Attached to this exam are 5 pages of Minitab regression analysis printout. Use them in answering the following questions. You need NOT compute by hand anything that you can get from the printout. (Indeed, you will be wise to avoid wasting time doing hand calculations for things that are obtainable from the printout.)

The data used in creating the printout come from a study of a nitride etch process on a single wafer plasma etcher. The process variables studied were

- \( x_1 \) = power applied to the cathode (W)
- \( x_2 \) = pressure in the reaction chamber (mTorr)
- \( x_3 \) = gap between the anode and the cathode (cm)
- \( x_4 \) = flow of the reactant gas C\(_2\)F\(_6\)

and the response variable was

- \( y \) = selectivity of the process (SiN/polysilicon)

The first 2\( \frac{1}{2} \) pages of the printout concern a simple linear regression analysis based on the model

\[
y_i = \beta_0 + \beta_3 x_{3i} + \epsilon_i.
\]

Until further notice, give answers based on (or related to) this SLR model.

(a) What fraction of the raw variability in \( y \) is explained using \( x_3 \) as a predictor variable?

(b) What is the sample correlation between \( y \) and \( x_3 \)? (Give a number.)

(c) There are two plots on page 2 of the printout. These are plots of standardized residuals versus \( \hat{y} \) and standardized residuals versus \( x_3 \). What difficulty with the simple linear regression model do they reveal? Explain.

For purposes of answering the questions (d)-(h) ignore the difficulty discovered in part (c).

(d) What is indicated by the plot on the top of page 3 on the printout?
(e) If one uses $\alpha = .05$, will one reject $H_0: \mu_{y|x_3} = \beta_0 + \beta_3 x_3$ based on a formal lack of fit test? Explain.

Reject? yes no (circle one)
Explanation:

(f) Give a 90% two-sided confidence interval for the increase in mean value of selectivity ($y$) that accompanies a 1 cm increase in the gap between the anode and the cathode. (No need to simplify after plugging in.)

(g) Give a 95% two-sided prediction interval for the next selectivity ($y$) that accompanies a .8 cm gap ($x_3$).

(h) As it turns out, the data have $\bar{x}_3 = .9636$ and $s_{x_3}^2 = .030545$. Use these facts and find a 95% confidence interval for the mean selectivity that accompanies a .9 cm gap. (No need to simplify.)

Beginning in the middle of page 3 of the printout, there is a multiple linear regression analysis of the data. Note that there are both an "all possible subsets" regression and a regression of $y$ on $x_1, x_2, x_3$ and $x_4$ (and some plots and some sample correlations). Use these to answer questions (i) through (r).

(i) Based on the "all possible subsets regression" output, what reduced model seems like one that should be investigated as a possible "simple" explanation of $y$? Explain in terms of $R^2, s$ and $C_p$. 

-2-
Model: __________________
Explanation:

(j) What degrees of freedom would be used in a formal test for lack of fit to the multiple linear regression model including all 4 predictors $x_1, x_2, x_3$ and $x_4$?

\[
\text{numerator } df = \_\_\_\_\_, \quad \text{denominator } df = \_\_\_\_.
\]

(k) Give the value of an $F$ statistic, its degrees of freedom and the corresponding $p$-value for testing whether the variables $x_1, x_2, x_3$ and $x_4$ together provide any ability to predict or explain $y$.

\[
f = \_\_\_\_, \quad df = \_\_\_, \_\_\_, \quad p\text{-value} = \_\_\_.
\]

(l) Find the value of an $F$ statistic and its degrees of freedom for testing whether after taking $x_1$ and $x_3$ into account, the variables $x_2$ and $x_4$ provide significant additional ability to explain or predict $y$. (This is not an easy question, but there is enough information on the printout to allow you to find this $f$.)

\[
f = \_\_\_, \quad df = \_\_, \_\_,
\]

(m) Find the value of a $t$ statistic, its degrees of freedom and the corresponding $p$-value for testing whether after taking $x_1, x_2,$ and $x_3$ into account, $x_4$ adds important explanatory power to the multiple regression model.

\[
t = \_\_\_, \quad df = \_\_, \quad p\text{-value} = \_\_\_.
\]

(n) Give a 97.5% lower prediction bound for the next value of $y$ when $x_1 = 325, x_2 = 500, x_3 = .8$ and $x_4 = 200$. 

-3-
(o) Comment on the appearance of the plot of $e^*$ versus $\hat{y}$ given at the bottom of page 4 of the printout.

(p) The first plot on page 5 of the printout is a plot of $y$ versus $\hat{y}$. It appears to be fairly linear. What sample correlation should be associated with this plot? (Give a number.)

(q) Based on the MLR model involving all 4 $x$'s, give a 90% two-sided confidence interval for increase in mean $y$ that accompanies a 1 cm increase in gap ($x_3$) if $x_1$, $x_2$ and $x_4$ are held fixed. (There is no need to simplify after plugging in.)

(r) Notice that from the first (SLR) regression run $b_3 = -1.0458$ while from the last (MLR) regression run $b_3 = -1.0412$. This difference is not simply numerical error in Minitab. Explain why you are not surprised that there is a difference in these two figures. (What do the correlations given on page 5 of the printout have to say about this?)
MTB > print c1-c5

Data Display

Row y x1 x2 x3 x4
1 1.63 275 450 0.8 125
2 1.37 275 500 1.0 160
3 1.10 275 550 1.2 200
4 1.58 300 450 1.0 200
5 1.26 300 500 1.2 125
6 1.72 275 450 0.8 125
7 1.65 300 550 0.8 160
8 1.42 325 450 1.2 160
9 1.26 325 500 1.2 200
10 1.54 325 550 1.0 125
11 1.72 275 450 0.8 125

MTB > Name c6 = 'FITS1' c7 = 'SRES1' c8 = 'PFIT1'
MTB > Regress 'y' 1 'x3';
SUBC> Fits 'FITS1';
SUBC> SResiduals 'SRES1';
SUBC> Constant;
SUBC> Pure;
SUBC> Predict 'x3';
SUBC> PFits 'PFIT1'.

Regression Analysis

The regression equation is
y = 2.52 - 1.05 x3

Predictor Coef StDev T P
Constant 2.5242 0.1711 14.75 0.000
x3 -1.0458 0.1750 -5.98 0.000

$S = 0.09670$ R-Sq = 79.9% R-Sq(adj) = 77.6%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.33410</td>
<td>0.33410</td>
<td>35.73</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>0.08416</td>
<td>0.00935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>0.41825</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>x3</th>
<th>y</th>
<th>Fit</th>
<th>StDev</th>
<th>Fit Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.20</td>
<td>1.1000</td>
<td>1.2692</td>
<td>0.0506</td>
<td>-0.1692</td>
<td>-2.05R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual

<table>
<thead>
<tr>
<th>Fit</th>
<th>StDev</th>
<th>CI 95.0%</th>
<th>PI 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6875</td>
<td>0.0409</td>
<td>(1.5950, 1.7800)</td>
<td>(1.4500, 1.9250)</td>
</tr>
<tr>
<td>1.4783</td>
<td>0.0298</td>
<td>(1.4108, 1.5459)</td>
<td>(1.2493, 1.7073)</td>
</tr>
<tr>
<td>1.2692</td>
<td>0.0506</td>
<td>(1.1547, 1.3837)</td>
<td>(1.0222, 1.5161)</td>
</tr>
<tr>
<td>1.2692</td>
<td>0.0506</td>
<td>(1.1547, 1.3837)</td>
<td>(1.0222, 1.5161)</td>
</tr>
<tr>
<td>1.6875</td>
<td>0.0409</td>
<td>(1.5950, 1.7800)</td>
<td>(1.4500, 1.9250)</td>
</tr>
<tr>
<td>1.6875</td>
<td>0.0409</td>
<td>(1.5950, 1.7800)</td>
<td>(1.4500, 1.9250)</td>
</tr>
<tr>
<td>1.2692</td>
<td>0.0506</td>
<td>(1.1547, 1.3837)</td>
<td>(1.0222, 1.5161)</td>
</tr>
<tr>
<td>1.6875</td>
<td>0.0409</td>
<td>(1.5950, 1.7800)</td>
<td>(1.4500, 1.9250)</td>
</tr>
<tr>
<td>1.4783</td>
<td>0.0298</td>
<td>(1.4108, 1.5459)</td>
<td>(1.2493, 1.7073)</td>
</tr>
<tr>
<td>1.6875</td>
<td>0.0409</td>
<td>(1.5950, 1.7800)</td>
<td>(1.4500, 1.9250)</td>
</tr>
</tbody>
</table>

Pure error test - $F = 0.14$ P = 0.7214 DF(pure error) = 8
MTB > Plot 'SRES1' 'FITS1';
SUBC> Symbol '***'.

Character Plot

MTB > Plot 'SRES1' 'x3';
SUBC> Symbol '***'.

Character Plot

MTB > Sort 'SRES1' c8;
SUBC> By 'SRES1'.
MTB > Set c9
DATA> 1c 1:11 / 1
DATA> End.
MTB > sub .5 c9 c9
MTB > div c9 11.0 c9
MTB > InvCDF c9 c9;
SUBC> Normal 0.0 1.0.
MTB > Plot c8 c9;
SUBC> Symbol '***';
SUBC> XLabel 'standardized residual quantile';
SUBC> YLabel 'std normal quantile'.

-2-
Character Plot

![Character Plot](image)

MTB > BReg 'y' 'x1' 'x2' 'x3' 'x4';
SUBC> NVars 1 4;
SUBC> Best 4.

Best Subsets Regression

Response is y

<table>
<thead>
<tr>
<th>Vars</th>
<th>R-Sq</th>
<th>R-Sq (adj)</th>
<th>C-p</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79.9</td>
<td>77.6</td>
<td>26.6</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>17.5</td>
<td>8.3</td>
<td>130.7</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>6.7</td>
<td>0.0</td>
<td>148.7</td>
<td>X</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.0</td>
<td>158.8</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>87.6</td>
<td>84.3</td>
<td>15.7</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>85.3</td>
<td>81.6</td>
<td>19.6</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>80.4</td>
<td>75.5</td>
<td>27.7</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>20.5</td>
<td>0.7</td>
<td>127.6</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>95.7</td>
<td>93.9</td>
<td>4.1</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>89.1</td>
<td>84.4</td>
<td>15.2</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>85.4</td>
<td>79.1</td>
<td>21.4</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>24.8</td>
<td>0.0</td>
<td>122.5</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>96.4</td>
<td>94.0</td>
<td>5.0</td>
<td>X</td>
</tr>
</tbody>
</table>

MTB > Name c10 = 'FITS2' c11 = 'SRES2'
MTB > Regress 'y' 4 'x1' 'x2' 'x3' 'x4';
SUBC> Fits 'FITS2';
SUBC> SResiduals 'SRES2';
SUBC> Constant;
SUBC> Predict 'x1' 'x2' 'x3' 'x4'.

Regression Analysis

The regression equation is
\[ y = 2.29 + 0.00327 \, x_1 - 0.00133 \, x_2 - 1.04 \, x_3 - 0.000532 \, x_4 \]
Predictor Coef StdDev T P
Constant 2.2896 0.2544 9.00 0.000
x1 0.0032704 0.0007621 4.29 0.005
x2 -0.0013315 0.0003811 -3.49 0.013
x3 -1.04120 0.09526 -10.93 0.000
x4 -0.0005321 0.0005094 -1.04 0.337

S = 0.05007 R-Sq = 96.4% R-Sq(adj) = 94.0%

Analysis of Variance
Source DF SS MS F P
Regression 4 0.40321 0.10080 40.21 0.000
Error 6 0.01504 0.00251
Total 10 0.41825

Source Seq SS
x1 1 0.00267
x2 1 0.08305
x3 1 0.31476
x4 1 0.00273

Fit StDev Fit 95.0% CI 95.0% PI
1.6903 0.0276 (1.6228, 1.7578) (1.5504, 1.8302)
1.3968 0.0232 (1.3401, 1.4536) (1.2618, 1.5319)
1.1007 0.0433 (0.9949, 1.2066) (0.9388, 1.2627)
1.5239 0.0333 (1.4424, 1.6054) (1.3767, 1.6710)
1.2490 0.0328 (1.1988, 1.3692) (1.1425, 1.4355)
1.6903 0.0276 (1.6228, 1.7578) (1.5504, 1.8302)
1.6203 0.0330 (1.5396, 1.7010) (1.4735, 1.7670)
1.4187 0.0391 (1.3230, 1.5144) (1.2632, 1.5742)
1.7403 0.0393 (1.6511, 1.8435) (1.5915, 1.9031)
1.5124 0.0388 (1.4175, 1.6074) (1.3574, 1.6674)
1.6903 0.0276 (1.5628, 1.7578) (1.5504, 1.8302)

MTB > Plot 'SRES2' 'FITS2';
SUBC> Symbol '***';
SUBC> XLabel 'yhat';
SUBC> YLabel 'standardized residual'.

Character Plot

MTB > Plot 'y' 'FITS2';
SUBC> Symbol '***';
SUBC> Xlabel 'yhat';
SUBC> Ylabel 'y'.

Character Plot

1.60+
1.40+
1.20+
1.08 1.32

MTB > Correlation 'x1' 'x2' 'x3' 'x4'.

Correlations (Pearson)

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x2</td>
<td>0.214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x3</td>
<td>0.214</td>
<td>0.214</td>
<td></td>
</tr>
<tr>
<td>x4</td>
<td>0.210</td>
<td>0.210</td>
<td>0.210</td>
</tr>
</tbody>
</table>
...whether after taking X3.

...standardized residuals versus 

...- Give the value of an F statistic, its degrees of freedom and the corresponding 

...p-value.

...Based on the "all possible subsets regression" output, what reduced model seems like one that should 

...be investigated as a possible "simple" explanation of y? Explain in terms of 

...R².6

...Find the value of a t statistic, its degrees of freedom and the corresponding 

...p-value.

...What difficulty with the simple linear regression model do they reveal?

...As it turns out, the data have b3 = -0.9036 and b2 = -0.000145. Use these facts and find a 95% 

...confidence interval for the mean selectivity that accompanies a 1 cm gap. (Be sure to simplify after 

...plugging in.) Use b3 ± t (s.e. of b3) to find this 

...interval.

...based on a formal lack or fit test? Explain.

...adds important explanatory power to the multiple 

...regression model. Explanation: It appears that the Var (Y given the X's) is much smaller than the total 

...variance, even with 3 or 4. That is, the "constant term" part of the SLR model assumptions appears to be 

...violated.

...Based on the "all possible subsets regression" output, what reduced model seems like one that should 

...be investigated as a possible "simple" explanation of y? Explain in terms of R², s and C."}

...Model: y = b0 + b1x1 + b2x2 + b3x3 + b4x4. 

...Explaination: This model has R² such as we use it in the full model and 

...essentially the same as the full model. Further, p = 3 

...is less than R² = 4 (same as the "ideal value of R²") with one less 

...predictor, than the full model.

...What degree of freedom would be used in a formal test for lack of fit to the multiple linear regression 

...model including all 4 predictors x1, x2, x3, and x4? What are the 95% and 99% confidence intervals for 

...the slope of x3 on x4? (Give a number.)

...Multiplier

...desired

...numerator df = 4. 

denominator df = 2.

...Give the value of an F statistic, its degrees of freedom and the corresponding p-value for testing 

...whether the variables x1, x2, and x4 together provide any ability to predict or explain y.

...Find the value of an F statistic, its degrees of freedom and the corresponding p-value for testing 

...whether after taking x1, x2, and x4 into account, x3 adds important explanatory power to the multiple 

...regression model.

...What is indicated by the plot on the top of page 3 on the printout?

...This is a normal plot for the standardized residuals. Hopefully, it is quite linear and therefore provides no 

...indication of obvious problems with the "normality" part of the SLR model assumptions.

...If one uses x1 = 0.05, will one reject H₀: b1 = 0.0 + b2x3 based on a formal lack of fit test? Explain. 

...Reject yes (circle one)

...Explanation: the p-value for the lack of fit test given at the bottom of page 1 is .72 which is very large (in particular it is larger than .05)

...Give a 90% two-sided confidence interval for the increase in mean value of selectivity (γ) that 

...accompanies a 1 cm increase in the gap between the anode and the cathode. (No need to simplify after 

...plugging in.) Use b3 ± t (s.e. of b3) to find this interval.

...Give a 95% two-sided prediction interval for the next selectivity (γ) that accompanies a .8 cm gap 

...in the gap.

...Give a 95.3% lower prediction bound for the next value of y when x1 = 330, x2 = 500, x3 = -3 and x4 = 100.

...Comment on the appearance of the plot of y versus x1 given at the bottom of page 4 of the printout. 

...This plot (apparently) looks like "polynomial regression", and thus indicates no problems with the 

...full MLR model.

...The first plot on page 5 of the printout is a plot of y versus x1. It appears to be fairly linear. What 

...sample correlation should be associated with this plot? (Give a number.)

...Based on the MLR model involving all 4 x's, give a 90% two-sided confidence interval for increase in 

...mean of y that accompanies a 1 cm increase in gap (x3). (No need to simplify after plugging in.) Use b3 ± t (s.e. of b3) to find this interval.

...Notice that from the first (SLR) regression run 

...b3 = -1.0458 while from the last (MLR) regression 

...run b3 = -1.0453. This difference is not simply numerical error in Minitab. Explain why you are not 

...surprised that there is a difference in these two figures. (What do the correlations given on page 5 of the 

...printout have to say about this?)

...true degree of multicollinearity in this problem. Correlations between predictors are nonzero, so estimates of 

...β's are expected to change depending upon what 

...terms are included in a model.