Generalized linear models I
Linear models

Peter McCullagh

Department of Statistics
University of Chicago

Polokwane, South Africa
November 2013

Outline

Exchangeability, covariates, relationships,..

Linear Gaussian models

Qualitative factors

Factorial subspaces

Linear projections

Model formulae

What is a statistical model?

Ans: Family of probability distributions on the observation space

The observation space:

- Usually a set of statistical units (the population)
- Observation recorded on a finite subset of size n
- Observation space for one unit is $\mathcal{S}$
- Observation space for sample is $\mathcal{S}^n$

Standard examples of observation spaces:

- $\mathcal{S} = \{0, 1\}$ or $\mathbb{Z} = \{0, 1, 2, \ldots\}$ or $\mathbb{R}$
- $\mathcal{S} = \mathbb{R}^k$ (multivariate response)
- $\mathcal{S} = \{z \in \mathbb{R}^2 : \|z\|^2 = 1\}$
- $\mathcal{S} = \{\text{Brand X}, \text{Brand Y}, \text{Brand Z}, \text{Other}\}$
- $\mathcal{S} = \mathrm{seq}(\mathcal{R}) = \bigcup_{r \geq 0} \mathcal{R}^r$ (survival studies)

Restrictions on distributions

Structure among the units:

Labels $i \mapsto u_i$: names of sampled units (patients)
covariate: $i \mapsto x_i$, e.g. sex, age, treat, variety
Relationship: $(i, j) \mapsto V(i, j)$: $\quad V(i, i) = V(j, j)$
Ménages à trois: $(i, j, k) \mapsto W(i, j, k)$

Exchangeability: Rules of the game

1. If $x_i = x_j$ then $Y_i \sim Y_j$.
   (one-dimensional distributions only)
2. If $V(i, j) = V(k, l)$ then $(Y_i, Y_j) \sim (Y_k, Y_l)$
   provided that $(x_i, x_j) = (x_k, x_l)$.
   (two-dimensional distributions only)
3. Similarly for 3- and 4-way relationships
4. Principle: Differences in distribution, marginal or joint, are associated with specific inhomogeneities in the experimental material (the units)

Restrictions on distributions

Structure among the units:

Labels $i \mapsto u_i$: names of sampled units (patients)
covariate: $i \mapsto x_i$, e.g. sex, age, treat, variety
Relationship: $(i, j) \mapsto V(i, j)$: $\quad V(i, i) = V(j, j)$
Ménages à trois: $(i, j, k) \mapsto W(i, j, k)$

Exchangeability: Rules of the game

1. If $x_i = x_j$ then $Y_i \sim Y_j$.
   (one-dimensional distributions only)
2. If $V(i, j) = V(k, l)$ then $(Y_i, Y_j) \sim (Y_k, Y_l)$
   provided that $(x_i, x_j) = (x_k, x_l)$.
   (two-dimensional distributions only)
3. Similarly for 3- and 4-way relationships
4. Principle: Differences in distribution, marginal or joint, are associated with specific inhomogeneities in the experimental material (the units)

Restrictions on distributions

Structure among the units:

Labels $i \mapsto u_i$: names of sampled units (patients)
covariate: $i \mapsto x_i$, e.g. sex, age, treat, variety
Relationship: $(i, j) \mapsto V(i, j)$: $\quad V(i, i) = V(j, j)$
Ménages à trois: $(i, j, k) \mapsto W(i, j, k)$

Exchangeability: Rules of the game

1. If $x_i = x_j$ then $Y_i \sim Y_j$.
   (one-dimensional distributions only)
2. If $V(i, j) = V(k, l)$ then $(Y_i, Y_j) \sim (Y_k, Y_l)$
   provided that $(x_i, x_j) = (x_k, x_l)$.
   (two-dimensional distributions only)
3. Similarly for 3- and 4-way relationships
4. Principle: Differences in distribution, marginal or joint, are associated with specific inhomogeneities in the experimental material (the units)

Restrictions on distributions

Structure among the units:

Labels $i \mapsto u_i$: names of sampled units (patients)
covariate: $i \mapsto x_i$, e.g. sex, age, treat, variety
Relationship: $(i, j) \mapsto V(i, j)$: $\quad V(i, i) = V(j, j)$
Ménages à trois: $(i, j, k) \mapsto W(i, j, k)$

Exchangeability: Rules of the game

1. If $x_i = x_j$ then $Y_i \sim Y_j$.
   (one-dimensional distributions only)
2. If $V(i, j) = V(k, l)$ then $(Y_i, Y_j) \sim (Y_k, Y_l)$
   provided that $(x_i, x_j) = (x_k, x_l)$.
   (two-dimensional distributions only)
3. Similarly for 3- and 4-way relationships
4. Principle: Differences in distribution, marginal or joint, are associated with specific inhomogeneities in the experimental material (the units)

Linear Gaussian families

Observation space $\mathcal{S}^n = \mathbb{R}^n$
Gaussian distribution $N_n(\mu, \Sigma)$ specified by

- mean vector $\mu \in \mathcal{R}^n$
- covariance matrix $\Sigma \in \mathrm{PD}_n$

Typical linear Gaussian model (family of distns)

- Mean-value subspace $\mathcal{X} \subset \mathbb{R}^n$ spanned by covariate vectors $\mathcal{X} = \{X\beta : \beta \in \mathcal{R}^p\}$
- Covariance cone $\mathcal{V} \subset \mathrm{PD}_n$ spanned by relationships e.g. $\mathcal{V} = \{\tau I_n : \tau > 0\}$ or $\mathcal{V} = \{\tau_0 I_n + \tau_1 V_1 + \tau_2 V_2\}$

Linear Gaussian families

Observation space $\mathcal{S}^n = \mathbb{R}^n$
Gaussian distribution $N_n(\mu, \Sigma)$ specified by

- mean vector $\mu \in \mathcal{R}^n$
- covariance matrix $\Sigma \in \mathrm{PD}_n$

Typical linear Gaussian model (family of distns)

- Mean-value subspace $\mathcal{X} \subset \mathbb{R}^n$ spanned by covariate vectors $\mathcal{X} = \{X\beta : \beta \in \mathcal{R}^p\}$
- Covariance cone $\mathcal{V} \subset \mathrm{PD}_n$ spanned by relationships e.g. $\mathcal{V} = \{\tau I_n : \tau > 0\}$ or $\mathcal{V} = \{\tau_0 I_n + \tau_1 V_1 + \tau_2 V_2\}$

Linear Gaussian families

Observation space $\mathcal{S}^n = \mathbb{R}^n$
Gaussian distribution $N_n(\mu, \Sigma)$ specified by

- mean vector $\mu \in \mathcal{R}^n$
- covariance matrix $\Sigma \in \mathrm{PD}_n$

Typical linear Gaussian model (family of distns)

- Mean-value subspace $\mathcal{X} \subset \mathbb{R}^n$ spanned by covariate vectors $\mathcal{X} = \{X\beta : \beta \in \mathcal{R}^p\}$
- Covariance cone $\mathcal{V} \subset \mathrm{PD}_n$ spanned by relationships e.g. $\mathcal{V} = \{\tau I_n : \tau > 0\}$ or $\mathcal{V} = \{\tau_0 I_n + \tau_1 V_1 + \tau_2 V_2\}$

Linear Gaussian families

Observation space $\mathcal{S}^n = \mathbb{R}^n$
Gaussian distribution $N_n(\mu, \Sigma)$ specified by

- mean vector $\mu \in \mathcal{R}^n$
- covariance matrix $\Sigma \in \mathrm{PD}_n$

Typical linear Gaussian model (family of distns)

- Mean-value subspace $\mathcal{X} \subset \mathbb{R}^n$ spanned by covariate vectors $\mathcal{X} = \{X\beta : \beta \in \mathcal{R}^p\}$
- Covariance cone $\mathcal{V} \subset \mathrm{PD}_n$ spanned by relationships e.g. $\mathcal{V} = \{\tau I_n : \tau > 0\}$ or $\mathcal{V} = \{\tau_0 I_n + \tau_1 V_1 + \tau_2 V_2\}$

Qualitative factor as a list of levels

Qualitative covariates:

variety: Kerr's Pink, King Edward, Yukon Gold, Lumper, Idaho,...
ethnicity: Xhosa, Zulu, Sotho, European,...
genotype: AA, Aa, aa
occupation: BLS classification
treatment: control, active$_1$, active$_2$,...
dose: none, low, medium, high
ship type: tanker, ore carrier, ferry, container ship,..

Some distinctions:

Set of levels may be finite or infinite
may be ordered or unordered or partially ordered
may be tree-structured (types and sub-types)
may contain a residue class "none of the above"

Classification factors, treatment factor, block factor

Qualitative factor as a binary matrix

Operationally A is a list of levels (not numbers)

one level for each unit one column in a spreadsheet

$$A = (\text{KP, YG, KP, KE, LP, YG, KP, LP}, \ldots)$$

$$A = \begin{array}{ccccc} \text{KP} & \text{KE} & \text{YG} & \text{LP} & \text{ID} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}$$

Indicator matrix A of order $n \times k$

Row sums: $A_{i,\bullet} = 1$ for every i

Col sums: $A_{\bullet,l}$ number of occcurrences of l in sample

Empty column: value does not occur in sample

Qualitative factor as a subspace

A trinity of representations:

$A = (\text{KP, YG, KP, KE, LP, YG, KP, LP}, \ldots)$ (*list of levels*)

$A(i)$ is the level of A on unit i

A is an *indicator matrix* of order $n \times k$

$A_{i,l} = 1$ if $A(i) = l$; zero otherwise

$A \subset \mathcal{R}^n$ is also a *subspace* of $\mathcal{R}^n$

span of colunms of indicator matrix

The vector $\mathbf{1} \equiv \mathbf{1}_n$ (all ones)

The subspace $\mathbf{1} \subset \mathcal{R}^n$ (constant functions)

$A_{i\bullet} = 1$ implies $\mathbf{1} \subset A$

Multiplication: $A^2 = A$

Definition:
A *variable* v is a real-valued *function* on the units:

$$i \mapsto v(i) \equiv v_i$$

Consequence: the product vw of two variables is a variable:
$(vw)(i) = v(i)w(i)$
Also $v^2, \exp(v), \log(v), \ldots$ are variables

The space $\mathcal{R}^n$ is the space of [all] variables
technically a commutative ring

For a *factor* A, the subspace $A \subset \mathcal{R}^n$ is a sub-ring
closed under multiplication of functions
$v, w \in A$ implies $vw \in A$
$A^2 = A$

Several factors

Experimental design

plot	101	102	103	104	105	106	107	108	109	110	111	112
A	KP	YG	KP	KE	LP	YG	KP	LP	YG	KE	LP	KE
B	CCL	CCL	CCL	BDM	BDM	BDM	CCL	CCL	CCL	BDM	BDM	BDM
y	40.3	33.5	41.2	40.3	41.5	31.8	38.4	43.8	28.7	44.5	40.0	38.0

Associated subspaces in $\mathbb{R}^{12}$:

const functions **1**; dimension 1

Variety A; levels KP, KE, LP, YG; dimension 4

Blight treatment B; 2 levels; dimension 2

span $A + B$; dimension $4 + 2 - 1 = 5$

Interaction: $A.B \equiv A : B \equiv A * B$ is also a *factor*

Levels of $A.B$: $\{\mathrm{KP}, \mathrm{KE}, \mathrm{LP}, \mathrm{YG}\} \times \{\mathrm{CCL}, \mathrm{BDM}\}$
Sample levels: (KP,C), (KE, B), (LP, B), (LP,C), (YG,B), (YG, C)
$\dim(A.B) = 6$ (could be up to 8)

Linear projections in Euclidean space

x, y are points in $\mathbb{R}^n$:

Inner product: $\langle x, y \rangle = x' W y$ is symmetric bi-linear

inner product matrix W is symmetric pos def

$\langle x, x \rangle = \|x\|^2$: squared Euclidean length

Orthogonality: $x \perp y$ iff $\langle x, y \rangle = 0$

$\mathcal{X} \subset \mathbb{R}^n$ is a subspace with basis vectors X (given)

P is the *orthogonal projection* onto $\mathcal{X}$

matrix representation: $P = X(X'WX)^{-1}X'W$

$\hat{y} = Py = X(X'WX)^{-1}X'Wy = X\hat{\beta}$

Coefficients of $\hat{y}$ with respect to given basis in $\mathcal{X}$:

$\hat{\beta} = (X'WX)^{-1}X'Wy$

Euclidean geometry:

orthogonality: $\langle y - Py, Py \rangle = \langle y - \hat{y}, \hat{y} \rangle = 0$

Pythagoras: $\|y\|^2 = \|Py\|^2 + \|y - Py\|^2$

Computation of projections

On the assumption that $W = \text{diag}\{w_1, \ldots, w_n\}$:

$$\langle x, y \rangle = x' W y$$

$\mathcal{X} = \text{span}\{x_1, \ldots, x_p\} \subset \mathcal{R}^n$

```
fit <- lm(y~x1+...+xp, weights=w)
```

Orthogonal projection $\hat{y} = Py$ in `fit$fitted`

Complementary projection $Qy = y - \hat{y} = (I - P)y$ in `fit$res`

Pythagoras theorem:

$$\langle Py, Qy \rangle = 0$$

$$\langle y, y \rangle = \langle Py, Py \rangle + \langle Qy, Qy \rangle$$

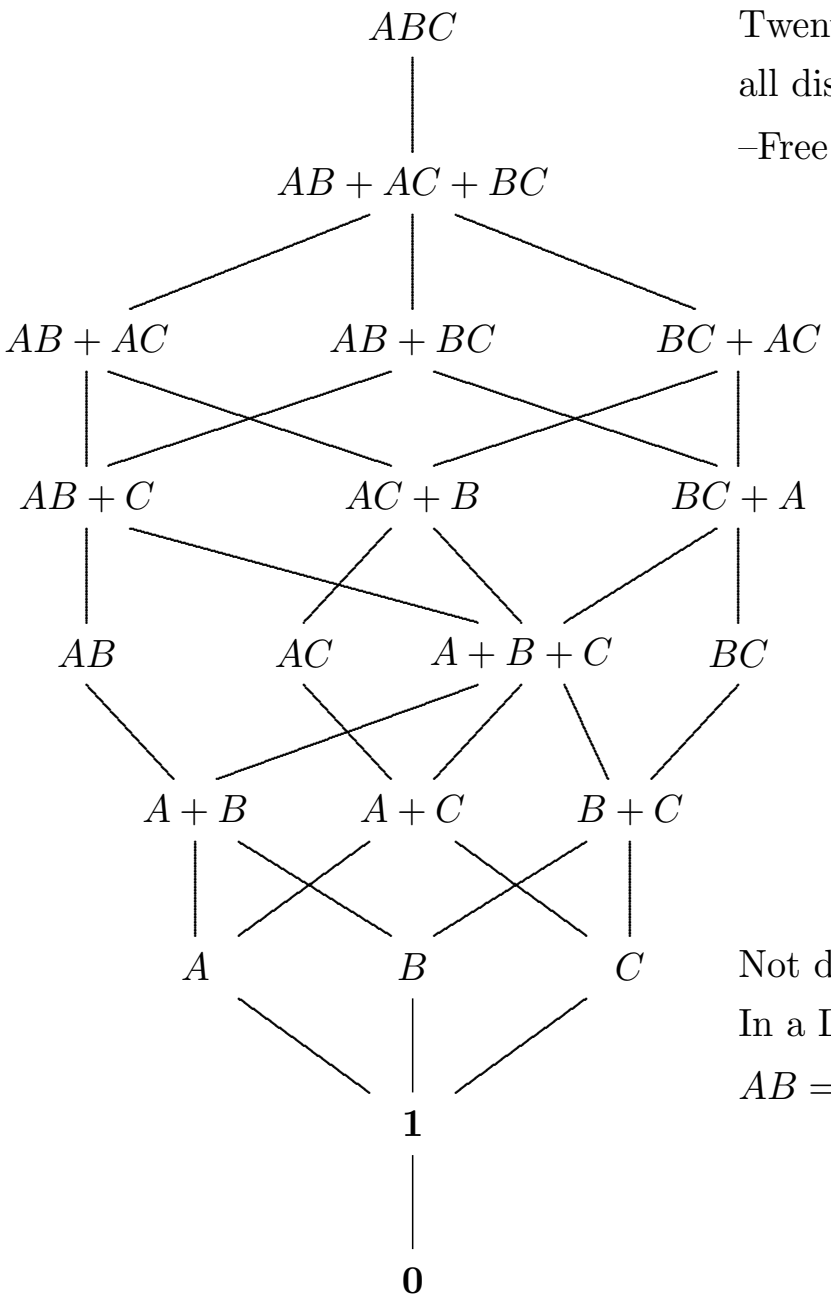
ANOVA decomposition for factorial subspaces

Hasse diagram of factorial subspaces
with dimensions and mean squares for this design

$$\begin{tikzcd}[column sep=small]
\|y\|^2 \\
= 6.41 \times 6 \\
+ 4.69 \times 1 \\
+ 1.84 \times 1 \\
+ 70.1 \times 3 \\
+ 17778
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
 & \mathcal{R}^n(12) \arrow[d, dash, "6.41"] & \\
 & A.B(6) \arrow[d, dash, "4.69"] & \\
 & A+B(5) \arrow[dl, dash, "1.84"'] \arrow[dr, dash, "67.8"] & \\
A(4) \arrow[dr, dash, "70.1"'] & & B(2) \arrow[dl, dash, "8.67"] \\
 & \mathbf{1} \arrow[d, dash, "n\bar{y}^2 = 17787"'] & \\
 & \mathbf{0} &
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=small]
\|y\|^2 \\
= 6.41 \times 6 \\
+ 4.69 \times 1 \\
+ 67.8 \times 3 \\
+ 8.67 \times 1 \\
+ 17778
\end{tikzcd}$$

F-ratios:
Non-additivity: $F_{1,6} = 4.69/6.41$
Variety elim Trt: $F_{3,6} = 67.8/6.41$
Trt elim Variety: $F_{1,6} = 1.84/6.41$

Hasse diagram for three factors



Twenty factorial subspaces

all distinct in a full factorial design

–Free distributive lattice

Not distinct in general

In a Latin square design

$AB = AC = BC$ is maximal

Model formulae

A, B are factors; X is a covariate
$B(i)$ is the level of factor B for unit i

$E(Y_i)$	Model formula	Alternative
$\alpha_{A(i)}$	A	$A - 1$
$\alpha + \beta x_i$	X	$1 + x$
βx_i	$X - 1$	through origin
$\alpha_{A(i)} + \beta x$	$A + X$	
$\alpha + \beta_{A(i)} x$	$A{:}X$	$1 + A{:}X$
$\alpha_{A(i)} + \beta_{A(i)} x$	$A + A{:}X$	$A * X$
$\alpha_{A(i)} + \beta_{B(i)}$	$A + B$	
$\alpha_{A(i)} + \beta_{B(i)} + \gamma x_i$	$A + B + X$	
$\alpha_{A(i),B(i)} + \gamma x_i$	$A{:}B + X$	$A * B + X$

Usually, the first level of each factor is taken as the reference level, meaning that the coefficient is set to zero
Alternative versions do not affect the subspace, but may affect the parameterization (choice of basis)

Interpretation of model formulae

Terms: A, B, C, $X1$, $X2$,...
Compound terms $A.B \equiv B.A, \quad A.C \equiv C.A, \quad A.X$,...
Each term $A, X, A.B \equiv A{:}B \equiv A * B$ is a vector subspace
  $\dim(X) = 1$ (but **1** included by default)
  $\dim(A) = \#\{\text{levels of } A \text{ in design}\}$
  $\dim(A.B) = \#\{\text{ordered pairs in design}\}$
  $A \subset A.B$ and $B \subset A.B$ regardless of parameterization
  $+$ means vector span $A + B, \quad A + X$,...

Examples:
$A * B \equiv B * A$ same basis vectors but different order
$A{:}B \equiv A * B$ same subspace, different basis
$A * B \equiv A + B + A{:}B$ includes **1** and uses reference levels
$A{:}B$ uses indicator vector for each ordered pair

```
lm(y ~ A:B-1)$coef
```
versus
```
tapply(y, list(A,B), mean)
```

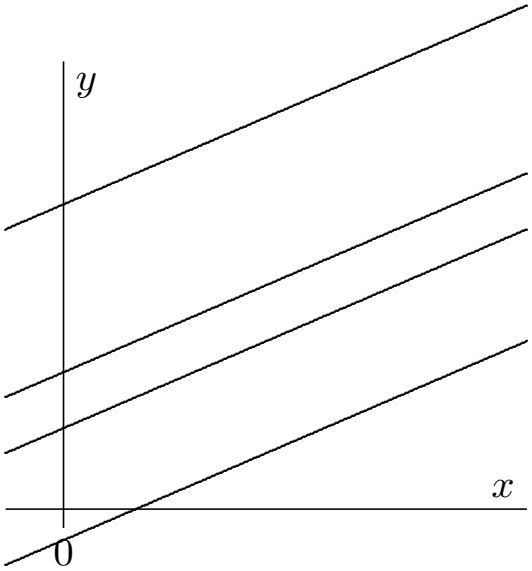
Beware the reference level!

Interpretation of model formulae

Models $A + X$, $A{:}X \equiv 1 + A{:}X$ and $A * X$

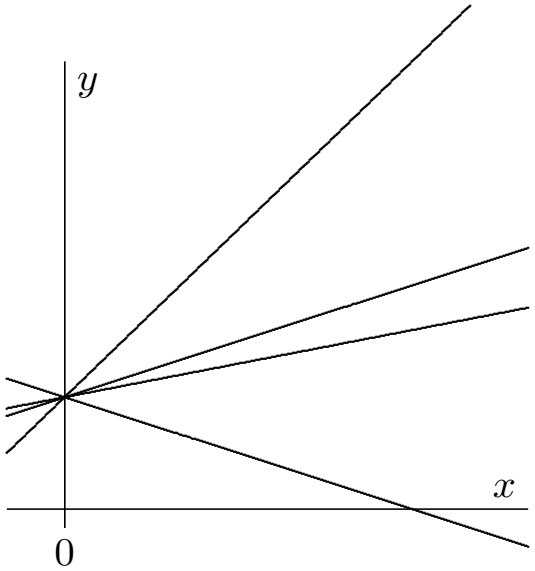
`y ~ A + X`

$\alpha_{A(i)} + \beta x_i$



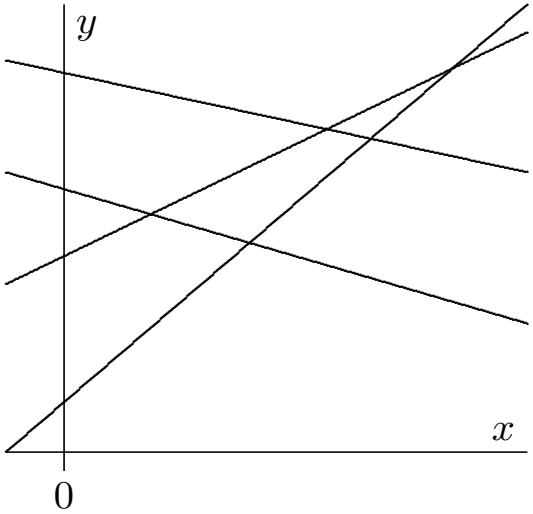
`y ~ A:X`

$\alpha + \beta_{A(i)} x_i$



`y ~ A + A:X`

$\alpha_{A(i)} + \beta_{A(i)} x_i$



Example: decay of ascorbic acid

Ascorbic acid concentrations of snap-beans after cold storage

	Weeks of storage				
Temp	2	4	6	8	*Total*
0°F	45	47	46	46	184
10°F	45	43	41	37	166
20°F	34	28	21	16	99

Transform: $Y = \log(\text{concentration})$ — why?

M0: `y ~as.factor(temp):time` RSS0=0.613 on 8df

M1: `y ~as.factor(temp)*time` RSS1=0.391 on 6df

Mean square ratio: $F_{2,6} = (0.222/2)/(0.391/6) = 1.70$

`pf(1.7, 2, 6) = 0.74;` `qf(0.95, 2, 6) = 5.14`

—M0: mean concentrations equal at time zero!

Further routine techniques for linear models

- Polynomial bases for factorial contrasts
- Non-additivity, non-linearity; Tukey's 1DOFNA

```
fit0 <- lm(y~ ...);   fv2 <- fit0$fitted^2
fit1 <- lm(~ ...+fv2)
rms1 <- sum(fit1$res^2)/fit1$res.df
F <- sum(fit0$res-fit1$res)^2 / rms1
T <- (fit1$coef / sqrt(diag(vcov(fit1))))
```

- Transformation for interaction removal
  Box-Cox transformation
- Confidence intervals for non-linear functions $1/\beta$
  Half-life = $\log(2)/\beta$ in exponential decay model
- Confidence intervals for parameter ratios
  $-\beta_0/\beta_1$ is x-intercept
  $-\beta_1/(2\beta_2)$ is stationary point of a quadratic
  Fieller method,...

Further routine techniques for linear models

- Polynomial bases for factorial contrasts
- Non-additivity, non-linearity; Tukey's 1DOFNA

```
fit0 <- lm(y~ ...);   fv2 <- fit0$fitted^2
fit1 <- lm(~ ...+fv2)
rms1 <- sum(fit1$res^2)/fit1$res.df
F <- sum(fit0$res-fit1$res)^2 / rms1
T <- (fit1$coef / sqrt(diag(vcov(fit1))))
```

- Transformation for interaction removal
  Box-Cox transformation
- Confidence intervals for non-linear functions $1/\beta$
  Half-life = $\log(2)/\beta$ in exponential decay model
- Confidence intervals for parameter ratios
  $-\beta_0/\beta_1$ is x-intercept
  $-\beta_1/(2\beta_2)$ is stationary point of a quadratic
  Fieller method,...

Further routine techniques for linear models

- Polynomial bases for factorial contrasts
- Non-additivity, non-linearity; Tukey's 1DOFNA

```
fit0 <- lm(y~ ...);   fv2 <- fit0$fitted^2
fit1 <- lm(~ ...+fv2)
rms1 <- sum(fit1$res^2)/fit1$res.df
F <- sum(fit0$res-fit1$res)^2 / rms1
T <- (fit1$coef / sqrt(diag(vcov(fit1))))
```

- Transformation for interaction removal
  Box-Cox transformation
- Confidence intervals for non-linear functions $1/\beta$
  Half-life = $\log(2)/\beta$ in exponential decay model
- Confidence intervals for parameter ratios
  $-\beta_0/\beta_1$ is x-intercept
  $-\beta_1/(2\beta_2)$ is stationary point of a quadratic
  Fieller method,...

Further routine techniques for linear models

- Polynomial bases for factorial contrasts
- Non-additivity, non-linearity; Tukey's 1DOFNA

```
fit0 <- lm(y~ ...);   fv2 <- fit0$fitted^2
fit1 <- lm(~ ...+fv2)
rms1 <- sum(fit1$res^2)/fit1$res.df
F <- sum(fit0$res-fit1$res)^2 / rms1
T <- (fit1$coef / sqrt(diag(vcov(fit1))))
```

- Transformation for interaction removal
  Box-Cox transformation
- Confidence intervals for non-linear functions $1/\beta$
  Half-life = $\log(2)/\beta$ in exponential decay model
- Confidence intervals for parameter ratios
  $-\beta_0/\beta_1$ is x-intercept
  $-\beta_1/(2\beta_2)$ is stationary point of a quadratic
  Fieller method,...

Further routine techniques for linear models

- Polynomial bases for factorial contrasts
- Non-additivity, non-linearity; Tukey's 1DOFNA

```
fit0 <- lm(y~ ...);   fv2 <- fit0$fitted^2
fit1 <- lm(~ ...+fv2)
rms1 <- sum(fit1$res^2)/fit1$res.df
F <- sum(fit0$res-fit1$res)^2 / rms1
T <- (fit1$coef / sqrt(diag(vcov(fit1))))
```

- Transformation for interaction removal
  Box-Cox transformation
- Confidence intervals for non-linear functions $1/\beta$
  Half-life = $\log(2)/\beta$ in exponential decay model
- Confidence intervals for parameter ratios
  $-\beta_0/\beta_1$ is x-intercept
  $-\beta_1/(2\beta_2)$ is stationary point of a quadratic
  Fieller method,...