THE ROYAL STATISTICAL SOCIETY

2005 EXAMINATIONS – SOLUTIONS

GRADUATE DIPLOMA

APPLIED STATISTICS

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.

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

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 1

(i) Stationary time series do not show any trend or periodic variation, and do not have any systematic change in variance.

For a strictly stationary series, the observation at time t, $X(t)$, has this property: the joint distribution of $X(t_1)$, …, $X(t_k)$ is the same as that of $X(t_1 + \tau)$, …, $X(t_k + \tau)$ for all t_1, …, t_k and all τ.

For a weakly stationary series, $E[X_t] = \mu$ (a constant) and $E[X_tX_s]$ is a function of $|t - s|$ alone.

> This means that the first two moments of the joint distribution are the same for time shifts, no conditions being placed on higher moments (whereas for strict stationarity, the entire joint distribution is unchanged if the times are shifted by the same amount). In particular, the autocovariance function $\text{Cov}(X(t), X(t + \tau)) = E[(X(t) - \mu)(X(t + \tau) - \mu)]$ is a function of the lag τ alone, often denoted by $\gamma(\tau)$.

(ii) An MA(1) process is $X_t = \mu + \phi\varepsilon_{t-1} + \varepsilon_t$, in which $\{\varepsilon_t\}$ is a purely random process with mean 0 and variance σ_ε^2 and where the $\{\varepsilon_i\}$ are uncorrelated.

Hence $E[X_t] = \mu + (\phi \times 0) + 0 = \mu$ for all t.

$$\begin{aligned}\text{Var}(X_t) &= \text{Var}(\mu) + \phi^2\text{Var}(\varepsilon_{t-1}) + \text{Var}(\varepsilon_t)\\ &= 0 + \phi^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2\\ &= (1+\phi^2)\sigma_\varepsilon^2 \qquad \text{(constant for all } t\text{)}.\end{aligned}$$

$$\text{Cov}(X_t, X_{t-k}) = \text{Cov}\{(\mu + \phi\varepsilon_{t-1} + \varepsilon_t), (\mu + \phi\varepsilon_{t-k-1} + \varepsilon_{t-k})\}.$$

Here, μ is of course a constant; and the covariance of any ε_i with any <u>other</u> ε_j is zero because the $\{\varepsilon_i\}$ are uncorrelated. The only non-zero contribution arises from the covariance of ε_{t-1} with itself (i.e. $\text{Var}(\varepsilon_{t-1})$) in the case $k = 1$. Thus

$$\text{Cov}(X_t, X_{t-k}) = \begin{cases}\phi\sigma_\varepsilon^2 & \text{if } k = 1\\ 0 & \text{otherwise}\end{cases}$$

The autocorrelation function is thus $\dfrac{\phi}{1+\phi^2}$ for $k = 1$ and otherwise 0.

(iii) A partial autocorrelation coefficient measures the correlation between observations k steps apart that is not accounted for by the autocorrelations in between. For an MA(1) process, the PACF is damped cosine or exponential decay.

(iv) For MA(1), the ACF should have a spike at lag 1 and then tail off towards 0. The PACF gives little extra information but decays after lag 1 (where it is the same as the ACF).

(v) Series B seems to match these requirements.

Series C may just be white noise, as there are no significant values in ACF or PACF.

Series A has exponential decay for ACF, and the PACF tails off cut-off after lag 1. It could be AR(1) with a negative coefficient.

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 2

(a) (i) A factor is a categorical variable in which values are simply codes for each category, as in types of house. A continuous variable is an observation recorded on a scale on which any real value (within some range) is possible.

(ii) There are 2 d.f. for regression, and both "age" and "type" are used as regressor variables. Thus "type" must have been treated as a continuous variable because 2 d.f. would be needed for a factor variable with 3 levels, leaving none for age. Type of house could be regarded as a proxy for "number of detached sides", but there seems no good reason to assume a linear scale for it. Factor coding would be better.

(iii) Package B: $\begin{matrix} 1 & 58 & 0 & 0 \\ 1 & 19 & 0 & 1 \\ 1 & 10 & 1 & 0 \end{matrix}$ would be the first three rows of the design matrix.

The coefficient of "age" is the same in each package, –0.4180.

For type 1, A gives 78.8603 + 11.2249 + 0 = 90.0852
and B gives 68.212 + 21.873 + 0 = 90.085 (same)

For type 2, A gives 78.8603 – 0.5764 = 78.2839
and B gives 68.212 + 10.072 = 78.284 (same)

For type 3, A gives 78.8603 – 11.2249 + 0.5764 = 68.2118
and B gives 68.212 = 68.212 (same)

ANOVA tables will show identical values for DF, SS, MS, F and p. SEs, p-values and confidence intervals for coefficients for factor and constant will be different; those for the continuous variable "age" will be identical. (Type is in fact roughly linear as the results show.)

(b) With 62 d.f. all the given correlation coefficients are significant (at 1%). Scales A, B are strongly negatively correlated – fear of falling goes with lack of confidence doing risky tasks. The anxiety scale C is positively related to A, as would be expected, and is rather weakly opposed to B – very anxious people have less confidence in undertaking tasks.

Thus if a simple linear regression of B on C were to be calculated, the regression coefficient would be negative. But in the multiple regression of B on A and C, there is already a component of anxiety modelled by scale A, and partial correlations are required to give the complete picture of relationships.

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 3

(i) Assumptions are that $\{\varepsilon_i\}$ have mean 0 and are uncorrelated with constant variance σ^2.

(a)

$$\begin{aligned}
E\left(\hat{\boldsymbol{\beta}}\right) &= E\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}\right] \\
&= E\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left\{\mathbf{X}^T\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\right\}\right] \\
&= E\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}+\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right] \\
&= E\left[\mathbf{I}\boldsymbol{\beta}+\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\right] \\
&= E\left[\boldsymbol{\beta}\right]+\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left[\boldsymbol{\varepsilon}\right] \\
&= \boldsymbol{\beta} \quad \text{since } E\left(\boldsymbol{\varepsilon}\right)=\mathbf{0}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) &= \mathrm{Var}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}\right] = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{Var}\left(\mathbf{Y}\right)\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)^T \\
&= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)^T \quad \text{as } \mathrm{Var}(\mathbf{Y})=\sigma^2\mathbf{I} \\
&\qquad\qquad \text{and } \sigma^2 \text{ is a scalar constant} \\
&= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \quad \text{as } \mathbf{X}^T\mathbf{X} \text{ is symmetrical} \\
&\qquad\qquad \text{and therefore so also is } (\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}.
\end{aligned}$$

(b) Subject to the same assumptions as above, the Gauss-Markov theorem states that the least squares estimator $\hat{\boldsymbol{\beta}}$, as defined above, is unbiased and has minimum variance in the class of linear unbiased estimators – a "BLUE" (best linear unbiased estimator) among linear combinations of $\{Y_i\}$.

(c) With $\{\varepsilon_i\}$ Normally distributed, $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$. It follows that it is a minimum variance unbiased estimator ("MVUE"). Estimators of the parameters (regression coefficients) β_i are also Normally distributed, which allows p-values and confidence intervals to be calculated.

Solution continued on next page

(ii) (a) Transformations aim to stabilise the variance of σ^2 (make it constant) and to achieve Normality (at least approximately). Transformations can also sometimes be used in modelling non-linear relationships.

(b) Graph A has no extreme residuals but, as the fitted value increases, so does the variability; so it appears that $\{\varepsilon_i\}$ here do not have constant variance. A logarithmic transformation is useful for this "fan" shape, or perhaps a square root.

Graph B shows a non-random, curved pattern of residuals, which seems to require an extra (quadratic) term in the linear model. A transformation as such would be unlikely to be of any help.

Graph C has extreme residuals, positive and negative, and an unusual non-constant pattern of variability. More information is needed about the variables; the largest residuals should also be studied. (A transformation that might be useful is $\sin^{-1}\sqrt{y}$ where y has first been scaled so as to lie in (0, 1), but this could be difficult to understand.)

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 4

(i) There is moderate positive correlation between emit and all of the predictor variables. There seem to be two main clusters of points, which will exaggerate the correlation and the appearance of linearity. Also there are positive correlations between all the four predictor variables; not all of them may be needed in an analysis.

(ii) The model using gastemp and gasvp gives the largest adjusted R^2, the smallest s, the smallest C_p, and an R^2 that is one of the best four. These two predictors are clearly needed but there seems little advantage in using more.

(iii) Stepwise regression is an automatic method of model selection. It begins with a constant term, then enters the best single predictor, then finds the best one to add to the existing model, and goes on in this way. (Usually the F test of the "extra sum of squares" is used to decide whether there has been an improvement.) Predictors are removed from the model if inclusion of an alternative appears better. The procedure continues until it is not possible to find a predictor that improves the fit by its inclusion or that could be removed.

In situations like the graphs in (i), it may not be possible to find a good model by an automatic method, due to the relations among all the variables. The element of multicollinearity which appears in the graphs implies that there may be several possible models, all almost equally good. Predictors which are easy to measure, and those of physical importance, may be preferred, as may those cheapest and quickest to measure. These practical considerations will help choice between models that are equally good statistically.

(iv) The Normal plot suggests systematic non-Normality, perhaps needing a quadratic term in one or more predictors. The residuals appear skew (too many large negative values), and perhaps there is a suggestion of fan-shape too. This model does not seem satisfactory.

Investigate the observations having large residuals, perhaps try a log or square root transformation of the observed variable, check if there are any practical technical difficulties in making any of the measurements, note any influential observations mentioned in the analysis output, see whether there really are clusters of observations and, if so, how to allow for them.

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 5

(i) If an event has a probability π of occurring, then the *odds* of it are $\frac{\pi}{1-\pi}$.

In a logistic model, the *log odds* or *logit* function $\log\frac{\pi}{1-\pi}$ is modelled as a function of predictor variables, typically using $\log\frac{\pi}{1-\pi} = \beta_1 + \beta_2 x$. This equivalently gives $\pi = \frac{\exp(\beta_1 + \beta_2 x)}{1+\exp(\beta_1 + \beta_2 x)}$. In the context of part (ii), we have $\log\frac{\pi_i}{1-\pi_i} = \beta_1 + \beta_2 x_i$ where x_i is the dose (actually $\log_{10}$(dose)) applied to the *i*th group of insects and π_i is the probability that an insect in that group will "respond". The second set of models in part (ii) includes also a quadratic term $\beta_2 x_i^{\,2}$.

Thus, for each *i*, we have that the log odds is given by the linear predictor $\mathbf{x_i}^T\boldsymbol{\beta}$ where $\mathbf{x_i}^T$ represents the predictors and $\boldsymbol{\beta}$ represents the parameters. (As we are taking β_1 as a constant term, we take the first predictor to be identically 1.) That is, for each *i*, we have $\text{odds} = \exp\left(\mathbf{x_i}^T\boldsymbol{\beta}\right)$.

(ii) (a)

Terms in model	*df*	*Scaled deviance / df*	*df change*	*Scaled deviance change*	
Constant	7	237.20 / 7 = 33.9			
Constant, *x*	6	28.32 / 6 = 4.7	1	208.9	both sig.
Constant, *x*, x^2	5	19.32 / 5 = 3.9	1	9.0	

The quadratic model is the best of these, as the reduction in the scaled deviance is significant for each term as it is included. But the deviance is still high, so the fit is not very good.

(b) Using the parameter estimates given in the table:

$\mathbf{x}^T\boldsymbol{\beta} = 444.1290 - 532.4059x + 158.9245x^2$; so for x = 1.6905 we have 444.1290 – 900.0322 + 454.1729 = –1.7303 as given.

$\therefore$ odds = exp (–1.7303).

An approximate 95% confidence interval for the linear predictor when x = 1.6905 is –1.7303 ± 1.96 × 0.3452, from the table of results.

This interval is –1.7303 ± 0.6766, i.e. –2.4069 to –1.0537. So the interval for the odds is $e^{-2.4069}$ to $e^{-1.0537}$, or 0.090 to 0.349.

Solution continued on next page

(c) At the 50% kill, $\pi = 1 - \pi = ½$ and so $\frac{\pi}{1-\pi} = 1$ and log odds = 0.

Hence we solve

$$444.1290 - 532.4059x + 158.9245x^2 = 0$$

to give

$$x = \frac{532.4059 \pm \sqrt{(532.4059)^2 - 4 \times 158.9245 \times 444.1290}}{2 \times 158.9245}$$

$$= \frac{532.4059 \pm 33.5280}{317.8490}$$

$$= 1.7805 \text{ or } 1.5695.$$

The second of these is below the range of x in the given data, so we ignore it and take $x = 1.7805$. Thus "dose" is $10^{1.7805} = 60.33$ units.

(d) The second item seems to have an exceedingly low N_i or a very high death rate $\frac{R_i}{N_i}$, and does not fit the pattern of the remaining data at all. The progression of R_i values looks reasonable, so it seems that this N_i is likely to be wrong. Either there was a recording error or some serious fault in the experiment. It would be better to recalculate the model without this item of data.

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 6

(i) A set of measurements (x_1, x_2, …, x_n) is taken on each of k objects (people, items etc). Cluster analysis looks for natural groupings among these k items, based on the available measurements. Items within a group should be similar to each other but different from those in other groups.

(ii)(a) Cluster analysis is an exploratory process whose results depend considerably on the methods used.

Distance measure. The "distance apart" of two of the k objects may be measured by any suitable function of $\{x_{ij}: \ i = 1 \text{ to } n\}$, the j here referring to the objects ($j = 1$ to k). (Simple Euclidean distance is one possibility.) Use of different measures can give very different results.

Linkage method. This is the method used to group objects into existing clusters or to combine clusters. We will wish to group according to whether a new object is in some sense "near" to all of the items in one of the existing clusters, i.e. the distance measures are in some sense "small". Even if the same distance measure is used for the objects, different linkage methods can give different results depending on how the distance between clusters is defined.

Standardisation. If the units of measurement on one dimension are changed, there can be a substantial change in the distance measure, which may become dominated by one of the dimensions simply because of the units of measurement, eg metres instead of cm. Scale independence can be produced by standardising continuous measurements to have mean 0 and variance 1. Categorical data are less easy to adjust in this way.

(ii)(b) This distance measure can be found by coding x_{iA} as 1 if idea i is present in answer A, 0 if absent, and doing the same with B. Then we have

$$d_{AB} = \sum_{i=1}^{n} \left(x_{iA} - x_{iB} \right) \equiv \sum_{i=1}^{n} \left(x_{iA} - x_{iB} \right)^2 \text{ as all differences are } -1 \text{ or } 0 \text{ or } 1.$$

This is Euclidean distance which is widely used and is theoretically sound as a distance measure.

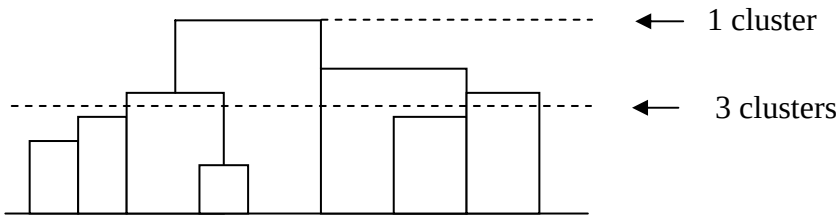
(ii)(c) Consider two objects (essays) which are the same except for one idea; but this idea is actually represented in equivalent ways in the two essays to give (1, 0, x_3, …, x_n) and (0, 1, x_3, …, x_n). Since the two x_3 data items are the same (both 0 or both 1) and so on for all up to x_n, the distance is 1 + 1 = 2.

Merging x_1 and x_2 into one single idea now gives identical sets of x_i and the distance is 0 as required.

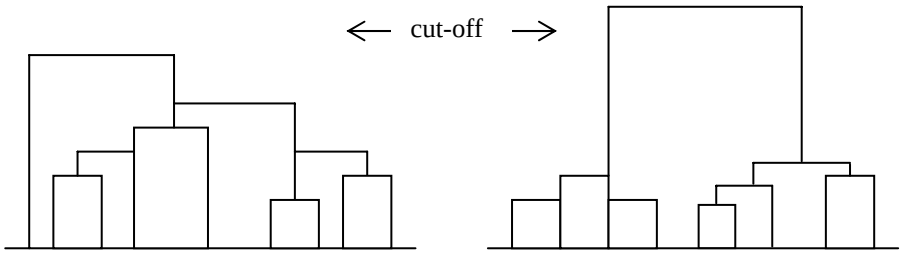
Failure to do such merges exaggerates distances between similar (even identical) objects and distorts the analysis. Re-examining the original transcripts to check the data is the best way of looking into the suggestion that the nurse's colleague has made. Features included should be independent and also important (clinically or practically).

Solution continued on next page

(ii)(d) Assuming that a dendrogram has been drawn, the effect of choosing different vertical cut-off points may be investigated.



Some knowledge about the objects being studied is helpful. At the same height, dendrogram A below suggest no clear clusters while B suggests 2. (Remember that different dendrograms may arise due to which options are chosen – see above.)



A B

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 7

(i) There are 9 observations on each of the response variables X1, X2 and X3. Column vector **X** contains these observations, with $\mathbf{X} = \begin{bmatrix} \mathbf{X1} \\ \mathbf{X2} \\ \mathbf{X3} \end{bmatrix}$ where each **Xi** is a column vector of the respective 9 observations. The model has an overall grand mean component ($\boldsymbol{\mu}$ say) for each response and the usual terms ($\boldsymbol{\tau}_i$ and $\boldsymbol{\beta}_j$ say) for treatments and blocks; these also appear in the model as column vectors accordingly. Also there is the usual column vector of residuals. We may write this as $\mathbf{X}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_{ij}$ where $\boldsymbol{\varepsilon}_{ij} \sim N_3(\mathbf{0}, \boldsymbol{\Sigma})$, the errors for different experimental units being uncorrelated.

(ii) Assuming that residuals have been calculated as part of an analysis, similar checks as in univariate analyses are possible. A necessary (but not sufficient) condition for multivariate Normality is univariate Normality in each dimension, so this should be checked. Similar variance-covariance structure is difficult to examine and (as in the univariate Bartlett test) is affected by non-Normality. Sample estimates of variances may be useful, and graphs of residuals against fitted values.

(iii) There is only one overall test, which eliminates the theoretical problem of multiple testing. Also a combined MANOVA analysis may show significant/ important effects more clearly.

(iv) (a) Consider the records on X1. The data may be set out as follows.

		Treatment			
		1	2	3	Total
Block	1	.	.	.	
	2	.	.	.	
	3	.	.	.	
	Total	T_{11}	T_{12}	T_{13}	G_1

The first subscript (1) refers to X1, the second subscript to treatment 1, 2, or 3.

The first diagonal entry in the SSCP matrix (2.149) is obtained as $\frac{T_{11}^{\ 2}}{3} + \frac{T_{12}^{\ 2}}{3} + \frac{T_{13}^{\ 2}}{3} - \frac{G_1^{\ 2}}{9}$.

A similar table for the records on X2 gives totals T_{21}, T_{22}, T_{23} and G_2; and for X3 will give $T_{31,}$ T_{32}, T_{33} and G_3. The other two diagonal entries are found in the same way.

Off-diagonal elements use pairs of totals; for example the (1, 2) or (2, 1) entry in the SSCP matrix, i.e. 1.582, comes from $\frac{T_{11}T_{21}}{3} + \frac{T_{12}T_{22}}{3} + \frac{T_{13}T_{23}}{3} - \frac{G_1G_2}{9}$ (a similar process to analysis of covariance).

Solution continued on next page

(b) Matrices have to be found for treatments, blocks and residual. Thus there is no single number to represent a "sum of squares", and different tests have been proposed for comparing SSCP(trt) with SSCP(resid). Three of the common ones are given here. The distributions of the test statistics are not the same (although transformations of them lead to approximate F distributions), so need not give the same, or similar, p-values. Tests vary in power and robustness against assumptions not being met e.g. homogeneity of covariance structure or multivariate Normality.

(v) Univariate analyses could be carried out on each of the three variables X1, X2 and X3. Also, if the effects are best described as linear combinations of X1, X2 and X3, a canonical variates analysis may be a good approach.

Graduate Diploma, Applied Statistics, Paper I, 2005. Question 8

(i) An appropriate model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

In this model:

μ is the grand mean

α_i refers to machine i (i = 1, 2, 3). This is a fixed effect since we are interested only in these three machines; thus $\sum_{i=1}^{3} \alpha_i = 0$

β_j refer to employee j (j = 1, 2, …, 6). This is a random effect since the employees were chosen at random from all those available. The $\{\beta_j\}$ are uncorrelated with mean 0 and variance σ_B^2

$\{(\alpha\beta)_{ij}\}$ are interactions. For each level of A (i = 1, 2, 3), they are uncorrelated with one another, with $\{\varepsilon_{ijk}\}$ or with $\{\beta_j\}$, and have mean 0 and variance σ_{AB}^2. For each level of B (j = 1 to 6) the $\{(\alpha\beta)_{ij}\}$ are constants and $\sum_i (\alpha\beta)_{ij} = 0$.

$\{\varepsilon_{ijk}\}$ are random variables, uncorrelated with one another, with $\{\beta_j\}$ or $\{(\alpha\beta)_{ij}\}$ and with mean 0 and variance σ^2.

(ii) $E[MS_A] = \sigma^2 + 2\sigma_{AB}^2 + 6\sum_{i=1}^{3} \alpha_i^2$

$E[MS_B] = \sigma^2 + 6\sigma_B^2$

$E[MS_{AB}] = \sigma^2 + 2\sigma_{AB}^2$

Solution continued on next page

(iii) The (corrected) total SS is given in the question: 2071.99.

$$\text{SS machines} = \frac{656.8^2}{12} + \frac{710.7^2}{12} + \frac{797.1^2}{12} - \frac{2164.6^2}{36}$$

$$= 130987.4283 - 130152.5878 = 834.84$$

$$\text{SS employees} = \frac{365.8^2}{6} + \frac{346.4^2}{6} + \frac{383.5^2}{6} + \frac{359.8^2}{6} + \frac{401.7^2}{6} + \frac{307.4^2}{6} - \frac{2164.6^2}{36}$$

$$= 131031.4233 - 130152.5878 = 878.84$$

Thus, using the information in the question,

$$\text{SS interaction} = \frac{264308.12}{2} - 130152.5878 - 834.84 - 878.84 = 287.79$$

and

residual SS = 2071.99 – (sum of all above SSs) = 70.52.

Hence the analysis of variance is

Source of variation	d.f.	Sum of squares	Mean square	*F* value
Machines (M)	2	834.84	417.42	417.42/28.78 = 14.50
Employees (E)	5	878.84	175.77	175.77/3.92 = 44.86
M × E interaction	10	287.79	28.78	28.78/3.92 = 7.35
Residual	18	70.52	3.92	
Total	35	2071.99		

To test the null hypothesis "all $\alpha_i = 0$", 14.50 is referred to the $F_{2,10}$ distribution. This is significant at the 1% level, so there is strong evidence against this null hypothesis. We may assume that there are differences among the means for machines; machine 3 is the best to buy.

There is however very strong evidence of an interaction between machines and employees; 7.35 is significant at the 0.1% level when referred to $F_{10,18}$, so we reject the null hypothesis that $\sigma^2_{AB} = 0$.

However, the table of totals shows that machine 3 is best for <u>all</u> employees, though less so for some employees than for others. So machine 3 still appears best overall. It appears that machine 2 is better than machine 1 for some employees but not for others.

There is also very strong evidence (refer 44.86 to $F_{5,18}$) that there are differences within the population of employees (the null hypothesis that $\sigma_B{}^2 = 0$ is rejected).