

475.101/102/107/108
STATISTICS

Assignment 2, Semester 1, 2000
Due: 4pm Thursday, 30th March

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- 6 as shown below and 6 marks for communication, style and presentation. Please refer to Section C, Questions 5 to 9, in your Lecture Workbook 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

Friday 14th April

Question 1. [13 marks]

- (a) Express $\frac{156}{624}$ (i) as a decimal (ii) as a percentage
- (b) Use the appropriate tables in your Lecture Workbook to find the following when $X \sim \text{Binomial}(n = 11, p = 0.40)$
- (i) $\text{pr}(X = 4)$ (ii) $\text{pr}(X \geq 6)$ (iii) $\text{pr}(X < 8)$ (iv) $\text{pr}(4 \leq X \leq 7)$
- (c) Use the appropriate tables in your Lecture Workbook to find the following when $X \sim \text{Poisson}(\lambda = 5.8)$
- (i) $\text{pr}(X \leq 4)$ (ii) $\text{pr}(X > 2)$ (iii) $\text{pr}(X = 11)$ (iv) $\text{pr}(8 \leq X \leq 12)$
- (d) Suppose X the number of races required to win the **next** America's Cup has a probability function given by the following table but **one** of the given probabilities is in error:

X	5	6	7	8	9
$\text{pr}(X = x)$	1.00	0.20	0.37	0.24	0.09

- (i) Which one of the probabilities is in error?

Give the correct value and use the correct value in answering the following questions.

- (ii) What is the probability that X is at least 7?
- (iii) What is the probability that X is no more than 8?
- (iv) Calculate the expected value and standard deviation of X .

Notes for Questions 2 to 6:

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) appropriate tables
- (iii) calculator
- (iv) graphics calculator

Question 2. [7 marks]

In (a) to (c), which one of the following statements best describes the probability distribution of the random variable X ? Exactly Binomial; approximately Binomial; exactly Poisson; approximately Poisson; neither Binomial nor Poisson.

When you choose exactly/approximately Binomial or Poisson then:

- (i) briefly justify your choice
- (ii) give the value(s) of the parameter(s) for the probability distribution.
- (a) A long distance truck driver drives logging trucks on the Napier to Taupo highway. Over many journeys he has observed numerous dead possums. According to his calculations, 4 dead possums occur on average every 5 kilometres. Let X be the number of dead possums found in a given kilometre of the Napier to Taupo highway.
- (b) Over the 9 years of his life an aged possum has observed many logging trucks driving the Napier to Taupo highway. The possum has noticed that during the day, on average a stream of 9 vehicles builds up behind each truck just prior to a passing lane. Let X be the number of vehicles in this stream of traffic. (Note: the possum is an insomniac and during the day attempts to fall asleep by sitting by the road counting cars).
- (c) A recent survey of all American hotels found that 80% offer complimentary shampoo in their guest rooms. Let X be the number of American hotels not offering complimentary shampoo in a random sample of 200 hotels.

Question 3. [8 marks]

A study of New Zealanders was conducted in order to determine peoples preferences for various brands of soft drink. 190 participants in the study reported that they drank either Pepsi or Coke.

Those who preferred Pepsi comprised 39.47% of these 190 participants. Of those who preferred Coke, 52.17% drank regular drinks, while those who preferred Pepsi, 42.67% drank diet drinks.

- Construct a 2x2 table displaying the above results as proportions. Complete the table, giving all calculated proportions to four decimal places (eg. 0.1234).
- If a person who drinks diet soft drinks is randomly selected what is the probability that he or she prefers Coke?
- What percentage of regular drinkers preferred Pepsi?
- From the results above, what do the observations in this study suggest about the brand preferences of Pepsi and Coke drinkers? Briefly justify your answer.

Question 4. [7 marks]

To better target its market, Alfa Romeo conducted a market research study. A random sample of 669 respondents was chosen, each of whom was asked to select one of four qualities that he or she believed best described him or her as a driver. The respondents were also asked to choose one of three Alfa Romeo models as her or his choice of the most suitable car. The following table summarises the respondents answers.

Alfa Romeo Model	Self-Image				Total
	Defensive	Aggressive	Enjoying	Prestigious	
Alfasud	22	21	34	56	133
Giulia	39	45	42	68	194
Spider	77	89	96	80	342
Total	138	155	172	204	669

- Which Self-Image group accounted for the:
 - largest number of respondents overall?
 - lowest number of respondents among those who said that the Giulia model is their model of choice?
- What percentage of respondents who describe themselves as prestigious drivers, selected Alphasud or the Spider as his or her model of choice?
- What proportion of respondents in the survey:
 - said the Spider was his or her model of choice and described him or herself as an aggressive driver?
 - selected either the Alphasud or the Giulia model as their model of choice?
- Among respondents who chose the Giulia or Spider as his or her model of choice what proportion describe him or herself as either an "enjoying" or "prestigious" type of driver?
- Given that a respondent describes him or herself as a defensive driver, what is the probability that the Alphasud is his or her model of choice?

Question 5. [9 marks]

According to the data base of a large multinational airline company, 35% of all its 475 pilots are over 40 years of age. The company is about to purchase the latest model Boeing and is planning to select a random sample of 25 pilots to receive training in flying this new plane.

Let X be the number of pilots over 40 years of age in this sample.

- Discuss the validity of using the Binomial distribution in this situation. (Hint: consider the conditions for this distribution and consider how well these conditions are satisfied here).
- State the value of the parameter(s) of this distribution.
- Assuming that the Binomial distribution you have described above is an appropriate model for X , Find the probability that:
 - more than 7 of the pilots selected are over 40 years of age.
 - exactly 5 of the pilots selected are over 40 years of age.
 - between 13 and 18 (inclusive) of the pilots selected are over 40 years of age.
- How many of the pilots selected would you expect to be over 40 years of age? What is the standard deviation of X ?
- Why would the airline company be interested in the variable “pilots over 40 years of age”? Suggest another variable the company may be interested in measuring? Briefly justify your answer.

Question 6. [10 marks]

The number of “hits” on (visits to) the stage I statistics web page is, on average, 7.5 per day

Let X be the number of “hits” on the stage I web page on a particular day.

- Discuss the validity of using the Poisson distribution in this situation. (Hint: consider the conditions for this distribution and consider how well these conditions are satisfied here).
- State the value of the parameter(s) of this distribution.
- Assuming that the Poisson distribution you have described above is an appropriate model for X , find the probability that there are at most 9 “hits” on a particular day.

Let Y be the number of “hits” on the stage I web page in a two-day period.

- Assuming that the Poisson distribution is an appropriate model for Y , find the probability that there are:
 - exactly 12 “hits” on the stage I web page in a two-day period.
 - between 11 and 19 (inclusive) “hits” on the stage I web page in a two-day period.
- How many “hits” would you expect to find in a randomly chosen two-day period? What is the standard deviation of Y ?
- Under this model, the number of “hits” on the stage I web page in a two-day period is very likely to be no further than three standard deviations away from the mean. Within what range is the number of “hits” on the stage I web page in a two-day period very likely to be?