Statistics 215a – 10/20/03 – D.R. Brillinger

Exploratory analysis of variance.

Multiway array data, $y_{ijk\ldots}$, i=1,…,I; j=1,…,J; k=1,…,K; …

Paradigm

response = fit + residual

becomes

response = main effect + two-factor effect + three-factor effect +…+ residual

Some decompositions

$y_i = m + r_I$

$= \bar{y} + (y_i - \bar{y})$

$y_{ij} = m + a_i + r_{ij}$

$y_{ij} = m + a_i + b_j + r_{ij}$

$y_{ijk} = m + a_i + b_j + r_{ijk}$

$y_{ijk} = m + a_i + b_j + (ab)_{ij} + r_{ijk} =$

$\bar{y}_{\dots} + (\bar{y}_{i..} - \bar{y}_{\dots}) + (\bar{y}_{.j.} - \bar{y}_{\dots}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{\dots}) + (y_{ijk} - \bar{y}_{ij.})$

(value splitting)

$y_{ijk} = m + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} +$

$(bc)_{jk} + r_{ijk}$

cp. Taylor expansion of function of several variables

y = **Xb** + **r**

X made up of 0’s and 1’s

dummy variables

side conditions

experiment

designed study

often comparative

treatment

things being compared

experimental unit

smallest division of the experimental material such that any 2 units may receive different treatments

factor

a basic treatment

levels

possible forms of a factor

factorial experiment

all factor combinations are of interest

data: multi-way array

purpose: comparison, optimization, …

cell

particular combination of factor levels

main effects: a_i, b_j, c_k, …

interaction

departure from additivity

two-factor interactions (effects): $(ab)_{ij}$, $(ac)_{ik}$, $(bc)_{jk}$, …

three-factor interactions: …

observational study

latin square

graeco-roman latin square

replication

runs carried out several times

sweeping

overlays

same size and shape as original table

recover original data by adding all overlays

ANOVA table

Source	SS	DF	MS=SS/DF
Mean	$\Sigma_{ijk}\ (\bar{y}_{\dots})^2$	1	
Rows	$\Sigma_{ijk}\ (\bar{y}_{i..}-\bar{y}_{\dots})^2$	(I-1)	
columns	$\Sigma_{ijk}\ (\bar{y}_{.j.}-\bar{y}_{\dots})^2$	(J-1)	
error	$\Sigma_{ijk}\ (\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{\dots})^2$	(I-1)(J-1)	
total	$\Sigma_{ijk}\ y_{ijk}^2$	IJK	

Other

transformation

Diagnostics

residuals

block

collection of units expected to be homogeneous

estimates more precise

nesting

pooling

robust/resistant methods

Statistics 215a – 10/20/03 – D.R. Brillinger

Diet experiment.

The questions

Effectiveness on cattle of 6 protein feeding treatments?

Which largest gain?

Additive level and type effects?

Experimental unit

rat

Factor A levels

beef, pork, cereal

Factor B levels

high, low

Response

weight gain (gm)

Replications
[important]
K = 10

The data (Snedecor and Cochran, 1980)

	Beef	pork	cereal
H	118	120	111
L	95	106	107
H	117	108	98
L	90	97	98
H	111	105	95
L	90	86	97
H	107	102	92
L	90	82	95
H	104	102	88
L	86	82	89
H	102	98	86
L	78	81	80
H	100	96	82
L	76	73	74
H	87	94	77
L	72	70	74
H	81	91	74
L	64	61	67
H	73	79	56
L	51	49	58

Full factorial

A1 B1 A2 B1 A3 B1

A1 B2 A2 B2 A3 B2

[parallel boxplots]

Fit

$y_{ijk} = m + a_i + b_j + (ab)_{ij} + r_{ijk}$

$m = \bar{y}_{\dots} =$ 87.9 gm

$\{a_i\} = \{\bar{y}_{i..} - \bar{y}_{\dots}\} =$ 1.73, 1.23, -2.96

$\{b_j\} = \{\bar{y}_{.j.} - \bar{y}_{\dots}\} =$ 7.27, -7.27
[amount having larger effect]

$\{(ab)_{ij}\}$

3.13 3.13 -6.26

-3.13 -3.13 6.26
[larger than type values]

[side-by-side plot, effects small wrt residuals]

$\{r_{ijk}\}$

N = 60 Median = 2.2
Quartiles = -8.8, 10.8

Decimal point is 1 place to the right of the colon

```
  -3 : 00
  -2 : 876
  -2 : 1
  -1 : 9875
  -1 : 3200
  -0 : 999766
  -0 : 444321
   0 : 0022222334
   0 : 55677789
   1 : 11111234
   1 : 6788
   2 : 03
   2 : 57
```

Conclusions:

High-protein beef seems best

Interactions

Overlays

$\{\bar{y}_{\dots}\}$

	beef	pork	cereal
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9
H	87.9	87.9	87.9
L	87.9	87.9	87.9

$\{\bar{y}_{i..} - \bar{y}_{...}\}$

	beef	pork	cereal
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3
H	4.9	4.4	-9.2
L	-1.4	-1.9	3.3

$\{\bar{y}_{.j.} - \bar{y}_{...}\}$

	beef	pork	cereal
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0
H	10.4	10.4	1.0
L	-10.4	-10.4	-1.0

$$\{\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\}$$

	beef	pork	cereal
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3
H	3.1	3.1	-6.3
L	-3.1	-3.1	6.3

$\{y_{ijk} - \bar{y}_{ij.}\}$

beef	pork	cereal

ANOVA table

Source	SS	DF	MS=SS/DF
mean	463233	1	463233.1
columns	266.53	2	133.27
rows	3168.27	1	3168.27
interaction	1178.13	2	589.01
error	11586.00	54	215.56
total	479432	60	

OLS

```
Coefficients:
 (Intercept)  $A1  $A2   $B   $A1$B $A2$B
    87.86     1.73 1.23 7.26   3.13  3.13


Degrees of freedom: 60 total; 54 residual
Residual standard error: 14.64772
```

Robust/resistant

```
Coefficients:
 (Intercept)  $A1   $A2  $B   $A1$B  $A2$B
    88.39     2.16 0.85 7.24   3.12   3.00
```

Degrees of freedom: 60 total; 54 residual
Scale estimate: 14.3

Examine final weights in search for outliers and patterns