NAME: ______________Solution__________________________________ RPI ID #: _____________________

ENGR 2600 Modeling and Analysis of Uncertainty Section 4 Fall 2008 Exam 3

Rules of the Game:

1. Work entirely alone. Do not give or solicit assistance from any other student.

2. Do not open any IM or email programs or web browsers. Turn off cell phones. Running such software will be regarded as evidence of cheating.

3. This exam is open-book/open-notes/open-computer. Use any written materials you wish.

4. There are 5 questions, weighted equally. Answer all 5. Tip: Do the easier questions first.

5. Feel free to use the restrooms as necessary.

6. If you have a question, bring it down front so as to minimize disruption.

Got to pay your dues
If you want to sing the blues
And you know it don't come easy.
-- Ringo

Question 1: Copy from the list below the method or technique you would use to respond to the following situations.

a) Graphically test whether a sample of 15 data values are from an exponential distribution.
probability plot

b) Estimate a 95% confidence interval for the mean value of a sample of 4 values from a normal distribution.
CI based on one-sample t-test

c) Test whether the number of defective units produced in one day predicts the number produced the next day.
chi square test of independence

d) Determine whether college GPA predicts starting salary.
t-test on slope of regression line

e) Estimate how much each GPA point contributes to estimated starting salary.
estimate slope of regression model

f) Test whether the weight of a 1/4" metal screw is stable over time.
S and Xbar charts

g) Determine whether a new gasoline additive increases mean fuel efficiency (MPG).
two sample t-test with one-sided alternative

h) Test whether the number of defects in 10 square meter plastic panels has a Poisson distribution.
chi square goodness of fit test

i) Identify the most unusual sample value of tensile strength, taking account of % carbon in the alloy.
outliers from regression model

j) Test whether all majors have the same proportions of sophomores, juniors and seniors.
chi square test of homogeneity

- Bayes Theorem
- Binomial distribution
- Boxplot
- Central Limit Theorem
- Chi square goodness of fit test
- Chi square test of homogeneity
- Chi square test of independence
- CI based on one-sample t-test
- Estimate slope of regression model
- Joint probability distribution
- Normal distribution
- Outliers from regression model
- Poisson process
- Probability plot
- S and Xbar charts
- Student t-test on slope of regression line
- Two sample t-test with one-sided alternative

Question 2: Open the Minitab dataset *Studnt14\CollMass.mtw*.

Focus on the mean SATM (SAT Math) scores of applicants to private and public colleges. Test Ho: mean scores are equal versus Ha: Private schools have a higher mean score.

a) ____50.4718___________ = point estimate of $\mu_{private} - \mu_{public}$

b) ____3.05___________ = observed T value

c) ______52_________ = d.f.

d) ______.002_________ = p-value

e) ______reject_________ = reject or retain Ho for $\alpha = 0.01$

Two-Sample T-Test and CI: SATM, PubPriv

```
Two-sample T for SATM
PubPriv    N   Mean  StDev  SE Mean
Private   45  563.9   91.7       14
Public    13  513.5   33.6      9.3
Difference = mu (Private) - mu (Public)
Estimate for difference:  50.4718
95% lower bound for difference:  22.7706
T-Test of difference = 0 (vs >): T-Value = 3.05  P-Value = 0.002  DF = 52
```

The above two sample t-test works with existing data. If you were planning a future comparison of two means, you would need to estimate the sample size required to tell them apart if they differ by a certain amount. Suppose you want to distinguish a difference in means of 10 seconds. What sample size do you need in each group to detect this difference with probability 0.9 if the standard deviations are 8 seconds, the desired Type I error rate is 5%, and the alternative hypothesis is two sided?

f) _____15__________ = sample size in each group

Power and Sample Size

```
2-Sample t Test
Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 8
            Sample  Target
Difference    Size   Power  Actual Power
        10      15     0.9      0.910482
The sample size is for each group.
```

Question 3: Read in the Minitab datset *Data/Exh_qc.mtw*. Focus on the measurements of Faults in C4, as collected by shift (in C5). Create Xbar and S charts for Faults using shift as a rational subgroup. Be sure to use Sbar to estimate the standard deviation of Faults and be sure that all tests are applied.

a) Why do the control limits change throughout the charts?

subgroup sizes differ

b) What is the best point estimate of the mean value of Faults?

Xbarbar = 0.153

c) Does the S chart indicate a process that is in control? (yes/no)

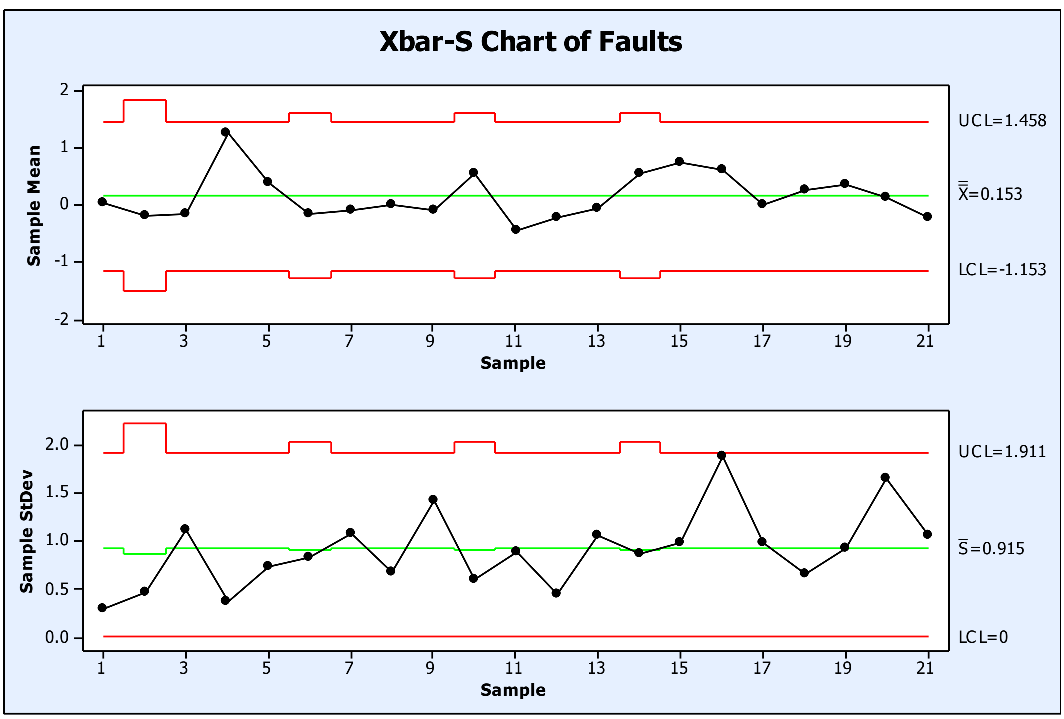
yes

d) Does the Xbar chart indicate a process that is in control? (yes/no)

yes

e) If the next shift in the future produces a subgroup with mean = 2, what would you recommend?

Since the process appears to be out of control, search for assignable causes.



Question 4. Consider the null hypothesis that a series of 200 counts were generated by simultaneously flipping 5 fair coins (thus, a binomial model) a total of 200 times. The observed counts are:

data	Count
0	7
1	23
2	67
3	66
4	35
5	2
N=	200

a) What is the expected number of 5's?

6.25

b) How many degrees of freedom are there for a chi-square goodness of fit test?

6-1-0 = 5

c) What is the critical value of the chi-square test of goodness of fit in this case, assuming $\alpha = 0.05$?

11.0705

d) The observed chi-square statistic in this case is 6.128. Would you reject or retain the null hypothesis?

retain

e) What is the p-value of the observed chi-square statistic?

p=0.294

MINITAB Student - Untitled - [Worksheet 3 ***]

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	data	value	binomial	expected	observed	chi-square	test statistic	df	p-value	
1	4	0	0.03125	6.25	7	0.090	6.128	5	0.294	
2	3	1	0.15625	31.25	23	2.178				
3	3	2	0.31250	62.50	67	0.324				
4	3	3	0.31250	62.50	66	0.196				
5	3	4	0.15625	31.25	35	0.450				
6	2	5	0.03125	6.25	2	2.890				
7	3									
8	3									

To test independence, the following table and analysis were created in Minitab. Fill in the 4 blacked out numbers, then draw a conclusion about the null hypothesis of independence.

__________________ reject/ not reject?

Data support the null hypothesis of independence, so do not reject it.

Tabulated statistics: lagged, coded

Rows: lagged   Columns: coded

	1	2	3	4	All
1	9 4.37	6 9.76	10 9.62	4 5.25	29 29.00
2	11 10.10	27 22.56	17 22.22	12 12.12	67 67.00
3	7 9.95	21 22.22	25 21.89	13 11.94	66 66.00
4	3 5.58	13 12.46	14 12.27	7 6.69	37 37.00
Missing	0 *	0 *	0 *	1 *	* *
All	30 30.00	67 67.00	66 66.00	36 36.00	199 199.00

Cell Contents: Count
                  Expected count

Pearson Chi-Square = 11.795, DF = 9, P-Value = 0.225

Question 5: Open the Minitab dataset *Stdnt14\Lakes.mtw*. Predict the historical PH level PHHist from the current PH level, PHCurnt.

a) Record the estimated regression model.

`PHHist = 0.940 + 0.844 PHCurnt`

b) Record the predicted R-square and briefly explain why predicted R-square is useful.

`R-Sq(pred) = 75.62%; this statistic approximates the R-square that will be obtained when the model is used on new data, i.e., data not used in estimating the model`

c) Record the 95% prediction interval for a lake whose current PH = 5.0

`(4.3615, 5.9577)`

d) Would such a lake be an outlier among the predictors? (yes/no) How do you know?

Yes; it is flagged with an "X" in the output.

e) For which lake (observation number) does the model most <u>over</u>-predict the historical PH? By how much?

Lake (observation #) 131, overpredicted by 0.9628

Regression Analysis: PHHist versus PHCurnt

```
The regression equation is
PHHist = 0.940 + 0.844 PHCurnt
Predictor      Coef  SE Coef      T      P
Constant     0.9397   0.2752   3.41  0.001
PHCurnt     0.84397  0.03884  21.73  0.000
S = 0.394681   R-Sq = 76.3%   R-Sq(adj) = 76.1%
PRESS = 23.5120   R-Sq(pred) = 75.62%
Analysis of Variance
Source           DF      SS      MS       F      P
Regression        1  73.554  73.554  472.19  0.000
Residual Error  147  22.899   0.156
Total           148  96.453
Unusual Observations
Obs  PHCurnt  PHHist     Fit  SE Fit  Residual  St Resid
 22     6.13  6.9333  6.1133  0.0478    0.8200     2.09R
 38     4.90  4.9732  5.0731  0.0892   -0.0998    -0.26 X
 47     7.78  6.7027  7.5058  0.0433   -0.8031    -2.05R
 48     6.45  7.5097  6.3833  0.0396    1.1264     2.87R
 82     7.56  8.1093  7.3201  0.0382    0.7891     2.01R
130     6.99  6.0109  6.8391  0.0324   -0.8282    -2.11R
131     7.45  6.2645  7.2273  0.0361   -0.9628    -2.45R
139     5.13  5.8956  5.2693  0.0808    0.6263     1.62 X
Predicted Values for New Observations
New
Obs     Fit  SE Fit        95% CI             95% PI
  1  5.1596  0.0855  (4.9906, 5.3285)  (4.3615, 5.9577)X
X denotes a point that is an outlier in the predictors.
Values of Predictors for New Observations
New
Obs  PHCurnt
  1     5.00
```