STA442F   Oct 20/09

1. Final Exam Dec 16 2–5 pm
   Beatty In Course Scholarships

2. HW #2 due Nov. 10.   (HW 3 due Dec 1)

3. Designed experiments – quantitative & qualitative factors

<u>Example</u> Table 8.2

3 factors

seat height	26 / 30 "	−	+	quant
tire pressure	40 / 55 lbs / sq"	−	+	quant
dynamo	on off	+	−	qual

(analyzed in § 9.3, p. 444)

<u>Example</u> J x_1 = length, x_2 = amplitude x_3 = load
3 quantitative factors, each at 3 levels

<u>Example</u> oneway layout (HO from last day)

		50	75	100	125	150
factor: seedrate	levels	1	2	3	4	5

response # grains / head
of barley   5 obs$^=$ per level

<u>Example</u> one way layout Table 9.3   45 – 5 groups of 9
A B C D E

$$y_{tr} = \beta_t + \varepsilon_{tr} \qquad t = 1, \ldots, 5 \qquad \varepsilon_{tr} \sim (0, \sigma^2)$$

$$r = 1, \ldots, 9 \qquad \beta_t \stackrel{=}{} E(y_{tr})$$

OR $\alpha + \gamma'_t + \varepsilon_{tr}$ $E(y_{tr}) = \alpha + \gamma'_t$ (over-parametrized)

If factor is qualitative (or treated as), then analysis of variance takes the form

Source	SS	df	MS
betw. groups	$R\sum_t (\bar{y}_{t\cdot} - \bar{y}_{\cdot\cdot})^2$	$T-1$	ratio
within groups	$\sum_{tr} (y_{tr} - \bar{y}_{t\cdot})^2$	$T(R-1)$	
total (corr)		$RT-1$	

Eg. on HO, mid p. 2

Source	df	SS	MS	F	p
seedrate	4	35.057	8.764	148.17	10^{-14}
resids	20	1.183	0.059		

Note: 1st fit has $\hat{\alpha} = 21.0930$ $\qquad E y_{1r} = \alpha \qquad \hat{\alpha} = \bar{y}_{1\cdot}$

$\gamma_1 = 0 \qquad \hat{\gamma}_1 = 0 \qquad \hat{\gamma}_2 = -1.0624 \qquad E y_{2r} = \alpha + \gamma_2 \qquad \hat{\gamma}_2 = \bar{y}_{2\cdot} - \bar{y}_{1\cdot}$

$\vdots$

$\hat{\gamma}_5 = -3.3514$

2nd fit has $\hat{\alpha} = 19.27954$

$\Sigma \gamma_t = 0 \qquad \hat{\gamma}_1 = 1.81341$

$\vdots$

$\hat{\gamma}_4 = -0.87837$

$\hat{\gamma}_5 = -(\hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3 + \hat{\gamma}_4 + \hat{\gamma}_5)$

Now we know groups are different, what next?

1) compare treat avg's of particular interest

2) Partition ~~diff~~ treatment SS. - use contrasts

which to use depends on application.

1) Example teaching methods A B C D E

usual usual praise reprove ignore

compare $\bar{y}_A$ to $\bar{y}_B$ $\quad \dfrac{\bar{y}_A - \bar{y}_B}{\{\hat{\mathrm{var}}(\bar{y}_A - \bar{y}_B)\}^{1/2}} \sim t$

$\mathrm{var}(\bar{y}_A) = \dfrac{\sigma^2}{9} \quad \mathrm{var}\,\bar{y}_B = \dfrac{\sigma^2}{9} \quad \mathrm{var}(\mathrm{diff}) = \dfrac{2\sigma^2}{9}$

$\tilde{\sigma}^2 =$ residual mean sq.

compare $\bar{y}_C$ to $\frac{1}{2}(\bar{y}_A + \bar{y}_B)$

2) Example sedatives numerical factor ∴ polynomial regression seems appropriate

- this can be done using linear regression

$$y_{tr} = \beta_0 + \beta_1 x_{tr} + \beta_2 x_{tr}^2 + \beta_3 x_{tr}^3 + \beta_4 x_{tr}^4 + \varepsilon_{tr}$$

- or by orthogonal polynomial regression

$$y_{tr} = \gamma_0 + \gamma_1 P_1(x_{tr}) + \gamma_2 P_2(x_{tr}) + \cdots + \gamma_4 P_4(x_{tr}) + \varepsilon_{tr}$$

where $P_j(x) = a_{0j} + a_{1j} x + \cdots + a_{jj} x^j$, $\quad a_{0j}, a_{1j}, a_{jj}$ chosen

so that in the data, columns are $\perp$

- this is easily fitted for ~~factors~~ or ordered factors by anova or by lm

& the contrast matrix can be extracted using contrasts (orderedfactor)

- see last pages : note $F = t^2$

3) Example 2^3 factorial experiment cycling data

		df	SS	MS
2x2x2	tmt	7	560.92	80.13
~~2x2x2~~	resid	8	33.50	4.19
		15		

partition according to 3 factors & their interactions

Table 9.16

		SS	df	F
tmt	seat	473.06	1	112.9
	dynamo	39.06	1	9.32
	tire	39.06	1	9.32
	s x d	1.56	1	0.37
	s x t	5.06	1	1.21
	d x t	0.06	1	0.01
	s x d x t	3.06	1	0.73

model $y_{tijr} = \mu + \alpha_t + \beta_i + \gamma_j + (\alpha\beta)_{ti} + (\alpha\gamma)_{tj} + (\beta\gamma)_{ij}$

$$+ (\alpha\beta\gamma)_{tij} + \varepsilon_{tijr}$$

interaction terms $(\alpha\beta)_{ti}$, etc.

y ~ a + b + c + a:b + a:c + b:c + a:b:c
~ a*b*c

Finally, sometimes we have a mixture of

regression on a cont^s variable either obs'd or designed

+ a factor variable either qual. or quant.

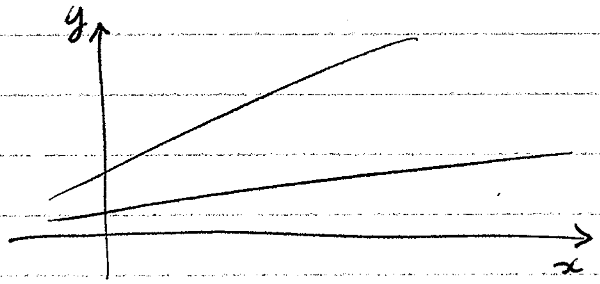
e.g. $\cancel{y_{tir} = \mu + \alpha_t + x_i + \varepsilon_{tir}}$ HW1 #3

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 (x_i z_i) + \varepsilon_i$$

$$z_i = -1, +1$$

another way to write it of could be

$$y_{1i} = \beta_0 + \beta_1 x_i - \beta_2 - \beta_3 x_i + \varepsilon_i$$
$$= \gamma_0 + \gamma_1 x_i + \varepsilon_{1i} \qquad (\beta_0 - \beta_2) + (\beta_1 - \beta_3) x_i$$
$$y_{2i} = \beta_0 + \beta_1 x_i + \beta_2 + \beta_3 x_i + \varepsilon_{2i}$$
$$(\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_i$$



$2\beta_2$ = difference betw 2 gps.

$2\beta_3$ = difference betw effect of x in 2 gps

i.e $X^=$

Also treated in § 9.3.3, with a more complex design.

Algebra for contrasts § 9.3.2

$$\underset{n\times 1}{y} = \underset{n\times p}{X}\,\underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}$$

Let $A_{p\times p}$ be s.t. $XA = C$, where $C = \begin{pmatrix} 1 & & & \\ \vdots & c_1 & \cdots & c_{p-1} \\ 1 & & & \end{pmatrix}$

& $c_i^T c_j = 0, \quad i \neq j$

$$y = XAA^{-1}\beta + \varepsilon$$

$$= C\gamma + \varepsilon, \text{ say} \qquad C^TC = \text{diag}(n, c_1^Tc_1, \ldots, c_{p-1}^Tc_{p-1})$$

$$\hat{\gamma} = (C^TC)^{-1}C^Ty, \quad \text{cov}(\hat{\gamma}) = \sigma^2 (C^TC)^{-1}$$

$$\cancel{\sigma^2}$$

i.e. $\text{cov}(\hat{\gamma}_j, \hat{\gamma}_k) = 0 \quad \text{var}\,\hat{\gamma}_j = \sigma^2 (c_j^Tc_j)^{-1}$

$$y^T\{I - C(C^TC)^{-1}C^T\}y = \underset{\sim}{y^Ty} - \hat{\gamma}^T(C^TC)^{-1}\hat{\gamma}$$

$$= y^Ty - n\bar{y}^2 - \hat{\gamma}_1^2\, c_1^Tc_1 - \cdots - \hat{\gamma}_{p-1}^2\, c_{p-1}^Tc_{p-1}$$

SS_{resid} $\qquad\qquad$ $SS_{\text{regression}}$