

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

1. The following question appears on a survey of high risk behavior: "Have you ever used methamphetamine? (yes/no)."

(a) The responses to this question provide a variable of which data type? (Circle the one best answer.)

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

(b) The data from this question could be represented by which kinds of plots? (Circle all that apply.)

- i. bar chart
- ii. histogram
- iii. line plot
- iv. pie chart
- v. stem-and-leaf plot
- vi. scatterplot

(c) The data from this question could be used to draw inference about which kind of population parameter? (Circle the one best answer.)

- i. a population mean
- ii. a population proportion
- iii. a population standard deviation
- iv. none of the above

(d) Which is the most appropriate type of confidence interval for the parameter you selected in the previous question? (Circle the one best answer.)

- i. a chi-square interval
- ii. a t interval
- iii. a z interval

(e) The standard symbol for the *point estimate* of the population parameter that you selected is: (Circle the one best answer.)

- i. μ
- ii. \bar{X}
- iii. s
- iv. σ
- v. p
- vi. \hat{p}
- vii. π

2. This question is based on a dataset described as follows:

NAME: Televisions, Physicians, and Life Expectancy
TYPE: Sample
SIZE: 40 observations, 6 variables

DESCRIPTIVE ABSTRACT:

For each of the forty largest countries in the world (according to 1990 population figures), data are given for the country's life expectancy at birth, number of people per television set, and number of people per physician.

SOURCE:

The World Almanac and Book of Facts 1993 (1993), New York: Pharos Books.

VARIABLE DESCRIPTIONS:

Columns
1 - 15 Country
23 - 26 Life expectancy
29 - 33 People per television
36 - 40 People per physician
43 - 44 Female life expectancy
47 - 48 Male life expectancy

Additional information about these data can be found in the "Datasets and Stories" article "Televisions, Physicians, and Life Expectancy" in the *Journal of Statistics Education* (Rossman 1994).

SUBMITTED BY:

Allan J. Rossman
Dickinson College Department of Mathematics and Computer Science
P.O. Box 1773
Carlisle, PA 17013
rossman@dickinson.edu

The variable names that I used were:

• Country - name of country

• life - life expectancy at birth

• pplper tv - number of people per television set

• pplper dr - number of people per doctor

• log tv - log of number of people per tv

• log dr - log of number of people per doctor

(a) We wish to develop a linear regression model to predict life expectancy. Look at the scatterplots of life versus pplper tv and life versus log tv. Briefly explain why we should use log tv rather than pplper tv in a simple linear regression model to predict life.

The relationship between life and log tv is much more linear than the one between life and pplper tv, which is curved.

(b) Based on the scatterplot, is the relationship between life and log tv positive or negative?

negative

(c) Consider the two regression models. Which of the two predictor variables, log tv or log dr, explains more of the variability in life? Cite specific SAS output to justify your answer.

log tv has $R^2 = .7312$, thus log dr has $R^2 = 0.6385$. Thus, log tv explains more of the variability in life.

(d) The SAS output includes a test of the hypothesis that there is no linear relationship between life and log tv in the population of all large countries.

1. Write this null hypothesis as a statement about population parameters, using conventional statistical symbols.

$$H_0: \beta = 0$$

ii. At the .05 significance level, should you reject the null hypothesis? (yes or no) Justify your answer by citing numerical values from the SAS output for:

A. a test statistic and p-value
 Yes! t statistic = -9.90 with p-value $< .0001$.
 Yes, reject H_0 since p-value $< \alpha$.

B. a confidence interval for the parameter

95% c.i. for β is $(-5.13, -3.39)$. This does not contain 0, so reject H_0 . There is strong evidence that β is negative.

iii. Does your answer to the previous question mean that giving more television sets to people in a country will cause life expectancy to go up in the country? Briefly explain.

No! Association is not causation!

iv. Based on the regression model, what is the point estimate of the average life expectancy in all countries with log tv = 1.386?

$$A. \text{ Calculate this value by hand; show your work. } 77.89 - 4.26(1.386) = 71.98$$

B. Can this estimated value be found in the SAS output? If so, explain where to find it.

First row "Output Statistics"

3. Researchers designed an experiment to assess the effects of different antihistamines on drivers' alertness. The experiment was carried out using a driving simulator. The 80 participants were randomly assigned to 4 different groups, each of 20 people. Participants in treatment groups 1, 2, and 3 received antihistamines 1, 2, and 3 respectively, and participants in treatment group 4 received placebo.

The procedure was that each participant was given a single dose of the type assigned and participants in treatment group 4 received placebo.

At some point during the session, the image of a dog running into the street in front and told to drive normally in response to the scene shown in the simulators windshield. Thirty minutes later the participant was placed in the driving simulator to his group. Thirty minutes later the participant was placed in the driving simulator to his group. Thirty minutes later the participant was placed in the driving simulator to his group. Thirty minutes later the participant was placed in the driving simulator to his group.

(a) What were the experimental units in this experiment? (Circle the one best answer.)

i. the 80 people

ii. the 4 groups

iii. the 3 antihistamines and placebo

iv. the time to apply the brakes

(b) What were the treatments? (Circle the one best answer.)

i. the 80 people

ii. the 4 groups

iii. the 3 antihistamines and placebo

iv. the time to apply the brakes

(c) What was the response variable? (Circle the one best answer.)

i. the 80 people

- ii. the 4 groups
- iii. the 3 antihistamines and placebo
- iv. the time to apply the brakes

(d) Which statistical test would be most appropriate for determining whether there were any differences among the effects on alertness of the three antihistamines and placebo? (Circle the one best answer.)

- i. ANOVA
- ii. Chi square test
- iii. paired t-test
- iv. two-independent-sample t-test

4. Researchers wish to estimate the mean head circumference in the population of male history professors. They believe that the population is normal with standard deviation equal to 1.5 inches.

(a) How large a simple random sample of male history professors will they need in order to get a 90% confidence interval of width no greater than 1 inch? (Numeric answer; show your work.)

If width < 1 , then margin of error < 0.5

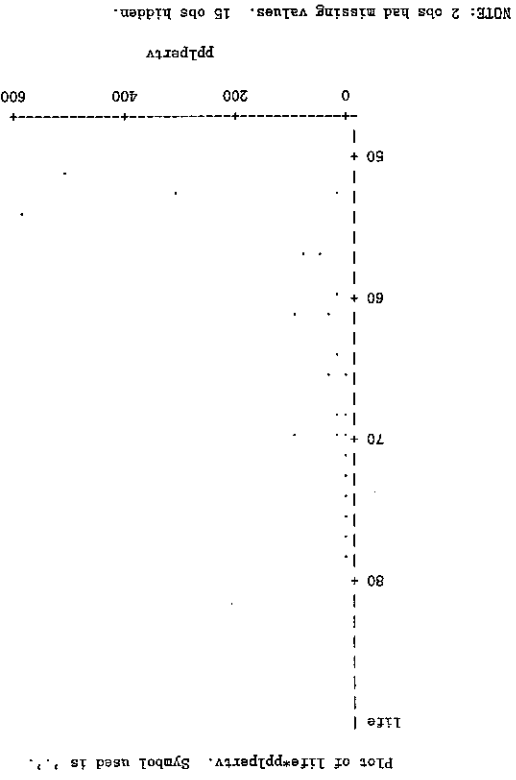
$$n = \left(\frac{z^* \sigma}{m} \right)^2 = \left[\frac{1.645 (1.5)}{0.5} \right]^2 = 24.35$$

Round up to 25.

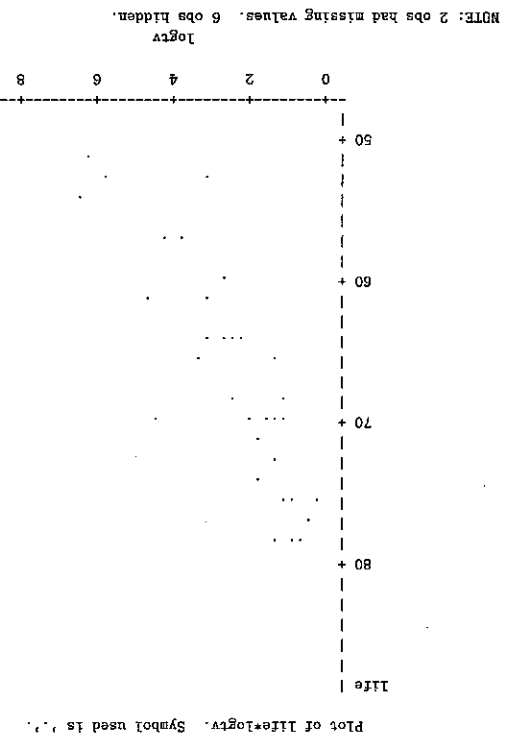
(b) Do they need to make sure that their sample is no greater than 1/10 the size of the population? Why or why not?

No. That requirement is for using normal approximations in computing confidence intervals for proportions. This study was about a population mean.

Obs	Country	life	pplpertv	pplperdr	female	male	logtv	logdr
1	Argentina	70.5	4.0	370	74	67	1.38629	5.9135
2	Bangladesh	53.5	315.0	6166	53	54	5.75257	8.7268
3	Brazil	65.0	4.0	684	68	62	1.38629	6.5280
4	Canada	76.5	1.7	449	80	73	0.53063	6.1070
5	China	70.0	8.0	643	72	68	2.07944	6.4661
6	Colombia	71.0	5.6	1551	74	68	1.72277	7.3467
7	Egypt	60.5	15.0	616	61	60	2.70805	6.4232
8	Ethiopia	51.5	503.0	36660	53	50	6.22059	10.5094
9	France	78.0	2.6	403	82	74	0.95551	5.9989
10	Germany	76.0	2.6	346	79	73	0.95551	5.8464
11	India	57.5	44.0	2471	58	57	3.78419	7.8124
12	Indonesia	61.0	24.0	7427	63	59	3.17805	8.9129
13	Iran	64.5	23.0	2992	65	64	3.13549	8.0037
14	Italy	78.5	3.8	233	82	75	1.33500	5.4510
15	Japan	79.0	1.8	609	82	76	0.58779	6.4118
16	Kenya	61.0	96.0	7615	63	59	4.56435	8.9379
17	Korea, North	70.0	90.0	370	73	67	4.49981	5.9135
18	Korea, South	70.0	4.9	1066	73	67	1.58924	6.9717
19	Mexico	72.0	6.6	600	76	68	1.88707	6.3969
20	Morocco	64.5	21.0	4873	66	63	3.04452	8.4915
21	Myanmar (Burma)	54.5	592.0	3485	56	53	6.38351	8.1562
22	Pakistan	56.5	73.0	2364	57	56	4.29046	7.7681
23	Peru	64.5	14.0	1016	67	62	2.63906	6.9236
24	Philippines	64.5	8.8	1062	67	62	2.17475	6.9679
25	Poland	73.0	3.9	480	77	69	1.36098	6.1738
26	Romania	72.0	6.0	559	75	69	1.79176	6.3261
27	Russia	69.0	3.2	259	74	64	1.16315	5.5568
28	South Africa	64.0	11.0	1340	67	61	2.39790	7.2004
29	Spain	78.5	2.6	275	82	75	0.95551	5.6168
30	Sudan	53.0	23.0	12550	54	52	3.13549	9.4375
31	Taiwan	75.0	3.2	965	78	72	1.16315	6.8721
32	Tanzania	52.5			55	50		
33	Thailand	68.5	11.0	4983	71	66	2.39790	8.4935
34	Turkey	70.0	5.0	1189	72	68	1.60944	7.0809
35	Ukraine	70.5	3.0	226	75	66	1.09861	5.4205
36	United Kingdom	76.0	3.0	611	79	73	1.09861	6.4151
37	United States	75.5	1.3	404	79	72	0.26236	6.0014
38	Venezuela	74.5	5.6	576	78	71	1.72277	6.3561
39	Vietnam	65.0	29.0	3096	67	63	3.36730	8.0379
40	Zaire	54.0			56	52		



7



8