Additional Exercise – Hypothesis Testing (Normal Distribution)

Question 1

In previous years, the marks obtained in a French test by students attending Topnotch College have been modelled satisfactorily by a normal distribution with a mean of 65 and a standard deviation of 9. Teachers in the French department suspect that this year their students are, on average, underachieving. In order to investigate this suspicion, the teachers selected a random sample of 35 students to take the French test and found that their mean score was 61.5.

(a) Investigate, at the 5% level of significance, the teachers' suspicion using the p-value method. (6 marks)

(b) Explain, in the context of this question, the meaning of a Type I error. (2 marks)

Question 2

It is known that the lifetime of a certain species of animal in the wild is 13.3 years. A zoologist reads a study of 50 randomly chosen animals of this species that have been kept in zoos. According to the study, for these 50 animals the sample mean lifetime is 12.48 years and the population variance is 12.25.

Test, at the 5% significance level, whether these results provide evidence that animals of this species that have been kept in zoos have a shorter expected lifetime than those in the wild. Calculate and state the p-value clearly. (7 marks)

Question 3

In the past, the time spent in minutes, by customers in a certain library had mean 32.5 and standard deviation 8.2. Following a change of layout in the library, the mean time spent in the library by a random sample of 50 customers is found to be 34.5 minutes.

(a) Assuming that the standard deviation remains at 8.2, test at the 5% significance level whether the mean time spent by customers in the library has changed. State your p-value.

(7 marks)

Question 4

The times taken by employees to complete a task are normally distributed with standard deviation 2.6 minutes. A manager claims that the mean time is 15.5 minutes but an employee suspects that the mean time is greater than this. He intends to carry out a hypothesis test at the 5% significance level to test this claim. He records the times taken by a random sample of 12 employees.

(a) Find the critical region for the test. (3 marks)

(b) The total of the times taken by the 12 employees was 202.1 minutes. Carry out the test. (5 marks)

Question 5

In the past, the time spent by customers in a certain shop had mean 10.5 minutes and standard deviation 4.2 minutes. Following a change of layout in the shop, the mean time spent in the shop by a random sample of 50 customers is found to be 12.0 minutes.

(a) Assuming that the standard deviation is unchanged, test at the 1% significance level whether the mean time spent by customers in the shop has changed. (7 marks)

(b) Another random sample of 50 customers is chosen and a similar test at the 1% significance level is carried out. Given that the population mean time has not changed, state the probability that the conclusion of the test will be that the population mean time *has* changed. (1 mark)

Question 6

A study was made of adult men from region A of a country. It was found that their heights were normally distributed with a mean of 175.4 cm and standard deviation 6.8 cm.
(a) Find the proportion of these men that are taller than 180 cm. (1 mark)

A student claimed that the mean height of adult men from region B of this country was different from the mean height of adult men from region A. A random sample of 52 adult men from region B had a mean height of 177.2 cm. The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region A and region B .
(b) Use a critical region test to assess the student's claim. You should state your hypotheses clearly and use a 5% level of significance. (4 marks)
(c) Find the p-value for the test in part (b) and explain how this supports your conclusion given in part (b). (2 marks)

Question 7

The volume of liquid in bottles of sparkling water is known to be normally distributed with a standard deviation of 4.5 ml. The manufacturer claims the mean volume is 500 ml. A trading standards officer suspects the filling machine is underfilling the bottles. She takes a random sample of 25 bottles and finds their mean volume is 498.2 ml.

(a) Find the critical region for a hypothesis test at the 2.5% significance level. (3 marks)
(b) Hence, state your conclusion to the test. (2 marks)

Question 8

A machine fills bags with flour. The weights of the bags are normally distributed with standard deviation 8 g. The target mean weight is 1000 g. To check if the machine is properly calibrated, a random sample of 40 bags is taken and the mean weight is found to be 1003 g.
Carry out a hypothesis test at the 10% significance level to determine if the mean weight differs from 1000 g. Use the critical region method. (6 marks)

Question 9

The reaction times of individuals taking a specific test are normally distributed with a standard deviation of 0.15 seconds. The historical mean reaction time is 2.40 seconds. A researcher believes that a new energy drink alters the reaction time. She tests a random sample of 15 individuals who have consumed the drink.

(a) Find the critical region for a two-tailed test at the 5% level of significance. (4 marks)

(b) The sample yields a mean reaction time of 2.31 seconds. What is the conclusion of the test? (2 marks)

(c) Calculate the p -value for this sample mean. (2 marks)

Question 10

A farmer grows a certain type of tomato and the weights are known to be normally distributed with a standard deviation of 12 g. Historically, the mean weight has been 85 g. The farmer uses a new fertilizer and wishes to test, at the 1% significance level, whether the mean weight has increased. He picks a random sample of 60 tomatoes.

(a) Find the critical value for the sample mean. (3 marks)

(b) The sum of the weights of the 60 tomatoes is 5280 g. Determine the p-value for this test and state your conclusion. (4 marks)

Fully Worked Solutions

Note for teachers/students: In accordance with the WJEC specification, p-values for two-tailed tests are found by calculating the probability in the observed tail and doubling it.

Solution 1
(a)

$$H_0: \mu = 65$$

$$H_1: \mu < 65$$

Under H_0, the sample mean $\bar{X} \sim N(65, \frac{9^2}{35}) \Longrightarrow \bar{X} \sim N(65, \frac{81}{35})$

Using a calculator (Normal CD, $\mu = 65, \sigma = 9/\sqrt{35}$): p-value $= P(\bar{X} \leq 61.5) = 0.0107$

Comparing to the significance level: $0.0107 < 0.05$

Conclusion: Reject H_0. There is sufficient evidence at the 5% significance level to support the suspicion that students are underachieving. (b) A Type I error is rejecting H_0 when it is actually true.
In this context, it means concluding that the students are underachieving when, in reality, their average mark has not dropped and is still 65.

Solution 2
(a)

$$H_0: \mu = 13.3$$

$$H_1: \mu < 13.3$$

Population variance $\sigma^2 = 12.25$, so standard deviation $\sigma = \sqrt{12.25} = 3.5$.
Under H_0, $\bar{X} \sim N(13.3, \frac{12.25}{50})$
Using a calculator (Normal CD, $\mu = 13.3, \sigma = 3.5/\sqrt{50}$): p-value $= P(\bar{X} \leq 12.48) = 0.0488$

Comparing to the significance level: $0.0488 < 0.05$

Conclusion: Reject H_0. There is sufficient evidence at the 5% level to suggest that animals of this species kept in zoos have a shorter expected lifetime.

Solution 3

$$H_0: \mu = 32.5$$

$$H_1: \mu \neq 32.5$$

Under H_0, $\bar{X} \sim N(32.5, \frac{8.2^2}{50})$

Since $34.5 > 32.5$, we find the probability in the upper tail: $P(\bar{X} \geq 34.5) = 0.0423$

Because this is a two-tailed test, WJEC requires us to double this probability:
$$p\text{-value} = 2 \times 0.0423 = 0.0846$$
Comparing to the significance level: $0.0846 > 0.05$

Conclusion: Do not reject H_0. There is insufficient evidence at the 5% level to suggest the mean time spent by customers in the library has changed.

Solution 4

(a)

$$H_0: \mu = 15.5$$

$$H_1: \mu > 15.5$$

Under H_0, $\bar{X} \sim N(15.5, \frac{2.6^2}{12})$

We want to find the critical value c such that $P(\bar{X} \geq c) = 0.05$ (or $P(\bar{X} \leq c) = 0.95$).

Using Inverse Normal (Area = 0.95, $\mu = 15.5, \sigma = 2.6/\sqrt{12}$): $c = 16.73$

Critical Region: $\bar{X} \geq 16.73$

(b) Sample mean

$$\bar{x} = \frac{202.1}{12} = 16.84 \text{ minutes}$$

Since $16.84 > 16.73$, the sample mean lies inside the critical region.
Conclusion: Reject H_0. There is sufficient evidence at the 5% level to support the employee's claim that the mean task time is greater than 15.5 minutes.

Solution 5

(a)

$$H_0: \mu = 10.5$$
$$H_1: \mu \neq 10.5$$

Under H_0, $\bar{X} \sim N(10.5, \frac{4.2^2}{50})$

Since $12.0 > 10.5$, we find the upper tail probability: $P(\bar{X} \geq 12.0) = 0.00578$

$$p\text{-value} = 2 \times 0.00578 = 0.0116$$

Comparing to the significance level: $0.0116 > 0.01$

Conclusion: Do not reject H_0. There is insufficient evidence at the 1% level to conclude that the mean time spent in the shop has changed.

(b) The probability of concluding the mean has changed when it actually has not changed is the probability of making a Type I error. This is exactly equal to the significance level of the test.

Probability: 0.01 (or 1%).

Solution 6

(a) Let $X \sim N(175.4, 6.8^2)$.
    $P(X > 180) = 0.249$ (to 3 s.f.)

(b)

$$H_0: \mu = 175.4$$
$$H_1: \mu \neq 175.4$$

Under H_0, $\bar{X} \sim N(175.4, \frac{6.8^2}{52})$

For a 5% two-tailed test, the critical region is split (2.5% in each tail).
Lower critical value: InvNorm(Area = 0.025) = 173.55
Upper critical value: InvNorm(Area = 0.975) = 177.25

Critical Region: $\bar{X} \leq 173.55$ or $\bar{X} \geq 177.25$.
Sample mean $\bar{x} = 177.2$. Because $173.55 < 177.2 < 177.25$, it does *not* fall into the critical region. Do not reject H_0. Insufficient evidence to suggest the mean height differs.

(c) $P(\bar{X} \geq 177.2) = 0.0281$, so p-value $= 2 \times 0.0281 = 0.0562$. Since this is > 5%, this would lead to us not rejecting H_0 at the 5% level, supporting our conclusion in (b).

Solution 7

(a)

$$H_0: \mu = 500$$

$$H_1: \mu < 500$$

Under H_0, $\bar{X} \sim N(500, \frac{4.5^2}{25}) \implies \bar{X} \sim N(500, 0.81)$

We need value c where $P(\bar{X} \leq c) = 0.025$.
Using Inverse Normal, $c = 498.24$

Critical Region: $\bar{X} \leq 498.24$

(b) The observed sample mean is 498.2 ml. $498.2 \leq 498.24$, so the value is inside the critical region.
Conclusion: Reject H_0. There is sufficient evidence at the 2.5% level that the machine is underfilling the bottles.

Solution 8

$$H_0: \mu = 1000$$

$$H_1: \mu \neq 1000$$

Under H_0, $\bar{X} \sim N(1000, \frac{8^2}{40})$

For a 10% two-tailed test, we place 5% in each tail.
Lower value: InvNorm(Area = $0.05, \mu = 1000, \sigma = 8/\sqrt{40}$) = 997.92 g
Upper value: InvNorm(Area = $0.95, \mu = 1000, \sigma = 8/\sqrt{40}$) = 1002.08 g

Critical region: $\bar{X} \leq 997.92$ or $\bar{X} \geq 1002.08$
Observed $\bar{x} = 1003$ g.
Since $1003 \geq 1002.08$, it lies in the critical region.
Conclusion: Reject H_0. There is sufficient evidence at the 10% level that the mean weight of flour bags differs from 1000 g.

Solution 9

(a)

$$H_0: \mu = 2.40$$

$$H_1: \mu \neq 2.40$$

Under H_0, $\bar{X} \sim N(2.40, \frac{0.15^2}{15})$

For a 5% two-tailed test, we put 2.5% in each tail.

Lower value: InvNorm(Area = $0.025, \mu = 2.40, \sigma = 0.15/\sqrt{15}) = 2.324$

Upper value: InvNorm(Area = $0.975, \mu = 2.40, \sigma = 0.15/\sqrt{15}) = 2.476$

Critical Region: $\bar{X} \leq 2.324$ or $\bar{X} \geq 2.476$

(b) Sample mean $\bar{x} = 2.31$.

Since $2.31 \leq 2.324$, the result is in the critical region.

Conclusion: Reject H_0. There is sufficient evidence to suggest the new energy drink changes the mean reaction time.

(c) $P(\bar{X} \leq 2.31) = 0.0101$ p-value $= 2 \times 0.0101 = 0.0202$

Solution 10

(a)

$$H_0: \mu = 85$$

$$H_1: \mu > 85$$

Under H_0, $\bar{X} \sim N(85, \frac{12^2}{60})$

Find c such that $P(\bar{X} \geq c) = 0.01$ (or $P(\bar{X} \leq c) = 0.99$).

Using Inverse Normal (Area = $0.99, \mu = 85, \sigma = 12/\sqrt{60}$): **Critical Value:** $= 88.60$ g.

(b) Sample mean $\bar{x} = \frac{5280}{60} = 88$ g.

p-value $= P(\bar{X} \geq 88) = 0.0264$

Comparing to significance level:

$$0.0264 > 0.01$$

Conclusion: Do not reject H_0. There is insufficient evidence at the 1% level to conclude that the new fertilizer has increased the mean weight of the tomatoes.