7
9
8
4
11
39

22S:30/105, Statistical Methods and Computing
Spring 2014, Instructor: Cowles
Midterm 3

Show your work on any problems that involve calculations.

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

1. We wish to estimate p, the population proportion of likely voters who believe the state of the economy is the most urgent national concern. We want a 99% confidence interval with a margin of error no greater than .02. How many likely voters must be surveyed? (Assume that you have no idea of the value of p. Show your work.)

3

$$n = \left(\frac{2.58}{0.02}\right)^2 .5(.5)$$

$$= 4160.25 \text{ round up to } 4161$$

2. Starting in the 1970s, medical technology allowed babies with very low birth weight (VLBW, less than about 3.3 pounds) to survive without major handicaps. A long-term study has followed 242 VLBW babies to age 20 years, along with a control group of 233 babies from the same population who had normal birth weight. At age 20, 179 of the VLBW group and 193 of the control group had graduated from high school.

   (a) This is an example of (circle one):

   1

      i. an observational study
      ii. a randomized controlled study
      iii. a nonrandomized experiment
      iv. a matched-pairs experiment

   (b) You wish to use these data to assess whether a smaller proportion of VLBW people than normal-birth-weight people graduate from high school by age 20. Write the hypotheses that you will test. Use conventional statistical symbols. Define the symbols that you use.

   3

$$H_0 : p_{VLBW} = p_{NBW}$$

$$H_A : p_{VLBW} < p_{NBW}$$

p_{VLBW} = proportion of VLBW people who graduate from h.s. by age 20

7

(c) Which of the following statistical tests could be used to test the hypotheses? Circle all that apply.

i. Chi square test

3 ii. F test

iii. one sample t test

iv. paired t test

v. two-independent-sample t test

vi. z test

(d) In the same study, IQ scores were available for 113 men in the VLBW group and for 106 men in the normal-birth-weight group. You wish to use these data to test whether the average IQ is lower in the population of VLBW men than among control men from similar backgrounds. Write the hypotheses that you will test. Use conventional statistical symbols. Define the symbols that you use.

3

$H_0: \mu_{VLBW} = \mu_{NBW}$

$H_A: \mu_{VLBW} < \mu_{NBW}$

(e) Which of the following statistical tests could be used to test the hypotheses about IQs? Circle all that apply.

i. Chi square test

ii. F test

iii. one sample t test

iv. paired t test

3 v. two-independent-sample t test

vi. z test

9

3. Drug-detection dogs have long been used to sniff out illegal drugs at border crossing, airports, and similar locations. Dogs are big and expensive to maintain. Rats are small and cheap. A study investigated whether rats could be trained to replace dogs in drug-detection work. In one study, rats were trained to rear up on their hind legs when they smelled simulated cocaine. After training, each rat was let loose on a surface with many cups sunk in it, one of which contained simulated cocaine. Four out of six trained rats succeeded in 80 out of 80 trials.

   This problem concerns how we should estimate the long-term success rate p of an individual rat that succeeds in every one of 80 trials.

   (a) Find the rat's sample proportion $\hat{p}$ and the 95% confidence interval for p. (Numeric answer; show your work.)

$$\hat{p} = \frac{80}{80} = 1$$

3

$$s.e. = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{1 \cdot 0}{80}} = 0$$

$$95\%\ c.i. = 1 \pm 1.96(0) = (1, 1)$$

   (b) Find the plus-four estimate $\hat{p}$ and the plus-four 95% confidence interval for p. (Numeric answers; show your work.)

$$\hat{p} = \frac{82}{84} = .976$$

3

$$s.e. = \sqrt{\frac{.976(1-.976)}{84}} = .0167$$

$$95\%\ c.i. = .976 \pm 1.96(0.0167) = (0.943, 1.009)$$

   (c) Comment briefly on which result is more reasonable.

   The plus-four is more reasonable. It is not believable that the true p is exactly 1 as suggested by the large-sample c.i.

2

4. Researchers asked children in 4th, 5th, and 6th grade whether good grades, athletic ability, or popularity was their most important goal. We wish to use the data they gathered to assess whether the 4th graders' choices are different from 5th graders' choices.

   Reference: Chase, M.A and Dummer, G.M. (1992), "The Role of Sports as a Social Determinant for Children," Research Quarterly for Exercise and Sport, 63, 418-424.

   The table below summarizes the data for the 4th and 5th graders in the study.

8

```
                    Table of grade by choice

          grade     choice

          Frequency|
          Expected |
          Percent  |
          Row Pct  |
          Col Pct  |Grades  |Popular |Sports  |  Total
          ---------+--------+--------+--------+
          4        |     49 |     24 |     19 |     92
                   |  45.54 |   27.6 |  18.86 |
                   |  24.50 |  12.00 |   9.50 |  46.00
                   |  53.26 |  26.09 |  20.65 |
                   |  49.49 |  40.00 |  46.34 |
          ---------+--------+--------+--------+
          5        |     50 |     36 |     22 |    108
                   |  53.46 |   32.4 |  22.14 |
                   |  25.00 |  18.00 |  11.00 |  54.00
                   |  46.30 |  33.33 |  20.37 |
                   |  50.51 |  60.00 |  53.66 |
          ---------+--------+--------+--------+
          Total          99       60       41      200
                      49.50    30.00    20.50   100.00


               Statistics for Table of grade by choice

     Statistic                      DF       Value      Prob
     ------------------------------------------------------
     Chi-Square                      2      1.3583    0.5070
     Likelihood Ratio Chi-Square     2      1.3647    0.5054
     Mantel-Haenszel Chi-Square      1      0.3584    0.5494
     Phi Coefficient                        0.0824
     Contingency Coefficient                0.0821
     Cramer's V                             0.0824
```

(a) In each of the two grades, what is the sample proportion that chose Popularity? (Numeric answers from SAS output.)

2 $\hat{p}_{4th} = .2609$ $\hat{p}_{5th} = .3333$

(b) What is the expected count of 5th graders who chose Popularity? (Numeric answer from SAS output.)

$\frac{60}{200}(108) = 32.4$

2

4

(c) Do these data provide evidence that the proportions choosing the various goals are different in the two grades? Use the alpha = .10 significance level. Tell what test you used and justify your answer by citing specific SAS output.

3 No. The p-value for the Chi square test is 0.5070 > 0.10. There is very little evidence of a difference between the two proportions.

5. For each statement below, mark "T" for true or "F" for false. If the statement is false, briefly explain why.

(a) The power of a hypothesis test is the probability of correctly rejecting H_0 when H_0 is false.

T

2

(b) The power of a hypothesis test depends on the specific alternative hypothesis of interest.

2

T

larger

(c) You can increase the power of a test by using a ~~smaller~~ sample size.

2 F

one minus II

(d) The power of a hypothesis test is the probability of making Type ~~I~~ error.

F

2

11