1. **[ACJC Prelims 17]**
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 users 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 10% significance level. [6]
      ii. Explain if it is necessary for the manufacturer to assume 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 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]

2. **[AJC Prelims 17]**
   (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 Boards 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_0 hours, find the range of values of μ_0. [4]

3. **[CJC Prelims 17]**
A computer hard drive manufacturer claims that the mean usage hours before failure of their R series hard drives is 50 thousand hours. A technology columnist wishes to investigate this claim and collected the usage hours, t thousand hours for each of the 50 randomly chosen hard drives which were submitted to the local service centre for drive failures. The data is summarized as follows:

$$n = 50, \quad \sum t = 2384.5, \quad \sum t^2 = 115885.23.$$

The technology columnist wants to use hypothesis testing to test whether the mean usage hours before failure of a hard drive is different from what the manufacturer has stated.

(a) Explain whether it is necessary for the columnist to know about the distribution of the usage hours before failure of the drives in order to carry out a hypothesis test. [1]

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

The columnist published the data and the results of the hypothesis testing in an online article.

(c) Suggest a reason why the test result might not be useful to a reader of the article who is deciding whether to buy an R series hard drive from the manufacturer. [1]

(d) State an alternative hypothesis that is more relevant to the decision making process and explain whether the result will differ from the earlier test carried out by the columnist at 1% level of significance. [2]

(e) State a necessary assumption that was made for all the tests carried out. [1]

4. **[DHS Prelims 17]**
The company Snatch provides a ride-hailing service comprising taxis and private cars in Singapore. Snatch claims that the mean waiting time for a passenger from the booking time to the time of the vehicles arrival is 7 minutes.

To test whether the claim is true, a random sample of 30 passengers waiting times is obtained. The standard deviation of the sample is 2 minutes. A hypothesis test conducted concludes that there is sufficient evidence at the 1% significance level to reject the claim.

(a) State appropriate hypotheses and the distribution of the test statistic used. [3]

(b) Find the range of values of the sample mean waiting time, $\bar{t}$. [3]

(c) A hypothesis test is conducted at the 1% significance level whether the mean waiting time of passengers is more than 7 minutes. Using the existing sample, deduce the conclusion of this test if the sample mean waiting time is more than 7 minutes. [2]

5. **[HCI Prelims 17]**
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 te 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]

6. **[IJC Prelims 17]**
The mass of strawberry jam in a randomly chosen jar follows a normal distribution and has a mean mass of 200 grams. A retailer suspects that the mean mass of the strawberry jam is being overstated. He takes a random sample of 30 jars of strawberry jam and weighs the content, x grams, in each jar. The results are summarized as follows:

$$\sum(x-200) = -66 \quad \text{and} \quad \sum(x-200)^2 = 958.$$

(a) Test at 2% significance level, whether the retailer's suspicion is justifiable. [6]

(b) Explain, in this context, the meaning of 'at 2% significance level'. [1]

(c) Suppose the retailer now decides to test whether the mean mass differs from 200 grams at 2% significance level. Without carrying out the test, explain whether the conclusion would change in part (a). [1]

The manufacturing process has now been improved and the population standard deviation is 3.5 grams. The retailer selects a new random sample of 20 jars of strawberry jam and the sample mean is found to be k grams. Find the range of possible values of k so that the retailers suspicion that the mean mass differs from 200 grams is not justified at the 2% significance level. Give your answer correct to one decimal place. [4]

7. **[TPJC Prelims 17]**
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.

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

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

(b) 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]

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

(d) Use your results in part (c) 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]

8. **[TJC Prelims 17]**

(a) College students intending to further their studies overseas have to sit for a mandatory Overseas Universities Test (OUT). Researcher Mr Anand wishes to find out if male college students tend to score higher for OUT compared to female college students. Mr Anands colleague randomly selects 150 male and 150 female students from the combined student population of three particular colleges near his home to form a sample of 300 college students for the research. Explain whether this sample is a random sample.

(b) The mean OUT score for all college students in 2016 is 66. Mr Anand randomly selects 240 college students taking OUT in 2017 and their scores, x, are summarised in the following table:

Score, x	60	65	68	70	75	80
Frequency, f	40	90	63	27	18	2

i. Write down the unbiased estimates of the population mean and variance of the OUT scores for the college students in 2017. [1]

ii. Test, at the 10% level of significance, whether the mean OUT score for all college students in 2017 is higher than the mean score attained in 2016. [4]

iii. Explain what is meant by the phrase "10% level of significance" in this context. [1]

iv. Mr Anand draws a new sample of 240 male college students. Using the unbiased estimate for the population variance computed in (i), find the range of values for the sample mean $\overline{x}$ that is required for this new sample to achieve a different conclusion from that in (ii). [4]

(c) The 2017 OUT scores of the male and female college students are independent and assumed to be normally distributed with means and standard deviations as shown in the table on the next page:

	Mean	Standard deviation
Male college students	64	5.5
Female college students	66	3.5

Mr Beng and Miss Charlene both scored 70. Explain who performed better relative to their respective gender cohort. [2]

9. **[YJC Prelims 17]**
In a factory, the average time taken by a machine to assemble a smartphone is 53 minutes. A new assembly process is trialled and the time taken to assemble a smartphone, x minutes, is recorded for a random sample of 60 smartphones. The total time taken was found to be 3129 minutes and the variance of the time was 18.35 minutes^2.

The engineer wants to test whether the average time taken by a machine to assemble a smartphone has decreased, by carrying out a hypothesis test.

(a) Explain why the engineer is able to carry out a hypothesis test without assuming anything about the distribution of the times taken to assemble a smartphone. [1]

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

(c) Explain, in the context of the question, the meaning of at 10% level of significance. [1]

After several trials, the engineer claims that the average time taken by a machine to assemblea smartphone is 45 minutes using the new assembly process. The internal control manager wishes to test whether the engineers claim is valid. The population variance of the time taken to assemble a smartphone using the new assembly process may be assumed to be 9 minutes^2. A random sample of 50 smartphones is taken.

(d) Find the range of values of the mean time of this sample for which the engineers claim would be rejected at the 10% significance level. [4]

Answers

1. (a) $5.94, 6.21$.
   $p = 0.334$. Do not reject H_0.
   (b) i. $z < -1.64$ or $z > 1.64$.
   ii. 90.

2. (a) Do not reject H_0.
   (b) $5.55 < \mu_0 < 6.73$.

3. (b) 47.69, 44.3.
   (d) Yes.

4. (a) $H_0 : \mu = 7, H_1 : \mu \neq 7$.
   Under H_0, $\overline{T} \equiv N(7, \frac{4}{29})$ approxmately by CLT since n is large.
   (b) $\bar{t} \leq 6.04$ or $\bar{t} \geq 7.96$.

5. (a) i. $\overline{x} = 124.4, s^2 = 7.43$.
   ii. p-value $= 0.0598$. Reject H_0.
   No assumption about masses of boxes of blueberries are needed since $n = 50$ is sufficient large for Central Limit Theorem to apply so the mean mass of boxes of raspberries will follow a normal distribution approximately.
   (b) 35.

6. (a) p-value $= 0.01142$. Reject H_0.
   (b) There is a probability of 0.02 that the hypothesis test will indicate that the mean mass of strawberry jam in the jar is less than 200g, when in fact it is 200g.
   (c) This will result in a different conclusion.
   $198.2 < k < 201.8$.

7. (a) Every dustbin has an equal probability of being selected and the selections are made independently.
   (b) Since $n = 50$ is large, by the Central Limit THeorem, the man mass of rubbish will be approximately normally distributed.
   (c) $\overline{x} = 18.49, s^2 = 23.6$.
   $p = 0.013937$. Do not reject H_0.
   (d) $n \geq 56, n \in \mathbb{Z}$.

8. (a) Non-random since students from other colleges do not have any chance of being selected.
   (b) i. $\overline{x} = 66.391, s^2 = 4.1048^2$
   $p = 0.0697$. Reject H_0.
   ii. There is a probability of 0.1 of wrongly concluding that the mean OUT score of all college students in 2017 is higher than the mean score attained in 2016.

iii. $\overline{x} > 66.3396$.

(c) $P(M \leq 70) = 0.862, P(F \leq 70) = 0.873$.
Hence Miss Charlene performed better relative to her gender.

9. (a) Since sample size is large, by Central Limit Theorem, the sample mean time for 60 smartphones is approximately normal. Hence the assumption that the time taken by a machine to assembly a smartphone is not necessary.

(b) $\overline{x} = 52.15$.
$s^2 = 18.7$.
p-value=0.0637.
Since p-value < 0.1, we reject H_0 and conclude at 10% level that there is sufficient evidence that average time taken by a machine to assembly a smartphone has reduced.

(c) There is a probability of 0.1 of concluding that the average time taken by a machine to assembly a smartphone has decreased when the average time taken by a machine to assembly a smartphone is 53 minutes.

(d) $\overline{x} < 44.3$ or $\overline{x} > 45.7$.