22S:30/105, Statistical Methods and Computing
Spring 2011, Instructor: Cowles
Final Exam

Name: _____ Course no. (30 or 105) _____

1. A physical therapist wishes to study the effectiveness of ice and prescribed exercises in reducing knee pain for patients with arthritis in their knees. She recruits 40 patients with knee arthritis into her study, and randomly divides them into 4 groups. At their first appointment in the study, each patient is assessed as to degree of knee pain, and:

   - Patients in group 1 are instructed simply to carry on their normal daily activities. They are not taught to ice their knees or to do special knee exercises.
   - Patients in group 2 are taught to apply ice packs to their knees for 15 minutes each day. They are not taught to do special exercises.
   - Patients in group 3 are taught to perform special knee exercises to strengthen the muscles supporting their knees. They are not taught to use ice packs.
   - Patients in group 4 are taught both to apply ice packs and to perform the special knee exercises.

   One month later, all patients are again assessed as to degree of knee pain, and their change from baseline is recorded.

   (a) What are the factors in this experiment? (circle one)
      i. the 40 individual patients
      ii. the 4 groups
      iii. ice and exercise
      iv. no ice + no exercise; ice + no exercise; no ice + exercise; ice + exercise
      v. change in degree of knee pain

   (b) What are the treatments in this experiment?
      i. the 40 individual patients
      ii. the 4 groups
      iii. ice and exercise
      iv. no ice + no exercise; ice + no exercise; no ice + exercise; ice + exercise
      v. change in degree of knee pain

   (c) What is the response variable in this experiment?
      i. the 40 individual patients
      ii. the 4 groups
      iii. ice and exercise
      iv. no ice + no exercise; ice + no exercise; no ice + exercise; ice + exercise
      v. change in degree of knee pain

2. For each of the following variables, state which data type it is (binary, nominal, ordinal, quantitative continuous, or quantitative discreet).

   (a) hair color (evaluated on a sample of human beings)

   (b) boiling temperature of water (evaluated at a number of different elevations in the mountains)

   (c) number of pets owned (evaluated on a sample of households in Iowa)

3. A dentist gathers data to determine whether people who floss their teeth at least three times a week get fewer new cavities than people who floss less frequently. During a one-month period, she asks each patient who comes in for a dental appointment whether they floss their teeth at least three times a week. She also counts how many new cavities they have developed since their last appointment.

   Is the dentist's investigation an observational study or an experiment? Briefly justify your answer.

4. Below are 12 quiz scores. Find the interquartile range (IQR) of these values. (Numeric answer; show your work.)

   ```
   85  79  92  98  67  81  88  71  95  77  91  90
   ```

5. An automotive magazine editor was interested in characteristics of car purchasers. In particular, he was interested in whether there were differences among the mean ages of purchasers of U.S.-made cars, purchasers of Asian-made cars, and purchasers of European-made cars. He obtained simple random samples of 10 purchasers of U.S. made cars, 10 purchasers of Asian-made cars, and 10 purchasers of European-made cars. He learns the ages of all of the purchasers.

   SAS output for an analysis of the editor's dataset is attached. You will need to refer to it in answering some of the following questions.

   (a) In this study, the populations of interest are (circle one):

      i. the 3 groups of purchasers in the dataset

      ii. all purchasers of U.S.-made cars, all purchasers of Asian-made cars, all purchasers of European- made cars

      iii. the mean age of all purchasers of U.S.-made cars, the mean age of all purchasers of Asian-made cars, the mean age of all purchasers of European-made cars

iv. the mean ages in each group in the dataset

v. none of the above

(b) The SAS output includes results of the ANOVA procedure. Why was ANOVA used instead of a chi-square test or a t-test?

(c) One assumption required for ANOVA is that the data are independent simple random samples from the populations of interest. Is there any way to assess this assumption by examining the data values? Briefly explain.

(d) The SAS output includes efforts to assess whether two other assumptions required for ANOVA have been met. Briefly state each assumption, and tell what the SAS output says as far as whether it has been met.

(e) State the null hypothesis being tested by the ANOVA procedure. Use conventional statistical symbols.

(f) At the .05 significance level, should we reject the null hypothesis? (yes/no) Why or why not?

(g) Briefly explain what this result means about the average ages of buyers of the three different classifications of cars.

6. Suppose that the weight of adult tigers follows a normal distribution with population mean 185 kg and population standard deviation 60 kg.

   If we get a simple random sample of 25 adult tigers, what is the probability that the sample mean of their weights will be greater than 200 kg? (Numeric answer; show your work).

7. We wish to estimate the population mean of length in meters of adult boa constrictors. Based on a sample of 15 boa constrictors, we obtain a sample mean of 3.9 m and a 95% confidence interval of (3.5, 4.3) meters.

   Circle all of the following statements that are true:

   (a) We are 95% confident that the sample mean lies in the interval.

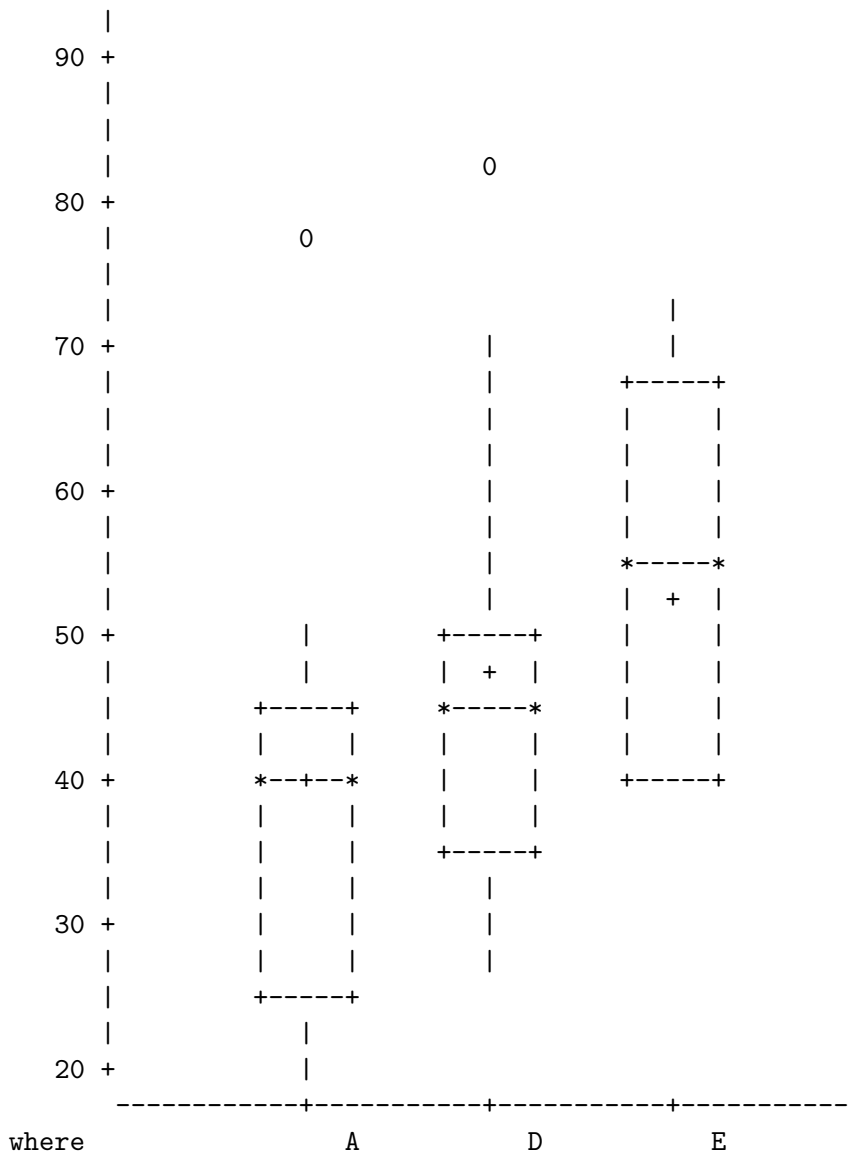
   (b) We should reject the null hypothesis because the sample mean lies in the interval.

   (c) We are 95% confident that the population mean lies in the interval.

   (d) We are 95% confident that the length of any randomly selected boa constrictor will lie in the interval.

The UNIVARIATE Procedure
Variable: age

Schematic Plots



Analysis Variable : age

where	N Obs	N	Mean	Std Dev	Minimum
A	10	10	39.6000000	17.3089572	21.0000000
D	13	13	47.3846154	15.7402799	27.0000000
E	7	7	53.1428571	13.5330284	39.0000000

Analysis Variable : age

where	N Obs	Maximum
A	10	78.0000000
D	13	83.0000000
E	7	72.0000000

The ANOVA Procedure

Class Level Information

Class	Levels	Values
where	3	A D E

Number of Observations Read	30
Number of Observations Used	30

Dependent Variable: age

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	791.132601	395.566300	1.58	0.2248
Error	27	6768.334066	250.679039		
Corrected Total	29	7559.466667			

R-Square	Coeff Var	Root MSE	age Mean
0.104655	34.31975	15.83285	46.13333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
where	2	791.1326007	395.5663004	1.58	0.2248