NAME OF CANDIDATE: ....
STUDENT NUMBER: ....

UNSW SYDNEY

SCHOOL OF MATHEMATICS AND STATISTICS

Semester 2, 2018

MATH2089
Numerical Methods and Statistics

(1) TIME ALLOWED – 2 Hours

(2) TOTAL NUMBER OF QUESTIONS – 4

(3) ANSWER ALL QUESTIONS

(4) THE QUESTIONS ARE OF EQUAL VALUE

(5) THIS PAPER MAY **NOT** BE RETAINED BY THE CANDIDATE

(6) **ONLY** CALCULATORS WITH AN AFFIXED "UNSW APPROVED" STICKER MAY BE USED

(7) STATISTICAL FORMULAE ARE ATTACHED AT END OF PAPER

Part A – Numerical Methods consists of questions 1 – 2

Part B – Statistics consists of questions 3 – 4

Both parts must be answered

All answers must be written in ink. Except where they are expressly required pencils may only be used for drawing, sketching or graphical work.

Part A – Numerical Methods

1. **[20 marks]**

Answer in a separate book marked Question 1

2. **[20 marks]**

Answer in a separate book marked Question 2

Part B – Statistics

3. Answer in a separate book marked Question 3

a) Let X follow the Bernoulli distribution:

$$p(x) = \begin{cases} 1-\pi, & \text{if } x = 0 \\ \pi, & \text{if } x = 1 \end{cases}$$

where $0 < \pi < 1$.

i) [**1 mark**] Show that $\mathbb{E}(X) = \pi$.

ii) [**2 marks**] Show that $\mathbb{V}\text{ar}(X) = \pi(1-\pi)$.

iii) [**3 marks**] Assume we have a random sample $\{X_1, \dots, X_n\}$ of size n from X. The probability π can be estimated from this sample as

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Show that the standard error of $\hat{p}$ is

$$\sqrt{\frac{\pi(1-\pi)}{n}}.$$

Make sure you include your reasoning.

b) *(Matlab output relevant to this question can be found overleaf.)*
In August this year, Roy Morgan Research published a poll on Rugby viewership of New Zealanders. The poll, of 6,422 randomly selected New Zealanders, found that 43.6% of them watch Rugby on the television.

i) [**3 marks**] Find a 95% confidence interval for the true proportion of New Zealanders who watch Rugby on the television.

ii) [**3 marks**] What assumptions did you make in the above? Where possible, check if these assumptions are reasonable.

iii) [**2 marks**] Let's say Roy Morgan Research wanted to estimate the proportion of rugby viewers to within 5% of its true value (with 95% confidence). How many people did they need to sample to achieve this?

c) *(Matlab output relevant to this question can be found below.)*

Assume Rugby New Zealand (the organising body for the sport) want to be able to demonstrate that Rugby viewership is in excess of 40% of New Zealanders, using a sample of size n.

i) [**1 mark**] What are the appropriate null and alternative hypotheses for this test?

ii) [**1 mark**] What is the distribution of the sample proportion $\hat{p}$, if the null hypothesis is true? Please write your answer as a function of n.

iii) [**2 marks**] Show that, for the relevant hypothesis test at the 0.05 significance level, the rejection region for $\hat{p}$ can be expressed as

$$\left(0.4 + \frac{0.806}{\sqrt{n}},\, 1\right]$$

where n is the sample size of the study.

iv) [**2 marks**] Hence find the sample size (n) required to ensure power would be at least 80% if the true proportion were 0.45.

In answering the above questions you can make use of the following Matlab output:

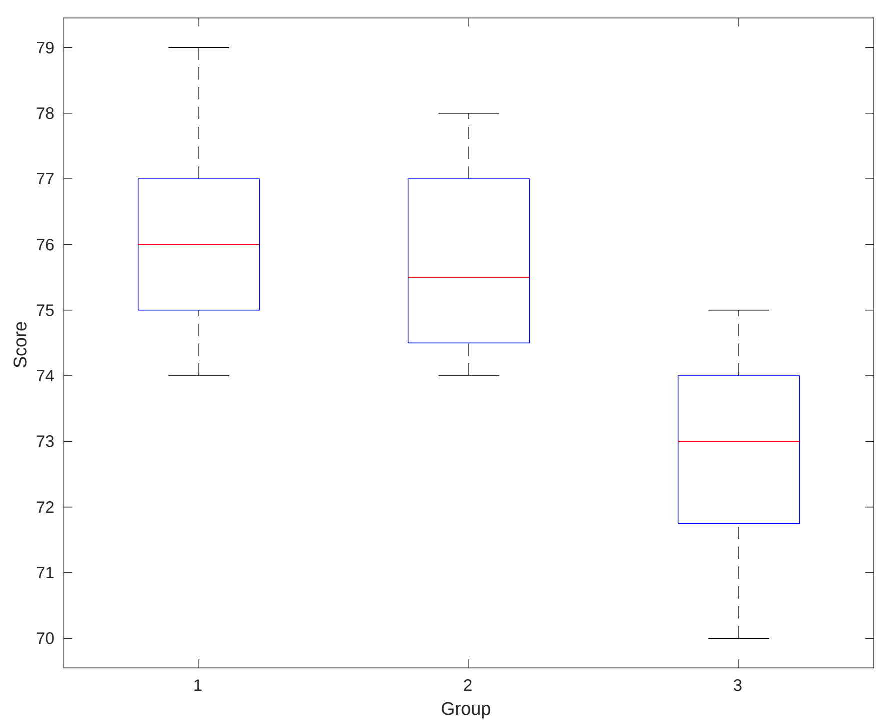
```
>> norminv([0.8 0.85 0.9 0.95 0.975])

ans =

    0.8416    1.0364    1.2816    1.6449    1.9600
```

4. Answer in a separate book marked Question 4

In order to assess the efficiency of a certain 'Memory drug', allegedly improving the capacity of students to process and memorise huge volumes of material, the following experiment was run. Students of a given class were randomly split into three groups. Before revising for a given exam, the first group was given the 'Memory drug', the second group was given a placebo drug (i.e., a candy) and the third group was given nothing at all. All the students sat the exam. The scores (out of 100) are shown below for the three different groups:

Memory	Placebo	Nothing
75	74	73
77	76	74
76	75	72
79	78	74
74	74	70
77	77	73
75	75	74
	77	71
		75
$n_1 = 7$	$n_2 = 8$	$n_3 = 9$
$\bar{x}_1 = 76.14$	$\bar{x}_2 = 75.75$	$\bar{x}_3 = 72.89$
$s_1 = 1.68$	$s_2 = 1.49$	$s_3 = 1.62$

The sample mean and the sample standard deviation of the exam scores have been computed for each group. Comparative boxplots are given in the figure below (Group 1 = 'Memory', Group 2 = 'Placebo', Group 3 = 'Nothing').



a) **[3 marks]** What do the boxplots tell you about the distribution of the exam scores within and between the three different groups of students?

b) **[3 marks]** In order to compare the exam scores across the three groups of students, we would like to carry out an Analysis of Variance (ANOVA).

   i) State three assumptions that need to be valid for ANOVA to be an appropriate analysis.

   ii) Comment on the suitability of these assumptions, where applicable.

   Assume from now on that these assumptions are valid.

c) **[3 marks]** An ANOVA table was partially constructed to summarise the data:

Source	df	SS	MS	F
Treatment	**(1)**	**(2)**	**(3)**	10.39
Error	**(4)**	**(5)**	2.536	
Total	**(6)**	105.96		

Copy the ANOVA table in your answer booklet. Complete the table by determining the missing values (1)–(6), and explain how they are obtained.

d) **[5 marks]** Now carry out the ANOVA F-test:

   i) **[1 mark]** State the appropriate null and alternative hypotheses. Define your notation properly.

   ii) **[1 mark]** What would be the distribution of the test statistic $F = MS_{\text{Tr}}/MS_{\text{Er}}$ if the null hypothesis was true?

   iii) **[1 mark]** What is the observed value f_0 of the test statistic?

iv) [**1 mark**] Give an approximation of the p-value of this test, making use of the following Matlab output:

```
>> norminv([0.90 0.95 0.975 0.99 0.999 0.9999])
ans =
        1.2816 1.6449 1.9600 2.3263 3.0902 3.7190

>> tinv([0.90 0.95 0.975 0.99 0.999 0.9999],23)
ans =
        1.3195 1.7139 2.0687 2.4999 3.4850 4.4152

>> finv([0.90 0.95 0.975 0.99 0.999 0.9999],2,21)
ans =
       2.5746 3.4668 4.4199 5.7804 9.7723 14.7430

>> chi2inv([0.90 0.95 0.975 0.99 0.999 0.9999],2)
ans =
       4.6052 5.9915 7.3778 9.2103 13.8155 18.4207
```

v) [**1 mark**] What is your conclusion in plain language? Use $\alpha = 0.05$ as the significance level.

e) We will now study the difference in mean scores between the 'Memory drug' and placebo groups.

i) [**2 marks**] Construct a 95% two-sided confidence interval for the difference between the 'true' mean exam score for students taking the 'Memory drug' and students taking the placebo.

ii) [**1 mark**] Does this confidence interval suggest that there is an effect of the 'Memory drug', as compared to a placebo? Explain.

f) [**3 marks**] Two-sample t-tests were carried out comparing all possible pairs of treatments, as summarised in the following table:

Pairwise comparison	p-value
Memory−Placebo	0.6418
Memory−Nothing	0.0018
Placebo−Nothing	0.0017

Do these results allow you to confirm the conclusion that you have made in d)-v) above? Explain.

END OF EXAMINATION