THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 6

SPECIMEN PAPER B

AND SOLUTIONS

The time for the examination is 1½ hours. The paper contains four questions, of which candidates are to attempt **three**. Each question carries 20 marks. An indicative mark scheme is shown within the questions, by giving an outline of the marks available for each part-question. The pass mark for the paper as a whole is 50%.

The solutions should not be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids. For this reason, they do not carry mark schemes. Please note that in many cases there are valid alternative methods and that, in cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of the questions and solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of the questions or solutions.

Note. In accordance with the convention used in all the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

1. In an experiment in which the subject of interest was the effect on weight gain of different amounts and types of protein, six groups of 10 male rats each were given diets that contained two levels of protein and three different types of protein. The mean weight gains (in grams) are shown in the table below.

		Protein type		
		A	*B*	*C*
Protein level	*High*	100.0	89.9	99.5
	Low	79.2	83.9	78.7

(i) Explain what is meant by *interaction*. Draw a suitable plot for the data above and comment on whether there appears to be any interaction between the type and level of protein used in the animals' diet.

(5)

(ii) Based on the animals' individual weights, the analysis of variance for this experiment is shown below. Complete the table. Explain carefully the interpretation of the F ratios. Comment on the relationship between your conclusions and the plot in (i).

```
Analysis of Variance

Source           DF       SS     MS     F
Level             *   3776.3      *     *
Type              *     82.5      *     *
Level*Type        *    730.1      *     *
Error             *        *      *
Total             *  16174.9
```

(12)

(iii) Calculate the proportion of overall variation in the data explained by the fitted model, and the estimated underlying residual variance. Comment on your answers.

(3)

2. A machine drills holes in steel plates. It is set to drill sets of three holes in a plate with the same drill bit. Samples from this process are taken at regular intervals. The table shows the diameters (mm) of holes in 12 such samples. The target mean diameter is 22.5 mm, and when the process is in control the standard deviation of the diameter (computed from a large amount of historical data) is 0.3 mm. The distribution of the random variable X underlying the observations is assumed to be Normal.

Sample	*Diameters*			*Mean*	*Range*
1	22.7	22.8	23.1	22.87	0.4
2	23.4	23.6	23.0	23.33	0.6
3	23.1	22.7	22.6	22.80	0.5
4	22.4	22.5	22.9	22.60	0.5
5	22.4	22.8	23.0	22.73	0.6
6	22.1	22.5	22.3	22.30	0.4
7	22.6	22.8	22.1	22.50	0.7
8	22.0	22.2	22.5	22.23	0.5
9	22.2	22.0	22.1	22.10	0.2
10	22.0	21.9	22.3	22.07	0.4
11	22.3	22.4	21.9	22.20	0.5
12	21.6	21.8	22.2	21.87	0.6

(i) State the distribution of X when the system is in control and hence write down the distribution of $\overline{X}$, the mean of a sample of three observations when the system is in control.

(2)

(ii) Construct a Shewhart control chart for the mean $\overline{x}$ of a set of three observations. Include suitably defined *warning limits* and *action limits*, explaining carefully your choice of these and how they should be used.

(10)

(iii) Variability may be measured by using the range. Let R_i be the range of the ith sample. Plot the values of R_i for i = 1 to 12 on a control chart. You may use the results that the target for R, based on the historical data, is 0.508, and that 95% limits for R are 0.09 and 1.46.

(4)

(iv) Comment on the behaviour of the machinery, based on the information from these two control charts.

(4)

3. Explain briefly how a randomly chosen 5 × 5 Latin Square can be constructed for use in an experiment.

(4)

Five different aptitude tests A – E are applied on five successive days to five different subjects who are considered comparable in intelligence. None of them has previously attempted tests of this type, and so it is required to remove any possible differences between days which could be attributed to a learning effect. This is done by using the layout shown in the following table, in which the scores obtained are also given.

	Day 1	*Day 2*	*Day 3*	*Day 4*	*Day 5*	*Subject totals*	*Totals for Tests*
Subject 1	E 56	B 62	A 65	D 59	C 76	318	A: 331
2	C 74	A 65	D 60	E 61	B 70	330	B: 332
3	B 63	E 59	C 80	A 66	D 64	332	C: 393
4	A 64	D 63	B 67	C 81	E 64	339	D: 310
5	D 64	C 82	E 64	B 70	A 71	351	E: 304
Day totals	321	331	336	337	345	1670	1670

Construct an analysis of variance for these data.

[Note. The uncorrected sums of squares are as follows.

The sum of the squares of all 25 observations is 112754.

That for the subjects is $318^2 + 330^2 + 332^2 + 339^2 + 351^2 = 558370$.

That for days is $321^2 + \ldots + 345^2 = 558092$.

That for tests is $331^2 + \ldots + 304^2 = 562750$.]

With the aid of appropriate statistical tests, investigate the validity of the claim that the aptitude tests measure the same qualities, and also examine whether there are differences between subjects and between days. Write a brief report on the results.

(16)

4. (i) Describe the backwards elimination procedure for selecting a multiple regression model to explain an observed variable Y when r possible predictor variables $x_1, x_2, \dots, x_r$ are available.

(5)

(ii) A set of $n = 13$ observations of (Y, x_1, x_2, x_3, x_4) is available. The observed variable Y is to be explained using as few as possible of (x_1, x_2, x_3, x_4). Some information from fitting all possible regressions is given in the table. Describe the steps in a backwards elimination procedure for this problem and state the model that results from it.

(10)

Predictors used	*df*	*Residual (error)*		*F value for fitted regression, with df*	
		Mean square	R^2 (%)		
x_1	11	115.062	53.39	12.60	1, 11
x_2	11	82.394	66.63	21.96	1, 11
x_3	11	176.309	28.58	4.40	1, 11
x_4	11	80.352	67.45	22.80	1, 11
x_1 x_2	10	5.790	97.87	229.50	2, 10
x_1 x_3	10	122.707	54.82	6.07	2, 10
x_1 x_4	10	7.476	97.25	176.63	2, 10
x_2 x_3	10	41.544	84.70	27.69	2, 10
x_2 x_4	10	86.888	68.01	10.63	2, 10
x_3 x_4	10	17.574	93.53	72.27	2, 10
x_1 x_2 x_3	9	5.346	98.23	166.35	3, 9
x_1 x_2 x_4	9	5.330	98.23	166.83	3, 9
x_1 x_3 x_4	9	5.648	98.13	157.27	3, 9
x_2 x_3 x_4	9	8.202	97.28	107.37	3, 9
x_1 x_2 x_3 x_4	8	4.983	98.53	134.25	4, 8

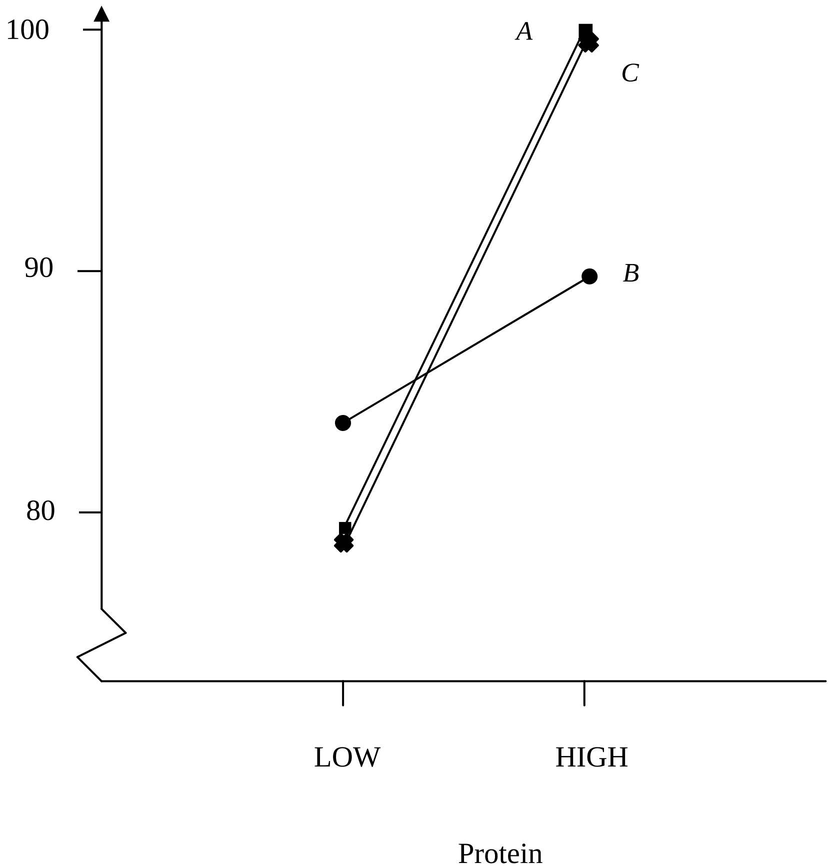
(iii) Explain how the information in the table indicates that the best single predictor to use (i.e. the best model containing just one of (x_1, x_2, x_3, x_4)) is x_4. Discuss briefly the anomaly that arises between this and the result of the backwards elimination procedure.

(5)

SOLUTIONS

Question 1

(i)



Interaction is when factors (here, the level and type of protein) do not appear to function independently. Here all the types – *A*, *B*, *C* – give an increase in mean weight gain with increasing level (although the behaviour of *B* is rather different from *A* and *C*). There is unlikely to be a large interaction – if there is any.

(ii) The analysis of variance table is as follows. Entries in *italics* are given in the question. The others need to be calculated.

SOURCE	DF	SS	MS	*F* value
Level	1	*3776.3*	3776.30	17.60 Compare $F_{1,54}$
Type	2	*82.5*	41.25	0.19 Compare $F_{2,54}$
Level * Type (Interaction)	2	*730.1*	365.05	1.70 Compare $F_{2,54}$
Error (Residual)	54	11586.0	214.56	$= \hat{\sigma}^2$
TOTAL	59	*16174.9*		

Solution continued on next page

Upper critical points of $F_{1,50}$ and $F_{2,50}$ are taken from the Society's statistical tables for use in examinations. Values for (1, 54) and (2, 54) will be very similar.

	5%	1%	0.1%
$F_{1,50}$	4.03	7.17	12.22
$F_{2,50}$	3.18	5.06	7.96

The *F* value for level is very highly significant; we have very strong evidence that the two levels of protein do not result in the same overall mean weight gain.

The *F* value for type is insignificant. We have no evidence to suggest that the three protein types are different in terms of the overall mean weight gain.

Similarly, we have no evidence that there is any interaction, i.e. that any protein type behaves differently as level is increased (even though the *B* responses are somewhat different from those of *A* and *C*). The graph in part (i) shows the pattern, and the analysis here confirms which are the significant sources of variation.

(iii) The 5 degrees of freedom for factors and interaction explain only 28.4% of the total variation (SS total). This is uncomfortably small.

We note also that the estimate of experimental error is $\hat{\sigma}^2$ = 214.56 ($\hat{\sigma}$ = 14.65), which is quite large compared with the values of the observations themselves (of order 100).

Perhaps it is simply the case that the weight gains are naturally very variable; or perhaps they are influenced by other covariates (e.g. initial weight).

The dependence of weight gain on protein level appears strong and is intuitively appealing. But there may be more to learn about the response variable.

Question 2

(i) When the system is in control, X ~ N(22.5, 0.3^2) and so $\overline{X}$ ~ N(22.5, $0.3^2/3$). i.e. N(22.5, 0.03).

(ii) Two systems of warning and action limits are in common use.

In one (sometimes referred to as the British system), warning limits are set at "Target ± 1.96 SE" and action limits at "Target ± 3.29 SE". These use the exact double-tailed 5% and 0.1% points of the Normal distribution, so these are limits that exclude, respectively, 5% and 0.1% of values of the sample mean when the system is in control. In the other system (sometimes referred to as the (current) American system), 1.96 and 3.29 are replaced by 2 and 3 respectively. [Thus "six sigma", corresponding with the value 3 here, refers to about 0.25%.]

If the system gives a sample mean outside the warning limits but within the action limits, the usual procedure is to continue sampling and, at the next sample, take no action if the mean is inside the warning limits but stop the process for inspection if it is outside the warning limits. If the system gives a sample mean outside the action limits, the process is stopped immediately.

Using 1.96 and 3.29, the limits are

warning:	$22.5 \pm 1.96\sqrt{0.03} = 22.5 \pm 0.34 = 22.16, 22.84$
action:	$22.5 \pm 3.29\sqrt{0.03} = 22.5 \pm 0.57 = 21.93, 23.07$.

The chart is shown on the next page.

(iii) The chart for the range is shown on the page after next.

(iv) The problem with the machinery is in the behaviour of the mean, not the variability which remains well in control throughout.

On the chart for the mean, the first point is (almost exactly) on the warning limit, which by itself is not enough to stop the process, but the second is well beyond the action limit and the process should have been stopped at this time and the machinery adjusted. Proceeding with the chart as it stands, a fairly steady drift downwards in the mean now occurs, until samples 9 and 10 both give means below the warning limit so that the process should have been stopped. The mean from sample 12 is below the action limit.

Overall, the mean is clearly not in control, though the variability is.

[These simple basic charts and tests are sometimes supplemented by more complicated "stopping rules" based on runs of points having a pattern – such as the steady downward drift in the mean here.]

The charts are on the next two pages

Chart for part (ii)

Control Chart for means of three values in drilling process

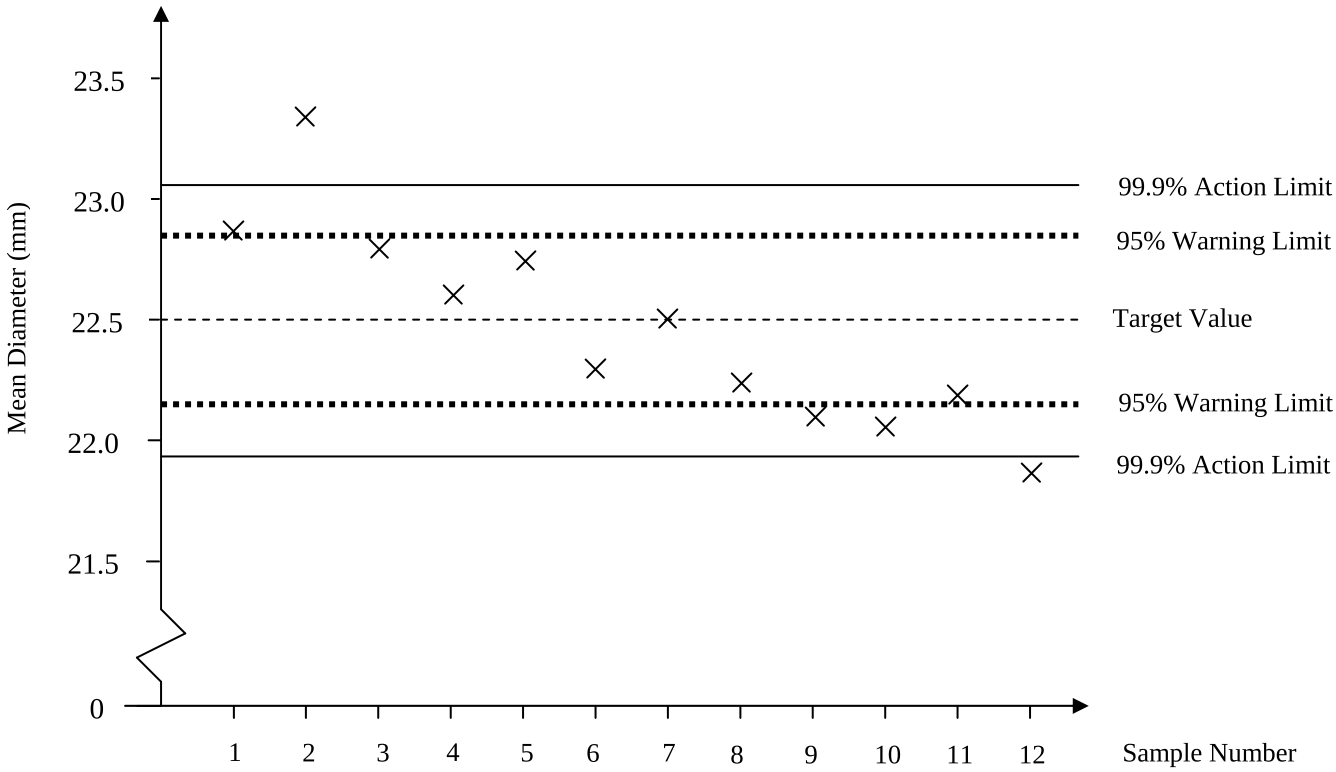
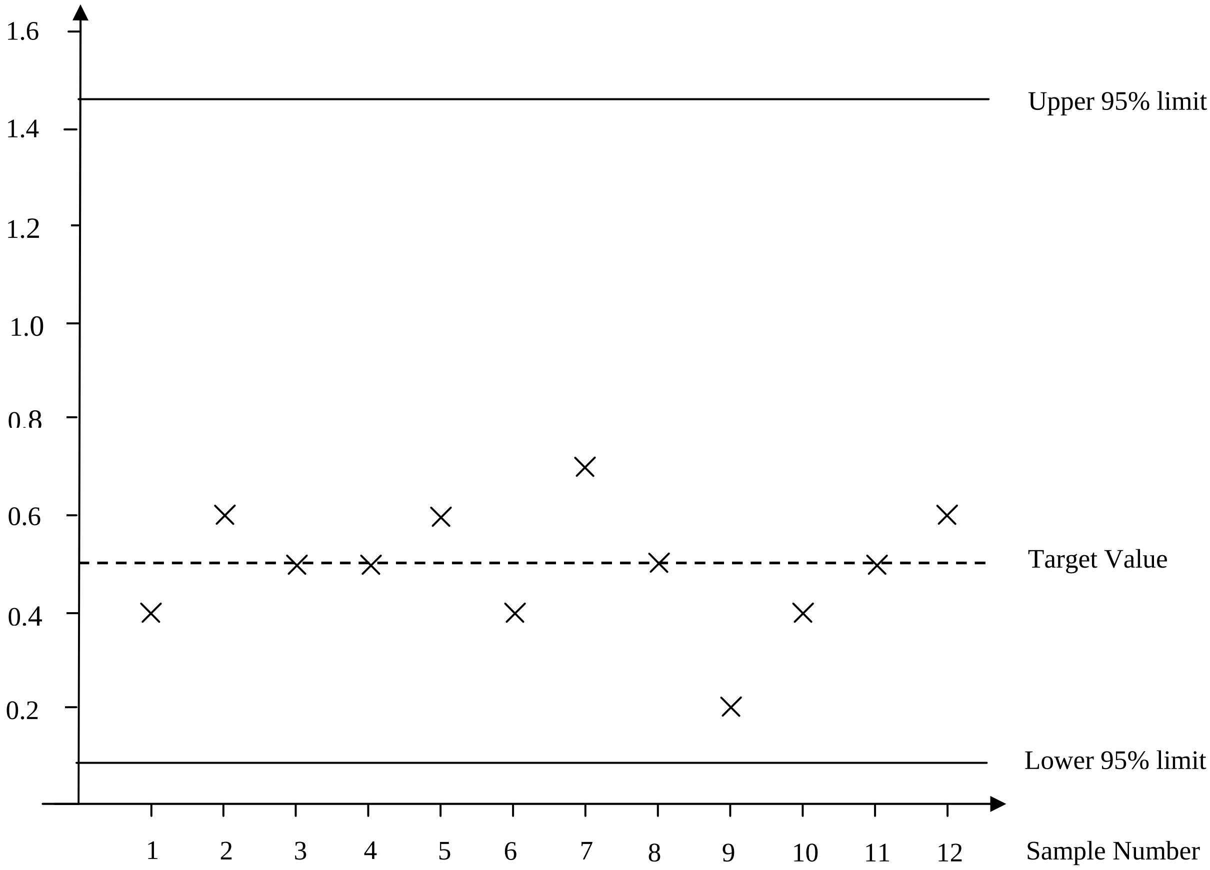


Chart for part (iii)

Control Chart for ranges of three values in drilling process



Question 3

There are tables (e.g. those by Fisher & Yates) of the "standard" 5×5 Latin squares, of which there are 56. These are the squares with the letters in alphabetical order in the first row and the first column. First, choose one of these at random. Then (using tables of random permutations, or otherwise) permute the order of the rows and the order of the columns at random, and allocate the treatments to the letters at random.

The "correction factor" is $1670^2/25 = 111556.0$. Hence the corrected sums of squares are as follows, each having 4 degrees of freedom.

Total:	112754 – 111556.0 = 1198.0
Subjects:	558370/5 – 111556.0 = 118.0
Days:	558092/5 – 111556.0 = 62.4
Tests:	562750/5 – 111556.0 = 994.0.

The residual is found by subtraction. Thus the analysis of variance table is as follows.

SOURCE	DF	SS	MS	F value
Subjects	4	118.0	29.500	15.00 Compare $F_{4,12}$
Days	4	62.4	15.600	7.93 Compare $F_{4,12}$
Tests	4	994.0	248.500	126.36 Compare $F_{4,12}$
Residual	12	23.6	1.967	$= \hat{\sigma}^2$
TOTAL	24	1198.0		

The upper critical points of $F_{4,12}$ are 3.26 at 5%, 5.41 at 1% and 9.63 at 0.1%.

There is extremely strong evidence that there are differences between the true means of the tests, suggesting that they do not measure the same qualities. The observed test means, arranged in ascending order, are

E 60.8 D 62.0 A 66.2 B 66.4 C 78.6.

Each of these is the mean of 5 observations. The least significant difference between two such means is given by

$$t\sqrt{\frac{2\times 1.967}{5}} = 0.887t$$

where 1.967 is the residual mean square from the analysis of variance and t is a critical point from the t_{12} distribution: 2.179 for 5%, 3.055 for 1% and 4.318 for 0.1%. Thus the least significant differences are 1.93 for 5%, 2.71 for 1% and 3.83 for 0.1%. So it appears that tests E and D give similar results; A and B also give similar results but higher than those of E and D; and C is considerably higher again.

There is also very strong evidence of differences between the subjects and strong evidence of differences between the days. These can also be investigated in detail using least significant difference analysis. Less formally, we can see that the means for days seem to show a fairly steady increase, suggesting that a learning effect might well exist. This could be investigated via linear regression of total against day number.

Question 4

(i) The procedure begins by carrying out a multiple regression of Y on all of x_1, x_2, …, x_r. The residual mean square from this is the best estimate of σ^2, the underlying residual variance, that can be obtained from the data.

Each possible regression of $r - 1$ predictors (i.e. leaving out just one of the x_i) is then examined. The regression that leads to the least increase in the residual mean square is chosen as the "best" at this stage, and the variable left out in that regression is dropped from the model.

The procedure continues in the same way, starting now with this "best" model with $r - 1$ predictors and leaving out each of these in turn. The resulting regressions using $r - 2$ predictors are examined, and another variable is dropped using the same rule.

The procedure again continues in the same way. It stops when it is no longer possible to leave out a variable without substantially increasing the residual mean square. Various criteria can be used for assessing when to stop.

Note that once a variable is dropped, it is never used again. This is a potential drawback of the procedure. It may be that a variable dropped at an early stage would in fact give a *better* model in combination with a smaller number of other variables than the models that are subsequently examined in the backwards elimination procedure.

(ii) In working through the backwards elimination procedure for this problem, we do not make overt use of the R^2 and F columns in the table; these columns are commonly given as part of standard computer output. We focus on the column of residual mean squares.

The full model with 4 predictors gives residual mean square 4.983. Omitting each predictor in turn, the smallest increase in residual mean square is for the model with x_3 omitted (this is *slightly* better than the models with either x_4 or x_2 omitted). So we leave out x_3 at this stage and examine the 2-predictor models which do not include x_3 and which have each other variable omitted in turn – i.e. the 2-predictor models (x_1, x_2), (x_1, x_4) and (x_2, x_4). Clearly (x_1, x_2) is the best of these, so we drop x_4 as well as x_3. We should now consider dropping each of these in turn, but the resulting models with just 1 predictor lead to very large increases in residual mean square and are not to be entertained. Thus our model is the 2-predictor model with x_1 and x_2.

(iii) The model with just x_4 has a smaller residual mean square than that for any of the other 1-predictor models; it also has the highest R^2 and the highest F value. So it is the best 1-predictor model. However, x_4 was dropped at the second stage of the backwards elimination procedure, so it will not appear in any subsequent models that are proposed by backwards elimination. [There are in fact several criteria, based largely on R^2 and the residual mean square, for choosing a "best" model.]