John Ashworth Nelder 1924–2010
Contributions to Statistics

R. A. Bailey

Queen Mary
University of London

r.a.bailey@qmul.ac.uk

58th World Statistics Congress of the
International Statistical Institute,
Dublin, August 2011

Early life

- Born in 1924.
- Grew up on the edge of Exmoor, exploring the local countryside, and became a life-long birdwatcher.
- Studied Mathematics at Sidney Sussex College, Cambridge, 1942–43.
- RAF navigator in World War II, 1943–1946.
- Returned to Cambridge, graduated (with first class honours) in 1948, followed by a diploma in Mathematical Statistics in 1949.
- 1949–1968: (National) Vegetable Research Station, Wellesbourne.

National Vegetable Research Station

- Negative components of variance
- Nelder–Mead simplex algorithm
- The idea that there should be a general framework for describing designed experiments and analysing data from them,
  rather than a collection of standard designs, each with its own recipe for analysis.

1954 paper on negative components of variance

J. A. Nelder:
The interpretation of negative components of variance.
Biometrika **41** (1954), pp. 544–548.

Random block effects: Model I

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)} \qquad \text{where} \quad f(\omega) = \text{treatment on plot } \omega.$$

Random block effects: Model I

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)} \qquad \text{where} \quad f(\omega) = \text{treatment on plot } \omega.$$

There is a random effect for each plot (variance σ^2),
and a random effect for each block (variance σ_B^2),
and these are all independent.

Random block effects: Model I

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)} \qquad \text{where} \quad f(\omega) = \text{treatment on plot } \omega.$$

There is a random effect for each plot (variance σ^2),
and a random effect for each block (variance σ_B^2),
and these are all independent. Then

$$\mathrm{Cov}(Y) = \sigma^2 I + \sigma_B^2 B,$$

$$\text{where} \quad B_{\alpha,\omega} = \begin{cases} 1 & \text{if } \alpha \text{ and } \omega \text{ are in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

Random block effects: Model I

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)} \qquad \text{where} \quad f(\omega) = \text{treatment on plot } \omega.$$

There is a random effect for each plot (variance σ^2),
and a random effect for each block (variance σ_B^2),
and these are all independent. Then

$$\mathrm{Cov}(Y) = \sigma^2 I + \sigma_B^2 B,$$

$$\text{where} \quad B_{\alpha,\omega} = \begin{cases} 1 & \text{if } \alpha \text{ and } \omega \text{ are in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

so

$$\begin{aligned} \mathrm{Cov}(Y) &= \sigma^2 \left(I - \frac{1}{k} B \right) + (k\sigma_B^2 + \sigma^2) \left(\frac{1}{k} B \right), \\ &= \xi_0 \left(I - \frac{1}{k} B \right) + \xi_1 \left(\frac{1}{k} B \right), \end{aligned}$$

with $0 \le \xi_0 \le \xi_1$.

Random block effects: Model II

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)}$$

Random block effects: Model II

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)}$$

$$\mathrm{Cov}(Y_\alpha, Y_\omega) = \begin{cases} \sigma^2 & \text{if } \alpha = \omega \\ \gamma & \text{if } \alpha \neq \omega \text{ but } \alpha \text{ and } \omega \text{ in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

Random block effects: Model II

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)}$$

$$\mathrm{Cov}(Y_\alpha, Y_\omega) = \begin{cases} \sigma^2 & \text{if } \alpha = \omega \\ \gamma & \text{if } \alpha \neq \omega \text{ but } \alpha \text{ and } \omega \text{ in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

so

$$\begin{aligned} \mathrm{Cov}(Y) &= \sigma^2 I + \gamma(B - I) \\ &= (\sigma^2 - \gamma)\left(I - \frac{1}{k}B\right) + ((k-1)\gamma) + \sigma^2)\left(\frac{1}{k}B\right), \\ &= \xi_0\left(I - \frac{1}{k}B\right) + \xi_1\left(\frac{1}{k}B\right), \end{aligned}$$

with $0 \leq \xi_0$ and $0 \leq \xi_1$.

Random block effects: Model II

$$\mathbb{E}(Y_\omega) = \tau_{f(\omega)}$$

$$\mathrm{Cov}(Y_\alpha, Y_\omega) = \begin{cases} \sigma^2 & \text{if } \alpha = \omega \\ \gamma & \text{if } \alpha \neq \omega \text{ but } \alpha \text{ and } \omega \text{ in the same block} \\ 0 & \text{otherwise,} \end{cases}$$

so

$$\begin{aligned} \mathrm{Cov}(Y) &= \sigma^2 I + \gamma(B - I) \\ &= (\sigma^2 - \gamma)\left(I - \frac{1}{k}B\right) + ((k-1)\gamma) + \sigma^2)\left(\frac{1}{k}B\right), \\ &= \xi_0\left(I - \frac{1}{k}B\right) + \xi_1\left(\frac{1}{k}B\right), \end{aligned}$$

with $0 \leq \xi_0$ and $0 \leq \xi_1$. (Weaker assumption than $0 \leq \xi_0 \leq \xi_1$.)

1954 paper: negative components of variance

In some experiments, it is reasonable that responses on plots in the same block will be negatively correlated, making

$$\text{blocks stratum variance } \xi_1 < \xi_0 \text{ plots stratum variance.}$$

1954 paper: negative components of variance

In some experiments, it is reasonable that responses on plots in the same block will be negatively correlated, making

$$\text{blocks stratum variance } \xi_1 < \xi_0 \text{ plots stratum variance.}$$

This implies that the component of variance σ_B^2 from Model I is negative.

1954 paper: negative components of variance

In some experiments, it is reasonable that responses on plots in the same block will be negatively correlated, making

$$\text{blocks stratum variance } \xi_1 < \xi_0 \text{ plots stratum variance.}$$

This implies that the component of variance σ_B^2 from Model I is negative.

Some statistical software does not allow you to estimate ξ_1 to be smaller than ξ_0.

1965 paper with Mead

J. A. Nelder & R. Mead:
A simplex method for function minimization.
Computer Journal **7** (1965), pp. 303–333.

1965 paper with Mead

J. A. Nelder & R. Mead:
A simplex method for function minimization.
Computer Journal **7** (1965), pp. 303–333.

This is still called “the Nelder–Mead simplex algorithm”,
and is still widely used.

1965 Royal Society papers

J. A. Nelder:
The analysis of randomized experiments with orthogonal block structure.
I. Block structure and the null analysis of variance.
Proceedings of the Royal Society of London, Series A **283** (1965) pp. 147–162.

J. A. Nelder:
The analysis of randomized experiments with orthogonal block structure.
II. Treatment structure and the general analysis of variance.
Proceedings of the Royal Society of London, Series A **283** (1965) pp. 163–178.

1965 Royal Society papers: first important idea

Distinguish the set of experimental units from the set of treatments, and think about structure on each before you think of which treatment to put on which experimental unit.

- building on approach of F. Yates
  but explicitly much more general
- subtly, and importantly, different from approach of
  O. Kempthorne
- still not widely appreciated in North America.

1965 Royal Society papers: first important idea

Distinguish the set of experimental units from the set of treatments, and think about structure on each before you think of which treatment to put on which experimental unit.

- building on approach of F. Yates
  but explicitly much more general
- subtly, and importantly, different from approach of
  O. Kempthorne
- still not widely appreciated in North America.

Example

There are 3 fields, each containing 10 plots.
10 varieties of wheat are planted on the plots, in a randomized complete-block design.
The experimental units are the 30 plots, not the 30 field-variety combinations.

1965 Royal Society papers: second important idea

The most common structure,
on either experimental units or treatments, is
simple orthogonal block structure,
made by iterated crossing and nesting.

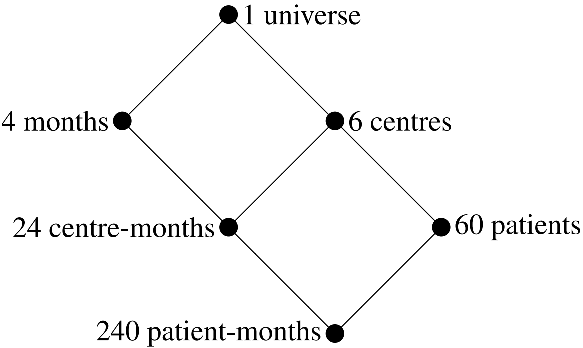
- established quite general theory
- building on ideas of G. Wilkinson
- still the basis of much statistical software.

Example

$$(6 \text{ centres}/10 \text{ patients}) * (4 \text{ months})$$

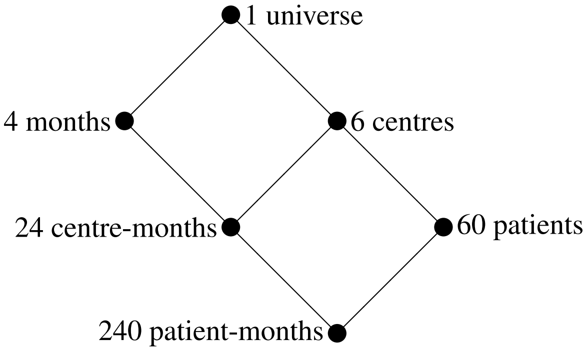
1965 Royal Society papers: randomization

$$(6 \text{ centres}/10 \text{ patients}) * (4 \text{ months})$$



1965 Royal Society papers: randomization

$$(6 \text{ centres}/10 \text{ patients}) * (4 \text{ months})$$



- Randomly permute the labels of the 6 centres;
- within each centre independently, randomly permute the labels of the 10 patients;
- randomly permute the labels of the 4 months.

1965 Royal Society papers: covariance structure

$$(6 \text{ centres}/10 \text{ patients}) * (4 \text{ months})$$

This randomization justifies the covariance model which is essentially scalar within each of these strata:

Between centres	5 df
Between patients within centres	54 df
Between months	3 df
Between centre-months within centres and months	15 df
Between units within all the above	162 df

1965 Royal Society papers: third important idea

A useful relationship between the structure on the experimental units and the structure on the treatments is **general balance**;
this depends on the design as well as on the two structures.

- extended the ideas of A. James and G. Wilkinson
- still not widely understood.

1965 Royal Society papers: third important idea

A useful relationship between the structure on the experimental units and the structure on the treatments is **general balance**;
this depends on the design as well as on the two structures.

- extended the ideas of A. James and G. Wilkinson
- still not widely understood.

$$\mathbb{E}(Y) = \sum T_i \beta_i \qquad \text{for known orthogonal idempotent matrices } T_i\ ;$$
$$\mathrm{Cov}(Y) = \sum \xi_i Q_i \qquad \text{for known orthogonal idempotent matrices } Q_i\ ;$$

there are constants λ_{ij} such that

$$T_i Q_j T_{i'} = \begin{cases} \lambda_{ij} T_i & \text{if } i = i' \\ 0 & \text{otherwise.} \end{cases}$$

Waite Agricultural Institute, Adelaide

Visit to Adelaide

John Nelder spent 1965–1966 at the Waite Agricultural Institute in Adelaide.

Here he worked with Graham Wilkinson, and the foundations for the statistical software GenStat were laid.

J. A. Nelder, Head of Statistics Department, Rothamsted

Broadbalk, Rothamsted

Rothamsted, 1968–1984

As head of department, he encouraged

- genuine collaboration with Rothamsted scientists
- the protocol for an experiment must include a dummy data analysis
- careful data input and routine analyses, with everything checked
- development of GenStat
- development of relevant statistical theory.

GenStat

- Separate `blockstructure` and `treatmentstructure`.
- Syntax for crossing and nesting.
- Possibliity to randomize according to a formula with crossing and nesting.
- A general algorithm, covering a large class of generally balanced designs.
- Users allowed (indeed encouraged) to write their own procedures.

Marginality

Twelve treatments are all combinations of:

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

Marginality

Twelve treatments are all combinations of:

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

$$E(Y_\omega) = \tau_{C(\omega),F(\omega)}$$

$$E(Y_\omega) = \lambda_{C(\omega)} + \mu_{F(\omega)}$$

$$E(Y_\omega) = \lambda_{C(\omega)} \qquad E(Y_\omega) = \mu_{F(\omega)}$$

$$E(Y_\omega) = \kappa$$

$$E(Y_\omega) = 0$$

Marginality

Twelve treatments are all combinations of:

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

$$
\begin{tikzcd}[column sep=small]
 & E(Y_\omega) = \tau_{C(\omega),F(\omega)} \arrow[d, dash, "\leftarrow\text{Interaction}"] & \\
 & E(Y_\omega) = \lambda_{C(\omega)} + \mu_{F(\omega)} \arrow[dl, dash] \arrow[dr, dash] & \\
E(Y_\omega) = \lambda_{C(\omega)} \arrow[dr, dash] & & E(Y_\omega) = \mu_{F(\omega)} \arrow[dl, dash] \\
 & E(Y_\omega) = \kappa \arrow[d, dash] & \\
 & E(Y_\omega) = 0 &
\end{tikzcd}
$$

Marginality

Twelve treatments are all combinations of:

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

$$E(Y_\omega) = \tau_{C(\omega),F(\omega)}$$

←Interaction

$$E(Y_\omega) = \lambda_{C(\omega)} + \mu_{F(\omega)}$$

$$E(Y_\omega) = \lambda_{C(\omega)} \qquad E(Y_\omega) = \mu_{F(\omega)}$$

$$E(Y_\omega) = \kappa$$

$$E(Y_\omega) = 0$$

Nelder always argued that the full model is marginal to the additive model, so that there is no sense in fitting an 'interaction' without main effects.

Marginality

Twelve treatments are all combinations of:

factor	levels
Cultivar (C)	Cropper, Melle, Melba
Fertilizer (F)	0, 80, 160, 240 kg/ha

$$
\begin{tikzcd}
 & E(Y_\omega)=\tau_{C(\omega),F(\omega)} \arrow[d, dash, "\leftarrow\text{Interaction}"] & \\
 & E(Y_\omega)=\lambda_{C(\omega)}+\mu_{F(\omega)} \arrow[dl, dash] \arrow[dr, dash] & \\
E(Y_\omega)=\lambda_{C(\omega)} \arrow[dr, dash] & & E(Y_\omega)=\mu_{F(\omega)} \arrow[dl, dash] \\
 & E(Y_\omega)=\kappa \arrow[d, dash] & \\
 & E(Y_\omega)=0 &
\end{tikzcd}
$$

Nelder always argued that the full model is marginal to the additive model, so that there is no sense in fitting an ‘interaction’ without main effects. Many in N. America disagreed, especially creators of statistical software, but recently the ‘strong heredity effect principle’ has begun to be recognized.

Towards more general models

$$
\begin{aligned}
\mathbb{E}(Y) &= \sum T_i\beta_i && \text{for known orthogonal idempotent matrices } T_i \text{ ;} \\
\mathrm{Cov}(Y) &= \sum \xi_i Q_i && \text{for known orthogonal idempotent matrices } Q_i \text{ ;}
\end{aligned}
$$

the distribution of Y is approximately normal;

and the covariance does not depend on the expectation.

Towards more general models

$$\mathbb{E}(Y) = \sum T_i \beta_i \qquad \text{for known orthogonal idempotent matrices } T_i \text{ ;}$$
$$\mathrm{Cov}(Y) = \sum \xi_i Q_i \qquad \text{for known orthogonal idempotent matrices } Q_i \text{ ;}$$

the distribution of Y is approximately normal;

and the covariance does not depend on the expectation.

What should we do if Y is Poisson, binomial, etc?

Towards more general models

$$\mathbb{E}(Y) = \sum T_i \beta_i \qquad \text{for known orthogonal idempotent matrices } T_i \text{ ;}$$
$$\mathrm{Cov}(Y) = \sum \xi_i Q_i \qquad \text{for known orthogonal idempotent matrices } Q_i \text{ ;}$$

the distribution of Y is approximately normal;

and the covariance does not depend on the expectation.

What should we do if Y is Poisson, binomial, etc?
Previously, statisticians transformed their data to make it approximately normal, but this did not get round the problem of covariance defined by the same parameter(s) as the mean.

1972 paper on generalized linear models

J. A. Nelder and R. W. M. Wedderburn:
Generalized linear models.
Journal of the Royal Statistical Society, Series A **135** (1972),
pp. 370–384.

1972 paper on generalized linear models

J. A. Nelder and R. W. M. Wedderburn:
Generalized linear models.
Journal of the Royal Statistical Society, Series A **135** (1972), pp. 370–384.

- The variables Y_i are independent, with distributions from an exponential family, not necessarily normal.
- $\mathbb{E}(Y_i) = g^{-1}(\eta_i)$, where η is an unknown linear combination of known covariates and the link function g is usually non-linear.

1972 paper on generalized linear models

J. A. Nelder and R. W. M. Wedderburn:
Generalized linear models.
Journal of the Royal Statistical Society, Series A **135** (1972),
pp. 370–384.

- The variables Y_i are independent, with distributions from an exponential family, not necessarily normal.
- $\mathbb{E}(Y_i) = g^{-1}(\eta_i)$, where η is an unknown linear combination of known covariates and the link function g is usually non-linear.

Parameters are estimated by maximum likelihood using iteratively reweighted least squares.

1972 paper on generalized linear models

J. A. Nelder and R. W. M. Wedderburn:
Generalized linear models.
Journal of the Royal Statistical Society, Series A **135** (1972), pp. 370–384.

- The variables Y_i are independent, with distributions from an exponential family, not necessarily normal.
- $\mathbb{E}(Y_i) = g^{-1}(\eta_i)$, where η is an unknown linear combination of known covariates and the link function g is usually non-linear.

Parameters are estimated by maximum likelihood using iteratively reweighted least squares.

(Wedderburn died in 1975 following a bee-sting.)

Generalized linear models

Software GLIM was developed for fitting generalized linear models, and eventually incorporated within GenStat.

Generalized linear models

Software GLIM was developed for fitting generalized linear models, and eventually incorporated within GenStat.

P. McCullagh & J. A. Nelder: *Generalized Linear Models*. Chapman and Hall, 1983 (second edition 1989).

Rothamsted, 1968–1984

- Awarded the Guy Medal in Silver by the Royal Statistical Society in 1971.
- President of the International Biometric Society 1978–1979.
- Elected a Fellow of the Royal Society on 19 March 1981.
- Retired at age 60 in 1984.

Imperial College, London, 1972–2009

- Visiting professor from 1972.
- From 1984, he considered Imperial College as his base, commuting by train from Harpenden about 4 days per week.
- President of the Royal Statistical Society 1985–1986.
- Developed GLIMPSE, an expert system for interactive analysis of data.
- Collaboration with Youngjo Lee, who initially did not realise that Nelder had officially retired!
- Awarded the Guy Medal in Gold by the Royal Statistical Society in 2005.
- Retired for the second time, October 2009.

Hierarchical generalized linear models

Generalized linear models with extra random effects.

Hierarchical generalized linear models

Generalized linear models with extra random effects.

Y. Lee & J. A. Nelder:
Hierarchical generalized linear models (with discussion).
Journal of the Royal Statistical Society, Series B **58** (1996), pp. 619–678.

Y. Lee & J. A. Nelder:
Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions.
Biometrika **88** (2001), pp. 987–1006.

Y. Lee & J. A. Nelder:
Double hierarchical generalized linear models (with discussion).
Applied Statistics **55** (2006), pp. 139–185.

Y. Lee, J. A. Nelder &Y. Pawitan:
Generalized Liner Models with Random Effects: Unified Analysis via H-likelihood.
CRC Press, London, 2006.

J. A. Nelder's influence on: R. A. Bailey

He had sufficient confidence in me that he appointed me to a position in the Statistics Department at Rothamsted in 1981,
even though the position was initially explicitly linked to A.D.A.S. (Agricultural Development Advisory Service)
and my background had originally been as a pure mathematician.

J. A. Nelder's influence on: R. A. Bailey

He had sufficient confidence in me that he appointed me to a position in the Statistics Department at Rothamsted in 1981,
even though the position was initially explicitly linked to A.D.A.S. (Agricultural Development Advisory Service)
and my background had originally been as a pure mathematician.

- scrutinze data before analysing it;
- the reality of agricultural field trials;
- (with Praeger, Rowley and Speed) proved his randomization claim for simple orthgonal block structures;
- generalized simple orthgonal block structures to poset block structures;
- emphasis on the collection of expectation models rather than a single equation in which some coefficients may be zero.

J. A. Nelder's influence on: C. J. Brien

"John Nelder inspired me from my beginnings in the statistical profession. ... His ability to synthesize a statistical topic into a coherent theory [was] without peer. That he replicated this several times is amazing."

J. A. Nelder's influence on: C. J. Brien

"John Nelder inspired me from my beginnings in the statistical profession. … His ability to synthesize a statistical topic into a coherent theory [was] without peer. That he replicated this several times is amazing."

- generalized the idea of needing two structures
  (such as plots and treatments)
  to needing three or more
  (such as field-phase plots, lab-phase runs, and treatments).

J. A. Nelder's influence on: R. W. Payne

- Nelder's co-twitcher;
- took over as chief of GenStat after Nelder's retirement from Rothamsted, and eventually took GenStat to the company VSN. Although it has expanded into many other areas, its core is still the analysis based on Nelder's 1965 Royal Society papers, which no other software does as well.

Acknowledgements

Obituary in the *Guardian* by Roger Payne and Stephen Senn, 23 September 2010.

Obituary in *Journal of the Royal Statistical Society, Series A* **174** (2011) by Peter McCullagh, Bob Gilchrist and Roger Payne.

Obituary and condolences on web page of VSN International.

John Nelder at the VSN conference in 2004

John at the Rothamsted Conference in 2004