

## FINAL EXAM

**Problem 1:** (20 percent;  $a = 5\%$   $b=5\%$ ;  $c=10\%$ )

WCAS assigns students to freshman seminars based on their preferences (and based on a program that I wrote – really!) Suppose we have the following data for 67 seminars, each of which has exactly 16 students.

No Women	No Seminars
0	0
1	0
2	2
3	0
4	2
5	2
6	11
7	9
8	13
9	11
10	5
11	1
12	6
13	2
14	2
15	1
16	0

<b>Total</b>	<b>67</b>
--------------	-----------

In other words, there were 2 seminars that had only 2 women, 2 seminars that had 4 women, 2 seminars with 5 women, 11 seminars that had 6 women, and so on. There were a total of 556 women assigned in this particular quarter out of a total of 1,072 students.

- a) Which of the following probability distributions would best reflect the number of women in a seminar assuming choices are random with respect to gender. (Check one only)

\_\_\_\_\_ Uniform

\_\_\_\_\_ Geometric

\_\_\_\_\_ Binomial

\_\_\_\_\_ Exponential

\_\_\_\_\_ Normal

\_\_\_\_\_ Chi-Squared

- b) Suppose you believe that the number of women in each seminar should follow a Binomial distribution (and this **may** or **may not** be the correct answer to part (a)). In that case, find the parameters of the Binomial distribution assuming that men and women have the same probability of being assigned to any individual seminar.

- c) To test the hypothesis that the underlying distribution is Binomial, you set up the following table.

No. Women	No. Seminars	Probability	Expected	Aggregate Groups -- No. Women	Expected	Observed	(Ei-Oi)2/Ei
0	0	0.000008	0.0006				
1	0	0.000143	0.0096				
2	2	0.001157	0.0775				
3	0	0.005818	0.3898				
4	2	0.020373	1.3650				
5	2	0.052685	3.5299	≤ 5	5.3723	6	0.073338
6	11	0.104077	6.9731	6	6.9731	11	2.3254451
7	9	0.160207	10.7338	7	10.7338	9	0.2800676
8	13			8		13	
9	11	0.186008	12.4625	9	12.4625	11	0.1716282
10	5	0.140299	9.4000	10	9.4000	5	2.0595834
11	1	0.082459	5.5247	11	5.5247	1	3.7057469
12	6	0.037021	2.4804	≥ 12	3.5218	11	
13	2	0.012274	0.8224				
14	2	0.002834	0.1899				
15	1	0.000407	0.0273				
16	0	0.000027	0.0018				
<b>Total</b>	<b>67</b>	1.0000000	67			67	

Find the values of the six shaded (highlighted) cells and test the null hypothesis that the number of women in a seminar follows a Binomial distribution. Use  $\alpha=0.05$ . Use the next page for your work, but fill in the answers in the table above.

Place your name on ALL pages NOW.  
THANKS!

Name \_\_\_\_\_

**Blank work page**

**Problem 2:** (20 percent;  $a=10\%$ ;  $b=10\%$ )

The Center for Disease Control (CDC) reports on a question that asked “Do you have any kind of health care coverage?” Results of the question by education level are shown below for Illinois.

Education	Yes	No	Total
Less than H.S.	295	64	359
H.S. or G.E.D.	1053	124	1177
Some post-H.S.	1084	93	1177
College Graduate	1233	50	1283

Education	Yes	No
Less than H.S.	82.2%	17.8%
H.S. or G.E.D.	89.5%	10.5%
Some post-H.S.	92.1%	7.9%
College Graduate	96.1%	3.9%

(Source = <http://apps.nccd.cdc.gov/brfss/education.asp?cat=HC&yr=2001&qkey=868&state=IL>)

- a. For the data shown above, test the following hypothesis with  $\alpha=0.05$ .

$$H_0: p_{\text{HS or GED}} = p_{\text{Some post HS}}$$

$$H_1: p_{\text{HS or GED}} \neq p_{\text{Some post HS}}$$

where  $p_{\text{HS or GED}}$  is the fraction of the population who have a High School or GED degree who answered yes to this question and other probabilities have the obvious meaning. Clearly show your work.

b. For the data shown above, test the following hypothesis with  $\alpha=0.05$ .

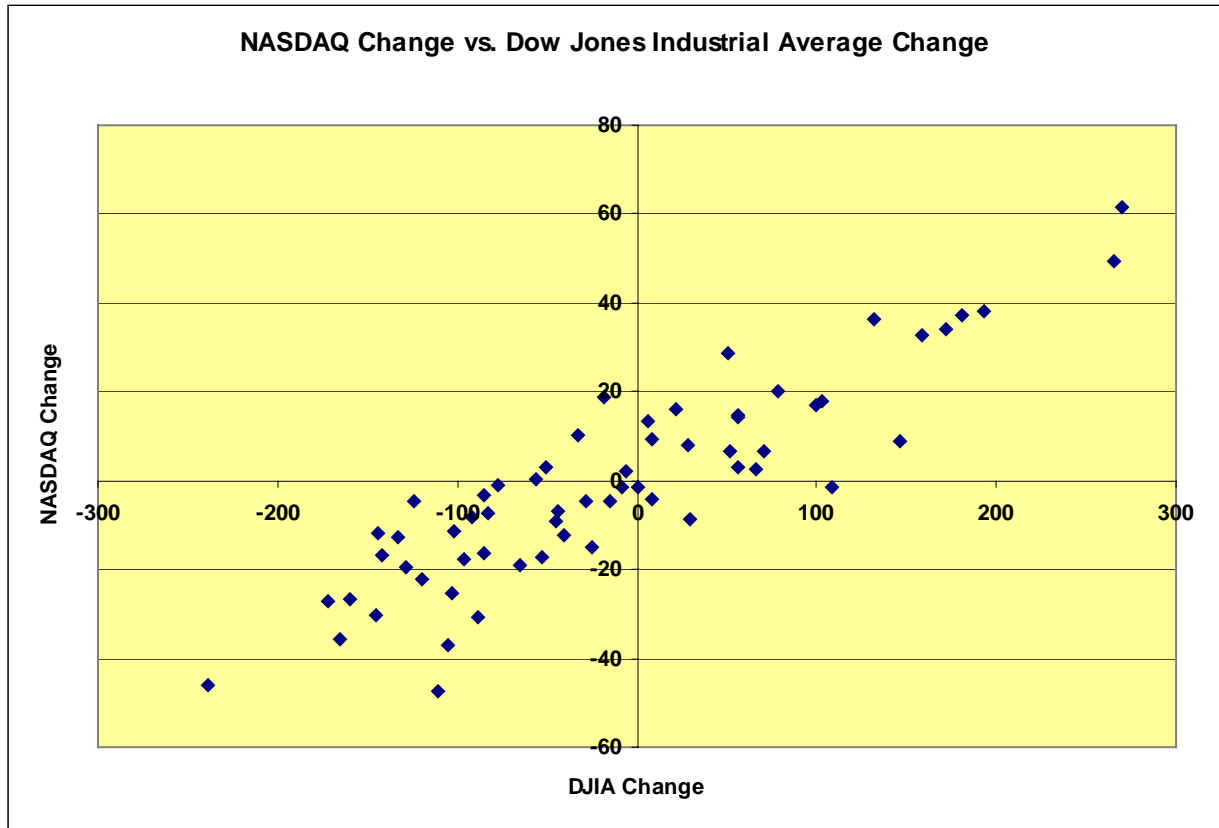
$$H_0: p_{\text{Less than HS}} = p_{\text{College Grad}}$$

$$H_1: p_{\text{Less than HS}} \neq p_{\text{College Grad}}$$

Clearly show your work.

**Problem 3:** (40%; a=12%; b=12%; c=8%; d=8%)

The closing averages for the NASDAQ (a stock exchange with a large number of high tech stocks) and the Dow Jones Industrial Average (a weighted average of 30 very large companies) were recorded from December 11, 2002 through March 13, 2003. There were a total of 63 business days during this interval during which the markets were open. The change in the two indices were computed from these 63 closing values, resulting in 62 **changes**. These changes are plotted below



(source = <http://table.finance.yahoo.com/k?s=^dji&g=d> and <http://table.finance.yahoo.com/k?s=^ixic&g=d>)

A summary of the relevant data is as follows, where  
Y = NASDAQ Change and  
X = DOW JONES Change.

$$\sum_{i=1}^{62} Y_i = -55.82$$

$$\sum_{i=1}^{62} Y_i^2 = 30,752.79$$

$$\sum_{i=1}^{62} X_i = -767.39$$

$$\sum_{i=1}^{62} X_i^2 = 755,864.41$$

$$\sum_{i=1}^{62} X_i Y_i = 135,987.53$$

- a) The regression output (from EXCEL) for these data is shown below. Fill in the 12 boxes that are shown with heavy lines. If a P-value is very small, just indicate that it is smaller than a number whose value you can readily estimate.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.893766
R Square	
Adjusted R Square	Ignore this cell
Standard Error	
Observations	62

ANOVA

	df	SS	MS	F	Significance F
Regression	1		24525.734		
Residual		6176.796			
Total	61	30702.531			

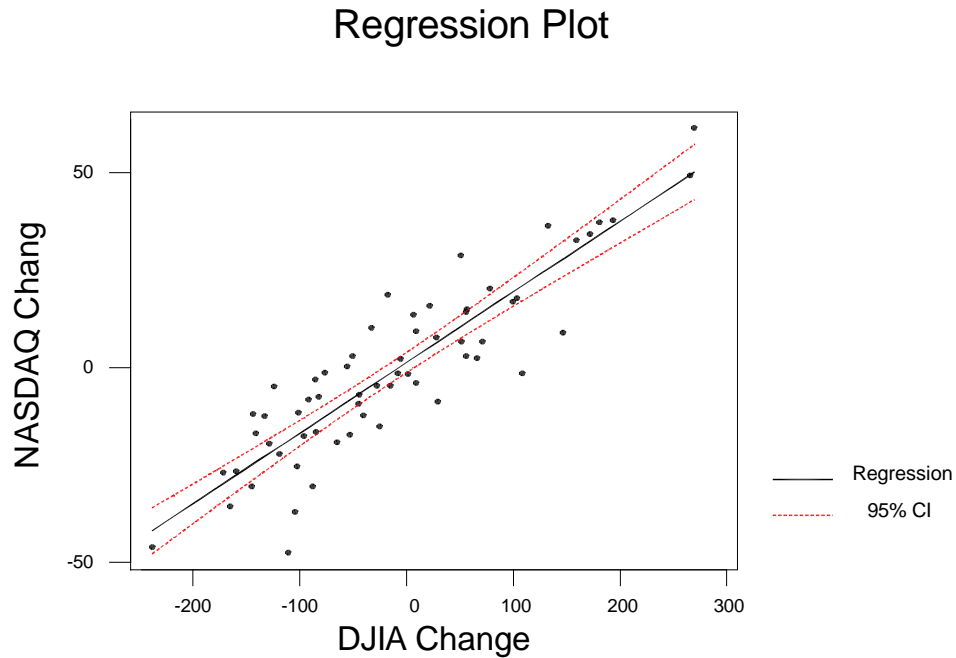
	Coefficients	Standard Error	t Stat	P-value
Intercept		1.296750	1.035936	
DJIA Change		0.011744		

- b) In listening to the reports on the closing averages, it seemed to me that you could reasonably well predict the change in the NASDAQ by dividing the change in the Dow Jones Average by 4. Test whether or not this is true using  $\alpha=0.05$ . You will need to perform **two** different hypothesis tests to see if this is true. Clearly state the null and alternative hypotheses and show your work.

Do you believe this is a good model? If not, propose an alternative model that you think would work well and that would be easy to use (since using this regression is not easy if I just want to guess the change in the NASDAQ after I hear the change in the Dow Jones. I would only have about 10 seconds to do the guestimation in my head.)



- c) When I plot the data and the regression line and the 95% confidence intervals for the line, I found that many of the observations fell outside of the 95% confidence interval. Specifically, you get the following (from MiniTab).



Is this a surprising result? Briefly explain your answer.

- d) Find a 95% prediction interval for the change in the NASDAQ on a day when the Dow Jones Industrial average lost 100 points.

