University of Toronto Scarborough
STAB22 Midterm Examination

June 2007

Name:	
ID:	
Tutorial day and time	

For this examination, you are allowed one letter-sized sheet of notes (both sides) prepared by you, a non-programmable, non-communicating calculator, and writing implements.

This examination has 14 numbered pages; before you start, check to see that you have all the pages.

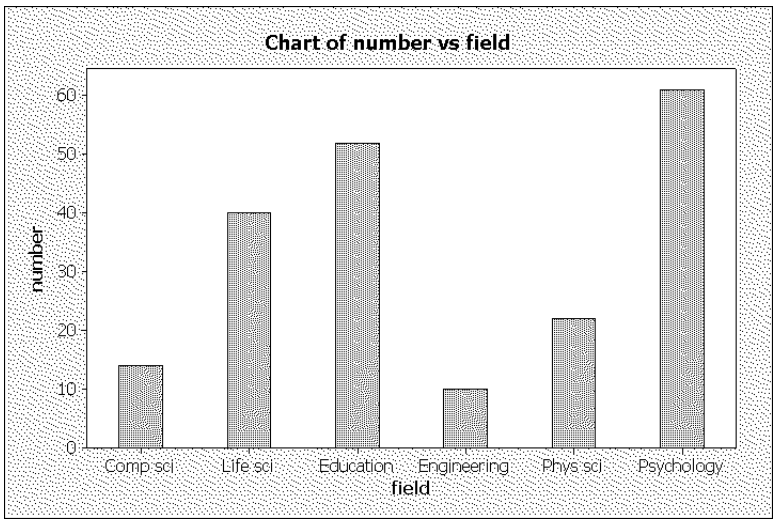
This examination contains a mixture of short-answer and multiple choice questions. For the short-answer questions, write your answers in the spaces provided; for the multiple-choice questions, circle the best answer out of the alternatives given.

Marks for each question are as shown.

The table below is for the markers' use only.

Question	Max	Mark
1	5	
2	8	
3	5	
4	7	
5	3	
6	9	
7	8	
8	2	
9	12	
10	3	
11	3	
12	3	
13	3	
14	7	
15	3	
16	7	
17	7	
18	5	
Total	100	

1. (5 marks) At a certain university in 1993, a record was kept of the number of women graduating with doctoral degrees in each of six different fields of study.



(a) What is this kind of graph usually called?

i. Histogram
ii. Bar chart
iii. Stem-and-leaf plot
iv. Pie chart

(b) Which field of study had the most women graduating with doctoral degrees in 1993?

(c) Approximately how many women graduated with doctoral degrees in life sciences in 1993?

2. (8 marks) A political scientist takes a large sample of registered voters, and measures a number of variables, as shown below. Are each of the variables listed below categorical or quantitative?

   (a) Gender
      i. Categorical
      ii. Quantitative

   (b) Age
      i. Categorical
      ii. Quantitative

   (c) Household income
      i. Categorical
      ii. Quantitative

   (d) Party voted for at last election
      i. Categorical
      ii. Quantitative

3. (5 marks) What is the median of the following set of numbers: 7, 6, 10, 9, 5?

   (a) 7.4
   (b) 7
   (c) 10
   (d) 6

4. (7 marks) The speeds of 57 cars were measured (in km/h) on a city street. A numerical summary of the results is given below:

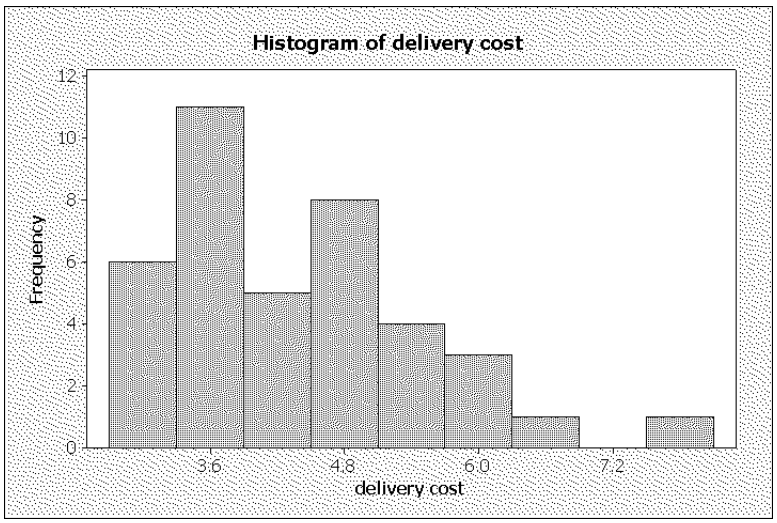
```
Descriptive Statistics: speed

Variable   N  N*   Mean  SE Mean  StDev
speed     57   0  45.53     1.64  12.38

Variable   Minimum     Q1  Median     Q3  Maximum
speed        25.60  36.80   43.20  51.20    83.20
```

   Do a calculation to decide whether the highest recorded speed, 83.2 km/h, is an outlier in this data set. What do you conclude?

5. (3 marks) A delivery company recorded the delivery costs for 39 small packages it delivered one day last week. A histogram of the delivery costs is shown below.



(a) Describe the shape of this histogram.

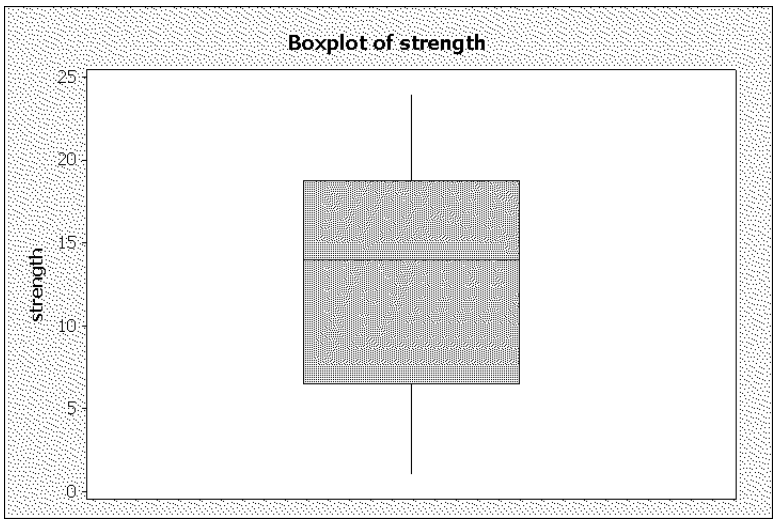
(b) A manager at the delivery company recorded the mean and median delivery cost, but unfortunately lost the paper on which the values were recorded. However, the manager can remember that one value was about \$4.00 and the other was about \$4.40. Which value is the mean and which is the median?

i. \$4.00 is the mean and \$4.40 is the median.

ii. \$4.40 is the mean and \$4.00 is the median.

iii. It's impossible to tell which is which.

6. (9 marks) All the third-graders at a certain elementary school were given a physical-fitness strength test. The test scores are shown in the boxplot below.



(a) What is the median test score, approximately?

(b) What is the interquartile range of test scores, approximately? Show your calculations.

(c) Are there any unusually high or low test scores? If there are, indicate the unusual values on the boxplot.

7. (8 marks) A set of exam marks has mean 70, median 65, inter-quartile range 25 and SD 15 marks. It is decided to subtract 10 from all the marks. For the new set of marks,

   (a) what is the mean?

   (b) what is the median?

   (c) what is the inter-quartile range?

   (d) what is the SD?

8. (2 marks) A set of data has quartiles 30 and 75, and median 40. What would you conclude about the shape of the data distribution?

9. (12 marks) For a particular group of adult males, the distribution of cholesterol readings is normal with mean 210 and SD 15. Use Table A to find the following, showing your calculation in each case:

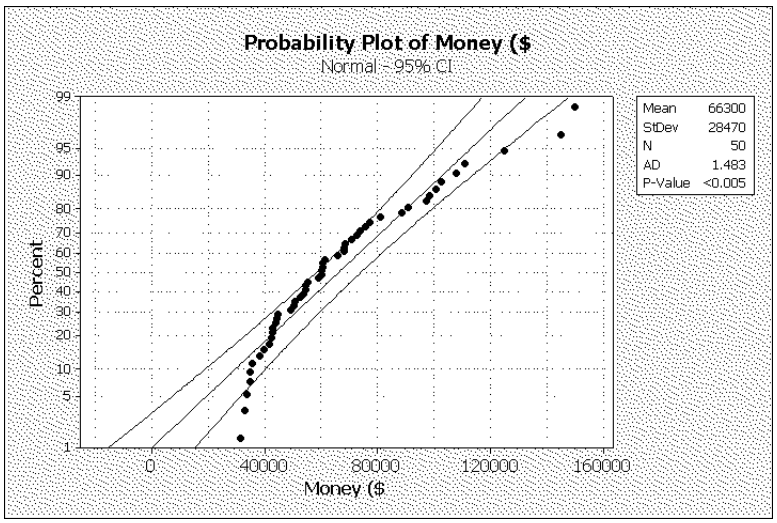
   (a) The proportion of males in this group with cholesterol reading less than 240.

   (b) The proportion of males in this group with cholesterol readings between 200 and 240.

   (c) The cholesterol reading that 20% of males in this group are higher than.

   (d) The first quartile of cholesterol readings for males in this group.

10. (3 marks) The 1998 winnings of 50 professional golfers were recorded. A normal quantile plot of the distribution is shown below.



From the plot, do you conclude that these data are described well by a normal distribution? In one sentence, explain your conclusion.

11. (3 marks) The correlation between two variables is very close to 0.

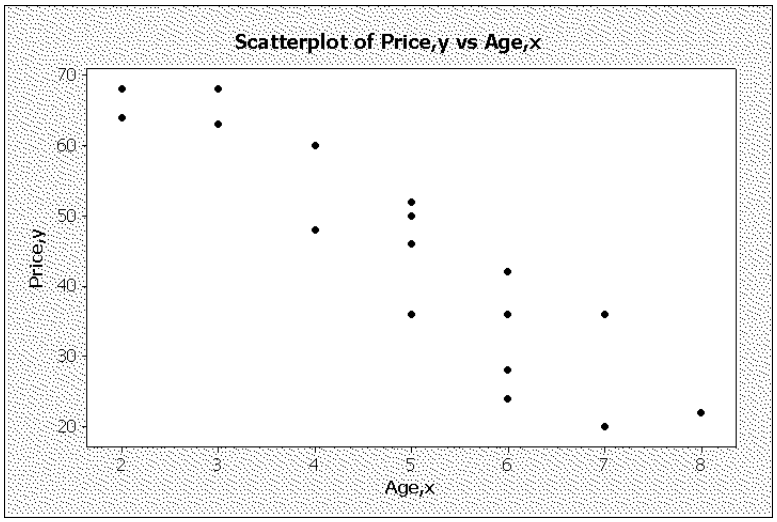
    (a) Does it follow that there is no relationship of any kind between these two variables?

        i. yes, there cannot be any relationship

        ii. no, there could be some kind of relationship

    (b) If you answered "no" to part (a), what kind of relationship could there be? Answer in one sentence only.

12. (3 marks) The scatterplot below shows, for 19 used foreign compact cars, the age of the car, and the asking price (in hundreds of dollars).



Choose the best value for the correlation between these two variables from the list below.

(a) -0.9

(b) -0.4

(c) 0

(d) 0.4

(e) 0.9

13. (3 marks) Water flowing across farmland washes away soil. Researchers released water across a test area at different flow rates and measured how much soil was washed away (amount of eroded soil). In this case, which is the explanatory variable and which is the response?

(a) Eroded soil is the explanatory variable and flow rate is the response.

(b) Eroded soil is the response and flow rate is the explanatory variable.

(c) There is no explanatory variable or response in this situation.

14. (7 marks) A study was made of some popular fast-food items. In particular, a regression analysis was done for predicting calorie content from the amount of fat in a food item. Some regression output from Minitab is given below:

```
Regression Analysis: Calories versus Fat

The regression equation is
Calories = 242 + 7.35 Fat


Predictor    Coef  SE Coef     T      P
Constant   242.03    68.44  3.54  0.003
Fat         7.353    2.860  2.57  0.021


S = 152.030   R-Sq = 29.2%   R-Sq(adj) = 24.8%
```

(a) What is the value of the slope of this regression line?

(b) Which is the best interpretation of the slope of this regression line?

i. The calorie content of an item that has no fat

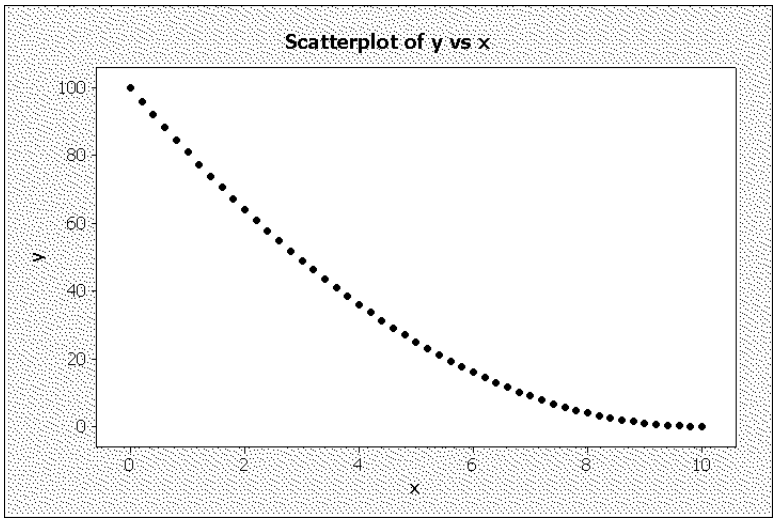
ii. The fat content of an item that has no calories

iii. The increase in calories associated with a one-unit increase in fat

iv. The increase in fat associated with a one-calorie increase.

(c) Predict the calorie content of an item containing 15 units of fat.

15. (3 marks) The plot below shows a scatterplot for two variables x and y.



Which of the statements below about the correlation between x and y is most accurate?

(a) The correlation is $+1$ because there is a perfect positive association.

(b) The correlation is a little less than $+1$ because the trend is slightly curved.

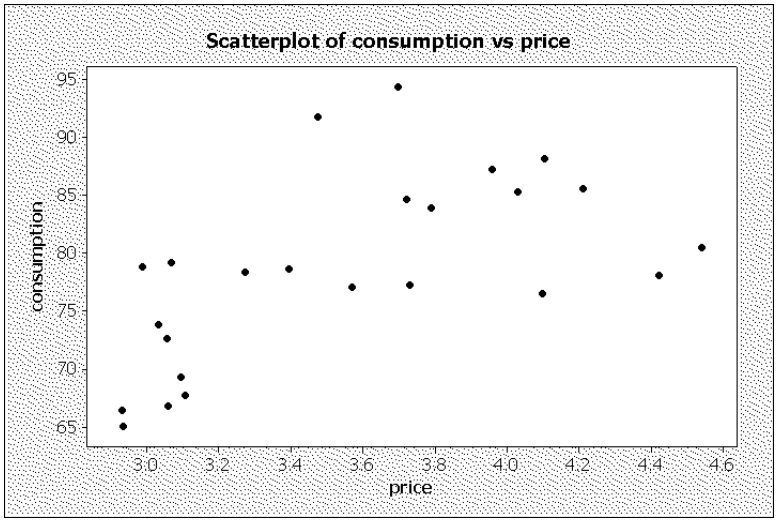
(c) The correlation is 0 because the trend is not linear.

(d) The correlation is a little greater than -1 because the trend is slightly curved.

(e) The correlation is -1 because there is a perfect negative association.

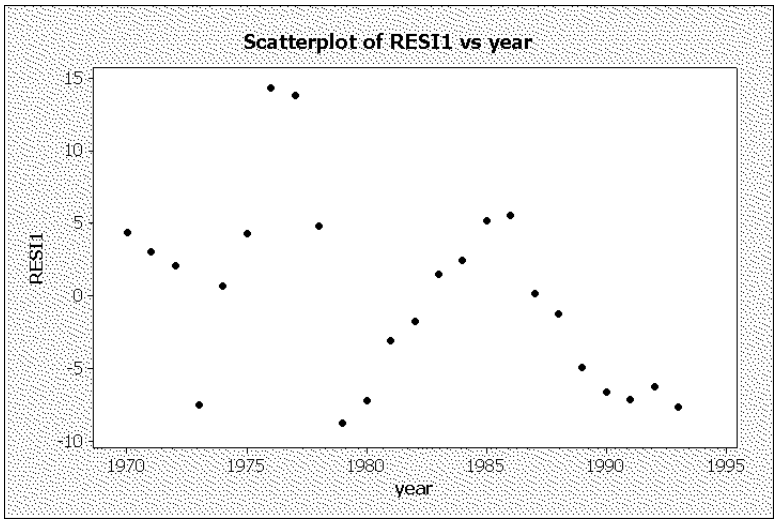
16. (7 marks) A study was made of the price (dollars per pound) and consumption (pounds per person per year) of beef in the United States for each year 1970–1993. The prices were adjusted to 1993 dollars, because everything got more expensive in this time period.

(a) Economic theory says that if the price of an item is higher, consumption of that item will be less. A scatter plot for predicting consumption from price is shown below.



Does this scatterplot support the economic theory? Explain briefly (1 sentence).

(b) A regression was done for predicting consumption from price. The residuals were plotted against the year, as shown below.



Do you see any problems in this plot? Explain briefly (1 sentence).

17. (7 marks) Breast cancer that is detected in its early stages can be treated. In the past, the preferred treatment was mastectomy (removal of the breast); now, it is usual to remove the tumour, and have the patient undergo radiation. A medical team compares the survival times after surgery of all women who have had either treatment.

    (a) What is the explanatory variable here?

    (b) What is the response variable here?

    (c) Is this study (choose one):

        i. an observational study

        ii. an experiment?

    (d) Will the medical team be able to conclude that the better of the two treatments causes a longer average survival time (circle your preferred answer)?

        i. yes

        ii. no

18. (5 marks) A statistical experiment is to be done to compare three treatments. 9 subjects are available, named: Alomar, Bikalis, Cranston, Durr, George, Han, Imrani, Lawless, Zhang. Three equal-sized groups of subjects will be used. The numbers 1–9 are randomly rearranged as follows:

    6 2 9 1 8 5 3 4 7

    Name the subjects that will receive the 2nd treatment, and explain how you came to your conclusion.