

University of Toronto Scarborough STAB22 Midterm Examination

November 2007

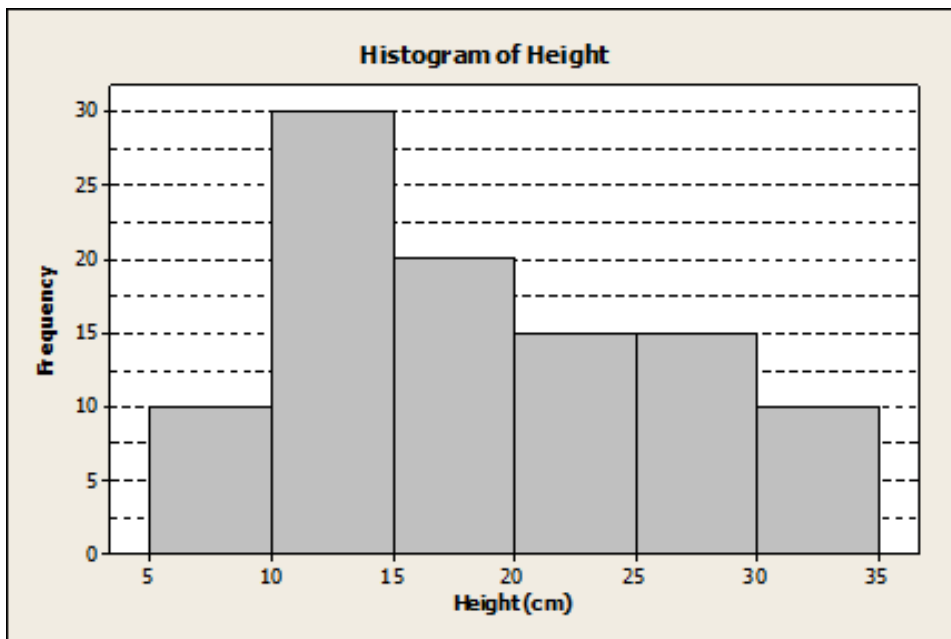
For this examination, you are allowed one handwritten letter-sized sheet of notes (both sides) prepared by you, a non-programmable, non-communicating calculator, and writing implements.

This examination has 13 numbered pages, plus copies of Table A (normal distribution) and Table C (binomial distribution) which are included at the back of this exam. Before you start, check to see that you have all pages and tables.

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), student number (in “Identification”) and tutorial number (in “Grade or Educ”). For your reference, a listing of tutorials is below.

1	Tue 14:00	6	Wed 13:00	13	Thu 14:00
2	Tue 16:00	7	Wed 15:00	9	Thu 15:00
3	Tue 17:00	11	Wed 16:00	10	Thu 16:00
4	Tue 18:00	8	Thu 12:00	14	Fri 11:00
5	Wed 12:00	12	Thu 13:00	15	Fri 12:00

1. A description of different houses on the market includes the following three variables. Which of these variables is/are quantitative?
 - (a) The square footage of the house
 - (b) The monthly gas bill
 - (c) The monthly electric bill
 - (d) All of the above
2. A histogram of the heights of 85 plants is shown below. The histogram will be used for this question and the following two questions.



- Approximately what percent of the plants in this sample are taller than 20 cm?
- (a) 55%
 - (b) 47%
 - (c) 37%
 - (d) 30%
 - (e) 15%
3. Again looking at the histogram in Question 2, which of the following will provide a better summary of the heights of these 85 plants?
 - (a) The five number summary
 - (b) The mean and the standard deviation

4. Some Minitab output is given below. It gives the stemplot and some summary statistics of IQ scores of a group of students. Use the information shown to answer this question and the three questions following.

Stem-and-Leaf Display: IQ

Stem-and-leaf of IQ N = 78
Leaf Unit = 1.0

```
2  7  24
4  7  79
4  8
6  8  69
10 9  0133
14 9  6778
24 10 0022333344
36 10 555666777789
(19) 11 0000111122223334444
23 11 55688999
15 12 003344
9  12 677888
3  13 02
1  13 6
```

Descriptive Statistics: IQ

Variable	N	N*	SE Mean	StDev	Q1	Q3
iq	78	0	1.49	13.17	103.00	118.00

Looking at the stemplot, which of the following numbers is the closest to the median IQ score?

- (a) 108
 - (b) 110
 - (c) 114
 - (d) 116
 - (e) 118
5. Looking at the output given in Question 4, which of the following numbers is the closest to the IQR of the IQ score?
- (a) 10
 - (b) 13
 - (c) 15
 - (d) 103
 - (e) 118

6. Looking again at the output given in Question 4, how many of these IQ scores are identified as outliers, using the standard rule?
- (a) No outliers.
 - (b) Only one outlier
 - (c) Only two outliers
 - (d) Only three outliers
 - (e) More than three outliers
7. We want to display the distribution of IQ scores, shown in Question 4, using another graphical display (other than the stemplot shown). Which of the following displays is the most appropriate for displaying the distribution of these IQ scores?
- (a) a pie chart
 - (b) a histogram
 - (c) bar chart
8. What is the median of the numbers 4, 8, 7, 2, 2?
- (a) 7
 - (b) 4.6
 - (c) 4
 - (d) 2
9. A set of numbers has mean 10 and standard deviation 3. A new set of numbers is created from this set by subtracting 2. What are the mean and standard deviation of the new set of numbers?
- (a) Mean is 8, SD is 1.
 - (b) Mean is 8, SD is 3.
 - (c) Mean is 10, SD is 1.
 - (d) Mean is 10, SD is 3.
 - (e) Cannot tell from the information available.

10. Which of the following sets of four numbers has the largest standard deviation?

(a) 100, 101, 102, 103

(b) 7, 8, 9, 10

(c) 5, 5, 5, 5

(d) 0, 1, 2, 3

(e) 0, 0, 10, 10

11. The time taken by students to complete a STAB22 test has a normal distribution with mean 90 minutes and standard deviation 12 minutes. Use the normal distribution tables to answer this question and the two following questions, which are based on this information.

If the students are given 1 hour and 45 minutes to complete the exam, what proportion of students will complete the exam within this given time? Mark the closest answer among the alternatives given.

(a) 0.70

(b) 0.75

(c) 0.80

(d) 0.85

(e) 0.90

12. Using the information in Question 11, what proportion of students will take longer than 95 minutes to complete the test? Mark the closest answer among the alternatives given.

(a) 0.05

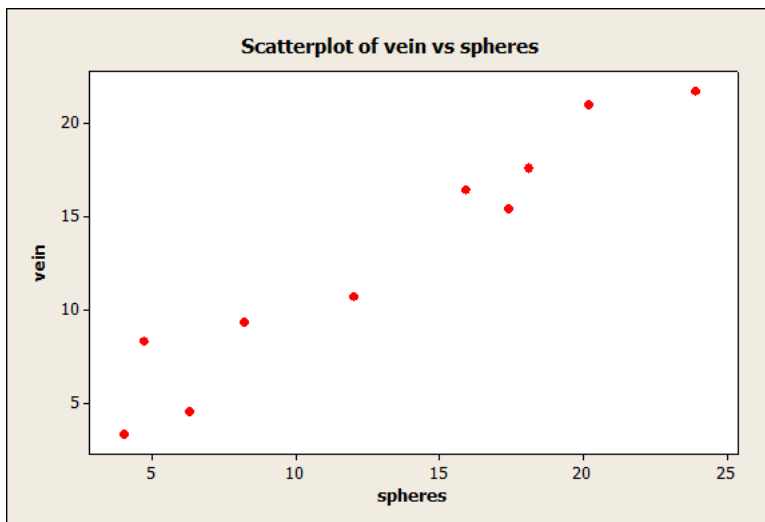
(b) 0.25

(c) 0.34

(d) 0.50

(e) 0.68

13. Using the information in Question 11, we want to determine the length of time we should allow for the test so that 95% of the students are able to complete the exam in the time allowed. Which of the following is the closest to this time?
- (a) 80 minutes
 - (b) 90 minutes
 - (c) 95 minutes
 - (d) 100 minutes
 - (e) 110 minutes
14. Research on digestion requires accurate measurements of blood flow through the lining of the stomach. One way to achieve this is to inject microscopic spheres into the blood stream. The spheres lodge in blood vessels at a rate proportional to blood flow. Medical researchers compared blood flow in the stomachs of dogs, measured by the microscopic spheres (“spheres”) and directly using a catheter inserted into a vein (“vein”). A scatterplot of the results is shown below.



How would you describe the relationship between “vein” and “spheres”?

- (a) There is no apparent relationship between the two variables.
- (b) The relationship is clearly curved.
- (c) The relationship is approximately a straight line.
- (d) We need to calculate the regression line before we can comment on the form of the relationship.

15. The correlation between two variables x and y is -0.7 . Which of these statements is true?
- (a) A larger value of x is the cause of a larger value of y .
 - (b) A larger value of x is the cause of a smaller value of y .
 - (c) Larger values of x tend to go with larger values of y .
 - (d) Larger values of x tend to go with smaller values of y .
16. It is believed that applying zinc to soil can affect the amount of herbage (leaves and stems) of plants grown. In an experiment, the regression line for predicting the amount of herbage (y) from the amount of zinc applied (x) is $y = 3.29 - 3.825x$. Use this information for this question and the following one.
- Which of the following can you conclude from this regression line?
- (a) When the amount of zinc is zero, the amount of herbage is predicted to be 3.29 units.
 - (b) When the amount of zinc is zero, the amount of herbage is predicted to be -3.825 units.
 - (c) There is an error because -3.825 is less than -1 .
 - (d) When the amount of zinc is increased by one unit, the amount of herbage is predicted to increase by 3.29 units.
 - (e) When the amount of zinc is increased by one unit, the amount of herbage is predicted to increase by 3.825 units.
17. Using the information in Question 16, what is the predicted amount of herbage when the amount of zinc is 0.5? Choose the closest answer from the alternatives below.
- (a) 3.3
 - (b) 0.0
 - (c) 4.2
 - (d) 0.7
 - (e) 1.4

18. A political scientist believes that, in the United States, women are more likely than men to vote Democrat. The political scientist selects a large sample of registered voters, both men and women, and asks each person sampled whether they voted Democrat or Republican at the last election. Which of the statements below applies to this situation?

- (a) This experiment will enable the political scientist to tell whether a person's gender is the cause of their voting for a particular party.
- (b) This is an observational study because no treatment was imposed on the people sampled.
- (c) A small sample would have been just as good as a large sample, and much less expensive to collect.
- (d) In this study, the response is gender and the explanatory variable is political party voted for.

19. Sickle cell disease affects the red blood cells. It can cause severe pain and other complications. A study was done on 300 (randomly chosen) sickle cell sufferers; 150 of these (randomly chosen) were given the drug hydroxyurea, and the other 150 were given a placebo. The number of episodes of pain was recorded for each sickle cell sufferer.

Use the information given above to answer this question and the two questions following.

In this study, what are the experimental units or subjects?

- (a) The patients who were given hydroxyurea
- (b) The patients who received the placebo
- (c) The patients suffering from sickle cell disease
- (d) The number of episodes of pain
- (e) Sickle cell disease

20. Read again the information in Question 19. In this study, what are the treatments?

- (a) Hydroxyurea and placebo
- (b) Having sickle cell disease and not having the disease
- (c) The number of episodes of pain
- (d) The patients suffering from sickle cell disease
- (e) The patients who were given hydroxyurea

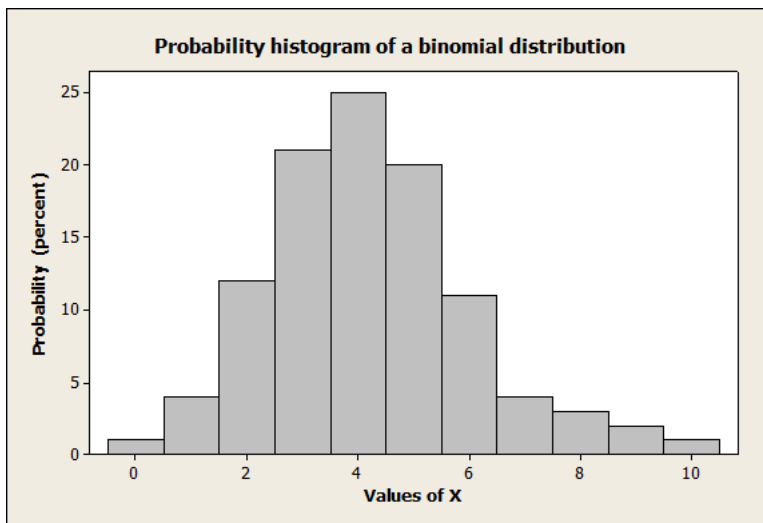
21. In the study of Question 19, what is the response variable?
- (a) The number of episodes of pain
 - (b) The drug hydroxyurea
 - (c) Having or not having sickle cell disease
 - (d) Red blood cell count
 - (e) Whether the patients received hydroxyurea or a placebo
22. A university has 2000 male and 500 female faculty members. The Human Resources department is commissioning a survey on childcare, and wants to sample 100 faculty members. What kind of sampling method would best represent the possibly differing opinions of males and females?
- (a) Simple random sample
 - (b) Stratified sample
 - (c) Multistage sample
 - (d) Systematic sample
23. A telemarketing company in Denver uses equipment that dials residential phone numbers in that city at random. Of the first 200 numbers dialled, 67 are unlisted. This is not surprising because 30% of all Denver residential phone numbers are unlisted. Are the numbers 67 and 30% parameters or statistics?
- (a) 67 is a parameter and 30% is a statistic.
 - (b) 30% is a parameter and 67 is a statistic.
 - (c) 67 and 30% are both parameters.
 - (d) 67 and 30% are both statistics.
24. A psychologist studied the effect of running on personality, by giving a questionnaire to a random sample of 231 runners (who run more than 30 km per week). A news report on the study said “The researcher found statistically significant personality differences between the runners and the 30-year-old male population as a whole”. What does “statistically significant differences” mean here?
- (a) The differences were meaningful in terms of personality.
 - (b) The differences could have occurred by chance because of the random sampling.
 - (c) Larger differences were observed than would be expected by chance alone.
 - (d) Because the runners were chosen at random, we cannot judge how big a difference between the runners and the other males there might be due to chance.

25. Two events A and B are such that $P(A) = 0.3, P(B) = 0.4, P(A \text{ and } B) = 0.12$. Are A and B independent, and why (or why not)?
- (a) Yes, because $P(A) \times P(B) = P(A \text{ and } B)$.
 - (b) No, because $P(A) + P(B)$ is not equal to $P(A \text{ and } B)$.
 - (c) Cannot say, because we don't know whether A and B are disjoint.
 - (d) Cannot say, because we need to know what A and B represent in order to figure out whether they are independent.
26. The probability that any one person develops a certain disease is 0.1. People develop the disease independently of other people. What is the probability that, out of a group of 3 people, at least 1 develops the disease?
- (a) 0.9
 - (b) 0.1
 - (c) 0.4
 - (d) 0.271
 - (e) 0.729
27. A random variable W takes values 1 and 2; $P(W = 1) = 0.7, P(W = 2) = 0.3$. Which of the values below is closest to the mean (expected value) of W ?
- (a) 0.3
 - (b) 0.7
 - (c) 1.0
 - (d) 1.3
28. A random variable X has $P(X = 0) = 0.6, P(X = 1) = 0.4$. The mean of X is 0.4. Which of the values below is the closest to the variance of X ?
- (a) 0.24
 - (b) 0.50
 - (c) 0.40
 - (d) 0.60

29. A young driver is going to take the driving test repeatedly until he passes. Let X be the number of tests taken by the young driver. Does X have a binomial distribution, and if not, what is the best reason why?

- (a) Yes, X has a binomial distribution.
- (b) No, because the number of trials n is not fixed.
- (c) No, because the observations do not fall into two categories.
- (d) No, even though p is the same for each observation.

30. A probability histogram of a binomial distribution is shown below.



What are the number of trials n and the success probability p for this binomial distribution?

- (a) $n = 10, p = 0.4$
- (b) $n = 10, p = 0.5$
- (c) $n = 20, p = 0.2$
- (d) $n = 20, p = 0.4$
- (e) $n = 100, p = 0.04$

31. A small sample of size $n = 6$ is to be taken from a large population in which the proportion of “successes” is 0.40. Use this information for this question and the following question.

What is the probability of observing exactly 3 successes in the sample?

- (a) 0.8208
 - (b) 0.4672
 - (c) 0.3110
 - (d) 0.2765
32. Using the information given in Question 31, what is the probability of observing 2 successes or fewer in the sample?
- (a) 0.7782
 - (b) 0.5443
 - (c) 0.3110
 - (d) 0.2333
33. Suppose you are considering using the normal approximation to the binomial. Which of the cases below would give you the best approximation?
- (a) $n = 10, p = 0.5$
 - (b) $n = 100, p = 0.49$
 - (c) $n = 100, p = 0.6$.
 - (d) $n = 120, p = 0.1$
34. Suppose that a population has exactly a normal-distribution shape. A sample of size n is taken from this population. What is the sampling distribution of the sample mean?
- (a) A normal distribution approximately.
 - (b) An approximate normal distribution if n is large.
 - (c) A normal distribution exactly.
 - (d) It is impossible to say anything about the sampling distribution.
35. Suppose that a sample of size n is taken from a population whose shape is not known. What is the sampling distribution of the sample mean?
- (a) A normal distribution approximately.
 - (b) An approximate normal distribution if n is large.
 - (c) A normal distribution exactly.
 - (d) It is impossible to say anything about the sampling distribution.

36. On a certain standardized test, the mean score is 25, and the standard deviation of scores is 5, and scores have a normal-distribution shape. A random sample of 4 students is taken. What is the probability that the mean score of these 4 students will be 28 or bigger? Choose the closest answer from the alternatives given below.
- (a) 0.12
 - (b) 0.27
 - (c) 0.50
 - (d) 0.73
 - (e) 0.88