University of Cincinnati, Dept. of Mathematical Sciences

MATH 533: Applied Regression Analysis, Spring 2001

Final Exam, June 04, 2001

Name Key

1. A researcher wants to evaluate the difference in the mean film thickness of a coating placed on silicon wafers using three different processes (Process 1, 2, and 3). Eight wafers are randomly assigned to each of the processes. The film thickness (Y) and the temperature (X) in the lab during the coating process are recorded on each wafer. The researcher is concerned that fluctuations in temperature may affect the thickness of the coating. We are going to the use regression approach discussed in class. Parts of the SAS output are given below.

```
                              The REG Procedure
                                Model: MODEL1
                           Dependent Variable: Y

                            Analysis of Variance

                                      Sum of           Mean
Source                   DF          Squares         Square    F Value    Pr > F

Model                     3            13246     4415.42589      62.79    <.0001
Error                    20       1406.34732       70.31737
Corrected Total          23            14653

            Root MSE              8.38555    R-Square     0.9040
            Dependent Mean      119.87500    Adj R-Sq     0.8896
            Coeff Var             6.99524

                            Parameter Estimates

                          Parameter       Standard
        Variable    DF     Estimate          Error    t Value    Pr > |t|

        Intercept    1    121.22791        1.71785      70.57      <.0001
        I1           1     -1.08823        2.42097      -0.45      0.6579
        I2           1     20.65443        2.42240       8.53      <.0001
        X_adj        1      4.05874        0.43580       9.31      <.0001

                        Covariance of Estimates

Variable        Intercept              I1                 I2                X_adj

Intercept     2.9509923096      -0.005275517       -0.013188792       0.0633062037
I1            -0.005275517      5.8610993626       -2.926593044       -0.015826551
I2            -0.013188792      -2.926593044       5.8680234786       -0.039566377
X_adj         0.0633062037      -0.015826551       -0.039566377       0.1899186112
```

(a) Using the SAS output, write the least squares regression line. Also find estimated mean thickness (at X_adj = 0) for each of the three processes. Note the X_adj is mean-adjusted X variable. Is the model reasonable? MODEL : $Y_{ij} = \mu_. + \beta_1 I_{ij1} + \beta_2 I_{ij2} + \gamma x_{ij} + \varepsilon_{ij}$

$$\hat{Y}_{ij} = 121.23 - 1.088\, I_{ij1} + 20.65\, I_{ij2} + 4.06\, x_{ij}, \quad i = 1, 2, 3 \quad j = 1, \cdots, 8$$

$$\begin{cases} 1, & i=1 \\ 0, & i=2 \\ -1, & i=3 \end{cases} \qquad \begin{cases} 1, & i=2 \\ 0, & i=1 \\ -1, & i=3 \end{cases}$$

proc 1 $(i=1)$: $\hat{Y}_{ij} = 121.23 - 1.088 = 120.14$

proc 2 $(i=2)$: $\hat{Y}_{ij} = 121.23 + 20.65 = 141.88$

proc 3 $(i=3)$: $\hat{Y}_{ij} = 121.23 + 1.088 - 20.65 = 101.67$

(b) Using the model and the variance-covariance matrix of the estimates, calculate the estimated standard error of the estimated mean thickness in (a) for each of the three processes. Also construct 95% confidence intervals for the population mean response for each of the three processes.

proc 1 : $Y_{ij} = \mu_{.} + \beta_1 + \varepsilon_{ij}$

$s^2(\hat{Y}_{ij}) = s^2(\hat{\mu}_{.}) + s^2(\hat{\beta}_1) + 2\,\widehat{Cov}(\hat{\mu}_{.}, \hat{\beta}_1) = 2.95 + 5.86 + 2(-.0053) = 8.8$

CI : $\hat{Y}_1 \pm t_{.975}(20) \cdot \sqrt{8.8} = 120.14 \pm 2.086(2.97) = 120.14 \pm 6.19$

proc 2 : Similarly.

proc 3 : $Y_{ij} = \mu_{.} - \beta_1 - \beta_2$

$s^2(\hat{Y}_{ij}) = s^2(\hat{\mu}_{.}) + s^2(\hat{\beta}_1) + s^2(\hat{\beta}_2) - 2\,\widehat{Cov}(\hat{\mu}_{.}, \hat{\beta}_1) - 2\,\widehat{Cov}(\hat{\mu}_{.}, \hat{\beta}_2) + 2\,\widehat{Cov}(\hat{\beta}_1, \hat{\beta}$

$= \ldots\ldots$

(c) Find a 95% confidence interval comparing process 1 and 3 only. Interpret the interval.

$\hat{Y}_1 - \hat{Y}_2 \pm t_{.975}(20)\sqrt{s^2(\hat{Y}_1 - \hat{Y}_2)}$

$s^2(\hat{Y}_1 - \hat{Y}_2) = s^2(\hat{\beta}_1) + s^2(\hat{\beta}_2) - 2\,\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 5.86 + 5.87 - 2(-2.93)$

$= \ldots$

(d) To test for treatment (process) effect the researcher fitted an ANOVA model without the treatment effect and found the SSE for the reduced model was 7874.4. Write the hypotheses and perform the test.

$H_0: \beta_1 = \beta_2 = 0$ $H_1:$ at least one is not zero

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{2}}{\frac{SSE(F)}{20}} = \frac{\frac{7874.4 - 1406.3}{2}}{\frac{1406.3}{20}} = \ldots$$

(e) Following is the SAS output of the one-way ANOVA for the treatment effect (factor A).

```
                              The GLM Procedure

Dependent Variable: Y

                                            Sum of
 Source                      DF          Squares     Mean Square    F Value    Pr > F

 Model                        2       7147.00000      3573.50000      10.00    0.0009

 Error                       21       7505.62500       357.41071

 Corrected Total             23      14652.62500


               R-Square     Coeff Var      Root MSE        Y Mean

               0.487762      15.77085      18.90531      119.8750


 Source                      DF        Type I SS     Mean Square    F Value    Pr > F

 A                            2      7147.000000     3573.500000      10.00    0.0009
```

Note that the sample level mean responses are 119.125, 141.125, and 99.125 for process 1, 2, and 3, respectively. Calculate the estimated standard errors of the mean response estimates and construct 95% confidence intervals for the population mean responses for each of the three processes. Compare the results with your answer in (b). What did you find? Do you see the advantage of using ANCOVA?

proc 1 : $\bar{Y}_{1\cdot} = 119.125$, $s^2(\bar{Y}_{1\cdot}) = \frac{MSE}{n_1} = \frac{357.4}{8} = 44.675$

95% CI : $\bar{Y}_{1\cdot} \pm t_{.975}(21)\sqrt{44.675} = 119.125 \pm 2.08\,(6.68)$

$= 119.125 \pm 13.8$

Similarly for proc 2 and 3.

CI's from ANCOVA are much narrower.

ANCOVA is prefered.

2. A researcher conducted an experiment to compare the effects of three different insecticides on a variety of string beans. To obtain a sufficient amount of data, it was necessary to use four different plots of land. Since the plots had somewhat different soil fertility, drainage characteristics, and sheltering from winds, the researcher decided to conduct a randomized complete block design with the plots serving as the blocks. Each plot was subdivided into three rows. A suitable distance was maintained between rows within a plot so that the insecticides could be confined to a particular row. The insecticides were randomly assigned to the rows within a plot so that each insecticide appeared in one row within all four plots. The response of interest was the number of seedlings that emerged per row. The data and parts of SAS output for this study are given below.

	Insecticide			
Plot	1	2	3	Plot Mean
1	56	83	80	73
2	48	78	72	66
3	66	94	83	81
4	62	93	85	85
Insecticide Mean	58	87	80	Overall mean=75

```
                    General Linear Models Procedure
                        Class Level Information

                     Class     Levels    Values

                     INS            3    1 2 3

                     PLOT           4    1 2 3 4

                Number of observations in data set = 12


                            The SAS System

                    General Linear Models Procedure
```

```
Dependent Variable: Y
                                        Sum of              Mean
Source                  DF             Squares            Square    F Value      Pr > F

Model                    5        2270.0000000       454.0000000     104.77      0.0001

Error                    6          26.0000000         4.3333333

Corrected Total         11        2296.0000000

                  R-Square                C.V.          Root MSE                 Y Mean

                  0.988676            2.775555         2.0816660              75.000000


Source                  DF           Type I SS       Mean Square    F Value      Pr > F

INS                      2        1832.0000000       916.0000000     211.38      0.0001
PLOT                     3         438.0000000       146.0000000      33.69      0.0004
```

(a) Write an appropriate statistical model for this experimental situation and estimate all parameters in the model including the common variance of the error terms.

$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij}$, $i = 1, \ldots, 4$ $j = 1, \ldots, 3$ $\varepsilon_{ij} \sim N(0, \sigma^2)$

block effect

factor effect

$\hat{\mu}_{..} = \bar{Y}_{..} = 75$

$\hat{\tau}_1 = \bar{Y}_{.j} - \bar{Y}_{..} = 58 - 75 = -17$

$\vdots$

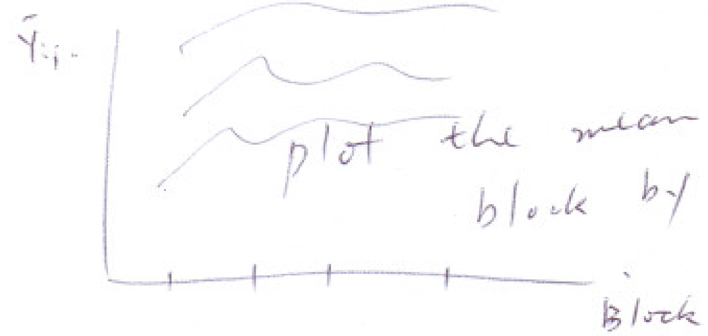
$\hat{\rho}_1 = \bar{Y}_{1.} - \bar{Y}_{..} = 73 - 75 = -2$

$\hat{\rho}_2 = \bar{Y}_{2.} - \bar{Y}_{..} = 66 - 75 = -9$

$\vdots$

$\hat{\sigma}^2 = MSE = 4.33$

(b) Provide a plot that will show no interaction between the two factors.



plot the mean response over block by Insecticide

(c) Perform a hypothesis testing for the main effect.

$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$

$H_1 :$ not all zero

$F = \frac{MSTR}{MSE} = \frac{916}{4.3} = 213$ huge!! reject H_0

Similarly for Block effect

(d) Write a statistical model for the completely randomized design with Insecticide effect only and draw a complete ANOVA table. What is the estimated error variance? Compare the estimate with that in (a).

$$Y_{ij} = \underbrace{\mu_{\cdot\cdot} + \tau_i}_{\mu_i} + \varepsilon_{ij} \quad , \quad \begin{matrix} i=1,2,3 \\ j=1,\cdots,4 \end{matrix} \qquad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Source	df	SS	MS	F	p-val
Insecticide	2	1832	916	17.77	0
Error	9	464	51.55		
total	11	2296			

$\hat{\sigma}^2 = 51.55 \gg 4.33$ ($\hat{\sigma}^2$ from RCBD)

(e) From the ANOVA table in (d), is the Insecticide effect significant? If so, provide 95% confidence interval for all possible pairwise comparisons using Bonferroni multiple comparison procedure.

Yes, p-value ≈ 0.

$$\left.\begin{matrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \mu_2 - \mu_3 \end{matrix}\right) \quad g = 3$$

$$B = t_{1-\frac{.05}{6}}(9) = t_{.9917}(9) \approx 2.9$$

CI for $\mu_1 - \mu_2$: $\bar{Y}_1 - \bar{Y}_2 \pm 2.9 \left(\sqrt{MSE\left(\frac{1}{4} + \frac{1}{4}\right)} \right)$

$58 - 87 \pm 2.9 \sqrt{51.55 \left(\frac{1}{2}\right)}$

$\vdots$

Similarly for others.

GOOD LUCK AND HAVE A GOOD BREAK!