiased estimates of the population mean and variance of the time taken to manufacture a batch of pills. **[2]**

(iv) Carry out a hypothesis test on the manager's claim. Find the p-value for this test and explain what it indicates about the manager's claim. **[4]**

It was found that the testing process had not been rigorous enough and the same hypothesis test was conducted in another random sample of 80 observations. The sample mean was found to be k hours and the sample standard deviation was found to be 1.5 hours. It was found that the manager's claim was not supported at 1% level of significance.

(v) Find the range of values of k, correct to 3 decimal places. **[3]**

Solution

(i) Billy is interested in the stress levels of employees in the research department, and surveyed **all 300 employees** in the research department, thus the 300 employees constitute a population.

(ii) Probability of selecting an employee from research department = $\frac{100}{300}=\frac{1}{3}$

Probability of selecting an employee from manufacturing department = $\frac{100}{330}=\frac{10}{33}$

Since not every employee has the same chance of being selected, the method does not give us a random sample.

(iii) An unbiased estimate of the population mean is $\overline{x}=\dfrac{\sum(x-12)}{80}+12=\dfrac{-12.8}{80}+12=11.84$

An unbiased estimate of the population variance is

$$s^2=\frac{1}{80-1}\left[\sum(x-12)^2-\frac{\left(\sum(x-12)\right)^2}{80}\right]$$

$$=\frac{1}{79}\left[16.19-\frac{(-12.8)^2}{80}\right]$$

$$=0.17901=0.179 \text{ (3 s.f.)}$$

(iv) $H_0: \mu=12$

$H_1: \mu<12$, where μ is the population mean time taken

Under H_0, $\overline{X}\sim N\left(12,\dfrac{0.17901}{80}\right)$ approximately by Central Limit Theorem, since sample size of 80 is large.

From GC, p-value $=P\left(\overline{X}\le 11.84\right)=0.000359$ (3 s.f.)

Since the ***p*-value is very small,** there is very strong evidence to reject H_0 and conclude that the **manager's claim is valid.**

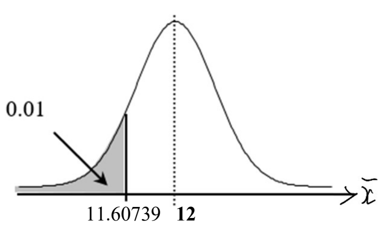
$\overline{x}=k,\ n=80$

$$s^2=\frac{n}{n-1}(\text{sample variance})=\frac{80}{79}\left(1.5^2\right)$$

Under H_0, $\overline{X}\sim N\left(12,\dfrac{1.5^2}{79}\right)$ approximately by Central Limit Theorem since $n=80$ is large.

Manager's claim not supported $\Rightarrow$ we do not reject H_0

<u>Method 1 (critical region)</u>



Critical region : $\overline{x}\le 11.60739$

Since H_0 is not rejected, $\overline{x}=k$ does not fall inside the critical region.

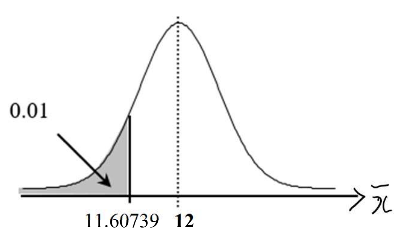
Thus $k>11.608$ (3 dp)

<u>Method 2 (*p*-value)</u>

Since H_0 is not rejected,

p-value >0.01

$P(\overline{X}\le k)>0.01$



$k>11.608$ (3 dp)

8 A factory makes tennis balls, and sends them to distributors in batches. Before they are sent, a random sample of 5 balls is taken from each batch and tested for bounce. Each ball is dropped from a standard height onto a hard surface, and the height X cm of its bounce is measured. It is assumed that X has a normal distribution with standard deviation 10.0. The sample mean is denoted by $\overline{X}$ cm and the population mean by μ cm. The null hypothesis $\text{H}_0 : \mu = 250$ is to be tested, at the 5% significance level, against the alternative hypothesis $\text{H}_1 : \mu \neq 250$.

(i) Find the value of a, correct to 2 decimal places, such that H_0 should be rejected if $\overline{X}$ lies outside the interval $(250-a,\ 250+a)$.

(ii) Find the probability of not rejecting H_0 if, in fact, $\mu = 247$.

(iii) The alternative hypothesis is now changed to $\text{H}_1 : \mu < 250$. Find the largest value of b, correct to 2 decimal places, such that H_0 should be rejected if $\overline{X} \leq b$.

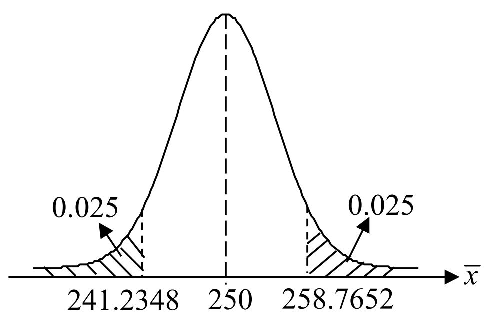
Solution

$\text{H}_0 : \mu = 250$ vs $\text{H}_1 : \mu \neq 250$ at 5% significance level.

Under H_0, $\overline{X} \sim \text{N}\left(250, \dfrac{10^2}{5}\right)$.

(i) $\text{P}\left(\overline{X} \leq 250 - a\right) = 0.025$

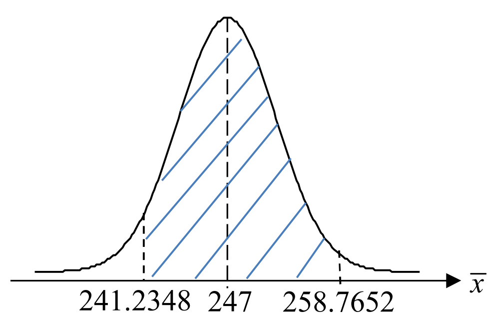
$$250 - a = 241.2348$$
$$a \approx 8.7652$$
$$a = 8.77 \quad \text{(2dp)}$$



[Note that comparing $250 + a$ with 258.7652 gives the same answer.]

(ii) If $\mu = 247$, $\overline{X} \sim N\left(247, \dfrac{10^2}{5}\right)$

$\text{P}(\text{not rejecting H}_0)$
$= \text{P}\left(250 - a < \overline{X} < 250 + a\right)$
$= \text{P}\left(241.2348 < \overline{X} < 258.7652\right)$
$= 0.897 \quad \text{(3sf)}$



(iii) To test $\text{H}_0 : \mu = 250$ vs $\text{H}_1 : \mu < 250$

Perform a 1-tail test at 5% significance level.

Under H_0, $\overline{X} \sim \text{N}\left(250, \frac{10^2}{5}\right)$.	To reject H_0, p-value $= \text{P}\left(\overline{X} \le b\right) \le 0.05$ $\Rightarrow b \le 242.643991$ $\Rightarrow$ largest $b = 242.64$ (2dp) [Note that $\text{P}\left(\overline{X} \le 242.643991\right) = 0.05$]

9 HCI/2006/02/Q22(i) (modified)

Clara, who has diabetes, has to monitor her blood glucose levels, which vary throughout the day. The results from a sample of 75 readings, x, taken at random times over a week, are summarized by

$$\sum x = 511.5 \quad \text{and} \quad \sum x^2 = 4027.89.$$

Assuming a normal distribution, find the range of values of μ_0 for which it would be accepted that $\mu \ne \mu_0$ at the 10% significance level. **[5]**

Solution

$\bar{x} = \dfrac{511.5}{75} = 6.82$, $s^2 = \dfrac{1}{74}\left(4027.89 - \dfrac{511.5^2}{75}\right) = 7.29$

$\text{H}_0 : \mu = \mu_0$ vs $\text{H}_1 : \mu \ne \mu_0$

Perform a 2-tail test at 10% significance level.

Under H_0, $\overline{X} \sim \text{N}\left(\mu_0\ ,\ \dfrac{7.29}{75}\right)$ approximately.

[Now with μ_0 unknown, we can only define the p-value if we consider the two possible scenarios, with either $\bar{x} = 6.82 < \mu_0$, or $\bar{x} = 6.82 \ge \mu_0$]

Case 1: $\bar{x} = 6.82 < \mu_0$

To reject H_0, p-value ≤ 0.10

$2\text{P}\left(\overline{X} \le 6.82\right) \le 0.10$

$\text{P}\left(\overline{X} \le 6.82\right) \le 0.05$

$$\text{P}\left(Z \le \frac{6.82 - \mu_0}{\sqrt{\frac{7.29}{75}}}\right) \le 0.05$$

$$\frac{6.82 - \mu_0}{\sqrt{\frac{7.29}{75}}} \le -1.6449$$

$$\mu_0 \ge 6.82 + 1.6449\sqrt{\frac{7.29}{75}} = 7.33 \quad \text{(3sf)}$$

Case 2: $\bar{x} = 6.82 \ge \mu_0$

To reject H_0, p-value ≤ 0.10

$2\text{P}\left(\overline{X} \ge 6.82\right) \le 0.10$

$\text{P}\left(\overline{X} \ge 6.82\right) \le 0.05$

$$\text{P}\left(Z \ge \frac{6.82 - \mu_0}{\sqrt{\frac{7.29}{75}}}\right) \le 0.05$$

$$\frac{6.82 - \mu_0}{\sqrt{\frac{7.29}{75}}} \ge 1.6449$$

$$\mu_0 \le 6.82 - 1.6449\sqrt{\frac{7.29}{75}} = 6.31 \quad \text{(3sf)}$$

$\therefore$ $\mu_0 \le 6.31$ or $\mu_0 \ge 7.33$

10 9205/N1990/02/Q10b

For this question, give non-exact answers to 3 decimal places.

A normal distribution has unknown mean μ and known variance σ^2. A random sample of n observations from the distribution has sample mean $\bar{x}$. The null hypothesis $\mu = \mu_0$ is being tested. Find, in terms of μ_0, σ and n, the set of values of $\bar{x}$ for which $\mu = \mu_0$ is rejected in favour of $\mu \neq \mu_0$ at the 1% level of significance. **[5]**

Find also, in terms of $\bar{x}$, σ and n, the set of values of μ_0 for which $\mu = \mu_0$ is rejected in favour of $\mu < \mu_0$ at the 5% level of significance. **[3]**

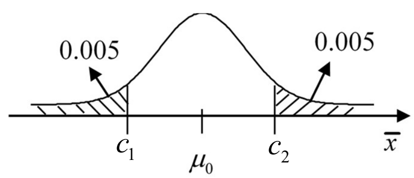
Solution

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

Perform a 2-tail test at 1% significance level.

Under H_0, $\bar{X} \sim N\left(\mu_0, \dfrac{\sigma^2}{n}\right)$.

To get the critical values c_1 and c_2:



$P\left(\bar{X} \le c_1\right) = 0.005$

$P\left(Z \le \dfrac{c_1 - \mu_0}{\sigma/\sqrt{n}}\right) = 0.005$

$\dfrac{c_1 - \mu_0}{\sigma/\sqrt{n}} = -2.57583$

$c_1 = \mu_0 - \dfrac{2.576\sigma}{\sqrt{n}}$

Thus $c_2 = \mu_0 + \dfrac{2.576\sigma}{\sqrt{n}}$ (Note that c_1 and c_2 are symmetric about μ_0)

The set of values of $\bar{x}$ for which $\mu = \mu_0$ is rejected in favour of $\mu \neq \mu_0$ at the 1% level of significance is $\left(-\infty, \mu_0 - \dfrac{2.576\sigma}{\sqrt{n}}\right] \cup \left[\mu_0 + \dfrac{2.576\sigma}{\sqrt{n}}, \infty\right)$. [The critical region]

$H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

Perform a 1-tail test at 5% significance level.

Under H_0, $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$.

To reject H_0, $p-\text{value} = P\left(\bar{X} \le \bar{x}\right) \le 0.05$

$$P\left(Z \le \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right) \le 0.05$$

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \le -1.64485$$

$$\mu_0 \ge \bar{x} + \frac{1.64485\sigma}{\sqrt{n}}$$

Set of values of μ_0 for which $\mu = \mu_0$ is rejected in favour of $\mu < \mu_0$ at the 5% level of significance is $\left[\bar{x} + \frac{1.645\sigma}{\sqrt{n}}, \infty\right)$

11 SAJC/9758/2018/02/Q11 modified

A road named Spring Avenue has a speed limit of 40 km/h in a housing estate.

The residents were concerned that many vehicles travelled too fast along the road and they decided to set up a speed tracking device to monitor the speed of vehicles travelling along this road. The data generated from the device indicated that the mean speed of vehicles travelling through this road was 44.1 km/hour.

In an attempt to reduce the mean speed of vehicles travelling through Spring Avenue, life-size photographs of a police officer were put up next to the road. The speed, X km/hour of a sample of 100 vehicles was then measured and the following data obtained.

$$\sum x = 4327.0, \qquad \sum\left(x-\bar{x}\right)^2 = 925.71.$$

(i) Calculate the unbiased estimates of the population mean and variance of the speed of vehicles travelling along Spring Avenue. **[2]**

(ii) State an assumption that must be made about the sample in order to carry out a hypothesis test to investigate whether the desired reduction in mean speed had occurred. **[1]**

(iii) Given that the assumption that you stated in part **(ii)** is valid, carry out such a test, using the 5% level of significance. **[5]**

(iv) Explain what is meant by "5% level of significance" in the context of this question. **[1]**

(v) Subsequently, the residents detected that a measurement error has occurred when measuring the speed of the 100 randomly selected vehicles. To rectify the error, a multiplication of a positive constant k to each reading for the 100 randomly selected cars is recommended. Find the greatest possible value of k, to 3 significant figures, for the conclusion obtained in **(iii)** to remain the same. **[3]**

Solution	
(i)	An unbiased estimate of population mean = $\bar{x} = \frac{\sum x}{100} = \frac{4327.0}{100} = 43.27$ An unbiased estimate of population variance, $s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{925.71}{99} = 9.3506 = 9.35$ (to 3 sf)
(ii)	The 100 vehicles are selected randomly, i.e. each vehicle travelling through Spring Avenue has an equal chance of being selected and each vehicle is independently chosen.
(iii)	Let μ be the population mean speed Test H_0: $\mu = 44.1$ H_1: $\mu < 44.1$ Carry out one-tailed z-test at 5% significance level. Under H_0, $\overline{X} \sim N(44.1, \frac{9.3506}{100})$ approximately by Central Limit Theorem, since sample size 100 is sufficiently large. p-value = $P\left(\overline{X} \le 43.27\right)$=0.00332. Since p-value $\le$ 0.05, we reject H_0 and conclude that at the 5% significance level, there is sufficient evidence that the mean speed of vehicles has been reduced after life sized police photos have been erected along Spring Avenue.
(iv)	5% significance level refers to the probability of 0.05 of wrongly concluding that the mean speed of vehicles is less than 44.1 km/hour when it is in fact 44.1 km/hour.
(v)	New unbiased estimate of population mean $= 43.27k$ New unbiased estimate of population variance = $\frac{30857}{3300}k^2$ Test H_0: $\mu = 44.1$ H_1: $\mu < 44.1$ Under H_0, $\overline{X} \sim N(44.1, \frac{\frac{30857}{3300}k^2}{100})$ approximately by Central Limit Theorem, since sample size 100 is sufficiently large.

	To have the same conclusion, i.e. to reject H_0, p value=$\mathrm{P}\left(\bar{X} \le 43.27k\right) \le 0.05$ $\mathrm{P}\left(Z \le \dfrac{43.27k-44.1}{k\sqrt{\frac{30857}{3300}}\Big/10}\right) \le 0.05$ $\dfrac{(43.27k)-44.1}{k\sqrt{\frac{30857}{3300}}\Big/10} \le -1.64485$ $43.27k-44.1 \le -\dfrac{1.64485\sqrt{\frac{30857}{3300}}}{10}k$ $k \le 1.0159$ Since $k>0,\ 0<k\le 1.01$ Greatest possible value of k is 1.01.

12 EJC JC2 Prelim 9758/2019/02/Q10

A company manufactures packets of potato chips with X mg of sodium in each packet. It is known that the mean amount of sodium per packet is 1053 mg. After some alterations to the production workflow, 50 randomly chosen packets of potato chips were selected for analysis. The amount of sodium in each packet was measured, and the results are summarised as follows:

$$\sum(x-1050)=58.0,\ \sum(x-1050)^2=2326$$

(i) Test at 5% level of significance whether the mean amount of sodium in a packet of potato chips has changed, after the alterations to the workflow. **[6]**

(ii) Explain what ‘5% level of significance’ means in this context. **[1]**

(iii) A second tester conducted the same test at $\alpha\%$ level of significance, for some integer α. However, he came to a different conclusion from the first tester. What is the range of α for which the second tester could have taken? **[1]**

(iv) Without performing another hypothesis test, explain whether the conclusion in part **(i)** would be different if the alternative hypothesis was that the mean amount of sodium had decreased after the alterations to the workflow. **[1]**

It is given instead that the standard deviation of amount of sodium in a packet of potato chips is 6.0 mg. The mean amount of sodium of a second randomly chosen sample of 40 packets of potato chips is $\bar{y}$.

(v) A hypothesis test on this sample, at 5% level of significance, led to a conclusion that the amount of sodium has decreased.
Find the range of values of $\bar{y}$, to 1 decimal place.

Explain if it is necessary to make any assumptions about the distribution of the amount of sodium in each packet of potato chips. **[3]**

Solution
(i) $\bar{x}=1050+\dfrac{58.0}{50}=1051.16$ $s^2=\dfrac{1}{n-1}\left(\sum(x-1050)^2-\dfrac{\left[\sum(x-1050)\right]^2}{n}\right)$ $=\dfrac{1}{49}\left(2326-\dfrac{58.0^2}{50}\right)$ $=46.096$ (5 s.f.) To test H_0: $\mu=1053$ against H_1: $\mu\neq1053$ Perform 2-tail test at 5% level of significance Since $n=50$ is large, by Central Limit Theorem, under H_0, $\bar{X}\sim N\left(1053,\dfrac{46.096}{50}\right)$ approximately p-value = $2P\left(\bar{X}<1051.16\right)=0.055322$ Since p-value > 0.05, we do not reject H_0 and conclude at 5% level of significance that there is insufficient evidence that the population mean amount of sodium per packet has changed after alterations to the workflow.
(ii) The probability of wrongly concluding that the mean amount of sodium is not 1053mg, when it is in fact 1053mg, is 0.05.
(iii) Since p-value is 0.055322, $\alpha\geq6$
(iv) If we tested H_1: $\mu<1053$, p-value$=P\left(\bar{X}<1051.16\right)=0.027661<0.05$ So we may reject H_0 and conclude there is sufficient evidence at 5% level of significance that the population mean amount of sodium had decreased.
(v) To test H_0: $\mu=1053$ against H_1: $\mu<1053$ at 5% level of significance Since $n=50$ is large, by Central Limit Theorem, under H_0, $\bar{Y}\sim N\left(1053,\dfrac{6.0^2}{40}\right)$ approximately To reject H_0, p-value = $P\left(\bar{Y}<\bar{y}\right)<0.05$ $\Rightarrow\bar{y}<1051.4$ (to 1 d.p.) It is not necessary to assume anything about the population distribution, as sample size (= 40) is large enough, so the Central Limit Theorem says the sample mean amount of sodium approximately follows a normal distribution.

The End