STATISTICS 174: APPLIED STATISTICS

MAKEUP MIDTERM EXAM

HANDED OUT: NOVEMBER 4, 2004

RETURN ON: NOVEMBER 9, 2004

This is an optional, take-home exam. You are welcome to look over it before deciding whether to do it. However, if you choose to do so, you must hand it in no later than the class of Tuesday, November 9.

You are allowed to consult all course materials and to use such computational aids as you require, including Excel, SAS, R and S-PLUS, for both questions. However, Question 1 is intended as a "pocket calculator" question and even if you use a computer for some part of the calculation, your final answer should give full details of the solution. For example, if it involves solving a least squares equation, say explicitly what is $X^T X$ and how you calculate its inverse. You will not get full credit if you fail to do this. Question 2 is intended as a computer exercise and your submitted answer need not contain full details of all the numerical calculations you did, but you should write a verbal description of the steps you went through (for example, if you used a variable selection procedure, say what procedure it was and how you reached your final conclusion about which model to recommend). Detailed computer output should not be handed in, but you may (physically or electronically) cut and paste tables and figures from the computer output if they are directly relevant to the answer.

You are not allowed to consult with each other or with any other person other than myself. Below, I ask you to sign a "pledge" that you have abided with this rule. I take this seriously and I remind you that the university's Honor Code is in effect.

You are welcome to contact me (by email, telephone or visiting my office) if you need clarification about the meaning of the question or if you suspect there may be an error somewhere. As usual during exams, I will not assist you directly with answering the question, but I will try to be helpful in resolving difficulties.

Each of the two questions is worth 50 points total.

Pledge: I certify that this is my own work and I have not discussed this exam with any other person except the instructor.

Signed: Date:

1. A gold bar known to weigh exactly 10 kilograms is cut into five pieces. Subsequently, somebody wants to know the weights of the individual pieces as accurately as possible. A scale is available and up to ten weighings are allowed. Three schemes are proposed:

   (i) Weigh each object twice.

   (ii) Weigh each pair of objects once.

   (iii) Each of objects 1,2,3,4 is weighed once; then each pair of objects 1,2,3,4 is weighed once (making 4+6=10 weighings altogether). Once the weights of objects 1,2,3,4 are determined, the weight of object 5 is calculated by subtracting the sum of objects 1–4 from 10 kg.

   Assume that the error on any individual weighing has a normal distribution with mean 0 and common variance σ^2, and that the errors are independent from one weighing to the next.

   (a) For each of methods (i), (ii), (iii), show how the problem may be represented as a least squares estimation problem, calculate the least squares estimators as a function of the observations $y_1, ..., y_{10}$, and find the variances of the estimators. Which of the three methods would you recommend? (Note that all three methods, not just (iii), should use the information that the sum of all five weights is known.)

   (b) The experimenter decides to use scheme (ii), collects the data, and estimates the five weights. Then, however, it is suggested that the original information (that the sum of all five weights is 10 kg.) may have been wrong. Show how to recalculate the weights without assuming the sum of weights is known, and construct an F test of the hypothesis

   H_0: the sum of the weights is 10 kg.

   against

   H_1: the sum of the weights is not 10 kg.

   (The final answer may consist of an algebraic formula for the F statistic, together with its distribution when H_0 is true, or it could consist of a sequence of steps with verbal descriptions. I will accept either form of answer, so long as the end result is an explicit set of directions for calculating the F statistic.)

   (c) Suppose $\sigma = 0.2$ kg., and the true sum of the weights is 12 kg. State the power of the test in part (b) when the test is of size $\alpha = 0.05$, and when the test is of size $\alpha = 0.01$.

   (Your final answer should preferably include a numerical answer, e.g. if $\alpha = 0.05$ the power is 0.82. However, whether you get a numerical answer or not, the most important thing is to explain clearly how the power is calculated for this problem.)

2. The table on the next page (also available through the course web page) gives the yield of corn in a midwestern state for each of 33 years, together with several meteorological variables that are believed to affect the yield. The yields for the last three years are represented (in SAS notation) by a decimal point alone, indicating that these values are not yet observed.

   (a) Find the best regression model to predict yield as a function of year and the 8 meteorological variables. (For this exercise you need only consider linear regressions of yield on the first 9 columns; no need to consider transformations, interactions, etc. However, you should describe in detail the method of variable selection you use.)

   (b) Compute the diagnostics for outliers and influence. Is there evidence of either outliers or influential values in this data set? Explain your answer.

   (c) Is there evidence of significant multicollinearity in this data set? Explain your answer.

   (d) For the final three years, obtain (a) a 95% confidence interval for the mean yield, (b) a 95% prediction interval. Also obtain *simultaneous* confidence and prediction intervals for these three years, using either the Bonferroni or Scheffé method (explaning which, and why you used the method you chose).

   (e) Write a short (up to 10 lines) discussion. Would you say that this is an effective example of linear regression? If not, explain what you might do differently (including the possibility of collecting different data).

Year	Spring	May	June	June	July	July	Aug	Aug	Yield
	PCP	Temp	PCP	Temp	PCP	Temp	PCP	Temp	
1	17.75	60.2	5.83	69.0	1.49	77.9	2.42	74.4	34.1
2	14.76	57.5	3.83	75.0	2.72	77.2	3.30	72.6	32.8
3	27.99	62.3	5.17	72.0	3.12	75.8	7.10	72.2	43.1
4	16.76	60.5	1.64	77.8	3.45	76.1	3.01	70.5	39.9
5	11.36	69.5	3.49	77.2	3.85	79.7	2.84	73.4	23.1
6	22.71	55.0	7.00	65.9	3.35	79.4	2.42	73.6	38.3
7	17.91	66.2	2.85	70.1	0.51	83.4	3.48	79.2	20.1
8	23.31	61.8	3.80	69.0	2.63	75.9	3.99	77.8	44.5
9	18.53	59.5	4.67	69.2	4.24	76.5	3.82	75.7	46.4
10	18.56	66.4	5.32	71.4	3.15	76.2	4.72	70.7	52.1
11	12.45	58.4	3.56	71.3	4.57	76.7	6.44	70.7	52.4
12	16.05	66.0	6.20	70.0	2.24	75.1	1.94	75.1	50.9
13	27.10	59.3	5.93	69.7	4.89	74.3	3.17	72.2	60.0
14	19.05	57.5	6.16	71.6	4.56	75.4	5.07	74.0	54.6
15	20.79	64.6	5.88	71.7	3.73	72.6	5.88	71.8	52.1
16	21.88	55.1	4.70	64.1	2.96	72.1	3.43	72.5	43.4
17	20.02	56.5	6.41	69.8	2.45	73.8	3.56	68.9	56.8
18	23.17	55.6	10.39	66.3	1.72	72.8	1.49	80.6	30.4
19	19.15	59.2	3.42	68.6	4.14	75.0	2.54	73.9	60.6
20	18.28	63.5	5.51	72.4	3.47	76.2	2.34	73.0	46.0
21	18.45	59.8	5.70	68.4	4.65	69.7	2.39	67.7	48.3
22	22.00	62.2	6.11	65.2	4.45	72.1	6.21	70.5	43.0
23	19.05	59.6	5.40	74.2	3.84	74.7	4.78	70.0	62.3
24	15.67	60.0	5.31	73.2	3.28	74.6	2.33	73.2	52.8
25	15.92	55.6	6.36	72.9	1.79	77.4	7.10	72.1	54.0
26	16.75	63.6	3.07	67.2	3.29	79.8	1.79	77.2	48.3
27	12.34	62.4	2.56	74.7	4.51	72.7	4.42	73.0	52.9
28	15.82	59.0	4.84	68.9	3.54	77.9	3.76	72.9	62.0
29	15.24	62.5	3.80	66.4	7.55	70.5	2.55	73.0	66.1
30	21.72	62.8	4.11	71.5	2.29	72.3	4.92	76.3	64.1
31	25.08	59.7	4.43	67.4	2.76	72.6	5.36	73.2	.
32	17.79	57.4	3.36	69.4	5.51	72.6	3.04	72.4	.
33	26.61	66.6	3.12	69.1	6.27	71.6	4.31	72.5	.

SOLUTIONS

Mark Scheme:
25+15+10 for Question 1.
14+9+9+9+9 for Question 2.

1. (a) (i) Suppose y_1 and y_2 are weights of piece 1, y_3 and y_4 of piece 2, and so on; y_9 and y_{10} are the weights of piece 5 minus 10 kg. (so that the expected sum of $y_1, ..., y_{10}$ is 0). The expected values are therefore $\beta_1, \beta_1, \beta_2, \beta_2, ..., \beta_5, \beta_5$, where $\beta_5 = -\beta_1 + \beta_2 + \beta_3 + \beta_4$. With that substitution, the X matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}.$$

We then have

$$X^TX = \begin{pmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix}, \quad (X^TX)^{-1} = \frac{1}{10}\begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix},$$

where we used $(aI_n + bJ_n)^{-1} = \frac{1}{a}I_n - \frac{b}{a(a+nb)}J_n$ with $a = 2,\ b = 2,\ n = 4$.

Also

$$X^TY = \begin{pmatrix} y_1 + y_2 - y_9 - y_{10} \\ y_3 + y_4 - y_9 - y_{10} \\ y_5 + y_6 - y_9 - y_{10} \\ y_7 + y_8 - y_9 - y_{10} \end{pmatrix},$$

from which

$$\hat{\beta}_1 = \frac{1}{10}\{4(y_1 + y_2) - (y_3 + y_4 + ... + y_9 + y_{10})\} = \frac{y_1 + y_2}{2} - \bar{y}$$

with symmetrical formulae for $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$.

A couple of things you can do to cross-check the result (though this was not a required part of the answer): (1) If you work out $Var(\hat{\beta}_1)$

directly from the last formula, the answer is $\frac{\sigma^2}{100}(2 \times 16 + 8 \times 1) = \frac{40}{100}\sigma^2$, the same answer as obtained from $(X^TX)^{-1}$; (2) $\hat{\beta}_5 = -(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4)$ has variance $\frac{4\times 4 - 12\times 1}{10}\sigma^2 = \frac{4}{10}\sigma^2$ which is the same as the variance of $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ — this makes sense, because the ordering of the five objects is arbitrary, so the estimates should all have the same variance.

(ii) Assume the ten weighings contain pieces 1+2; 1+3; 1+4; 1+5; 2+4; 2+5; 3+4; 3+5; 4+5. To make the definitions of the β's consistent with those in (i), subtract 10kg. from each of the weights that include piece 5, i.e. observations 4, 7, 9, 10. With this notation, for example, we have $E\{y_4\} = \beta_1 + \beta_5 = -\beta_2 - \beta_3 - \beta_4$. The matrices X, X^TX, $(X^TX)^{-1}$ are respectively

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 6 & 3 & 3 & 3 \\ 3 & 6 & 3 & 3 \\ 3 & 3 & 6 & 3 \\ 3 & 3 & 3 & 6 \end{pmatrix}, \quad \frac{1}{15}\begin{pmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ -1 & -1 & -1 & 4 \end{pmatrix}.$$

We also have

$$X^TY = \begin{pmatrix} y_1 + y_2 + y_3 - y_7 - y_9 - y_{10} \\ y_1 - y_4 + y_5 + y_6 - y_9 - y_{10} \\ y_2 - y_4 + y_5 - y_7 + y_8 - y_{10} \\ y_3 - y_4 + y_6 - y_7 + y_8 - y_9 \end{pmatrix}.$$

Based on this,

$$\hat{\beta}_1 = \frac{1}{15}\{3(y_1 + y_2 + y_3 + y_4) - 2(y_5 + y_6 + y_7 + y_8 + y_9 + y_{10})\}$$

with symmetrical expressions for $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$. Each of the $\hat{\beta}_i$ (including $i = 5$, defined by subtraction) has variance $\frac{4}{15}\sigma^2$.

(iii) In this case $X,\ X^TX$ and $(X^TX)^{-1}$ are

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 \\ 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 4 \end{pmatrix}, \quad \frac{1}{21}\begin{pmatrix} 6 & -1 & -1 & -1 \\ -1 & 6 & -1 & -1 \\ -1 & -1 & 6 & -1 \\ -1 & -1 & -1 & 6 \end{pmatrix}.$$

We also have

$$X^TY = \begin{pmatrix} y_1 + y_5 + y_6 + y_7 \\ y_2 + y_5 + y_8 + y_9 \\ y_3 + y_6 + y_8 + y_{10} \\ y_4 + y_7 + y_9 + y_{10} \end{pmatrix}$$

and

$$\hat{\beta}_1 = \frac{1}{21}(6y_1 - y_2 - y_3 - y_4 + 5y_5 + 5y_6 + 5y_7 - 2y_8 - 2y_9 - 2y_{10})$$

with $\hat{\beta}_i$ for $i = 2, 3, 4$ handled symmetrically (all four estimates are invariant to permutations among the labels). In this case $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ all have variance $\frac{2}{7}\sigma^2$, but $\hat{\beta}_5 = -(\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4)$ has variance $\frac{4}{7}\sigma^2$ (it doesn't have to be the same, because in this case the five objects are not all handled the same way).

Comparing methods (i), (ii), (iii), if the objective is to minimize the variance of the $\hat{\beta}_i$'s, it seems clear we should choose method (ii).

(b) To answer this question we first need to calculate the least squares estimates of $\beta_1, ..., \beta_5$ without assuming $\sum \beta_i = 10$. This is the same as the weighing problems of Section 3.2.4, and leads to the following solution: defining $y_1, ..., y_{10}$ to be the actual measured weights (differently from the above), the $X,\ X^TX$ and $(X^TX)^{-1}$ matrices are

$$\begin{pmatrix} 1&1&0&0&0\\ 1&0&1&0&0\\ 1&0&0&1&0\\ 1&0&0&0&1\\ 0&1&1&0&0\\ 0&1&0&1&0\\ 0&1&0&0&1\\ 0&0&1&1&0\\ 0&0&1&0&1\\ 0&0&0&1&1 \end{pmatrix}, \quad \begin{pmatrix} 4&1&1&1&1\\ 1&4&1&1&1\\ 1&1&4&1&1\\ 1&1&1&4&1\\ 1&1&1&1&4 \end{pmatrix}, \quad \frac{1}{24}\begin{pmatrix} 7&-1&-1&-1&-1\\ -1&7&-1&-1&-1\\ -1&-1&7&-1&-1\\ -1&-1&-1&7&-1\\ -1&-1&-1&-1&7 \end{pmatrix},$$

and hence

$$\hat{\beta}_1 = \frac{3(y_1+y_2+y_3+y_4)-(y_5+y_6+y_7+y_8+y_9+y_{10})}{12}$$

with corresponding expressions (derived by permuting the indexes) for $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$.

Hence, a verbal description of the F test is as follows:

i. Assuming H_0 true, compute the parameter estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ as in part (a)(ii), and hence the fitted values, e.g. $\hat{y}_1 = \hat{\beta}_1 + \hat{\beta}_2$. Then set $SSE_0 = \sum(y_i - \hat{y}_i)^2$. Note that SSE_0 has 10–4=6 degrees of freedom.

ii. Calculate SSE_1 with 5 degrees of freedom in exactly the same way, but using the different definitions of $y_1, ..., y_{10}$ and the estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$ under H_1.

iii. Compute
$$F = \frac{SSE_0 - SSE_1}{1} \cdot \frac{5}{SSE_1}.$$
Under H_0, we have $F \sim F_{1,5}$. Reject H_0 at level α if $F > F_{1,5;1-\alpha}$.

(c) We use the substitution rule to calculate $\sigma^2\delta^2$. To do this, choose values of $\beta_1, ..., \beta_5$ that sum to 12, e.g. $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 2, \beta_5 = 4$. It's a mathematical property of the procedure that so long as $\sum_1^5 \beta_i$ is held fixed, the answer that follows does not depend on the individual values of $\beta_1, ..., \beta_5$ — you should verify that by trying different values of $\beta_1, ..., \beta_5$ in the calculations.

Fixing $\beta_1, ..., \beta_5$, define $y_1 = \beta_1 + \beta_2$, $y_2 = \beta_1 + \beta_3$, etc — these are the "expected values of the observations when H_1 is correct". Now calculate the F test in part (b) for these y-values. You should find $SSE_1 = 0$. This is true because under H_1 there is no error, so the least squares estimates recover exactly the true $\beta_1, ..., \beta_5$, and $\hat{y}_i = y_i$ for each i.

After redefining $y_i = y_i - 10$ for $i = 4, 7, 9, 10$ to be consistent with the notation for (a), calculate SSE_0. You should get the answer 6.4. This is $\sigma^2\delta^2$. Recall $\sigma = 0.2$. Also $\nu_1 = 1, \nu_2 = 5$. Hence compute $\phi = \frac{\delta}{\sqrt{1+\nu_1}} = \sqrt{\frac{SSE_0}{2\sigma^2}} = 8.944$. Using the `pearsonhartley` function in S-PLUS for $\alpha = 0.05, 0.01$, we find the power is 1 (to at least four decimal places) in either case.

Alternatively: Use the formula

$$\sigma^2\delta^2 = (h - h')^T \left\{C(X^TX)^{-1}C^T\right\}^{-1} (h - h').$$

Here, $h - h' = 2$, $C = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}$, so $C(X^TX)^{-1}C^T$ is the sum of all the entries in $(X^TX)^{-1}$; this is $5 \times \frac{7}{24} - 20 \times \frac{1}{24} = \frac{5}{8}$. Then $\sigma^2\delta^2 = 4 \times \frac{8}{5} = 6.4$. The rest is the same as in the previous paragraph.

Remark. If I'd worked out the numerical answer fully before setting the question I would have said $\sigma = 1$. In that case, you would find $\phi = 1.789$ and the power of the test is 0.53 when $\alpha = 0.05$, and 0.21 when $\alpha = 0.01$. The latter answer illustrates much more clearly the limitations of the F test in this kind of problem.

2. (a) We perform regression analyses using the "yield" variable `y` (col. 10 of the table) as the response, and covariates `yr,pcp1,t1,pcp2,t2, pcp3,t3,pcp4,t4` (cols. 1–9). The three leading models selected by C_p are

Number of covariates (=p-1)	Cp	RSquare	Variables
4	1.7758	0.6245	yr pcp1 pcp3 t4
3	1.9474	0.5871	yr pcp3 t4
2	1.9720	0.5522	yr t4

Forward and backward selection both select model with just `yr` and `t4` — in this case, `pcp3` and `pcp1` are not statistically significant. Choosing the best five models of each model order up to 5, we have:

Variables	p	PRESS	AIC	BIC
none	1	4121.23	148.302	149.703
yr	2	2456.51	132.675	135.477
yr, t4	3	2241.55	128.201	132.405
yr, t4, pcp3	4	2170.64	127.766	133.371
yr, t4, pcp3, pcp1	5	2168.88	126.915	133.921

(The PRESS and AIC values were obtained directly from SAS; we computed `BIC=AIC+p*1.401197` where $1.401197 = \log(n) - 2$; $n = 30$.) Thus, BIC confirms the selection of `yr, t4` as the best model, but AIC or PRESS would prefer the model with `yr, t4, pcp3, pcp1`. (However, adding more variables does not improve the AIC or PRESS scores.)

(b) For the next part, I use the model with `yr, t4` though it would also be acceptable to use the model with `yr, t4, pcp3, pcp1`. There are no serious outliers — the largest RStudent in magnitude is –2.1476 for observation 5, which is not significant. The h_i statistics show rows 18 and 7 as possibly of high leverage, and Cook's D and DFFITS rank observations 7 and 18 (in that order) as the most influential. Specific influences as assessed by DFBETAS include observations 7, 18, 21 (all influential on the intercept and t4) and observation 5 (influential on yr). Overall, observations 7 and 18 seem most influential because of high leverage, but none looks a very serious problem.

[If you use the model with `yr, t4, pcp3, pcp1`: possible outliers are in rows 22 (RStudent=–2.616) and 5 (–2.04). Rows 5 and 18 (followed by 22) are most influential judged by Cook's D or DFFITS. DFBETAS suggests influence in observation 5 (yr, pcp1), 18 (intercept, t4), 21 (intercept, t4), 22 (pcp1, t4). The only point of high leverage is row 29, though 5 and 18 are close.]

(c) VIFs for both covariates are 1.00354, clearly indicating no problem there. However the largest condition index is 61.677, which does indicate a multicollinearity problem. Study of the variance proportions attributable to the different singular vectors suggests that the problem lies with an association between the intercept and t4, which (similar to the nuclear power example discussed in the text) may indicate simply a failure to center the t4 variable.

[Model with `yr, t4, pcp3, pcp1`: largest VIF is 1.38 for pcp3, does not indicate a problem. Largest condition index is 90.01, seems again to be caused by collinearity between the intercept and t4.]

(d) The predicted values, CIs and PIs obtained by SAS are:

```
Obsn.  Predicted    CI                   PI
 31    61.2381    (55.0258,67.4503)  (43.5340,78.9422)
 32    63.1553    (56.6115,69.6992)  (45.3322,80.9785)
 33    63.8790    (57.0349,70.7230)  (45.9434,81.8145)
```

If we assume the CI and PI limits are of the form: $\hat{y} \pm 2.051831 SE$ where $\hat{y}$ is the predicted value, SE is the standard error (different for confidence intervals and for prediction intervals) and $2.051831 = t_{27;.975}$, then the confidence SEs are 3.0276, 3.1893, 3.3356 and the prediction SEs are 8.6284, 8.6865, 8.7412. For simultaneous confidence and prediction intervals, the Bonferroni bound $t_{27,1-\frac{.025}{3}} =$ 2.5525 beats the Scheffé bound $\sqrt{3F_{3,27;.95}} = 2.9801$, so we use Bonferroni and the simultaneous confidence and prediction intervals are:

Obsn.	CI	PI
31	(53.51,68.97)	(39.21,83.26)
32	(55.01,71.30)	(40.98,85.33)
33	(55.37,72.39)	(41.57,86.19)

[Model with `yr, t4, pcp3, pcp1`: The above tables become

Obsn.	Predicted	CI	PI
31	62.7608	(54.0011,71.5205)	(44.6645,80.8572)
32	65.7888	(58.7695,72.8080)	(48.4678,83.1097)
33	73.1679	(62.1925,84.1433)	(53.9012,92.4346)

for the regular CIs and PIs, and

Obsn.	CI	PI
31	(51.85,73.67)	(40.21,85.31)
32	(57.04,74.53)	(44.21,87.37)
33	(59.49,86.84)	(49.16,97.17)

if we include a Bonferroni correction.]

(e) The confidence and prediction intervals are very wide in comparison ewith the range of variability of the predicted values themselves, implying that the regression is not doing a particularly good job at predicting these variable. This is supported by the not especially high value of R^2 for the regression (0.55 or 0.62 depending on which model you used) and the relatively low influence of the meteorological variables (the linear trend in year is by far the most significant effect). It seems that trying to predict yield from meteorology is not very successful; maybe we need to find better meteorological predictors, or find other predictors altogether.