Queen Mary
University of London

B. Sc. Examination by course unit 2014

MTH4106 Introduction to Statistics

Duration: 2 hours

Date and time: 30 April 2014, 10.00-12.00

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. The unauthorized use of material stored in pre-programmable memory constitutes an examination offence. 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 which is not to be assessed.

Important note: the Academic Regulations state that possession of unauthorized material at any time by a student who is under examination conditions is an assessment offence and can lead to expulsion from QMUL.

Please check now to ensure you do not have any notes, mobile phones or unauthorised electronic devices on your person. If you have any, then please raise your hand and give them to an invigilator immediately. Please be aware that if you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will 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. Disruption caused by mobile phones is also an examination offence.

Exam papers must not be removed from the examination room.

Examiner(s): L. I. Pettit and H. Maruri-Aguilar

 TURN OVER

Question 1 (20 marks) In a study of the American black bear researchers wanted to develop a model that could be used to predict a bear's weight from its length. The following data represent the lengths and weights of 12 American black bears.

Total Length (cm)	Total Weight (kg)
139.0	110
138.0	60
139.0	90
120.5	60
149.0	85
141.0	100
141.0	95
150.0	85
166.0	155
151.5	140
129.5	105
150.0	110

(a) What sort of variables are length and weight? [2]

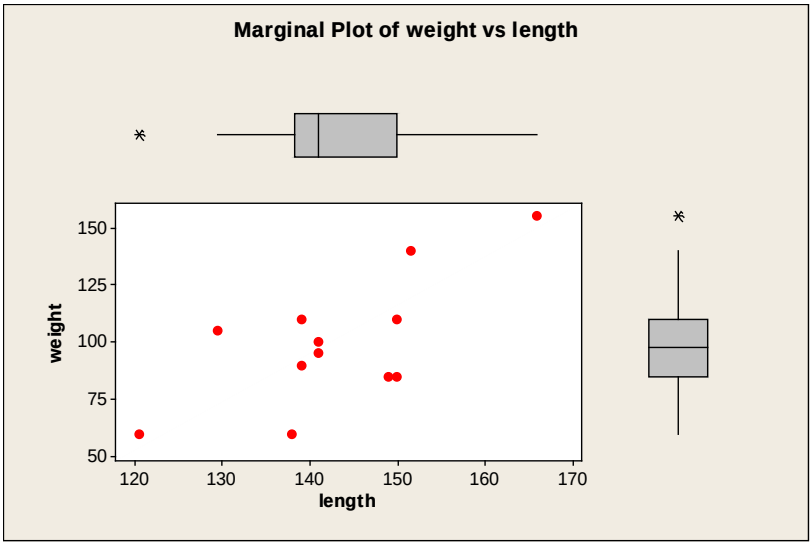
(b) Find the sample correlation coefficient between weight and length. [6]

(c) A marginal plot of weight versus length was found using Minitab and is shown on the next page. Comment on whether the scatter diagram is consistent with the correlation coefficient you have found. [2]

(d) Comment on the distribution of weights by using the information in the boxplot. [4]

(e) Comment on the distribution of lengths by using the information in the boxplot. [4]

(f) One of the researchers noted that the relationship between length and weight may differ for male and female bears. If it was available how could you include information about the sex of the bears on a single scatter diagram? [2]



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 when X is a geometric random variable $X \sim \mathrm{Geom}(p)$ with parameter p, $0 < p < 1$ then its probability generating function is

$$G(t) = \frac{pt}{1-qt},$$

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

Question 3 (15 marks) The random variables $X_i, i = 1, \ldots, n$ are independent and each has a normal distribution with mean μ_i and variance σ_i^2. The sample total is given by $T = \sum_{i=1}^{n} X_i$.

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

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

(iii) Which result allows you to deduce that T is normally distributed? [2]

(b) A company packs parcels of books. Suppose the weights of hardback books are normally distributed with mean 2kg and standard deviation 0.5 kg and paperback books are normally distributed with mean 0.8kg and standard deviation 0.3kg. Assume the weights are all independent of each other. A parcel has three hardback books and six paperback books. Find the probability that the parcel weighs more than 11kg. [7]

Question 4 (5 marks) State the Central Limit Theorem. [5]

Question 5 (10 marks) Let X, Y and Z be random variables with a joint distribution.

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

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

(c) Suppose that X and Y are independent, with $\mathrm{Var}(X) = 9$ and $\mathrm{Var}(Y) = 10$, and that $Z = 2X + Y$. Find the correlation between X and Z. [5]

Question 6 (10 marks) Using Minitab, a student enters the numbers 1–300 in order into column `C1`. She simulates 300 values from the distribution Poisson(6) and puts them in column `C2`. Call these values $x_1, \ldots, x_{300}$. She uses the calculator to find `Partial Sum(C2)` and store it in column `C3`, and then to calculate `C3/C1` and store it in column `C4`. Finally she creates a scatterplot of `C4` against `C1`.

(a) Write down the entries in row 5 of columns `C1`, `C2`, `C3` and `C4`. [4]

(b) Describe the important features of the scatterplot, either in words or with a labelled sketch. [4]

(c) What theorem does this scatterplot illustrate? [2]

Question 7 (15 marks) A company produces steel rods which are supposed to be 36cm long. When the machine making these rods is working properly the rods are normally distributed with mean 36cm and standard deviation 0.035cm. Every week the quality control takes a sample of 20 rods and finds the sample mean length. The sample mean this week is 36.015cm. They wish to test if the machine is working properly this week.

(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 99% confidence interval for the mean length of rod this week. [5]

Question 8 (10 marks) (a) Let T be an estimator for a proportion θ. Define the *bias* and *mean squared error* of T. [2]

(b) Prove that

$$\mathrm{MSE}(T) = \mathrm{Var}(T) + [\mathrm{Bias}(T)]^2.$$

[8]

End of Paper