

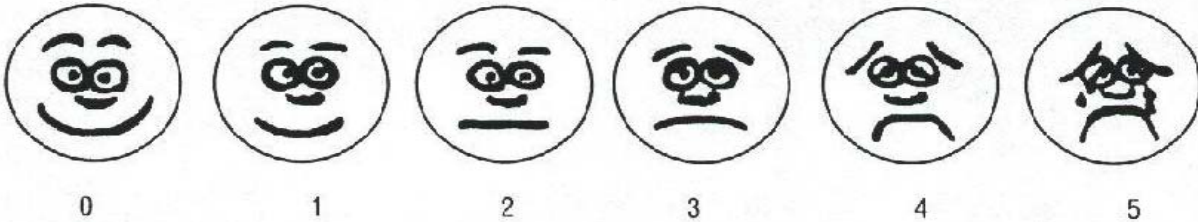
**E370 Exam Two – Evening  
Fall, 2004-05  
Version Blue**

**INSTRUCTIONS**

1. The only items allowed within your reach during an exam are calculators, the exam paper, pencils, erasers, a highlighter pen and your tool cards. Take off your hat and put it under your seat.
2. In the upper left corner of your NCS answer sheet **PRINT YOUR LAST NAME**. Skip a square and print your first name. Bubble in the corresponding spaces beneath those letters.
3. **DO NOT FILL IN YOUR STUDENT ID NUMBER.**
4. In spaces KLMNO, **PRINT THE FIVE DIGITS OF YOUR TEAM NUMBER**. Bubble in the corresponding spaces beneath those numbers.
5. **WHOEVER PERFORMS #2, 3 & 4 ABOVE CORRECTLY WILL RECEIVE ONE POINT CREDIT!!!!**
6. Remember, a student is to avoid even the appearance of cheating. Keep your eyes on your exam or on the ceiling. If any member of the teaching team observes questionable behavior on your part, he or she has the right to confiscate your exam and ask you to leave the room.
7. **ANY** talking between students will be interpreted as cheating and all parties will fail the course.
8. Absolutely **ALL** cell phones must be turned off and out of reach. **ANY** cell phone usage for any purpose will be interpreted as cheating and the user will fail the course.
9. **ANY TOOL CARDS WHICH HAVE PHOTOCOPIED, COMPUTER GENERATED OR TYPE-WRITTEN INFORMATION ATTACHED TO THEM WILL BE CONFISCATED.**
10. There are 33 multiple choice questions on this exam. Each question is worth 3 points. **NOTE: EVEN THE HARD QUESTIONS ARE ONLY WORTH 3 POINTS!**
11. Don't let yourself get stuck on one question. Get all the answers you are sure of, then go back to the ones you are not sure of.
12. I believe that there is only one completely correct answer to each question on this test. Look for it and select it as your best choice.
13. Stay calm and do your best!

1. Which of the following statements about the mean is not always correct?
  - A. The mean is a measure of the center of a distribution.
  - B. The sum of the deviations from the mean is zero.
  - C. The value of the mean times the number of observations equals the sum of the observations.
  - D. The mean is the most common observation in a population.
  
2. The human resources director of a large corporation wishes to develop an employee benefits plan and selected 500 employees from the list of all 40,000 employees in order to study their preferences for the various components of a potential package. The proportion who preferred a Point of Service plan was 0.58. The companies insurance provider reports such plans typically cost \$375 per person. Which of the numbers mentioned is/are statistics?
  - A. 500 and 40,000
  - B. 40,000 and \$375
  - C. \$375 and 0.58
  - D. 0.58

3. When students go to the Health Center these days complaining of pain they are asked to assess their pain using the following emoticons:



One person's assessment is an example of what kind of data?

- A. Categorical-Nominal
  - B. Categorical-Ordinal
  - C. Numerical-Discrete
  - D. Numerical-Continuous
- Use this information to answer the next **THREE** questions. The monthly amounts spent for food by families of four receiving food stamps approximates a symmetrical distribution. The mean is \$100 and the variance is \$400. **NOTE: The questions continue onto the next page.**
4. Which is wider, the range over which Chebyshev predicts at least 68% or the range over which the Empirical Rule predicts about 68%?
    - A. The range predicted by Chebyshev's Rule is wider.
    - B. The range predicted by the Empirical Rule is wider.
    - C. The ranges are the same size.
    - D. It is not possible to perform this calculation.

5. According to Chebyshev's Theorem, within what values will at least 55% of the families' monthly food expenditures fall?
- \$56 and \$144
  - \$0 and \$400
  - \$70 and \$130
  - Chebyshev's Theorem is not appropriate in this instance.
6. Using the Empirical Rule, the range over which approximately 95% of families' monthly food expenditures falls is
- \$60 and \$140
  - \$0 and \$400
  - \$80
  - \$120
7. A study of the scores on an in-plant course in management principles and the years of service of the employees enrolled in the course yielded the following statistics:
- |                   |          |                |
|-------------------|----------|----------------|
| Test Scores:      | mean=900 | variance = 196 |
| Years of Service: | mean=9   | variance = 16  |
- Of test scores and years of service, which measure has the greater dispersion?
- Years of service
  - Test scores
  - Both test scores and years of service have the same relative dispersion.
  - It is impossible to tell.
8. Which of the following statements is **TRUE?**
- The sum of cumulative percentages in a distribution always equals one.
  - The sum of relative frequencies in a distribution always equals one.
  - The sum of cumulative frequencies in a distribution always equals one.
  - None, that is, they are all false.
9. The correlation coefficient between two variables X and Y is 0.73. If each observation of the variable X increases by 3 and if each observation of Y doubles, then the new correlation coefficient between X+3 and 2Y is
- 0.365
  - 1.46
  - 0.146
  - 0.73
10. Which of the following statements is **FALSE?**
- A summary measure calculated from all items of interest is a parameter.
  - A summary measure calculated from some of the items of interest is a statistic.
  - The main goal of descriptive statistics is to estimate characteristics of populations.
  - Statistical inference is the process of making a prediction based on only a small piece of information.
11. A stratified random sample is considered appropriate when which of the

following conditions holds?

- A. The population from which the sample is drawn consists of several heterogeneous groups.
  - B. The population from which the sample is drawn consists of several homogeneous groups.
  - C. When the strata possess identical means and standard deviations.
  - D. When no more than three strata can be identified.
12. The host of a dinner dance reached into a bowl, mixed all the tickets around, and selected the ticket to award the grand door prize. What sampling method was used?
- A. cluster sample
  - B. stratified sample
  - C. systematic sample
  - D. simple random sample

The following stem and leaf diagram represents the scores on a calculus exam for a sample of 40 students. Use this display to answer the next ***THREE*** questions.

```

3 | 6 8
4 | 1 2 7 8
5 | 1 4 4 5 6 7 8 9
6 | 1 2 2 3 4 5 7 8 8 8 9
7 | 1 1 4 5 6 9 9 9
8 | 1 2 3 4 6
9 | 1 4

```

13. If scores between 70 and 79 were C's, what is the mean score of students earning a C?
- A. 75.00
  - B. 75.25
  - C. 75.50
  - D. 76.38
  - E. 76.50
14. What is the Interquartile Range of this sample?
- A. 20.5
  - B. 23.0
  - C. 24.5
  - D. 30.75
15. If the data from the stem and leaf diagram above were graphed, which of the following types of graphs would show the Interquartile Range?
- A. Pareto diagram
  - B. histogram
  - C. pie chart
  - D. ogive

On September 30, 2004 CBS.com reported results of their on-line poll taken immediately after the close of the first debate between President Bush and Senator

Kerry. The following table contains part of those results, specifically the age groups and preferences between the two participants in the debate of the registered voters who participated in the poll. The observations in the cells are absolute frequencies of undecided voters. Use this information to answer the next **FOUR** questions.

<b>CBS.com Post Debate Poll Results</b>				
<b>Age of Undecided Voter by Debate "Winner"</b>				
	<b>Bush</b>	<b>Kerry</b>	<b>Undecided</b>	<b>Total</b>
<b>18-23 year olds</b>	3		7	
<b>23 to 65 year olds</b>	40	41	32	
<b>65 to 75 year olds</b>	11		21	72
<b>Total</b>	54	86		200

16. The sample above is definitely plagued with which of the following errors associated with sampling?
- Sampling error and error from self-selection bias
  - Error from interviewer bias and selection bias
  - Error from non-response bias, selection bias, and sampling error.
  - Non-sampling error only.
17. To the nearest whole year, estimate the mean age of the voters who participated in the poll.
- 43
  - 45
  - 47
  - 51
18. If the undecided voter's age is independent of his or her opinion of the debate's winner, how many voter's would you expect would be at least 65 and thought Kerry won the debate? (To the nearest whole voter.)
- 20
  - 31
  - 40
  - 43
19. Which is more likely, that a random voter was undecided about the debate winner given that the voter was between 18 and 23 years of age, or that a random voter was at least 65 years old given that the voter was undecided about the winner of the debate?
- The two likelihoods are identical.
  - The oldest voters among the undecideds is more likely.
  - The undecided voters among the youngest voters is more likely.
  - There is insufficient information to answer this question.
20. The line of best fit as determined by Excel is considered the best fit because

- A. the sum of squared distances between observations and predictions is as small as possible.
- B. the sum of distances between observations and predictions is as small as possible.
- C. it passes through the highest number of observations in the data set.
- D. it is the line that has the fewest squared distances between observations and predictions.

21. Which of the following is **FALSE?**

- A. If a measure of central tendency is to be compared with averages already determined for other distributions, it must be the same measure as those with which it is being compared.
- B. One problem with the mode is that it is affected by the numerical magnitude of the extreme data points in a data set.
- C. Some distributions may have more than one mode, but no distributions will have more than one median and one mean.
- D. A mode may be determined for qualitative data for which it is not practical or possible to determine a mean or median.

22. Based on the following data values {7, 5, 6, 4, 8, 8, 12}, which of these statements is **TRUE?**

- A. median = mode < mean
- B. mean = mode > median
- C. mean = median < mode
- D. median < mean < mode

23. A sample of 20 observations has a variance of 9. The sum of the squared deviations from the sample mean is

- A. 20                      B. 57                      C. 60                      D. 171

A study of factors affecting corn yield was conducted by Purdue University. The study yielded the following information. The covariance between rainfall amounts and corn yield is 1118.57 bushel inches. The covariance between fertilizer amounts and corn yield is 333.39 bushel pounds. The covariance between rainfall and fertilizer amounts is 835 pound-inches. Rainfall was measured in inches, fertilizer in pounds, and corn yield in bushels. Use this information to answer the next **FOUR** questions.

**Note: the questions continue onto the next page.**

24. Between which of these pairs would you reject the idea of a true linear relationship?

- A. Fertilizer and Corn Yield
- B. Rainfall and Fertilizer
- C. Rainfall and Corn Yield
- D. Not enough information to tell.

25. The least squares line for corn yield and rainfall is **Bushels = 89.54 + 0.13(Inches)**. How much would corn yield change should rainfall decrease by 20 inches?
- A. - 2.6 bushels  
 B. + 2.14 bushels  
 C. -0.13 bushels  
 D. A starting point is required to calculate this.
26. According to this model, what will be the sign on the correlation coefficient?
- A. zero  
 B. positive  
 C. negative  
 D. Not enough information to tell.
27. Suppose that the study was repeated the following year and rainfall amounts doubled. How would this change affect the covariance of corn yield and rainfall?
- A. It will double.  
 B. It will quadruple.  
 C. It will not change.  
 D. Not enough information to tell.

A sample of families was asked how many pets they owned. Their responses are summarized in the following table.

# Pets	# Families
0	3
1	10
2	5
3	4
4	2
5	1

28. Suppose you created a histogram of this sample. Describe it.
- A. The histogram would have five columns with each column increasingly higher than the first one, peaking at 5. The histogram is very right skewed.
- B. The histogram would have five columns with a peak at 1 and the remaining columns would rapidly decrease to the low point at 5. The histogram is highly left skewed.
- C. The histogram would have six columns with a peak at 1, a sharp decline on the left, and a gradual decline on the right. The histogram is pretty right skewed.
- D. The histogram would have six columns with a peak at 2.5, and the other columns would very regularly decrease in height as one moved away from the peak.

The following is Excel output generated by Descriptive Statistics. It represents monthly values of Canadian Imports and Exports between 1974 and 2001 in constant dollars. Use the data to answer the next **FOUR** questions. **NOTE: the questions continue onto the next page.**

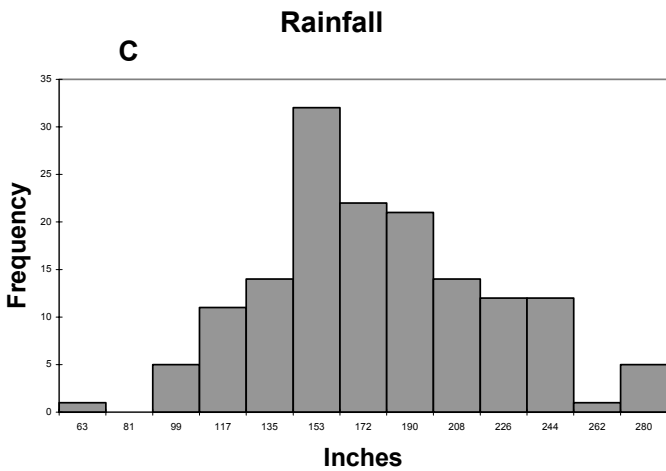
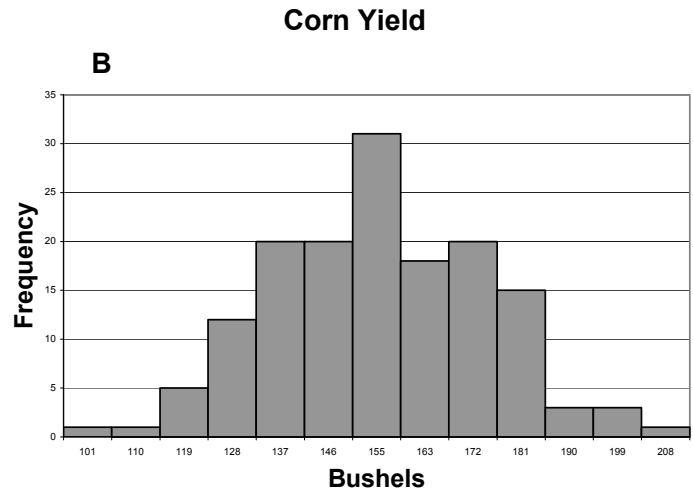
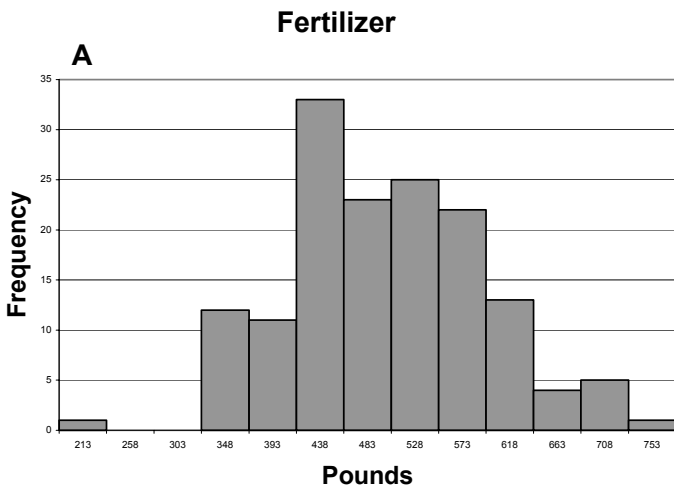
<i>Canadian Imports</i>		<i>Canadian Exports</i>	
Mean	6441.98	Mean	
Standard Error		Standard Error	273.91
Median	5215.50	Median	6159.47
Mode	2634.50	Mode	#N/A
Standard Deviation		Standard Deviation	4975.76
Sample Variance	17806397.81	Sample Variance	
Range	15689.20	Range	
Minimum		Minimum	1396.60
Maximum	17085.70	Maximum	20761.50
Sum		Sum	2509874.00
Count	330	Count	

29. Of Canadian Imports and Canadian Exports, which distribution is more skewed?
- Imports is more skewed.
  - Exports is more skewed.
  - The skewness of both variables is identical.
  - Since there is no mode for Imports, this question cannot be answered.
30. The data sets contain observations for all months between 1974 and 2001. Based on your knowledge of Excel, what would you expect the true standard deviation of Canadian Exports to be?
- 4975.76
  - Greater than 4975.76
  - Smaller than 4975.76
  - This answer requires the actual data to answer.
31. The Canadian government discovered a systematic record keeping error in the Exports data base. Each month's level of Exports was actually twice the size of the recorded level. For example, exports in January 1974 were actually \$3,500, not the \$1,750 recorded. Assuming the variance in the output above is that of the data set containing the error, how does the variance of the correct data set relate?
- The correct variance is four times as large as that reported.
  - The correct variance is one-quarter the size as that reported.
  - The correct variance is twice as large as that reported.
  - The correct variance is one-half the size as that reported.

32. Of the two variables, Canadian Exports and Canadian Imports, which is most unstable?

- A. Imports are more unstable.
- B. Exports are more unstable.
- C. The stability of Imports and Exports is identical.
- D. There is insufficient information to determine this.

33. Which of the following variables has the least skewness?



**E370 Fall 2004-05  
Exam Two Answers**

**Blue Version**

1 D	6 C	11 B	16 A	21 B	26 B	31 A
2 D	7 A	12 D	17 D	22 D	27 A	32 A
3 B	8 B	13 C	18 B	23 D	28 C	33 B
4 A	9 D	14 B	19 C	24 B	29 C	
5 C	10 C	15 D	20 A	25 A	30 C	