Department of Statistics

Stage I Statistics
STATS 191
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 **6** 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. **Frequency** is classified as a:

(1) qualitative variable.

(2) response variable.

(3) discrete variable.

(4) quantitative variable.

(5) continuous variable.

3. 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, 38, 22
(2) 38, 50, 38, 09, 13
(3) 38, 35, 02, 22, 40
(4) 38, 68, 35, 02, 79
(5) 38, 68, 50, 27, 22

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

(1) dot plot of **Circulation 1** for each value of **Price**.
(2) scatter plot of **Circulation 1** against **Price**.
(3) histogram of **Circulation 1** for each value of **Price**.
(4) box plot of **Circulation 1** against **Price**.
(5) two-way table of counts with **Circulation 1** for the row values and **Price** for the column values.

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

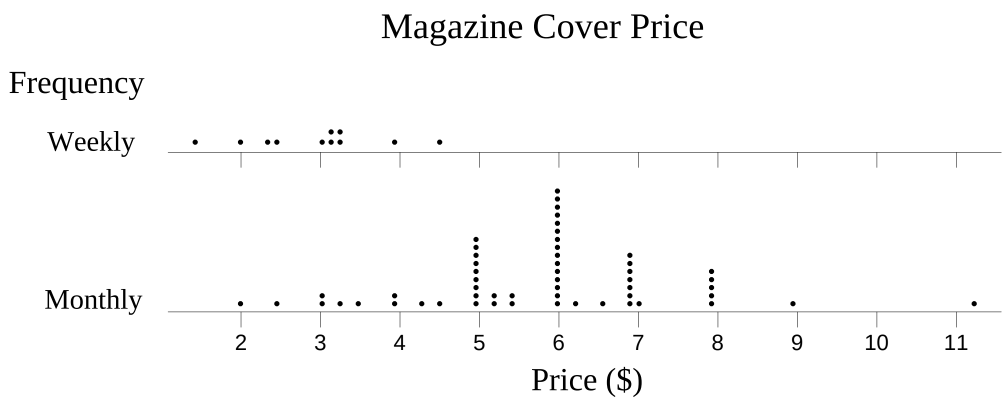

Figure 1: Magazine cover price by frequency of sale

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

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

(2) 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).

(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 standard deviation for the monthly magazines is greater than \$3.00.

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

$$(2.00,\quad 4.95,\quad 5.95,\quad 6.95,\quad 11.25)$$

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) The interquartile range is \$2.00.

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

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

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

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

Questions 7 to **11** 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.

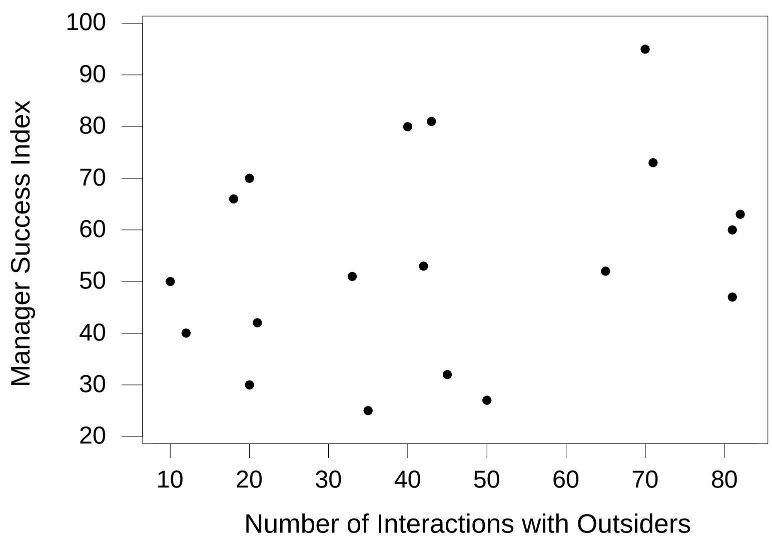

Figure 2: Relationship between success as a manager and networking

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

(3) 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.

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

(5) 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.

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

Units: $2 \mid 5 = 25$ 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 median number of interactions with outsiders for these data are:

(1) 41
(2) 40
(3) 42.5
(4) 42
(5) 43

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

(1) 50
(2) 47
(3) 51
(4) 48
(5) 49

10. 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 blinding used in this study.

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

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

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

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

11. The mean number of interactions with outsiders for this random sample of 19 managers is $\overline{x} = 44.16$, with a standard deviation of $s = 24.52$. Using $t = 2.101$ as the t-multiplier, a 95% confidence interval for the underlying mean number of interactions with outsiders, μ, is:

(1) (41.45, 46.87)

(2) (32.34, 55.98)

(3) (33.13, 55.19)

(4) (38.53, 49.79)

(5) (32.91, 55.41)

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

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

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

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

(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) 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.

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

(1) 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.

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

(3) 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.

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

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

Questions 14 to **16** 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

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

(1) $\overline{x} = 47.1$cm $s = 12.07$cm
(2) $\overline{x} = 55.1$cm $s = 12.07$cm
(3) $\overline{x} = 51.1$cm $s = 11.96$cm
(4) $\overline{x} = 51.1$cm $s = 12.07$cm
(5) $\overline{x} = 47.1$cm $s = 11.96$cm

Questions 15 and **16** 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).

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

(1) (0.587, 0.831)

(2) (0.518, 0.900)

(3) (0.648, 0.770)

(4) (0.701, 0.717)

(5) (0.653, 0.765)

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

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

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

(3) 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.

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

(5) 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.

Questions 17 to **20** 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 **17** to **19**.

```
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
```

17. The probability that the compound return for shares in a randomly selected large US company is negative (less than 0 per cent) is approximately:

(1) 0.2940
(2) 0.2000
(3) 0.5000
(4) 0.1100
(5) 0.7060

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

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

19. 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) 10.3 per cent
(2) 32.9 per cent
(3) 28.1 per cent
(4) -6.1 per cent
(5) 100.0 per cent

20. 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) -29.6 per cent and 51.6 per cent
(2) -9.3 per cent and 31.3 per cent
(3) 0.9 per cent and 21.2 per cent
(4) 9.7 per cent and 12.3 per cent
(5) 5.9 per cent and 16.1 per cent

Questions 21 to **24** refer to the following information.

A Gallup Poll was conducted on Wednesday night, 14 March 2001, after the Dow Jones Industrial Average dropped more than 300 points during the day, closing at its lowest level in almost two years. Two of the questions asked in a telephone interview of American adults were:

> **Stock Market Question:**
> Do you think that the stock market will get better, stabilise, or get worse?

and

> **Economy Question:**
> Do you think that the American economy will get better, stabilise, or get worse?

There were 463 stockholders and 193 non-stockholders interviewed. The results (as percentages) for these two questions are shown in Table 3 below.

Question	**Response**				**Sample size**
	Better %	**Stabilise %**	**Worse %**	**Unsure %**	
Stock Market					
Stockholders	29%	54%	14%	3%	463
Non-Stockholders	29%	43%	19%	9%	193
Economy					
Stockholders	19%	53%	26%	2%	463
Non-Stockholders	30%	43%	23%	4%	193

Table 3: Americans' reaction to the fall in the stock market

Let:

> p_{better} be the proportion of **stockholders** who think that the **economy** will get **better**

and

> p_{worse} be the proportion of **stockholders** who think that the **economy** will get **worse**.

21. An estimate of the difference between p_{better} and p_{worse} is:

(1) -0.25
(2) -0.07
(3) 0.10
(4) 0.15
(5) -0.34

22. For the purpose of calculating se($\widehat{p}_{better} - \widehat{p}_{worse}$), the sampling situation can be described as:

(1) one sample of size 463, many yes/no items.

(2) one sample of size 463, several response categories.

(3) one sample of size 656, several response categories.

(4) two independent samples of sizes 463 and 193.

(5) one sample of size 656, many yes/no items.

Questions 23 and **24** refer to the following additional information.

Let:

$p_{stockmarket}$ be the proportion of **stockholders** who think that the **stock market** will **stabilise**

and

$p_{economy}$ be the proportion of **stockholders** who think that the **economy** will **stabilise**.

Information from Table 3 is used to construct a 95% confidence interval for the difference $p_{stockmarket} - p_{economy}$.

23. The formula for the standard error of the estimate, se($\widehat{p}_{stockmarket} - \widehat{p}_{economy}$), is:

(1) $\sqrt{\dfrac{\widehat{p}^2_{stockmarket}}{463} - \dfrac{\widehat{p}^2_{economy}}{463}}$

(2) $\sqrt{\dfrac{\widehat{p}^2_{stockmarket}}{463} + \dfrac{\widehat{p}^2_{economy}}{463}}$

(3) $\sqrt{\dfrac{\widehat{p}_{stockmarket}(1 - \widehat{p}_{stockmarket})}{463} + \dfrac{\widehat{p}_{economy}(1 - \widehat{p}_{economy})}{463}}$

(4) $\sqrt{\dfrac{(1 - \widehat{p}_{stockmarket}) + (1 - \widehat{p}_{economy}) - (\widehat{p}_{stockmarket} - \widehat{p}_{economy})^2}{463}}$

(5) $\sqrt{\dfrac{(\widehat{p}_{stockmarket} + \widehat{p}_{economy}) - (\widehat{p}_{stockmarket} - \widehat{p}_{economy})^2}{463}}$

24. A 95% confidence interval for $p_{stockmarket} - p_{economy}$ is $(-0.078, 0.098)$. The best interpretation of this interval is:

(1) With 95% confidence, $p_{stockmarket}$ is somewhere between 7.8% higher than and 9.8% lower than $p_{economy}$.

(2) With 95% confidence, $p_{stockmarket}$ is somewhere between 7.8% lower than and 9.8% higher than $p_{economy}$.

(3) With 95% confidence, $p_{stockmarket}$ is either 7.8% higher than or 9.8% lower than $p_{economy}$.

(4) With 95% confidence, the true mean of the difference $p_{stockmarket} - p_{economy}$ is in the interval $(-0.078, 0.098)$.

(5) With 95% confidence, $p_{stockmarket}$ is either 7.8% lower than or 9.8% higher than $p_{economy}$.

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

(1) All values in a confidence interval are plausible as values of the parameter that could have produced the sample estimate.

(2) The level of confidence for an interval gives the percentage of such intervals that contain the true value of the parameter, under long-run repeated sampling.

(3) A confidence interval for a parameter is a set of plausible values for the parameter.

(4) Generally, the width of a confidence interval decreases as the level of confidence we have in it increases.

(5) Once the confidence interval has been calculated, the true value of the parameter is either in the interval or not (but we do not usually know which is true).

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)				