BIRKBECK
(University of London)

BSc EXAMINATION
SCHOOL OF BUSINESS, ECONOMICS AND INFORMATICS

Probability and Statistics
EMMS098S5

30 credits

Thursday, 2nd June, 2016
Morning, 10:00 a.m. - 1:00 p.m.

This examination contains two sections: Section A (8 questions) and Section B (4 questions). Questions in Section A are worth 5 *marks each and questions in Section B are worth* 20 *marks each.*
Candidates should attempt **all** *of the questions in Section A and* **two** *questions out of the four in Section B.*
New Cambridge Statistical Tables are provided.
Candidates can use their own calculator, provided the model is on the circulated list of authorised calculators or has been approved by the chair of the Mathematics and Statistics Examination Sub-board.

Please turn over

Section A

1. Consider the following events in the toss of three fair coins:

   A: Observe an odd number of heads

   B: Observe two or more heads

   (a) Find $\Pr(A)$, $\Pr(B)$, and $\Pr(A \cup B)$ [1]

   (b) Calculate $\Pr(B|A)$ [2]

   (c) State whether events A and B are independent, justifying your answer. [2]

2. A gardener decides to plant 10 bulbs. It is known that 80% of the bulbs will bloom.

   (a) Suggest a probability distribution for X, the number of bulbs that bloom, including the values of any parameters. [2]

   (b) What is the probability that at least 4 out the 10 bulbs planted will bloom. [3]

3. A salesman assumes that the demand for a certain product is distributed according to a Poisson distribution with mean of two items per day.

   How many items does he have to buy in order to have 95% probability of not having a demand higher than the supply for 7 days? [5]

4. In a science test the marks were normally distributed with mean 68 and standard deviation 10.

   (a) What is the probability that a student scored 75 or more? [3]

   (b) What is the probability that a student scored between 70 and 75? [2]

Please turn over

5. The efficacy of a treatment for adolescent smokers that combines Nicotine Patch with Bupropion is to be studied using a clinical trial. Sixteen patients where randomly allocated to either Group 0 (placebo control group) and Group 1 (treatment group).
At the end of a three-month follow-up study, the change (in ml) in the amount of Carbon monoxide transfer (an indicator of improved lung function) is measured and the sample means $\bar{x}$ and variances s^2 for each group are reported in the table below.

	n (number of patients)	$\bar{x}$	s^2
Group 0	7	-0.208	4.101^2
Group 1	9	3.953	4.630^2

(a) Suggest a hypothesis test to investigate whether there is evidence of significant improvement in the treatment group (higher mean in the treatment Group 1 with respect to the placebo control Group 0), specifying any assumptions and appropriate null and alternative hypotheses. [2]

(b) Carry out the hypothesis test, stating your conclusions clearly. Use a 1% significance level. [3]

6. James wants to test whether the number of smokers frequenting his bar is aligned or not with the national statistics that says that 20% of people smoke. He takes a random sample of 120 people, and 30 of them smoke.

(a) State the null and the alternative hypothesis clearly. Calculate an appropriate test statistic to test whether or not 20% of the people attending the bar smoke, specify any assumption that must be made in order to carry out the test. [2]

(b) Draw your conclusions at 5% significance level. [3]

7. The following results are obtained from rolling 1000 times a six-faced die:

Score:	1	2	3	4	5	6
Frequency:	212	140	156	170	172	150

(a) Suggest a hypothesis test to test if the die is fair or not, specifying any assumptions and appropriate null and alternative hypotheses. [2]

(b) Carry out the hypothesis test, stating your conclusions clearly. Use a 5% significance level. [3]

Please turn over

8. In the 19th century Francis Galton argued that the eye color are inherited. He collected the data from a sample size of 5008 individuals.

	Parent Light	Parent Dark
Child Light	2524	1060
Child Dark	528	896

(a) Suggest a hypothesis test to investigate whether there is association between the eye color of the parents and of the children, specifying any assumptions and appropriate null and alternative hypotheses. [2]

(b) Carry out the relevant test, stating your conclusions clearly. Use $\alpha = 0.05$. [3]

Please turn over

Section B

9. (a) The probability that a dog is affected by a certain rare disease is 0.002. We have collected a random sample of 3000 dogs. We are interested in the probability that there is at most one dog affected by that disease.

   (i) Calculate the exact probability. [4]

   (ii) Calculate the probability by using the Poisson approximation. [4]

   (iii) Calculate the probability by using the Normal approximation. [4]

   (b) Suppose there is a blood test to detect this disease. This test is very accurate: in fact, the probability of the test resulting positive given that the dog has the disease is 0.99, and the probability of the test resulting negative given that the dog does not have the disease is 0.95.

   (i) Find the probability that the test results positive given that the dog does not have the disease. [3]

   (ii) Find the probability that the dog is affected by the disease given that the test results positive. [5]

Please turn over

10. Consider the continuous random variable X with p.d.f.:

$$f(x) = \begin{cases} \theta\, x^{\theta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Show that $f(x)$ is a density function. [4]

(b) Find an expression for the c.d.f. $F(x)$ of this distribution. [4]

(c) Find the mean of X. [4]

(d) Find the variance of X. [5]

(e) Fix $\theta = 2$ and find $\Pr(X \geq 0.5)$. [3]

Please turn over

11. (a) A sample of 20 high-powered copper coils yields the following tensile strength values (in units of kilogram per square centimetre):

280, 155, 329, 140, 307, 116, 202, 262, 130, 131,
187, 187, 292, 83, 207, 197, 134, 294, 163, 217

We assume that the data are normally distributed, with unknown mean and unknown variance.

(i) Calculate a point estimate of the mean. [3]

(ii) Calculate a point estimate of the variance. [4]

(iii) Calculate a 95% confidence interval for the mean. [4]

(iv) Calculate a 95% confidence interval for the variance. [5]

(b) State (without proof) the *Central Limit Theorem*. [4]

Please turn over

12. In a trial to evaluate a new drug, which it is hoped will reduce the cholesterol levels, 9 people were randomly selected to try the new drug. Their cholesterol levels in mg/l were measured before (column `NoDrug`), and after taking the drug (column `Drug`). The measurement are set out below.

Person	`NoDrug`	`Drug`
1	206.13	166.81
2	203.62	181.14
3	226.43	211.65
4	139.81	96.99
5	137.40	141.41
6	131.80	166.91
7	145.41	101.25
8	141.64	169.05
9	216.86	237.90

(a) A test has been performed using `R`:

```
Paired t-test

data:  NoDrug and Drug
t = 0.80838, df = 8, p-value = 0.2211
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -10.97919       Inf
sample estimates:
mean of the differences
               8.443333
```

(i) State precisely the statistical model and assumption that are being used in the `R` output, specify the null and alternative hypotheses in terms of the model parameters. [4]

(ii) Write down the general formula of the test statistics used to perform the test, and the distribution under the null hypothesis. [3]

(iii) State your conclusions clearly, specifying the significance level you used. [3]

(b) A second test has been performed using `R`:

```
Wilcoxon signed rank test

data:  NoDrug - Drug
V = 30, p-value = 0.2129
alternative hypothesis: true location is greater than 0
```

. . . continued

Please turn over

(i) State the assumption that are being used in the `R` output, specify the null and alternative hypotheses. [4]

(ii) State your conclusions clearly, specifying the significance level you used. [3]

(iii) Are the two tests testing the same parameter? Are the two test both useful to test if the drug decreasing the level of cholesterol? Explain briefly. [3]

——————— End of examination paper———————