A Level H2 Math

Hypothesis Testing Test 3

Q1

Yummy Berries Farm produces blueberries and raspberries packed in boxes.

(a) Yummy Berries Farm claims that the mass, x grams, of each box of blueberries is no less than 125 grams. After receiving ~~a~~ complaints from consumers, the Consumers Association of Singapore (CASE) took a random sample of 50 boxes of blueberries from Yummy Berries Farm and the mass of each box was recorded. The data obtained are summarised in the table.

x (grams)	120	121	122	123	124	125	126	127	128	129	130
No. of boxes	3	6	6	6	3	10	3	4	6	2	1

(i) Find unbiased estimates of the population mean and variance. [2]

(ii) Test, at the 10% level of significance, whether Yummy Berries Farm has overstated its claim.
State, giving a reason, whether any assumptions about the masses of boxes of blueberries are needed in order for the test to be valid. [6]

(b) The masses of boxes of raspberries, each of y grams, are assumed to have a mean of 170 grams with standard deviation 15 grams. CASE took a random sample of n boxes of raspberries and the mean mass of boxes of raspberries from the sample is found to be 165 grams. A test is to be carried out at the 5% level of significance to determine if the mean mass of the boxes of raspberries is not 170 grams. Find the minimum number of boxes of raspberries to be taken for which the result of the test would be to reject the null hypothesis. [4]

Q2

At a canning factory, cans are filled with potato puree. The machine which fills the cans is set so that the volume of potato puree in a can has mean 420 millilitres. After the machine is recalibrated, a quality control officer wishes to check whether the mean volume has changed. A random sample of 30 cans of potato puree is selected and the volume of the puree in each can is recorded. The sample mean volume is $\overline{x}$ millilitres and the sample variance is 12 millilitres^2.

(i) Given that $\overline{x} = 418.55$, carry out a test at the 1% level of significance to investigate whether the mean volume has changed. State, giving a reason, whether it is necessary for the volumes to have a normal distribution for the test to be valid. [6]

(ii) Use an algebraic method to calculate the range of values of $\overline{x}$, giving your answer correct to 2 decimal places, for which the result of the test at the 1% level of significance would be to reject the null hypothesis. [3]

Q3

(a) The Health Promotion Board of a certain country claims that the average number of hours of sleep of working adults is at most 6 hours per day. To investigate this claim, the editor of a magazine plans to conduct a survey on a sample of adults travelling to work by train.

(i) Explain why this method of sampling will not give a random sample for the purpose of the investigation. [1]

The editor of another magazine interviewed a random sample of 50 working adults and their number of hours of sleep per day, x, are summarised as follows:

$$\sum x = 320, \sum x^2 = 2308.5$$

(ii) Test at the 5% level of significance whether there is any evidence to doubt the Health Promotion Board's claim. State with a reason, whether it is necessary to assume that the number of hours of sleep per day follows a normal distribution. [5]

(b) The Health Promotion Board carried out their own survey on another random sample of 50 working adults. The sample yielded an average of 6.14 hours of sleep per day and a standard deviation of 2.1 hours.

If the sample does not provide significant evidence at the 5% level of significance that the mean number of hours of sleep per day of working adults differs from μ_o hours, find the range of values of μ_o [4]

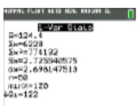
Answers

Hypothesis Testing Test 3

Q1

(a)(i)
Using GC,
Unbiased estimate of the population mean,
$\bar{x} = 124.4$ g

Unbiased estimate of the population variance,
$s^2 = 2.725540575^2$
$= 7.428571429$
$= 7.43 \quad (3 \text{ s.f.})$



(a)(ii)
Let μ g be the population mean mass of a box of blueberries.
$H_0 : \mu = 125$
$H_1 : \mu < 125$
Under H_0, test statistic
$$Z = \frac{\bar{X} - 125}{\sqrt{\frac{7.428571429}{50}}} \sim N(0,1) \text{ approximately by CLT}$$
Level of significance: 10%
Critical region: Reject H_0 if p-value ≤ 0.1

Since p-value $= 0.0598 < 0.1$, we reject H_0 and conclude that at the 10% level of significance, there is sufficient evidence that Yummy Berries Farm has overstated its claim.

No assumptions about masses of boxes of blueberries are needed. Since $n = 50$ is sufficiently large, by Central Limit Theorem, the <u>mean</u> mass of boxes of raspberries will follows a normal distribution approximately.

(b)

Let μ_1 g be the population mean mass of a box of raspberries.

$H_0 : \mu_1 = 170$

$H_1 : \mu_1 \neq 170$

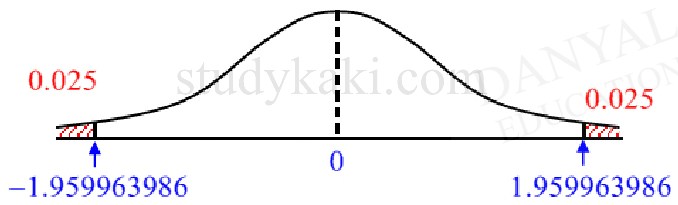
Under H_0, assuming n is large,

test statistic $Z = \dfrac{\overline{Y} - 170}{\frac{15}{\sqrt{n}}} \sim N(0,1)$ approximately by CLT

Level of significance: 5%

Critical region: Reject H_0 if p-value ≤ 0.05

i.e. Reject H_0 if z-value ≤ -1.959963986 or z-value ≥ 1.959963986



$$\frac{165-170}{\frac{15}{\sqrt{n}}} \leq -1.959963986 \quad \text{or} \quad \frac{165-170}{\frac{15}{\sqrt{n}}} \geq 1.959963986$$

$$\sqrt{n} \geq 5.87989 \quad \text{or} \quad \sqrt{n} \leq -5.87989 \text{ (rejected)}$$

$\therefore\ n \geq 34.573$

Hence least n is 35.

Q2

(i) $H_0 : \mu = 420$

$H_1 : \mu \neq 420$

$s^2 = \frac{30}{29}(12) = 12.414$

Under H_0, since $n = 30$ is large, by Central Limit Theorem,

$\bar{X} \sim N\left(420, \frac{12.414}{30}\right)$ approximately.

Hence it is <u>not necessary</u> for the volumes to have a normal distribution for the test to be valid.

Test statistic $Z = \frac{\bar{X} - 420}{\sqrt{\frac{12.414}{30}}} \sim N(0,1)$ approximately

$\alpha = 0.01$

From GC, $z = \frac{418.55 - 420}{\sqrt{\frac{12.414}{30}}} = -2.2541$

p-value = 0.0242 (3 sf)

Since p-value = 0.0242 > $\alpha = 0.01$, we <u>do not reject</u> H_0 at 1% level of significance and conclude that there is <u>insufficient</u> evidence that the population mean volume has changed.

(ii) $\alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$

Reject H_0 if $z \leq -2.5758$ or $z \geq 2.5758$

$\frac{\bar{x} - 420}{\sqrt{\frac{12.414}{30}}} \leq -2.5758$ or $\frac{\bar{x} - 420}{\sqrt{\frac{12.414}{30}}} \geq 2.5758$

$\bar{x} \leq 420 - 2.5758\sqrt{\frac{12.414}{30}}$ or $\bar{x} \geq 420 + 2.5758\sqrt{\frac{12.414}{30}}$

$\underline{\underline{\bar{x} \leq 418.34}}$ or $\underline{\underline{\bar{x} \geq 421.66}}$

Q3

a(i) Only working adults travelling by train will have a chance of being selected. Those who do not travel by train will have no chance of being chosen. Hence not every working adult in the country has an equal chance to be selected – therefore the sample is not a random sample.

a(ii) Let X hours be the number of hours of sleep of a randomly chosen adult and μ be the mean of X.

To test $H_o : \mu = 6$ vs $H_1 : \mu > 6$

Since sample size is large, by CLT, $\overline{X} \sim N\left(6, \frac{\sigma^2}{50}\right)$

Since population variance σ^2 is unknown, it is replaced by s^2

Under H_o, test statistic $Z = \frac{\overline{X} - 6}{\frac{s}{\sqrt{50}}} \sim N(0,1)$

We use a one-tailed test at 5% level of significance,
that is, reject H_o if p-value < 0.05

Sample readings: $\overline{x} = \frac{320}{50} = 6.4$,

$s^2 = \frac{1}{49}\left(2308.5 - \frac{(320)^2}{50}\right) = 5.31633$

From GC, p-value $= 0.109967 = 0.110 > 0.05$
$\Rightarrow$ we **do not reject** H_o.
Hence we conclude that there is **insufficient** evidence at the 5% level of significance to doubt the Health Promotion Board's claim.

It is not necessary to assume that the number of hours of sleep follow a normal distribution because since the sample size is large, by the Central Limit Theorem, the sample mean follows a normal distribution approximately.

b To test $H_o : \mu = \mu_o$ vs $H_1 : \mu \neq \mu_o$

$s^2 = \frac{50}{49}(2.1)^2, \overline{x} = 6.14$,

Since H_o is not rejected at the 5% level,

$$-1.95996 < \frac{\overline{x} - \mu_o}{\frac{s}{\sqrt{n}}} < 1.95996$$

$$\Rightarrow -1.95996 < \frac{6.14 - \mu_o}{\frac{2.1}{\sqrt{49}}} < 1.95996$$

$$\Rightarrow 6.14 - 1.95996\frac{2.1}{\sqrt{49}} < \mu_o < 6.14 + 1.95996\frac{2.1}{\sqrt{49}}$$

$$\Rightarrow 5.552012 < \mu_o < 6.727988$$

$$\Rightarrow 5.55 < \mu_o < 6.73$$