475.101/102/107/108
STATISTICS

Assignment 2, Semester 2, 2000
Due: 4pm Thursday, 17th August

Instructions for handing in: PLEASE

1. Use (standard) **A4** sized paper.
2. Number each page in the top centre and **print** your name legibly at the **top right hand corner** of each page with the surname or family name underlined.
3. Attach a white *Department of Statistics Assignment Cover Sheet* to the front of the assignment. Staple or clip all pages together in the extreme top left hand corner.
4. Fold the paper length-wise so that the printed side of the cover sheet faces out.
5. **Print** your name, paper number and assignment number on the outside of the cover sheet. **Print** your ID number on the inside of the cover sheet.
6. **Hand the assignment in** to the appropriate hand-in box in the basement of the Maths/Physics building.

Notes:

- Statistics is about summarising, analysing and communicating information. Communication is an important part of statistics. For this reason you will be expected to write answers which clearly communicate your thoughts. **The mark you receive will be based on your written English** as well as your statistical/technical work.
- Assignment 2 will be marked out of 60 marks, 54 marks for questions 1- 7 as shown below and 6 marks for communication, style and presentation. Please refer to Section B, Questions 5 to 9, in your Course Resource book for examples of how to set out your assignment answers. Your final mark will be converted to a mark out of 10 which will be recorded towards your course work.

TERM TEST

The 475.101 / 102 / 107 /108 term test will be held at 6.30-7.30pm

on

Tuesday 12th September

Question 1. [13 marks]

(a) Express $\frac{47}{64}$ **(i)** as a decimal **(ii)** as a percentage

(b) Use the appropriate tables in your Lecture Workbook to find the following when $X \sim \text{Binomial}(n = 20, p = 0.70)$

(i) $\text{pr}(X = 10)$ **(ii)** $\text{pr}(X > 12)$ **(iii)** $\text{pr}(X < 11)$ **(iv)** $\text{pr}(13 \le X \le 17)$

(c) Use the appropriate tables in your Lecture Workbook to find the following when $X \sim \text{Poisson}(\lambda = 11.5)$

(i) $\text{pr}(X \le 6)$ **(ii)** $\text{pr}(X \ge 16)$ **(iii)** $\text{pr}(X = 5)$ **(iv)** $\text{pr}(9 \le X \le 14)$

(d) An appliance store has the following probabilities for X, the number of major appliances sold on a given day.

X	0	1	2	3	4	5
$\text{pr}(X = x)$	0.150	0.333	0.350	0.090	0.067	0.010

(i) What is the probability that X is at most 2?
(ii) What is the probability that X is more than 3?
(iii) What is the probability that X is between 1 and 4 (inclusive)?
(iv) Calculate the expected value and standard deviation of X.

Notes for Questions 2 to 7:

In order to answer some (or part thereof) of the remaining questions you can choose any one of the following methods to calculate probabilities:
(i) computer software such as EXCEL or MINITAB
(ii) calculator
(iii) graphics calculator

Question 2. [7 marks]

In (a) to (c), choose the most likely distribution which best describes the probability distribution of the random variable X? Make your choice from any one of exactly Binomial; approximately Binomial; approximately Poisson; neither Binomial nor Poisson.

When you choose exactly/approximately Binomial or approximately Poisson then:
(i) critically evaluate the relevant assumptions
(ii) give the value(s) of the parameter(s) for the probability distribution.

(a) An investment company employs 52 researchers with PhD. degrees, 14 of whom have their doctorates in finance. A delegation of seven researchers is to be randomly selected to attend an international investments conference. Let X be the number of researchers with PhD. degrees in finance who are selected to attend the conference.

(b) Customer surveys for power companies in New Zealand have indicated that customer satisfaction is strongly related to a repair response rate time of no more than 2 hours from the initial call requesting service. The arrival of the repair person within 2 hours is considered an acceptable waiting period. Data collected over a number of years by the electricity industry indicate that the likelihood is 60% that a repair person will reach the customer's home within the acceptable waiting period. Let X be the number of acceptable service calls out of 13 randomly monitored service calls in a given month.

(c) Over a long period the manager of a bank located in the central business district of a city has observed customers visiting the bank during the 12:00pm to 1:00pm lunch hour. According to the manager, on average, 55 customers visit the bank during this period. Let X be the number of customers who visit the bank during the 12:00pm to 1:00pm lunch hour on any given day.

Question 3. [10 marks]

Smokeless tobacco (chewing tobacco) is considered to be one of the most addictive substances known to man and more difficult to quit than smoking. Mouth and throat cancer, mouth lesions and tooth loss are a few of the possible health risks involved with the use of smokeless tobacco.

A study of smokeless tobacco use among high school students in two Arkansas communities was conducted. Researchers classified 901 students by use of smokeless tobacco and by grade at school (Marty, McDermott and Williams, 1986).

In the study 730 of the high school students did not use smokeless tobacco, while 330 were grade 11 students. 262 grade 11 students did not use smokeless tobacco. Of those students who use smokeless tobacco 33.92% were in grade 12, while of those that did not use smokeless tobacco, 38.5% were in grade 10.

(a) Construct a 3×2 table displaying the above results as proportions. Complete the table, giving all calculated proportions to four decimal places (e.g. 0.1234).

(b) From each of the three grades, what percentage of randomly selected students use smokeless tobacco? Show your working.

(c) From the results above (including those shown in the table), what do the observations in this study suggest about smokeless tobacco use among high school students in the two Arkansas communities? Briefly justify your answer.

(d) Is it valid to rely on any conclusions from this study when discussing smokeless tobacco use among high school students in Arkansas? Briefly justify your answer.

Question 4. [7 marks]

The U.S. Bureau of the Census recently published statistics on educational attainment of the non-institutional population of the United States, based on the March 1998 Current Population Survey. 172,214 people were surveyed and classified by age group and highest educational qualification attained. The following table summarises the results of the survey.

	Age					
Level of Education	25 – 34	35 – 44	45 – 54	55 – 64	Over 64	Total
Did not complete high school	4,754	5,326	4341	4,558	10,580	29,559
Completed high school	12,569	15,136	10,943	8,311	11,215	58,174
Attended university for between 1 and 3 years	19,587	20,450	14,921	7,379	8,478	70,815
Attended university for 4 or more years	2,444	3,548	3,854	2,007	1,813	13,666
Total	39,354	44,460	34,059	22,255	32,086	172,214

(a) Which age category accounted for the:
(i) lowest number of Americans who did not complete high school?
(ii) the highest percentage of Americans who attended university for at least one year. Show your working.

(b) What percentage of Americans in this survey:
(i) did not attend university?
(ii) who only completed high school or were aged between 35 and 44 years?

(c) Given that a randomly chosen American in this survey had only completed high school, what is the probability that he or she is at most 54 years of age?

(d) Among Americans in this survey aged between 25 and 34 years what proportion did not attend university?

Question 5. [6 marks]

(a) A student studies the Poisson distribution using computer-assisted instruction. The computer then presents 10 problems which are of differing degrees of difficulty. X is the number of problems that the student solves correctly. Briefly discuss how the Binomial distribution assumption of "constant probability of getting a success" is violated.

(b) An archer in an archery competition has 3 shots aiming to hit the bulls-eye (i.e. the centre of the target). Over many years of competing the archer has determined that she hits the bulls-eye 62% of the time. The archer uses the result of the previous shot to try and improve the result of the next shot. X is the number of times the archer hits the bulls-eye in the three shots. Briefly discuss how the Binomial distribution assumption of "independence" is violated.

(c) Buses on a busy Auckland route are regularly timetabled to arrive on the half hour. Passengers arrive at the bus stop at a rate of 3.4 every 10 minutes. Let X be the number of people who arrive at the bus stop in the half hour prior to the bus arriving.

(i) Briefly discuss how the Poisson distribution assumptions "events occur at a constant average rate per unit of time" and "independent occurrences" are violated.

(ii) Describe the following events as any of the following: likely, less likely or extremely unlikely:
- more than one person arrives at the bus stop in a 10 minute interval.
- more than one person arrives at the bus stop in a 2 minute interval.
- more than one person arrives at the bus stop in a 30 second interval.

In relation to the Poisson distribution assumptions what do the events in part (ii) suggest? Briefly explain.

Question 6. [6 marks]

This question relates to the description given in **question 5 part(c)**. Assume that the Poisson distribution is an appropriate model for X.

(a) State the value of the parameter(s) of this distribution.

(b) Find the probability that:

(i) at most 6 people arrive at the bus stop in the half hour prior to the bus arriving.

(ii) no one arrives at the bus stop in the half hour prior to the bus arriving.

(iii) between 11 and 19 people (inclusive) arrive at the bus stop in the half hour prior to the bus arriving.

(c) **(i)** How many people would you expect to arrive at the bus stop in the half hour prior to the bus arriving.

(ii) What is the standard deviation of X?

(d) Under this model, the number of people arriving at the bus stop in the half hour prior to the bus arriving is very likely to be no further than three standard deviations away from the mean. Within what range is the number of people arriving at the bus stop in the half hour prior to the bus arriving very likely to be?

Question 7. [5 marks]

This question relates to the description given in **question 5 part(b)**. Assume that the Binomial distribution is an appropriate model for X,

(a) State the value of the parameter(s) of this distribution.

(b) Find the probability that:

(i) all 3 shots hit the bulls-eye.

(ii) at least 2 shots hit the bulls-eye.

(iii) more than 0 but at most 2 shots hit the bulls-eye.

(c) How many shots would you expect to hit the bulls-eye? What is the standard deviation of X?