MTH4106: Introduction to Statistics

Duration: 2 hours

Date and time: 6 May 2016, 2.30-4.30

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 awarded are shown next to the questions.

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 workings 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.

Examiner(s): D. S. Coad and L. I. Pettit

Question 1 (20 marks). A biologist has measured the lengths of 12 female green lynx spiders. The lengths, in mm, are as follows:

8.3, 10.0, 5.9, 7.1, 8.5, 7.6, 9.8, 10.8, 6.6, 7.6, 8.1, 9.1

(a) What sort of variable is this? [**2**]

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

(c) Determine whether there are any outliers in the data. [**3**]

(d) Draw the corresponding boxplot and comment briefly on the distribution of the data. [**5**]

(e) Calculate the sample mean and the sample standard deviation of the data. [**5**]

Question 2 (15 marks). Let X be a discrete random variable all of whose values are non-negative integers. Let $G(t)$ be the probability generating function of X.

(a) Prove that

$$\left.\frac{\mathrm{d}G(t)}{\mathrm{d}t}\right|_{t=1} = \mathbb{E}(X)$$

and

$$\left.\frac{\mathrm{d}^2G(t)}{\mathrm{d}t^2}\right|_{t=1} = \mathbb{E}(X^2) - \mathbb{E}(X).$$

[**8**]

(b) It is known that the probability generating function of $X \sim \mathrm{Bin}(n, p)$ is

$$G(t) = (q + pt)^n,$$

where $q = 1 - p$. Find $\mathbb{E}(X)$ and $\mathrm{Var}(X)$. [**7**]

Question 3 (15 marks). Let $X_1, \ldots, X_n$ be independent random variables for which $\mathbb{E}(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$ for $i = 1, 2, \ldots, n$. Let

$$Y = \sum_{i=1}^{n} a_i X_i,$$

where $a_1, \ldots, a_n$ are any real numbers.

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

(ii) Find $\mathrm{Var}(Y)$. [**3**]

(iii) What extra assumption would guarantee that Y is normally distributed? State any relevant result. [**2**]

(b) The Colosseum in Rome can be divided into several thousand small sections. Suppose that the earthquake force that a section can withstand has mean and standard deviation, respectively, of 3.4 and 1.5 on the Richter scale. A random sample of 100 sections is taken. Find an approximation to the probability that the mean earthquake force that these can withstand is greater than 3.6 on the Richter scale. [**7**]

Question 4 (15 marks). (a) Let $X \sim \text{Exp}(6)$ and put $Y = \sqrt[3]{X} - 1$. Find the probability density function of Y. [**7**]

(b) Let $U \sim U(5, 15)$ and put $V = 2U - 5$. Find the probability density function of V. State the expectation and variance of V. [**8**]

Question 5 (10 marks). Let X, Y and Z be jointly distributed random variables.

(a) Define the **correlation** $\text{corr}(X, Y)$. [**3**]

(b) Suppose that $Y = aX + b$, where a and b are constants and $a \neq 0$. State the values of $\text{corr}(X, Y)$ in this case. [**2**]

(c) Suppose that Y and Z are independent, with $\text{Var}(Y) = 3$ and $\text{Var}(Z) = 8$, and that $X = Y + 4Z$. Find the correlation between X and Y. [**5**]

Question 6 (15 marks). The average Big Mac contains 20g of fat. In a study reported in 1993, a random sample of 200 portions of moo shoo pork was analysed, giving a sample mean of 64g of fat and a sample standard deviation of 18g. A food scientist wants to know if the average fat content of a portion of moo shoo pork is higher than that of three average Big Macs. Assume that the fat content of a portion of moo shoo pork is normally distributed.

(a) State the null and alternative hypotheses. [**2**]

(b) Carry out the appropriate hypothesis test by finding the P-value and report the conclusion. [**8**]

(c) Find a 90% confidence interval for the mean fat content of a portion of moo shoo pork. [**5**]

Question 7 (10 marks). Let T be an estimator for a parameter θ.

(a) Define the **bias** and the **mean squared error** of T. [**2**]

(b) Prove that

$$\text{MSE}(T) = \text{Var}(T) + \{\text{Bias}(T)\}^2.$$

[**8**]

End of Paper.