30 points

22S:30/105, Statistical Methods and Computing
Spring 2011, Instructor: Cowles
Midterm 3

Show your work on any problems that involve calculations.

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

1. While a graduate student at the University of Washington, W.J. Kleindl performed a study of how the growth of cities affects the ability of streams to support plant and animal life. He reported two variables on each of 17 streams:

   `bibiscore` -- `an index of the ability of the stream to support plant and animal life`
   `impervpct` -- `a measure of how urbanized the area around the stream was`

   Reference: Kleindl, W. J. 1995. A Benthic Index of Biotic Integrity for Puget Sound Lowland Streams, Washington, USA. M.S. Thesis, University of Washington. Seattle, Washington.

   Refer to the SAS output in answering the following questions about these data.

   (a) In the scatterplot, the response variable is (circle one):

   2/2

   i. `bibiscore` (circled)
   ii. ~~`impervpct`~~
   iii. impossible to determine

   (b) We wish to use linear regression to test the null hypothesis that there is no linear relationship between `bibiscore` and `impervpct`. Write this hypothesis as a statement about an unknown population parameter. Use the conventional statistical symbol.

   2/2 $H_0: \beta = 0$

   (c) Find the following values in the SAS output, and copy the numbers here:

   i. the point estimate of the parameter you named in the previous question

   2/2 -0.43054

   ii. a 95% confidence interval for the parameter you named in the previous question

   2/2 (-0.64380, -0.21729)

   iii. the p-value for a two-sided test of the hypothesis you stated in the previous question

   2/2 0.0006

10

(d) Based on the SAS results you just cited, would you reject the null hypothesis at significance level $\alpha = .05$? Briefly justify your answer.

2/2 Yes p-value = 0.0006 < α = .05 and 0 is not in C.I.

(e) Suppose that `impervpct` in stream A is 5 units higher than the `impervpct` in stream B. Would you expect `bibiscore` to be higher or lower in stream A than stream B, and by how much?

2/2 5(-0.43054) = -2.1527

2. We wish to determine whether the proportion of left-handed people is the same in four different regions of the U.S. (West Coast, Heartland, Northeast, and South). We will draw a simple random sample of 50 people from each of the 4 regions. We will ask each person in each of the 4 samples whether they are left-handed or right-handed.

(a) The statistical test that is best-suited to our study is (circle one)

i. ANOVA
ii. Chi-square test
iii. correlation analysis
iv. linear regression
v. paired t-test
vi. two independent sample t-test
vii. none of the above

1/1

(b) Briefly justify your choice in the previous question.

2/2 Parameter is proportion, and we have more than 2 populations.

3. We wish to estimate the linear relationship between women's heights in inches and their shoe sizes, with the aim of being able to use height to predict shoe size. We will obtain a random sample of women and will determine the height and shoe size of each one.

(a) The statistical test that is best-suited to our study is (circle one)

i. ANOVA
ii. Chi-square test
iii. correlation analysis
iv. linear regression
v. paired t-test
vi. two independent sample t-test
vii. none of the above

1/1

2/2 We want linear association between 2 quantitative variables, and we want to use an explanatory variable to predict a response variable.

(b) Briefly justify your choice in the previous question.

4. We wish to determine whether the population mean of head circumference is the same in all philosophy majors as in all engineering majors in U.S. universities. We obtain a simple random sample of 25 philsophy majors from U.S. universities and another simple random sample of 25 engineering majors from U.S. universities. We measure the head circumference in inches of each individual in each sample.

(a) The statistical test that is best-suited to our study is (circle one)

i. ANOVA
ii. Chi-square test
iii. correlation analysis
iv. linear regression
v. paired t-test
vi. two independent sample t-test
vii. none of the above

(b) Briefly justify your choice in the previous question.

1

2

We are interested in comparing means of a quantitative variable in 2 populations. The samples are not paired

5. We wish to estimate the proportion of left-handed people in the population of all professional musicians. We plan to obtain a simple random sample of professional musicians and ask each of them whether they are left-handed or right handed. We want to calculate a 90% confidence interval for the unknown population proportion.

How large a sample of professional musicians will we need if we want our confidence interval to have width no greater than 0.1? (Show your work.)

3 1/2

margin of error is $\frac{1}{2}$ width of interval

z for 90% is 1.645

We have no guess of likely value of p, so we'll do worst-case scenario:

$$n = \left(\frac{1.645}{.05}\right)^2 .5(.5)$$

$$= 270.6$$

round up to 271

3

6 1/2

6. A study compared the effectiveness of two different kinds of flu vaccines (live attenuated influenza vaccine or LAIV, and inactivated influenza vaccine or IIV) in preventing flu in children.

The study results are shown in the table below.

	got flu	did not get flu	total
LAIV	4	25	29
IIV	9	23	32
total	13	48	61

Let p_L represent the proportion who will get flu in the population of all children who receive LAIV, andlet p_I represent the proportion who will get flu in the population of all children who receive IIV.

What is the *expected count* in the upper left cell under the null hypothesis:

$$H_0 : p_L = p_I$$

(numeric answer; show your work)

If $p_L = p_I$, our best estimate of that common value is $\frac{13}{61}$.

So the expected count in upper left cell is

$$\frac{13}{61}(29) = 6.18$$

Noninteger result is ok: this not the expected result for any one table, but the expected average over all tables with these margins.