E370 Exam One – Evening
Fall, 2006-07
Version A

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.

2. Take off your hat and put it under your seat.

3. 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.

4. DO NOT FILL IN YOUR STUDENT ID NUMBER.

5. In spaces KLMNO, PRINT THE THREE DIGITS OF YOUR TEAM NUMBER. Bubble in the corresponding spaces beneath those numbers. **Do not leave any spaces between numbers.**

6. **WHOEVER PERFORMS #2, 3 & 4 ABOVE CORRECTLY WILL RECEIVE ONE POINT CREDIT!!!!!**

7. 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.

8. **ANY** talking between students will be interpreted as cheating and all parties will fail the course.

9. 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.

10. **ANY TOOL CARDS WHICH HAVE PHOTOCOPIED, COMPUTER GENERATED OR TYPE-WRITTEN INFORMATION ATTACHED TO THEM WILL BE CONFISCATED**.

11. 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!**

12. 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.

13. 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.

14. Stay calm and do your best!

1. The human resources director of a large corporation wishes to develop an employee benefits plan and selected 800 employees from the list of all 80,000 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 0.7. The companies insurance provider reports such plans typically cost $425 per person. Which of the numbers mentioned would be represented by Greek letters?

   A. $425
   B. 0.7
   C. $425 and 0.7
   D. 800 and 0.7

2. Which of the following is an example of performing statistical inference?

   A. Examine the weights of a sample of 10 candy bars to see if their average weight is 6 ounces.
   B. Examine the weights of a sample of 25 manufacturer parts to see if the average weight of all the parts produced by the process is 3 pounds.
   C. Calculate the mean number of fruit trees damaged by Mediterranean fruit flies in California last year.
   D. None of the above.

3. Of the following groups of variables, for which does the order of the variables follow the hierarchy of data types discussed in class?

   A. unemployment rate, postal zip codes, restaurant ratings
   B. number of insurance claims filed, Sports Illustrated's ranking of NFL teams, gallons of gas pumped.
   C. postal zip codes, number of insurance claims filed, restaurant rankings.
   D. numbers on football players' jerseys, restaurant rankings, number of insurance claims filed

4. Which of the following is false?

   A. If the values of the seventh and eighth class in a cumulative frequency distribution are the same, we know that there are no observations in the eighth class.
   B. Determining the class boundaries of a frequency distribution is highly subjective.
   C. The sum of relative frequencies in a distribution always equals one.
   D. Excel's class limits are the same as most statisticians.

5. Which of the following is NOT a mutually exclusive situation?

   A. When Event A and event B cannot occur at the same time.
   B. When Event A and Event B can occur at the same time.
   C. A frequency distribution with classes 0-10 and 10-20, with open upper limits and closed lower limits.
   D. A frequency distribution with classes 0-10 and 10-20, with open upper and lower limits.

6. A model of corn yield in bushels per acre based on total rainfall in inches was estimated. The least squares line generated was **Bushels = 0.77 Inches + 89.54**. One observation in the data set had a total rainfall of 80 inches and a corn yield of 171 bushels per acre. What is the value of the error for this observation, rounded to two decimals?

   A. 19.86
   B. 109.40
   C. 394.42
   D. Insufficient information to answer this question.

7. Why is it inappropriate to construct a Pareto Diagram out of interval data?

   A. Interval data does not have an inherent order.
   B. Interval data has an inherent order.
   C. Interval data cannot be put into classes.
   D. Interval data can be put into classes.

8. In what circumstances might scatter plots be useful?

   A. When you know two variables have a dependent relationship.
   B. When you think three variables have a dependent relationship.
   C. When you wonder if two variables have a dependent relationship.
   D. None of the above.

9. In a histogram, the proportion of the total area which must be to the left of the mean is

   A. 50%.
   B. >50%.
   C. <50%.
   D. not A, B, or C.
   E. A, B, or C.

10. The least squares line is the line guaranteed to

    A. be the one line of all possible lines around which the smallest square can be drawn.
    B. be the line of all possible lines that connects the most observations with the fewest turns.
    C. be the line of all possible lines that has the smallest squared sum of the distance between observations and predictions.
    D. be the line of all possible lines that has the smallest sum of squared distance between observations and predictions.

11. Which of the following is NOT a collectively exhaustive situation?

    A. When Event A and event B cannot occur at the same time.
    B. When Event A and Event B can occur at the same time.
    C. A frequency distribution with classes 0-10 and 10-20, with open upper limits and closed lower limits.
    D. A frequency distribution with classes 0-10 and 10-20, with open upper and lower limits.

12. Which one of the following statements is FALSE?

    A. Pie charts are the best type of chart for comparing two sets of categorical data.
    B. Data that are nominal scale are presented using summary tables.
    C. Means and standard deviations of ordinal data are meaningless.
    D. The scatter-plot is the basic graphic tool for investigating relationships between two interval variables.

13. Below is data on the leaders in the number of sacks in the NFL during the 2005 Regular Season. (Source: www.nfl.com)

Sacks in the NFL

Number of Sacks	Number of players
8 to 11	20
11 to 14	7
14 to 17	3

What is the mean number of sacks by the league's leaders?

A. 9.5 B. 10.8 C. 12.5 D. 11.2

The following data, in the form of a stem-and-leaf plot, represents the hourly wage (in US$) earned by a sample of railway employees. Use this information to answer the following ***TWO*** questions.

2	2 7 5
3	4 9 7 6
4	9 7 3 2
5	0 2 3 2 5 7
6	2 1 0

14. What is the Inter-Quartile Range of this sample?

A. 18.25 B. 10.5 C. 36.25 D. 54.5

15. If the stem-and-leaf plot were transformed into a frequency distribution with five classes, what would be the upper limit on the fourth class?

A. 57.9 B. 59 C. 59.9 D. 60

16. Which of the following statements about the median is WRONG?

A. The median can be calculated for ordinal data.
B. The median is unique to a data set.
C. The sum of the deviations from the median is zero.
D. The median is not sensitive to outliers.

17. In a left-skewed distribution, which of the following is most likely?

A. The distance from the first quartile to the second is smaller than the distance from the second to the third quartile.
B. The distance from the minimum value to the first quartile is larger than the distance from the third quartile to the maximum value.
C. The distance from the minimum value to the second quartile is smaller than the distance from the second quartile to the maximum value.
D. The distance from the first quartile to the third is twice the distance from the first to the second quartile.

An ice cream vendor sells three flavors of ice cream, Chocolate, Vanilla and Pistachio. Sales are by single dip cone, double dip cone or triple dip cone. Information for sales for a one-week period are found in the following table. Use this information to answer the following ***TWO*** questions.

Ice Cream Sales
Number of Dips per cone by Flavor

# of dips per cone	Chocolate	Vanilla	Pistachio	Total
1	15	75	100	190
2	170	35	40	245
3	40	15	10	65
Total	225	125	150	500

18. What is the mean number of dips of Pistachio Ice Cream sold in a cone this week?

A. 2.5 B. 2 C. 1.4 D. 1

19. Assume the mean number of dips of Vanilla Ice Cream sold in a cone this week was 1.5. What is the standard deviation of the number of dips of Vanilla Ice Cream sold in a cone this week?

A. 0.49 B. 0.7 C. 0.92 D. 0.96

20. A sample of data has a mean of 24 and a standard deviation of 7. The removal of which of the following numbers from the data set will change the sum of the squared deviations the most?

A. 14 B. 7 C. 24 D. 0

21. Suppose the amount of time (in seconds) needed to assemble a carburetor on an assembly line at Rover has a mean of 54 seconds and a standard deviation of 6 seconds. According to Chebyshev's Theorem, at least 80% of the on-time tasks are, to the nearest second, between

A. 41 and 67 seconds
B. 42 and 66 seconds
C. 36 and 72 seconds
D. 39 and 69 seconds

The following table of data was calculated using Descriptive Statistics in Excel. The data are observations of Age at Purchase, Household Income and Years of Education for owners of Mercedes luxury automobiles. Use this information to answer the following ***FOUR*** questions.

Mercedes Owners	*Age*	*Income*	*Education*
Mean	52.34	184214.86	
Median	52	186070	17
Mode	53		
Standard Deviation		47553.71	1.76
Sample Variance	59.76		3.08
Skewness			
Range	42		9
Minimum		49941	13
Maximum	77	334823	
Sum		13447685	1260
Count	73		

22. The characteristic in which Mercedes owners were most diverse is

    A. Age at Purchase
    B. Household Income
    C. Years of Education
    D. They are not diverse in any characteristic.

23. You want to check your variance calculation for Age at Purchase, so you use the =VAR(array) command in Excel. The value of the variance you will get is

    A. larger than 59.76
    B. smaller than 59.76
    C. equal to 59.76
    D. This calculation requires the use of Excel.

24. Of Age at Purchase and Years of Education, which is the least symmetric?

    A. Age at Purchase
    B. Years of Education
    C. There are equally asymmetric.
    D. This calculation requires the use of Excel.

25. If the mean of 72 of the Age at Purchase observations was 52.375, what was the age of the 73rd Mercedes owner when the car was purchased?

    A. 53 B. 52 C. 50 D. 35

26. Suppose the Union Board has determined that the daily demand of IU students for The New York Times is bell-shaped with a mean of 300 and a variance of 225. If the Union Board orders 330 copies of The New York Times every day, what percentage of days will they sell out?

    A. about 2.5%
    B. about 5%
    C. about 95%
    D. about 97.5%.

27. For which of the following pairs of variables would you feel confident in calculating a least squares line, that is, to have a causal relationship?

    A. Age of Husband and Wife at marriage; r = 0.94
    B. For the 50 states, Rate of Births and Rate of Deaths; r = -0.48
    C. A country's Life Expectancy in Years and Number of People per TV; r = -0.80
    D. The percent of weight borne by the front axle of a car and the fuel capacity of the car; r = -0.16

28. The correlation coefficient for Golf Score and Time to Complete Course is 0.69. Assuming that a valid relationship exists between the two variables, what does this say about better golfers and their on-course time? (Better golfers have lower scores.)

    A. They take more time to play.
    B. They take less time to play.
    C. They take about the same time as poorer golfers.
    D. Another correlation coefficient is required to answer this question.

29. Under which of the following conditions might stratified random sampling be appropriate?

    A. The population can be divided into a large number of strata so that each stratum contains only a few individuals.
    B. The population can be divided into strata so that individuals in each stratum are as different as possible.
    C. The population can be divided into strata of equal sizes so that each individual still has the same chance of being selected.
    D. The population can be divided into a small number of strata so that each stratum contains a large number of individuals.

The following is a covariance matrix calculated by Excel using data from a sample of buyers of Mercedes. The buyers age at purchase, household income and years of education were recorded. Use this information to answer the following ***TWO*** questions.

	Age	*Income*	*Education*
Age	58.94		
Income	-13164.49	2230377821.68	
Education	1.39	-1311.79	3.04

30. Of the three variable pairs, which has the strongest linear relationship?

    A. Age and Income
    B. Income and Education
    C. Education and Age
    D. There is not enough information to answer the question.

31. Suppose that the incomes recorded had not been corrected for inflation, and the actual incomes were twice as large as the data set reported. What effect would this have on the covariance of Age and Income?

    A. The covariance would be twice as large.
    B. The covariance would be four times as large.
    C. The covariance would not change.
    D. There is insufficient information to answer the question.

32. A statistician is constructing a systematic random sample of instructors. All 640 instructors have been identified, sorted by last name and numbered from 1 to 640 in ascending order. The first instructor in this sequence is assigned as number 1. The statistician wishes to estimate the mean score of the instructors on student evaluations, and has calculated an appropriate sample size of n=40. The 500th of all the instructors was selected as the third observation in the sample. What is the number of the first observation in the sample?

    A. 420 B. 460 C. 468 D. 484

33. The covariance between two variables is -53.5. Which of the following could be the correlation coefficient between the two variables?

    A. -0.357
    B. 0.05
    C. -1.13
    D. None of the above could be the correlation coefficient.

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