A Level H2 Math

Hypothesis Testing Test 1

Q1

It has been suggested that the optimal pH value for shampoo should be 5.5, to match the pH level of healthy scalp. Any pH value that is too low or too high may have undesirable effects on the user's hair and scalp. A shampoo manufacturer wants to investigate if the pH level of his shampoo is at the optimal value, by carrying out a hypothesis test at the 10% significance level. He measures the pH value, x, of n randomly chosen bottles of shampoo, where n is large.

(a) In the case where $n = 30$, it is found that $\sum x = 178.2$ and $\sum x^2 = 1238.622$.

(i) Find unbiased estimates of the population mean and variance, and carry out the test at the 10% significance level. [6]

(ii) Explain if it is necessary for the manufacturer to assume that the pH value of a bottle of shampoo follows a normal distribution. [1]

(b) In the case where n is unknown, assume that the sample mean is the same as that found in **(a)**.

(i) State the critical region for the test. [1]

(ii) Given that n is large and that the population variance is found to be 6.5, find the greatest value of n that will result in a favourable outcome for the manufacturer at the 10% significance level. [3]

Q2

The town council is investigating the mass of rubbish in domestic dustbins. In 2016, the mean mass of rubbish in domestic dustbins was 20.0 kg per household per week. The town council starts a recycling initiative and wishes to determine whether there has been a reduction in the mass of rubbish in domestic dustbins.

The mass of rubbish in a domestic dustbin is denoted by X kg. A random sample of 50 domestic dustbins is selected and the results are summarised as follows.

$$n = 50 \qquad \sum x = 924.5 \qquad \sum x^2 = 18249.2$$

(i) Explain what is meant in this context by the term 'a random sample'. [2]

(ii) Explain why the town council is able to carry out a hypothesis test without knowing anything about the distribution of the mass of rubbish in domestic dustbins. [2]

(iii) Find the unbiased estimates of the population mean and variance and carry out the test at the 1% level of significance for the town council. [6]

(iv) Use your results in part **(iii)** to find the range of values of n for which the result of the test would be that the null hypothesis is rejected at the 1% level of significance. [3]

Q3

Physicists are conducting an experiment involving collisions between protons and anti-protons. The mean amount of energy, $\bar{x}$ MeV, released in n collisions is found to be 1864 MeV.

One model predicts the energy released would be 1860 MeV with standard deviation 40 MeV. This is tested at a 1% level of significance against a newer model that claims a higher value.

(i) Find the least value of n such that the hypothesis that the mean amount of energy released is 1860 MeV is rejected. [5]

Given instead that $n = 600$.

(ii) Calculate the p-value and state its meaning in context of the question. [3]

(iii) State, with a reason, whether it is necessary to assume the amount of energy released in collisions to be normally distributed for this test to be valid. [1]

Two-sigma is an indicative of how confident researchers feel their results are. For researchers to feel confident, they must be able to produce a "two-sigma" result – that is the experimental result must be at least two standard deviations away from the predicted mean under the null hypothesis.

(iv) Calculate the level of significance that corresponds to a "two-sigma" test. Hence, using your answer from part **(ii)** determine whether the experiment has met the "two-sigma" threshold. [3]

Answers

Hypothesis Testing Test 1

Q1

(a) (i) Let X be the random variable "pH value of a randomly chosen bottle of shampoo".

Unbiased estimate of population mean

$= \bar{x}$

$= \dfrac{178.2}{30}$

$= \underline{5.94}$

Unbiased estimate of population variance

$= s^2$

$= \dfrac{1}{29}\left(1238.622 - \dfrac{178.2^2}{30}\right)$

$= 6.21083$

$= \underline{6.21}$ (3 s.f.)

To test $H_0: \mu = 5.5$

against $H_1: \mu \neq 5.5$ at 10% significance level

Under H_0, since $n = 30 > 20$ is large,

$\bar{X} \sim N\left(5.5, \dfrac{6.21083}{30}\right)$ approx. by Central Limit Theorem

Test statistic $Z = \dfrac{\bar{X} - 5.5}{\sqrt{\dfrac{6.21083}{30}}} \sim N(0,1)$ approx.

Value of test statistic $z = \dfrac{5.94 - 5.5}{\sqrt{\dfrac{6.21083}{30}}} = 0.967$ (3 s.f.)

Either Since $-1.64 < 0.967 < 1.64$, z lies <u>outside</u> the critical region

$\Rightarrow$ Do <u>not</u> reject H_0

Or p-value $= 0.334 > 0.1 \Rightarrow$ Do <u>not</u> reject H_0

$\therefore$ There is insufficient evidence at 10% significance level to conclude that the mean pH value of the shampoo is not 5.5.

	Comments The best solutions for this question are a result of careful attention to the way students phrase their working and calculate the required values. If students take some time to understand the rationale for writing things a certain way, they would be able to appreciate the principles behind a statistical hypothesis test. Students are encouraged to spell out "unbiased estimate of" rather than just writing $\bar{x}$ or s^2. Some students even wrote "pop. mean/variance" or μ and σ^2 instead of the unbiased estimates. The correct alternative hypothesis has been hinted in the question ("…too high or too low…"). Presentation wise, a number of students wrote subscripts on μ, which is not necessary. Many students are still writing the wrong mean in the distribution for $\bar{X}$. The phrase "Under H_0" implies that we're assuming that the population mean $\mu = 5.5$, therefore $E(\bar{X}) = 5.5$. Students should also be aware of whether CLT is used. An alarming number of students attempted to write down the formula of the *p*-value, and then seemed to calculate the *p*-value using normalcdf instead of the Z-test. Students should only attempt to do this if they're very sure of the correct formula for the *p*-value in the respective tests; otherwise, they're better off using the Z-test function in the GC and letting it do its work. Some students keyed in the wrong σ into the GC, which resulted in an extremely low *p*-value. The final part of comparing *p*-value to significance level and the conclusion was also horribly done. Students generally made some permutation of the following mistakes: 1. Dividing the *p*-value by 2, or using the *p*-value for the one-tail test 2. Comparing *p*-value to 0.05 instead of 0.1 3. Comparing wrongly (e.g. $0.334 < 0.1$) 4. Mixing up the results of the test (e.g. $0.334 > 0.1$, hence reject H_0) 5. Mixing up "sufficient/insufficient evidence" and "H_0 / H_1 is true/not true". In particular, students should learn that the purpose of the test is to use the evidence to try and prove that **H_1 is true**, and hence the final conclusion must reflect this (i.e. is there sufficient evidence to conclude that **H_1 is true**?).	
(a)(ii)	It is not necessary to assume X is normally distributed. As the sample size is large, by Central Limit Theorem, $\bar{X}$ is approximately normally distributed.	This question has highlighted a fundamental conceptual error that many students have about CLT: that CLT allows us to approximate X as a normal distribution. It therefore results in answers

		ranging from "No, CLT says X is normal" to "Yes, since CLT says X is normal". Because it is very easy for students to simply give the correct answer "No" with a superficial explanation, the marking of this part is very much stricter. Many students simply said "It is not necessary, since n is large, **it** is approximately normal by CLT". These are important concepts that need to be corrected so students can have a better picture of how CLT is used.
(b)(i)	Critical region of the test is $z<-1.64485$ or $z>1.64485$ $\Rightarrow$ $\underline{z<-1.64 \text{ or } z>1.64}$ (3 s.f.)	The phrases "critical value" and "critical region" are added into the new syllabus, so students must know and distinguish between them. A number of students gave just the critical values. Also, critical region is usually expressed in terms of the test statistic (in our case, z). Finally, there are also students who gave the non-critical region as the critical region. One way to rectify this is to reinforce the fact that the critical region is also known as the rejection region (i.e. rejection of H_0).
(b)(ii)	Value of test statistic $z=\dfrac{5.94-5.5}{\sqrt{\dfrac{6.5}{n}}}=\dfrac{0.44\sqrt{n}}{\sqrt{6.5}}$ For a favourable outcome at 10% significance level, do not reject H_0 $\Rightarrow$ z lies <u>outside</u> the critical region $\Rightarrow -1.64485<\dfrac{0.44\sqrt{n}}{\sqrt{6.5}}<1.64485$ $\Rightarrow \dfrac{-1.64485\sqrt{6.5}}{0.44}<\sqrt{n}<\dfrac{1.64485\sqrt{6.5}}{0.44}$ $\Rightarrow n<\left(\dfrac{1.64485\sqrt{6.5}}{0.44}\right)^2$ $\Rightarrow n<90.837$ Hence largest $n=\underline{90}$	Students who are careless with reading the questions would have used either $\frac{178.2}{n}$ as the sample mean or 6.2108 as s^2. Some students were confused about what the "favourable outcome" meant about the rejection of H_0. This involves understanding the context of the problem. A significant portion of students only wrote down $z<1.64485$ and not the full non-critical region. Credit was only given if the correct inequality with the p-value was given earlier; the assumption is that with the correct inequality, students would be able to use invNorm to find the correct critical value. Otherwise, the

full region should be written down. It is actually also possible to obtain the correct answer with $z > -1.64485$, but the earlier inequality would have been more appropriate since the test statistic here is positive.

Students were also generally very careless with solving inequalities.

Q2

(i) Every dustbin has **an equal probability of being selected** and the selections of each dustbin are **made independently**.

(ii) Since $n = 50$ is large, by Central Limit Theorem, the **mean** mass of rubbish in dustbins will be approximately normally distributed.

(iii) Unbiased estimate of population mean, $\bar{x} = \frac{924.5}{50} = 18.49$

Unbiased estimate of population variance, $s^2 = \frac{1}{49}\left[18249.2 - \frac{924.5^2}{50}\right] = 23.575$ (5 s.f.)

$= 23.6$ (3 s.f.)

Let μ be the population mean mass of rubbish, in kg, in a domestic dustbin.

To test: H_0: $\mu = 20$

against H_1: $\mu < 20$

at 1% level of significance

Since $n = 50$ is large, by Central Limit Theorem,

$\bar{X} \sim N\left(20, \frac{23.575}{50}\right)$ approximately under H_0.

Test Statistic: $Z = \frac{\bar{X} - 20}{\sqrt{23.575/50}} \sim N(0,1)$ approximately under H_0.

Using GC, $[\bar{x} = 18.49, s^2 = 23.575, n = 50]$

$z_{test} = -2.199$, p-value $= 0.013937$ (5 s.f.)

Since p-value $= 0.013937 > 0.01$, we do **not** reject H_0 and conclude that there is **insufficient** evidence at 1% level of significance to claim that there has been a reduction in the mass of rubbish in dustbins.

(iv) For H_0 to be rejected, $z_{test} = \frac{18.49 - 20}{\sqrt{23.575}} \times \sqrt{n} < -2.3263$

$n > 55.954$

Range of values of n is $n \geq 56,\ n \in \mathbb{Z}^+$

[Also Accept: $n > 55,\ n \in \mathbb{Z}$ (or equivalent form)]

Q3

(i)

Let μ denote the population mean amount of energy released in the collisions.

Test H₀: $\mu = 1860$

Against H₁: $\mu > 1860$

Using a one-tail test at 1% significance level.

Under H_0, $\overline{X} \sim N\left(1860, \frac{40^2}{n}\right)$ approx

Test statistic: $Z = \frac{\overline{X} - 1860}{40/\sqrt{n}} \sim N(0,1)$

$$z_{calc} = \frac{1864 - 1860}{40/\sqrt{n}} = \frac{\sqrt{n}}{10}$$

To reject H_0 at 1% level of significance, the critical region is:

$$z_{calc} > 2.32635$$

Hence,

$$\frac{\sqrt{n}}{10} > 2.32635$$

$$n > 541.189$$

Thus, the least value of n is 542.

(ii)

Test H_0: $\mu = 1860$

Against H_1: $\mu > 1860$

Using a one-tailed test at 1% significance level.

Under H_0, $\overline{X} \sim N\left(1860, \frac{40^2}{600}\right)$ approx

Test statistic: $Z = \frac{\overline{X} - 1860}{40/\sqrt{600}} \sim N(0,1)$

From GC,

$$p\text{-value} = 0.00715$$

The p-value means that the lowest level of significance at which we would reject the hypothesis that the mean amount of energy released is 1860 MeV in favour of the hypothesis that the amount is greater than 1860 MeV is 0.715 %.

(iii)

No assumption needed. This is because the sample size of 600 is large and thus by Central Limit Theorem, $\overline{X}$ follows a normal distribution.

(iv)
Let $Z \sim N(0, 1)$
$P(Z \geq 2) = 0.0228$

Hence, lowest level of significance for which the experiment meets the "two sigma" threshold is 2.28%.

Since p-value = 0.00715 < 0.0228, the result meets the "two sigma" threshold.

Alternative:

Under H_0, $\overline{X} \sim N\left(1860, \frac{40^2}{600}\right)$ approx

$$2\sigma = 2\sqrt{\frac{1600}{600}} = 3.265986$$

So $\mu + 2\sigma = 1863.265986$

$$P\left(\overline{X} > 1863.265986\right) \approx 0.0228$$

Hence, lowest level of significance for which the experiment meets the "two sigma" threshold is 2.28%.

Since $\bar{x} = 1864 > 1863.265986$ the test meets the "two sigma" threshold.