22S:30/105, Statistical Methods and Computing
Spring 2009, Instructor: Cowles
Midterm 2

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

Show your work on any problems that involve calculations.

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

1. The primary purpose of a research study is to estimate the mean number of hours that first-year UI students work per week at paid jobs. Both a point estimate and a 90% confidence interval are desired. The investigators plan to draw a simple random sample of first-year UI students and interview them about their employment.

   (a) What is the population of interest in this study? (Circle one)
   - i. all first-year UI students
   - ii. the students who are selected at random and interviewed
   - iii. the mean number of hours worked per week by all UI students
   - iv. the mean number of hours worked per week by the students who are selected and random and interviewed

   (b) What is the parameter of interest in this study? (Circle one)
   - i. all first-year UI students
   - ii. the students who are selected at random and interviewed
   - iii. the mean number of hours worked per week by all UI students
   - iv. the mean number of hours worked per week by the students who are selected and random and interviewed

   (c) Is it likely that the distribution of hours worked per week follows a normal distribution among all UI first-year students? Briefly state why or why not.

   No. Many students work 0 hours. The distribution is likely to be skewed to the right because a few students will work a very large number of hours. Normal distributions are symmetric.

   (d) Suppose that the investigators strongly believe that the distribution of hours worked among all UI first-year students follows a normal distribution with standard deviation $\sigma = 6$ hours.

   If the investigators want to obtain a 90% confidence interval of width no greater than 2 hours, how many students should they enroll in their study? (Numeric answer; show your work.)

   so margin of error is 1

   $$1 = 1.645\left(\frac{6}{\sqrt{n}}\right)$$

   $$\sqrt{n} = 6(1.645)$$

   $$n = 97.4$$

(e) The conventional statistical symbol for the quantity that they will be 90% confident lies in the interval they obtain is:

i. $\bar{x}$

ii. y

iii. μ

iv. σ

v. s

(f) Suppose that a secondary purpose of the study is to determine whether the mean weekly work hours for all first-year UI students is greater than 12 hours. Suppose also that they have 36 students in their study, and that they still believe that $\sigma = 6$.

i. Write the null and alternative hypotheses, using the standard symbols from your textbook and lectures.

$H_0: \mu = 12$

$H_A: \mu > 12$

ii. (1) ~~This is a~~ (~~circle~~ one):

A. one-sided test

~~B. two-sided test~~

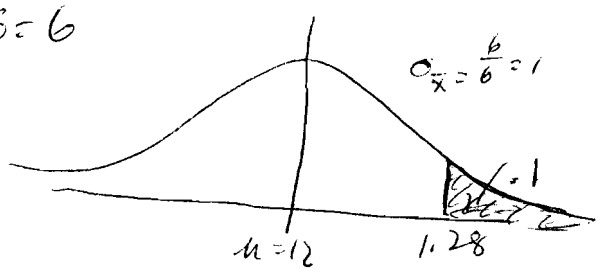
C. Chi square test

D. none of the above

iii. ~~The~~ researchers choose to conduct their hypothesis test at significance level $\alpha = .1$. What is the critical value of $\bar{x}$ that would enable them to reject H_0? Give a numeric answer. Show your calculations.

$n = 36$

$6 = 6$

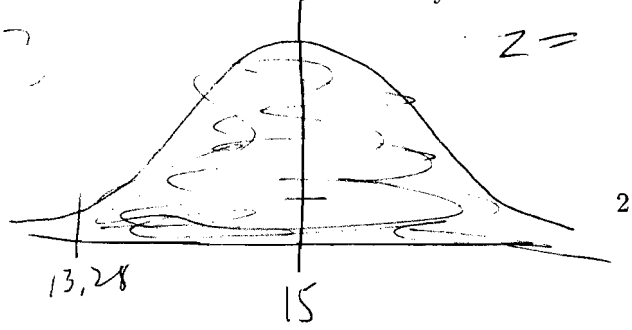


$\bar{x} = \mu + z\,\sigma_{\bar{x}}$

$= 12 + 1.28\,(1)$

$= 13.28$

iv. What is the power of the test against the alternative $\mu = 15$? Give a numeric answer. Show your calculations.



$z = \frac{13.28 - 15}{1} = -1.72$

$Pr(z > -1.72) = .957$

(g) Now suppose the investigators select and interview the students; the average weekly hours worked by students is 11.5 hours. The number 11.5 hours in this case is: (Circle one)

i. a population
ii. a sample
iii. a parameter
iv. a statistic
v. a standard deviation

2. The results of an experiment are said to be "statistically significant" if

(a) they are important to statisticians, regardless of their importance to the investigators
(b) both researchers and statisticians agree that the results are meaningful and important
(c) the observed effect is too large to attribute plausibly to chance
(d) they support the findings of previous, similar studies

3. The density curve for a continuous random variable X has which of the following properties?

(a) The probability of any event is the area under the density curve and above the values of X that make up the event.
(b) The total area under the density curve for X must be exactly 1.
(c) The probability of any event of the form $X =$ some exact number is 0.
(d) All of the above.
(e) None of the above.

4. The attached SAS output is based on a dataset giving the numbers of community hospital beds per 1000 population that are available in each state in the U.S. and in the District of Columbia in a particular year. These data are used to test the null hypothesis that the mean number of community hospital beds per 1000 population in all states at all times is 5.

(a) Proc univariate has been used to carry out a 1-sample t-test. What is the p-value of the test? (Numeric answer). .0030

(b) Briefly explain what this p-value means in this application.

If H_0 were true (i.e. $\mu = 5$) we would have only 3/1000 chance of getting a dataset with this much evidence against the null or more (that is, a t statistic as far away from 0 as ours).

(c) Should the null hypothesis be rejected at significance level $\alpha = .05$? Briefly justify your answer.

Yes. .003 < .05 so reject.

(d) Do you see any evidence in the SAS output that a one-sample t-test might not be appropriate for these data? Explain briefly.

Outliers!

(e) Is there any other reason (not shown in the SAS output) that this t-test might not give valid conclusions about the research question of interest? Explain briefly.

Yes. One year of data is not a simple random sample from all possible years.

Basic Statistical Measures

Location		Variability	
Mean	4.556863	Std Deviation	1.01277
Median	4.500000	Variance	1.02570
Mode	3.600000	Range	4.70000

Tests for Location: Mu0=5

Test	-Statistic-	-----p Value------
Student's t	t -3.12473	Pr > |t| 0.0030

Stem	Leaf	#	Boxplot
74	0	1	0
72	0	1	0
70			
68			
66			
64			
62			
60	0	1	|
58	00	2	|
56	000	3	|
54	000	3	|
52	0	1	|
50	0000	4	+------+
48	000	3	| |
46	00000	5	| |
44	000000	6	*--+--*
42	0000	4	| |
40	0	1	| |
38	000	3	| |
36	000000	6	+------+
34	00	2	|
32			|
30	0000	4	|
28			|
26	0	1	|

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

Multiply Stem.Leaf by 10**-1