

15
8
9
7

35

22S:30/105, Statistical Methods and Computing
Spring 2015, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

Name: Solutions
Course no. (30, 105, or 197) _____

1. The head of the Statistics department needs to form a committee of three faculty members. He wants to randomly select the committee members from among all the faculty in the department, but he wants to make sure that both women and men are represented. In particular, he wants 2 women and 3 men to serve.
 - (a) What kind of sampling design should the department head use? (Circle one.)
 - i. convenience sampling
 - ii. judgment sampling
 - iii. randomized block design
 - iv. simple random sampling
 - v. stratified random sampling
 - vi. voluntary response sampling
 - (b) Below are the last names of the Statistics faculty. Use Table B starting at line 103 to select the faculty to serve on the committee. Mark up the listing so that it is clear exactly what you did.

- Women faculty
- 1 Cowles
 - 2 Ghosh
 - 3 Stramer
 - 4 Tan

- Men faculty
- 1 Chan
 - 2 Huang
 - 3 Lang
 - 4 Russo
 - 5 Tierney
 - 6 Zimmerman

6

2. The Framingham Heart study included an investigation of the relationship between systolic blood pressure (SBP) and body mass index (BMI). The regression equation from a simple linear regression model fit to data on 3539 participants in the Framingham study was

$$\widehat{SBP}_i = 108.28 + 0.67(BMI_i)$$

- 2 (a) Jenny's BMI is 2 units higher than Fred's. How much higher or lower would you expect Jenny's SBP to be than Fred's? Explain briefly.

$$2 \times 0.67 = 1.34 \text{ mmHg}$$

The slope 0.67 is the difference in SBP expected with a 1-unit difference in BMI, so double it.

- 2 (b) In the Framingham Heart study data, was the correlation coefficient between SBP and BMI positive, negative, or 0? How could you tell?

Positive - same sign as slope.

- 4 3. Researchers conduct an experiment to investigate the effects of dietary fat and exercise on speed of learning in laboratory rats. They design diet and exercise regimens as follows:

- Diets
 - high fat
 - low fat
- Exercise
 - none
 - moderate
 - high intensity

The researchers randomly assign 25 rats to each of the 6 combinations of diet and exercise. After 30 days, each rat is run through a maze twice and is timed each time. The learning score is the difference in seconds between the first and second running times (second time - first time).

- (a) What are the factors in this experiment? (Circle one.)

- i. the individual rats
- ii. the groups of 25 rats
- iii. dietary fat and exercise
- iv. high fat and lowfat diet; none, moderate, and high intensity exercise
- v. low fat + no exercise; low fat + moderate exercise; low fat + high intensity exercise; high fat + no exercise; high fat + moderate exercise; high fat + high intensity exercise
- vi. learning score

(b) What are the experimental units? (Circle one.)

- i. the individual rats
- ii. the groups of 25 rats
- iii. diet and exercise
- iv. high fat and lowfat diet; none, moderate, and high intensity exercise
- v. low fat + no exercise; low fat + moderate exercise; low fat + high intensity exercise; high fat + no exercise; high fat + moderate exercise; high fat + high intensity exercise
- vi. learning score

(c) What are the treatments? (Circle one.)

- i. the individual rats
- ii. the groups of 25 rats
- iii. diet and exercise
- iv. high fat and lowfat diet; none, moderate, and high intensity exercise
- v. low fat + no exercise; low fat + moderate exercise; low fat + high intensity exercise; high fat + no exercise; high fat + moderate exercise; high fat + high intensity exercise
- vi. learning score

(d) What is the response variable? (Circle one.)

- i. the individual rats
- ii. the groups of 25 rats
- iii. diet and exercise
- iv. high fat and lowfat diet; none, moderate, and high intensity exercise
- v. low fat + no exercise; low fat + moderate exercise; low fat + high intensity exercise; high fat + no exercise; high fat + moderate exercise; high fat + high intensity exercise
- vi. learning score

4. Gretchen had midterms in French and calculus last week. On the French test, scores for the whole class followed a normal distribution with mean 78 points and standard deviation 14. On the calculus test, class scores followed a normal distribution with mean 72 points and standard deviation 6.

Gretchen got an 85 on the French test and a 78 on the calculus test. On which exam did she do better compared to the rest of the class? Justify your answer using the fact that the score distributions were normal, and show your work.

$$Z_F = \frac{85 - 78}{14} = 0.5$$

$$Z_C = \frac{78 - 72}{6} = 1.0$$

5

She did better with respect to the calculus test - larger positive z-score so more standard deviations above the mean.

5. Many graduate schools require all applicants to submit their scores on the Graduate Record Examination (GRE) as part of their applications. One section of the GRE General Test is called Quantitative Reasoning.

- (a) Consider scores on the GRE Quantitative Reasoning test obtained by undergraduate math majors in the Honors Program at The University of Iowa. What shape is most likely for the distribution of their scores? (Circle one).

- 1
 i. left skewed
 ii. right skewed
 iii. roughly symmetric

- (b) Briefly explain why the shape you chose is most likely.

2 Most math majors in the Honors Program would get very high Quant Reasoning scores but there likely would be a few low outliers.

- (c) Given your thought about the shape of the distribution, which numeric summary would be most appropriate to describe Quantitative Reasoning scores of math majors in the Honors Program?

- 2
 i. mean and standard deviation
 ii. correlation and regression
 iii. five number summary
 iv. mode and range
 v. frequency table

6. Duncan (1961) studied the prestige of various occupations as perceived in the U.S. in 1950. He classified occupations into the following types:

prof = Professional and Managerial
 wc = White Collar
 bc = Blue Collar

The table below summarizes the distribution of the type variable in Duncan's data.

type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
bc	21	46.67	21	46.67
prof	18	40.00	39	86.67
wc	6	13.33	45	100.00

- (a) How many observations were there in Duncan's dataset?

45

2

- (b) What proportion of the occupations were white collar?

.1333

2

(c) Which kinds of graphs could be used to summarize this variable? Circle all that apply.

3
i. bar graph

ii. histogram

iii. line plot

iv. pie chart

v. scatterplot

vi. stem and leaf plot

7. Researchers wish to investigate whether a new dietary supplement can prevent balding in men over age 50 who haven't yet started to go bald. They recruit 500 men who are not yet losing their hair into the study and randomly assign half of the men to receive the new supplement and the other half to receive a placebo. After 5 years on the assigned treatment, each man is evaluated as to whether he is starting to become bald, and the proportions in the two treatment groups are compared.

(a) The population of interest is (circle one):

2
i. the 500 men in the study

ii. all men over age 50 who haven't yet started to go bald

iii. the new supplement and the placebo

iv. all of the men in the study who are not going bald after 5 years

(b) What additional information would you need in order to determine whether the study was single blind or double blind?

2
Were the study personnel who dealt directly with the participants blinded as to treatment assignment.

7