SECTION A: Compulsory questions

QUESTION 1

The National Lottery runs two Lotto prize funds: a Jackpot fund, for people who guess all the numbers correctly, and a non-jackpot fund, for those who only have some numbers right. The amount of money placed into each fund before a draw is proportional to the number of tickets sold for that draw; if there were no jackpot winners on the previous draw, the previous jackpot fund is added to the current fund.

The table below shows the total amount paid out (in units of £1 million) in the 18 draws in December 2010 and January 2011, split into jackpot and non-jackpot payouts.

Wednesdays				Saturdays			
Date	Total	Jackpot	Non-jackpot	Date	Total	Jackpot	Non-jackpot
01 Dec	17.8	11.3	6.5	04 Dec	14.0	4.6	9.4
08 Dec	7.8	2.4	5.4	11 Dec	13.9	4.6	9.3
15 Dec	7.8	2.6	5.2	18 Dec	13.5	4.8	8.7
22 Dec	5.3	0	5.3	25 Dec	16.8	7.2	9.6
29 Dec	5.2	0	5.2	01 Jan	16.6	7.3	9.3
05 Jan	4.9	0	4.9	08 Jan	17.8	8.2	9.6
12 Jan	5.7	0	5.7	15 Jan	10.5	0	10.5
19 Jan	17.8	10.5	7.3	22 Jan	14.4	4.5	9.9
26 Jan	8.0	2.9	5.1	29 Jan	14.3	4.7	9.6

(i) Use a stem-and-leaf diagram to compare the size of the total payouts on Wednesdays with those on Saturdays. **[3 marks]**

(ii) Use a box-and-whisker plot to compare the size of the non-jackpot payouts on Wednesdays with those on Saturdays. **[5 marks]**

(iii) Comment on your findings. **[2 marks]**

[Total: 10 marks]

QUESTION 2

X and Y are discrete random variables with joint probability function as shown in the following table:

$p_{X,Y}(x,y)$	$x=0$	$x=1$	$x=2$
$y=0$	0.2	0.1	0.1
$y=1$	0	0.1	0.4
$y=2$	0	0	0.1

(i) Calculate the marginal probability functions of X and Y. **[1 mark]**

(ii) For each value of $x \in \{0,1,2\}$ calculate the conditional expectation $\mathbb{E}[Y \mid X=x]$. **[2 marks]**

(iii) Evaluate $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ and verify that the tower property holds. **[3 marks]**

(iv) Obtain the probability function of the random variable Z given by $Z = XY$ and use this to evaluate the covariance of X with Y. **[3 marks]**

[Total: 9 marks]

QUESTION 3

A Quantitative Methods module is taken by 25 students on an Underwriting course and by 24 students on a Valuation course. The marks obtained by the Underwriting students are normally distributed with mean 66, standard deviation 12; those obtained by the Valuation students are also normally distributed, with mean 63 and standard deviation 18. Let $\bar{U}$ and $\bar{V}$ denote the sample mean scores obtained by the students on the Underwriting and Valuation courses respectively.

(i) Write down the expectation and variance of $\bar{U}$ and of $\bar{U} - \bar{V}$. **[3 marks]**

(ii) Calculate the probability that $\bar{U} > \bar{V}$. **[3 marks]**

(iii) Marks are rounded to the nearest integer before being presented to the Board of Examiners. The pass mark for the module is 50%.

(a) What is the probability that a Valuation student passes the module?

(b) Let X represent the number of Valuation students who pass the module. Write down the distribution of X, giving the name of the distribution and the value(s) of the parameter(s).

(c) How likely is it that no more than two of the Valuation students will fail the module? **[5 marks]**

[Total: 11 marks]

QUESTION 4

A pack of 30 cards consists of 10 different values (1 to 10) in each of 3 different colours: red, green and blue. A single card is about to be drawn from the pack. Let R, G and B represent the event that the card is red, green or blue respectively, and denote by D_2 and D_3 the event that the number on the card is divisible by 2 or 3 respectively.

Player X wins a prize if the card is red and the number is even; player Y wins a prize if the card is blue or the number is divisible by 3; player Z wins a prize if the card is green.

(i) Using set-theoretic notation, express the events "X wins a prize", "Y wins a prize" and "Z wins a prize" in terms of R, G, B, D_2 and D_3. **[2 marks]**

(ii) For each player, calculate the probability of winning a prize. **[2 marks]**

(iii) For each pair of players, calculate the probability that both players win a prize simultaneously. **[3 marks]**

(iv) How many possible outcomes of the experiment result in no prize being awarded? **[2 marks]**

(v) Draw a Venn diagram to illustrate your results. **[1 mark]**

[Total: 10 marks]

QUESTION 5

(i) Explain the difference between a type I error and a type II error. **[1 mark]**

(ii) X is a Poisson random variable with unknown mean μ. A test is required of $H_0 : \mu = 2$ against $H_1 : \mu > 2$ based on the single value of X. A test which is suggested is to reject H_0 if the observed value of X is greater than 4.

(a) Calculate the probability of a type I error.

(b) Calculate the probability of a type II error if in fact $\mu = 5$. **[4 marks]**

(iii) $X_1, X_2, \ldots, X_{20}$ is a sequence of independent Poisson random variables, each having unknown mean μ. Suggest a test at the 5% level of significance of $H_0 : \mu = 2$ against $H_1 : \mu > 2$ based on this sample of size 20. You should answer by writing down the test statistic, the (approximate) distribution of the test statistic if the null hypothesis is true and the critical region for this test. **[4 marks]**

[Total: 9 marks]

QUESTION 6

Feline Friends magazine runs a reader survey in which readers are asked to assess their overall level of happiness (y) on a scale from 1 to 10 and to report how many cats they have at home (x). The magazine then undertakes a linear regression to determine how much extra happiness is attributable to one cat. The model fitted is $y_i = \alpha + \beta x_i + E_i$, where the E_i are assumed to be independent random variables with mean 0, variance σ^2. The parameter estimates resulting from the 42 responses received are $\hat{\alpha} = 4.473$, $\hat{\beta} = 0.211$, $\hat{\sigma}^2 = 5.977$.

(i) Carry out a test to determine whether β is significantly different from zero at the 5% level. [Note: $S_{xx} = 788.5$.] **[4 marks]**

(ii) Derive a 95% confidence interval for the expected level of happiness of someone who has 6 cats at home. [Note: $\bar{x} = 4.52$.] **[4 marks]**

(iii) State any additional assumptions which are necessary for the validity of the procedures in (i) and (ii). **[1 mark]**

(iv) Figure 1 below is a scatter diagram for this data set. Does this diagram, or any other aspect of the experiment, cast doubt on whether the assumptions are satisfied? **[2 marks]**

[Total: 11 marks]

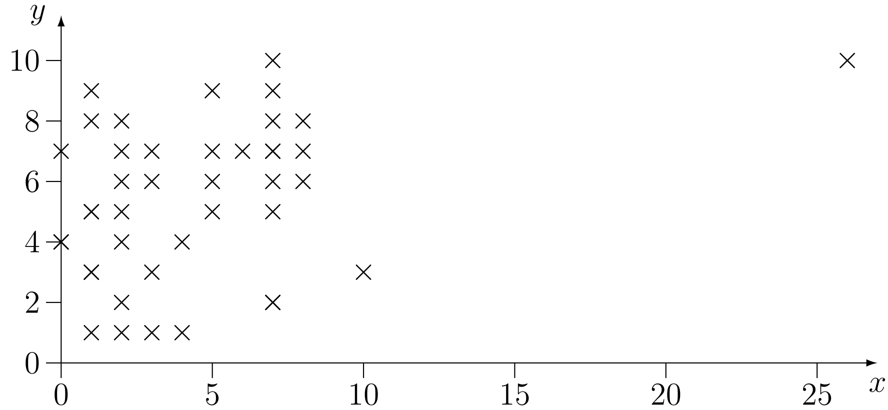

Figure 1: Scatter diagram of happiness (y) against cats (x)

SECTION B: Optional questions

QUESTION 7

Athanasius is a student who likes to go to the cinema in the evening but also likes to submit coursework on time. On a typical day the number, X, of urgent courseworks is a Poisson random variable with mean 0.25, and the number, Y, of non-urgent courseworks is a Poisson random variable with mean 1.25, independent of X. If Athanasius has any urgent courseworks to do, or if he has three or more non-urgent courseworks to do, he will stay at home in the evening. If, however, $X = 0$ and $Y = y < 3$, he will go to the cinema with probability $(3 - y)/4$, or otherwise stay at home and work.

(i) (a) Calculate the probabilities of the event $\{X > 0\}$, the event $\{X = 0 \cap Y \geq 3\}$ and the event $\{X = 0 \cap Y = y\}$ for each $y = 0, 1, 2$.

(b) Draw a tree diagram to illustrate the situation.

(c) What is the probability that Athanasius goes to the cinema on a given evening?
[8 marks]

(ii) Athanasius becomes interested in football and decides that he will go to watch City Academicals whenever they have a home match, which happens with probability 10% on any given evening.

(a) Expand your tree diagram to include this information.

(b) If Athanasius is not at home in the evening, what is the probability that he is watching the football? **[4 marks]**

(iii) Athanasius decides to ask his friends on his course how many evenings a week they normally go out, rather than staying at home, and to compare their answers with those supplied by his friends at the football matches. The answers he receives are as follows:

Course friends	1	0	3	2	2	1	4	0	1	5
Football friends	3	7	5	2	4	1	6	6		

(a) Why would it not be a good idea to use a two-sample t test to compare these data sets?

(b) Use a rank sum test to test whether Athanasius' friends from City Academicals tend to spend more evenings out than his friends on his course. **[8 marks]**

[Total: 20 marks]

QUESTION 8

(i) State the characteristic properties of the distribution function (c.d.f.) of a continuous random variable. **[2 marks]**

(ii) Verify that the function $F(x)$ given by

$$F(x) = \begin{cases} 1 - (1 + \lambda x)e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

is a distribution function, and calculate the associated density function $f(x)$. **[5 marks]**

(iii) Verify that the expectation of this distribution is $2\lambda^{-1}$ and evaluate the variance of the distribution. **[4 marks]**

(iv) A statistician has a data set consisting of 50 observations, tabulated below, and wishes to determine whether the observations come from the distribution whose distribution function is $F(x)$, for some value of λ.

Range of values	Frequency
0.0 to 0.99	20
1.0 to 1.99	10
2.0 to 2.99	9
3.0 to 3.99	6
4.0 to 4.99	4
5.0 to 5.99	1

(a) Calculate the sample mean of the observations.

(b) Obtain an estimate of λ by equating the sample mean to the expectation of the distribution.

(c) Use this estimate for λ to calculate the expected number of observations falling in each of the given ranges.

(d) Carry out a goodness-of-fit test to determine whether it is reasonable to assume that the observations are taken from the distribution whose distribution function is $F(x)$. **[9 marks]**

[Total: 20 marks]

QUESTION 9

(i) Observations x_i represent the percentage daily changes in the exchange rate between the pound sterling and the euro for first six months of 2010, excluding days on which little or no trading took place. Observations y_i represent the same information for the last six months of 2010. The observations are summarised in the following table:

Data set	Number of observations	Sample mean	Sample variance	Number of positive values
First 6 months (x)	163	-0.030	0.147	75
Last 6 months (y)	167	0.054	0.185	95

(a) Test at the 5% level of significance whether the variance of the percentage daily changes is the same for the last six months as it was for the first six months. Ensure that you state the null and alternative hypotheses explicitly.
[Note: $F_{162,166,0.025} = F_{166,162,0.025} = 1.359$, $F_{162,166,0.05} = F_{166,162,0.05} = 1.294$.]

(b) Obtain a 95% confidence interval for the difference between the mean daily change over the last six months and the mean daily change over the first six months.

(c) Test whether the second half of 2010 contained a greater proportion of days when the exchange rate moved in the positive direction than the first half of 2010. Report the result of this test in the form of a p-value.

[12 marks]

(ii) An investor has £100, which he wants to invest in shares in company A and company B, each of which currently cost £1. The value of a share in company A in one year's time is a random variable X with mean 110p, standard deviation 12p. The value of a share in company B at that time is a random variable Y with mean 105p, standard deviation 5p. In addition, the correlation coefficient ρ_{XY} is equal to 0.1.

(a) Suppose the investor buys a shares in company A and b shares in company B, keeping the remaining $\pounds(100 - a - b)$ as cash. Write down the value, Z, of this portfolio in one year's time, as a linear combination of X and Y.

(b) Obtain expressions for the expectation and variance of Z, in terms of a and b.

(c) The investor wants his portfolio to have an expected value of £107.50 in one year's time, with as little risk as possible. Find the values of a and b which minimise the variance of Z subject to this condition, and comment on your solution.

[8 marks]

[Total: 20 marks]

QUESTION 10

(i) If $X_1, X_2, \ldots, X_n$ are independent observations from a distribution with unknown mean μ and unknown variance σ^2, show that the sample variance S_X^2 has expectation equal to σ^2. **[4 marks]**

(ii) The percentage scores in an examination taken by 81 students in 2010 are tabulated as follows:

Range	0–29	30–39	40–49	50–59	60–69	70–79	80-100
Frequency	9	6	9	14	16	14	13

(a) Use a suitable diagram to illustrate the data set and indicate whether you think the data exhibit skewness.

(b) Estimate the median and quartiles of the data set. Do your estimates confirm your impression as to the skewness of the data set? Give a reason for your answer. **[7 marks]**

(iii) The examination scores have sample mean 59.55 and sample variance 376.5.

(a) Derive 95% confidence intervals for the mean, μ, and the variance, σ^2, of the distribution from which the observations were taken.

(b) List the assumptions you are making in producing the confidence intervals.

(c) Over many previous years the mean score on the examination has been 51%, with standard deviation 19%. Without carrying out any formal tests, comment on any differences between the 2010 exam results and those from previous years. **[9 marks]**

[Total: 20 marks]