STAT:2010/4200, Statistical Methods and Computing
Spring 2016, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

Name: Solutions

Course no. (STAT:2010 or STAT:4200) _____

1. Which data type is each of the following variables? Circle the one best answer for each one.

   (a) The Motion Picture Association of American ratings of movies. The ratings are:

rating	definition
G General Audiences	All ages admitted. Nothing that would offend parents for viewing by children.
PG Parental Guidance Suggested	Some material may not be suitable for children. Parents urged to give "parental guidance."
PG-13 Parents Strongly Cautioned	May be inappropriate for children under 13.
R Restricted	Under 17 requires accompanying parent or guardian.
NC-17 Adults Only	No One 17 and Under Admitted.

   i. binary
   ii. nominal
   iii. ordinal ✓ 2
   iv. discrete quantitative
   v. continuous quantitative

   (b) the number of parking tickets a driver received in 2015

   i. binary
   ii. nominal
   iii. ordinal
   iv. discrete quantitative ✓ 2
   v. continuous quantitative

2. The next set of questions is based on a dataset described as follows:

```
Data on Vocabulary and Education from the 1989 General Social Survey

[1] Observation Index

[2] Education, in years

[3] Vocabulary Test Score, 10-Item Test

Source: 1989 General Social Survey, National Opinion Research Center.
Distributed by the Inter-University Consortium for Political and Social
Research.
```

(a) Below is some SAS output describing the Vocabulary Test Score variable. Which SAS procedure was used to produce it?

proc univariate

(2)

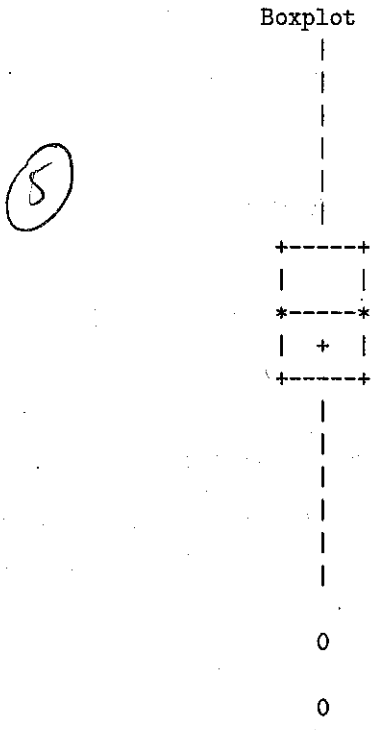
(1 for 5-number summary)

Quantiles (Definition 5)

Level	Quantile
100% Max	10
99%	10
95%	10
90%	9
75% Q3	7
50% Median	6
25% Q1	5
10%	3
5%	2
1%	0
0% Min	0

(b) Below is a boxplot of the Vocabulary Test Score variable. On each of the lines labeled with a letter – (a), (b), etc. – write the numeric value represented by the corresponding row of the boxplot.

(5)



(a) 10

(b) 7

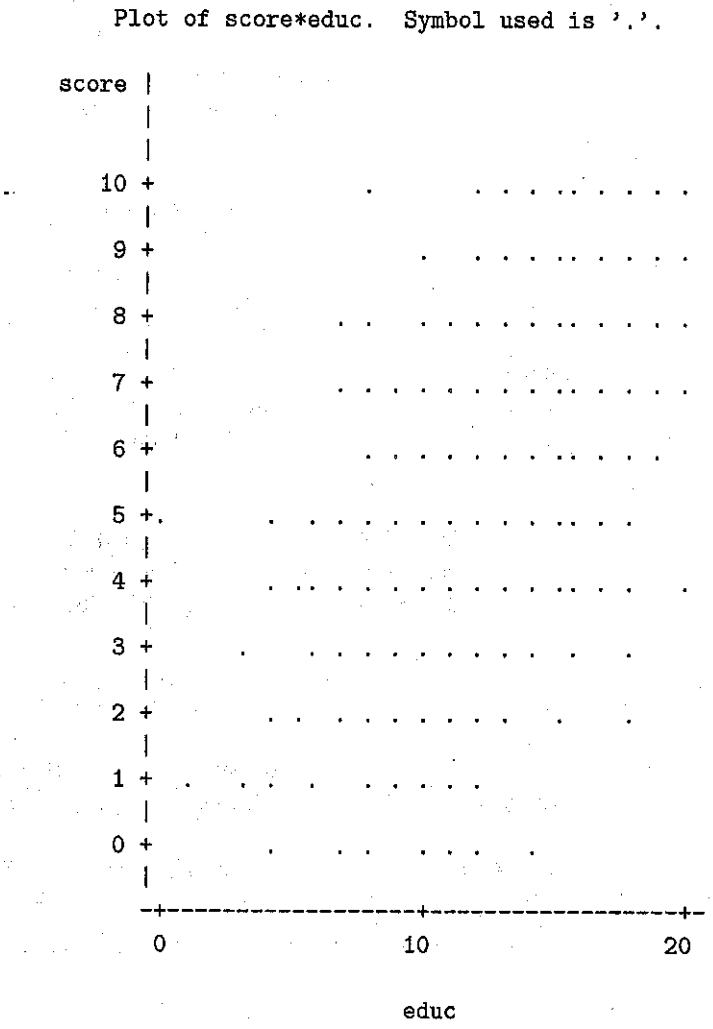
(c) 5

(d) 2

(e) 0

adjacent value
$5 - 1.5(7-5) = 2$

(c) Below is a scatterplot of Vocabulary Test Score (score) versus education (educ). The points lie in straight rows and columns because both variables are reported only as whole numbers.



i. Which variable, score or educ, is the response variable?

2 score

ii. Based on the scatterplot, to which value below is the sample correlation coefficient, r, likely to be closest? (Circle one).

A. -1
B. -0.5
C. 0
2 (D. 0.5)
E. 1

(d) Below is the SAS output for running a regression of score on educ.

The REG Procedure
Model: MODEL1
Dependent Variable: score

Number of Observations Read	968
Number of Observations Used	968

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1175.11129	1175.11129	318.92	<.0001
Error	966	3559.41351	3.68469		
Corrected Total	967	4734.52479			

Root MSE	1.91956	R-Square	0.2482
Dependent Mean	5.94008	Adj R-Sq	0.2474
Coeff Var	32.31530		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.13481	0.27606	4.11	<.0001
educ	1	0.37413	0.02095	17.86	<.0001

i. What is the numeric value of the sample slope?

2 0.37413

ii. Explain what the slope means in terms of the relationship between educ and score.

3 For each 1-year increase in education, we expect on average a 0.374-point increase in score.

iii. What is the predicted value of score for a person with 12 years of education? (Numeric answer; show your work.)

3 1.13481 + 0.37413 (12) = 5.62 points

3. A researcher wishes to evaluate the effect of caffeine and exercise on the performance of students on math exams. He recruits 80 students in a large calculus class to participate in an experiment. He randomly assigns 20 students to each of 4 groups.

   On the day of the next exam in the calculus class,

   - The 20 students in Group 1 will be required to meet 1 hour before the exam and will be taken for a vigorous, 30-minute walk. They will also be given a cup of strong coffee 30 minutes before the exam.
   - The 20 students in Group 2 will also be taken for the vigorous 30-minute walk. They will be given a cup of decaffeinated coffee (decaf) 30 minutes before the exam.
   - The 20 students in Group 3 will be required to sit quietly for 30 minutes before the exam and will be given a cup of strong coffee.
   - The 20 students in Group 4 will be required to sit quietly for 30 minutes before the exam and will be given a cup of decaf.

   The students are not told whether they are receiving coffee or decaf. After the exam, each student's score will be recorded and the mean scores in the 4 groups will be compared.

   (a) What are the experimental units in the study? (Circle one.)
   i. the 80 students
   ii. the 4 groups
   iii. caffeine and exercise
   iv. coffee and vigorous walking, coffee and sitting quietly, decaf and vigorous walking, decaf and sitting quietly
   v. exam score

   (b) What are the factors in the experiment? (Circle one.)
   i. the 80 students
   ii. the 4 groups
   iii. caffeine and exercise
   iv. coffee and vigorous walking, coffee and sitting quietly, decaf and vigorous walking, decaf and sitting quietly
   v. exam score

   (c) What is the response variable in the study (Circle one.)
   i. the 80 students
   ii. the 4 groups
   iii. caffeine and exercise
   iv. coffee and vigorous walking, coffee and sitting quietly, decaf and vigorous walking, decaf and sitting quietly
   v. exam score

   (d) In this study, the decaffeinated coffee is serving as a (Circle one):
   i. placebo
   ii. response variable
   iii. nonresponse bias
   iv. stimulant