INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

23 April 2015 (pm)

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. *Mark allocations are shown in brackets.*

4. *Attempt all 11 questions, beginning your answer to each question on a new page.*

5. *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 Two groups of students sat the same exam. The marks in the first group of 64 students had an average of 52 and a standard deviation of 9. The marks in the second group of 42 students had an average of 45 and a standard deviation of 8.

Calculate the average and standard deviation of the combined data set of 106 students. [4]

2 A random sample of size n consists of k distinct observations $x_1, x_2, \ldots, x_k$ which have been observed with frequencies $f_1, f_2, \ldots, f_k$ where $n = \Sigma_{i=1}^{k} f_i$. Consider the deviations of x from a constant A, giving the observations $d_i = x_i - A$ for $i = 1, \ldots, k$.

Show that the sample variance of the x_i values is given by:

$$s_x^2 = (\Sigma_{i=1}^{k} f_i d_i^2 - (\Sigma_{i=1}^{k} f_i d_i)^2 / n) / (n-1) .$$ [4]

3 Assume that in a large portfolio of insurance contracts the claim size is a normally distributed random variable with expected value 1000. Also assume that the number of claims is a random variable following a Poisson distribution with parameter $\lambda = 400$.

(i) Calculate the expected value of the total claim amount from contracts in this portfolio. [1]

(ii) Calculate a lower limit for the standard deviation of the total amount of claims from contracts in this portfolio. [3]
[Total 4]

4 An insurance company experiences claims from 290 insurance policies in a year on a portfolio of 900 policies. Only one claim can be made on a policy in a year. The company assumes that all policies are independent of each other.

Determine a 90% confidence interval for the proportion of policies on which a claim is made in a year. [3]

5 A random sample of 30 observations is drawn from a normal distribution with unknown variance.

(i) Write down an expression for the distribution of S, the population standard deviation. [1]

The sample standard deviation, s, is 7.5.

(ii) Calculate a 95% confidence interval for the population standard deviation. [4]
[Total 5]

6 Let $X_1, X_2, \ldots, X_6$ be a random sample from a population following a Gamma(2,1) distribution. Consider the following two estimators of the mean of this distribution:

$$\hat{\theta}_1 = \bar{X} \text{ and } \quad \hat{\theta}_2 = \frac{9}{30}(X_1 + X_2 + X_3) + \frac{1}{30}(X_4 + X_5 + X_6)$$

where $\bar{X}$ is the mean of the sample.

(i) Determine the sampling distribution of $\bar{X}$ using moment generating functions. [5]

(ii) Derive the bias of each estimator $\hat{\theta}_1$ and $\hat{\theta}_2$. [2]

(iii) Derive the mean square error of each estimator $\hat{\theta}_1$ and $\hat{\theta}_2$. [5]

(iv) Compare the efficiency of the two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$. [1]

[Total 13]

7 A continuous random variable X has the cumulative distribution function $F_X(x)$ given by:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}x^3, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

(i) Determine the probability density function of X. [2]

(ii) Calculate $P(0.5 < X < 1)$. [2]

Let $Y = \sqrt{X}$.

(iii) Determine the cumulative distribution function and the probability density function of Y. [4]

(iv) Calculate the expected values of X and Y. [4]

[Total 12]

8 The random variables X and Y have a joint probability distribution with density function:

$$f_{xy}(x,y) = \begin{cases} 3x, \ 0 < y < x < 1 \\ 0, \ \text{otherwise} \end{cases}$$

(i) Determine the marginal densities of X and Y. [4]
(ii) State, with reasons, whether X and Y are independent. [2]
(iii) Determine $E[X]$ and $E[Y]$. [2]
[Total 8]

9 An insurance company has calculated premiums assuming that the average claim size per claim for a certain class of insurance policies does not exceed £20,000 per annum. An actuary analyses 25 such claims that have been randomly selected. She finds that the average claim size in the sample is £21,000 and the sample standard deviation is £2,500. Assume that the size of a single claim is normally distributed with unknown expectation α and variance σ^2.

(i) Calculate a 95% confidence interval for α based on the sample of 25 claims. [3]

(ii) Perform a test for the null hypothesis that the expected claim size is not greater than £20,000 at a 5% significance level. [3]

(iii) Discuss whether your answers to parts (i) and (ii) are consistent. [2]

(iv) Calculate the largest expected claim size, $\alpha_{0,}$ for which the hypothesis $\alpha \leq \alpha_0$ can be rejected at a 5% significance level based on the sample of 25 claims. [2]

The insurer is also concerned about the number of claims made each year. It is found that the average number of claims per policy was 0.5 during the year 2011. When the analysis was repeated in 2012 it was found that the average number of claims per policy had increased to 0.6. These averages were calculated on the basis of random samples of 100 policies in each of the two years. Assume that the number of claims per policy per year has a Poisson distribution with unknown expectation λ and is independent from the number of claims in any other year or for any other policy.

(v) Perform a test at 5% significance level for the null hypothesis that $\lambda = 0.6$ during the year 2011. [3]

(vi) Perform a test to decide whether the average number of claims has increased from 2011 to 2012. [3]
[Total 16]

10 A factory manager is measuring the times that different teams are taking to assemble an electronic device. For each team, A, B and C, the following times are measured on a given day (in minutes):

A	52.1	51.4	55.8	53.6	50.3
B	46.6	49.0	51.2	48.3	71.0
C	54.2	56.3	57.4	59.3	58.2

$\sum y_A = 263.2, \sum y_B = 266.1, \sum y_C = 285.4$

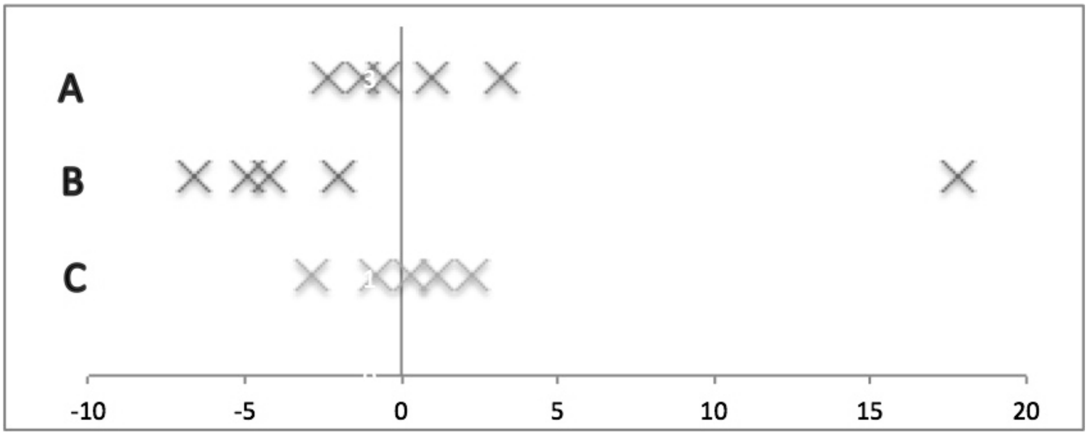
$\sum y_A^2 = 13,873.06, \sum y_B^2 = 14,567.89, \sum y_C^2 = 16,305.82$

The manager decides to use analysis of variance (ANOVA) to assess whether there is a difference between the teams and produces the following table:

Source of Variation	SS	df	MSS
Between Teams	58.25	2	29.12
Residual	439.45	12	36.62
Total	497.70	14	

(i) Test the hypothesis that there is no difference between team times against a general alternative. [3]

Plotting the residuals from the ANOVA gives the following chart:



(ii) Explain why the ANOVA may or may not be valid. [1]

On further investigation it is discovered that the time of 71.0 minutes for Team B was an error and the observation is removed from the sample.

(iii) (a) Prepare an updated ANOVA table.

(b) Test the hypothesis that there is no difference between the team times, using your updated ANOVA table. [7]

(iv) Comment on the results of parts (i) and (iii). [2]

[Total 13]

11 Consider the set of paired data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ to which we fit the linear regression model:

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2),$$

where the Y_i are independent random variables, and α, β and σ^2 are unknown parameters.

(i) (a) Show that the least squares estimator $\hat{\beta}$ is an unbiased estimator for parameter β.

(b) Show that the variance of $\hat{\beta}$ is given by:

$$V(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$$

where $S_{xx} = \sum (x_i - \bar{x})^2$.

You are given that the covariance of the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ is given by

$$\text{cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{x}\sigma^2}{S_{xx}} .$$

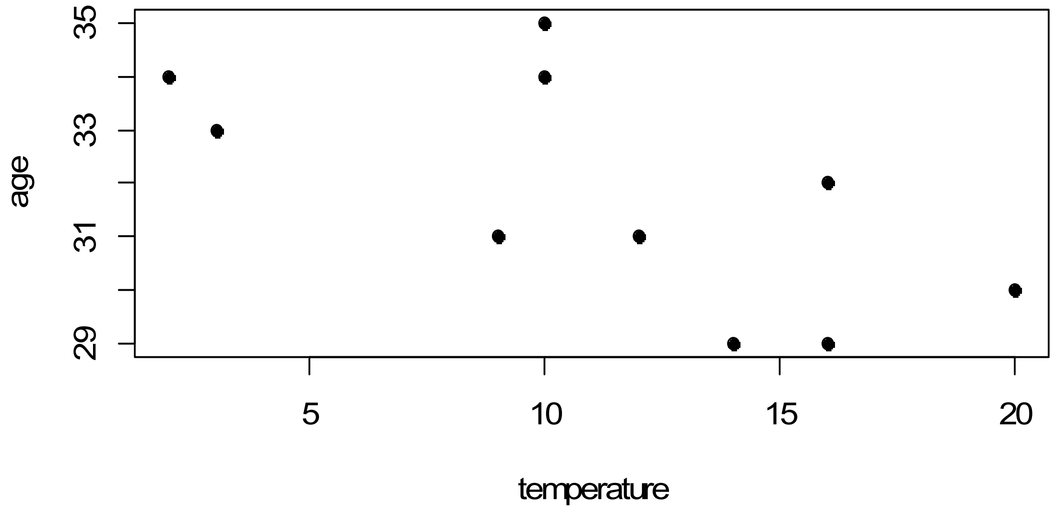
(c) Explain why a data transformation of the form $z_i = x_i - \bar{x}$ would result in a better model.

[7]

A study investigated whether there is an association between babies' first crawling age and the average temperature during the month they first try to crawl (about 6 months after birth). It is thought that in colder months heavier clothing may restrict a baby's movement more than in warmer months and so the age at which the baby first crawls would be expected to be greater. A random sample of babies was taken and the results of the study are given in following table:

Baby	1	2	3	4	5	6	7	8	9	10
Average temperature , °C (x)	16	2	10	20	12	16	14	9	3	10
Crawling Age, weeks (y)	32	34	34	30	31	29	29	31	33	35

A scatterplot of the data is given below:



(ii) Comment on the relationship between crawling age and average temperature based on these observations. [2]

The linear regression model $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ was fitted to these data, and some of the results are given below:

	Estimate	*Standard error*
α	34.5501	1.2617
β	−0.2455	0.1015
σ	1.733	

Also, $R^2 = 0.4226$.

Based on these results:

(iii) (a) Perform a statistical test to investigate the hypothesis that there is no linear relationship between crawling age and average temperature.

(b) Comment on the fit of the model. [5]

Consider the data transformation mentioned in part (i)(c), giving the model $Y_i \sim N(\gamma + \delta z_i, \sigma^2)$.

(iv) (a) Show that the estimators of the parameters in this model are given by:

$$\hat{\gamma} = \hat{\alpha} + \hat{\beta}\bar{x}, \ \hat{\delta} = \hat{\beta}$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the parameter estimators of the model using the original data.

(b) Write down the fitted model under this transformation. [4]

[Total 18]

END OF PAPER