A generalization of the
Young-Whittle model

32

Jan de Leeuw

RB - 006 - 73

Summary

In this paper we analyse a model which is somewhere on the border line of factor analysis and fixed effects analysis of variance, combining an additive structure of means and a multiplicative structure of interactions with an additive structure of variances. This generalizes some recent developments in ANOVA, and it also extends some old ideas from the factor analysis area. We derive least squares estimates of the parameters and MINQUE estimates of the variances. The possibility of simultaneous maximum likelihood estimation of all parameters is briefly discussed. In general our treatment of the model is sketchy and incomplete.

0 Introduction

0.1 Notational conventions

Lower case Greek letters are used for scalars, lower case Latin letters for column vectors, and capital Latin or Greek letters for matrices. An important exception a the lower case Latin subscripts and superscripts, which are used for indices and their ranges. A tilde under a symbol indicates that the symbol stands for a random variable (scalar, vector, or matrix), no tilde means that the symbol is a constant (parameter). We use the notation $\underset{\sim}{\chi} \sim \mathcal{N}(\mu, \sigma^2)$ to indicate that (the scalar rv) $\underset{\sim}{\chi}$ is normally distributed with mean μ and variance σ^2. For vectors the corresponding notation is $\underset{\sim}{x} \sim \mathcal{N}_p(u, \Sigma)$, indicating that the p-dimensional random vector x is multinormally distributed with vector of means u and dispersion matrix Σ. If no confusion is possible the subscript p is dropped. Unless otherwis specified all random variables are assumed to be mutually independent in this pape Primes are used to denote row vectors and transposed matrices.

0.2 Factor analysis

There are two basically different linear structural models called 'factor analysis in the statistical and psychometric litera ture. The first model is

A1: $\underset{\sim}{\chi}^k_{ij} = \gamma_j + \sum_{s=1}^{p} \alpha_{is} \beta_{js} + \underset{\sim}{\epsilon}^k_{ij}$,

A2: $\underset{\sim}{\epsilon}^k_{ij} \sim \mathcal{N}(0, \sigma^2_j)$.

The second model is

B1: $\underset{\sim}{\chi}^k_{ij} = \gamma_j + \sum_{s=1}^{p} \underset{\sim}{\varphi}_s \beta_{js} + \underset{\sim}{\epsilon}^k_{ij}$,

B2: $\underset{\sim}{\epsilon}^k_{ij} \sim \mathcal{N}(0, \sigma^2_j)$,

B3: $\underset{\sim}{f} \sim \mathcal{N}_p(0, \Phi)$.

In both models the range of the indices is

$i = 1, \ldots, n$;

$j = 1, \ldots, m$;

$k = 1, \ldots, l_{ij}$.

In this paper we study a straightforward generalization of model A. Other generalizations, and similar generalizations of model B will be discussed in a related series of papers on factor analysis.

1 Model

1.1 Assumptions

In this paper we investigate properties of the model

A1: $\underset{\sim}{\chi}_{ij}^{k} = \delta + \lambda_i + \mu_j + \sum_{s=1}^{p} \alpha_{is}\beta_{js} + \underset{\sim}{\omega}_i^k + \underset{\sim}{\eta}_j^k,$

A2: i) $\underset{\sim}{\omega}_i^k \sim \mathcal{N}(0,\ \rho_i^2),$

ii) $\underset{\sim}{\eta}_j^k \sim \mathcal{N}(0,\ \tau_j^2).$

Observe that this implies (and for all practical purposes is equivalent to)

A1': $\underset{\sim}{\chi}_{ij}^{k} = \delta + \lambda_i + \mu_j + \sum_{s=1}^{p} \alpha_{is}\beta_{js} + \underset{\sim}{\epsilon}_{ij}^{k}$,

A2': $\underset{\sim}{\epsilon}_{ij}^{k} \sim \mathcal{N}(0,\ \rho_i^2 + \tau_j^2).$

1.2 Indices

The range of the indices is

$i = 1, \ldots , n$;

$j = 1, \ldots , m$;

$k = 1, \ldots , l_{ij}$.

Let

$l_{i*} = \sum_{j=1}^{m} l_{ij},$

$l_{*j} = \sum_{i=1}^{n} l_{ij},$

$l_{**} = \sum_{i=1}^{n} \sum_{j=1}^{m} l_{ij},$

and

$\pi_{ij} = l_{ij}/l_{**},$

$\pi_{i.} = l_{i*}/l_{**},$

$\pi_{.j} = l_{*j}/l_{**},$

$\pi_{i|j} = l_{ij}/l_{*j},$

$\pi_{j|i} = l_{ij}/l_{i*}.$

In an important special case $\pi_{ij} = \pi_{i.}\pi_{.j}$ for all $i=1,\ldots,n; j=1,\ldots,m$. This is the proportional frequencies (PF) case. In fact the general case leads to not very interesting complications and we concentrate on the PF-case.

1.3 Identification

If the equations E given by

$$x_{ij} = \delta + \lambda_i + \mu_j + \sum_{s=1}^{p} \alpha_{is}\beta_{js}$$

has a solution, then this solution may not be unique. In this section we describe a set of conditions I with the property that if E has a solution, then it also has a solution satisfying I. In the first place we can choose the λ_i and μ_j in such a way that

I1: $\sum \lambda_i \pi_{i.} = 0,$

I2: $\sum \mu_j \pi_{.j} = 0.$

For α_{is} and β_{js} we use the conditions

I3: $\sum \pi_{i.} \alpha_{is} \alpha_{it} = 0$ for all $s,t=0,\dots,p$ with $s \neq t$,

I4: $\sum \pi_{.j} \beta_{js} \beta_{jt} = 0$ for all $s,t=0,\dots,p$ with $s \neq t$.

Here

$\alpha_{i0} = 1$ for all $i=1,\dots,n$,

$\beta_{j0} = 1$ for all $j=1,\dots,m$.

Another type of identification can be achieved by rewriting

$$\theta_{ij} = \sum_{s=1}^{p} \alpha_{is}\beta_{js}$$

as

$$\theta_{ij} = \sum_{s=1}^{p} \alpha_{is}\beta_{js}\gamma_s,$$

and by requiring

I5: $\sum \pi_{i.} \alpha_{is}^2 = 1$ for all $s=1,\dots,p$,

I6: $\sum \pi_{.j} \beta_{js}^2 = 1$ for all $s=1,\dots,p$

(of course these equations are automatically satisfied for $s = 0$). Finally we require

I7: $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_p \geq 0.$

The number of parameters is now

$N_p = 1 + m + n + pm + pn + p = (p+1)(n+m+1).$

The number of identification equations is

$N_I = 1 + 1 + \frac{1}{2}p(p+1) + \frac{1}{2}p(p+1) + p + p = (p+1)(p+2).$

There remain

$N_F = N_P - N_I = (n+m-p-1)(p+1)$

free parameters. Because equation E is symmetric in both modes of classification we can suppose that $m \leq n$. It follows that the number of unknowns (free variables) is less than the number of equations if $p < m-1$.

1.4 Solvability

The equations

$$\chi_{ij} = \delta + \lambda_i + \mu_j + \theta_{ij},$$

where the parameters are identified by

$$\sum \lambda_i \pi_{i.} = 0,$$

$$\sum \lambda_j \pi_{.j} = 0,$$

$$\sum \theta_{ij} \pi_{i.} = 0 \text{ for all } j=1,\dots,m,$$

$$\sum \theta_{ij} \pi_{.j} = 0 \text{ for all } i=1,\dots,n,$$

can always be solved by setting

$$\tilde{\delta} = x_{..},$$

$$\tilde{\lambda}_i = x_{i.} - x_{..},$$

$$\tilde{\mu}_j = x_{.j} - x_{..},$$

$$\tilde{\theta}_{ij} = x_{ij} - x_{i.} - x_{.j} + x_{..}.$$

Consequently a sufficient condition for the solvability of

$$\chi_{ij} = \delta + \lambda_i + \mu_j + \sum_{s=1}^{p} \alpha_{is} \beta_{js} \gamma_s$$

is that the rank of $\tilde{\Theta}$ does not exceed p. The condition is also necessary. Because $\text{rank}(\tilde{\Theta}) \leq m-1 \leq n-1$, the condition $p \geq m-1$ is sufficient for solvability.

1.5 Uniqueness

It is obvious that a NASC for the uniqueness of the solution of our set of equations E and I is that both the decomposition

$$\chi_{ij} = \delta + \lambda_i + \mu_j + \theta_{ij},$$

and the decomposition

$$\hat{\theta}_{ij} = \sum_{s=1}^{p} \alpha_{is} \beta_{js} \gamma_s$$

have unique solutions satisfying the conditions I. There remains only a single type of non-uniqueness that we usu ally do not eliminate in practice: for each s we can change the signs of all α_{is} en β_{js} simultaneously. This can be eliminated by the requirement

I8: Let k_s be the smallest integer such that $\alpha_{k_s s} \neq 0$. Then $\alpha_{k_s s} > 0$ for all $s=1,\ldots,p$.

A NASC for the uniquess of both decompositions (given that $\pi_{ij} > 0$ for all $i=1,\ldots,n$; $j=1,\ldots,m$) is that $\gamma_1 > \gamma_2 > \cdots > \gamma_p > 0$. A necessary condition is $\mathrm{rank}(\tilde{\Theta}) = p$.

2 Estimation

2.1 Estimation of variances by MINQUE methods

Consider the slightly more general model

$$\text{A1:}\quad \underset{\sim}{\chi}_{ij}^{k} = \delta + \lambda_i + \mu_j + \sum_{s=1}^{p} \alpha_{is}\beta_{js} + \underset{\sim}{\omega}_i^{k} + \underset{\sim}{\eta}_j^{k} + \underset{\sim}{\kappa}^{k},$$

where

$$\text{A2: iii)}\quad \underset{\sim}{\kappa}^{k} \sim \mathcal{N}(0, \varepsilon^2).$$

This implies that if

$$\underset{\sim}{\epsilon}_{ij}^{k} = \underset{\sim}{\omega}_i^{k} + \underset{\sim}{\eta}_j^{k} + \underset{\sim}{\kappa}^{k},$$

then

$$\underset{\sim}{\epsilon}_{ij}^{k} \sim \mathcal{N}(0, \rho_i^2 + \tau_j^2 + \varepsilon^2).$$

Suppose ϵ_{ij}^{k} are observations on $\underset{\sim}{\epsilon}_{ij}^{k}$. Define

$$\epsilon_{ij} = \sum (\epsilon_{ij}^{k})^2.$$

Obviously

$$\epsilon_{ij}/(\rho_i^2 + \tau_j^2 + \varepsilon^2) \sim \chi^2(l_{ij}),$$

and thus

$$E(\epsilon_{ij}) = (\rho_i^2 + \tau_j^2 + \varepsilon^2)l_{ij},$$

$$V(\epsilon_{ij}) = 2(\rho_i^2 + \tau_j^2 + \varepsilon^2)^2 l_{ij}.$$

An estimator of the form

$$\hat{\eta} = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij}\epsilon_{ij}$$

has expected value

$$E(\hat{\eta}) = \sum \rho_i^2 \sum l_{ij}\gamma_{ij} + \sum \tau_j^2 \sum l_{ij}\gamma_{ij} + \varepsilon^2 \sum\sum l_{ij}\gamma_{ij},$$

and variance

$$V(\hat{\eta}) = 2\sum\sum \gamma_{ij}^2 l_{ij}(\rho_i^2 + \tau_j^2 + \varepsilon^2)^2.$$

If $\hat{\eta}$ is to be an unbiased estimator of

$$\eta = \sum \underline{\alpha}_i \rho_i^2 + \sum \underline{\beta}_j \tau_j^2 + \underline{\gamma}\,\varepsilon^2$$

(the bar under these symbols is used to distinguish them from our parameters α, β, and γ) we must have

$$\sum l_{ij}\gamma_{ij} = \underline{\alpha}_i \text{ for all } i=1,\ldots,n,$$

$\sum l_{ij} \tau_{ij} = \underline{\beta}_j$ for all $j=1,\ldots,m$,

$\sum\sum l_{ij} \tau_{ij} = \underline{\gamma}$.

For η we clearly must have

$\sum \underline{\alpha}_i = \sum \underline{\beta}_j = \underline{\gamma}$.

The MINQUE estimator minimizes

$$\phi = \sum\sum \tau_{ij}^2 l_{ij},$$

under these conditions. Using undetermined multipliers $\underline{\lambda}_i$, $\underline{\mu}_j$, and $\underline{\delta}$ (again the ba is used to distinguish them from the parameters of the model) we find

$$\tau_{ij} = \underline{\lambda}_i + \underline{\mu}_j + \underline{\delta},$$

where

$$\underline{\lambda}_i + \sum \pi_{j|i} \underline{\mu}_j + \underline{\delta} = \underline{\alpha}_i / l_{i*},$$

$$\sum \pi_{i|j} \underline{\lambda}_i + \underline{\mu}_j + \underline{\delta} = \underline{\beta}_j / l_{*j},$$

$$\sum \pi_{i.} \underline{\lambda}_i + \sum \pi_{.j} \underline{\mu}_j + \underline{\delta} = \underline{\gamma} / l_{**}.$$

Thus

$$\tilde{\tau}_{ij} = \underline{\alpha}_i / l_{i*} + \underline{\beta}_j / l_{*j} - \underline{\gamma} / l_{**} + \sum (\pi_{j|i} - \pi_{.j}) \underline{\mu}_j + \sum (\pi_{i|j} - \pi_{i.}) \underline{\lambda}_i,$$

and in the PF-case

$$\tilde{\tau}_{ij} = \underline{\alpha}_i / l_{i*} + \underline{\beta}_j / l_{*j} - \underline{\gamma} / l_{**}.$$

If

$$\eta = \rho_i^2 + \partial_j^2 + \epsilon^2,$$

then

$$\hat{\eta} = (\sum_{l=1}^{m} \epsilon_{il}) / l_{i*} + (\sum_{k=1}^{n} \epsilon_{kj}) / l_{*j} - (\sum_{k=1}^{n} \sum_{l=1}^{m} \epsilon_{kl}) / l_{**}.$$

If

$$\eta = \rho_i^2 - \rho_.^2,$$

then

$$\hat{\eta} = (\sum_{l=1}^{m} \epsilon_{il}) / l_{i*} - (\sum_{k=1}^{n} (\sum_{l=1}^{m} \epsilon_{kl}) / l_{k*}) / n.$$

A similar formula can be derived for

$$\eta = \partial_j^2 - \partial_.^2.$$

We find

$$\hat{\eta} = \epsilon_{.j} - (\sum_{l=1}^{m} \epsilon_{.l})/m.$$

In the special case in which

$\underset{\sim}{\omega}_i^k \equiv 0$ for all $i=1,\ldots,n$; $j=1,\ldots,m$; $k=1,\ldots,l_{ij}$

i.e.

$$\underset{\sim}{\epsilon}_{ij}^k \sim \eta(0, \delta_j^2 + \varepsilon^2)$$

we minimize ϕ under the conditions

$$\sum_{i=1}^{n} l_{ij} \tau_{ij} = \underline{\beta}_j \text{ for all } j=1,\ldots,m,$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} l_{ij} \tau_{ij} = \chi,$$

where

$$\sum_{j=1}^{m} \beta_j = \chi.$$

It follows that

$$\tau_{ij} = \underline{\mu}_j + \underline{\delta} = \underline{\beta}_j / l_{*j},$$

and thus

$$\hat{\eta} = \sum_{j=1}^{m} \underline{\beta}_j \epsilon_{.j}$$

estimates

$$\eta = \sum_{j=1}^{m} \underline{\beta}_j \; \delta_j^2 + \chi \varepsilon^2.$$

If

$$\eta = \delta_j^2 + \varepsilon^2$$

we find

$$\hat{\eta} = \epsilon_{.j}.$$

Contrasts of the δ_j^2 are estimated by corresponding contrasts of the $\epsilon_{.j}$. In a similar way if

$\underset{\sim}{\eta}_j^k \equiv 0$ for all $i=1,\ldots,n$; $j=1,\ldots,m$; $k=1,\ldots,l_{ij}$

then

$$\eta = \sum_{i=1}^{n} \underline{\alpha}_i \rho_i^2 + \chi \varepsilon^2$$

is estimated by

$$\hat{\eta} = \sum_{i=1}^{n} \alpha_i \epsilon_{i.}.$$

If

$$\eta = \rho_i^2 + \varepsilon^2$$

we find

$$\hat{\eta} = \epsilon_{i.}.$$

Contrasts of the ρ_i^2 are estimated by corresponding contrasts of the $\epsilon_{i.}$. If both

$$\underset{\sim}{\omega}_j^k \equiv 0 \text{ for all } i=1,\dots,n \ ; \ j=1,\dots,m \ ; \ k=1,\dots,l_{ij}$$

$$\underset{\sim}{\eta}_i^k \equiv 0 \text{ for all } i=1,\dots,n \ ; \ j=1,\dots,m \ ; \ k=1,\dots,l_{ij}$$

then

$$\hat{\eta} = \not{l}\,\epsilon_{..}$$

estimates

$$\eta = \not{l}\,\varepsilon^2.$$

2.2 Estimation of parameters by SLS methods

We want to minimize

$$\underline{\eta} = \sum\sum\sum (x_{ij}^k - \delta - \lambda_i - \mu_j - \sum_{s=1}^{p} \alpha_{is}\beta_{js})^2$$

under the conditions that the parameters satisfy the identification constraints from the previous section. Any set of SLS-estimates of the parameters satisfies

$$\hat{\delta} = x^{\bullet}_{..} - \sum_{i=1}^{n} \hat{\lambda}_i \pi_{i.} - \sum_{j=1}^{m} \hat{\mu}_j \pi_{.j} - \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{ij} \sum_{s=1}^{p} \hat{\alpha}_{is}\hat{\beta}_{js},$$

$$\hat{\lambda}_i = x^{\bullet}_{i.} - \hat{\delta} - \sum_{j=1}^{m} \hat{\mu}_j \pi_{j|i} - \sum_{j=1}^{m} \pi_{j|i} \sum_{s=1}^{p} \hat{\alpha}_{is}\hat{\beta}_{js}, \text{ for all } i=1,\dots,n,$$

$$\hat{\mu}_j = x^{\bullet}_{.j} - \hat{\delta} - \sum_{i=1}^{n} \hat{\lambda}_i \pi_{i|j} - \sum_{i=1}^{n} \pi_{i|j} \sum_{s=1}^{p} \hat{\alpha}_{is}\hat{\beta}_{js}, \text{ for all } j=1,\dots,m,$$

and

$$\sum_{j=1}^{m} \pi_{ij} \left[x^{\bullet}_{ij} - (\hat{\delta} + \hat{\lambda}_i + \hat{\mu}_j + \sum_{s=1}^{p} \hat{\alpha}_{is}\hat{\beta}_{js})\right] \hat{\beta}_{jt} = 0 \text{ for all } i=1,\dots,n \ ; \ t=1,\dots$$

$$\sum_{i=1}^{n} \pi_{ij} \left[x^{\bullet}_{ij} - (\hat{\delta} + \hat{\lambda}_i + \hat{\mu}_j + \sum_{s=1}^{p} \hat{\alpha}_{is}\hat{\beta}_{js})\right] \hat{\alpha}_{it} = 0 \text{ for all } j=1,\dots,m \ ; \ t=1,\dots$$

In the PF-case these equations simplify considerably. Using the identification conditions we find

$$\hat{\delta} = x^{\bullet}_{\cdot\cdot},$$

$$\hat{\lambda}_i = x^{\bullet}_{i.} - x^{\bullet}_{\cdot\cdot},$$

$$\hat{\mu}_j = x^{\bullet}_{.j} - x^{\bullet}_{\cdot\cdot}.$$

To simplify the second set we define the matrix T by

$$\tau_{ij} = \pi^{\frac{1}{2}}_{i.}(x^{\bullet}_{ij} - x^{\bullet}_{i.} - x^{\bullet}_{.j} + x^{\bullet}_{\cdot\cdot})\pi^{\frac{1}{2}}_{.j},$$

and we define

$$\tilde{\alpha}_{is} = \pi^{\frac{1}{2}}_{i.}\hat{\alpha}_{is},$$

$$\tilde{\beta}_{js} = \pi^{\frac{1}{2}}_{.j}\hat{\beta}_{js}.$$

The stationary equations can now be rewritten in matrix form as

$$T\tilde{B} = \tilde{A}\hat{\Lambda},$$

$$T'\tilde{A} = \tilde{B}\hat{\Lambda},$$

where $\hat{\Lambda}$ is the diagonal matrix with the values $\hat{\gamma}_s$, and $\tilde{B}'\tilde{B} = \tilde{A}'\tilde{A} = I$, as required by the identification conditions. It follows that

$$T'T\tilde{B} = T'\tilde{A}\hat{\Lambda} = \tilde{B}\hat{\Lambda}^2,$$

$$TT'\tilde{A} = T\tilde{B}\hat{\Lambda} = \tilde{A}\hat{\Lambda}^2,$$

and consequently $\tilde{A}$ and $\tilde{B}$ are normalized eigenvectors of, respectively, TT' and T'T. The $\hat{\gamma}_s$ are the square roots of the corresponding eigenvalues. Assuming $m \leqslant n$ again we find for these values

$$\underline{\partial} = \sum_{s=p+1}^{m} \lambda_s(T'T),$$

and consequently the SLS-estimators correspond with the p largest eigenvalues.

2.3 Estimation of parameters by WLS methods

If

$$\underline{\partial} = \sum\sum \underline{\omega}_{ij} \sum (x^k_{ij} - \delta - \lambda_i - \mu_j - \sum_{s=1}^{p} \alpha_{is}\beta_{js})^2,$$

where the $\underline{\omega}_{ij} \geqslant 0$ are nonconstant weights, the situation becomes more complicated. The methods from the previous section can be applied only if $\underline{\omega}_{ij} = \underline{\omega}_{i\cdot}\underline{\omega}_{\cdot j}/\underline{\omega}_{\cdot\cdot}$, but this case is not very interesting. In this section we develop a more general method to minimize $\underline{\partial}$, which can of course also be used for SLS estimation in the non-PF case.

Define the vectors $\hat{y}$ and $\hat{z}$ by

$$(\hat{y})_i = \sum_{j=1}^{m} \pi_{ij} w_{ij}(x^{\bullet}_{ij} - \hat{\theta}_{ij}) / \sum_{j=1}^{m} \pi_{ij} w_{ij} - \hat{\tau},$$

$$(\hat{z})_j = \sum_{i=1}^{n} \pi_{ij} w_{ij}(x^{\bullet}_{ij} - \hat{\theta}_{ij}) / \sum_{i=1}^{n} \pi_{ij} w_{ij} - \hat{\tau},$$

where the number $\hat{\tau}$ is defined by

$$\hat{\tau} = \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{ij} w_{ij}(x^{\bullet}_{ij} - \hat{\theta}_{ij}) / \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{ij} w_{ij}.$$

Define the matrix

$$(P)_{ij} = \pi_{ij} w_{ij},$$

and the diagonal matrices

$$(Q)_{ii} = \sum_{j=1}^{m} \pi_{ij} w_{ij},$$

$$(R)_{jj} = \sum_{i=1}^{n} \pi_{ij} w_{ij}.$$

We identify the λ_i and μ_j by

$$\sum_{i=1}^{n} (Q)_{ii} \lambda_i = \sum_{j=1}^{m} (R)_{jj} \mu_j = 0.$$

Then the stationary equations are

$$\hat{\delta} = \hat{\tau},$$

$$\hat{y} = \hat{\lambda} + Q^{-1} P \hat{\mu},$$

$$\hat{z} = \hat{\mu} + R^{-1} P' \hat{\lambda}.$$

Let

$$\tilde{\mu} = R^{\frac{1}{2}} \hat{\mu},$$

$$\tilde{\lambda} = Q^{\frac{1}{2}} \hat{\lambda},$$

$$\tilde{y} = Q^{\frac{1}{2}} \hat{y},$$

$$\tilde{z} = R^{\frac{1}{2}} \hat{z},$$

$$\tilde{P} = Q^{-\frac{1}{2}} P R^{-\frac{1}{2}},$$

then

$$\tilde{y} = \tilde{\lambda} + \tilde{P} \tilde{\mu},$$

$$\tilde{z} = \tilde{\mu} + \tilde{P}' \tilde{\lambda},$$

or

$(I - \tilde{P}\tilde{P}')\tilde{\lambda} = \tilde{y} - \tilde{P}\tilde{z},$

$(I - \hat{P}'\hat{P})\tilde{\mu} = \tilde{z} - \hat{P}'\hat{y}.$

If $\hat{\alpha}_{is}$ and $\hat{\beta}_{js}$ (and thus $\hat{\partial}_{ij}$) are known we can compute the corresponding $\hat{\lambda}_i$, $\hat{\mu}_j$, and $\hat{\delta}$ using these formula's. As a next step define the vectors $\hat{a}_i$ and $\hat{b}_j$ by

$$(\hat{a}_i)_j = (\hat{b}_j)_i = w_{ij}\pi_{ij}(x^{\cdot}_{ij} - (\hat{\delta} + \hat{\lambda}_i + \hat{\mu}_j)),$$

and the matrices $G^{(i)}$ and $H^{(j)}$ by

$$g^{(j)}_{st} = \sum_{j=1}^{m} w_{ij}\pi_{ij}\hat{\beta}_{js}\hat{\beta}_{jt},$$

$$h^{(j)}_{st} = \sum_{i=1}^{n} w_{ij}\pi_{ij}\hat{\alpha}_{is}\hat{\alpha}_{it}.$$

Assuming that the matrices are nonsingular, we can write the stationary equations as

$$\hat{\alpha}_i = \left[G^{(i)}\right]^{-1}\hat{B}'\hat{a}_i,$$

$$\hat{\beta}_j = \left[H^{(j)}\right]^{-1}\hat{A}'\hat{b}_j.$$

This suggests a simple iterative process to compute $\hat{A}$ and $\hat{B}$ if $\hat{\lambda}_i$, $\hat{\mu}_j$, and $\hat{\delta}$ a known, which is closely related to the power method. The complete computational procedure is now obvious. We minimize $\underset{\sim}{\ell}$ over α_{is} and β_{js} for fixed $\hat{\delta}$, $\hat{\lambda}_i$, an $\hat{\mu}_j$ by the iterative procedure we just mentioned. Reasonable initial estimates o δ, λ_i, and μ_j are $x^{\cdot}_{..}$, $x^{\cdot}_{i.} - x^{\cdot}_{..}$, $x^{\cdot}_{.j} - x^{\cdot}_{..}$. The next step is to minimize $\underset{\sim}{\ell}$ o δ, λ_i, and μ_j for the current values of $\hat{\alpha}_{is}$ and $\hat{\beta}_{js}$, and so on.

2.4 Alternating WLS and MINQUE methods

Both WLS and SLS give consistent estimates (if $l_{ij} \to \infty$ for all $i=1,\ldots,n$; $j=1,\ldots$ of the parameters. Thus

$$\hat{\epsilon}^k_{ij} = x^k_{ij} - \hat{\delta} - \hat{\lambda}_i - \hat{\mu}_j - \sum_{s=1}^{p}\hat{\alpha}_{is}\hat{\beta}_{js} \xrightarrow{\mathcal{L}} \eta(0, \rho^2_i + \partial^2_j + \varepsilon^2),$$

and the MINQUE methods can be applied to the LS residuals. This suggest starting with SLS, apply MINQUE to the residuals to estimate the variances, use the variances to compute WLS estimates, use the new residuals to find new MINQUE estimates of the variances, and so on. Convergence and distribution properties of the resulting sequence of estimates are unknown, their theoretical investig seems extremely complicated. Of course in each cycle the MINQUE estimates of the variances are unbiased, the WLS estimates of the parameters are CAN. ~~Decreased~~

2.5 Estimation of parameters and variances by ML methods

If we want to minimize

$$\underline{\mathcal{L}} = \sum\sum \frac{\sum_k (x_{ij}^k - \delta - \lambda_i - \mu_j - \sum_{s=1}^{p} \alpha_{is}\beta_{js})^2}{\rho_i^2 + \zeta_j^2 + \xi^2} + \sum\sum l_{ij} \ln (\rho_i^2 + \zeta_j^2 + \xi^2$$

over both parameters and variances the situation becomes even more complicat
It is easy enough to computed derivatives, but simple examples (for example when $l_{ij} = 1$ for all i,j) show that the LF is unbounded and that the station equations are not sufficient conditions for a minimum. The general formula f derivatives is

$$g(\underline{\theta}) = \frac{\partial \mathcal{L}}{\partial \underline{\theta}} = \sum\sum \frac{l_{ij}\hat{\omega}_{ij} - \hat{\tau}_{ij}}{\hat{\omega}_{ij}^2} \frac{\partial \hat{\omega}_{ij}}{\partial \underline{\theta}},$$

$$V(\underline{\theta}, \overline{\theta}) = \frac{\partial^2 \mathcal{L}}{\partial \underline{\theta} \partial \overline{\theta}} = \sum\sum \frac{2\hat{\tau}_{ij} - l_{ij}\hat{\omega}_{ij}}{\hat{\omega}_{ij}^3} \frac{\partial \hat{\omega}_{ij}}{\partial \underline{\theta}} \frac{\partial \hat{\omega}_{ij}}{\partial \overline{\theta}},$$

with

$$\sum (\hat{\epsilon}_{ij}^k)^2 = \hat{\tau}_{ij} = \sum (x_{ij}^k - \hat{\delta} - \hat{\lambda}_i - \hat{\mu}_j - \sum \hat{\alpha}_{is}\hat{\beta}_{js})^2,$$

$$\hat{\omega}_{ij} = \hat{\rho}_i^2 + \hat{\zeta}_j^2 + \hat{\xi}^2.$$

If for all $i=1,\dots,n; j=1,\dots,m$ iid($\hat{\epsilon}_{ij}^1, \dots, \hat{\epsilon}_{ij}^{l_{ij}}$) with expectations zero and variances $\omega_{ij} < \infty$, then

$$\frac{\hat{\tau}_{ij}}{l_{ij}} \xrightarrow{a.s} \omega_{ij} \geq \tfrac{1}{2}\omega_{ij}.$$

Consequently the matrix V is psd with probability tending to unity. Obviously

$$E(g(\underline{\theta})) = 0,$$

$$E(V(\underline{\theta}, \overline{\theta})) = \sum\sum \frac{l_{ij}}{\hat{\omega}_{ij}^2} \frac{\partial \hat{\omega}_{ij}}{\partial \underline{\theta}} \frac{\partial \hat{\omega}_{ij}}{\partial \overline{\theta}}.$$

If, in addition,

$$\hat{\epsilon}_{ij}^k \sim \mathcal{N}(0, \omega_{ij})$$

then

$$\hat{\tau}_{ij}/\omega_{ij} \sim \chi^2(l_{ij}),$$

and

$$\text{prob}\left\{\text{PSD}(V)\right\} \geq \prod\prod \text{prob}\left\{\chi^2(l_{ij}) \geq \tfrac{1}{2} l_{ij}\right\} .$$

If $l_{ij} \to \infty$ for all i,j the right side tends to unity again. For the second partials we find the expressions

$$V(\rho^2_{k'}, \rho^2_{k'}) = \delta^{kk'} \sum_{j=1}^{m} \frac{2\hat{\tau}_{kj} - l_{kj}\hat{\omega}_{kj}}{\hat{\omega}^3_{kj}} ,$$

$$V(\lambda^2_{l'}, \lambda^2_{l'}) = \delta^{ll'} \sum_{i=1}^{n} \frac{2\hat{\tau}_{il} - l_{il}\hat{\omega}_{il}}{\hat{\omega}^3_{il}} ,$$

$$V((\varepsilon^2)^2) = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{2\hat{\tau}_{ij} - l_{ij}\hat{\omega}_{ij}}{\hat{\omega}^3_{ij}} ,$$

$$V(\rho^2_{k'}, \omega^2_{l}) = \frac{2\hat{\tau}_{kl} - l_{kl}\hat{\omega}_{kl}}{\hat{\omega}^3_{kl}} ,$$

$$V(\rho^2_{k'}, \varepsilon^2) = \sum_{j=1}^{m} \frac{2\hat{\tau}_{kj} - l_{kj}\hat{\omega}_{kj}}{\hat{\omega}^3_{kj}} ,$$

$$V(\lambda^2_{l'}, \varepsilon^2) = \sum_{i=1}^{n} \frac{2\hat{\tau}_{il} - l_{il}\hat{\omega}_{il}}{\hat{\omega}^3_{il}} .$$

If the l_{ij} are such that the maximum likelihood (or likelihood equation) estimates have desirable asymptotic properties, then these partials can be used to solve for the variances, given the parameters. This can be alternated with solving for the parameters, given the variances (for example by WLS). This shows the desirability of obtaining replications in factor analytic situations (or, equivalently, of grouping the subjects and/or variables according to a priori defined criteria).

3 Historical

The history of models of this form is quite complicated. In factor analysis the model

$$\underset{\sim}{x}_{ij} = \lambda_j + \alpha_i \beta_j + \underset{\sim}{\epsilon}_{ij}$$

was proposed by Young (1941). He derived estimates under the assumption that

$$\underset{\sim}{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma^2).$$

Lawley (1942) tried to generalize this to

$$\underset{\sim}{x}_{ij} = \lambda_j + \sum_{s=1}^{p} \alpha_{is} \beta_{js} + \underset{\sim}{\epsilon}_{ij},$$

with

$$\underset{\sim}{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma_j^2).$$

It was pointed out by Anderson and Rubin (1956, p 130) that Lawley's procedur were invalid. They also showed that it was possible to estimate some of the parameters of the model efficiently, but by using quite different methods. Whittle (1952) and Lawley (1953) made the obvious step of estimating the mode

$$\underset{\sim}{x}_{ij} = \lambda_j + \sum_{s=1}^{p} \alpha_{is} \beta_{js} + \underset{\sim}{\epsilon}_{ij},$$

with

$$\underset{\sim}{\epsilon}_{ij} \sim \mathcal{N}(0, \sigma^2).$$

It is clear that these factor analytic models differ from our model in two important respects. In the first place replications within cells are not take into account, in the second place there is an essential assymmetry between th two modes of the design (tests and subjects, for example). In model B from section 0.2 this asymmetry is natural, but in model A it is considerably less natural. Of course within our models asymmetry can be introduced by setting subsets of the parameters equal to zero.

In the ANOVA area the first model in this direction seems to be due to Tukey (1949). He studied

$$\underset{\sim}{x}_{ij} = \mu_i + \lambda_j + \theta \mu_i \lambda_j + \underset{\sim}{\epsilon}_{ij}.$$

Compare also Ward and Dick (1952) for generalizations to incomplete designs. Tukey's model was generalized by Mandel (1961) to

$$\underset{\sim}{\chi}_{ij} = \mu_i + \lambda_j + \theta_i \lambda_j + \underset{\sim}{\varepsilon}_{ij}.$$

Finally Williams (1952), Gollob (1968), Mandel (1971, 1972), Corsten and Eynsbergen (1972), Johnson and Graybill (1972) generalized (with varying degrees of expliciteness, sophistication, completeness, and erroneousness) to

$$\underset{\sim}{\chi}_{ij} = \mu_i + \lambda_j + \alpha_i \beta_j + \underset{\sim}{\varepsilon}_{ij},$$

or

$$\underset{\sim}{\chi}_{ij} = \mu_i + \lambda_j + \sum_{s=1}^{p} \alpha_{is} \beta_{js} + \underset{\sim}{\varepsilon}_{ij}.$$

All these ANOVA models assume that

$$\varepsilon_{ij} \sim \mathcal{N}(0, \delta^2).$$

The relationship between these models and the 'vacuum cleaner' of Tukey (1962 has been discussed by Mandel (1971, 1972), and Linssen (1972). The first one to connect factor analysis and ANOVA with multiplicative decomposition of the interaction was Gollob (1968). Of course there have been earlier contribution by Burt and his school to the comparison of these two techniques, but they compared factor analysis (and essentially model B) with variance component analysis (this approach is generalized in the currently popular analysis of covariance structures).

Although additive decomposition of both means and variances seems quite natur from a computational point of view there is a lot to be said for the models of Bechhofer (1960) who supposes

$$\underset{\sim}{\varepsilon}_{ij} \sim \mathcal{N}(0, \varepsilon^2 \rho_i^2 \delta_j^2).$$

It is easy to seen that the computations are simplified considerably, but I don't know of any situation where this model can be naturally applied to psychometric data.

The MINQUE method to estimate variances is due to Rao (1970, 1972 a,b).

Young, G.	Maximum likelihood estimation and factor analy Psychometrika, 1941, 6, 49-53
Lawley, D.N.	Further investigations in factor estimation Proc. Roy. Soc. Edinburgh, 1942, 61, 176-185
Anderson, T.W. & Rubin, H.	Statistical inference in factor analysis. Proc. III Berkeley Symp, volume V, 111-150.
Whittle, P.	On principal component and leastsquares metho of factor analysis. Skandinavisk Aktuarietidskrift, 1952, 35, 223-
Lawley, D.N.	A modified method of estimation in factor anal sis and some large sample results. Uppsala Symp. Psychol. factor analysis. Stockholm, Almqvist-Wicksell, 1953.
Tukey, J.W.	One degree of freedom for non-additivity Biometrika, 1949, 15, 232-242.
Ward, G.C. & Dick, I.D.	Non-additivity in randomized block designs an balanced incomplete block designs. New Zealand J. of Science and Technology, 195 33, 430-436.
Mandel, J.	Non-additivity in analysis of variance. J. American Statist. Assoc., 1961, 56, 878-88
Williams, E.J.	The interpretation of interactions in factori experiments. Biometrika, 39, 1952, 65-81.
Gollob, H.F.	A statistical model which combines features o factor analytic and analysis of variance techniques. Psychometrika, 1968, 33, 73-116.
Mandel, J.	A new analysis of variance model for nonaddit data. Technometrics, 13, 1971, 1-18.
Mandel, J.	Principal components, analysis of variance, and data structure. Statistica Neerlandica, 26, 1972, 119-129.
Corsten, L.C.A. & Eynsbergen, A.C. van	Multiplicative effects in two-way analysis of variance. Statistica Neerlandica, 26, 1972, 61-68.
Johnson, D.E. & Graybill, F.A.	Analysis of a two-way model with interaction and no replication. J. American Statist. Assoc., 1972, 67, 862-
Tukey, J.W.	The future of data analysis. Ann. Math. Statist. , 1962, 33, 1-67.
Linssen, H.N.	The vacuum cleaner and attachment. Statistica Neerlandica, 26, 1972, 203-210.

Bechhofer, R.E. A multiplicative model for analyzing variances which are affected by several factors.
J. Amer. Statist. Assoc., 1960, 55, 245-

Rao, C.R. Estimation of heteroscedastic variances linear models.
J. Amer. Statist. Assoc., 65,1970,161-17

Rao, C.R. Estimation of variance and covariance components in linear models.
J. Amer. Statist. Assoc.,67, 1972, 112-1

Rao, C.R. Minimum variance quadratic unbiased estimation of variance components.
J. Multivariate Analysis, 1, 1972,