475.101/102/107/108
STATISTICS

Assignment 4, Semester 1, 2000
Due: 4pm Thursday, 11th May

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 4 will be marked out of 60 marks, 54 marks for questions 1 - 4 as shown below and 6 marks for communication, style and presentation. Please refer to Section C, Questions 14 to 16, in your Lecture Workbook / 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.

Question 1. [13 marks]

A university lecturer, lecturing in "Sport and Society" was interested in examining the extent of approval of unsporting play in New Zealand secondary schools. Twenty-six male and thirty female randomly chosen New Zealand secondary school hockey players completed a questionnaire that was used to arrive at an approval score, with higher scores indicating greater approval rating for unsporting play. The results are given in below:

Male:	2.90	2.55	2.75	2.31	2.80	3.09	3.31	2.34	3.01	2.67	2.46	2.45	3.33	2.38	3.22
	2.19	1.99	3.03	2.43	2.59	2.28	3.00	2.64	2.49	2.99	3.00				
Female:	2.52	1.90	2.36	1.87	1.78	1.57	2.41	2.63	2.10	1.69	2.17	2.44	1.54	1.91	1.68
	1.61	1.94	2.03	2.30	2.24	2.49	1.39	2.04	1.68	0.97	2.51	2.56	2.01	1.62	1.73

(a) Write down the method you would use to choose these New Zealand secondary school hockey players if you had been conducting the study. Briefly describe.

(b) What are the mean approval score and standard deviation for each group of hockey players?

(c) **By hand**, calculate a 95% confidence interval for the difference in the mean approval score between male and female hockey players. Interpret your result using plain English in 1-2 sentences.

(d) Does your confidence interval from part (c) contain the true difference? Explain your answer briefly using plain English in 2-3 sentences.

Question 2. [19 marks]

(a) In each of the following cases indicate whether a hypothesis test or a confidence interval is the more appropriate test to initially carry out. Briefly explain.

(i) A lecturer in the statistics department receives junk e-mail. These are unsolicited e-mails usually offering (unbelievable) deals. Records of how many e-mails were received were kept for a sample of weeks preceding the February 1998 power crisis in central Auckland and for a sample of weeks after the power crisis. During the crisis the university e-mail system was out of action for a week and e-mail messages sent to the lecturer at this time were sent back with a message that they were undeliverable. You need to determine if there is any evidence of a difference in the average number of junk e-mails before and after the power crisis.

(ii) The Manager of a large metropolitan city's planning department is interested in examining the city's air quality. Measurements of the nitrogen oxide levels were taken over a number of weeks. You are asked to estimate the true mean nitrogen oxide level.

(b) For each of the following *P-values* how would you describe the strength of evidence against the null hypothesis:

(i) 0.0465 **(ii)** 0.1253 **(iii)** 0.0689 **(iv)** 0.465

(c) Using your tables for the Student's *t*-distribution obtain upper and lower values that a *P-value* can take in the following cases:

(i) $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ where $t_0 = 1.709$ and $df = 27$

(ii) $H_0 : p_1 = p_2$ vs $H_1 : p_1 > p_2$ where $t_0 = 3.433$ and $df = 60$

(d) In each of the following explain why the hypotheses are not valid. Rewrite each hypothesis correctly.

(i) $H_0 : \mu_1 - \mu_2 < 0$ v $H_1 : \mu_1 - \mu_2 > 0$

(ii) $H_0 : \hat{p}_1 - \hat{p}_2 > 0$ v $H_0 : \hat{p}_1 - \hat{p}_2 = 0$

(e) For each of the following statements, state whether they are true or false. If they are false, rewrite correctly.

(i) For a great many situations, an (approximate) confidence interval is given by *estimate* ± *t standard errors*. The size of the multiplier, *t*, depends on either the desired confidence level, or the degrees of freedom (*df*).

(ii) It is important never to quote a *P-value* about the existence of an effect without also supplying a confidence interval for the likely size of the effect.

(iii) A value of t_0 equalling 2.15 implies that the estimated value is 2.15 standard deviations above the hypothesised value.

(iv) A two-sided *t*-test will reject (at the 5% level) values for the null hypothesis that are outside the corresponding 95% confidence interval.

(v) It makes no sense to try to interpret a *P-value* without explicitly stating the relevant hypotheses.

(vi) Differences or effects seen in data that are easily explainable in terms of sampling variation provide convincing evidence that real differences or effects exist.

(vii) If H_0 is rejected at the 5% level, it will also be rejected at the 1% level.

NOTES:

- If you are asked to carry out a *t*-test, you will need to:
  1. state the parameter in words.
  2. state the null and alternative hypotheses.
  3. when instructed use either *Excel*, *MINITAB* or a graphics calculator.
     - Make sure you are prepared for both questions 3 and 4 before you begin to use the computer.
     - ONLY hand in computer output if you are asked to.
     - When using the computer report *P-values* to 4 decimal places.

Question 3. [7 marks]

This question refers to the details in Question 1.

(a) Is there a significant difference in the average approval score between male and female hockey players? **By hand** carry out a *t*-test to investigate this. Interpret your result using plain English in 1-2 sentences.

(b) Write a **brief** report (of approximately 100 to 150 words) to the university lecturer on the extent of approval of unsporting play using your conclusions from **question 3(a)** and **question 1(b**). Include recommendation(s) about what the university lecturer should do next.

Question 4. [15 marks]

The National Centre on Addiction and Substance Abuse (CASA) at Columbia University regularly surveys attitudes of teens and those who most influence them. As part of the 1999 CASA "National Survey of American Attitudes on Substance Abuse V: Teens and Their parents", 2000 teenagers aged between 12 and 17 (1000 boys and 1000 girls), 536 mothers and 464 fathers were interviewed by telephone on many lifestyle issues. The data in the table below is given for three questions answered by the teenagers in the survey.

Have you ever had a serious discussion with both your parents about the risks of using illegal drugs? [IF NO] *Was it with your Mum only or Dad only?* [IF NO] *Would you like to have such a discussion?*

	Total	**Male**	**Female**
Yes, both parents	890	500	390
Yes, Mum only	290	120	170
Yes, Dad only	70	40	30
No, but would have liked such a discussion	160	70	90
No, and would not have wanted such a discussion	560	260	300
Don't know	30	10	20

Would you describe your relationship with your mother as excellent, very good, good, fair, or poor?

	Total	**Male**	**Female**
Excellent	770	380	390
Very Good	680	350	330
Good	350	190	160
Fair	160	70	90
Poor	40	10	30

Would you describe your relationship with your father as excellent, very good, good, fair, or poor?

	Total	**Male**	**Female**
Excellent	590	330	260
Very Good	580	310	270
Good	460	210	250
Fair	210	90	120
Poor	130	50	80
Don't Know	30	10	20

(a) Identify the sampling situation as (a) *Two independent samples*, (b) *Single sample, several response categories* or (c) *Single sample, two or more Yes/No items* in the following cases:

(i) We want to compare the proportion of teenagers who have had serious discussions about the risks of using illegal drugs with their mum only and the proportion of teenagers who have had serious discussions about the risks of using illegal drugs with their dad only.

(ii) We are interested in comparing the proportion of teenagers who describe their relationship with their mother as excellent and the proportion of teenagers who describe their relationship with their father as excellent.

(iii) We wish to compare the proportion of male teenagers who have had serious discussions about the risks of using illegal drugs with both parents and the corresponding proportion of female teenagers.

(b) Calculate the standard errors for the cases described in parts **(i) (ii)** and **(iii)** in **(a)** above.

(c) Is there a significant difference in the proportion of male teenagers who have had serious discussions about the risks of using illegal drugs with both parents and the corresponding proportion of female teenagers? Using either *Excel*, *MINITAB* or a graphics calculator, investigate this question. Interpret your results.

Notes for question 5(c):
- If you wish to use *Excel* for this question refer to the *Lecture Workbook Chapters 8 –12*, Chapter 9/5.
- If you use either *Excel* or *MINITAB* include the computer output.

(d) Below in figure 1 is the *Excel* computer output for the hypothesis test and 95% confidence interval for the case described in part **(ii)** above. State the hypotheses and interpret the results.

Fig. 1 Question 4(a) part (ii)

	A	B
1	Test of No Difference Between	
2	Two Population Proportions	
3		
4	Input data	
5	p1_ratio	*****
6	p2_ratio	*****
7	p_diff	0.09
8	X1_sample	770
9	X2_sample	590
10	n_total	2000
11		
12		
13	Alpha	0.05
14	Calcualated Value	
15	se	*****
16	Zo	4.910267891
17		
18	Two-Tail Test	
19	pvalue	9.10837E-07
20		
21	Confidence Intervals (95%)	
22	Lower Limit	Upper Limit
23	0.05407528	0.12592472