

University of Toronto Scarborough

STAB22 Midterm Examination

October 2008

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

This question paper has 12 numbered pages; before you start, check to see that you have all the pages. There is also a signature sheet at the front and statistical tables at the back.

This examination is multiple choice. Each question has equal weight. On the Scantron answer sheet, ensure that you enter your last name, first name (as much of it as fits), and student number (in “Identification”).

Mark in each case the best answer out of the alternatives given (which means the numerically closest answer if the answer is a number and the answer you obtained is not given.)

Before you begin, check that the colour printed on your Scantron sheet matches the colour of your question paper. If it does not, get a new Scantron from an invigilator.

Also before you begin, complete the signature sheet, but *sign it only when the invigilator collects it*. The signature sheet shows that you were present at the exam.

1. There are two sets of numbers. Set A is 5, 5, 5, 5. Set B is 4, 5, 6, 7. Without doing any calculations, what can you say about the standard deviations of the two sets of numbers?
 - (a) Sets A and B have about the same standard deviation.
 - (b) Set B has a smaller standard deviation.
 - (c) The only way to compare the standard deviations is to calculate them.
 - (d) * Set A has a smaller standard deviation.

2. What is the median of the values 7, 3, 0, 1, 6?
 - (a) 3.4
 - (b) * 3
 - (c) 6
 - (d) 0

3. A stringing machine is a device for adjusting the tension of strings in a tennis racket. Some tennis players ask for a string tension of 65 pounds. The actual tension produced by the machine for the rackets belonging to these players varies according to a normal distribution with mean 65 pounds and standard deviation 2 pounds. Use this information to answer this question and the 2 following questions.

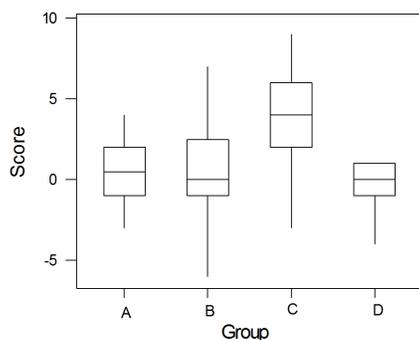
What proportion of rackets (belonging to players asking for a string tension of 65 pounds) will have a string tension of 67.5 pounds or more?

 - (a) 0.00
 - (b) about 0.50
 - (c) 1.00
 - (d) 0.89
 - (e) * 0.11

4. Using the information in Question 3, what proportion of rackets (belonging to players asking for a string tension of 65 pounds) will have a string tension between 64 pounds and 67.5 pounds?
 - (a) * 0.59
 - (b) 0.89
 - (c) 0.11
 - (d) 0.70
 - (e) 0.31

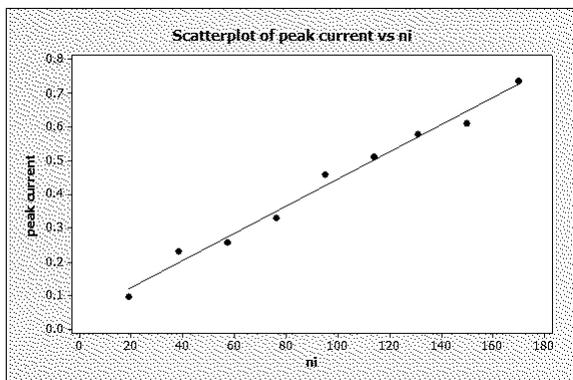
5. For all rackets (belonging to players asking for a string tension of 65 pounds), there is a value x such that 20% of rackets will have an actual string tension less than x pounds. What is x ?
 - (a) 60.9
 - (b) 69.1
 - (c) 66.7
 - (d) 65
 - (e) * 63.3

6. Below is a comparison of scores in a spatial-temporal reasoning test for four groups of children. Higher scores indicate superior reasoning. Use this information for this question and the following one.



Which of the following is true?

- (a) * The range and IQR for group C are both smaller than the corresponding measures for group B.
 - (b) More than 75% of those in group C scored higher than the best child in group A.
 - (c) Fewer than half of the children in group A scored above 0.
 - (d) The distribution of scores in group A is right skewed.
7. Using the information in Question 6 above, which group has a right-skewed distribution of scores?
- (a) group A
 - (b) group D
 - (c) * group B
 - (d) group C
8. In a chemical procedure called differential pulse polarography, a chemist measured the peak current produced (in microamperes) when a solution containing a given amount of nickel is used. A scatterplot showing the peak current against amount of nickel (Ni), together with the least-squares regression line, is shown below. Use this information for this question and the question following.



What is the correlation between peak current and amount of nickel? Choose the best alternative from those given.

- (a) * 0.99
- (b) -1.00

- (c) negative, but not -1
 - (d) 1.00
 - (e) 0.50
9. Based on the scatterplot given in Question 8, what can you say about the slope of the regression line?
- (a) it is negative
 - (b) it is close to -1
 - (c) it is close to 0
 - (d) * it is positive
 - (e) it is close to 1
10. In a long-term study, the height of children at age 6 is recorded. For those same children, their height is again measured at age 16. Which of these variables should go on the vertical scale (*y*-axis) of a scatterplot?
- (a) the age 6 height because it is the explanatory variable
 - (b) the age 16 height because it is the explanatory variable
 - (c) the age 6 height because it is the response variable
 - (d) * the age 16 height because it is the response variable
11. A subject was given some small household objects (such as a pencil or a dinner plate) and asked to estimate their size (length or diameter) without measuring them. The estimated sizes and the actual sizes of the objects were recorded, and a regression was run to predict the actual size from the estimated size. Some output is shown below. All sizes are in inches. Use this information to answer this question and the following question.

Descriptive Statistics: estimated, actual

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
estimated	9	0	6.94	1.30	3.90	2.75	3.63	7.00	9.75	14.50
actual	9	0	7.31	1.29	3.87	3.75	4.50	6.00	9.75	15.75

Regression Analysis: actual versus estimated

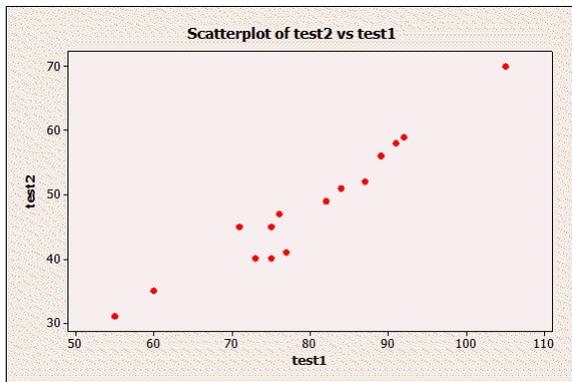
The regression equation is
 actual = 0.627 + 0.962 estimated

Predictor	Coef	SE Coef	T	P
Constant	0.6267	0.7165	0.87	0.411
estimated	0.96185	0.09119	10.55	0.000

S = 1.00540 R-Sq = 94.1% R-Sq(adj) = 93.2%

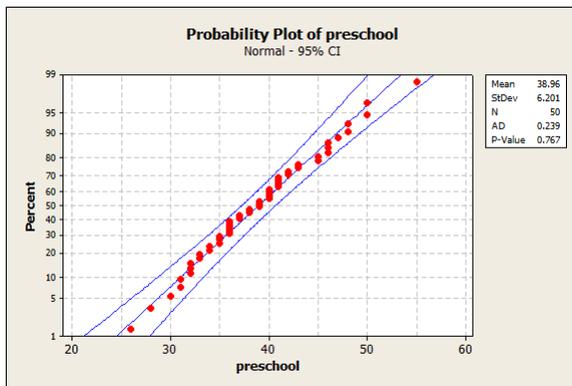
- Based on these data, how accurately do you think you could predict the actual measurement from the estimated measurement?
- (a) Prediction may not be very accurate at all.
 - (b) Prediction will only be moderately accurate.
 - (c) * Prediction will be very accurate.
 - (d) It is impossible to tell from this output how accurate a prediction might be.

12. Using the information in Question 11, if the estimated size of an object is 20 inches, what do you predict the actual size of the object to be?
- (a) 19.87 inches
 - (b) 20.00 inches
 - (c) 0.63 inches
 - (d) 0.96 inches
 - (e) * Cannot predict because using the regression line for prediction would be extrapolation.
13. A company is evaluating two job interview techniques. Test 1 is a short interview, while Test 2 contains an interview and other assessments. The company is interested in whether individuals' scores on Test 2 can be predicted from those on Test 1 (which would mean that the extra time taken to administer Test 2 is unnecessary). A scatterplot of the scores on the two tests for 15 individuals is shown below.



How would you describe the association between the two test scores?

- (a) * Positive and more or less linear
 - (b) Negative and more or less linear
 - (c) Negative and definitely curved
 - (d) Positive and definitely curved
 - (e) There is no association between the test scores.
14. For 50 children, the ages were recorded (in months) at which they were first enrolled in preschool. A normal quantile plot is shown below for these data.

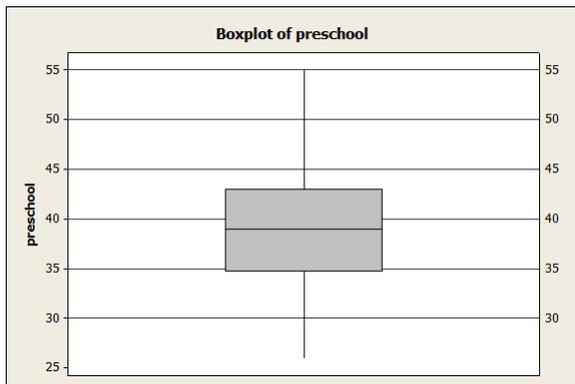


What do you conclude from this plot?

- (a) The data are skewed to the left.
- (b) There is a strong linear relationship between age and gender.

- (c) * A normal distribution describes these data reasonably well.
- (d) The data are skewed to the right.
- (e) There are outliers.

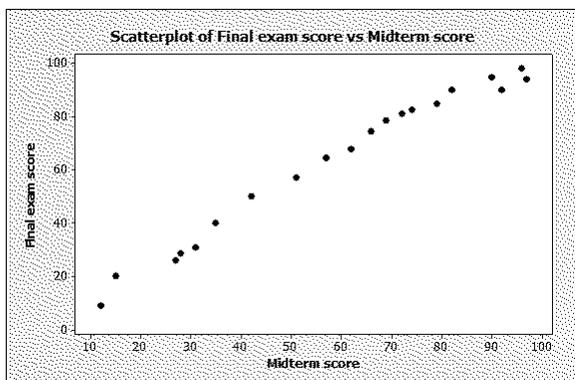
15. A boxplot is drawn for the data in Question 14. The boxplot is shown below. What is the interquartile range for these data?



- (a) 39
- (b) * 8
- (c) 43
- (d) 15
- (e) 35

16. The relationship between midterm score and final exam score in a certain course was examined. The data was recorded on a sample of 20 students selected from this class. Use this information for this question and the following 4 questions.

A scatterplot of these data is shown below.



According to the plot, the relationship between the midterm score and final exam scores can be described as

- (a) there is no relationship between these two variables.
- (b) * positive linear association
- (c) negative linear association
- (d) positively skewed

17. Refer to the information given in Question 16. Some Minitab output from a regression predicting final exam score from midterm score is shown below.

Regression Analysis: Final exam score versus Midterm score

The regression equation is

Final exam score = 2.62 + 1.03 Midterm score

Predictor	Coef	SE Coef	T	P
Constant	2.619	2.334	1.12	0.277
Midterm score	1.02811	0.03613	28.46	0.000

S = 4.30358 R-Sq = 97.8% R-Sq(adj) = 97.7%

The intercept and slope of this regression line are:

- (a) slope 0.036 and intercept 2.334
 - (b) slope 2.334 and intercept 0.036
 - (c) slope 2.62 and intercept 1.03
 - (d) * slope 1.03 and intercept 2.62
18. Refer to the output shown in Question 17. According to that output, if the midterm score increases by one mark, the final exam score is predicted to:
- (a) not change
 - (b) increase by 2.62 marks
 - (c) * increase by 1.03 marks
 - (d) decrease by 2.62 marks
 - (e) decrease by 1.03 marks
19. Refer to the output shown in Question 17. The percent of variation in the final exam scores that can be explained by the regression analysis above is
- (a) 87.9%
 - (b) 80.0%
 - (c) *97.8%
 - (d) 67.8%
 - (e) 98.9%
20. Refer to the output shown in Question 17. The correlation between final exam scores and midterm scores is
- (a) 0.724
 - (b) * 0.989
 - (c) 0.879
 - (d) 0.978
21. Leonardo da Vinci believed that a person's height could be predicted from a person's armspan (the distance, measured across the back, between that person's fingertips when the arms are held straight and horizontally). To test this belief, height and armspan were measured for 8 people, with the results shown below:

Descriptive Statistics: armspan, height

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
armspan	8	0	166.36	3.25	9.18	153.00	157.65	168.90	174.65
height	8	0	166.70	2.80	7.91	157.50	158.13	167.65	174.03

Variable	Maximum
armspan	176.50
height	177.80

Correlations: armspan, height

Pearson correlation of armspan and height = 0.946
P-Value = 0.000

What is the slope of the regression line for predicting height from armspan?

- (a) 31.10
 - (b) 15.18
 - (c) * 0.82
 - (d) -17.00
 - (e) 1.10
22. A data set about some students is shown below. Use the data set for this question and the following question.

Student	Gender	Major	GPA
1	female	Psychology	3.8
2	male	Biology	3.2
3	female	Chemistry	2.9
4	female	Sociology	3.6
5	male	Biology	2.7

GPA is which of the following?

- (a) a categorical variable
 - (b) * a quantitative variable
 - (c) an individual
23. In the data set of Question 22, Student 3 is which of the following?
- (a) * an individual
 - (b) a categorical variable
 - (c) a quantitative variable
24. The time it takes students to complete a STAB22 final exam is normally distributed with a mean of 160 minutes and standard deviation of 15 minutes. How much time should be allowed if we wish to ensure that 90% of the students can complete it? (round to the nearest minute)
- (a) * 179 minutes
 - (b) 177 minutes
 - (c) 182 minutes

- (d) 181 minutes
25. A study was conducted to investigate the effects of sleep deprivation on ability to solve problems. 10 subjects took part in an experiment. Each subject was deprived of sleep for a certain number x of hours, and the number of errors made in solving a set of simple addition problems y was recorded. Use this information for this question and the following question.
- The regression equation turned out to be $y = 3 + 0.475x$. In this equation, what is the interpretation of the number 0.475?
- (a) the average increase in the number of hours of sleep deprivation when the number of errors increases by 1.
(b) the predicted number of errors for a person who was not deprived of sleep at all.
(c) * the average increase in the number of errors when the number of hours of sleep deprivation is increased by 1.
(d) the predicted number of hours of sleep deprivation for a person who made no errors.
26. In the study of Question 25, what is the predicted number of errors made by a person who is deprived of sleep for 10 hours?
- (a) * 7.75
(b) 14.74
(c) 10.25
(d) cannot answer this question from the regression given in Question 25
(e) 5.25
27. Scores on a test have mean 70 and standard deviation 15. Let x denote a student's score on this test. The professor wants to adjust the scores so that they have mean 75 and standard deviation 12. Let y denote a student's adjusted score. y is related to x by $y = a + bx$, and the professor wants to find a and b so that the adjusted scores have the correct mean and standard deviation. What does a have to be?
- (a) 5
(b) 0.8
(c) * 19
(d) -10
(e) 12
28. The number of times an adult human breathes per minute while at rest varies considerably from one person to another. Suppose the number of breaths has approximately a normal distribution with mean 16 and standard deviation 4. Approximately what percentage of adults take between 4 and 28 breaths per minute while at rest?
- (a) 2.5%
(b) * 99.7%
(c) 68%
(d) 16%
(e) 95%
29. Scores on a test have a normal distribution with mean 20. 95% of the test scores are between 13 and 27. What is the standard deviation of test scores?

- (a) 5.5
 (b) 7
 (c) * 3.5
 (d) 1
30. The mean price of Nike shoes is \$100 with a standard deviation of \$15. During the holiday season, Nike had a promotion of a 20% discount on any pair of shoes. The mean and standard deviation of Nike shoes during the holiday season are
- (a) mean 100 and SD 15
 (b) mean 120 and SD 18
 (c) mean 80 and SD 15
 (d) *mean 80 and SD 12
31. The prices of 19 different brands of walking shoe were recorded. A stemplot of the prices is shown below. Use this information to answer this question and the 2 following questions.

Stem-and-leaf of shoeprice N = 19
 Leaf Unit = 1.0

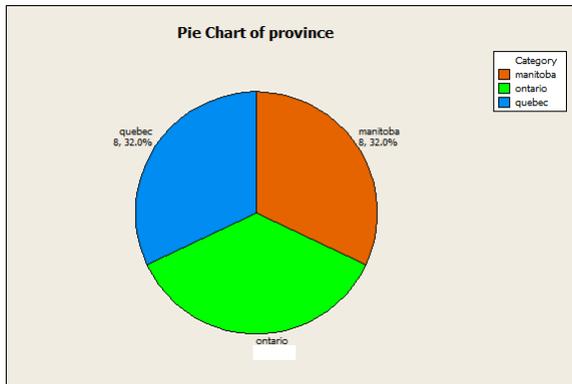
```

1  4  0
1  4
1  5
1  5
2  6  0
9  6  5557889
(5) 7  01234
5  7  556
2  8
2  8
2  9  0
1  9  5

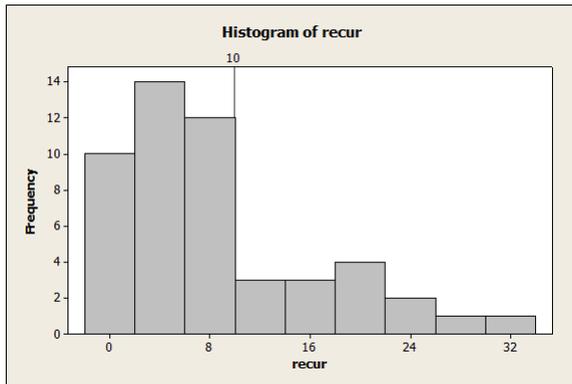
```

- What is the median price of these walking shoes?
- (a) impossible to find the median from a stemplot
 (b) * 70
 (c) 72
 (d) 7.0
32. Using the information in Question 31, what is the 3rd quartile (Q3) of the prices of walking shoes?
- (a) 7.4
 (b) * 75
 (c) 74
 (d) 7.5
33. In addition to the information in Question 31, the interquartile range of shoe prices is 10, and the first quartile is 65. Using the standard rule for outliers, any data values lower than what value would be considered outliers?
- (a) 9.0

- (b) * 50
 (c) 6.0
 (d) 40
 (e) 90
34. Sarah got 87 on her STAB22 final exam where the class average was 77 and standard deviation 10 while in PSY257 she got 75 and the class average was 65 with standard deviation 8. In which course did she do better, relative to the rest of the class?
- (a) STAB22.
 (b) She performed the same in both courses.
 (c) *PSY257.
35. A manufacturer of jeans has plants in Ontario, Quebec and Manitoba. A sample of 25 jeans sold is taken, and for each pair of jeans in the sample, the manufacturing plant is recorded. A pie chart of the results is shown below. Use this information for this question and the one following. Unfortunately, a manager was careless with his cup of coffee, and some of the information for Ontario was lost.



- What percentage of jeans in the sample was manufactured in Ontario?
- (a) * 36
 (b) 9
 (c) 33
 (d) 32
 (e) 8
36. For the data of Question 35, which of the following graphs could also be used to display these data?
- (a) * bar chart
 (b) stemplot
 (c) histogram
 (d) scatter plot
 (e) boxplot
37. The length of time (in months) was recorded for 50 patients between the onset of a particular illness and its recurrence. A histogram of the results is shown below. Use this histogram to answer this question and the two questions following.



How would you describe the shape of the distribution of recurrence times?

- (a) * Skewed to the right
 - (b) Bell-shaped
 - (c) The distribution has outliers
 - (d) Symmetric
 - (e) Skewed to the left
38. Look again at Question 37. The value 10 months is marked (at the top of the plot). What percentage of patients had a recurrence time less than (or equal to) 10 months?
- (a) 36
 - (b) * 72
 - (c) 48
 - (d) 24
 - (e) 12
39. Look again at Question 37. What can you say about the median recurrence time?
- (a) cannot say anything about the median by looking at the histogram
 - (b) more than 10 months
 - (c) somewhere around 4 months
 - (d) * more than 6 months