STAT:2010/4200, Statistical Methods and Computing
Spring 2016, Instructor: Cowles
Midterm 3

Show your work on any problems that involve calculations.

Name: _____ Course no. (2010 or 4200) _____

1. CNN exit polls from the New York Republican primary showed the following results broken down by gender. http://www.cnn.com/election/primaries/polls/ny/Rep

Candidate	Females	Males	
Cruz	63	75	138
Kasich	118	123	241
Trump	240	338	578
	421	536	957

If the respondents to the exit poll were a simple random sample from the population of voters in the NY Republican primary, we can use these data to test the null hypothesis that the proportion of women is the same in the population of all NY Cruz voters, the population of all NY Kasich voters, and the population of all NY Trump voters. Refer to the SAS output for this problem in answering the following questions.

(a) Write the null hypothesis using conventional symbols.

(b) The expected count in cell for females who voted for Trump is 254.27. Show how this number was calculated using other numbers in the table.

(c) Are the rules of thumb met so that we can trust the Chi square test procedure? State each rule of thumb and how the data meets or does not meet it.

(d) At significance level $\alpha = .05$, does the data provide evidence against the null hypothesis? Explain, citing specific numbers from the SAS output.

2. In August 2010, the *Columbus Dispatch* newspaper tested water samples from 20 state park swimming areas in Ohio for bacteria that may make swimmers ill. Of the 20 swimming areas tested, 13 were found to have unsafe levels of the bacteria. Assume that the swimming areas tested represent a random sample of park swimming areas throughout the state. The newspaper investigators wished to draw conclusions about what proportion of all park swimming areas in the state have unsafe bacteria levels.

   (a) Are the rules of thumb for the normal approximation method of computing confidence intervals met? State each rule of them and how the data does or does not meet it.

   (b) Compute a 90% c.i. using the plus-four method. (Numeric answer; show your work.)

   (c) What quantity are we 90% confident lies in the interval? (Circle one.)

      i. The proportion of park swimming areas in the sample that were found to have unsafe bacterial levels.

      ii. The proportion of park swimming areas in the whole state that have unsafe bacteria levels.

      iii. Neither of the above.

   (d) How large a sample of swimming areas would have been needed to get a 90% confidence interval of width no greater than 0.04? (Numeric answer; show your work).

3. An experiment was done to compare the effects of two treatments on plant growth. A total of 30 plants were included in the study. Ten plants were randomly assigned to a control group, ten to treatment 1, and ten to treatment 2. At the end of the study, all the plants were dried. The response variable was the dry weight of each plant.

   The experimenters were interested in determining whether the population means of dry plant weight are equal in populations receiving treatment 1, treatment 2, or control.

   Refer to the attached SAS output in answering the following questions.

   (a) Why was ANOVA used in this analysis instead of a chi-square test?

   (b) Write the null hypothesis using conventional statistical symbols.

   (c) Is there anything in the SAS output that suggests that the results of the ANOVA analysis might not be trustworthy? Explain briefly.

   (d) At the .05 significance level, can the null hypothesis be rejected? Cite specific SAS output in justifying your answer.

   (e) Which population means (if any) does the analysis indicate are significantly different from each other (at the .05 significance level)?

SAS output for problem 1

Table of cand by gender

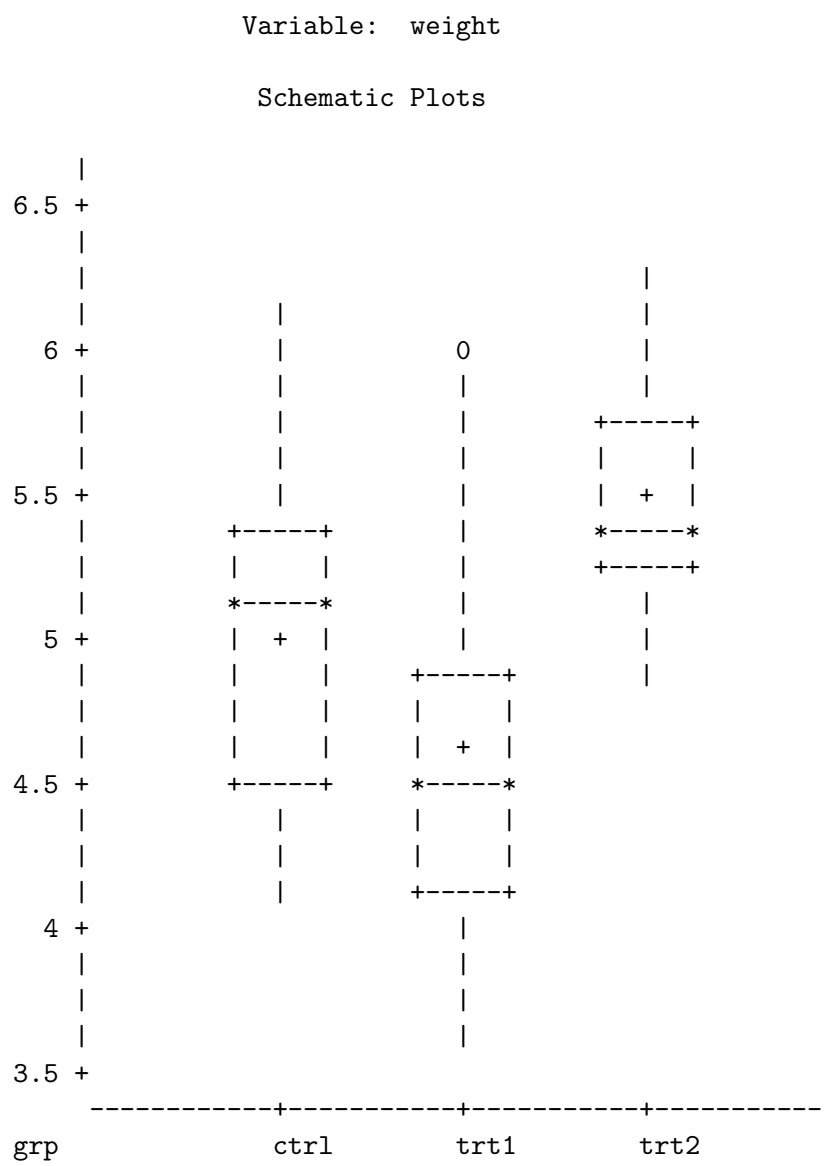
cand	gender		
Frequency|			
Expected |			
Percent |			
Row Pct |			
Col Pct |	F	M	Total
Cruz	63	75	138
	60.708	77.292	
	6.58	7.84	14.42
	45.65	54.35	
	14.96	13.99	
Kasich	118	123	241
	106.02	134.98	
	12.33	12.85	25.18
	48.96	51.04	
	28.03	22.95	
Trump	240	338	578
	254.27	323.73	
	25.08	35.32	60.40
	41.52	58.48	
	57.01	63.06	
Total	421	536	957
	43.99	56.01	100.00

Statistics for Table of cand by gender

Statistic	DF	Value	Prob
Chi-Square	2	4.0017	0.1352
Likelihood Ratio Chi-Square	2	3.9923	0.1359
Mantel-Haenszel Chi-Square	1	2.1652	0.1412
Phi Coefficient		0.0647	
Contingency Coefficient		0.0645	
Cramer's V		0.0647	

Sample Size = 957

SAS output for problem 3



grp	N Obs	N	Mean	Std Dev	Minimum	Maximum
ctrl	10	10	5.0320000	0.5830914	4.1700000	6.1100000
trt1	10	10	4.6610000	0.7936757	3.5900000	6.0300000
trt2	10	10	5.5260000	0.4425733	4.9200000	6.3100000

The ANOVA Procedure

Class Level Information

Class	Levels	Values
grp	3	ctrl trt1 trt2

Number of Observations Read	30
Number of Observations Used	30

Dependent Variable: weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3.76634000	1.88317000	4.85	0.0159
Error	27	10.49209000	0.38859593		
Corrected Total	29	14.25843000			

R-Square	Coeff Var	Root MSE	weight Mean
0.264148	12.28809	0.623375	5.073000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
grp	2	3.76634000	1.88317000	4.85	0.0159

Bonferroni (Dunn) t Tests for weight

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	27
Error Mean Square	0.388596
Critical Value of t	2.55246
Minimum Significant Difference	0.7116

Means with the same letter are not significantly different.

Bon Grouping		Mean	N	grp
	A	5.5260	10	trt2
	A			
B	A	5.0320	10	ctrl
B				
B		4.6610	10	trt1