7
8
4
9
3
31

22S:30/105, Statistical Methods and Computing
Spring 2014, Instructor: Cowles
Midterm 2

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

Show your work on any problems that involve calculations. If your answer to a multiple choice or true-false question would vary under different conditions, write an explanation.

I will grade on a curve and will give partial credit wherever possible.

1. Biologists are interested in the effects of diet and exposure to sunlight on the bone density of young rats. They will use a sample of rats who are 1 month old at the beginning of the study. They will randomly assign each rat to receive one of three diets (high calcium, moderate calcium, or low calcium) and one of two sunlight-exposure conditions (6 hours per day or 2 hours per day). After 6 weeks, they will measure the bone density in the left femur of each baby rat.

   (a) Are the biologists undertaking an experiment or an observational study? Explain briefly. experiment. They are controlling the treatments that each rat receives.

   (b) Name the response variable. bone density in left femur at 6 weeks

   (c) Name the factor(s). diet
   exposure to sunlight

   (d) What are the experimental units?
   each rat

   (e) Name the treatments.

   high calcium + 6 hrs sun
   high calcium + 2 hrs sun
   moderate calcium + 6 hrs sun
   moderate calcium + 2 hrs sun
   low calcium + 6 hrs sun
   low calcium + 2 hrs sun

2. An elementary school principal wants to find out whether the kindergarten students in her school feel that the tables and chairs in the school cafeteria are too large for them. She wants to draw a random sample of 8 kindergarten students to talk to, and she wants to make sure that that the sample includes 4 girls and 4 boys.

(a) The best method for the principal to use to draw her sample is (circle one):

i. convenience sampling
ii. judgment sampling
iii. randomized block design
iv. simple ~~random sampling~~
v. stratified random sampling
vi. voluntary response sampling
vii. none of the above

(b) Briefly explain your answer to the previous question.

Simple random sampling generally is the ideal. However, a SRS would not necessarily contain 4 girls and 4 boys. Stratified random sampling is needed to assure representation of subgroups of interest.

(c) The *population* in which the principal is interested is (circle one):

i. all kindergarten students in the school
ii. the 8 students she will interview
iii. only those students who answer her questions
iv. none of the above

3. Here is the probability model for the reasons for absence reported by employees of a business:

Reason	Illness	Family emergency	Car trouble	Weather	Other
Probability	.60	.08	.10	.05	?

(a) What is the probability that a randomly chosen absence is due to "Other" reasons? Numeric answer; show your work.

.60
.08
.10
.05
.83

1 - 0.83 = 0.17

(b) This probability model is (circle one):

i. continuous
ii. discrete
iii. randomized
iv. none of the above

(c) What is the probability that a randomly-chosen absence is due to either Car trouble or Weather? Numeric answer; show your work.

.10 + .05 = .15

4. After each statement below, circle **T** for True or **F** for false. For each false statement, briefly explain why it is false.

   (a) A parameter is a number that describes an entire population (T) F

   (b) Choose a simple random sample of size n from any population with mean μ and finite standard deviation σ. The Central Limit Theorem states that, when n is large, the population distribution of individual values is approximately Normal.
   T (F) The CLT says that when n is large, the sampling distribution of X-bar is approximately normal.

5. Students at the University of North Dakota in Grand Forks, ND, wanted to estimate the mean high temperature in Grand Forks on January 15 across all years. They selected a random sample of 10 years during the period that weather records have been kept for Grand Forks and recorded the reported high temperature for Jan. 15 in each year. They used SAS to plot their data and to compute a 90% confidence interval for the mean high temperature in Grand Forks on Jan. 15 across all years. Their SAS output is below.

```
Stem Leaf                    #          Boxplot
   3 1                       1             |
   2                                       |
   2                                       |
   1 6899                    4          +------+
   1 03                      2          *--+--*
   0                                    |     |
   0 0                       1          +------+
  -0 2                       1             |
  -0                                       |
  -1 4                       1             |
     ----+----+----+----+
Multiply Stem.Leaf by 10**+1
```

```
                    The MEANS Procedure

                  Analysis Variable : high

        Lower 90%          Upper 90%
       CL for Mean        CL for Mean
      ---------------------------------
        3.4616674          18.5383326
      ---------------------------------
```

$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

(a) Is there anything in the SAS output that suggests that the interval may not be valid? Explain briefly.

2 No. There are no extreme outliers and the distribution is roughly symmetric.

(b) From this confidence interval, is it possible to figure out what $\bar{x}$ was in their sample? If so, give the numeric value. If not, explain why it cannot be done.

2 Yes. $\bar{x}$ is in the middle of the interval. $\frac{3.462 + 18.538}{2} = 11$

(c) What quantity are the students 90% confident lies in the interval? (circle one)

i. $\bar{x}$

1 ii. s

iii. μ

iv. σ

v. the high temperature on Jan. 15 in Grand Forks in any future year

vi. p

(d) When we say that we are 90% confident that the quantity you selected lies in the interval, we mean (circle the best answer):

i. there is 90% probability that the true value of the quantity lies in the interval

2 ii. when we use this method to calculate confidence intervals, 90% of SRSs will produce an interval that contains the true value

iii. neither of the above

(e) When SAS computed the 90% confidence interval, it used a t distribution with how many degrees of freedom? (numeric answer)

1 $10 - 1 = 9$

(f) If these same data values were used to compute an 80% confidence interval, the 80% interval would be (circle one):

i. wider than the 90% interval

ii. narrower the 90% interval

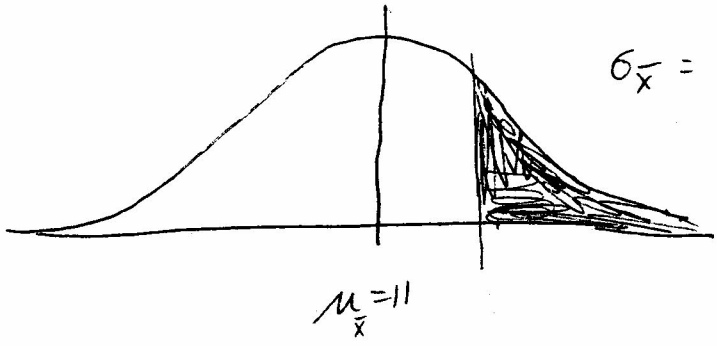
1 iii. the same width as the 90% interval

iv. any of the above; it's impossible to predict the comparative width.

9

(g) Suppose you had been told that high temperatures in Grand Forks, ND, on Jan. 15 follow a normal distribution with mean $\mu = 11$ degrees and known standard deviation $\sigma = 10$ degrees F. You plan to gather data on the high temperatures in Grand Forks on Jan. 15 from a random sample of 10 years. What is the probability the sample mean $\bar{x}$ from your data will be greater than 13 degrees? Numeric answer; show your work.

3



$$z = \frac{13-11}{10/\sqrt{10}} = .6325$$

$$Pr(Z > z) = 1 - .73$$

$$= .2643$$

7