Statistics 5371 Final Exam Review

Note: Most of this review can be done using a hand calculator and formulas from the text. However, you should know how to use SAS to obtain answers for each one of these problems if you had the appropriate data. And you should also know how to import data in to SAS!

Note 2: Homework problems and your previous exam problems are also fair game!

1. Suppose a researcher wants to design a new study with a power of 0.8 and a significance of 0.05 to test whether the caffeine content for a brand of coffee is really 100mg. A previous study gave a mean caffeine level for this brand of 110 mg and a standard deviation of 7 mg. Use PROC POWER to determine how many cups of coffee need testing.

2. A company did a study to estimate the effect of different promotional strategies on the market share of one of their products. Over a period of 36 months they varied their promotional strategy. There are four strategies: Ordinary (standard pricing and advertising); Discount (price discount with standard advertising); Promotion (standard pricing, enhanced advertising); and Both (price discount with enhanced advertising). Each month one of the strategies was employed and the market share (the percentage of purchasers of the given product type that selected the company's product). Higher market share is better for the company. The table below records the number of months in which each strategy was used (n) and a summary of the market share results in those months.

Strategy	Pricing	Advertising	n	mean	SD
Ordinary	standard	standard	8	2.40	0.12
Discount	Discount	standard	7	2.42	0.11
Promotion	Standard	enhanced	8	2.74	0.18
Both	Discount	enhanced	13	2.90	0.18

   a. What is the factor for this experiment and its levels?

   b. Suppose that a one-way ANOVA for the effect of strategy on market share is statistically significant. Can we conclude that strategy causes a change in market share? Explain.

   c. Perform a one-factor analysis of variance for the data. Write the table below and interpret the result.

   d. Regardless of the significance of your ANOVA, apply contrast defined by applying the weights (-0.5, 0.5, -0.5, 0.5) to the strategies (in the order listed) and determine the significance of the contrast.

   e. What does this contrast measure?

   f. Interpret the confidence interval for the contrast.

   g. Suppose we believe that the effect of a price discount may depend on whether there is enhanced advertising or not. Develop a set of contrast weights that would measure the

degree to which the effect of a price discount differs between the standard advertising and enhanced advertising settings.

3. It is suspected that an unnatural craving for substances such as paint might influence lead poisoning in children. A study was conducted to investigate this hypothesis. Ten of 20 rats were randomly assigned to a calcium-deficient diet (experimental group) and 10 to a regular diet (control group). Each of the rats occupied a separate cage and was monitored to determine the quantity of a 0.15% lead-acetate solution that they consumed during the study period. The amount each consumed is shown below.

Rat	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Control	5.4	6.2	3.1	3.8	6.5	5.8	6.4	4.5	4.9	4.0
Experiment	8.8	9.5	10.6	9.6	7.5	6.9	7.4	6.5	10.5	8.3

The mean and SD for the control group are 5.06 and 1.189, respectively. The mean and SD of the experiment group are 8.56 and 1.471. Perform the appropriate test for these data.

   a. State null and alternative hypotheses
   b. Find the rejection region of the test
   c. Calculate the value of the test statistic and the p-value.
   d. Carefully and completely write out any conclusions from the test.
   e. A journalist sees the results of this study and writes an article with the headline, “Lack of calcium in diet leads to lead poisoning in children.” Comment on this headline. Is it misleading? If so, explain. If not, explain why not.

4. As cheese ages, various chemical processes take place that determine the taste of the final product. This dataset contains concentrations of various chemicals in 30 samples of mature cheddar cheese, and a subjective measure of taste for each sample (variable taste – where a larger score is better). The variables "Acetic" and "H2S" are the natural logarithm of the concentration of acetic acid and hydrogen sulfide respectively.

```
/* Problem 5 */
data cheese;
input Case taste Acetic H2S;
datalines;
1     12.3  4.543 3.135
2     20.9  5.159 5.043
3     39    5.366 5.438
4     47.9  5.759 7.496
5     5.6   4.663 3.807
6     25.9  5.697 7.601
7     37.3  5.892 8.726
8     21.9  6.078 7.966
9     18.1  4.898 3.85
10    21    5.242 4.174
11    34.9  5.74  6.142
12    57.2  6.446 7.908
13    0.7   4.477 2.996
14    25.9  5.236 4.942
15    54.9  6.151 6.752
```

```
16    40.9  6.365 9.588
17    15.9  4.787 3.912
18    6.4   5.412 4.7
19    18    5.247 6.174
20    38.9  5.438 9.064
21    14    4.564 4.949
22    15.2  5.298 5.22
23    32    5.455 9.242
24    56.7  5.855 10.199
25    16.8  5.366 3.664
26    11.6  6.043 3.219
27    26.5  6.458 6.962
28    0.7   5.328 3.912
29    13.4  5.802 6.685
30    5.5   6.176 4.787
;
run;
```

a. Which variable is the response variable?

b. What is the equation of the regression line?

c. Is the overall regression equation significant?

d. Are the slopes of the regression line statistically significant? Explain your answer.

e. Interpret each slope.

f. Obtain partial correlations for each variable and interpret their meanings.

g. Obtain a 90% confidence interval for population slope of the variable "H2S". Interpret the interval.

5. In one weight-loss study 89 sedentary men were randomly assigned to either a special diet or exercise for a year. Forty-two men were placed on a diet and they lost an average of 7.2 kg with a standard deviation of 3.7 kg. The other 47 men were put on an exercise program and they lost an average of 5.3 kg with a standard deviation of 3.9 kg. For these data, researchers calculated three confidence intervals. They are given in the table below.

Confidence Level	Lower Limit	Upper Limit
90%	0.6	3.2
95%	0.3	3.5
99%	-0.2	4.0

a. Use the CI results to tell what p-value would be obtained in a two-sample t-test. (Hint: You can't give a precise value, only an interval. You can do the test to check but should report the answer to the question that is asked!)

b. A critic complains that beginning weight is strongly associated with the amount of weight lost and that this experimental design does not control for this important factor. The physician conducting the study says that he doesn't need to worry about the weight of a subject at the beginning of the study. What feature of the experimental design guarantees that the conclusion is still valid even though the study did not account for beginning weight? Explain.