Queen Mary
University of London

Main Examination period 2017

MTH4106/MTH4206: Introduction to Statistics

Duration: 2 hours

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

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

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: D. S. Coad, D. S. Stark

 Turn Over

Question 1. [20 marks]
A cherry farmer in Suttons Bay, Michigan, transports tart cherries in large tanks. The weights, in kg, of cherries in 15 tanks are as follows:

464	431	427	451	430
514	468	419	483	439
455	427	477	447	435

(a) Draw a stem-and-leaf plot of the data. [**3**]

(b) Find the five-number summary for these data. [**5**]

(c) Comment on the distribution of the data in the light of your answers to (a) and (b). [**2**]

(d) For these data, the sample mean is 451.13kg and the sample variance is 683.981kg^2. For a weight of x measured in kg, the corresponding weight measured in lb is $2.205x$. Find the sample mean and sample variance of the weights in lb, justifying your answers. [**5**]

(e) Explain the difference between an observational study and an experiment. Which of the two is described above? [**5**]

Question 2. [15 marks]
Let X and Y be random variables both taking values on the non-negative integers.

(a) Define the **probability generating function** of X. [**2**]

(b) State a theorem linking the probability generating functions of X, Y and $X + Y$ under a suitable condition. [**3**]

(c) Let $X \sim \mathrm{Geom}(p)$. Derive the probability generating function of X. [**5**]

(d) Suppose that $X \sim \mathrm{Geom}(p)$ and $Y \sim \mathrm{Geom}(p)$ are independent. Explain how you would find the distribution of $X + Y$. [**5**]

Question 3. [10 marks]
Let $X_1, \ldots, X_n$ be mutually independent random variables which all have the same distribution with finite mean μ and finite variance σ^2. Let T be the sample total.

(a) Find $\mathbb{E}(T)$. [**3**]

(b) Find $\mathrm{Var}(T)$. [**3**]

(c) State the Central Limit Theorem. [**4**]

Question 4. [15 marks]

(a) Suppose that a baseball team loses each away game independently with probability $p = 0.6$. Let X be the number of away games lost in a season in which $n = 81$ away games are played. Find an approximation to $\mathbb{P}(42 \leq X \leq 50)$. [**8**]

(b) Let Y denote the number of new customers per hour at Steve's Coffee Shop. Assume that the distribution of Y is Poisson with mean 15. Determine approximately $\mathbb{P}(Y > 11)$. [**7**]

Question 5. [15 marks]

(a) Let $X \sim \text{Exp}(4)$ and put $Y = 3X - 5$. Find the probability density function of Y. Without using integration, find the expectation of Y. [**7**]

(b) Let $U \sim U(2, 9)$ and put $V = \sqrt{U}$. Find the probability density function of V. Hence, find the expectation of V. [**8**]

Question 6. [10 marks]
Let X, Y and Z be jointly distributed random variables.

(a) Define the **covariance** $\text{Cov}(X, Y)$. [**2**]

(b) Prove that $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$. [**3**]

(c) Suppose that Y and Z are independent, with $\text{Var}(Y) = 7$ and $\text{Var}(Z) = 9$, and that $X = 3Y - 5Z + 2$. Find the correlation between X and Z. [**5**]

Question 7. [15 marks]
Students are weighed at the beginning and at the end of a term-long health-fitness programme. A researcher wants to know if the health-fitness programme is effective at reducing the average weight of students. Assume that the weight change for a student, postweight minus preweight, is normally distributed with a standard deviation of 1.5kg. The weight changes, in kg, of a random sample of 12 students were as follows:

2.0, -0.5, 1.4, -2.2, 0.3, -0.8, 3.7, -0.1, 0.6, 0.2, 0.9, -0.1

(a) Explain what are meant by type I and type II errors in a hypothesis test. [**4**]

(b) Write down the researcher's null and alternative hypotheses. [**2**]

(c) Find the sample mean of the data. [**2**]

(d) Carry out the appropriate hypothesis test at the 5% significance level and report the conclusion. [**7**]

End of Paper.