

22S:30/105, Statistical Methods and Computing
Spring 2006, Instructor: Cowles
PRACTICE PROBLEMS for MIDTERM 1, 2008
Midterm 1

Show your work on any problems that involve calculations.

There are 50 total points on this midterm. Point values for each question are shown in parentheses. I will grade on a curve.

Name: _____
Course no. (30 or 105) _____ Section (1 or 2) _____

1. In a study conducted in Italy, 10 patients with high levels of triglycerides in their blood were placed on a low-fat, high-carbohydrate diet. Before the start of the diet, cholesterol and triglyceride measurements were recorded for each subject. The variables in the dataset "cholesterol.dat" are

- patient number (**pat**)
- cholesterol level (units are mmol/l) (**chol**)
- triglyceride level (units are mmol/l) (**trig**)

Here are the data values:

pat	chol	trig
1	5.12	2.30
2	6.18	2.54
3	6.77	2.95
4	6.65	3.77
5	6.36	4.18
6	5.90	5.31
7	5.48	5.53
8	6.02	8.83
9	10.34	9.48
10	8.51	14.20

Refer to the attached SAS output in answering the following questions.

(a) (3) Give the numeric values of the mean and standard deviation of the variable **chol** and give the units for each.

(b) (3) Give the 5-number summary of the variable **chol**. (Numeric answer)

(c) (3) Which summary statistics (mean and standard deviation versus 5-number summary) provide a better description of this variable? Briefly justify your answer, referring to specific SAS output.

(d) (2) Write the regression equation for the regression of **trig** on **chol**.

(e) (2) Calculate the predicted value of **trig** for observation number 8. (Numeric answer; show your work).

(f) (2) Calculate the residual for observation number 8. (Numeric answer; show your work).

(g) (3) Is the linear relationship between **trig** and **chol** very strong? Cite both numeric and graphical evidence to justify your answer.

(h) (3) If patient B's cholesterol is 2 mmol/l higher than patient A's cholesterol, based on the regression model, how much higher or lower would you expect patient B's triglycerides to be than patient A's? (Numeric answer)

2. Heights of Dandie Dinmont terriers are known to follow a normal distribution with mean 9.5 inches and standard deviation 0.75 inches.

(a) (3) What proportion of Dandie Dinmont terriers are shorter than 8.5 inches? (Show your work!)

(b) (3) Find the height in inches such that only 20% of Dandie Dinmont terriers are taller than this height. (Show your work!)

3. The faculty in the department of Statistics at the UI are:

Professors:

- J Broffitt
- K-S Chan
- R Dykstra
- J Geweke
- J Huang
- R Lenth
- R Russo
- E Shiu
- L Tierney
- G Woodworth
- D Zimmerman

Associate Professors

- K Cowles
- J Lang
- O Stramer

Assistant Professors

- N Shyamalkumar
- Q Tang
- J Yan

(a) (5) Use the list of random digits below to choose three simple random samples (SRSs) as follows:

- one SRS of size 3 from the full professors
- one SRS of size 1 from the associate professors
- one SRS of size 1 from the assistant professors

Write enough on the list of names, and make markings on the list of random digits, so that I can tell what procedure you used.

62964 88145 83083 69453 46109 59505 69680 00900 19687 12633
 96746 12149 37823 71868 18442 35119 62103 39244 91596 27611
 90056 82226 29485 32863 00095 77558 71709 45467 73676 47150

Write the names of the five people you selected here:

(b) (1) When the three simple random samples are combined, the resulting sample is of which type? (Circle one.)

- i. simple random sample
- ii. stratified random sample
- iii. voluntary response sample
- iv. convenience sample

- v. judgment sample
- vi. none of the above

4. For each of the variables described below, identify its data type and all of the plots that would be appropriate to describe it.

- (a) body weight in grams measured on tadpoles
 - i. (1) Data type (circle one):
 - A. binary
 - B. nominal
 - C. ordinal
 - D. discrete quantitative
 - E. continuous quantitative
 - ii. (2) Appropriate plots (circle all that apply):
 - A. bar graph
 - B. boxplot
 - C. histogram
 - D. pie chart
 - E. scatterplot
 - F. stem and leaf plot
- (b) (1) number of arrests in the year 2005 measured on individual Iowa City residents
 - i. (1) Data type (circle one):
 - A. binary
 - B. nominal
 - C. ordinal
 - D. discrete quantitative
 - E. continuous quantitative
- (c) (1) whether or not a person had been arrested during 2005 (yes/no)
 - i. (1) Data type (circle one):
 - A. binary
 - B. nominal
 - C. ordinal
 - D. discrete quantitative
 - E. continuous quantitative
 - ii. (2) Appropriate plots (circle all that apply):
 - A. bar graph
 - B. boxplot
 - C. histogram
 - D. pie chart
 - E. scatterplot
 - F. stem and leaf plot

5. NOTE: A problem of this kind would *not* appear on midterm 1 this year, as we have not covered this material yet.

A statistics professor carries out an experiment to determine the effect of temperature and lighting on the proportion of students that fall asleep during her lectures. She teaches 4 classes per day. For one class, she sets the classroom temperature at 65 and keeps all the lights in the room on during her lectures. For her second class she sets the room temperature at 70 and keeps all the lights on. For her third class she sets the room temperature at 65 and dims the lights. For her fourth class she sets the temperature at 70 and dims the lights. She does this on 5 different days. Following each lecture, she calculates and records the proportion of students that fell asleep at some point during the lecture.

- (a) (1) Is this an experiment or an observational study? Briefly justify your answer.
- (b) (1) Identify the response variable in this study.
- (c) (1) Identify the experimental units in this study.
- (d) (3) Identify the factor(s) and levels in this study.
- (e) (1) Identify the treatments in this study.
- (f) (2) Identify two possible confounding variables in this study. Briefly explain.

The UNIVARIATE Procedure
Variable: chol

Moments

N	10	Sum Weights	10
Mean	6.733	Sum Observations	67.33
Std Deviation	1.56330881	Variance	2.44393444
Skewness	1.64364116	Kurtosis	2.57906393
Uncorrected SS	475.3283	Corrected SS	21.99541
Coeff Variation	23.2186071	Std Error Mean	0.49436165

Basic Statistical Measures

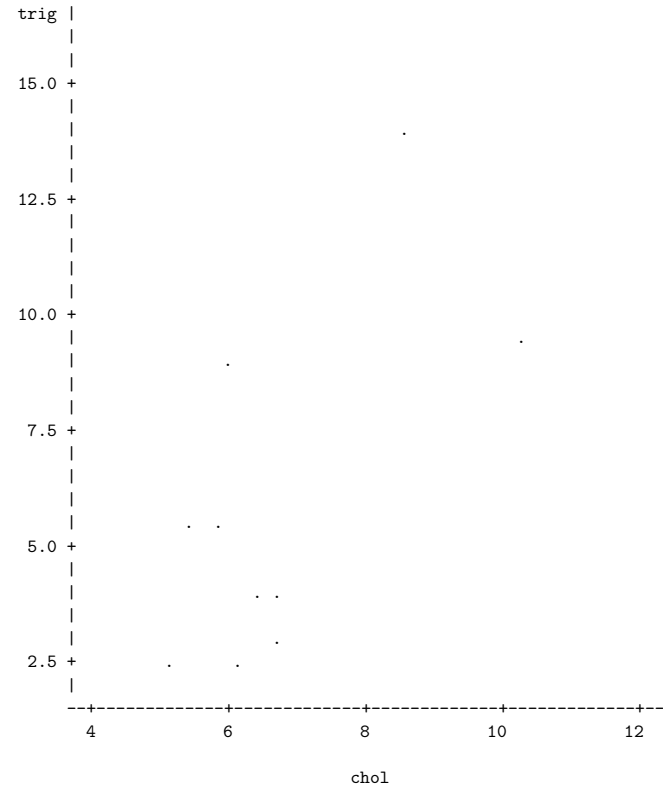
Location		Variability	
Mean	6.733000	Std Deviation	1.56331
Median	6.270000	Variance	2.44393
Mode	.	Range	5.22000
		Interquartile Range	0.87000

Quantile	Estimate
100% Max	10.340
99%	10.340
95%	10.340
90%	9.425
75% Q3	6.770
50% Median	6.270
25% Q1	5.900
10%	5.300
5%	5.120
1%	5.120
0% Min	5.120

Stem Leaf	No	Boxplot
10 3	1	*
9		
8 5	1	0
7		
6 02468	5	+++++
5 159	3	+++++

-----+-----+-----+

Plot of trig*chol. Symbol used is '.'.



The REG Procedure
 Model: MODEL1
 Dependent Variable: trig

Number of Observations Read 10
 Number of Observations Used 10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	55.38272	55.38272	5.84	0.0420
Error	8	75.84017	9.48002		
Corrected Total	9	131.22289			

Root MSE 3.07896 R-Square 0.4221
 Dependent Mean 5.90900 Adj R-Sq 0.3498
 Coeff Var 52.10635

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-4.77490	4.52622	-1.05	0.3223
chol	1	1.58680	0.65651	2.42	0.0420