UNIVERSITY of VIRGINIA	**NAME:**_____ **STAT 2120: Introduction to Statistical Analysis** Spring 2010

Midterm Exam, October 13, 2011

This exam is closed book and closed notes, but you are allowed to use the formula sheet that is provided for you. You may use a calculator but not your laptop computer. All cell phones, ipods, and other electronic devices must be turned off. You are not permitted to work in groups or discuss the exam with anyone. The work you submit must be your own.

Please fill out your identifying information on the top of the scantron form now. On the right side of your form, write and carefully bubble your UVA computing ID (the front part of your UVA email address, *i.e.*, djs4y) in the space provided. Left justify your computing ID (*i.e.*, leave space to the right). For each question, completely fill in the label of your answer in the corresponding space on the accompanying scantron form. Each question is designed to have a single correct answer. If it seems that more than one statement is correct, then you should select the statement that best answers the question. If you calculate a numerical answer, but its exact value is not listed as a choice of answer then you should select the listed value that is closest to yours. Also, if the question asks for a probability but states that a "good approximation" is acceptable, then you should assume the context is such that any relevant approximation formula is valid (*e.g.*, the sample size is large enough for validity of the central limit theorem).

Some problems will require an evaluation of one of the Excel functions **normsdist** and **normsinv**. For these you should refer to the list of selected evaluations of those functions that is provided with the exam.

Print your name clearly above, sign the honor pledge below, and write out this honor pledge on the space provided on the back of the scantron form. You have 90 minutes to complete the exam. There are thirty-three questions on the exam. Each counts the same amount toward your final score.

Honor Pledge:
I have neither given nor received unauthorized aid on this exam.

Signed: _____

1) In a questionnaire, respondents are asked to mark the number of times they have flown on an airplane in the past twelve months. Depending on how it is used, the number of airplane flights an example of a:

a. Categorical variable
b. Discrete variable
c. Quantitative variable that is not continuous
d. Any of the above

2) Suppose the distribution of tax owed by taxpayers in a certain population is Normal with a mean of $490 dollars and a standard deviation of $205. Tax collectors want to target the 10% of taxpayers in this population who owe the most. What value defines the minimum amount of tax owed by a taxpayer in the targeted subgroup.

a. 254.00
b. 510.50
c. 752.40
d. 832.20

3) The ages of seventeen people responding to a survey are:

26, 31, 19, 37, 18, 18, 18, 18, 19, 18, 20, 28, 19, 35, 18, 18, 18

What is the third quartile of these data? Hint: Stem plots work well for back-of-the-envelope calculations such this.

a. 18
b. 19
c. 26
d. 27

4) Refer to the data of Problem 3. The mean of these data is x-bar = 22.24. A mean-median comparison tells us that the data are:

a. Multi-modal
b. Right-skewed
c. Left-skewed
d. Symmetric

5) Suppose the distribution of annual profits per policy in a particular insurance company's population of homeowner policies is Normal with a mean of $150 and a standard deviation of $150. Use the 68-95-99.7 rule to deduce the proportion of homeowner policies that will not make a profit this year, but will lose no more than $150.

a. 0.1350
b. 0.1850
c. 0.2700
d. 0.3200

6) Which of the following is not a standard of ethical data production?

a. Participation only after informed consent.
b. Subjects are to be compensated for participation.
c. Oversight by an institutional review board.
d. Raw data are to be kept confidential.

7) A sampling study intends to generalize results to all residents of a certain town, but a simple random sample is collected only from those residents who are registered to vote. The bias in this setup is due to:

a. Probability sampling using unknown selection probabilities.
b. Non-response of the sampled individuals.
c. Under-coverage of the population list.
d. Voluntary sampling.

8) The coefficient of determination is:

a. The value of slope that minimizes the squared prediction errors, where a prediction error is the vertical distance between a response value and the regression line.
b. The proportion of variability in the response that is explained by the regression line.
c. A measure of strength and direction in the relationship between two variables.
d. A measure of the degree to which systematic patterns are present that might invalidate the use of linear regression.

9) In a cost-study of road resurfacing projects, a regression analysis examining the dependency of the length of road resurfaced (y, in miles) on cost (x, in millions of dollars) produces the least-squares regression line $y = 1.982 + 3.214\,x$. According to this line, what is to be expected of the length of road resurfaced for each one million dollars invested in a project?

a. It would increase by 1.982 miles
b. It would decrease by 1.982 miles
c. It would increase by 3.214 miles
d. It would decrease by 3.214 miles

10) A small-animal clinic is interested in the proportion of its 400 animal patients who are overdue for a vaccination. To investigate, veterinarians at the clinic checked the records of a random sample of 15 of its patients and found that 3 of the sampled patients (20%) were overdue for a vaccination. What is the population of interest in this problem?

a. The 400 animals that are patients of the clinic.
b. The hypothetical population of all animals of the kind that the clinic serves.
c. The 15 animal patients whose records were examined.
d. About 80 animals (20%) of the 400 who are patients of the clinic.

11) The use of comparisons is a principle of experimental design. What is the benefit of using comparisons?

a. Comparisons reduce random variability.
b. Comparisons cancel the effects of lurking variables.
c. Comparisons randomize the allocation of subject to treatments.
d. All of the above