Department of Statistics

Stage I Statistics
475.101/102/107/108
Term Test : Semester I 2000

VERSION 1

Instructions:

All questions have a single correct answer. Multiple answers to a question will ALL be marked wrong. Incorrect answers are not penalised. If you do not know the answer, then take a guess. All questions carry the same mark value.

There are 26 questions.

Formulae are on page 15.

Answer ALL questions on the ANSWER SHEET provided (appended to the front of the test paper).

- Hand in your answer sheet.
- Keep a personal record of your answers on the test paper - answers will be posted next week.

Questions 1 to **7** refer to the following information.

In 1998 the New Zealand House of Parliament sat for 92 days of the year. The attendance or absence of Members of Parliament (MPs) is recorded for each sitting day. Table 1 shows data for the 44 National party MPs in 1998. For each MP, measurements were made on the following variables:

MP: The name of the MP.

Type: The type of seat held by the MP, either electorate (Elect) or list (List).

Gender: The gender of the MP.

Abs: The number of sitting days the MP was absent.

t

No.	MP	Type	Gender	Abs	No.	MP	Type	Gender	Abs
01	Anae	List	M	13	23	McCully	Elect	M	33
02	Ardern	Elect	M	9	24	McKinnon	List	M	35
03	Banks	Elect	M	26	25	McLauchlan	List	F	3
04	Birch	Elect	M	6	26	McLean	Elect	M	17
05	Bradford	Elect	M	20	27	Neeson	Elect	M	16
06	Brownlee	Elect	M	10	28	O'Regan	List	F	12
07	Carter D.	Elect	M	3	29	Revell	Elect	M	18
08	Carter J.	Elect	M	14	30	Roy	List	M	8
09	Creech	Elect	M	2	31	Ryall	Elect	M	7
10	East	Elect	M	25	32	Shipley	Elect	F	38
11	English	Elect	M	20	33	Simcock	Elect	M	8
12	Fletcher	Elect	F	30	34	Simich	Elect	M	4
13	Graham	List	M	5	35	Smith L.	Elect	M	49
14	Gresham	List	M	7	36	Smith N.	Elect	M	10
15	Hasler	Elect	F	9	37	Sowry	List	M	16
16	Herlihy	Elect	M	14	38	Steel	Elect	M	2
17	Kidd	Elect	M	4	39	te Heu Heu	List	F	6
18	Kyd	Elect	M	12	40	Upton	List	M	35
19	Luxton	Elect	M	10	41	Vernon	Elect	F	4
20	Mapp	Elect	M	7	42	Williamson	Elect	M	13
21	Marshall	Elect	M	27	43	Wong	List	F	7
22	Maxwell	List	M	19	44	Young	List	F	4

Table 1: National MPs, 1998

1. **Gender** is classified as:

(1) a continuous variable.

(2) a response variable.

(3) a discrete variable.

(4) a quantitative variable.

(5) a qualitative variable.

2. We wish to randomly select and interview five of the MPs listed in Table 1 about Members' absences from the House of Parliament. The method for randomly selecting the five MPs is carried out by using the number (in the **No.** column of Table 1) associated with each MP, and random number digits. Use the row of random digits below to select a sample of five MPs. You must start at the beginning of the row and use consecutive digits.

38683 50279 38224 09844 13578

The five MPs randomly selected are:

(1)	Steel	Smith L.	Neeson	O'Regan	Carter J.
(2)	Steel	Vernon	Smith L.	Ardern	East
(3)	Steel	Smith L.	Ardern	Steel	Maxwell
(4)	Steel	Smith L.	Ardern	Maxwell	Upton
(5)	Steel	Vernon	Smith L.	Ardern	Maxwell

3. This type of sampling is called:

(1) systematic random sampling.

(2) stratified random sampling.

(3) simple random sampling.

(4) cluster random sampling.

(5) convenience random sampling.

4. Suppose the purpose of the table is to convey information so that the reader can make visual comparisons and see trends in the National party MPs' absences quickly and easily. The **best** improvement in the presentation of data in the table would be to:

(1) use code 0 for Elect and 1 for List for **Type**.

(2) round the measurements made on **Abs** to the nearest ten.

(3) order the table with the electorate seat MPs in alphabetical order followed by the list seat MPs in alphabetical order.

(4) order the table with the female MPs in alphabetical order followed by the male MPs in alphabetical order.

(5) order the table based on **Abs**, from highest to lowest (or vice versa).

5. The most appropriate way to begin to explore the relationship between **Type** and **Gender** is a:

(1) two-way table of counts of these two variables.

(2) dot plot for each of these two variables drawn on the same scale.

(3) back-to-back stem-and-leaf plot of these two variables.

(4) histogram for each of these two variables drawn on the same scale.

(5) scatter plot of these two variables.

Questions 6 and **7** refer to the following additional information.

Figure 1 shows a stem-and-leaf plot for **Abs** for the National party MPs in 1998.

Units: 2 | 5 = 25 absent days
$n = 44$

```
0| 2 2 3 3 4 4 4 4
0| 5 6 6 7 7 7 7 8 8 9 9
1| 0 0 0 2 2 3 3 4 4
1| 6 6 7 8 9
2| 0 0
2| 5 6 7
3| 0 3
3| 5 5 8
4|
4| 9
```

Figure 1: Number of sitting days absent for National MPs, 1998

6. The five-number summary for the **Abs** data set displayed in Figure 1 is:

(1) (2, 6.5, 10, 19.5, 49)

(2) (2, 7, 10, 20, 49)

(3) (2, 6, 11, 19, 49)

(4) (2, 7, 11, 20, 49)

(5) (2, 6.5, 11, 19.5, 49)

7. Which **one** of the following statements is **false** when referring to the data in Figure 1?

(1) On a box plot of the data, the observation 49 absent days would be an outside value.

(2) On a box plot of the data, the lower whisker would end at 2.

(3) This set of data is negatively or left skewed.

(4) The class interval used in the plot is 5 absent days.

(5) 25% of National party MPs are absent on 20 or more sitting days.

8. The number of sitting days absent from the House of Parliament in 1998 for all of the 121 MPs is displayed in Table 2.

Number of absences	**Frequency**
0–4	18
5–9	29
10–14	26
15–19	14
20–24	13
25–29	10
30–34	5
35–39	3
40–44	1
45–49	2
Total	121

Table 2: Number of sitting days absent for all MPs, 1998

These 121 observations form a sample of the number of sitting days absent for MPs in a year. The estimated mean and standard deviation of this sample, respectively, are:

(1) $\overline{x} = 14.48$ days, $s = 10.37$ days

(2) $\overline{x} = 14.48$ days, $s = 10.33$ days

(3) $\overline{x} = 14.98$ days, $s = 10.33$ days

(4) $\overline{x} = 12.00$ days, $s = 14.00$ days

(5) $\overline{x} = 14.98$ days, $s = 10.37$ days

Questions 9 to **11** refer to the following information.

Table 3 shows data for all of the 121 MPs in 1998. Each MP has been cross-classified according to the party they represented at the end of 1998 and their gender.

	Party							
Gender	Act	Alliance	Indep	Labour	National	NZ First	United	**Total**
Female	3	6	3	14	9	2	0	37
Male	5	6	7	24	35	6	1	84
Total	8	12	10	38	44	8	1	121

Table 3: Gender of MPs by party, 1998

9. The percentage of the 1998 MPs who were female is approximately:

(1) 30.6%
(2) 11.6%
(3) 57.8%
(4) 14.0%
(5) 37.0%

10. The percentage of the 1998 Labour party MPs who were female is approximately:

(1) 50.4%
(2) 36.8%
(3) 37.8%
(4) 14.0%
(5) 11.6%

11. The proportion of the 1998 MPs who were male National party MPs is approximately:

(1) 0.795
(2) 0.547
(3) 0.603
(4) 0.350
(5) 0.289

12. Recently 55 New Zealanders volunteered to join 6,500 others in an international study searching for a vaccine against the herpes virus. An article describing the study appeared in the *New Zealand Herald*. It stated:

> "Two out of three New Zealand volunteers would receive the vaccine and the others would get a placebo."

From the information given in the quote, this study could be **best** described as:

(1) a double-blind experiment.

(2) a controlled, blind experiment.

(3) a controlled experiment.

(4) a controlled, blind, observational study.

(5) a randomised, double-blind, observational study.

13. Transcendental meditation (TM) is a method of inducing deep relaxation which is supposed to be practised twice a day for 20 minutes. A five-year study looked at health insurance records of people who had various medical conditions. The study showed dramatically reduced hospital admission rates for practitioners of TM compared with non-practitioners. The results from this study alone should not be used to argue the case that increasing the number of TM practitioners will result in fewer people being admitted to hospitals with the listed problems **mainly** because the designers of this study:

(1) ignored the fact that many external factors would have changed over the five-year period.

(2) could not be sure that people who were practising TM were using the correct technique.

(3) included more than just one medical condition.

(4) did not decide who would practise TM.

(5) could not be sure that the records they studied were reliable.

14. The *Consumer* magazine conducted a poll in February last year in an attempt to measure how people felt about the quality of the service they were receiving from their local councils. In January and February 1999, *Consumer* sent out questionnaires to 12,000 of its subscribers and received 6,813 usable responses. The summary of attitudes expressed in this survey (*Consumer* July 1999) may not necessarily represent those of all New Zealanders **mainly** because of possible:

(1) selection and nonresponse biases.

(2) sampling errors.

(3) self-selection bias and question effects.

(4) random errors.

(5) transferring findings and interviewer effect.

15. An Auckland concrete manufacturer wanted to determine the way and extent that the hardness of a batch of concrete depends on the amount of cement used in making it. Forty batches of concrete were made up with varying amounts of cement in the mix. The variable **Cement** represented the amount of cement (in grams) used to make a batch of concrete, and the variable **Hardness** represented the hardness of the concrete (in kg/cm^2) after seven days. To investigate the relationship between **Hardness** and **Cement** the scatter plot shown in Figure 2 was constructed.

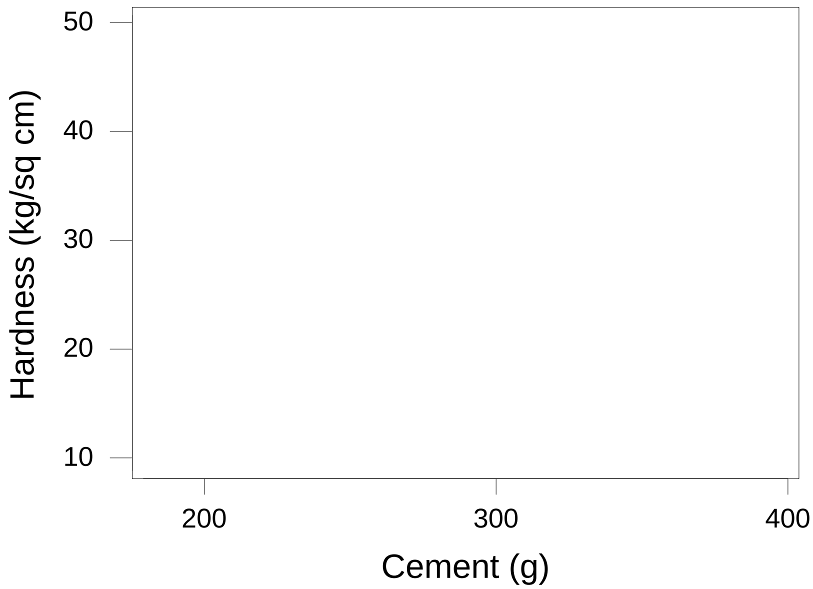

Figure 2: Scatter plot of Hardness versus Cement

Which **one** of the following statements is **false**?

(1) One of the reasons why it is appropriate to use regression to model the relationship between **Hardness** and **Cement** is that **Cement** is quantitative and **Hardness** is continuous and random.

(2) There is a positive association between **Hardness** and **Cement**.

(3) It is not appropriate to use this data to predict the hardness of a batch of concrete made by using 100 grams of cement.

(4) **Hardness** is the explanatory variable and **Cement** is the response variable.

(5) The scatter plot shows there is a linear trend between **Hardness** and **Cement**.

16. In 1998 the *Journal of the American Medical Association* reported a French study about a gene mutation which slows the progress of Aids in many HIV-infected newborns. The researchers studied 276 white children who had contracted the HIV virus from their mothers. Of these, 49 children had inherited the mutated gene from one of their parents. At the age of three, 9% of the children with the gene mutation had developed symptoms of Aids, while 28% of the children without the gene mutation had developed symptoms of Aids.

Of those children in this study who had developed symptoms of Aids, the proportion who had inherited the mutated gene from one of their parents is approximately:

(1) 0.09
(2) 0.28
(3) 0.06
(4) 0.40
(5) 0.24

17. Suppose x is an observation from a random variable X with mean $\mu_X = 100$ and standard deviation $\sigma_X = 15$. If a particular observation, x, has a z-*score* $= -2$, then the value of x is:

(1) 30
(2) 130
(3) -30
(4) 70
(5) 85

Questions 18 to **20** refer to the following information:

Competitive cyclists are thought to have a very low percentage of body fat due to intensive training methods. A study of competitive cyclists (*International Journal of Sport Nutrition* June 1995) found them to have a mean percent body fat of 9% with a standard deviation of 3%. Assume that the distribution of percent body fat is well modelled by a Normal distribution and that the mean and standard deviation found in this study apply to all competitive cyclists.

To answer Question 18 and Question 19, you will need some of the following output from MINITAB.

```
Normal with mean = 9.0000 and standard deviation = 3.0000

 x  P(X ≤ x)               P(X ≤ x)      x
 7    0.2525                  0.1       5.1153
 8    0.3694                  0.2       6.4571
                              0.8      11.5249
                              0.9      12.8447
```

18. The proportion of competitive cyclists who have a percent body fat of greater than 8% is approximately:

(1) 0.37
(2) 0.25
(3) 0.87
(4) 0.75
(5) 0.63

19. The interval encompassing the central 80% of competitive cyclists' percent body fat is approximately:

(1) from 5% to 12%.
(2) from 6% to 12%.
(3) from 5% to 13%.
(4) from 12%, upwards.
(5) from 6% to 13%.

20. Suppose the distribution of percent body fat for competitive **swimmers** is well modelled by a Normal distribution with a mean of 10% and a standard deviation of 4%. (Recall that the distribution of percent body fat for competitive **cyclists** is well modelled by a Normal distribution with a mean of 9% and a standard deviation of 3%.)

Which **one** of the following statements is **false**?

(1) It is less likely that a randomly selected competitive swimmer would have percent body fat of less than 12% than a randomly selected competitive cyclist would have percent body fat less than 11%.

(2) A larger proportion of competitive cyclists than competitive swimmers have percent body fat less than 6%.

(3) About 68% of competitive cyclists have percent body fat between 6% and 12%.

(4) Over 99% of competitive swimmers have percent body fat less than 22%.

(5) On average, competitive swimmers have a higher percent body fat than competitive cyclists.

21. A large US book warehouse sends books out to its subsidiary warehouses throughout the country. The books are sent out in batches of 50. From data collected over the years, it is known that the width, X, of all of its different titled books has mean $\mu_X = 18$mm and standard deviation $\sigma_X = 8.6$mm.

Let the random variable W represent the total width of 50 books. If 50 randomly selected books (all with different titles) are packed, then the **total** width W, of the books has a mean, μ_W, and a standard deviation, σ_W, given by:

(1) $\mu_W = 900$mm, $\sigma_W = 430$mm

(2) $\mu_W = 900$mm, $\sigma_W = 60.8$mm

(3) $\mu_W = 18$mm, $\sigma_W = 8.6$mm

(4) $\mu_W = 900$mm, $\sigma_W = 340$mm

(5) $\mu_W = 18$mm, $\sigma_W = 1.2$mm

Questions 22 and **23** refer to the following information.

An article in the *New Zealand Herald* (28 January 1998) stated:

> "Fifty-four people have drowned in the Waikato River since 1980, an average of three a year."

The same article also stated:

> "A fourth person has drowned in Waikato River waters within a week."

Let random variable, D, represent the number of people drowned in the Waikato River in a given **year**.

22. Which **one** of the following statements is **false**?

(1) If approximately 20% of the drownings in the Waikato River were incidents in which more than one person drowned, then the distribution of D should not be modelled by a Poisson distribution.

(2) If it is not appropriate to model the distribution of D with a Poisson distribution or a Binomial distribution, then it is not necessarily appropriate to model D with a Normal distribution.

(3) If drownings in the Waikato River tend to occur a lot more often in summer months than in winter months, then the distribution of D should not be modelled by a Poisson distribution.

(4) If drownings occur at about the same rate throughout the year, and if drownings are independent of each other, and if drownings rarely occur at the same time, then the distribution of D is approximately Poisson.

(5) Since there are a fixed number of trials (54), each trial has two outcomes (a person either drowns or does not), and river conditions are reasonably constant for each drowning, then the distribution of D is approximately Binomial.

23. Suppose it is appropriate to model D with a Poisson distribution with parameter, $\lambda = 3$. (**Note**: This may not be the case.) The probability that exactly 4 people will drown in any **one-week** period can be found by evaluating:

(1) $\mathrm{pr}(X = 4) \times 52$ where $X \sim \mathrm{Poisson}(\lambda = \frac{3}{52})$

(2) $\mathrm{pr}(D = 4 \times 52) \div 52$ where $D \sim \mathrm{Poisson}(\lambda = 3)$

(3) $\mathrm{pr}(X = 4)$ where $X \sim \mathrm{Poisson}(\lambda = \frac{3}{52})$

(4) $\mathrm{pr}(D = 4) \div 52$ where $D \sim \mathrm{Poisson}(\lambda = 3)$

(5) $\mathrm{pr}(D = 4 \times 52)$ where $D \sim \mathrm{Poisson}(\lambda = 3)$

24. U and V are two independent random variables. U has a mean of 20 and a standard deviation of 5. V has a mean of 15 and a standard deviation of 7. Suppose $Y = 9U - 6V$. The expected value of Y, $\text{E}(Y)$, and the standard deviation of Y, $\text{sd}(Y)$, are given by:

(1) $\text{E}(Y) = 270$, $\text{sd}(Y) = 87$
(2) $\text{E}(Y) = 270$, $\text{sd}(Y) = 3$
(3) $\text{E}(Y) = 90$, $\text{sd}(Y) = 61.6$
(4) $\text{E}(Y) = 90$, $\text{sd}(Y) = 3$
(5) $\text{E}(Y) = 90$, $\text{sd}(Y) = 57.3$

25. A random variable X is well modelled by a Binomial($n = 100$, $p = 0.6$) distribution. Let $X_1, X_2, X_3, \ldots, X_{64}$ represent a random sample of 64 observations from this distribution. The distribution of the sample mean, $\overline{X}$, is

(1) of unknown shape with $\mu_{\overline{X}} = 60$ and $\sigma_{\overline{X}} = 4.9$
(2) approximately Normal with $\mu_{\overline{X}} = 60$ and $\sigma_{\overline{X}} = 0.61$
(3) approximately Binomial with $\mu_{\overline{X}} = 60$ and $\sigma_{\overline{X}} = 0.61$
(4) approximately Normal with $\mu_{\overline{X}} = 60$ and $\sigma_{\overline{X}} = 4.9$
(5) approximately Binomial with $\mu_{\overline{X}} = 60$ and $\sigma_{\overline{X}} = 4.9$

26. Which **one** of the following statements is **false**?

(1) For large random samples, the true value of μ lies inside the interval $\overline{x} \pm 2\text{se}(\overline{x})$ for a little more than 95% of all samples taken.
(2) For random samples from a Normal distribution, $T = (\overline{X} - \mu)/\text{se}(\overline{X})$ is exactly distributed as Student($df = n - 1$).
(3) The precision of an estimate refers to its variability — one estimate is less precise than another if it has more variability.
(4) The Student($df = \infty$) distribution has 'fatter' or 'heavier' tails than the Normal($\mu = 0$, $\sigma = 1$) distribution.
(5) For large random samples, $T = (\overline{X} - \mu)/\text{se}(\overline{X})$ is distributed as approximately Normal($\mu_X = 0$, $\sigma_X = 1$).

ANSWERS:

1. (5)	2. (4)	3. (3)	4. (5)	5. (1)
6. (5)	7. (3)	8. (1)	9. (1)	10. (2)
11. (5)	12. (2)	13. (4)	14. (1)	15. (4)
16. (3)	17. (4)	18. (5)	19. (3)	20. (2)
21. (2)	22. (5)	23. (3)	24. (3)	25. (2)
26. (4)				