STOR 664: FINAL EXAM
DECEMBER 14 2009

This is an open book exam. Course text, personal notes and calculator are allowed. You have 3 hours to complete the exam. Answers should preferably be written in blue books.

SHOW ALL WORKING: You are not expected to provide lengthy explanations or to reproduce standard results that are in the text, but you should show all working in enough detail to make clear how your answers were obtained.

The University Honor Code is in effect during this exam and you should sign the "pledge" at the front of your exam book. If you have questions, please ask the instructor.

The total number of points on the exam is 100. To guide you in allocating your time, the points available for each question or part of a question are shown in boldface type.

1. We are given three rectangular metal objects, for which the volumes are given by the formula

$$Volume_j \quad = \quad Length_j \times Breadth_j \times Height_j, \quad j = 1, 2, 3.$$

However, it is known that all three objects were made using exactly the same amount of metal, therefore *the three volumes are equal.* If we define

$$\begin{aligned} \theta_1 &= \log Length_1, & \theta_2 &= \log Breadth_1, & \theta_3 &= \log Height_1, \\ \theta_4 &= \log Length_2, & \theta_5 &= \log Breadth_2, & \theta_6 &= \log Height_2, \\ \theta_7 &= \log Length_3, & \theta_8 &= \log Breadth_3, & \theta_9 &= \log Height_3, \end{aligned}$$

the problem becomes to estimate $\theta_1, ..., \theta_9$ under the constraint

$$\theta_1 + \theta_2 + \theta_3 \;=\; \theta_4 + \theta_5 + \theta_6 \;=\; \theta_7 + \theta_8 + \theta_9.$$

Suppose we are given measurements of the form $y_i = \theta_i + \epsilon_i$, $i = 1, 2, ..., 9$, where ϵ_i are independent $N[0, \sigma^2]$.

By recasting the problem in the form

$$\begin{aligned} \theta_1 &= \beta_1 + \beta_2, \\ \theta_2 &= \beta_1 + \beta_3, \\ \theta_3 &= \beta_1 - \beta_2 - \beta_3, \\ \theta_4 &= \beta_1 + \beta_4, \\ \theta_5 &= \beta_1 + \beta_5, \\ \theta_6 &= \beta_1 - \beta_4 - \beta_5, \\ \theta_7 &= \beta_1 + \beta_6, \\ \theta_8 &= \beta_1 + \beta_7, \\ \theta_9 &= \beta_1 - \beta_6 - \beta_7, \end{aligned}$$

write down the normal equations for $\hat{\beta}_1, ..., \hat{\beta}_7$ (note: you are *not* being asked to solve the normal equations). Find the variances of $\hat{\beta}_1, ..., \hat{\beta}_7$ and show theoretically that each of $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$, $\hat{\theta}_2 = \hat{\beta}_1 + \hat{\beta}_3$, etc., has variance $\frac{7}{9}\sigma^2$. **[20 points.]**

2. A recent paper (by G.-H. Lee and M.-G. Shen, *Journal of Food Science* **74**, E519–E525, 2009) considered the influence of three process variables (feeding rate FR, air pressure AP and product temperature PT) on the production of spherical red ginseng capsules. The experimental design (in standardized units for each of the three variables) is as follows, together with the yield Y:

Number	FR	AP	PT	Y	Number	FR	AP	PT	Y
1	-1	-1	-1	76.27	9	0	0	0	84.91
2	-1	-1	1	67.29	10	-2	0	0	74.51
3	-1	1	-1	74.84	11	2	0	0	76.33
4	-1	1	1	64.95	12	0	-2	0	58.65
5	1	-1	-1	60.67	13	0	2	0	78.41
6	1	-1	1	61.02	14	0	0	-2	56.17
7	1	1	-1	64.30	15	0	0	2	58.43
8	1	1	1	62.69					

We fit the model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 \mathrm{FR}_i^2 + \beta_2 \mathrm{AP}_i^2 + \beta_3 \mathrm{PT}_i^2 + \beta_4 \mathrm{FR}_i \mathrm{AP}_i + \beta_5 \mathrm{FR}_i \mathrm{PT}_i + \beta_6 \mathrm{AP}_i \mathrm{PT}_i \\ &\quad + \beta_7 \mathrm{FR}_i + \beta_8 \mathrm{AP}_i + \beta_9 \mathrm{PT}_i + \epsilon_i \end{aligned}$$

with ϵ_i independent $N(0, \sigma^2)$ for some unknown σ^2.

You are given the following fit of the model, where some of the entries have been intentionally replaced by ?:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80.4333     6.9468  11.579 8.43e-05 ***
FRFR         -1.8129       ?         ?        ?
APAP         -3.5354     2.3667  -1.494   0.1955
PTPT         -6.3429       ?         ?        ?
FRAP          1.1337       ?         ?        ?
FRPT           ?         2.7849      ?        ?
APPT         -0.3588       ?         ?        ?
FR           -1.9394       ?         ?        ?
AP            2.5656       ?         ?        ?
PT             ?           ?         ?        ?
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 7.877 on 5 degrees of freedom
Multiple R-Squared: 0.7165,     Adjusted R-squared: 0.2062
F-statistic: 1.404 on 9 and 5 DF,  p-value: 0.3703
```

(a) Write down the X matrix for this problem, and calculate $X^T X$. **[8 points.]**

(b) Fill in the entries marked by ? in the preceding table, as accurately as you are able to. **[12 points.]**

Hints: (i) This question does not require that you calculate the full matrix inverse of X^TX. (ii) For the last column of the table, use the following table of percentage points of the t_5 distribution. You are not expected to work out the last column to any greater precision than the numbers in this table.

x	0	0.13	0.27	0.41	0.56	0.73	0.92	1.16	1.48	2.02	2.57	4.03
$\Pr\{\|t_5\| > x\}$	1	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.01

(c) Based on the estimates $\hat{\beta}_0, ..., \hat{\beta}_9$, it is stated that the fitted surface has a local maximum at FR=–0.5566, AP=0.2828, PT=–0.1815. *Without doing any detailed calculations*, explain how this calculation was made. [**8 points.**]

(d) The claim is made that the true values at which the maximum expected yield is attained are FR=r, AP=s, PT=t (for given numbers $r,\ s,\ t$). *Without doing any detailed calculations*, explain how to construct an F test of this hypothesis. [**8 points.**]

(e) How would you use the result of part (d) to construct a joint 95% confidence region for the values of (FR, AP, PT) at which the maximum expected yield is attained? Once again, detailed numerical calculations are not required. [**4 points.**]

3. Much of the recent debate over climate change has focussed on the possibility of reconstructing historical temperatures using various proxies for global temperature. One of the most commonly used proxies is tree ring measurements. However, before such proxies can be reliably used, it is necessary to test their ability to reproduce observed global temperatures for the period for which direct observations are available.

To conduct such a test, annual tree ring measurements from 70 trees were compared with direct measurements of global mean temperature for the years 1850–1980. We performed a regression of global mean temperature (the y variable) on a subset of tree rings (the x variables) in order to see how well the temperature could be approximated by a linear combination of tree measurements. Two strategies were employed, with results shown in Appendices A–C.

(i) We first identified all the trees for which the correlation with the global temperature series was greater than 0.4, a total of six trees. Then, a conventional variable selection was performed with these six trees.

(ii) Prior to any regression, a principal components analysis was performed on the tree rings, and the scores of the leading principal components were calculated. Recall that in principal components analysis, it is customary to list the principal components in decreasing order of variance. A variable selection was performed on the scores of the leading 15 principal components.

(a) For strategy (i), which model would you select? Use F statistics to compare models of different order. (*Hint.* Since $1 - R^2 = SSE/SSTO$, it is possible to compute all the SSE values, modulo a constant, directly from the R^2 values.) [**10 points.**]

(b) Write a detailed report for the model containing just trees 1, 5 and 6 (the analysis in Appendix B). Using the detailed output statistics, highlight any features of the analysis that might indicate a poor fit of the model. [**20 points.**]

(c) For strategy (ii), comment on which principal components you would include in the model. Overall, do you think strategy (ii) is better than strategy (i)? [**10 points.**]

Appendix A: R-Square Values for Strategy (i)
(Leading 2 models for each of model orders 1 through 6)

Number in Model	R-Square	Variables in Model
1	0.3178	tree1
1	0.3134	tree2
2	0.4010	tree2 tree6
2	0.3957	tree1 tree6
3	0.4248	tree1 tree5 tree6
3	0.4169	tree2 tree5 tree6
4	0.4344	tree1 tree4 tree5 tree6
4	0.4311	tree1 tree2 tree5 tree6
5	0.4408	tree1 tree2 tree4 tree5 tree6
5	0.4399	tree1 tree2 tree3 tree5 tree6
6	0.4470	tree1 tree2 tree3 tree4 tree5 tree6

Appendix B: Detailed Analysis for Model Including Trees 1, 5, 6 Only

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.55269	0.51756	31.27	<.0001
Error	127	2.10233	0.01655		
Corrected Total	130	3.65501			

Root MSE	0.12866	R-Square	0.4248
Dependent Mean	-0.26125	Adj R-Sq	0.4112
Coeff Var	-49.24804		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-0.90654	0.06766	-13.40	<.0001	0
tree1	1	0.22897	0.06513	3.52	0.0006	1.65526
tree5	1	0.15728	0.06204	2.54	0.0125	1.50829
tree6	1	0.17705	0.05022	3.53	0.0006	1.41039

Output Statistics

<table>
<tr><th>Obs</th><th>-2-1 0 1 2</th><th>Cook's D</th><th>RStudent</th><th>Hat Diag H</th><th>Cov Ratio</th><th>DFFITS</th></tr>
<tr><td>1</td><td>| *| |</td><td>0.003</td><td>-0.5744</td><td>0.0296</td><td>1.0526</td><td>-0.1004</td></tr>
<tr><td>2</td><td>| |* |</td><td>0.011</td><td>0.8330</td><td>0.0616</td><td>1.0760</td><td>0.2135</td></tr>
<tr><td>3</td><td>| | |</td><td>0.001</td><td>0.3307</td><td>0.0470</td><td>1.0793</td><td>0.0735</td></tr>
<tr><td>4</td><td>| | |</td><td>0.001</td><td>-0.2783</td><td>0.0461</td><td>1.0793</td><td>-0.0612</td></tr>
<tr><td>5</td><td>| | |</td><td>0.000</td><td>0.0254</td><td>0.0363</td><td>1.0710</td><td>0.0049</td></tr>
<tr><td>6</td><td>| | |</td><td>0.000</td><td>0.1106</td><td>0.0367</td><td>1.0710</td><td>0.0216</td></tr>
<tr><td>7</td><td>| | |</td><td>0.000</td><td>0.0965</td><td>0.0342</td><td>1.0684</td><td>0.0182</td></tr>
<tr><td>8</td><td>| | |</td><td>0.000</td><td>-0.1832</td><td>0.0493</td><td>1.0845</td><td>-0.0417</td></tr>
<tr><td>9</td><td>| **| |</td><td>0.008</td><td>-1.1192</td><td>0.0264</td><td>1.0190</td><td>-0.1843</td></tr>
<tr><td>10</td><td>| | |</td><td>0.000</td><td>-0.1771</td><td>0.0173</td><td>1.0492</td><td>-0.0235</td></tr>
<tr><td>11</td><td>| *| |</td><td>0.001</td><td>-0.6845</td><td>0.0118</td><td>1.0291</td><td>-0.0748</td></tr>
<tr><td>12</td><td>| | |</td><td>0.001</td><td>-0.3625</td><td>0.0197</td><td>1.0485</td><td>-0.0514</td></tr>
<tr><td>13</td><td>| **| |</td><td>0.007</td><td>-1.4625</td><td>0.0138</td><td>0.9784</td><td>-0.1731</td></tr>
<tr><td>14</td><td>| |* |</td><td>0.012</td><td>0.9863</td><td>0.0452</td><td>1.0483</td><td>0.2147</td></tr>
<tr><td>15</td><td>| ***| |</td><td>0.018</td><td>-1.5731</td><td>0.0289</td><td>0.9833</td><td>-0.2714</td></tr>
<tr><td>16</td><td>| |*** |</td><td>0.031</td><td>1.6255</td><td>0.0451</td><td>0.9947</td><td>0.3531</td></tr>
<tr><td>17</td><td>| | |</td><td>0.000</td><td>-0.2234</td><td>0.0196</td><td>1.0511</td><td>-0.0316</td></tr>
<tr><td>18</td><td>| | |</td><td>0.000</td><td>-0.1672</td><td>0.0133</td><td>1.0452</td><td>-0.0194</td></tr>
<tr><td>19</td><td>| | |</td><td>0.000</td><td>-0.0651</td><td>0.0266</td><td>1.0602</td><td>-0.0108</td></tr>
<tr><td>20</td><td>| | |</td><td>0.001</td><td>0.3301</td><td>0.0250</td><td>1.0550</td><td>0.0529</td></tr>
<tr><td>21</td><td>| | |</td><td>0.000</td><td>-0.0398</td><td>0.0221</td><td>1.0554</td><td>-0.0060</td></tr>
<tr><td>22</td><td>| | |</td><td>0.000</td><td>-0.1837</td><td>0.0186</td><td>1.0505</td><td>-0.0253</td></tr>
<tr><td>23</td><td>| |* |</td><td>0.001</td><td>0.5885</td><td>0.0148</td><td>1.0361</td><td>0.0720</td></tr>
<tr><td>24</td><td>| | |</td><td>0.000</td><td>-0.1865</td><td>0.0221</td><td>1.0543</td><td>-0.0281</td></tr>
<tr><td>25</td><td>| **| |</td><td>0.018</td><td>-1.3817</td><td>0.0366</td><td>1.0088</td><td>-0.2692</td></tr>
<tr><td>26</td><td>| *| |</td><td>0.002</td><td>-0.6891</td><td>0.0182</td><td>1.0355</td><td>-0.0938</td></tr>
<tr><td>27</td><td>| *| |</td><td>0.001</td><td>-0.7086</td><td>0.0097</td><td>1.0257</td><td>-0.0700</td></tr>
<tr><td>28</td><td>| |*** |</td><td>0.011</td><td>1.7778</td><td>0.0136</td><td>0.9476</td><td>0.2085</td></tr>
<tr><td>29</td><td>| |***** |</td><td>0.046</td><td>3.0988</td><td>0.0201</td><td>0.7852</td><td>0.4438</td></tr>
<tr><td>30</td><td>| |* |</td><td>0.002</td><td>0.5691</td><td>0.0281</td><td>1.0511</td><td>0.0968</td></tr>
<tr><td>31</td><td>| |** |</td><td>0.009</td><td>1.1981</td><td>0.0255</td><td>1.0123</td><td>0.1939</td></tr>
<tr><td>32</td><td>| |*** |</td><td>0.039</td><td>1.7228</td><td>0.0504</td><td>0.9902</td><td>0.3968</td></tr>
<tr><td>33</td><td>| |* |</td><td>0.003</td><td>0.8047</td><td>0.0175</td><td>1.0292</td><td>0.1073</td></tr>
<tr><td>34</td><td>| |* |</td><td>0.005</td><td>0.8382</td><td>0.0279</td><td>1.0384</td><td>0.1421</td></tr>
<tr><td>35</td><td>| | |</td><td>0.001</td><td>0.3147</td><td>0.0390</td><td>1.0706</td><td>0.0634</td></tr>
<tr><td>36</td><td>| | |</td><td>0.000</td><td>0.1692</td><td>0.0348</td><td>1.0684</td><td>0.0321</td></tr>
<tr><td>37</td><td>| | |</td><td>0.000</td><td>0.1937</td><td>0.0102</td><td>1.0415</td><td>0.0197</td></tr>
<tr><td>38</td><td>| | |</td><td>0.000</td><td>-0.0926</td><td>0.0224</td><td>1.0555</td><td>-0.0140</td></tr>
<tr><td>39</td><td>| | |</td><td>0.000</td><td>0.1945</td><td>0.0173</td><td>1.0490</td><td>0.0258</td></tr>
<tr><td>40</td><td>| |** |</td><td>0.009</td><td>1.1876</td><td>0.0241</td><td>1.0116</td><td>0.1868</td></tr>
<tr><td>41</td><td>| | |</td><td>0.000</td><td>-0.1808</td><td>0.0382</td><td>1.0720</td><td>-0.0360</td></tr>
<tr><td>42</td><td>| *| |</td><td>0.001</td><td>-0.5750</td><td>0.0137</td><td>1.0355</td><td>-0.0677</td></tr>
<tr><td>43</td><td>| **| |</td><td>0.006</td><td>-1.3967</td><td>0.0132</td><td>0.9836</td><td>-0.1615</td></tr>
</table>

```
44  |    ***|       |    0.006   -1.5562   0.0096   0.9657  -0.1535
45  |      *|       |    0.003   -0.5528   0.0403   1.0651  -0.1132
46  |      *|       |    0.002   -0.7724   0.0153   1.0285  -0.0961
47  |       |*      |    0.002    0.8497   0.0135   1.0226   0.0993
48  |       |**     |    0.018    1.0781   0.0571   1.0551   0.2652
49  |     **|       |    0.007   -1.1764   0.0202   1.0083  -0.1687
50  |       |       |    0.000   -0.1792   0.0139   1.0456  -0.0213
51  |       |       |    0.000    0.2569   0.0129   1.0434   0.0293
52  |      *|       |    0.001   -0.5623   0.0124   1.0347  -0.0630
53  |     **|       |    0.006   -1.2637   0.0146   0.9959  -0.1536
54  |     **|       |    0.016   -1.3264   0.0352   1.0120  -0.2532
55  |    ***|       |    0.013   -1.5922   0.0206   0.9731  -0.2308
56  |      *|       |    0.003   -0.8567   0.0167   1.0255  -0.1116
57  |       |       |    0.000    0.0277   0.0179   1.0509   0.0037
58  |    ***|       |    0.009   -1.9389   0.0096   0.9265  -0.1906
59  |   ****|       |    0.012   -2.1463   0.0105   0.9035  -0.2210
60  |   ****|       |    0.047   -2.4874   0.0309   0.8792  -0.4443
61  |    ***|       |    0.010   -1.5725   0.0166   0.9710  -0.2041
62  |  *****|       |    0.029   -2.7304   0.0161   0.8335  -0.3495
63  |     **|       |    0.013   -1.2281   0.0335   1.0183  -0.2287
64  |       |       |    0.000   -0.0543   0.0565   1.0939  -0.0133
65  |       |       |    0.000   -0.0504   0.0208   1.0539  -0.0073
66  |       |*      |    0.007    0.9151   0.0319   1.0383   0.1662
67  |      *|       |    0.002   -0.5465   0.0204   1.0437  -0.0790
68  |     **|       |    0.014   -1.3119   0.0307   1.0085  -0.2333
69  |       |       |    0.000   -0.3093   0.0127   1.0423  -0.0351
70  |       |       |    0.000    0.0231   0.0384   1.0733   0.0046
71  |       |       |    0.001    0.2667   0.0273   1.0587   0.0447
72  |       |*      |    0.003    0.7668   0.0182   1.0318   0.1043
73  |     **|       |    0.005   -1.0046   0.0191   1.0192  -0.1403
74  |       |       |    0.000   -0.3086   0.0094   1.0388  -0.0301
75  |       |       |    0.000    0.1001   0.0222   1.0552   0.0151
76  |       |       |    0.000   -0.0851   0.0215   1.0545  -0.0126
77  |       |*      |    0.002    0.7543   0.0138   1.0279   0.0892
78  |      *|       |    0.001   -0.5723   0.0142   1.0362  -0.0687
79  |       |       |    0.000    0.1905   0.0297   1.0625   0.0333
80  |     **|       |    0.011   -1.0731   0.0383   1.0348  -0.2140
81  |       |       |    0.000    0.1765   0.0201   1.0522   0.0252
82  |       |*      |    0.003    0.7832   0.0185   1.0314   0.1076
83  |       |       |    0.000   -0.1957   0.0195   1.0514  -0.0276
84  |      *|       |    0.002   -0.8960   0.0122   1.0187  -0.0996
85  |       |*      |    0.003    0.7660   0.0191   1.0329   0.1069
86  |       |       |    0.000   -0.0476   0.0224   1.0557  -0.0072
87  |       |       |    0.002    0.4779   0.0343   1.0611   0.0901
88  |       |       |    0.000    0.1368   0.0437   1.0787   0.0292
89  |       |**     |    0.015    1.1913   0.0418   1.0299   0.2487
90  |       |**     |    0.011    1.3182   0.0241   1.0013   0.2073
91  |       |*      |    0.005    0.7396   0.0321   1.0480   0.1347
```

```
 92  |       |****  |     0.013    2.3610    0.0095    0.8764    0.2318
 93  |       |*     |     0.003    0.7224    0.0246    1.0408    0.1147
 94  |       |**    |     0.005    1.0112    0.0185    1.0181    0.1387
 95  |       |***** |     0.016    2.6537    0.0095    0.8383    0.2593
 96  |       |**    |     0.010    1.2818    0.0244    1.0045    0.2029
 97  |       |      |     0.003   -0.4254    0.0546    1.0855   -0.1022
 98  |       |      |     0.002    0.3802    0.0487    1.0800    0.0860
 99  |       |      |     0.001   -0.4008    0.0166    1.0442   -0.0521
100  |       |*     |     0.007    0.7432    0.0506    1.0683    0.1716
101  |     **|      |     0.005   -1.1555    0.0147    1.0043   -0.1413
102  |       |      |     0.001   -0.3489    0.0376    1.0683   -0.0689
103  |       |*     |     0.011    0.8827    0.0519    1.0621    0.2066
104  |       |**    |     0.013    1.1659    0.0369    1.0266    0.2282
105  |       |      |     0.000    0.0261    0.0252    1.0588    0.0042
106  |      *|      |     0.008   -0.9019    0.0391    1.0469   -0.1820
107  |     **|      |     0.006   -1.0852    0.0216    1.0164   -0.1612
108  |       |*     |     0.015    0.8953    0.0701    1.0822    0.2459
109  |       |**    |     0.020    1.3889    0.0398    1.0115    0.2827
110  |       |      |     0.000   -0.0495    0.0394    1.0744   -0.0100
111  |       |*     |     0.002    0.6601    0.0189    1.0376    0.0916
112  |       |*     |     0.019    0.9954    0.0698    1.0753    0.2726
113  |       |**    |     0.006    1.2073    0.0160    1.0018    0.1541
114  |       |*     |     0.004    0.7354    0.0292    1.0451    0.1274
115  |    ***|      |     0.018   -1.5154    0.0313    0.9912   -0.2722
116  |       |      |     0.001   -0.4296    0.0306    1.0585   -0.0763
117  |       |*     |     0.003    0.9503    0.0115    1.0147    0.1023
118  |       |      |     0.002   -0.3570    0.0450    1.0764   -0.0775
119  |      *|      |     0.023   -0.9661    0.0905    1.1018   -0.3048
120  |       |*     |     0.023    0.5678    0.2234    1.3155    0.3046
121  |       |      |     0.000    0.1070    0.0316    1.0654    0.0193
122  |       |      |     0.002   -0.4197    0.0393    1.0684   -0.0849
123  |       |***   |     0.016    1.5185    0.0275    0.9870    0.2553
124  |       |***   |     0.014    1.5993    0.0214    0.9732    0.2363
125  |      *|      |     0.003   -0.5842    0.0365    1.0597   -0.1138
126  |      *|      |     0.002   -0.5104    0.0345    1.0602   -0.0964
127  |     **|      |     0.009   -1.0133    0.0334    1.0337   -0.1884
128  |       |      |     0.006    0.4364    0.1176    1.1626    0.1593
129  |      *|      |     0.007   -0.6250    0.0627    1.0876   -0.1616
130  |       |**    |     0.010    1.3357    0.0227    0.9983    0.2036
131  |       |      |     0.002    0.2924    0.0718    1.1089    0.0813
```

Appendix C: Principal Components Regression for Tree Ring Data
(R-Squared analysis for each model order followed by
regression for modelwith all 15 principal components)

Number in Model	R-Square	Variables in Model
1	0.1417	sco4
2	0.2454	sco2 sco4
3	0.3249	sco2 sco4 sco8
4	0.3647	sco2 sco4 sco5 sco8
5	0.3928	sco2 sco4 sco5 sco8 sco15
6	0.4156	sco2 sco4 sco5 sco7 sco8 sco15
7	0.4262	sco2 sco3 sco4 sco5 sco7 sco8 sco15
8	0.4364	sco2 sco3 sco4 sco5 sco7 sco8 sco10 sco15
9	0.4418	sco1 sco2 sco3 sco4 sco5 sco7 sco8 sco10 sco15
10	0.4464	sco1 sco2 sco3 sco4 sco5 sco7 sco8 sco10 sco13 sco15
11	0.4506	sco1 sco2 sco3 sco4 sco5 sco7 sco8 sco10 sco12 sco13 sco15
12	0.4544	sco1 sco2 sco3 sco4 sco5 sco7 sco8 sco9 sco10 sco12 sco13 sco15
13	0.4559	sco1 sco2 sco3 sco4 sco5 sco7 sco8 sco9 sco10 sco11 sco12 sco13 sco15
14	0.4571	sco1 sco2 sco3 sco4 sco5 sco6 sco7 sco8 sco9 sco10 sco11 sco12 sco13 sco15
15	0.4577	sco1 sco2 sco3 sco4 sco5 sco6 sco7 sco8 sco9 sco10 sco11 sco12 sco13 sco14 sco15

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	1.67276	0.11152	6.47	<.0001
Error	115	1.98226	0.01724		
Corrected Total	130	3.65501			

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.26125	0.01147	-22.78	<.0001
sco1	1	-0.01024	0.00958	-1.07	0.2876
sco2	1	0.06228	0.01329	4.69	<.0001
sco3	1	-0.02057	0.01371	-1.50	0.1362
sco4	1	0.08645	0.01577	5.48	<.0001
sco5	1	0.04768	0.01642	2.90	0.0044
sco6	1	-0.01110	0.02170	-0.51	0.6099
sco7	1	-0.04896	0.02226	-2.20	0.0299
sco8	1	-0.09723	0.02368	-4.11	<.0001
sco9	1	0.02212	0.02464	0.90	0.3711
sco10	1	-0.03914	0.02661	-1.47	0.1441
sco11	1	-0.01582	0.02842	-0.56	0.5789
sco12	1	0.02931	0.03087	0.95	0.3443
sco13	1	0.03113	0.03150	0.99	0.3251
sco14	1	-0.01100	0.03298	-0.33	0.7394
sco15	1	0.08497	0.03482	2.44	0.0162

SOLUTIONS

1. We find X, X^TX and $(X^TX)^{-1}$ as

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}, \quad X^TX = \begin{pmatrix} 9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix},$$

$$(X^TX)^{-1} = \begin{pmatrix} \frac{1}{9} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Writing the normal equations in the form $X^TX\hat{\beta} - X^TY$, we deduce $\hat{\beta}_1 = (y_1 + ... + y_9)/9$, $2\hat{\beta}_2 + \hat{\beta}_3 = y_1 - y_3$, $\hat{\beta}_2 + 2\hat{\beta}_3 = y_2 - y_3$, $2\hat{\beta}_4 + \hat{\beta}_5 = y_4 - y_6$, $\hat{\beta}_4 + 2\hat{\beta}_5 = y_5 - y_6$, $2\hat{\beta}_6 + \hat{\beta}_7 = y_7 - y_9$, $\hat{\beta}_6 + 2\hat{\beta}_7 = y_8 - y_9$, from which the complete set of estimates are easily deduced.
$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{9}$ while the variances of $\hat{\beta}_2, ..., \hat{\beta}_9$ are each $\frac{2\sigma^2}{3}$.

We have $\mathrm{Var}(\hat{\theta}_1)=\mathrm{Var}(\hat{\beta}_1)+\mathrm{Var}(\hat{\beta}_2) = \sigma^2\left(\frac{1}{9} + \frac{2}{3}\right) = \frac{7}{9}\sigma^2$; $\mathrm{Var}(\hat{\theta}_2) = \frac{7}{9}\sigma^2$ by the same calculation; and $\mathrm{Var}(\hat{\theta}_3)=\mathrm{Var}(\hat{\beta}_1)+\mathrm{Var}(\hat{\beta}_2)+\mathrm{Var}(\hat{\beta}_3)$ $+2\ \mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2\left(\frac{1}{9} + \frac{2}{3} + \frac{2}{3} - 2 \times \frac{1}{3}\right) = \frac{7}{9}\sigma^2$. The corresponding results for $\hat{\theta}_4, ..., \hat{\theta}_9$ follow by exactly the same arguments.

2. (a) The matrices X and X^TX are respectively

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 1 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & -2 \\ 1 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 15 & 16 & 16 & 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 16 & 40 & 8 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 16 & 8 & 40 & 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 16 & 8 & 8 & 40 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 \end{pmatrix}.$$

(b) Write $M = (X^T X)^{-1}$ with entries $\{m_{i,j}\}$. We don't need to evaluate M, but we do note (i) by symmetry of cols. 2–4 of X, the values $m_{2,2}, m_{3,3}, m_{4,4}$ are all the same, (ii) each of $m_{5,5}, m_{6,6}, m_{7,7}$ is $\frac{1}{8}$ and each of $m_{8,8}, m_{9,9}, m_{10,10}$ is $\frac{1}{16}$. Therefore

$$\begin{aligned}
\hat{\beta}_5 &= \frac{y_1 - y_2 + y_3 - y_4 - y_5 + y_6 - y_7 + y_8}{8} = 2.201, \\
\hat{\beta}_9 &= \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8 - 2y_{14} + 2y_{15}}{16} = -0.9756.
\end{aligned}$$

The standard errors of $\hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$ are each $\frac{s}{\sqrt{8}} = 2.7849$ (since $s = 7.877$); the standard errors of $\hat{\beta}_7, \hat{\beta}_8, \hat{\beta}_9$ are each $\frac{s}{\sqrt{16}} = 1.9692$. Therefore the full table (as accurately as it can be filled) is

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80.4333     6.9468  11.579 8.43e-05 ***
FRFR         -1.8129     2.3667  -0.766   0.50
APAP         -3.5354     2.3667  -1.494   0.1955
PTPT         -6.3429     2.3667  -2.680   0.05
FRAP          1.1337     2.7849   0.4071  0.70
FRPT          2.201      2.7849   0.790   0.45
APPT         -0.3588     2.7849  -0.1288  0.90
FR           -1.9394     1.9692  -0.9849  0.35
AP            2.5656     1.9692   1.3029  0.25
PT           -0.9756     1.9692  -0.4954  0.65
```

(c) The stationary point of the surface

$$E\{y\} = \beta_0 + \beta_1 \mathrm{FR}^2 + \beta_2 \mathrm{AP}^2 + \beta_3 \mathrm{PT}^2 + \beta_4 \mathrm{FRAP} + \beta_5 \mathrm{FRPT} + \beta_6 \mathrm{APPT} + \beta_7 \mathrm{FR} + \beta_8 \mathrm{AP} + \beta_9 \mathrm{PT}$$

is given by

$$\begin{aligned}
2\beta_1 \mathrm{FR} + \beta_4 \mathrm{AP} + \beta_5 \mathrm{PT} + \beta_7 &= 0, \\
\beta_4 \mathrm{FR} + 2\beta_2 \mathrm{AP} + \beta_6 \mathrm{PT} + \beta_8 &= 0, \\
\beta_5 \mathrm{FR} + \beta_6 \mathrm{AP} + 2\beta_3 \mathrm{PT} + \beta_9 &= 0.
\end{aligned}$$

The sample solution $x = \begin{pmatrix} \mathrm{FR} \\ \mathrm{AP} \\ \mathrm{PT} \end{pmatrix}$ is therefore given by solving $Ax = b$, where

$$A = \begin{pmatrix} 2\hat{\beta}_1 & \hat{\beta}_4 & \hat{\beta}_5 \\ \hat{\beta}_4 & 2\hat{\beta}_2 & \hat{\beta}_6 \\ \hat{\beta}_5 & \hat{\beta}_6 & 2\hat{\beta}_3 \end{pmatrix}, \quad b = \begin{pmatrix} -\hat{\beta}_7 \\ -\hat{\beta}_8 \\ -\hat{\beta}_9 \end{pmatrix}.$$

Note 1: In R there is actually a function `solve(A,b)`, which would solve exactly this equation.

Note 2: This solution does not address the question of whether the resulting solution is in fact a maximum of the response surface. A number of students noted that $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are all < 0, which is certainly relevant to that point, but the actual condition that is required is that A be negative definite, equivalent to a statement that each of the eigenvalues of A be < 0. This statement is true but you were not expected to verify that as part of our answer.

(d) The desired null hypothesis is

$$\begin{aligned} 2\beta_1 r + \beta_4 s + \beta_5 t + \beta_7 &= 0, \\ \beta_4 r + 2\beta_2 s + \beta_6 t + \beta_8 &= 0, \\ \beta_5 r + \beta_6 s + 2\beta_3 t + \beta_9 &= 0. \end{aligned}$$

This is of the form $C\beta = h$, where $C = \begin{pmatrix} 0 & 2 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$, $h = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
The required test follows by Theorem 3.3 of the course text (note that $n = 15,\ p = 10,\ q = 3$ so the distribution of the F statistic, when H_0 is true, is $F_{3,5}$).

(e) Suppose we want a $100(1-\alpha)\%$ confidence region. For each possible (r, s, t), test the hypothesis of part (d) at significance level α. The required confidence region consists of all triples FR=r, AP=s, PT=t for which the test in (d) results in acceptance of the null hypothesis.

3. (a) The logical approach is to test the best model with k trees against the best model with $k+1$ trees, using an F test, for each of $k = 1, ..., 5$. For cases where the two models are not nested, I am arbitrarily choosing two models which are.

$k = 1$ against $k = 2$: Test $R^2 = .3134$ (tree 2) against $.4010$ (trees 2 and 6). The null hypothesis H_0 will be that the $k = 1$ model is true, H_1 will be $k = 2$. Then $SSE_0 = (1 - .3134) \times SSTO,\ SSE_1 = (1 - .4010) \times SSTO$, so $SSE_0/SSE_1 = 1.146$. The F statistic is $(SSE_0 - SSE_1)/(SSE_1/128) = 128 \times 0.146 = 18.7$ which is obviously significant (as an $F_{1,128}$ variable, the p-value is 3×10^{-5}.) Note that $n = 131$, since this is the total number of observations, which $p = 3$, the total number of parameters (including the intercept) under H_1.

$k = 2$ against $k = 3$: $SSE_0/SSE_1 = (1 - .3957)/(1 - .4248) = 1.0506$, $F = 127 \times .0506 = 6.43$, significant.

$k = 3$ against $k = 4$: $SSE_0/SSE_1 = (1 - .4248)/(1 - .4344) = 1.0170$, $F = 126 \times .0170 = 2.14$, not significant.

$k = 4$ against $k = 5$: $SSE_0/SSE_1 = (1 - .4344)/(1 - .4408) = 1.0114$, $F = 125 \times .0114 = 1.43$, not significant.

$k = 5$ against $k = 6$: $SSE_0/SSE_1 = (1 - .4408)/(1 - .4470) = 1.0112$, $F = 124 \times .0112 = 1.39$, not significant.

Therefore the best model has $k = 3$: variables tree1, tree5, tree6.

Note: Each of the F statistics may be compared with $3.84 = 1.96^2$, which would be the 5% rejection point in the limiting case that the degrees of freedom in the denominator tend to ∞.

(b) Each of the three variables is statistically significant with a p-value of .0125 or smaller. The variance inflation factors are all < 2, so it looks as though there is no problem with multicollinearity.

Leverage: with $p = 4,\ n = 131,\ 2p/n = .061$; observations 108, 112, 119, 120 (especially), 128, 129, 131 appear to have high leverage.

Outliers: Based on RStudent, observations 29, 59, 60, 62, 92, 95 have four or five stars and therefore seem to be outliers. We may also note that residuals appear clustered, with runs of positive or negative residuals, which might indicate serial correlation.

DFFITS: $2\sqrt{p/n} = .349$, exceeded in observations 16, 29, 32, 62; on the other hand, in none of these is DFFITS especially large so there does not seem to be too much of a problem with influential values.

COVRATIO: $1 \pm 3p/n$ are .908 and 1.092; the worst cases are observations 29, 62, 95, 120. Note that each of these has previously been identified as an outlier or a point of high leverage.

Overall: the presence of high leverage points and outliers casts some doubt on the appropriateness of the model (which may reflect inconsistencies in the tree ring record over time). Also, we probably should apply some correction for serial correlation.

(c) Without completely repeating the F tests of part (a), but based on the observation that an increase in R^2 of .01 is not statistically significant, whereas an increase of .02 is, we may deduce that the increases in R^2 are significant until about $k = 6$, which shows principal components 2, 4, 5, 7, 8, 15 as the most significant. These six PCs are also the most significant in the final model with $k = 15$. On the other hand, the R^2 obtained with 6 PCs in this model (.4156) is lower than the R^2 with 3 trees (.4248) in the model of part (b). Based on this, it looks as though directly selecting the most significant trees is better than doing the PC analysis.