

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

29 September 2014 (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. *Mark allocations are shown in brackets.*
4. *Attempt all 10 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.

<p><i>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.</i></p>
--

- 1** A sample of marks from an exam has median 49 and interquartile range 19. The marks are rescaled by multiplying by 1.2 and adding 6.

Calculate the new median and interquartile range. [4]

- 2** Consider an insurer that offers two types of policy: home insurance and car insurance. 70% of all customers have a home insurance policy, and 80% of all customers have a car insurance policy. Every customer has at least one of the two types of policies.

Calculate the probability that a randomly selected customer:

- (i) does not have a car insurance policy. [1]
(ii) has car insurance and home insurance. [1]
(iii) has home insurance, given that he has car insurance. [2]
(iv) does not have car insurance, given that he has home insurance. [2]

[Total 6]

- 3** Let N be a random variable describing the number of withdrawals from a bank branch each day. It is assumed that N is Poisson distributed with mean μ . Let X_i , the random variable describing the amount of each withdrawal, be exponentially distributed with mean $1/\lambda$. All X_i are independent and identically distributed. Let S denote the total amount withdrawn from that branch in a day i.e.

$$S = \sum_{i=1}^N X_i$$

with $S = 0$ if $N = 0$.

- (i) Derive the moment generating function of S . [4]
(ii) Calculate the mean and variance of S if $\mu = 100$ and $\lambda = 0.025$. [3]

[Total 7]

- 4** Consider six life policies, each on one of six independent lives. Each of four of the policies has a probability of $2/3$ of giving rise to a claim within the next five years, and each of the other two policies has a probability of $1/3$ of giving rise to a claim within the next five years. It is assumed that only one claim can arise from each policy.

- (i) Calculate the expected number of claims which will arise from the six policies within the next five years. [2]
(ii) Calculate the probability that exactly one claim will arise from the six policies within the next five years. [2]
(iii) Calculate the probability that two policies chosen at random from the six policies will both give rise to claims within the next five years. [4]

[Total 8]

- 5** Consider two random variables X and Y with $E[X] = 2$, $V[X] = 4$, $E[Y] = -3$, $V[Y] = 1$, and $\text{Cov}[X, Y] = 1.6$.

Calculate:

- (a) the expected value of $5X + 20Y$.
- (b) the correlation coefficient between X and Y .
- (c) the expected value of the product XY .
- (d) the variance of $X - Y$.

[4]

- 6** In a medical study conducted to test the suggestion that daily exercise has the effect of lowering blood pressure, a sample of eight patients with high blood pressure was selected. Their blood pressure was measured initially and then again a month later after they had participated in an exercise programme. The results are shown in the table below:

<i>Patient</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
<i>Before</i>	155	152	146	153	146	160	139	148
<i>After</i>	145	147	123	137	141	142	140	138

- (i) Explain why a standard two-sample t -test would not be appropriate in this investigation to test the suggestion that daily exercise has the effect of lowering blood pressure. [1]
- (ii) Perform a suitable t -test for this medical study. You should clearly state the null and alternative hypotheses. [7]

[Total 8]

- 7 Consider the following discrete distribution with an unknown parameter p for the distribution of the number of policies with 0, 1, 2, or more than 2 claims per year in a portfolio of n independent policies.

<i>number of claims</i>	0	1	2	<i>more than 2</i>
<i>probability</i>	$2p$	p	$0.25p$	$1 - 3.25p$

We denote by X_0 the number of policies with no claims, by X_1 the number of policies with one claim and by X_2 the number of policies with two claims per year. The random variable $X = X_0 + X_1 + X_2$ is then the number of policies with at most two claims.

- (i) Derive an expression for the maximum likelihood estimator \hat{p} of parameter p in terms of X and n . [5]
- (ii) Show that the estimator obtained in part (i) is unbiased. [3]

The following frequencies are observed in a portfolio of $n = 200$ policies during the year 2012:

<i>number of claims</i>	0	1	2	<i>more than 2</i>
<i>observed frequency</i>	123	58	13	6

A statistician proposes that the parameter p can be estimated by $\tilde{p} = 58/200 = 0.29$ since p is the probability that a randomly chosen policy leads to one claim per year.

- (iii) Estimate the parameter p using the estimator derived in part (i). [1]
- (iv) Explain why your answer to part (iii) is different from the proposed estimated value of 0.29. [2]

An alternative model is proposed where the probability function has the form

<i>number of claims</i>	0	1	2	<i>more than 2</i>
<i>probability</i>	p	$2p$	$0.25p$	$1 - 3.25p$

- (v) Explain how the maximum likelihood estimator suggested in part (i) needs to be adapted to estimate the parameter p in this new model. [1]
- (vi) Suggest a suitable test to use to make a decision about which of the two models should be used based on empirical data. [1]

[Total 13]

- 8** The promoter of a touring dance show wishes to analyse how the price per ticket affects the size of its audiences. She tests two prices, £14 and £16, over 10 shows each which give rise to the following attendances.

<i>Price</i>										
£14	120	115	130	127	124	110	121	129	118	122
£16	111	107	101	115	111	105	99	104	110	98

- (i) Calculate the mean and standard deviation of the attendance for each sample. [4]
- (ii) Perform a statistical test to determine whether the variances of the attendance are equal under the two prices. [3]
- (iii) Perform a t-test to determine whether the mean attendance is the same under the two prices. [4]
- (iv) Calculate a 95% confidence interval for the difference between the mean show revenue under each price. [4]
- (v) Comment on which price the promoter should choose. [3]
- [Total 18]

- 9 The following data (x) give the acidity (in appropriate units) of three different varieties of grape.

<i>Variety</i>					<i>Mean</i>	<i>Variance</i>
A	8	7	18	15	12.0	28.7
B	90	74	200	122	121.5	3137.0
C	897	493	812	365	641.8	64284.9

A wine maker wants to test whether there are differences in the mean acidity level of the three varieties and wishes to use analysis of variance (ANOVA) methodology.

- (i) Explain why ANOVA should not be used for the data as given in the table above. [2]

A statistician suggests two transformations of the original data:

- the natural logarithm, $y = \ln(x)$,
- and the square root, $z = \sqrt{x}$.

These give the following summary statistics:

<i>Variety</i>	$y = \ln(x)$		$z = \sqrt{x}$	
	<i>Mean</i>	<i>Variance</i>	<i>Mean</i>	<i>Variance</i>
A	2.4075	0.2136	3.3975	0.6046
B	4.7250	0.1892	10.8200	5.9242
C	6.4000	0.1800	24.9425	26.4567

The wine maker then decides to use the natural logarithm transformation (y) of the original data.

- (ii) Justify the wine maker's choice of data transformation for performing the analysis. [1]
- (iii) Perform ANOVA on the transformed data, y , to investigate possible differences in the mean acidity level of the three grape varieties and state your conclusions. [6]
- (iv) Calculate 95% confidence intervals for the mean values of each of the three varieties on the original scale, based on the ANOVA performed on the transformed values. [6]
- (v) Comment on the intervals obtained in part (iv) in relation to your conclusion in part (iii). [2]

[Total 17]

- 10** An insurer has collected data on average alcohol consumption (units per week) and cigarette smoking (average number of cigarettes per day) in eight regions in the UK.

<i>Region, i</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>Average</i>
<i>Alcohol units per week, x_i</i>	15	25	21	29	13	18	21	17	19.875
<i>Cigarettes per day, y_i</i>	4	8	8	10	6	9	7	5	7.125

For these observations we obtain:

$$\sum x_i y_i = 1,190; \quad \sum x_i^2 = 3,355; \quad \sum y_i^2 = 435$$

- (i) Calculate the coefficient of correlation between alcohol consumption and cigarette smoking. [4]
- (ii) Calculate a 95% confidence interval for the true correlation coefficient. You may assume that the joint distribution of the two random variables is a bivariate normal distribution. [6]
- (iii) Fit a linear regression model to the data, by considering alcohol consumption as the explanatory variable. You should write down the model and estimate the values of the intercept and slope parameters. [3]
- (iv) Calculate the coefficient of determination R^2 for the regression model in part (iii). [1]
- (v) Give an interpretation of R^2 calculated in part (iv). [1]
- [Total 15]

END OF PAPER