INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

20 September 2018 (am)

Subject CT3 – Probability and Mathematical Statistics Core Technical

Time allowed: Three hours

INSTRUCTIONS TO THE CANDIDATE

1. *Enter all the candidate and examination details as requested on the front of your answer booklet.*

2. *You must not start writing your answers in the booklet until instructed to do so by the supervisor.*

3. *You have 15 minutes of planning and reading time before the start of this examination. You may make separate notes or write on the exam paper but not in your answer booklet. Calculators are not to be used during the reading time. You will then have three hours to complete the paper.*

4. *Mark allocations are shown in brackets.*

5. *Attempt all 10 questions, beginning your answer to each question on a new page.*

6. *Candidates should show calculations where this is appropriate.*

Graph paper is NOT required for this paper.

AT THE END OF THE EXAMINATION

Hand in BOTH your answer booklet, with any additional sheets firmly attached, and this question paper.

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator from the approved list.

1 A data set of 20 observations has mean 45 and standard deviation 25.4. The data set is reviewed and one observation which was incorrectly recorded as 130 is now corrected to 30.

Determine the mean and standard deviation of the corrected data. [4]

2 A random variable, X, has the probability generating function $G_X(t)$ where

$$G_X(t) = 0.4096 + 0.4096t + 0.1528t^2 + 0.0256t^3 + 0.0016t^4$$

(i) Determine the probability $P(X = 3)$ using $G_X(t)$. [1]

You are now given that X follows a binomial distribution.

(ii) Determine the parameter values of the distribution of X. [3]

[Total 4]

3 A sports scientist is building a statistical model to describe the number of attempts a high jump athlete will have to make until she succeeds in clearing a certain height for the first time during an indoor sports event. For this model the scientist considers a geometric distribution with probability of success p. The cumulative distribution function of the geometric distribution is given as

$$F_X(x) = 1 - (1 - p)^x, \quad x = 1, 2, 3, \ldots$$

(i) (a) State the assumptions that the scientist needs to make for considering this distribution.

(b) Comment on the validity of the assumptions in part (i)(a). [3]

The athlete has tried n jumps without success.

(ii) (a) Determine the probability that the athlete will require more than x additional jumps to succeed in clearing the height.

(b) Comment on what the answer in part (ii)(a) means for the athlete. [3]

[Total 6]

4 We consider three groups of policyholders: A, B and C. We denote by X_A the random variable for the number of claims that a randomly chosen policyholder in group A submits during a calendar year. X_B and X_C denote the corresponding random variables for policyholders in groups B and C. We assume that X_A, X_B and X_C have Poisson distributions with parameters $\lambda_A = 0.2$, $\lambda_B = 0.1$ and $\lambda_C = 0.05$ depending on the group. Each policyholder belongs to exactly one group and group membership does not change during the lifetime of a policyholder. It is assumed that any individual policyholder submits claims during any year independently of claims submitted by other policyholders.

An insurance company has a portfolio of policies with 20% of policyholders belonging to group A, 20% belonging to group B and the remaining policyholders belonging to group C.

The insurance company chooses a policyholder at random.

(i) Determine the probability that this policyholder will submit at least two claims during a year given that he belongs to group A. [2]

The insurance company chooses another policyholder at random but does not know to which group he belongs.

(ii) Show that the probability this policyholder will submit exactly one claim during a year is approximately 0.0794. [3]

(iii) Calculate the probability that this policyholder belongs to group A given that he submitted exactly one claim in the previous year. [2]

[Total 7]

5 In a small empirical study 100 male and 100 female workers in a company are asked about their body weight and then classified into the following three categories: not overweight, overweight and obese. The observed numbers of workers in each category are shown in the following table.

	Not overweight	Overweight	Obese	Total
Females	45	32	23	100
Males	33	41	26	100
Total	78	73	49	200

Test the null hypothesis that weight classification is independent of gender, using a 5% significance level. [5]

6 A poll was conducted with respect to a future referendum, where voters will answer either "yes" or "no" to a question on a political issue. A random sample of 1,106 eligible voters were asked in the poll and 608 of them answered "no". Let p denote the true population proportion of "no" voters.

We wish to predict the result of the referendum, by testing the hypotheses H_0: $p = 0.5$ against H_1: $p > 0.5$.

(i) Perform a suitable test of these hypotheses at the 5% level of significance, stating your conclusion in terms of a predicted result for the referendum. [4]

(ii) (a) Determine a 90% central confidence interval for p.

(b) Explain the effect of using a larger sample on the confidence interval in part (ii)(a). [4]

(iii) (a) Give the definition of the P-value of a hypothesis test.

(b) Derive the P-value of the test in part (i) stating your conclusion in terms of a predicted result for the referendum.

(c) Comment, with reference to your answers in parts (i) and (iii)(b), on the advantages of using a P-value approach for testing, compared with a fixed significance level. [6]

[Total 14]

7 An analyst wishes to model the number of Initial Public Offerings (IPOs) on the local stock exchange each calendar month. Let X_i be a random variable denoting the number of IPOs in month i, where all X_i's are independent. The analyst wishes to model X_i using a Poisson distribution but is aware that there is less activity during the summer so uses the following rates:

$$\lambda_i = \begin{cases} \lambda, & i \neq \textit{July, August} \\ u\lambda, & i = \textit{July, August} \end{cases}$$

(i) Write down the probability function of X_i. [2]

The analyst wishes to estimate u and λ using data from the last 12 months.

(ii) Show that the log likelihood is given by

$$l(x;\lambda,u) = \sum_i x_i \ln\lambda - (10 + 2u)\lambda + (x_{Jul} + x_{Aug})\ln u + C$$

where C is a constant, independent of u and λ. [3]

(iii) Derive the maximum likelihood estimators for u and λ. You are not required to confirm that these estimators maximise the likelihood function. [5]

[Total 10]

8 An insurance company believes that claim amounts in a certain portfolio of policies follow a normal distribution. An analyst chose 61 policies at random which gave a sample mean of £523 and a sample standard deviation of £81.

(i) Determine a 95% confidence interval for the mean claim amount in the portfolio. [3]

(ii) Determine a 95% confidence interval for the variance of claim amounts in the portfolio. [3]

The company assumes that the true mean and standard deviation of claim amounts are the same as those in the sample.

The number of claims per month for the portfolio follows a Poisson process with mean 250.

(iii) Determine the approximate probability that the number of claims in a particular month exceeds 270, justifying any assumptions you use. [5]

(iv) Determine the mean and standard deviation for the total annual amount of claims in the portfolio. [4]

The company has changed its loss assessment processes in order to reduce claim sizes on average, targeting a reduction of £20 compared to the current mean. It does not expect a change to the variability of claim amounts. The company intends to verify whether the target has been met by using a sample of claims to test the null hypothesis that there is no change, against a one-sided alternative hypothesis. Company policy is to perform statistical tests at a significance level of 5%.

(v) Determine the smallest number of claims that would need to be sampled under the new processes for a £20 reduction to be statistically significant in the test. [5]

[Total 20]

9 For an investigation into drinking habits a random sample of men aged 16–90 is obtained. The following data are reported for men belonging to different age groups:

Age group	16–24	25–44	45–64	65 and over
Average units per week	3.5	4.8	5.1	4.2
Sample standard deviation	2.3	1.8	1.6	1.1
Sample size	50	65	60	35

(i) Calculate a 95% confidence interval for the expected value of the average units of alcohol per week consumed by men aged 16–24 based on the sample above. [2]

(ii) Calculate the overall average units of alcohol per week consumed by men aged 16–90 in the sample above. [3]

(iii) Test the hypothesis, using an analysis of variance, that the mean number of units of alcohol per week is the same for all age groups. [8]

(iv) Calculate a 95% confidence interval for the expected units of alcohol per week consumed by all men aged 16–90 based on the sample above. [2]

(v) Comment on your results in parts (iii) and (iv), in particular, whether the result in part (iv) should be used to draw inference about the drinking habits of an individual. [2]

[Total 17]

10 A statistician has a series of bivariate data $\{(x_1,y_1), (x_2,y_2), \ldots (x_n,y_n)\}$ and wishes to perform a linear regression on these data.

(i) State the equation that must be minimised to give the least squares estimates of the regression coefficients. [1]

(ii) Derive the least squares estimate of the slope coefficient from the equation in part (i). [5]

For a sample of 44 fish, the age (days) and length (millimetres) of each fish are measured. Denote age by X and length by Y. The following summary data are obtained:

$$\sum x_i = 3{,}660, \sum x_i^2 = 389{,}684, \sum y_i = 136{,}727, \sum y_i^2 = 500{,}813{,}951,$$
$$\sum x_i y_i = 13{,}609{,}918$$

(iii) Determine the coefficients for a linear regression of Y on X. [5]

(iv) Calculate the sample correlation coefficient between x and y. [2]

[Total 13]

END OF PAPER