

NATIONAL UNIVERSITY OF SINGAPORE
ST1131 – INTRODUCTION TO STATISTICS

Semester I : 2013–2014

Names of examiners : Associate Professor Gan Fah Fatt/Ms Wong Yean Ling

November / December 2013

Time allowed : 2 hours

INSTRUCTIONS TO CANDIDATES

1. Please write your matriculation/registration number and seat number only.
Do not write your name.
2. This examination paper contains a total of **FIVE (5)** questions and comprises **TWELVE (12)** printed pages.
3. Answer **ALL** questions. The total number of marks for this paper is 80.
4. Candidates may use calculators. However, they should lay out systematically the various steps in the calculations.
5. Statistical tables will be provided.
6. This is a closed book examination. Two A4-size help sheets are allowed.
7. Write your answers in this booklet.

Question	Marks	Maximum Marks
1		16
2		16
3		16
4		16
5		16

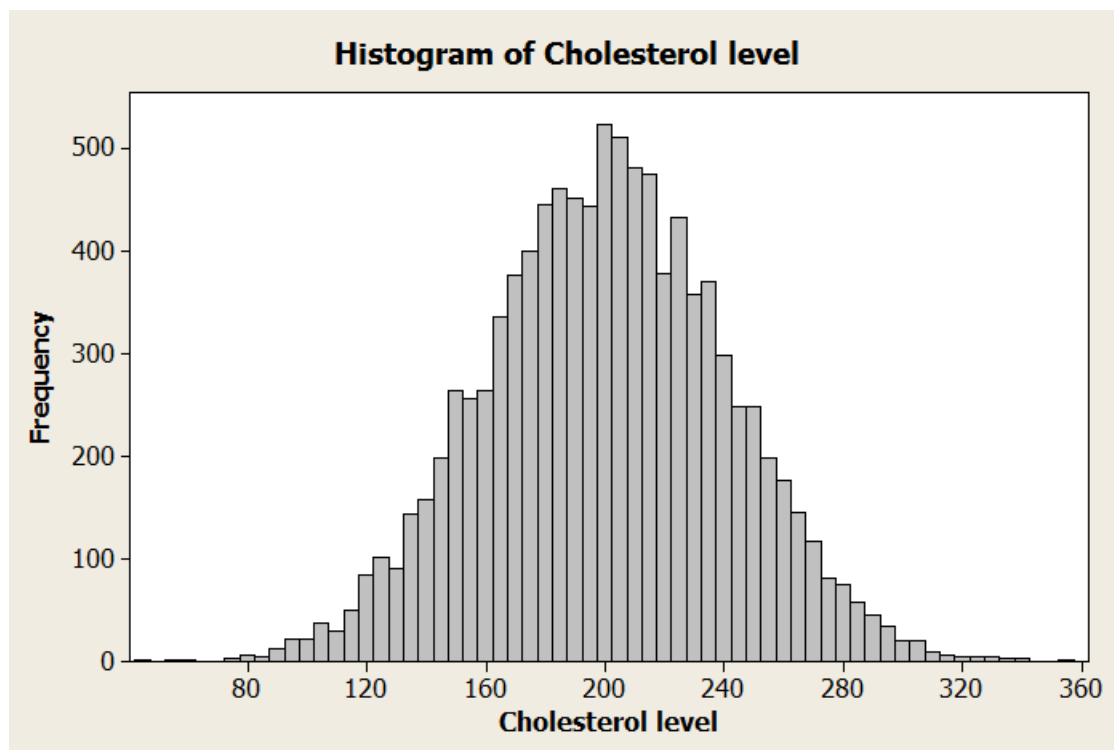
Total
80

Matriculation No. _____

Seat No. _____

Question 1

A random sample of adults was taken from a population and a histogram of their blood cholesterol levels (in mg/dL) is shown below.



- (i) Write a concise description (in one short sentence) of the distribution of cholesterol levels. [2 marks]

- (ii) Give an estimate of the sample median cholesterol level. [2 marks]

- (iii) John claims that “The sample mean cholesterol level is approximately the same as the sample median cholesterol level.” Is his claim correct? Give a reason (in one short sentence) to support your answer. [2 marks]

- (iv) John claims that “If I take 1000 random samples of 5 adults from this population and calculate the sample mean cholesterol level for each sample of 5 adults, then approximately 500 of the sample means will be less than 240 mg/dL.” Is his claim correct? Give a reason (in one short sentence) to support your answer. [2 marks]

- (v) John claims that “If I continue to sample randomly more and more adults from this population, then the sample mean cholesterol level will become less and less variable.” Is his claim correct? Give a reason (in one short sentence) to support your answer. [2 marks]

- (vi) John claims that “If I continue to sample randomly more and more adults from this population, then the sample mean cholesterol level will converge to 240 mg/dL approximately.” Is his claim correct? Give a reason (in one short sentence) to support your answer. [2 marks]

- (vii) Based on a random sample of 5 adults from this population, estimate the probability that the sample mean cholesterol level is more than 220 mg/dL. [4 marks]

Question 2

According to a study, 4% of a population in a region have lung disease. Of those people without lung disease, 13% are smokers. Of those people with lung disease, 92% are smokers.

- (i) If a person is randomly selected from the region, what is the probability that the person has lung disease and is a smoker? [2 marks]

- (ii) If a person is randomly selected from the region, what is the probability that the person is a smoker? [2 marks]

- (iii) If a person is randomly selected from the region, what is the probability that the person is a smoker or has lung disease? [2 marks]

- (iv) If two people are randomly selected from the region, what is the probability that both of them have lung disease? [2 marks]

- (v) If 20 people are randomly selected from the region, what is the probability that at least one of them has lung disease? [2 marks]

- (vi) If 1000 people are randomly selected from the region, what is an approximate probability that the sample proportion of people having lung disease is less than 0.05? [3 marks]

- (vii) Give the main reason (in one short sentence) that supports the validity of the answer obtained in part (vi). [3 marks]

Question 3

The tourism board of Sunshine City reveals that the mean and the standard deviation of the cost per person for a 3-day tour in the city, during the non-peak season are \$380 and \$220 respectively.

- (i) Sketch a reasonable density curve for the cost per person for a 3-day tour in the city during the non-peak season and describe the distribution (in not more than 2 short sentences). [3+2 marks]



- (ii) A random sample of 50 tourists was chosen from the departure hall during the non-peak season. Can the probability that they spent at least \$400 on average be estimated accurately? If yes, compute the probability, otherwise give an explanation (in not more than 2 short sentences). [3 marks]



- (iii) In order to know more about the cost of such a tour during the peak season, a random sample of 4 tourists from the departure hall during the peak season was chosen. The sample mean and sample standard deviation of the cost per person for a 3-day tour were found to be \$500 and \$180 respectively. Construct a 90% confidence interval for the mean cost of such a tour during the peak season. [2 marks]

- (iv) What are the necessary cautions when using the confidence interval in part (iii)? [4 marks]

- (v) How large should the sample size be so that the margin of error of the confidence interval in part (iii) is approximately \$20? [2 marks]

Question 4

A researcher conducted an experiment to study the effectiveness of two treatments, A and B in lowering the blood pressure (in mmHg) in human. The researcher used Minitab to produce the following outputs.

Two-Sample T-Test and CI: A, B

Two-sample T for A vs B

	N	Mean	StDev	SE Mean
A	5	147.4	14.3	6.4
B	5	129.80	7.16	3.2

Difference = μ (A) - μ (B)

Estimate for difference: 17.60

95% CI for difference: (-0.74, 35.94)

T-Test of difference = 0 (vs not =): T-Value = 2.47 P-Value = 0.057 DF = 5

Two-Sample T-Test and CI: A, B

Two-sample T for A vs B

	N	Mean	StDev	SE Mean
A	5	147.4	14.3	6.4
B	5	129.80	7.16	3.2

Difference = μ (A) - μ (B)

Estimate for difference: 17.60

95% CI for difference: (1.15, 34.05)

T-Test of difference = 0 (vs not =): T-Value = 2.47 P-Value = 0.039 DF = 8

Both use Pooled StDev = 11.2805

Paired T-Test and CI: A, B

Paired T for A - B

	N	Mean	StDev	SE Mean
A	5	147.40	14.26	6.38
B	5	129.80	7.16	3.20
Difference	5	17.60	11.55	5.16

95% CI for mean difference: (3.26, 31.94)

T-Test of mean difference = 0 (vs not = 0): T-Value = 3.41 P-Value = 0.027

- (i) State appropriate hypotheses for this experiment with the parameters clearly defined. [2 marks]

- (ii) If matched pairs were used in the experiment, how many patients participated in the experiment? [2 marks]

- (iii) If matched pairs were used in the experiment and the population standard deviation of difference is assumed to be 10 mmHg, calculate the probability of type II error of the test if the true difference is 20 mmHg, for a 0.05 significance level test. [3 marks]

- (iv) If the experiment was conducted by randomly assigning the treatments to the patients, how many patients participated in the experiment? [2 marks]

- (v) If the experiment was conducted by randomly assigning the treatments to the patients, state a 95% confidence interval for the mean difference in blood pressure.

[2 marks]

- (vi) Interpret (in not more than 2 short sentences) the confidence interval in part (v).

[3 marks]

- (vii) What assumptions do you need for the confidence interval in part (v) to be valid?

[2 marks]

Question 5

In a recent survey of commuters, it was found that 94% of them said that they would give up their seat to those who need it more. Assuming that the true percentage of commuters who would give up the seat to those who need it more is indeed 94%, you are tasked to find out if the percentage of NUS student commuters who would do so is different from that of the general population of commuters.

- (i) You sampled 10 NUS students and found that there were 7 students who said they would give up their seat to those who need it more. Compute the p -value to test the hypotheses of interest. [3 marks]

- (ii) State all the assumptions for the p -value in part (i) to be valid.

[2 marks]

- (iii) You managed to sample 200 NUS student commuters instead. How many students do you expect who would give up their seat to those who need it more if the percentage of NUS student commuters who would do so is the same as that of the general population of commuters? [2 marks]

- (iv) Out of the 200 students sampled, 168 students said they would give up their seat to those who need it more. Compute the standard error of the sample proportion.

[3 marks]

- (v) Construct an approximate 90% confidence interval for the true proportion of NUS student commuters who would give up their seat to those who need it more, based on the data in part (iv).

[3 marks]

- (vi) Based on the confidence interval in part (v), is there any evidence that the true proportion is different from that of the general population of commuters? Give a reason (in one short sentence) to support your answer.

[3 marks]