BU-652-M

COVARIANCE ANALYSIS OF DESIGNED EXPERIMENTS USING STATISTICAL PACKAGES

W. T. Federer and H. V. Henderson
Biometrics Unit, Cornell University

A comparative evaluation of analysis of covariance programs, for generally balanced designs with covariates, in several widely distributed statistical packages is reported. The specification, computation and output for the programs were evaluated using the criteria established by Heiberger. Several deficiencies of some of these programs are noted and suggestions are made for overcoming these.

COVARIANCE ANALYSES OF DESIGNED EXPERIMENTS USING STATISTICAL PACKAGES

Walter T. Federer and Harold V. Henderson, Cornell University

BU-652-M July 1978

1. Introduction

Experimenters and statisticians place considerable trust in statistical output from statistical computer package programs. In some cases, this trust is misplaced. One should always check to ascertain that one is receiving a correct and an appropriate statistical analysis for a set of data. If the statistical computations are incorrect and/or inappropriate, and if the results are published, the general scientific community suffers. The subject of covariance in itself appears not to be well understood by some experimenters and statisticians, and hence, one would not expect the statistical computer packages to be in any better shape. It would be better not to include covariance analyses in a package if there are errors in the program and/or if it is wrongly used a large proportion of the time.

As a result of a statistical consulting problem related to computer output, it was decided to study a number of statistical computer covariance programs. The adequacy, deficiencies, and correctness of computer program covariance analyses was investigated for a completely randomized design, a randomized complete block design, a latin square design, and a split plot design. A numerical example for each of these designs was obtained from statistical literature as follows:

i) Completely randomized design: S. R. Searle, Linear Models, pages 353-355, Tables 8.5, 8.6a, and 8.6b. The 3 treatments are less than high school education, high school education, and college education with 3, 2, and 2 observations, respectively. The dependent variate Y_{ij} is investment index and the covariate X_{ij} is number of children in a man's family.

ii) Randomized complete block design: G. W. Snedecor and W. G. Cochran, Statistical Methods, pages 427-428, Table 14.4.2. Six varieties of corn were grown in 4 blocks. The dependent variate Y_{ij} is pounds field weight of ear corn and the covariate X_{ij} is number of plants (stand) per plot.

iii) Latin square design: W. T. Federer, Experimental Design – Theory and Application, pages 490-493, Tables XVI.5 and XVI.6. Six double cross corn hybrids were grown in a 6 X 6 latin square design. The dependent variate Y_{hij} was pounds field weight of ear corn and the covariate X_{hij} was number of plants (stand) per plot.

iv) Split plot design: Rothamsted Experiment Station Reports, 1931, page 142. The 3 whole plot treatments were oat varieties Marvellous (M), Golden Rain II (G), and Victory (V), planted in 6 blocks of a randomized complete block design. Each variety whole plot was split into 4 split plots and 4 levels of nitrogen fertilizer were randomly allotted to the 4 split plots in each whole plot. The dependent variate Y_{hij} (rounded to whole pounds) is grain yield in pounds per split plot and the covariate X_{hij} (rounded to whole pounds) is straw weight in pounds per split plot.

The appropriateness of a covariance analysis for each of the above examples could be in question. A more appropriate analysis could be one in which regression coefficients vary from treatment to treatment (see Robson and Atkinson (1960)) or a bivariate analysis of variance (see Steel and Federer (1955)). This is not our concern here. We simply use these as examples to compare covariance analyses output from a number of widely distributed computer packages. To be specific, the packages investigated were:

1. BMDP – Biomedical Computer Program, version 1977.
2. GENSTAT – A General Statistical Program, version 4.01.
3. SAS – Statistical Analysis System, version 76.6.
4. SPSS – Statistical Packages for the Social Sciences, version H, release 7.2.

In section two, we present tables for the four experiment designs indicating the computations and statistics desired from covariance analyses. The statistical covariance linear models assumed for these designs are:

Completely randomized design

response model equation =

$$Y_{ij} = \mu + \tau_i + \beta_E(X_{ij}-\bar{x}_{..}) + \epsilon_{ij}, \quad i=1,2,\cdots,v;\ j=1,2,\cdots,r_i\ ;$$

$$E\bar{y}_{i.} = \mu + \tau_i + \beta_E(\bar{x}_{i.}-\bar{x}_{..})\ ;$$

ϵ_{ij} are NIID$(0,\sigma^2_\epsilon)$.

Randomized complete block design

response model equation =

$$Y_{ij} = \mu + \rho_j + \tau_i + \beta_E(X_{ij}-\bar{x}_{..}) + \epsilon_{ij}\ ; \quad i=1,\cdots,v;\ j=1,\cdots,r\ ;$$

$$E\bar{y}_{i.} = \mu + \tau_i + \beta_E(\bar{x}_{i.}-\bar{x}_{..})\ ;$$

ϵ_{ij} are NIID$(0,\sigma^2_\epsilon)$; ρ_j are IID$(0,\sigma^2_\beta)$.

Latin square design

response model equation =

$$Y_{hij} = \mu + \rho_h + \gamma_j + \tau_i + \beta_E(X_{hij}-\bar{x}_{...}) + \epsilon_{hij}\ ;$$

$$E\bar{y}_{.i.} = \mu + \tau_i + \beta_E(\bar{x}_{.i.}-\bar{x}_{...})\ ;$$

ϵ_{hij} are NIID$(0,\sigma^2_\epsilon)$; ρ_h are IID$(0,\sigma^2_\rho)$; γ_j are IID$(0,\sigma^2_\gamma)$.

Split plot design

response model equation =

$$Y_{hij} = \mu + \rho_h + \tau_i + \delta_{hi} + \beta_A(\bar{x}_{hi.}-\bar{x}_{...}) + \alpha_j + \alpha\tau_{ij} + \beta_E(X_{hij}-\bar{x}_{hi.}) + \epsilon_{hij}\ ;$$

$$E\bar{y}_{.i.} = \mu + \tau_i + \beta_A(\bar{x}_{.i.}-\bar{x}_{...})\ ; \quad h=1,\cdots,r\ ; \quad i=1,\cdots,a\ ; \quad j=1,\cdots,b\ ;$$

$$E\bar{y}_{\cdot ij} = \mu + \tau_i + \beta_A(\bar{x}_{\cdot i\cdot} - \bar{x}_{\cdots}) + \alpha_j + \alpha\tau_{ij} + \beta_X(\bar{x}_{\cdot ij} - \bar{x}_{\cdot i\cdot}) ;$$

$$E\bar{y}_{\cdot\cdot j} = \mu + \alpha_j + \beta_E(\bar{x}_{\cdot\cdot j} - \bar{x}_{\cdots}) ;$$

ϵ_{hij} are NIID$(0,\sigma^2_\epsilon)$; δ_{hi} are NIID$(0,\sigma^2_\delta)$; ρ_h are IID$(0,\sigma^2_\rho)$.

In section three, the numerical results for the four examples are presented. The Y-variable, X-variable, and adjusted Y-variable residuals were not presented in the textbooks from which the examples were taken. With the advances in data analytic procedures, we believe that residuals should be investigated as a regular feature of statistical analyses.

Attempts were made using the previously described packages to obtain all the computations obtained in section three. The success for each of the packages is described in section four. Some comments on the successes and deficiencies of the various packages are given in the last section.

The results obtained here represent an extension of papers by Heighberger (1976a, 1976b). The present paper is in the same spirit of these papers.

2. Covariance Analyses for Four Experiment Designs

A form of covariance analysis for each of the four selected experiment designs is given in Tables 2.1 to 2.4. The form of the analysis of covariance tables follows that in standard statistics textbooks (e.g., Snedecor and Cochran (1967), chapter 14, and Federer (1955), chapter XVI). In addition, the $R(\cdot/\cdot)$ notation described in Searle (1971) is used. For example, the correction for the mean equal to the total squared divided by the total number of observations, is designated as $R(\mu)$. The sum of squares for treatments corrected for the mean but ignoring all else in the response model equation is designated as $R(\tau/\mu)$, and is equal to

$R(\mu,\tau) - R(\mu)$. The sum of squares due to the mean, the treatments, and a linear regression coefficien is $R(\mu,\tau,\beta)$. The total sum of squares for any design is designated as $\underline{y}'\underline{y}$ where $\underline{y}$ is the column vector of all the observations in the experiment. The remaining computations are as described in the above reference.

Additional computations, e.g., the treatment regression coefficient $b_T = T_{xy}/T_{xx}$, are often desired. Also, it may be of interest to compare the treatment and error regressions. Federer (1955), page 493, gives one such test, but the error variances for treatment and error regressions may differ. For this case, the reader is referred to Smith (1958).

It should be noted that one form of covariance analysis for a split plot design was used here (see Federer (1955)). Another form has been described by Truitt and Smith (1956). They consider the situation wherein the whole plot and split plot regressions estimate the same parameters, β_E, and the terms in the split plot design response model equation combine into the single term $\beta_E(X_{hij}-\bar{x}_{...})$. (It was observed that in 6 of the 9 examples they considered, these regressions were significantly different.) They further show how to obtain the maximum likelihood estimate of β_E and how to make tests of significance. If only error (b) sums of products were used to estimate β_E and to adjust all other sums of squares including error (a), then the mean squares for error (a) and for main plots are not independent and the F-test is not valid. (Also, see Anderson (1946) and Bartlett (1937).)

Table 2.1. Covariance analysis for a completely randomized design.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2	F
Total	$r_{.}-1$	$S_{yy}=\underline{y}'\underline{y}-R(\mu)$	S_{xy}	S_{xx}	
Treatment	$v-1$	$T_{yy}=R(\tau/\mu)$	T_{xy}	T_{xx}	$T_{xx}(r_{.}-v)/(v-1)E_{xx}$
Error	$r_{.}-v$	$E_{yy}=\underline{y}'\underline{y}-R(\mu,\tau)$	E_{xy}	E_{xx}	

		Adjusted sum of squares	Mean square	F
Error adj.	$r_{.}-v-1$	$E'_{yy}=E_{yy}-E^2_{xy}/E_{xx}=\underline{y}'\underline{y}-R(\mu,\tau,\beta)$	$E'_{yy}/(r_{.}-v-1)=E^*_{yy}$	-
Error regression	1	$E^2_{xy}/E_{xx}=R(\beta/\mu,\tau)$	E^2_{xy}/E_{xx}	$E^2_{xy}/E^*_{yy}E_{xx}$
Treatment + error	$r_{.}-2$	$S'_{yy}=S^2_{yy}-S^2_{xy}/S_{xx}$	-	-
Treatment adj	$v-1$	$T'_{yy}=S'_{yy}-E'_{yy}=R(\tau/\mu,\beta)$	$T'_{yy}/(v-1)=T^*_{yy}$	T^*_{yy}/E^*_{yy}

Treatment means

unadjusted		adjusted y
$\bar{y}_{1.}$	$\bar{x}_{1.}$	$\bar{y}_{1.}-b_E(\bar{x}_{1.}-\bar{x}_{..})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{v.}$	$\bar{x}_{v.}$	$\bar{y}_{v.}-b_E(\bar{x}_{v.}-\bar{x}_{..})$
$\bar{y}_{..}$	$\bar{x}_{..}$	-

where $b_E = E_{xy}/E_{xx}$, $\bar{y}_{i.}$ and $\bar{x}_{i.}$ are treatment i means from r_i observations, and $\bar{y}_{..}$ and $\bar{x}_{..}$ are overall arithmetic means for the variates Y_{ij} and X_{ij}, respectively.

Standard error of a difference between 2 adjusted treatment means, i and i'

$$\sqrt{E^*_{yy}\left\{\frac{1}{r_i}+\frac{1}{r_{i'}}+\frac{(\bar{x}_{i.}-\bar{x}_{i'.})^2}{E_{xx}}\right\}}$$

Table 2.1. (Cont'd)

Average standard error of a difference between 2 adjusted treatment means

$r_i = r$: $\sqrt{\frac{2E^*_{yy}}{r}\left\{1 + \frac{T_{xx}/(v-1)}{E_{xx}}\right\}} = A^*$

$r_i \neq r$: $\sqrt{\text{average of } v(v-1)/2 \text{ variances of a difference between 2 adjusted means}} = A^*$

Efficiency of covariance

$2E_{yy}/rv(r-1)A^*$, $r_i = r$

Aver. unadjusted standard error of a difference$/A^*$, $r_i \neq r$.

Residuals

Residuals for Y_{ij}: $\hat{e}_{ijy} = Y_{ij} - \bar{y}_{i.}$

Residuals for X_{ij}: $\hat{e}_{ijx} = X_{ij} - \bar{x}_{i.}$

Residuals for adjusted Y_{ij}: $e'_{ij} = \hat{e}_{ijy} - b_E\hat{e}_{ijx} = Y_{ij} - \left(\hat{Y}_{ij} = \hat{\mu} + \hat{\tau}_i + b_E(X_{ij}-\bar{x}_{..})\right)$

Solutions for fixed effects, using usual constraints

$\hat{\mu} = \bar{y}_{...}$

$\widehat{\mu + \tau_i} = \bar{y}_{i.} - b_E(\bar{x}_{i.}-\bar{x}_{..})$ = adjusted i^{th} treatment mean

$\hat{\tau}_i = \bar{y}_{i.}-\bar{y}_{..} - b_E(\bar{x}_{i.}-\bar{x}_{..})$

Table 2.2. Covariance analysis for a randomized complete block design.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2	F
Total	rv-1	$S_{yy} = \underline{y}'\underline{y}-R(\mu)$	S_{xy}	S_{xy}	
Block	r-1	$B_{yy}=R(\rho/\mu)$	B_{xy}	B_{xx}	
Treatment	v-1	$T_{yy}=R(\tau/\mu)$	T_{xy}	T_{xx}	$(r-1)T_{xx}/E_{xx}$
Error	(r-1)(v-1)	$E_{yy}=\underline{y}'\underline{y}-R(\mu,\rho,\tau)$	E_{xy}	E_{xx}	
		Adjusted sum of squares		Mean square	F
Error adj.	(r-1)(v-1)-1	$E'_{yy}=E_{yy}-E^2_{xy}/E_{xy}=\underline{y}'\underline{y}-R(\mu,\tau,\rho,\beta)$		E^*_{yy}	-
Error regression	1	$E^2_{xy}/E_{xx}=R(\beta/\mu,\tau,\rho)$		E^2_{xy}/E_{xx}	$E^2_{xy}/E_{xx}E^*_{yy}$
Treatment + error	r(v-1)-1	$(T_{yy}+E_{yy})'=T_{yy}+E_{yy}-(T_{xy}+E_{xy})^2/(T_{xx}+E_{xx})$		-	-
Treatment adj	v-1	$T'_{yy}=(T_{yy}+E_{yy})'-E'_{yy}=R(\tau/\mu,\rho,\beta)$		T^*_{yy}	T^*_{yy}/E^*_{yy}

Treatment means unadjusted		adjusted y
$\bar{y}_{1.}$	$\bar{x}_{1.}$	$\bar{y}_{1.}-b_E(\bar{x}_{1.}-\bar{x}_{..})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{v.}$	$\bar{x}_{v.}$	$\bar{y}_{v.}-b_E(\bar{x}_{v.}-\bar{x}_{..})$
$\bar{y}_{..}$	$\bar{x}_{..}$	-

where $b_E = E_{xy}/E_{xx}$, $\bar{y}_{i.}$ and $\bar{x}_{i.}$ are treatment i means from r observations, and $\bar{y}_{..}$ and $\bar{x}_{..}$ are overall means for the variates Y_{ij} and X_{ij}, respectively.

Standard error of a difference between 2 adjusted treatment means, i and i'

$$\sqrt{E^*_{yy}\left\{\frac{2}{r}+\frac{(\bar{x}_{i.}-\bar{x}_{i'.})^2}{E_{xx}}\right\}}$$

Table 2.2. (Cont'd)

Average standard error of a difference between 2 adjusted treatment means

$$\sqrt{\frac{2E^{*}_{yy}}{r}\left\{1 + T_{xx}/(v-1)E_{xx}\right\}} = A^{*}$$

Efficiency of covariance

$$2E_{yy}/r(r-1)(v-1)A^{*}$$

Residuals

Residuals for Y_{ij}: $\hat{e}_{ijy} = Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$

Residuals for X_{ij}: $\hat{e}_{ijx} = X_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$

Residuals for adjusted Y_{ij}: $e'_{ij} = \hat{e}_{ijy} - b_E\hat{e}_{ijx}$

Solutions for fixed effects

$$\hat{\mu} = \bar{y}_{...}$$

$$\widehat{\mu + \tau_i} = \bar{y}_{i.} - b_E(\bar{x}_{i.} - \bar{x}_{..})$$

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - b_E(\bar{x}_{i.} - \bar{x}_{..})$$

Table 2.3. Covariance analysis for a latin square design.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2	F
Total	v^2-1	$S_{yy}=\underline{y}'\underline{y}-R(\mu)$	S_{xy}	S_{xx}	
Row	v-1	$R_{yy}=R(\rho/\mu)$	R_{xy}	R_{xx}	
Column	v-1	$C_{yy}=R(\gamma/\mu)$	C_{xy}	C_{xx}	
Treatment	v-1	$T_{yy}=R(\tau/\mu)$	T_{xy}	T_{xx}	$(v-2)T_{xx}/E_{xx}$
Error	(v-1)(v-2)	E_{yy}	E_{xy}	E_{xx}	
		Adjusted sum of squares		Mean square	F
Error adj.	(v-1)(v-2)-1	$E'_{yy}=E_{yy}-E^2_{xy}/E_{xx}=\underline{y}'\underline{y}-R(\mu,\rho,\gamma,\tau,\beta)$		E^*_{yy}	-
Error regression	1	$E^2_{xy}/E_{xx}=R(\beta/\mu,\rho,\gamma,\tau)$		E^2_{xy}/E_{xx}	$E^2_{xy}/E_{xx}E^*_{yy}$
Treatment + error	$(v-1)^2-1$	$(T_{yy}+E_{yy})'=T_{yy}+E_{yy}-(T_{xy}+E_{xy})^2/(T_{xx}+E_{xx})$		-	-
Treatment adj.	v-1	$T'_{yy}=(T_{yy}+E_{yy})'-E'_{yy}=R(\tau/\mu,\rho,\gamma,\beta)$		T^*_{yy}	T^*_{yy}/E^*_{yy}

Treatment means unadjusted		adjusted y
$\bar{y}_{.1.}$	$\bar{x}_{.1.}$	$\bar{y}_{.1.}-b_E(\bar{x}_{.1.}-\bar{x}_{...})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{.v.}$	$\bar{x}_{.v.}$	$\bar{y}_{.v.}-b_E(\bar{x}_{.v.}-\bar{x}_{...})$
$\bar{y}_{...}$	$\bar{x}_{...}$	-

where $b_E = E_{xy}/E_{xx}$, $\bar{y}_{.i.}$ and $\bar{x}_{.i.}$ are treatment i means from r observations, and $\bar{y}_{...}$ and $\bar{x}_{...}$ are overall means for the variates Y_{hij} and X_{hij}, respectively.

Standard error of a difference between 2 adjusted treatment means, i and i'

$$\sqrt{E^*_{yy}\left\{\frac{2}{v}+\frac{(\bar{x}_{.i.}-\bar{x}_{.i'.})^2}{E_{xx}}\right\}}$$

Table 2.3. (Cont'd)

Average standard error of a difference between 2 adjusted treatment means

$$\sqrt{\frac{2}{v} E^*_{yy} \left\{1 + T_{xx}/(v-1)E_{xx}\right\}} = A^*$$

Efficiency of covariance

$$2E_{yy}/v(v-1)(v-2)A^*$$

Residuals

Residuals for Y_{hij}: $\hat{e}_{hijy} = Y_{hij} - \bar{y}_{h\cdot\cdot} - \bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot\cdot j} + 2\bar{y}_{\cdot\cdot\cdot}$

Residuals for X_{hij}: $\hat{e}_{hijx} = X_{hij} - \bar{x}_{h\cdot\cdot} - \bar{x}_{\cdot i\cdot} - \bar{x}_{\cdot\cdot j} + 2\bar{x}_{\cdot\cdot\cdot}$

Residuals for adjusted Y_{hij}: $e'_{hij} = \hat{e}_{hijy} - b_E\hat{e}_{hijx}$

Solutions for fixed effects

$$\hat{\mu} = \bar{y}_{\cdot\cdot\cdot}$$

$$\widehat{\mu + \tau_i} = \bar{y}_{\cdot i\cdot} - b_E(\bar{x}_{\cdot i\cdot} - \bar{x}_{\cdot\cdot\cdot})$$

$$\hat{\tau}_i = \bar{y}_{\cdot i\cdot} - \bar{y}_{\cdot\cdot\cdot} - b_E(\bar{x}_{\cdot i\cdot} - \bar{x}_{\cdot\cdot\cdot})$$

Table 2.4. Covariance analysis for a split plot design.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2		F
Total	rab-1	S_{yy}	S_{xy}	S_{xx}		-
Block	r-1	R_{yy}	R_{xy}	R_{xx}		-
W.p.treat.=A	a-1	W_{yy}	W_{xy}	W_{xx}		$(r-1)W_{xx}/E_{xx}$
Error (a)	(r-1)(a-1)	A_{yy}	A_{xy}	A_{xx}		
S.p.treat.=B	b-1	T_{yy}	T_{xy}	T_{xx}		$a(r-1)T_{xx}/E_{xx}$
A × B	(a-1)(b-1)	I_{yy}	I_{xy}	I_{xx}		$a(r-1)I_{xx}/(a-1)E_{xx}$
Error (b)	a(r-1)(b-1)	E_{yy}	E_{xy}	E_{xx}		-
		Adjusted sum of squares			Mean square	F
Error(a) adj.	(r-1)(a-1)-1	$A'_{yy}=A_{yy}-A^2_{xy}/A_{xx}$			A^*_{yy}	-
Error(a) + w.p.tr.	r(a-1)-1	$(W_{yy}+A_{yy})'=W_{yy}+A_{yy}$ $-(W_{xy}+A_{xy})^2/(W_{xx}+A_{xx})$			-	-
A adj.	a-1	$W'_{yy}=(W_{yy}+A_{yy})'-A'_{yy}$ $= R(\alpha/\mu,\rho,\beta_A)$			W^*_{yy}	W^*_{yy}/A^*_{yy}
Error(b) adj.	a(r-1)(b-1)-1	$E'_{yy}=E_{yy}-E^2_{xy}/E_{xx}$			E^*_{yy}	-
Error(b) + B adj.	(b-1)(ar-a+1)-1	$(T_{yy}+W_{yy})'=T_{yy}+E_{yy}$ $-(T_{xy}+E_{xy})^2/(T_{xx}+E_{xx})$			-	-
B adj.	b-1	$T'_{yy}=(T_{yy}+E_{yy})'-E'_{yy}$			T^*_{yy}	T^*_{yy}/E^*_{yy}
Error(b)+A×B	(b-1)(ar-1)-1	$(I_{yy}+E_{yy})'=I_{yy}+E_{yy}$			-	-
A×B adj.	(a-1)(b-1)	$I'_{yy}=(I_{yy}+E_{yy})'-E'_{yy}$			I^*_{yy}	I^*_{yy}/E^*_{yy}
Error (a) regression	1	A^2_{xy}/A_{xx}			A^2_{xy}/A_{xx}	$A^2_{xy}/A_{xx}A^*_{yy}$
Error (b) regression	1	E^2_{xy}/E_{xx}			E^2_{xy}/E_{xx}	$E^2_{xy}/E_{xx}E^*_{yy}$

Table 2.4. (Cont'd)

Whole plot treatment means

unadjusted		adjusted y
$\bar{y}_{.1.}$	$\bar{x}_{.1.}$	$\bar{y}_{.1.}-b_A(\bar{x}_{.1.}-\bar{x}_{...})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{.a.}$	$\bar{x}_{.a.}$	$\bar{y}_{.a.}-b_A(\bar{x}_{.a.}-\bar{x}_{...})$
$\bar{y}_{...}$	$\bar{x}_{...}$	

where $b_A = A_{xy}/A_{xx}$, $\bar{y}_{.i.}$ and $\bar{x}_{.i.}$ are whole plot treatment i means from rb observations, and $\bar{y}_{...}$ and $\bar{x}_{...}$ are overall means for the variates Y_{hij} and X_{hij}, respectively.

Split plot treatment means

unadjusted		adjusted y
$\bar{y}_{..1}$	$\bar{x}_{..1}$	$\bar{y}_{..1}-b_E(\bar{x}_{..1}-\bar{x}_{...})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{..b}$	$\bar{x}_{..b}$	$\bar{y}_{..b}-b_E(\bar{x}_{..b}-\bar{x}_{...})$

where $b_E = E_{xy}/E_{xx}$, and $\bar{y}_{..j}$ and $\bar{x}_{..j}$ are split plot treatment j means from ra observations.

Split plot treatments within levels of whole plot treatments

unadjusted		adjusted y
$\bar{y}_{.11}$	$\bar{x}_{.11}$	$\bar{y}_{.11}-b_E(\bar{x}_{.11}-\bar{x}_{.1.})$ $-b_A(\bar{x}_{.1.}-\bar{x}_{...})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{.1b}$	$\bar{x}_{.1b}$	$\bar{y}_{.1b}-b_E(\bar{x}_{.1b}-\bar{x}_{.1.})$ $-b_A(\bar{x}_{.1b}-\bar{x}_{...})$
$\bar{y}_{.21}$	$\bar{x}_{.21}$	$\bar{y}_{.21}-b_E(\bar{x}_{.21}-\bar{x}_{.2.})$ $-b_A(\bar{x}_{.2.}-\bar{x}_{...})$
$\vdots$	$\vdots$	$\vdots$
$\bar{y}_{.ab}$	$\bar{x}_{.ab}$	$\bar{y}_{.ab}-b_E(\bar{x}_{.ab}-\bar{x}_{.a.})$ $-b_A(\bar{x}_{.a.}-\bar{x}_{...})$

where $\bar{y}_{.ij}$ and $\bar{x}_{.ij}$ are treatment ij means for j^{th} split treatment in i^{th} whole plot treatment from r observations for variates Y_{hij} and X_{hij}, respectively

Table 2.4. (Cont'd)

Standard error of a difference between 2 adjusted whole plot treatment means i and i'

$$\sqrt{A^*_{yy} \left\{ \frac{2}{rb} + \frac{(\bar{x}_{.i.}-\bar{x}_{.i'.})^2}{A_{xx}} \right.}$$

Average standard error of difference between 2 adjusted whole plot treatment means

$$\sqrt{\frac{2}{rb} A^*_{yy} \left\{1 + W_{xx}/(a-1)A_{xx} \right\}} = A^*_{w}$$

Standard error of a difference between 2 adjusted split plot treatment means, j and j'

$$\sqrt{E^*_{yy} \left\{ \frac{2}{ar} + \frac{(\bar{x}_{..j}-\bar{x}_{..j'})^2}{E_{xx}} \right\}}$$

Average standard error of a difference between 2 adjusted split plot treatment means

$$\sqrt{\frac{2}{ar} E^*_{yy} \left\{1 + R_{xx}/(a-1)E_{xx} \right\}} = A^*_{s}$$

Standard error of a difference between 2 adjusted split plot means at the same level of a whole plot treatment, ij and ij'

$$\sqrt{E^*_{yy} \left\{ \frac{2}{r} + \frac{(\bar{x}_{.ij}-\bar{x}_{.ij'})^2}{E_{xx}} \right\}}$$

Average standard error of a difference between 2 adjusted split plot means at the same level of a whole plot treatment

$$\sqrt{\frac{2}{r} E^*_{yy} \left\{1 + (T_{xx}+I_{xx})/a(b-1)E_{xx} \right\}} = A^*_{ws}$$

Table 2.4. (Cont'd)

Efficiency of covariance

Whole plot: $2A_{yy}/rb(a-1)(r-1)A^{*}_{w}$

Split plot: $2E_{yy}/ra^{2}(b-1)(r-1)A^{*}_{s}$

Split plot within whole plot: $2E_{yy}/ra(b-1)(r-1)A^{*}_{ws}$

Residuals

Residuals for Y_{hij}:

$$\hat{a}_{hiy} = \bar{y}_{hi.} - \bar{y}_{h..} - \bar{y}_{.i.} + \bar{y}_{...}$$

$$\hat{e}_{hijy} = Y_{hij} - \bar{y}_{hi.} - \bar{y}_{.ij} + \bar{y}_{.i.}$$

Residuals for X_{hij}:

$$\hat{a}_{hix} = \bar{x}_{hi.} - \bar{y}_{h..} - \bar{y}_{.i.} + \bar{y}_{...}$$

$$\hat{e}_{hijx} = X_{hij} - \bar{x}_{hi.} - \bar{x}_{.ij} + \bar{x}_{.i.}$$

Residuals for adjusted Y_{hij}:

$$a'_{hi} = \hat{a}_{hiy} - b_A\hat{a}_{hix}$$

$$e'_{hij} = \hat{e}_{hijy} - b_E\hat{e}_{hijx}$$

Solutions for fixed effects

$$\hat{\mu} = \bar{y}_{...}$$

$$\widehat{\mu + \tau_i} = \bar{y}_{.i.} - b_A(\bar{x}_{.i.}-\bar{x}_{...})$$

$$\hat{\tau}_i = \bar{y}_{.i.} - \bar{y}_{...} - b_A(\bar{x}_{.i.}-\bar{x}_{...})$$

$$\widehat{\mu + \alpha_j} = \bar{y}_{..j} - b_E(\bar{x}_{..j}-\bar{x}_{...})$$

$$\hat{\alpha}_j = \bar{y}_{..j} - \bar{y}_{...} - b_E(\bar{x}_{..j}-\bar{x}_{...})$$

$$\widehat{\alpha\beta}_{ij} = \bar{y}_{.ij} - \bar{y}_{.i.} - \bar{y}_{..j} + \bar{y}_{...} - b_E(\bar{x}_{.ij}-\bar{x}_{.i.}-\bar{x}_{..j}+\bar{x}_{...})$$

3. Numerical Examples of Covariance Analyses

The nature of the four numerical examples selected for the four experiment designs (the completely randomized, the randomized complete block, latin square, and split plot designs) has been discussed in the first section. A number of numerical results presented in Tables 3.1 to 3.4 are in fractions in order to eliminate any rounding errors due to lack of carrying an insufficient number of significant digits. In cases where fractions are not used, e.g., residuals for the Y variable adjusted for the covariate, a sufficient number of significant digits were carried to keep rounding errors small. The sum of squares of residuals can then be used to compute the error line sum of products and have exact or close agreement with the correct values.

It may be desirable to have the option of whether or not to compute the individual standard errors of a difference between two treatment means adjusted for error regression. If the treatment means $\bar{x}_{\cdot i \cdot}$ are not too variable or if the number of treatments v is large or moderately large, it may be desired not to compute the v(v-1)/2 individual standard errors. Instead, only the average standard error of a mean difference would be computed.

Table 3.1. Covariance analysis for completely randomized design from Searle, <u>Linear Models</u>, pages 353-355.

Source of variation	Degrees of freedom	Sum of products	xy	x^2	F
Total	6	$\underline{y}'\underline{y}-R(\mu)=392$	43	82/7	-
Education level	2	$R(\tau/\mu)=310$	40	40/7	1.91
Error	4	$\underline{y}'\underline{y}-R(\mu,\tau)=82$	3	6	-

		Adjusted sum of squares	Mean square	F
Error adj.	3	$\underline{y}'\underline{y}-R(\mu,\tau,\beta)=80.5$	26.833	-
Error regression	1	$R(\beta/\mu,\tau)=1.5$	1.500	0.06
Education + error	5	234.16	-	-
Education level adj.	2	$R(\tau/\mu,\beta)=153.66$	76.83	2.86

Education level means unadjusted		adjusted y	Regression coefficients	Standard errors of a difference between 2 adjusted means: Adjusted treatment mean	Adjusted treatment mean $88\frac{2}{7}$	Adjusted treatment mean $78\frac{2}{7}$
$\bar{y}_{1.}=73$	$\bar{x}_{1.}=3$	$73\frac{2}{7}$	$b_E=3/6=1/2$			
$\bar{y}_{2.}=78$	$\bar{x}_{2.}=3$	$78\frac{2}{7}$	$b_T=40/40/7=7$			
$\bar{y}_{3.}=89$	$\bar{x}_{3.}=5$	$88\frac{2}{7}$		$73\frac{2}{7}$	6.344	4.729
$\bar{y}_{..}=79$	$\bar{x}_{..}=25/7$	-		$78\frac{2}{7}$	6.687	-

<u>Average standard error of a difference between adjusted means</u>

$$\sqrt{\frac{6.344^2 + 4.729^2 + 6.687}{3}} = 35.778 = 5.98$$

<u>Efficiency of covariance</u>

$$\frac{1}{3}\left(\frac{82}{4}\right)\left[2\left(\frac{1}{3}+\frac{1}{2}\right)+\left(\frac{1}{2}+\frac{1}{2}\right)\right]\Big/ 35.778 = 51\%$$

Table 3.1. (Cont'd)

Residuals

Y_{ij}	X_{ij}	Adjusted $Y_{ij} = \hat{e}_{ijy} - b_E\hat{e}_{ijx} = e'_{ij}$
$\hat{e}_{11y} = 1$	$\hat{e}_{11x} = 0$	1
$\hat{e}_{12y} = -5$	$\hat{e}_{12x} = 1$	-11/2
$\hat{e}_{13y} = 4$	$\hat{e}_{13x} = -1$	9/2
$\hat{e}_{21y} = -2$	$\hat{e}_{21x} = -1$	- 3/2
$\hat{e}_{22y} = 2$	$\hat{e}_{22x} = 1$	3/2
$\hat{e}_{31y} = -4$	$\hat{e}_{31x} = -1$	- 7/2
$\hat{e}_{32y} = 4$	$\hat{e}_{32x} = 1$	7/2

Table 3.2. Covariance analysis for a randomized complete block design from Snedecor and Cochran, Statistical Methods, pages 427-428.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2	F
Total	23	$\underline{y}'\underline{y}-R(\mu)=18,678.50$	1485.00	181.33	-
Block	3	$R(\rho/\mu) = 436.17$	8.50	21.67	-
Variety	5	$R(\tau/\mu) = 9,490.00$	559.25	45.83	1.21
Error	15	$\underline{y}'\underline{y}-R(\mu,\rho,\tau)=8,752.33$	917.25	113.83	-

		Adjusted sum of squares	Mean square	F
Error adjusted	14	$\underline{y}'\underline{y}-R(\mu,\rho,\tau,\beta) = 1,361.07$	97.22	-
Error regression	1	$R(\beta/\mu,\rho,\tau) = 7,391.26$	7,391.26	76.03
Variety + error	19	4,587.99	-	-
Variety adjusted	5	$R(\tau/\mu,\rho,\beta) = 3,226.92$	645.38	6.64

Variety means unadjusted		adjusted y	Regression coefficients
$\bar{y}_{1.} = 173.00$	$\bar{x}_{1.} = 24.00$	191.8	$b_E = 917.25/113.83 = 8.058$
$\bar{y}_{2.} = 182.25$	$\bar{x}_{2.} = 25.25$	191.0	$b_T = 559.25/45.83 = 12.203$
$\bar{y}_{3.} = 194.50$	$\bar{x}_{3.} = 26.50$	193.1	
$\bar{y}_{4.} = 232.75$	$\bar{x}_{4.} = 28.00$	219.3	
$\bar{y}_{5.} = 201.00$	$\bar{x}_{5.} = 27.75$	189.6	
$\bar{y}_{6.} = 215.00$	$\bar{x}_{6.} = 26.50$	213.6	
$\bar{y}_{..} = 199.75$	$\bar{x}_{..} = 79/3$	-	

Table 3.2. (Cont'd)

Standard errors of a difference between 2 adjusted means

Adjusted variety mean	Adjusted variety mean				
	191.8	191.0	193.1	219.3	189.6
213.6	7.345	7.067	6.972	7.109	7.067
189.6	7.786	7.421	7.067	6.976	-
219.3	7.891	7.345	7.343	-	-
193.1	7.345	7.067	-	-	-
191.0	7.067	-	-	-	-

Average standard error of a difference between 2 adjusted means

$$\sqrt{\frac{2}{4}\,(97.22)\left\{1 + \frac{45.83}{5(113.83)}\right\} = 52.519} = 7.247$$

Efficiency of covariance

$$2(8{,}752.33)/4(15)(52.519) = 556\%$$

Table 3.2. (Cont'd)

Residuals times 12		Adjusted residuals
Y_{ij}	X_{ij}	$\hat{e}_{ijy} - b_E\hat{e}_{ijx} = e'_{ij}$
$\hat{e}_{11y} = 413$	$\hat{e}_{11x} = 40$	$e'_{11} = 7.557$
$\hat{e}_{12y} = -61$	$\hat{e}_{12x} = -10$	$e'_{12} = 1.632$
$\hat{e}_{13y} = 169$	$\hat{e}_{13x} = 22$	$e'_{13} = -0.690$
$\hat{e}_{14y} = -521$	$\hat{e}_{14x} = -52$	$e'_{14} = -8.499$
$\hat{e}_{21y} = -382$	$\hat{e}_{21x} = -35$	$e'_{21} = -8.331$
$\hat{e}_{22y} = 260$	$\hat{e}_{22x} = 23$	$e'_{22} = 6.222$
$\hat{e}_{23y} = 202$	$\hat{e}_{23x} = 19$	$e'_{23} = 4.075$
$\hat{e}_{24y} = -80$	$\hat{e}_{24x} = -7$	$e'_{24} = -1.966$
$\hat{e}_{31y} = -13$	$\hat{e}_{31x} = -2$	$e'_{31} = 0.260$
$\hat{e}_{32y} = -79$	$\hat{e}_{32x} = -16$	$e'_{32} = 4.161$
$\hat{e}_{33y} = -161$	$\hat{e}_{33x} = -8$	$e'_{33} = -8.045$
$\hat{e}_{34y} = 253$	$\hat{e}_{34x} = 26$	$e'_{34} = 3.625$
$\hat{e}_{41y} = -316$	$\hat{e}_{41x} = -56$	$e'_{41} = 11.270$
$\hat{e}_{42y} = 14$	$\hat{e}_{42x} = 14$	$e'_{42} = -8.234$
$\hat{e}_{43y} = 16$	$\hat{e}_{43x} = 10$	$e'_{43} = 5.382$
$\hat{e}_{44y} = 286$	$\hat{e}_{44x} = 32$	$e'_{44} = 2.346$
$\hat{e}_{51y} = 77$	$\hat{e}_{51x} = 19$	$e'_{51} = -6.342$
$\hat{e}_{52y} = -241$	$\hat{e}_{52x} = -7$	$e'_{52} = -15.383$
$\hat{e}_{53y} = -83$	$\hat{e}_{53x} = -35$	$e'_{53} = 16.586$
$\hat{e}_{54y} = 247$	$\hat{e}_{54x} = 23$	$e'_{54} = 5.139$
$\hat{e}_{61y} = 221$	$\hat{e}_{61x} = 34$	$e'_{61} = -4.414$
$\hat{e}_{62y} = 107$	$\hat{e}_{62x} = -4$	$e'_{62} = 11.603$
$\hat{e}_{63y} = -143$	$\hat{e}_{63x} = -8$	$e'_{63} = -6.545$
$\hat{e}_{64y} = -185$	$\hat{e}_{64x} = -22$	$e'_{64} = -0.644$

Table 3.3. Covariance analysis for a latin square design from Federer, Experimental Design – Theory and Applications, pages 490-495.

Source of variation	Degrees of freedom	Sum of products y^2	xy	x^2	F
Total	35	$\underline{y}'\underline{y}-R(\mu) = 73.268$	-10.175	134.750	-
Row	5	$R(\rho/\mu) = 1.906$	4.708	29.583	-
Column	5	$R(\gamma/\mu) = 10.010$	3.358	17.583	-
Hybrid	5	$R(\tau/\mu) = 32.413$	-25.208	19.917	1.18
Error	20	$\underline{y}'\underline{y}-R(\mu,\rho,\gamma,\tau)=28.939$	6.967	67.667	-
		Adjusted sum of squares		Mean square	F
Error	19	$y'y-R(\mu,\rho,\gamma,\tau,\beta)=28.222$		1.485	-
Error regression	1	$R(\beta/\mu,\rho,\gamma,\tau)=0.717$		0.717	0.48
Hybrid + error	24	57.553		-	-
Hybrid adjusted	5	$R(\tau/\mu,\rho,\gamma,\beta)=29.331$		5.866	3.95

Hybrid means: unadjusted		adjusted y	Regression coefficients
$\bar{y}_{.1.} = 7.08$	$\bar{x}_{.1.} = 17.00$	7.09	$b_E = 6.967/67.667 = 0.103$
$\bar{y}_{.2.} = 8.22$	$\bar{x}_{.2.} = 16.33$	8.30	$b_T = 25.208/19.917 = -1.266$
$\bar{y}_{.3.} = 7.08$	$\bar{x}_{.3.} = 17.00$	7.09	
$\bar{y}_{.4.} = 7.62$	$\bar{x}_{.4.} = 16.67$	7.66	
$\bar{y}_{.5.} = 7.57$	$\bar{x}_{.5.} = 16.83$	7.60	
$\bar{y}_{.6.} = 5.18$	$\bar{x}_{.6.} = 18.67$	5.02	
$\bar{y}_{...} = 7.125$	$\bar{x}_{...} = 17.08$	-	

Table 3.3. (Cont'd)

Standard errors of a difference between 2 adjusted means

	Hybrid mean				
Hybrid mean	7.09	8.30	7.09	7.66	7.60
5.02	0.746	0.711	0.706	0.705	0.754
7.60	0.704	0.705	0.704	0.763	-
7.66	0.706	0.708	0.746	-	-
7.09	0.704	0.784	-	-	-
8.30	0.711	-	-	-	-

Average standard error of a difference between 2 adjusted means

$$\sqrt{\frac{1.485}{3}\left\{1 + \frac{3.9834}{67.667}\right\}} = 0.724$$

Efficiency of covariance

$$2(28.939)/6(20)(0.5243) = 92\%$$

Table 3.3. (Cont'd)

Residuals times 6: Y_{hij}	Residuals times 6: X_{hij}	Adjusted residuals: $\hat{e}_{hijy} - b_E\hat{e}_{hijx} = e'_{hij}$
$\hat{e}_{111y} = 8.2$	$\hat{e}_{111x} = 13$	$e'_{111} = 1.144$
$\hat{e}_{122y} = -8.2$	$\hat{e}_{122x} = 2$	$e'_{122} = -1.401$
$\hat{e}_{133y} = -4.6$	$\hat{e}_{133x} = 5$	$e'_{133} = -0.852$
$\hat{e}_{144y} = -1.5$	$\hat{e}_{144x} = -12$	$e'_{144} = -0.044$
$\hat{e}_{155y} = 6.4$	$\hat{e}_{155x} = -10$	$e'_{155} = 1.238$
$\hat{e}_{166y} = -0.3$	$\hat{e}_{166x} = 2$	$e'_{166} = -0.084$
$\hat{e}_{221y} = -2.8$	$\hat{e}_{221x} = 2$	$e'_{221} = -0.501$
$\hat{e}_{232y} = 4.6$	$\hat{e}_{232x} = -11$	$e'_{232} = 0.955$
$\hat{e}_{243y} = 3.6$	$\hat{e}_{243x} = -8$	$e'_{243} = 0.737$
$\hat{e}_{254y} = -10.8$	$\hat{e}_{254x} = 14$	$e'_{254} = -2.040$
$\hat{e}_{265y} = 4.5$	$\hat{e}_{265x} = 12$	$e'_{265} = 0.544$
$\hat{e}_{216y} = 0.9$	$\hat{e}_{216x} = -9$	$e'_{216} = 0.304$
$\hat{e}_{331y} = -9.2$	$\hat{e}_{331x} = -5$	$e'_{331} = -1.448$
$\hat{e}_{352y} = 8.3$	$\hat{e}_{352x} = -1$	$e'_{352} = 1.400$
$\hat{e}_{313y} = -2.2$	$\hat{e}_{313x} = -7$	$e'_{313} = -0.247$
$\hat{e}_{364y} = 2.9$	$\hat{e}_{364x} = 6$	$e'_{364} = 0.380$
$\hat{e}_{325y} = -1.7$	$\hat{e}_{325x} = -7$	$e'_{325} = -0.163$
$\hat{e}_{346y} = 1.9$	$\hat{e}_{346x} = 14$	$e'_{346} = 0.076$
$\hat{e}_{441y} = 0.6$	$\hat{e}_{441x} = 0$	$e'_{441} = 0.100$
$\hat{e}_{462y} = -5.8$	$\hat{e}_{462x} = -15$	$e'_{462} = -0.709$
$\hat{e}_{423y} = 5.8$	$\hat{e}_{423x} = 6$	$e'_{423} = 0.864$
$\hat{e}_{414y} = 0.9$	$\hat{e}_{414x} = -5$	$e'_{414} = 0.236$
$\hat{e}_{435y} = 4.3$	$\hat{e}_{435x} = 16$	$e'_{435} = 0.442$
$\hat{e}_{456y} = -5.8$	$\hat{e}_{456x} = -2$	$e'_{456} = -0.932$
$\hat{e}_{551y} = 2.0$	$\hat{e}_{551x} = -5$	$e'_{551} = 0.419$

Table 3.3. (Cont'd)

$\hat{e}_{542y} = -3.7$	$\hat{e}_{542x} = 11$	$e'_{542} = -0.805$
$\hat{e}_{563y} = -2.5$	$\hat{e}_{563x} = 0$	$e'_{563} = -0.417$
$\hat{e}_{524y} = 10.2$	$\hat{e}_{524x} = 1$	$e'_{524} = 1.683$
$\hat{e}_{515y} = -12.6$	$\hat{e}_{515x} = -6$	$e'_{515} = -1.997$
$\hat{e}_{536y} = 6.6$	$\hat{e}_{536x} = -1$	$e'_{536} = 1.117$
$\hat{e}_{661y} = 1.2$	$\hat{e}_{661x} = -5$	$e'_{661} = 0.286$
$\hat{e}_{612y} = 4.8$	$\hat{e}_{612x} = 14$	$e'_{612} = 0.560$
$\hat{e}_{653y} = -0.1$	$\hat{e}_{653x} = 4$	$e'_{653} = -0.085$
$\hat{e}_{634y} = -1.7$	$\hat{e}_{634x} = -4$	$e'_{634} = -0.215$
$\hat{e}_{645y} = -0.9$	$\hat{e}_{645x} = -5$	$e'_{645} = -0.064$
$\hat{e}_{626y} = -3.3$	$\hat{e}_{626x} = -4$	$e'_{626} = -0.482$

Table 3.4. Covariance analysis for a split plot design from the Rothamsted Experiment Station Reports 1931, page 142.

Source of variation	Degrees of freedom	y^2	xy	x^2	F
Total	35	3,239.9444	1,959.8611	2,988.6528	-
Block	5	975.4444	219.3611	205.0695	-
Variety = V	2	118.0277	- 144.8056	224.1111	3.95
Error (a)	10	370.4723	184.8056	283.7222	-
Fertilizer = F	3	1,262.3888	1,435.2500	1,638.8194	40.81
F X V	6	23.1946	23.9167	34.5556	0.43
Error (b)	45	490.4166	241.3333	602.3750	-
		Adjusted sum of squares		Mean square	F
Error (a) adjusted	9	250.0971		27.7886	-
Error (a) + variety	11	485.3494		-	-
Variety adj. for error (a) reg.	2	235.2523		117.6261	4.23
Error (b) adjusted	44	393.7297		8.9484	-
Error (b) + fertilizer	47	498.5934		-	-
Error (b) + F X V	50	403.1477		-	-
Fertilizer adjusted	3	104.8637		34.9546	3.91
F X V adjusted	6	9.4180		1.5697	0.18
Error (a) regression	1	120.3752		120.3752	4.33
Error (b) regression	1	96.6869		96.6869	10.80

Regression coefficients

$b_A = 184.8056/283.7222 = 0.65136$; $b_V = -144.8056/224.1111 = -0.64613$;

$b_E = 241.3333/602.3750 = 0.40064$; $b_F = 1435.2500/1638.8194 = 0.8758$; and

$b_I = 23.9167/34.5556 = 0.6921$.

Table 3.4. (Cont'd)

Fertilizer	Oat Variety								
	M			G			V		
	unadjusted		adjusted	unadjusted		adjusted	unadjusted		adjusted
	$\bar{y}_{.1j}$	$\bar{x}_{.1j}$	$\bar{y}_{.1j}$	$\bar{y}_{.2j}$	$\bar{x}_{.2j}$	$\bar{y}_{.2j}$	$\bar{y}_{.3j}$	$\bar{x}_{.3j}$	$\bar{y}_{.3j}$
1	21.67	28.50	25.66	20.00	30.17	22.47	17.83	31.00	19.83
2	27.17	34.00	28.96	24.50	36.50	24.44	22.17	36.50	21.96
3	29.33	36.17	30.26	28.83	40.83	27.03	27.67	41.33	25.52
4	31.67	39.17	31.39	31.17	44.00	28.10	29.67	45.00	26.05
Variety mean	27.46	34.46	29.07	26.12	37.88	25.51	24.33	38.46	23.34

Fertilizer	Fertilizer mean		
	unadjusted		adjusted
	$\bar{y}_{..1}$	$\bar{x}_{..1}$	$\bar{y}_{..1}$
1	19.83	29.89	22.65
2	24.61	35.67	25.12
3	28.61	39.44	27.60
4	30.83	42.72	28.51
Variety mean	25.97	36.93	-

Standard error of a difference between two adjusted whole plot (variety) means

	M : 29.07	G : 25.51
V : 23.34	1.970	1.533
G : 25.51	1.860	-

Table 3.4. (Cont'd)

Average standard error of a difference between 2 adjusted variety means

$$\sqrt{\frac{2(27.7886)}{24}\left\{1 + \frac{224.1111}{2(283.7222)}\right\}} = 3.230305 = 1.797$$

Standard errors of a difference between 2 adjusted fertilizer means

	1 : 22.65	2 : 25.12	3 : 27.60
4 : 28.51	1.855	1.316	1.074
3 : 27.60	1.533	1.098	-
2 : 25.12	1.221	-	-

Average standard error of a difference between 2 adjusted fertilizer means

$$\sqrt{\frac{2}{18}(8.9484)\left\{1 + \frac{1638.8194}{3(602.375)}\right\}} = 1.895933 = 1.377$$

Standard error of an adjusted mean difference between 2 fertilizers for a given variety

Variety M	1 : 25.04	2 : 28.34	3 : 29.64
4 : 30.77	2.162	1.752	1.733
3 : 29.64	1.964	1.734	-
2 : 28.34	1.900	-	-
Variety G	1 : 22.71	2 : 24.67	3 : 27.27
4 : 28.33	2.413	1.954	1.770
3 : 27.27	2.161	1.806	-
2 : 24.67	1.892	-	-
Variety V	1 : 20.21	2 : 22.34	3 : 25.90
4 : 26.43	2.428	2.014	1.784
3 : 25.90	2.137	1.825	-
2 : 22.34	1.853	-	-

Table 3.4. (Cont'd)

Average standard error of a difference between 2 adjusted fertilizer means for one variety

$$\sqrt{\frac{2(8.9484)}{6}\left\{1 + \frac{(1638.8194+34.5556)}{(3+6)(602.3750)}\right\}} = 3.903478 = 1.976$$

Efficiency of covariance

Variety or whole plot:	2(370.4723)/24(10)(3.230305) = 96%
Fertilizer or split plot:	2(490.4166)/18(45)(1.895933) = 64%
Fertilizer within variety:	2(490.4166)/6(45)(3.903478) = 93%

Residuals for whole plots times 36 = $ra\sqrt{b}$		Residuals for adjusted $\bar{y}_{hi.}$
$\sqrt{b}\ \bar{y}_{hi.}$	$\sqrt{b}\ \bar{x}_{hi.}$	$\sqrt{b}(\hat{a}_{hiy} - b_A\hat{a}_{hix})$
$\hat{a}_{11y}$ = -203	$\hat{a}_{11x}$ = - 32	-5.0599
$\hat{a}_{12y}$ = 73	$\hat{a}_{12x}$ = 28	1.5212
$\hat{a}_{13y}$ = 130	$\hat{a}_{13x}$ = 4	3.5387
$\hat{a}_{21y}$ = 289	$\hat{a}_{21x}$ = 256	3.3959
$\hat{a}_{22y}$ = -173	$\hat{a}_{22x}$ = -206	-1.0783
$\hat{a}_{23y}$ = -116	$\hat{a}_{23x}$ = - 50	-2.3176
$\hat{a}_{31y}$ = 127	$\hat{a}_{31x}$ = 148	0.8500
$\hat{a}_{32y}$ = -119	$\hat{a}_{32x}$ = 46	-4.1378
$\hat{a}_{33y}$ = - 8	$\hat{a}_{33x}$ = -194	3.2879
$\hat{a}_{41y}$ = -197	$\hat{a}_{41x}$ = 22	-5.8703
$\hat{a}_{42y}$ = - 65	$\hat{a}_{42x}$ = -152	0.9446
$\hat{a}_{43y}$ = 262	$\hat{a}_{43x}$ = 130	4.9256
$\hat{a}_{51y}$ = -155	$\hat{a}_{51x}$ = -284	0.8330
$\hat{a}_{52y}$ = 175	$\hat{a}_{52x}$ = 226	0.7720
$\hat{a}_{53y}$ = - 20	$\hat{a}_{53x}$ = 58	-1.6050
$\hat{a}_{61y}$ = 139	$\hat{a}_{61x}$ = -110	5.8514
$\hat{a}_{62y}$ = 109	$\hat{a}_{62x}$ = 58	1.9784
$\hat{a}_{63y}$ = -248	$\hat{a}_{63x}$ = 52	-7.8297

Table 3.4. (Cont'd)

Residuals

	$24\ \hat{e}_{hijy}$			$24\ \hat{e}_{hijx}$			$\hat{e}_{hijy} - b_E\hat{e}_{hijx}$		
hj	i=M	i=G	i=V	i=M	i=G	i=V	i=M	i=G	i=V
11	7	51	- 12	- 31	5	89	0.809	2.042	-1.986
12	- 77	- 57	28	- 67	- 75	- 43	-2.090	-1.123	1.884
13	87	- 65	- 32	121	61	- 15	1.605	-3.727	-1.083
14	- 17	71	16	- 23	9	- 31	-0.324	2.808	1.184
21	- 41	- 15	66	17	- 61	131	-1.992	0.393	0.563
22	67	141	- 62	53	99	- 49	1.907	4.222	-1.765
23	39	- 59	94	- 47	- 5	75	2.410	-2.375	2.665
24	- 65	- 67	- 98	- 23	- 33	-157	-2.324	-2.241	-1.462
31	61	135	- 66	41	179	- 97	1.857	2.637	-1.131
32	- 47	- 21	- 50	- 43	- 69	59	-1.241	0.277	-3.068
33	21	- 77	106	49	- 5	15	0.057	-3.125	4.166
34	- 35	- 37	10	- 47	-105	23	-0.674	0.211	0.033
41	- 17	51	- 30	- 49	- 7	- 49	0.110	2.242	-0.432
42	67	- 81	- 38	35	- 15	179	2.207	-3.125	-4.571
43	-105	103	- 2	- 65	1	- 57	-3.290	4.275	0.868
44	55	- 73	70	79	21	- 73	0.973	-3.392	4.135
51	1	-117	48	- 19	11	- 1	0.359	-5.059	2.017
52	- 35	15	40	- 7	- 69	- 85	-1.341	1.777	3.086
53	9	79	-140	13	19	- 57	0.158	2.974	-4.882
54	25	23	52	13	39	143	0.825	0.308	-0.220
61	- 11	-105	- 6	41	-127	- 73	-1.143	-2.255	0.969
62	25	3	82	29	129	- 61	0.558	-2.028	4.435
63	- 51	19	- 26	- 71	- 71	39	-0.940	1.977	-1.734
64	37	83	- 50	1	69	95	1.525	2.306	-3.669

4. Adequacy of Package Programs to Obtain the Desired Computations

The desired computations from covariance analyses of four standard experiment designs have been discussed in previous sections. In this section we follow Heiberger's (1976a) format. The computations desired are listed on the left hand side of Table 4.1 for the randomized block and latin square designs with one covariate, and Table 4.2 for the split plot design with one covariate, with a summary of performance in Table 4.3.

Table 4.1. Printed output features of statistical package program for:
Randomized block design with one covariate
Latin square design with one covariate

	BMD	GENSTAT	SAS	SPSS
	P2V	ANOVA	GLM	ANOVA
Version/Date		4.01	76.6	H 7.02
	1977	1977	1978	1977
ANOVA table for unadjusted y	X	0	X	X
for x	X	0	X	X
Sums of products for xy	-	-	-	-
Adjusted ANOVA table for y	0	0	SS 2,3,4	Options default, 7,8,9
Sums of squares for covariates, Error regression	0	0	SS 2,3,4	Options 7,9
Significance tests	S	S	RU	R
Observed significance of test (Probability)	0	-	0	0
Treatment means	X	0	0	D (- with option 9)
Adjusted treatment means	C	0	0 (not in GLM earlier versions)	D (- with option 9)
Standard error of differences between adjusted means	-	-	T	-
Average standard error of differences between adjusted means	-	0	-	-
Single degree of freedom contrasts	-	0	-	-
Effects (coefficients, solutions)	-	0	Z	-
Regression coefficients for covariates	0	0	0	Options 7,8,9
Residuals for: unadjusted y	X	0	X	-
x	X	0	X	-
adjusted y	0	0	0	-
Estimate efficiency of covariate adj.	-	0	-	-

Notes: O = the program has the features in One procedure call

X = the program has the feature, but requires an eXtra procedure call, e.g., ANOVA without covariates

\- = the program lacks the feature

W = Wrong or inappropriate value given which the user would be tempted to use

R = all effects tested against Residual

S = appropriate test determined from Specifications

U = User-specified numerator and denominator for F-tests

D = expressed as Deviation from the mean

C = Cell means

T = does not give standard errors, but Tests the difference of 2 adjusted means and gives p values

Z = solution with Zero constraints (e.g., last factor level set to 0)

$[\;]\frac{1}{b}$ = the SS are for whole plot means and so are 1/b times the SS for observations. The F-tests are correct, with the scale factor cancelling.

P = Pool block by subplot interaction with residual to get subplot error

BMDP2V and GENSTAT ANOVA

Give the correct analysis from the design specifications.

SAS GLM options for sums of squares.

Type 2, 3, 4 are identical with orthogonal data models without interaction.

Type 1 gives sequential sums of squares and so is dependent on the order of variables specified in the model statement.

The default option gives types 1 and 4 sums of squares.

SPSS ANOVA options

Default and 10 fit covariate first and give regression coefficient for total line.

The default option for SPSS ANOVA would be more appropriately set to option 7.

The order of specification within the factor set and covariate set is irrelevent for the default, 7, 8 and 9 options, but is relevant for option 10.

It cannot handle nested designs, e.g., split plot.

Table 4.2 Printed output features of statistical package programs for split plot design with one covariate (with possibly different whole and subplot regressions).

		BMD P2V		GENSTAT ANOVA		SAS GLM			SPSS ANOVA
Version/Date		1977		4.01 1977		76.6 1978			H 7.02 1977
	Analysis 1		Analysis 2		Analysis 1		Analysis 2	Analysis 3	
ANOVA table for : unadjusted y		X		O		X		X	XP
: x		X		O		X		X	XP
Sums of products for xy		-		-		-		-	-
Adjusted ANOVA table for y:									
Whole plot adjusted: Whole plot treatment A	W		O	O	W		2,3	X[2,3,4]	W
Whole plot adjusted: Error A	W		O	O	W		O	X[0] $\frac{1}{b}$	W
Whole plot adjusted: Error (a) regression	-		O	O	-		W	X[2,3,4]	W
Subplot adjusted: Subplot treatment B	O		W	O		3,4		3,4	W
Subplot adjusted: A x B	O		W	O		2,3,4		2,3,4	W
Subplot adjusted: Error B	O		P	O		O		O	P
Subplot adjusted: Error (b) regression	O		W	O		2,3,4		2,3,4	W
Significance tests	R		S	S		RU		RU=S	SW
Observed significance test (Probability)		O		-		O		O	W
Treatment means: Whole plot		X		O		O		O	W
Treatment means: Split plot	X		O	O		O		O	W
Treatment means: Split plot within whole plot		X		O		O		O	-
Adjusted treatment means: Whole plot	-		WC	O		W		X	W
Adjusted treatment means: Split plot	C		WC	O		O		O	W
Adjusted treatment means: Split plot within whole plot		WC		O		W		W	-
Standard error of difference of means on: Whole plot		-		-		WT		XT	-
Standard error of difference of means on: Subplot		-		-		T		T	-
Standard error of difference of means on: Subplot within whole plot		-		-		WT		WT	-
Average standard error of difference of means on: Whole plot		-		O		-		-	-
Average standard error of difference of means on: Subplot		-		O		-		-	-
Average standard error of difference of means on: Subplot within whole plot		-		O		-		-	-
Single degree of freedom contrasts		-		O		-		-	-
Effects (coefficients solutions): Whole plot		-		O		W		Z	-
Effects (coefficients solutions): Subplot		-		O		Z		Z	-
Regression coefficients for covariates: Whole plot	-		O	O		-		X	-
Regression coefficients for covariates: Subplot	O		W	O		O		O	W
Residuals for Unadjusted Y: Whole plot		X		O		-		X	-
: Subplot		X		O		X		X	-
X: Whole plot		X		O		-		X	-
: Subplot		X		O		X		X	-
Adjusted Y: Whole plot	-		O	O		-		X	-
: Subplot	O		O	O		O		O	-
Efficiency of covariate: Whole plot		-		O		-		-	-
Efficiency of covariate: Subplot		-		O		-		-	-
Efficiency of covariate: Subplot within whole plot		-		O		-		-	-

Table 4.3. Summary of package capabilities for split plot design with covariate.

		BMD P2V		GENSTAT ANOVA	SAS GLM			SPSS ANOVA
Version/Date		1977		4.01 1977	76.6 1978			H 7.02 1977
		Analysis 1	Analysis 2		Analysis 1	Analysis 2	Analysis 3	
Features available	O's etc.	9	10	33	14	16	15	1
Available with extra procedure calls	X	9	8	0	4	4	13	2
Not available	-	17	13	5	13	12	8	22
Wrong or inappropriate calculation	W	3	7	0	7	6	2	13

SPECIFICATION OF SPLIT PLOT ANALYSIS WITH COVARIATE

BMDP2V

ANALYSIS 1 - COVARIATE ADJUSTED ON ERROR(B) LINE

```
/PROBLEM       TITLE IS 'SPLIT PLOT DESIGN WITH COVARIATE'.
/INPUT         VARIABLES ARE 5.
               FORMAT IS '(F1.0,2F2.0,2F3.0)'.
/VARIABLE      NAMES ARE BLOCK,VARIETY,NITROGEN,X,YIELD.
/DESIGN        GROUPING ARE 1,2,3.
               DEPENDENT IS 5.
               COVARIATE IS 4.
               INCLUDE IS 1,2,3,12,23.
               RESIDUAL = MEAN.
               PRINT.
/END
1 1 1 24 16
1 1 2 28 18
1 1 3 38 27
-----------
     .
    DATA
     .
-----------
6 3 2 36 23
6 3 3 45 24
6 3 4 51 25
```

ANALYSIS 2 - REPEATED MEASURES FORMULATION

```
/PROBLEM       TITLE IS 'SPLIT PLOT WITH COVARIATE USING REPEATED MEASURES'.
/INPUT         VARIABLES ARE 10.
               FORMAT IS '(F1.0,F2.0,8F3.0)'.
/VARIABLE      NAMES ARE BLOCK,VARIETY,X1,Y1,X2,Y2,X3,Y3,X4,Y4.
/DESIGN        GROUPING ARE BLOCK,VARIETY.
               DEPENDENT ARE 4,6,8,10.
               COVARIATE IS 3,5,7,9.
               LEVEL IS 4.
               NAME IS NITROGEN.
               EXCLUDE IS 12.
               RESIDUAL = MEAN.
               PRINT.
/END
1 1 24 16 28 18 38 27 35 25
1 2 28 20 31 20 41 24 42 32
1 3 32 16 32 22 38 25 41 29
---------------------------
             .
            DATA
             .
---------------------------
6 1 30 24 35 31 33 30 39 36
6 2 27 18 44 27 40 32 49 37
6 3 30 15 36 23 45 24 51 25
```

GENSTAT ANOVA

```
'REFERENCE' SPLIT_PLOT
'PAGE'
'CAPTION'
          ''
             SPLIT PLOT DESIGN
             WITH 1 COVARIATE
          ''
'UNITS'      $ 72
'NAMES'      VARLEVELS = MARVLOUS, GOLDRAIN, VICTORY
   :         NITLEVELS = 0-CWT, 0.2-CWT, 0.4-CWT, 0.6-CWT
'FACTOR'     BLOCKS    $ 6
   :         PLOTS     $ 3
   :         SUBPLOTS  $ 4
   :         VARIETY   $ VARLEVELS
   :         NITROGEN  $ NITLEVELS
'GENERATE'   BLOCKS, PLOTS,SUBPLOTS
'READ/P,PRIN=DEM,FLEV=F'  VARIETY, NITROGEN, X, YIELD $ S , 1X, 4
'BLOCKS'     BLOCKS / PLOTS / SUBPLOTS
'TREATMENTS' VARIETY * NITROGEN
'COVARIATES' X
'ANOVA/ PR=12313, PRX=10013, PRYU=10013'  YIELD
'PAGE'
'RUN'
1 1 1 24 16
1 1 2 28 18
1 1 3 38 27
-----------
     .
    DATA
     .
-----------
6 3 2 36 23
6 3 3 45 24
6 3 4 51 25
'EOD'
'CLOSE'
```

SAS GLM

```
COMMENT
                SPLIT PLOT DESIGN
                WITH 1 COVARIATE ;
DATA ORIGINAL;
INPUT BLOCK VARIETY NITROGEN X YIELD;
CARDS;
1 1 1 24 16
1 1 2 28 18
1 1 3 38 27
-----------
     .
    DATA
     .
-----------
6 3 2 36 23
6 3 3 45 24
6 3 4 51 25
PROC PRINT;
TITLE1          SPLIT PLOT DESIGN  ;
TITLE2          WITH 1 COVARIATE ;
PROC GLM DATA=ORIGINAL;
CLASSES         BLOCK VARIETY NITROGEN;
MEANS  VARIETY;
MODEL     X  YIELD = BLOCK VARIETY BLOCK*VARIETY NITROGEN VARIETY*NITROGEN /P ;
TEST            H = VARIETY  E = BLOCK*VARIETY;
PAGE;

            ANALYSIS 1 - COVARIATE ADJUSTED ON ERROR(B) LINE

PROC GLM;
CLASSES         BLOCK VARIETY NITROGEN;
MEANS  VARIETY;
MODEL           YIELD = BLOCK VARIETY BLOCK*VARIETY X NITROGEN VARIETY*NITROGEN
                                                   / XPX SOLUTION P SS1 SS2 SS3 SS4;
TEST            H = VARIETY  E = BLOCK*VARIETY;
LSMEANS  VARIETY / E STDERR PDIFF;
PAGE;

            ANALYSIS 2 - ADDITIONAL COVARIATE OF WHOLE PLOT MEANS

PROC SORT;  BY BLOCK VARIETY;
PROC MEANS; BY BLOCK VARIETY;
VAR X;
OUTPUT MEAN=XBAR;
PROC MATRIX;
FETCH XM;
ONE4 = 1/1/1/1;
XK = XM @ ONE4;
OUTPUT XK OUT=XMEAN(RENAME=(COL3=XA));
DATA COMPLETE;
MERGE ORIGINAL XMEAN;
PROC PRINT;

PROC GLM;
CLASSES         BLOCK VARIETY NITROGEN;
MEANS  VARIETY;
MODEL           YIELD = BLOCK XA VARIETY BLOCK*VARIETY X NITROGEN VARIETY*NITROGEN
                                                   / XPX SOLUTION P SS1 SS2 SS3 SS4;
TEST            H = VARIETY  E = BLOCK*VARIETY;
LSMEANS  VARIETY / E STDERR PDIFF;

            ANALYSIS 3 - ANALYSIS OF WHOLE PLOT MEANS AND OBSERVATIONS

PROC SORT;  BY BLOCK VARIETY;
PROC MEANS; BY BLOCK VARIETY;
VAR X YIELD;
OUTPUT OUT=W_PLOT MEAN=WP_X WP_YIELD;
PROC PRINT;
PROC GLM DATA=W_PLOT;
CLASSES         BLOCK VARIETY;
MEANS  VARIETY;
MODEL           WP_YIELD = BLOCK WP_X VARIETY
                                                   / SOLUTION P SS1 SS2 SS3 SS4;
LSMEANS  VARIETY / STDERR PDIFF;
PAGE;
PROC GLM DATA=ORIGINAL;
CLASSES         BLOCK VARIETY NITROGEN;
MEANS  VARIETY;
MODEL           YIELD = BLOCK*VARIETY X NITROGEN VARIETY*NITROGEN
                                                   / SOLUTION P SS1 SS2 SS3 SS4;
LSMEANS  VARIETY*NITROGEN NITROGEN / STDERR PDIFF;
```

SPSS ANOVA

DOES NOT HANDLE SLIT PLOT DESIGN AND OTHER NESTED DESIGNS

5. Recommendations

In general, the labelling of each SS in the output should be made more explicit and informative. Source A is not an acceptable label to describe A/μ; $A/\mu,\beta$; $A/\mu,X,\beta$; the SS for A from the weighted squares of means analysis, and many others, for factors A and B and covariate X. Ideally, the R( ) notation should be followed, where applicable, using the variable names rather than the corresponding parameters. The additional complication of restricted models, with different sets of constraints imposed on the model (rather than just on the solutions) can also be denoted by including a symbol to denote the constraint. For example, Searle (1977) uses $R^*(\alpha/\mu,\beta,\gamma)_\Sigma$, the Σ denoting the Σ or usual constraints and the * designating it is for a restricted model, to denote SSA_w, the SS from the weighted squares of means analysis. The corresponding variable names with U to denote usual constraints gives the equivalent (A/MU,B,AB)U which could be used in output. In designs with a large number of factors, interactions, or covariates further compromises might need to be made so that, for example, A/FACTORS, X/COVARIATES and A/COVARIATES, FACTORS could denote A adjusted for all other factors, X adjusted for all other covariates and A adjusted for all covariates and all other factors, respectively. When space limitations preclude the use of the variable name, use first letter as in BMDP2V, could be used.

Of the four packages investigated, the user is well advised to use GENSTAT ANOVA for an almost complete analysis of orthogonal designs and designs with balanced or partial confounding, with its block and treatment formulation giving a succinct description of the design.

Literature Cited

Anderson, R. L. (1946). Missing-plot techniques. Biometrics Bulletin 2:41-47.

Bartlett, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. Journal Royal Statistical Society, Suppl. 4:137-170.

Federer, W. T. (1955). *Experimental Design – Theory and Application*, Chapter XVI, Macmillan, New York (republished by Oxford and IBH Publishing Company, Calcutta and New Delhi in 1967, 3rd printing in 1977).

Heiberger, R. M. (1976a). Conceptualization of experimental designs and their specification and computation with ANOVA programs. Proceedings, Statistical Computing Section, American Statistical Association, pages 13-23.

Heighberger, R. M. (1976b). Criteria and considerations for computer programs for the analysis of designed experiments. Technical Report 4, Department of Statistics, The Wharton School, University of Pennsylvania, 12 pp.

Robson, D. S. and G. F. Atkinson. (1960). Individual degrees of freedom for testing homogeneity of regression coefficients in a one-way analysis of covariance. Biometrics 16:593-605.

Rothamsted Experiment Station Reports, 1931, page 142. Rothamsted Experiment Station, Harpenden Herts, England.

Searle, S. R. (1971). *Linear Models*, section 8.2, Wiley, New York, N. Y.

Searle, S. R. (1977). Illustrative calculations of sums of squares in the 2-way classification, unbalanced data, all cells filled. BU-608-M in the Biometrics Unit Mimeo Series, Cornell University.

Smith, H. F. (1957). Interpretation of adjusted treatment means and regressions in analysis of covariance. Biometrics 13:282-308.

Snedecor, G. W. and W. G. Cochran. (1967). *Statistical Methods*, Chapter 14. The Iowa State University Press, Ames, Iowa.

Steel, R. G. D. and W. T. Federer. (1955). Yield-stand analyses. Journal Indian Society Agricultural Statistics 7:27-45.

Truitt, J. T. and H. F. Smith. (1956). Adjustment by covariance and consequent tests of significance in split-plot experiments. Biometrics 12:23-39.