

Main Examination period 2018

MTH4106/MTH4206: Introduction to Statistics

Duration: 2 hours

Student number

--	--	--	--	--	--	--	--	--	--

Desk number

--	--	--

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

Write your solutions in the spaces provided in this exam paper. If you need more paper, ask an invigilator for an additional booklet and attach it to this paper at the end of the exam.

You should attempt ALL questions. Marks available are shown next to the questions.
--

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.

The New Cambridge Statistical Tables are provided.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it shall be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: H. Maruri-Aguilar, D. S. Coad

This page is for marking purposes only.
Do not write on it.

Question	Mark	Comments
1	/ 20	
2	/ 17	
3	/ 14	
4	/ 17	
5	/ 14	
6	/ 18	
Total		

Question 1. [20 marks]

Yield data (in bushels) after harvest were collected from several one-acre plots on two farms termed 'Farm 1' and 'Farm 2'. The data are given below.

Farm 1: 254 238 235 239 235 241 235 243 255 247

Farm 2: 253 251 246 240 242 262 260 240 242 242

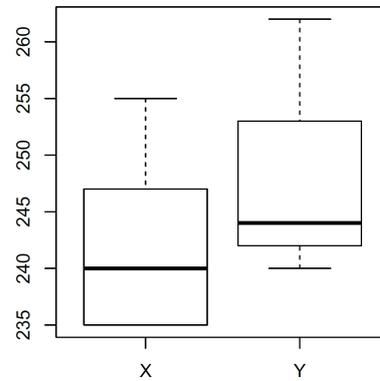
(a) Compute the five number summary for each of the farms.

[8]

(b) Compute the sample mean and sample variance for each farm.

[4]

- (c) The following boxplots were produced with these data, but it is not known which farm corresponds to each plot. Write in the space provided which of Farms 1 and 2 correspond to labels X and Y, and briefly say why. [2]



Boxplot	Farm and brief explanation
X	----- -----
Y	----- -----

- (d) Describe the distribution of yields for each farm, using the standard hierarchy of description: location, spread and tails of the distributions. Also briefly compare the two farms. [6]

Question 2. [17 marks]

- (a) Define the **probability generating function** of a discrete random variable X . [2]

- (b) Determine the probability generating function of $X \sim \text{Ber}(1/3)$. [4]

- (c) Let the random variable Y be independent to X and with the same distribution as X . Compute the probability generating function of the random variable $Z = X + Y$. [5]

- (d) Using your previous result or otherwise, compute $E(Z)$. [6]

Question 3. [14 marks]

(a) Let X_1, \dots, X_n be independent $\text{Exp}(\theta)$ random variables with $\theta = 3$. Let $T = \sum_{i=1}^n X_i$.

(i) State $E(X_i)$ and $\text{Var}(X_i)$.

[4]

(ii) Compute $E(T)$.

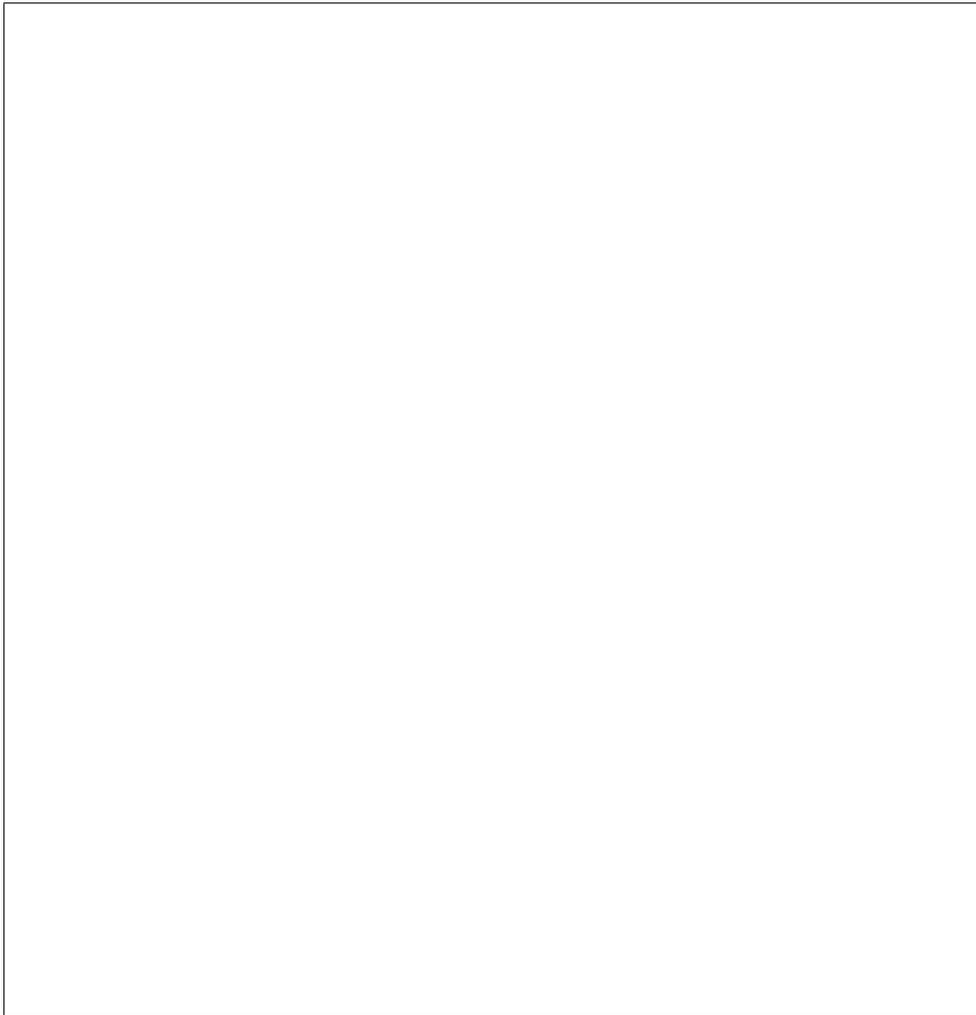
[3]

(iii) Compute $\text{Var}(T)$.

[3]

(b) State the central limit theorem.

[4]

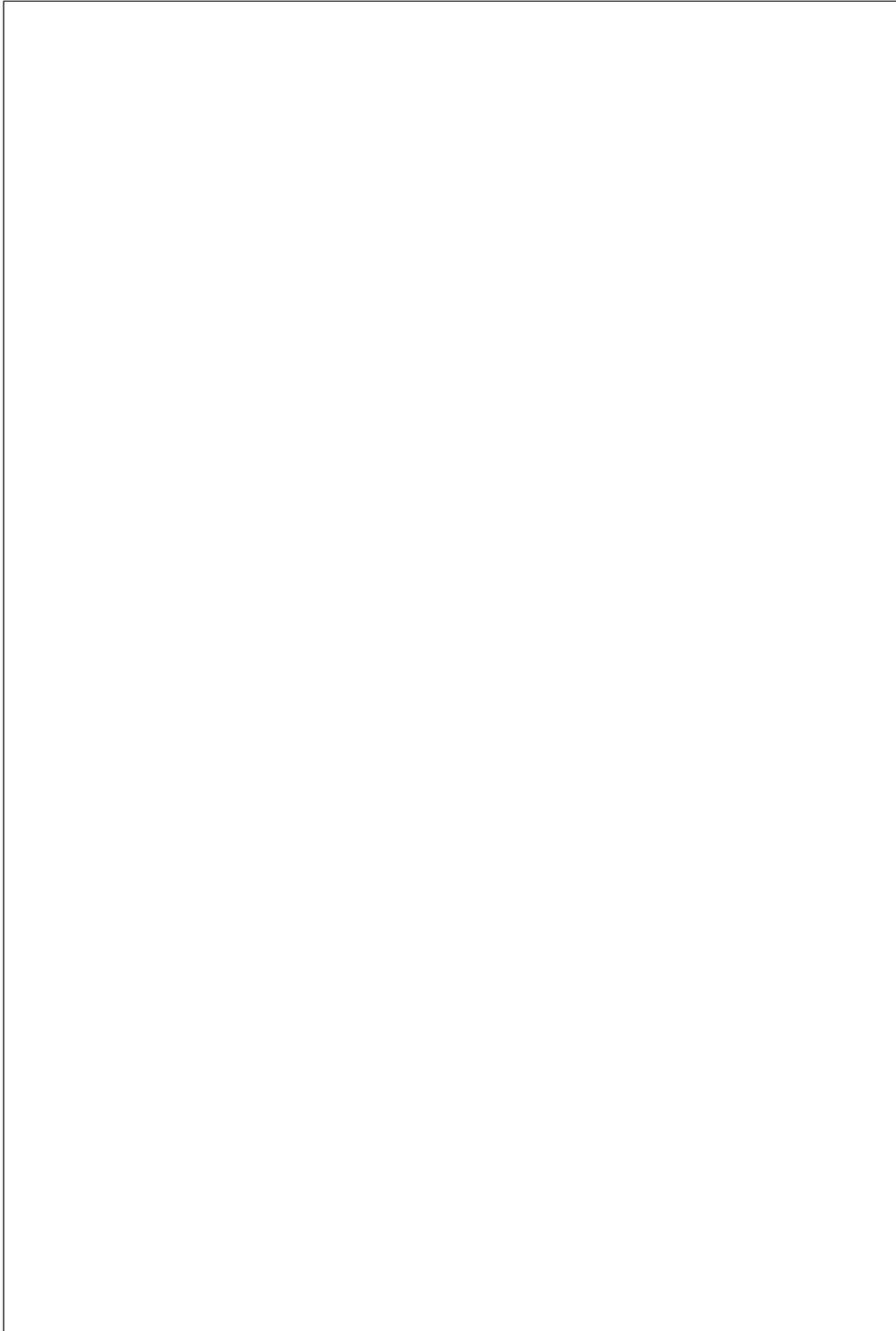


Question 4. [17 marks]

- (a) Suppose that X , the number of users in a busy library in a given hour, follows a Poisson distribution with mean 35. Using a normal approximation, compute $\Pr(X > 40)$. [8]

- (b) Miyota engines drive most quartz wristwatches. Suppose that 80% of Miyota engines stay within the standard accuracy of ± 15 seconds per month. Let W be the number of engines within the standard accuracy range from a batch of 200. Compute the normal approximation to $\Pr(W > 150)$.

[9]



Question 5. [14 marks]

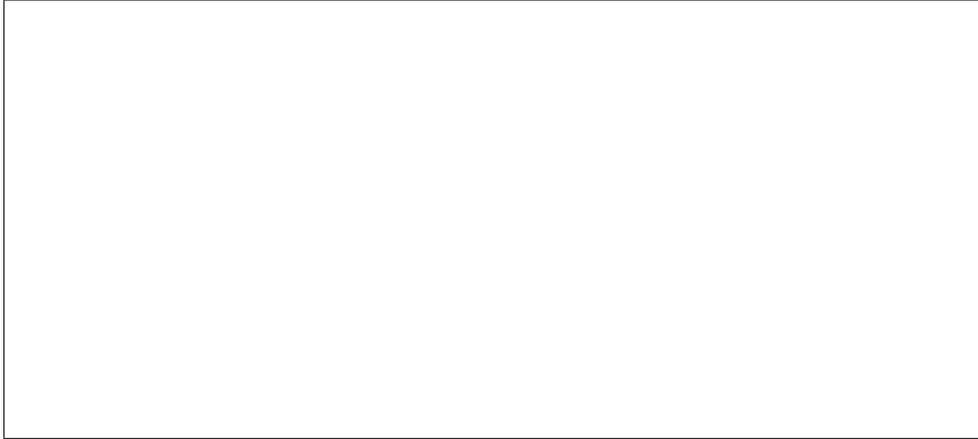
- (a) Let X be a continuous r.v., and let $Y = g(X)$. Write down the theorem that gives the p.d.f. of Y in terms of the p.d.f. of X and the function g . Include all the conditions required.

[6]

- (b) Let $X \sim \text{Exp}(1)$. Determine the cumulative distribution function of $Y = \log(X)$.

[4]

- (c) Show that the lower quartile of the random variable Y above is $Q_1 \approx -1.2459$. [4]



Question 6. [18 marks]

- (a) Explain what are meant by type I and type II errors in a hypothesis test.

[4]

- (b) A freight company is studying its current pricing policy for deliveries. Customers pay the current price (calculated from the current pricing policy), but the true cost of delivery varies from this price. For a random sample of deliveries, the company analyst computed the difference between what was paid and the true cost of the delivery. If the difference is negative, there was loss for the company and if the difference is positive, the customer paid more for the service. Below are the data consisting of 10 differences, rounded to the nearest one hundred pounds

Differences: -3 6 -5 0 7 -2 -2 -2 -1 0

- (i) Suppose that the analyst is interested in determining whether the price differs from the true cost. Write down the hypotheses H_0 and H_1 .

[4]

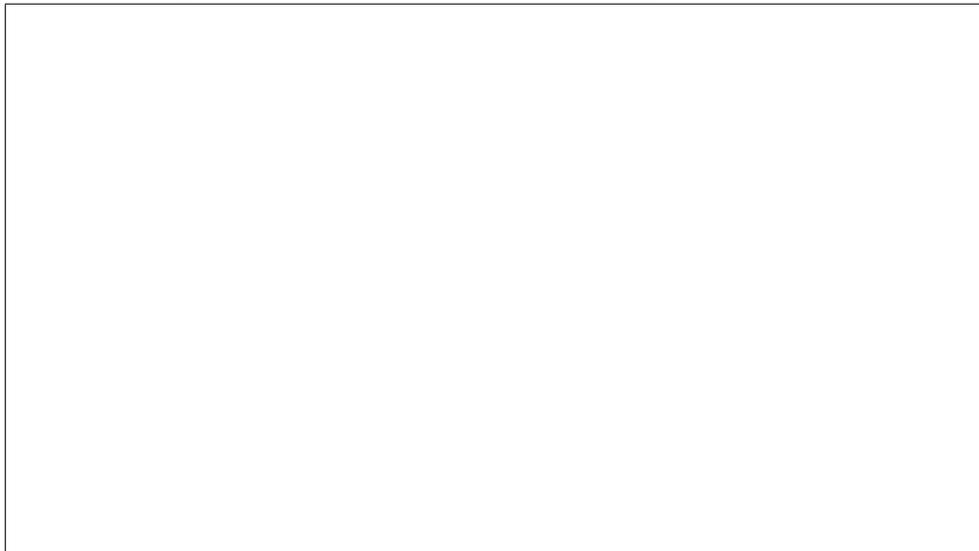
- (ii) If the standard deviation is believed to be $\sigma = 4$, carry out the hypothesis test at the 5% significance level.

[6]



- (iii) Compute the p-value for the above test and state your conclusions.

[4]



End of Paper.