STAT:2010/4200, Statistical Methods and Computing
Spring 2019, Instructor: Cowles
Midterm 3

Name: ____ Section _____________

1. A professor asked her sophomore students, "How many drinks do you typically have per session? (A drink is defined as one 12-ounce beer, one 4-ounce glass of wine, or one 1-ounce shot of hard liquor." Some of the students didn't drink. From the students who did drink, the professor obtained numeric data from 17 female students and 23 male students. She is tentatively willing to regard the students as a simple random sample of sophomore students are her college. She wishes to compare mean numbers of drinks in the populations of female students and male students at her college.

   (a) The professor's data is best described as (circle one):
      i. single sample
      ii. paired sample
      iii. two-independent sample
      iv. none of the above

   (b) The statistical test best suited to the professor's analysis is:
      i. z test
      ii. t test
      iii. chi-square test
      iv. ANOVA
      v. none of the above

      Briefly justify your choice.

2. You wish to study whether the proportions of women in three occupations (accountant, actuary, graphic designer) are equal. You randomly sample local companies and obtain the following employment data:

	women	men
accountants	17	26
actuaries	9	25
graphic designers	6	7

   (a) In the table above, fill in the margins to complete the contingency table.

(b) Compute the expected count for the number of women graphic designers under the null hypothesis of equal population proportions in all three occupations. (Numeric answer; show your work.)

(c) The test statistic value is 2.159. At significance level 0.05, should we reject the null hypothesis of equal population proportions? (yes/no) Briefly describe how you got your answer.

3. The Business Opportunities Handbook gives data on business startup costs in thousands of dollars for five types of businesses (pizza parlors, bakeries, shoe stores, gift shops, and pet stores).

   These sample data can be used to test whether the population means of startup costs are equal in these 5 types of businesses. SAS output for this problem is attached.

   (a) Why is ANOVA a better choice for this data analysis than the Chi square test?

   (b) List two assumptions that must be met in order for the results of ANOVA to be trustworthy. For each one, describe whether the SAS output suggests that the assumption is met.

   (c) Write the null hypothesis that will be tested. Use conventional symbols.

(d) We wish to test the null hypothesis at significance level alpha = 0.05. Give the numeric values of the test statistic and the p-value (from the SAS output).

(e) Should we reject H_0? (yes/no) Briefly explain.

(f) According to the SAS output, which pair or pairs of population means are different?

(g) In the following line from the ANOVA output, what quantity are we 95% confident falls in the interval? (Circle one)

```
                          Difference     Simultaneous
      type                   Between    95% Confidence
   Comparison                  Means        Limits

   gifts  - pet            35.38        -3.75   74.50
```

i. $\bar{x}_{gifts} - \bar{x}_{pet}$
ii. $\mu_{gifts} - \mu_{pet}$
iii. μ_{gifts}
iv. μ_{pet}
v. None of the above

```
                   The UNIVARIATE Procedure
                       Variable:  cost

                       Schematic Plots

     |
 160 +            |
     |            |
     |            |           |
     |            |           |
 140 +            |           |                       |
     |            |           |                       |
     |            |           |                       |
     |            |           |                       |
 120 +         +-----+        |                       |
     |         |     |        |                       |
     |         |     |        |           |        +-----+
     |         |     |        |           |        |     |
 100 +         |     |     *-----*        |        |     |
     |         |     |     |     |        |        |     |
     |         |  +  |     |     |        |        |     |
     |         *-----*     |  +  |        |        |  +  |
  80 +         |     |     |     |        |        *-----*
     |         |     |     |     |     +-----+     |     |
     |         |     |     |     |     |     |     |     |
     |         |     |     |     |     |     |     |     |
  60 +         +-----+     |     |     |     |     +-----+
     |            |        |     |     |     |        |
     |            |        +-----+     *--+--*        |
     |            |           |        |     |        |
  40 +            |           |        |     |        |
     |                        |        |     |        |
     |                                 +-----+
     |                                    |
  20 +                                    |
      ------------+-----------+-----------+-----------+------------
type            bakery      gifts         pet       pizza
```

```
                       Schematic Plots

     |
 160 +
     |
     |
     |
 140 +
     |
     |
     |            |
 120 +            |
     |            |
     |            |
     |            |
 100 +            |
     |         +-----+
     |         |     |
     |         |     |
  80 +         |     |
     |         |     |
     |         *--+--*
     |         |     |
  60 +         |     |
     |         |     |
     |         |     |
     |         +-----+
  40 +            |
     |            |
     |
     |
  20 +
      ------------+-----------
type             shoes
```

The MEANS Procedure

Analysis Variable : cost

type	N Obs	N	Mean	Std Dev	Minimum
bakery	11	11	92.0909091	38.8933273	40.0000000
gifts	10	10	87.0000000	35.9041935	35.0000000
pet	16	16	51.6250000	27.0748961	20.0000000
pizza	13	13	83.0000000	34.1345377	35.0000000
shoes	10	10	72.3000000	31.3654091	35.0000000

The ANOVA Procedure

Class Level Information

Class	Levels	Values
type	5	bakery gifts pet pizza shoes

Number of Observations Read	60
Number of Observations Used	60

The ANOVA Procedure

Dependent Variable: cost

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	14298.22424	3574.55606	3.25	0.0184
Error	55	60560.75909	1101.10471		
Corrected Total	59	74858.98333			

R-Square	Coeff Var	Root MSE	cost Mean
0.191002	44.13598	33.18290	75.18333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
type	4	14298.22424	3574.55606	3.25	0.0184

The ANOVA Procedure

Bonferroni (Dunn) t Tests for cost

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	55
Error Mean Square	1101.105
Critical Value of t	2.92470

Comparisons significant at the 0.05 level are indicated by ***.

type Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
bakery - gifts	5.09	-37.31	47.50	
bakery - pizza	9.09	-30.67	48.85	
bakery - shoes	19.79	-22.61	62.20	
bakery - pet	40.47	2.45	78.48	***
gifts - bakery	-5.09	-47.50	37.31	
gifts - pizza	4.00	-36.82	44.82	
gifts - shoes	14.70	-28.70	58.10	
gifts - pet	35.38	-3.75	74.50	
pizza - bakery	-9.09	-48.85	30.67	
pizza - gifts	-4.00	-44.82	36.82	
pizza - shoes	10.70	-30.12	51.52	
pizza - pet	31.38	-4.86	67.61	
shoes - bakery	-19.79	-62.20	22.61	
shoes - gifts	-14.70	-58.10	28.70	
shoes - pizza	-10.70	-51.52	30.12	
shoes - pet	20.67	-18.45	59.80	
pet - bakery	-40.47	-78.48	-2.45	***
pet - gifts	-35.38	-74.50	3.75	
pet - pizza	-31.38	-67.61	4.86	
pet - shoes	-20.67	-59.80	18.45	