Department of Statistics

Stage I Statistics
STATS 101 / 102 / 107 / 108
Term Test: First Semester, 2001

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

- Hand in your answer sheet **only**.
- Keep a personal record of your answers on this test paper - answers will be posted on Thursday.

Questions 1 to **5** refer to the following information.

Table 1 below shows part of a table published by *The New Zealand Audit Bureau of Circulations (Incorporated)* listing information about magazine sales in New Zealand. For each magazine, measurements were made on the following variables:

No.	A unique listing number associated with each magazine.
Publication	The name of the magazine.
Frequency	The frequency at which the magazine is published — weekly or monthly.
Price	The cover price of the magazine.
Circulation 1	The total net circulation for the 6 month period ending 30/06/00.
Circulation 2	The total net circulation for the 6 month period ending 31/12/99.

No.	Publication	Frequency	Price	Circulation 1	Circulation 2
01	Ad/Media	Monthly	$5.95	1276	1337
02	Australian Woman's Weekly	Monthly	$5.20	75285	73828
03	Auto Trader	Weekly	$3.00	19531	20028
04	B	Monthly	$6.60	21046	20060
05	Boating New Zealand	Monthly	$5.95	15579	13358
06	Buy, Sell & Exchange (Cant)	Weekly	$2.00	23481	22149
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
62	Truck Trader	Monthly	$2.40	9590	9456
63	TV Guide	Weekly	$1.40	244335	251824
64	TV Hits	Monthly	$5.40	22579	20622
65	Unlimited	Monthly	$7.95	7245	7240
66	Your Home & Garden	Monthly	$5.95	33908	36156

Table 1: Magazine sales data

1. Suppose one of the main purposes of the full table is to convey information about circulation numbers for these magazines over the six month period ending 30/06/00, **Circulation 1**. One change in the presentation of the table which would **not** be an improvement, with respect to this purpose, would be to:

(1) use a code of 1 for Monthly and a code of 2 for Weekly for the levels of the variable **Frequency**.

(2) round the values of the variables **Circulation 1** and **Circulation 2** to the nearest hundred.

(3) add a row at the bottom of the table showing the average circulation for **Circulation 1** and **Circulation 2**.

(4) round the values of the variables **Circulation 1** and **Circulation 2** to the nearest five hundred.

(5) list the magazines in order of the values of the variable **Circulation 1**.

2. We wish to take a simple random sample of five magazines appearing in the full list of 66 magazines.

Draw the sample using the listing number associated with each magazine (the value of the variable **No.**) and the following line of random digits. Start at the beginning of the line and use consecutive pairs of digits.

`38683 50279 38224 09844 13578`

The sample consists of those magazines with listing numbers:

(1) 38, 35, 02, 22, 40

(2) 38, 35, 02, 38, 22

(3) 38, 68, 50, 27, 22

(4) 38, 50, 38, 09, 13

(5) 38, 68, 35, 02, 79

3. To investigate the relationship between **Circulation 1** and **Price**, the most appropriate tool to use would be a:

(1) box plot of **Circulation 1** against **Price**.

(2) histogram of **Circulation 1** for each value of **Price**.

(3) scatter plot of **Circulation 1** against **Price**.

(4) two-way table of counts with **Circulation 1** for the row values and **Price** for the column values.

(5) dot plot of **Circulation 1** for each value of **Price**.

4. Figure 1 below shows a dot plot of **Price** for both levels of **Frequency**.

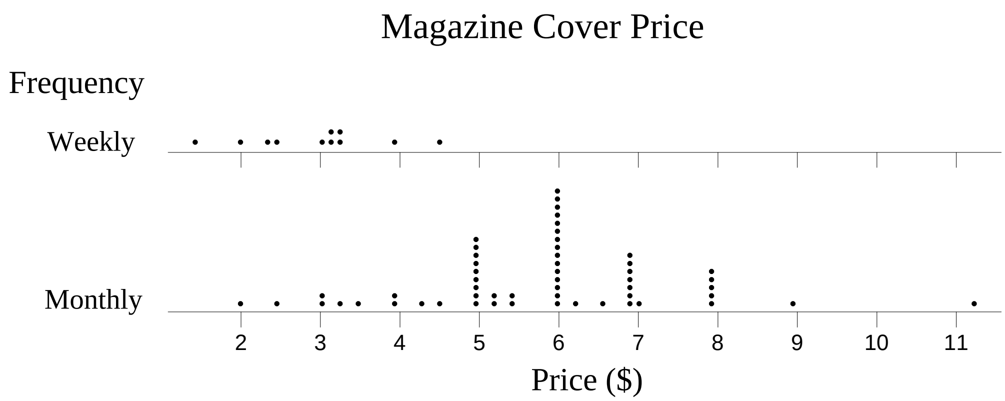

Figure 1: Magazine cover price by frequency of sale

Referring to Figure 1 above, which **one** of the following statements is **false**?

(1) Ignoring the monthly magazine that has a cover price of $11.25, (*Chartered Accountants Journal*), the cover price of the monthly magazines are slightly skewed to the left (negatively skewed).

(2) The standard deviation for the monthly magazines is greater than $3.00.

(3) Weekly magazines have a cover price, on average, approximately half that of monthly magazines.

(4) The cover prices of weekly magazines are less variable than those of monthly magazines.

(5) The monthly magazines have a mode of about $6.

5. The five-number summary for the cover price (in $) of the 55 monthly magazines is:

(2.00, 4.95, 5.95, 6.95, 11.25)

Using this five-number summary and the dot plot for the monthly magazines in Figure 1 above, which **one** of the following statements about the cover price of the **monthly** magazines is **false**?

(1) When drawing a box plot, $11.25 would be an outside value.

(2) Neither whisker would be longer than $3.00.

(3) The lower whisker would be longer than the upper whisker.

(4) The interquartile range is $2.00.

(5) When drawing a box plot, $2.00 would be an outside value.

Questions 6 to **9** refer to the following information.

Forest researchers interested in the growth rate and yield of pinus radiata trees commonly measure the diameter of trees at a height 1.3 metres above the ground. This variable is called the diameter at breast height (**dbh**).

The data in Table 2 below was taken from research reported in the *Journal of Applied Statistics* (Vol. 23, No. 6, 1996, 609 – 619). The frequency table below gives the diameters at breast height of a sample of 55 mature pinus radiata trees from a plot in a forest.

Diameter at breast height dbh (cm)	**Frequency**
20 – less than 28	2
28 – less than 36	4
36 – less than 44	10
44 – less than 52	12
52 – less than 60	12
60 – less than 68	12
68 – less than 76	3
Total	55

Table 2: Diameters at breast height of 55 pinus radiata trees

6. The **best** estimates of the sample mean, $\overline{x}$, and sample standard deviation, s, of these **dbh** values are:

(1) $\overline{x} = 55.1$cm $s = 12.07$cm

(2) $\overline{x} = 51.1$cm $s = 11.96$cm

(3) $\overline{x} = 51.1$cm $s = 12.07$cm

(4) $\overline{x} = 47.1$cm $s = 12.07$cm

(5) $\overline{x} = 47.1$cm $s = 11.96$cm

Questions 7 and **8** refer to the following additional information.

Trees with a **dbh** of 44cm or more are of great value to the forest company because they can be used to make high quality veneer plywood. The proportion of trees in this sample with a **dbh** of 44cm or more is 0.709. Let p be the population proportion of trees (grown under the same conditions) with a **dbh** of 44cm or more. Let $\widehat{P}$ be the proportion of trees with a **dbh** of 44cm or more in a random sample of 55 trees (grown under the same conditions).

7. Using $\text{se}(\widehat{p}) = \sqrt{\dfrac{\widehat{p}(1-\widehat{p})}{n}}$, a two-standard-error interval for p is:

(1) (0.701, 0.717)

(2) (0.653, 0.765)

(3) (0.518, 0.900)

(4) (0.587, 0.831)

(5) (0.648, 0.770)

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

(1) The variability of the sample proportion decreases as the sample size increases.

(2) If another random sample of 55 **dbh** values was taken from trees in the same plot in the forest, the sample proportion obtained would probably not be 0.709.

(3) The standard error of the sample proportion, $\text{se}(\widehat{p})$, estimates the variability of $\widehat{P}$.

(4) If another random sample of 55 **dbh** values was taken from trees in the same plot in the forest, the population proportion p would probably change.

(5) The sample proportion $\widehat{p} = 0.709$ is an estimate of p.

9. Let the random variable T represent the number of trees in a random sample of 55 trees from a plot of mature pinus radiata trees with a **dbh** of 44cm or more. Assume that T can be well modelled by a Binomial$(n = 55, p = 0.709)$ distribution.

Let $T_1, T_2, T_3, \ldots, T_{100}$ represent a random sample of 100 observations from the distribution of T.

The distribution of the mean number of trees with a **dbh** of 44cm or more per sample, $\overline{T}$, has a standard deviation, $\sigma_{\overline{T}}$, given by:

(1) $\sigma_{\overline{T}} = 0.34$

(2) $\sigma_{\overline{T}} = 3.37$

(3) $\sigma_{\overline{T}} = 33.69$

(4) $\sigma_{\overline{T}} = 11.35$

(5) $\sigma_{\overline{T}} = 0.03$

Questions 10 to **12** refer to the following information.

A management theorist believes that the success of a manager is related to the number of interactions a manager has with people outside his or her work unit (*networking*) over some specified time period. A random sample of nineteen managers from medium-sized manufacturing plants was measured on a success index — the higher the value of the index the more successful the manager. A scatter plot of **Manager Success Index** against **Number of Interactions with Outsiders** is given in Figure 2 below.

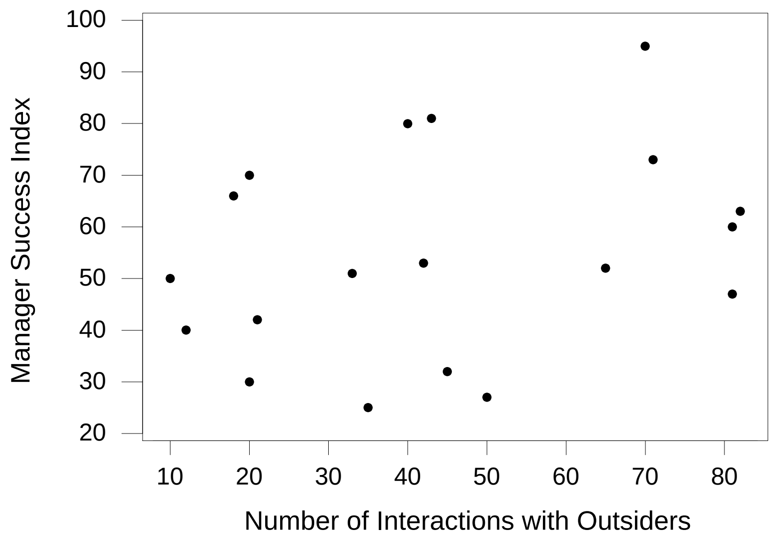

Figure 2: Relationship between success as a manager and networking

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

(1) The variables are on the correct axes, if the **Number of Interactions with Outsiders** is being used to predict the **Manager Success Index**.

(2) Using these data only, it would be dangerous to make a prediction of a manager's success index for a manager who made 110 interactions with outsiders over the specified time period.

(3) There is clearly non-constant scatter about a linear trend line.

(4) Using these data only, it would be difficult to make a precise prediction of a manager's success index for a manager who made 55 interactions with outsiders over the specified time period.

(5) There is very weak positive linear association in the data between **Manager Success Index** and **Number of Interactions with Outsiders**.

11. From this study alone, we should **not** conclude that a manager could increase his or her success index by increasing the number of interactions with outsiders **mainly** because:

(1) there has been no attempt by the researchers to use a control group.

(2) the researchers did not allocate the number of interactions each manager had with outsiders.

(3) there has been no blinding used in this study.

(4) some managers with a low score for **Number of Interactions with Outsiders** have a high score for **Manager Success Index**.

(5) the number of managers in the study was too small.

12. A stem-and-leaf plot of **Number of Interactions with Outsiders** for the 19 managers is given in Figure 3 below.

Units: 2 | 5 = 25 interactions with outsiders
$n = 19$

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

Figure 3: Number of interactions with outsiders for 19 managers

The interquartile range of the data in the stem-and-leaf plot in Figure 3 above is:

(1) 51

(2) 47

(3) 49

(4) 48

(5) 50

Questions 13 to **15** refer to the following information.

Let the random variable C represent the number of fatal air crashes per year in New Zealand.

13. Which **one** of the following statements about fatal air crashes in New Zealand would, if true, make the Poisson distribution **unsuitable** as a model for the distribution of C?

(1) The number of fatal air crashes per year is a discrete random variable.

(2) It is extremely unlikely for two or more fatal air crashes to occur at exactly the same time.

(3) Fatal air crashes are random occurrences.

(4) Each fatal air crash is independent of all other fatal air crashes.

(5) Fatal air crashes occur more often in summer than in winter.

Questions 14 and **15** refer to the following additional information.

Suppose it is appropriate to model C with a Poisson distribution with parameter $\lambda = 10$.

Use the appropriate part of the MINITAB output below to answer Questions **14** and **15**.

```
Poisson with mu = 2.50000

      x        P( X = x)
   2.00          0.2565
   3.00          0.2138
   4.00          0.1336
   5.00          0.0668
   6.00          0.0278
```

```
Poisson with mu = 2.50000

      x        P( X <= x)
   2.00          0.5438
   3.00          0.7576
   4.00          0.8912
   5.00          0.9580
   6.00          0.9858
```

```
Poisson with mu = 10.0000

      x        P( X = x)
  12.00          0.0948
  13.00          0.0729
  14.00          0.0521
  15.00          0.0347
  16.00          0.0217
```

```
Poisson with mu = 10.000

      x        P( X <= x)
  12.00          0.7916
  13.00          0.8645
  14.00          0.9165
  15.00          0.9513
  16.00          0.9730
```

14. The probability that fewer than 15 fatal air crashes occur in a year in New Zealand is approximately:

(1) 0.9513
(2) 0.0347
(3) 0.9730
(4) 0.9165
(5) 0.0521

15. The probability that there are exactly 3 fatal air crashes in New Zealand in the **first three months** of the year is approximately:

(1) 0.2138
(2) 0.7916
(3) 0.7576
(4) 0.0237
(5) 0.0948

Questions 16 to **18** refer to the following information.

During the US presidential election in 2000 it was claimed that many voters in the county of Palm Beach in the state of Florida were confused by the ballot paper. They may have inadvertently voted for Buchanan when they had intended to vote for Gore. Table 3 below shows the number of votes cast for presidential candidates for the five counties in south-east Florida.

Presidential Candidate

County	Bush	Gore	Buchanan	Other	**Total**
Broward	177,902	387,703	795	8743	575,143
Martin	33,970	26,620	112	1311	62,013
Miami-Dade	289,533	328,808	560	6548	625,449
Monroe	16,059	16,483	47	1298	33,887
Palm Beach	152,951	269,732	3411	7092	433,186
Total	670,415	1,029,346	4925	24,992	1,729,678

Table 3: Votes for US presidential candidates for five Florida counties

Use Table 3 above to answer **Questions 16** to **18**.

16. The percentage of voters who did **not** vote for Bush or Gore is approximately:

(1) 1.4%
(2) 61.2%
(3) 40.5%
(4) 0.3%
(5) 1.7%

17. The proportion of voters who voted for Bush and who lived in Miami-Dade county is approximately:

(1) 0.167
(2) 0.749
(3) 0.463
(4) 0.582
(5) 0.432

18. The percentage of Palm Beach voters who voted for Buchanan is approximately:

(1) 0.79%

(2) 69.26%

(3) 25.13%

(4) 0.20%

(5) 25.33%

19. On 13 January 2000 *The New Zealand Herald* published the results of a *Herald*-DigiPoll. The 750 people polled were asked if they thought it was unmanly for men to cry. The sample consisted of 500 men and 250 women. 10% of the men and 4% of the women agreed that men crying was unmanly.

Of the people in this survey who thought crying was **not** unmanly, the proportion who were women is approximately:

(1) 0.167

(2) 0.320

(3) 0.348

(4) 0.960

(5) 0.933

Questions 20 to **23** refer to the following information.

Based on US figures from 1926 to 1997 the (annual) compound return for large company shares in the United States has a mean of 11.0 per cent and a standard deviation of 20.3 per cent. Assume that the compound return for large US company shares has a Normal distribution with a mean of 11.0 per cent and a standard deviation of 20.3 per cent.

Use the appropriate part of the MINITAB output below to answer Questions **20** to **22**.

```
Normal with mean = 11.0000 and standard deviation = 20.3000

         x     P( X <= x)                    P( X <= x)         x
    0.0000       0.2940                        0.2000     -6.0849
   20.0000       0.6712                        0.4000      5.8571
   30.0000       0.8254                        0.5000     11.0000
   50.0000       0.9726                        0.6000     16.1429
   80.0000       0.9997                        0.8000     28.0849


Normal with mean = 0 and standard deviation = 28.7085

         x     P( X <= x)
  -80.0000       0.0027
  -20.0000       0.2430
    0.0000       0.5000
   20.0000       0.7570
   80.0000       0.9973


Normal with mean = 0 and standard deviation = 40.6000

         x     P( X <= x)
  -80.0000       0.0244
  -20.0000       0.3111
    0.0000       0.5000
   20.0000       0.6889
   80.0000       0.9756
```

20. The proportion of large US companies whose compound return for shares is between 20 per cent and 50 per cent is approximately:

(1) 0.1708
(2) -0.3014
(3) 0.8254
(4) 0.0271
(5) 0.3014

21. Consider the top 20% of large US companies ranked by compound returns for shares. The lowest possible compound return for any of these so-called "Top 20%" companies is approximately:

(1) 100.0 per cent
(2) 28.1 per cent
(3) 32.9 per cent
(4) 10.3 per cent
(5) −6.1 per cent

22. Two large US companies are randomly selected. Assume that the compound returns for shares in these two companies are independent. The probability that the compound returns for shares in these two companies differ by at least 20 per cent is approximately:

(1) 0.5140
(2) 0.4860
(3) 0.2430
(4) 0.3111
(5) 0.6222

23. If a random sample of 16 large US companies was selected there would be a 68% chance that the mean compound return for shares in these companies would be approximately between:

(1) 9.7 per cent and 12.3 per cent
(2) −29.6 per cent and 51.6 per cent
(3) −9.3 per cent and 31.3 per cent
(4) 5.9 per cent and 16.1 per cent
(5) 0.9 per cent and 21.2 per cent

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

(1) Slight changes in the wording of questions can make a measurable difference to the results of a survey.

(2) There are always statistical procedures available to correct results (at the completion of a survey) when the population from which a sample is taken is different from the population of interest.

(3) Bias can occur when too many respondents in a survey give an answer which does not reflect their actual behaviour.

(4) The outcome of a survey which uses personal interviews may be different from the outcome of the same survey if telephone interviews had been used.

(5) The outcome of a survey may be affected by the race and/or gender of the interviewer.

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

(1) An experiment involving human subjects can involve psychological effects. There should be an attempt to account for such effects in the design of the experiment.

(2) In experiments, blocking is a procedure used in an attempt to ensure that comparisons made between the treatments are fair.

(3) In a completely randomised design, the randomisation is carried out over all of the units together so that each unit is equally likely to be assigned to any one of the treatment groups.

(4) If blocking is used in the design of an experiment, then nothing is gained from using randomisation in that design.

(5) Having a control group is important when the effectiveness of a treatment is to be estimated.

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

(1) For very large values of the degrees of freedom, the distribution of $T = \dfrac{\overline{X} - \mu}{\text{se}(\overline{X})}$ is almost identical to the standard Normal distribution.

(2) $T = \dfrac{\overline{X} - \mu}{\text{se}(\overline{X})}$ measures the difference between $\overline{X}$ and μ in terms of the number of standard errors of the sample mean.

(3) The graph of the Student's t-distribution with 20 degrees of freedom ($df = 20$) has fatter tails than the graph of the Student's t-distribution with 10 degrees of freedom ($df = 10$).

(4) The graph of the Student's t-distribution with 50 degrees of freedom ($df = 50$) has fatter tails than the graph of the standard Normal distribution.

(5) Student's t-distribution describes a family of distributions indexed by a parameter called the degrees of freedom.

ANSWERS:

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