

THE ROYAL STATISTICAL SOCIETY

2002 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

APPLIED STATISTICS I

PAPER I

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many 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 these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 1

- (i) If Y_t = value at time t , and ε_t is a $N(0, \sigma_\varepsilon^2)$ random variable, and $\phi_0, \phi_1, \dots, \phi_p$ are coefficients (constants),

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

is the general equation for $AR(p)$.

Likewise, with $\mu, \theta_1, \theta_2, \dots, \theta_q$ a set of coefficients and the $\{\varepsilon_t\}$ uncorrelated, the general $MA(q)$ equation is

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} .$$

[Alternative expressions of these are possible.]

- (ii) (a) $Y_t = 95 + \varepsilon_t - 0.7\varepsilon_{t-1} .$

$$E[Y_t] = 95 \text{ since } E[\varepsilon_t] = 0 \text{ for all } t.$$

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(\varepsilon_t - 0.7\varepsilon_{t-1}) = \sigma_\varepsilon^2 (1 + 0.49) && (\text{since } E[\varepsilon_t \varepsilon_{t-1}] = 0) \\ &= 1.49\sigma_\varepsilon^2 . \end{aligned}$$

$$\gamma_k = \begin{cases} -0.7\sigma_\varepsilon^2, & |k|=1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$\therefore \rho_k = \frac{-0.7}{1.49} = -0.47 \text{ for } |k|=1, \text{ and } 0 \text{ elsewhere.}$$

- (b) $Y_t = 68 - 0.5Y_{t-2} + \varepsilon_t .$

$$E[Y_t] = \frac{68}{1+0.5} = 45.3 .$$

$$\text{Var}(Y_t) = \frac{\sigma_\varepsilon^2}{1-(0.5)^2} = 1.33\sigma_\varepsilon^2 .$$

$$\rho_k = -0.5\rho_{k-2} ;$$

$$\rho_1 = 0, \text{ since no term in } Y_{t-1} ;$$

$$\rho_2 = -0.5, \rho_4 = +0.25, \rho_6 = -0.125, \dots ; \quad \rho_3 = \rho_5 = \dots = 0 .$$

- (iii) $Y_t = 90 - 0.8Y_{t-1} + \varepsilon_t .$

Consider $Y_t = \mu + \alpha Y_{t-1} + \varepsilon_t = \mu + \alpha(\mu + \alpha Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$

$$= \mu(1 + \alpha) + \alpha^2 Y_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t$$

$$= \mu(1 + \alpha) + \alpha^2 (\mu + \alpha Y_{t-3} + \varepsilon_{t-2}) + \alpha \varepsilon_{t-1} + \varepsilon_t$$

$$= \mu(1 + \alpha + \alpha^2) + \alpha^3 Y_{t-3} + \alpha^2 \varepsilon_{t-2} + \alpha \varepsilon_{t-1} + \varepsilon_t$$

$$= \dots$$

$$\rightarrow \frac{\mu}{1-\alpha} + \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha^2 \varepsilon_{t-2} + \dots$$

Here $\alpha = -0.8, \mu = 90$, so $Y_t \rightarrow 50 + \varepsilon_t - 0.8\varepsilon_{t-1} + 0.64\varepsilon_{t-2} - 0.512\varepsilon_{t-3} + \dots .$

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 2

(i) The two classes may have different characteristics and, because a whole class either has feedback or does not, any difference between classes will be confounded with treatment (feedback) effect.

(ii) (a) The pre-test scores (x -values in diagram) are similar in the two groups but the post-test scores are different – see (c). Hence a one-way analysis of variance on post-test scores will estimate the treatment effects. The variances in the two groups appear to be similar, as is required for a valid analysis of variance.

(b)
$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where

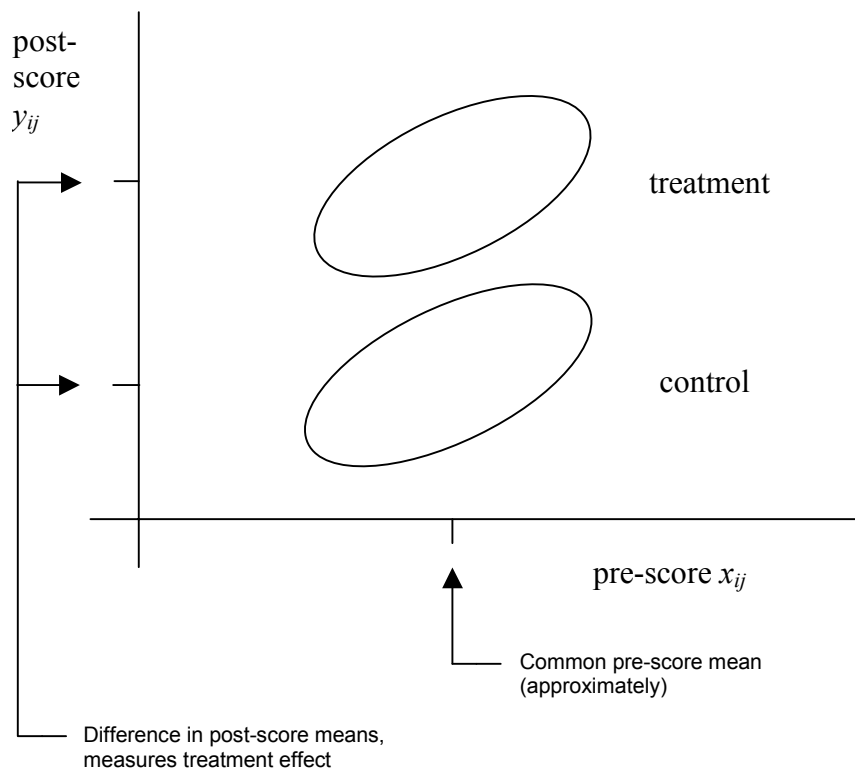
y_{ij} = post-test score for j th person in group i ($i = 1, 2$),

μ = overall mean,

τ_i is the effect due to treatment i ($i = 1, 2$),

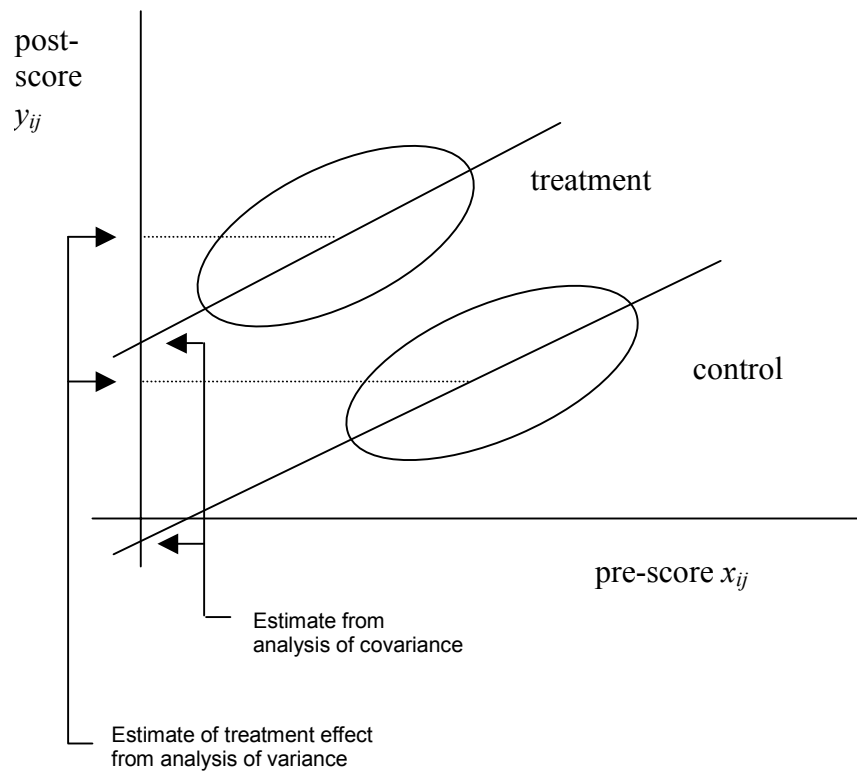
$\{\varepsilon_{ij}\}$ are independent identically distributed $N(0, \sigma^2)$ random variables representing the error (residual).

(c)



(iii) (a) $y_{ij} = \mu + \tau_i + \gamma(x_{ij} - \bar{x}) + \varepsilon_{ij}$
 where (in addition to the notation in (ii)(b)), x_{ij} is the pre-score, \bar{x} is its mean, and γ is a regression coefficient.

(b) Here, as an example, is a case showing some pre-scores for the control group higher than for the treatment group, and where analysis of variance will under-estimate the treatment effect.



(c) If $d_{ij} = y_{ij} - x_{ij}$ were used instead of y_{ij} , then we would have

$$d_{ij} [\equiv y_{ij} - x_{ij}] = \mu + \tau_i + \gamma(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

i.e. $y_{ij} = \mu + \tau_i + (1 + \gamma)(x_{ij} - \bar{x}) + \gamma\bar{x} + \varepsilon_{ij}$

$$= (\mu + \gamma\bar{x}) + \tau_i + (1 + \gamma)(x_{ij} - \bar{x}) + \varepsilon_{ij} .$$

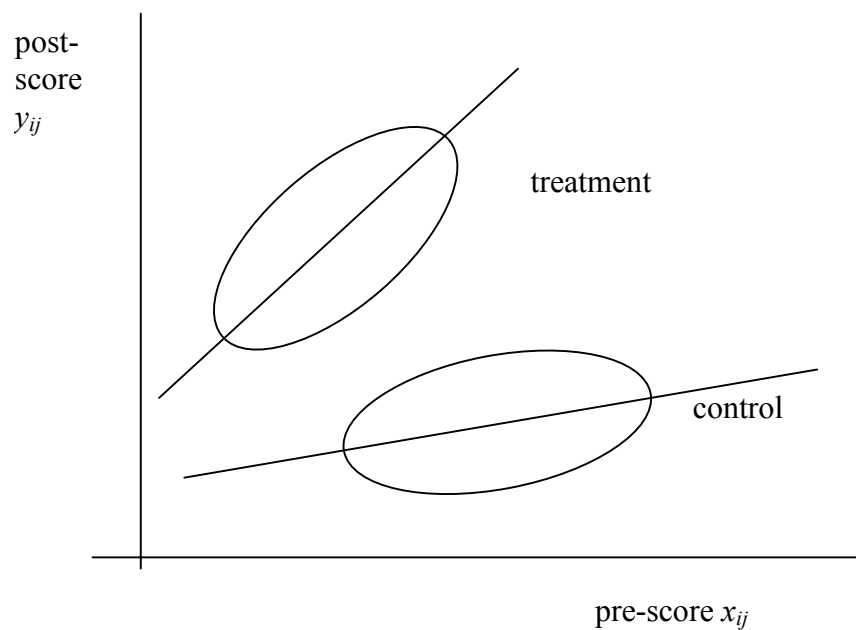
So the model is exactly the same, with two different constants, $\mu + \gamma\bar{x}$ for μ and $1 + \gamma$ for γ - i.e. equivalent to the model in (ii)(a).

(iv) (a) The regression slopes must be the same in both groups.

(b) $y_{ij} = \mu + \tau_i + \beta_i(x_{ij} - \bar{x}) + \varepsilon_{ij}, \quad i = 1, 2.$

[Note β_i for the two separate regression coefficients, not simply γ for both of them.]

(c) If $\beta_1 \neq \beta_2$, the lines are not parallel, and this allows for a possibly different relationship between y and x in the two groups.



(d) Treatment effect depends on pre-score x . The best estimate is usually to estimate the effects at \bar{x}_T, \bar{x}_C and compare these.

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 3

Case Study 1

- (i) There are two large standardised residuals. Although 2 out of 56 is not unreasonable, the largest residuals are mostly positive, the overall pattern being rather skew. There may be some lack of fit. For latitude there is a basic fan shape, suggesting that the model fits but variance is not constant as latitude changes. For longitude, the larger values are not satisfactorily modelled. It may not be a very suitable variable in itself, except perhaps as a proxy for distance from sea in the case of the USA.
- (ii) Normal probability plots may indicate other problems with residuals. Plot residuals against distance inland, height above sea level, and other geographical variables. Temperatures may be plotted against other variables (before beginning regression), although residual plots often give clearer pictures.
- (iii) Relationship with latitude makes geographical sense, although other relationships may turn out to be more important on further analysis. More than one variable can be included. A weighted regression or a (logarithmic) transformation could remove the fan-out problem. "Odd" observations can be studied by identifying the cities concerned and looking for any special reasons or demographic features which may affect the relationship.

Case Study 2

- (i) There is one outlier, but otherwise the plots are satisfactory.
- (ii) Make a Normal probability plot, though there is only a small sample of data. An initial plot of strength against specific gravity would clarify where the outlier was. Measures of influence may be worth examining.
- (iii) Check the data again for correctness. Recompute the model without the outlier – unlikely to be 'influence' but may be a 'leverage' point. So few data that is worth obtaining more if possible. Consider any other possible predictor variables.

Case Study 3

- (i) None of the plots looks random; there is (not surprisingly) evidence of autocorrelation in the residuals.
- (ii) A Durbin-Watson test would be the 'formal' way of assessing the problem. A runs test could also be useful.
- (iii) Autocorrelation must be allowed for in modelling. Then the pattern of trend etc can be studied more clearly.

[NOTE. The marking scheme reflected how much could usefully be said about each Study.]

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 4

(i) (a) If $\pi_i = P(\text{positive response for level } i \text{ of a factor})$ then the Odds Ratio for two levels j and k is

$$\frac{\frac{\pi_j}{1-\pi_j}}{\frac{\pi_k}{1-\pi_k}}, \quad \text{or} \quad \frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)}.$$

(b) The model will contain a term with parameter γ for the factor, present as γ_2 since γ_1 is set equal to 0 to avoid over-parametrization. The fitted model gives $\hat{\gamma}_2$ and its estimated variance.

$\hat{\gamma}_2$ is asymptotically Normal since it is a maximum likelihood estimator. A 95% confidence interval for the odds ratio is approximately $\exp\{\hat{\gamma}_2 \pm 1.96SE(\hat{\gamma}_2)\}$.

(ii) (a) There seems to be no difference between boys and girls, and no dependence of proportion successful on initial grading. Overall about 70% pass, and the apparently extreme observations are all based on very small numbers, and therefore unreliable.

(b)

<i>Predictor Variables in Model</i>	<i>Scaled Deviance</i>	<i>DF</i>
–	12.348	17
Sex	11.997	16
Initial grading	12.346	16
Initial grading + Sex	11.993	15

Forward selection first step gives constant plus term for sex; extra SS for sex = $12.348 - 11.997 = 0.351$, not significant as χ_1^2 ; extra SS for initial grading = 0.002, not significant.

No term other than constant required (clearly no point in adding initial grading after constant + sex).

The ratio $\frac{\text{scaled deviance}}{\text{d.f.}} = \frac{12.348}{17} = 0.726$ which is < 1 , so fit is satisfactory.

Sample sizes are quite small, so there may be some doubt whether the assumptions necessary for modelling do hold, since testing them is not possible to do precisely.

(c) The graph would be presented. The report should avoid technical jargon. It would make the point that the percentage passing is very similar for the two sexes, and likewise for the different initial gradings. It would state that the analysis showed that 68.6% of participants passed and that, statistically, there was no difference between the sexes, nor did the initial grading affect the results.

(d) One could point out that the manager is not an expert in statistics and would be reassured to know that a proper analysis has been done. In fact if the manager had worked out the percentages and found 70.1% for boys and 66.3% for girls, he/she would not have known whether this was important, without the formal analysis. There was of course statistical variability in the pattern on the graph, and the top and bottom ends of the picture were affected by having very few data. The analysis takes care of this point. The conclusion may be "obvious" to a statistician, although without a good appreciation of the basic variability it can be hard to say where the division comes between "obviously different" and "obviously the same".

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 5

(i) There is no information about the nature of the variables, their dimensions or variability. Good modelling uses the physical knowledge of what the measurements represent. Also there are too many predictor variables for too few data.

(ii) No x -variable is very highly correlated with y ; the highest correlations are x_{13} and x_{12} at 0.599 and 0.572 respectively.

However, the predictor variables are correlated among themselves. In fact they seem to fall into clusters of strongly positively correlated values:

x_1, x_2 , and x_3 ; x_8, x_9, x_{10} and x_{11} ; x_4, x_7 ; x_5, x_6 ; x_{12}, x_{13} ;
 x_{11} is also correlated with x_{12} ; x_{15} is also correlated with x_{14} and x_{16} ;
and so is x_5 with x_{14} .

(iii) Because of these strong correlations among predictors, principal components might summarise most of the variation using substantially fewer than 16 'variables' (components).

(iv) PC1 is an average of variables 1 – 4, 7, 12 – 16.
PC2 is a contrast of 5 and 6 with 8 – 11.
PC3 is a contrast of 1 – 3 with 15 – 16 (or, perhaps 1, 2, 3, 4, 7 with 12, 13, 15, 16).
PC4 is an average of 4 and 7.
PC1, PC2 explain $\frac{3}{4}$ of the total variation, PC3 and PC4 a further 17%.

We would certainly like to know what the variables are, so as to see whether it makes better sense to include or omit some of the doubtful ones – and also to check whether those that might be expected to work together actually do so.

(v) Backward elimination will not have been possible because there were too few observations. Forward selection began with x_{13} because it has the highest correlation with y . The two variables chosen have given a poor model, although the regression is significant so there is evidence that some relationship has been modelled.

The PCs seem to be of limited use, because including those which together explain 97.4% of the variation has only given $R^2 = 35.4\%$ (so no subset of the first 5 can even do as well as this). This is a very poor model.

(vi) (a) Apparently there is a perfect fit – most unusual for real data, even in highly controlled experimental conditions. This result is vastly better than anything in (v).

(b) Forward selection would never consider this combination, because it has to begin with x_{13} (see above) and this is never removed later in the process. Presumably while x_{13} is present, $x_4 - x_7$ individually are not significant. Also there is no linear combination of PCs 1 – 5 that gives the linear combination of 4 – 7 and 16 found in this model.

(c) Because of the excess of variables over observations, there is no basis for "backward elimination", with no estimate of residual variation. In principle a "best subsets" program would find the combination (4 – 7, 16), with an upper limit of 11 on the number of variables that could be included. This is a serious combinatorial problem. There may be other "perfect fits". But in view of the substantial correlations among groups of the predictor variables, including only one from each group might be a useful simplification to start with. This would be greatly helped by a knowledge of what the variables actually are.

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 6

(i) Some variables are more accurately measured than others, and some are less objective (more subjective) in nature.

Some will not be linear predictors, e.g. TIME. Some may be unstable, e.g. CHANGE. Some are correlated: there is a strong correlation of TIME with ACCTS, only one of which should be included in a model for SALES. Also CHANGE and RATING go together. AREA is highly subjective and so best omitted.

Most of the variables are sufficiently correlated with SALES to be worth including (except possibly WORK).

(ii) Backward elimination begins from a model containing all predictors, though this may be difficult to obtain if multicollinearity occurs.

Forward selection begins with the variable most highly correlated with the response (y) and adds others to it; but does not examine other combinations of variables that do not include the original one.

Best subsets examines all possible combinations of predictors but needs large-scale computing facilities, and there is the theoretical problem of the actual probabilities of Types I and II Error.

It is much better to use practical knowledge of the predictors and construct a model that makes physical sense, not relying only on significance tests for inclusion. A 'parsimonious' model is a good aim since it is easier to explain.

Fit may be checked by R^2 , adjusted R^2 , C_p etc; plots of residuals may be examined for misfit, presence of systematic variation, violation of necessary assumptions (e.g. constant variance).

(iii) (a) $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. We have

$$\mathbf{HY} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{Y}},$$

thus \mathbf{H} converts \mathbf{Y} to $\hat{\mathbf{Y}}$ and so \mathbf{H} is called the *hat matrix*.

Further results are

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y},$$

$$E[\hat{\boldsymbol{\varepsilon}}] = \mathbf{0} \quad \text{and} \quad \text{Var}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H}), \quad \text{i.e.} \quad \text{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - h_{ii}).$$

Residuals are correlated (unlike the unobservable errors); if errors are Normally distributed, so are residuals.

If the model contains a constant (intercept) term, residuals sum to 0.

We also have

$$\sum_{i=1}^n h_{ij} = \sum_{j=1}^n h_{ij} = 1, \quad \text{and} \quad \hat{y}_i = h_{ii}y_i + \sum_{j \neq i} h_{ij}y_j;$$

so as $h_{ii} \rightarrow 1$, $\hat{y}_i \rightarrow y_i$. h_{ii} is the leverage of the i th observation, indicating how heavily y_i contributes to \hat{y}_i . High leverage of points that are not outliers is not easy to explain. Points of influence are important, because a model ought not to depend too heavily on a few observations. A useful rule is that a point is "influential" if

$h_{ii} > 2(p + 1)/n$, where n = number of observations and p = number of predictor variables. In model construction, it is often sensible after fitting an initial model to remove points of influence and consider the effects on the model.

(b) All of these measures look at the influence of individual observations.

(1) A Studentised residual SR_i is defined as $\hat{\epsilon}_i / (\hat{\sigma} \sqrt{1 - h_{ii}})$, and in a Studentised deleted residual $\hat{\sigma}$ is replaced by $\hat{\sigma}_{(i)}$ from the variance estimate when observation i is not used. These can be plotted against fitted values \hat{y}_i , and in a time-series or process-control situation they may also be plotted against observation number (when i is listed in time order).

Points with high residuals are investigated for influence and lack of fit.

(2) Cook's distance $D_i = \frac{h_{ii} (SR_i)^2}{p(1 - h_{ii})}$, where p is rank (\mathbf{X}), or

$\frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})^T (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})}{p\hat{\sigma}^2}$, where $\hat{\mathbf{Y}}_{(i)}$ is the vector of fitted values when observation i has been omitted.

D_i is a measure of the influence of observation i . It may be plotted against i or against the fitted value. "High" values, usually >0.8 but it depends on the dataset, indicate influential observations.

(3) $DFFITs_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\hat{\sigma}_{(i)} \sqrt{h_{ii}}}$, or $\left(\frac{D_i p \hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} \right)^{1/2}$, is similar to Cook's distance, with $\hat{\sigma}_{(i)}^2$ instead of $\hat{\sigma}^2$. Plot it against fitted values, and again look for relatively large values.

(4) DFBETAS. This measures how an observation influences parameter values (estimates): $DFBETA_{j(i)} = \frac{(\hat{\beta}_j - \hat{\beta}_{j(i)})}{SE(\hat{\beta}_{j(i)})}$. High values are of interest (often taken as $> 2/\sqrt{n}$). Plot, for omission of each i , for each parameter.

(5) COVRATIO. This measure the impact of the i th observation on the variance-covariance matrix of estimated coefficients, i.e. the precision of parameter estimates:

$$COVRATIO_i = \frac{\det \left[\hat{\sigma}_{(i)}^2 (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \right]}{\det \left[\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} \right]}.$$

Value >1 indicates increase in precision, and <1 decrease in precision, i.e. undue influence on variation.

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 7

(i) Linear discriminant analysis can be used to produce a classification rule where the groups are known a priori, and data are described by several variables. Linear combinations of these variables x_i can show up relations not obvious from separate, univariate, analyses. Classifications so found can be applied to the new sites.

(ii) Multivariate Normal variance-covariance matrices are required to be equal for each group (but locations will be different). This is not easy to check; although formal tests exist, they are sensitive to non-Normality. Also, relatively small sample sizes do not help. Univariate Normality for each measurement can be checked in the usual ways (e.g. histograms, stem-and-leaf plots, Normal probability plots); univariate Normality is a necessary but not sufficient condition for multivariate Normality.

(iii) Variance-covariance matrices are apparently different, with changes in sign as well as size of individual entries. Normality cannot be checked on the information given.

Means of x_1 and x_4 (and possibly x_5) appear different for the two groups. Also x_4 has the smallest variance of the $\{x_i\}$.

(iv) Method 1. After constructing and applying the discriminant function, 14/17 (1) and 12/15 (2) are found to have been correctly classified. This is good, but is likely to be an overestimate of the future success rate (since the same data have been used to construct the function and to "check" it).

Cross-validation may be carried out by, for example, a jack-knife method: calculate the function omitting one observation, and use the function to predict class membership of that item; repeat this for each item in turn and observe the number of correct predictions. [In a large data-set, the discriminant function would be calculated on some of the data and then used to check the success rate of the remainder. Here we do not have enough data for that.] This gave 12/17 (1) and 9/15 (2) correct.

Method 2. Note that x_4 was identified in (iii) as a useful variate. This method correctly classifies 12/17 (1) and 12/15 (2), and the numbers on cross-validation are the same. This seems the better method.

With these sample sizes, using 5 variables (Method 1) may be over-fitting. The univariate (as it has turned out) Method (2) is more successful.

Graduate Diploma, Applied Statistics, Paper I, 2002. Question 8

- (i) (a) For all factors fixed, and the linear model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

each expected mean square will, on a null hypothesis that the corresponding $\tau_i = 0$, estimate $\sigma^2 [\equiv \text{var}(\varepsilon_{ij})]$ + a term in τ_i ; these are listed in the following table, and the set $\{\tau_i\}$ contains all the necessary main effect and

interaction terms. $\theta_a^2 = \frac{\sum \tau_a^2}{a-1}$ where $a - 1$ is the number of degrees of freedom for A , $\theta_{ab}^2 = \frac{\sum \tau_{ab}^2}{(a-1)(b-1)}$, and so on. There are 4 replicates.

SOURCE	$E(\text{MS})$	SS	DF(f)	MS	$F_{f,36}$
A	$\sigma^2 + 24\theta_a^2$	60.75	1	60.75	12.79 very highly sig
B	$\sigma^2 + 16\theta_b^2$	6.00	2	3.00	< 1
C	$\sigma^2 + 24\theta_c^2$	18.75	1	18.75	3.95 not sig.
AB	$\sigma^2 + 8\theta_{ab}^2$	0.00	2	0.00	< 1
AC	$\sigma^2 + 12\theta_{ac}^2$	0.75	1	0.75	< 1
BC	$\sigma^2 + 8\theta_{bc}^2$	24.00	2	12.00	2.53 not sig.
ABC	$\sigma^2 + 4\theta_{abc}^2$	6.00	2	3.00	< 1
Within	σ^2	171.00	36	4.75	
Total		287.25	47		

The only significant F value is that for A , at about 0.1%, so there is very strong evidence of an effect due to factor A , i.e. a difference in the means at high and low levels of A .

(b) When factor C is random, τ_c will be assumed to have variance σ_c^2 , τ_{ac} etc to have variances σ_{ac}^2 etc, as in the following table.

SOURCE	$E(MS)$	DF	MS	
A	$\sigma^2 + 12\sigma_{ac}^2 + 24\theta_a^2$	1	60.75	$F_{2,1} = \frac{60.75}{0.75} = 81$ not significant
B	$\sigma^2 + 8\sigma_{bc}^2 + 16\theta_b^2$	2	3.00	$F_{2,2} < 1$ not significant
C	$\sigma^2 + 24\sigma_c^2$	1	18.75	$F_{1,36} = 3.95$ not significant
AB	$\sigma^2 + 4\sigma_{abc}^2 + 8\theta_{ab}^2$	2	0.00	$F < 1$
AC	$\sigma^2 + 12\sigma_{ac}^2$	1	0.75	$\hat{\sigma}_{ac}^2 = 0$
BC	$\sigma^2 + 8\sigma_{bc}^2$	2	12.00	$F_{2,36} = 2.53$ not significant
ABC	$\sigma^2 + 4\sigma_{abc}^2$	2	3.00	$\hat{\sigma}_{abc}^2 = 0$
Within	σ^2	36	4.75	

A is tested against AC for the null hypothesis " $\tau_a = 0$ ", which cannot be rejected; B is tested against BC , and AB against ABC ; all others are tested against $\hat{\sigma}^2 = 4.75$. No τ_i differs significantly from 0.

(ii) A confidence interval for the difference between the means of the 2 levels of factor A in (a) would show how precisely the difference is estimated. No further action is needed in (b). In each case, some standard checking of assumptions (using descriptive statistics and other diagnostic programs) could be done.

(iii) This must happen at the planning stage, according to whether interest lies specifically in the 2 levels of factor C actually used or whether these were a random sample from a larger population of possible levels.