16. Test for non-additivity; 3 factor designs

Sometimes we can make only one observation per cell ($n = 1$). Then all $y_{ij1} - \bar{y}_{ij.} = 0$, so $SS_E = 0$ on $ab(n-1) = 0$ d.f. The interaction SS, which for $n = 1$ is

$$\sum_{i,j} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2, \qquad (*)$$

is what we should be using to estimate experimental error. There is still however a way to test for interactions, if we assume that they take a simple form:

$$(\tau\beta)_{ij} = \gamma\tau_i\beta_j.$$

We carry out 'Tukey's one d.f. test for interaction', which is an application of the usual 'reduction in SS' hypothesis testing principle. Our 'full' model is

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \varepsilon_{ij}.$$

Under the null hypothesis H_0: $\gamma = 0$ of no interactions, the 'reduced' model is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

in which the minimum SS (i.e. SS_{Red}) is (*) above. One computes

$$F_0 = \frac{SS_{Red} - SS_{Full}}{MS_E(Full)} \sim F^1_{(a-1)(b-1)-1}.$$

The difference

$$SS_N = SS_{Red} - SS_{Full}$$

is called the 'SS for non-additivity', and uses 1 d.f. to estimate the one parameter γ. The ANOVA becomes

Source	SS	df	MS
A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$
B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$
N	SS_N	1	$MS_N = \frac{SS_N}{1}$
Error	SS_E	$(a-1)(b-1)-1$	$MS_E = \frac{SS_E}{df(Err)}$
Total	SS_T	$ab-1$	

The error SS is SS_{Full}. To obtain it one has to minimize

$$\sum_{i,j}\left(y_{ij} - \left[\mu + \tau_i + \beta_j + \gamma\tau_i\beta_j\right]\right)^2.$$

After a calculation it turns out that

$$SS_N = \frac{ab\left\{\sum_{i,j} y_{ij}\bar{y}_{i.}\bar{y}_{.j} - \bar{y}_{..}\left(SS_A + SS_B + ab\bar{y}_{..}^2\right)\right\}^2}{SS_A \cdot SS_B}.$$

Then SS_E is obtained by subtraction: $SS_E = SS_{Red} - SS_N$.

An R function to calculate this, and carry out the F-test, is at "R commands for Tukey's 1 df test" on the course web site.

Example. For the experiment at Example 5.2 of the text there are $a = 3$ levels of temperature and $b = 5$ of pressure; response is $Y =$ impurities in a chemical product.

```
> h <- tukey.1df(y,temp,press)
    SS     df MS     F0     p
A   23.333 2  11.667 42.949 1e-04
B   11.6   4  2.9    10.676 0.0042
N   0.099  1  0.099  0.363  0.566
Err 1.901  7  0.272
Tot 36.933 14
```

A 3 factor example. Softdrink bottlers must maintain targets for fill heights, and any variation is a cause for concern. The deviation from the target (Y) is affected by %carbonation (A), pressure in the filler (B), line speed (C). These are set at $a = 3, b = 2, c = 2$ levels respectively, with $n = 2$ observations at each combination ($N = nabc = 24$ runs, in *random order*).

	y	carbon	press	speed
1	-3	10	25	200
2	-1	10	25	200
3	0	12	25	200
4	1	12	25	200
5	5	14	25	200
6	4	14	25	200
		....		
19	1	10	30	250
20	1	10	30	250
21	6	12	30	250
22	5	12	30	250
23	10	14	30	250
24	11	14	30	250

```
> plot.design(data)
> interaction.plot(carbon,press,y)
> interaction.plot(carbon,speed,y)
> interaction.plot(press,speed,y)
```

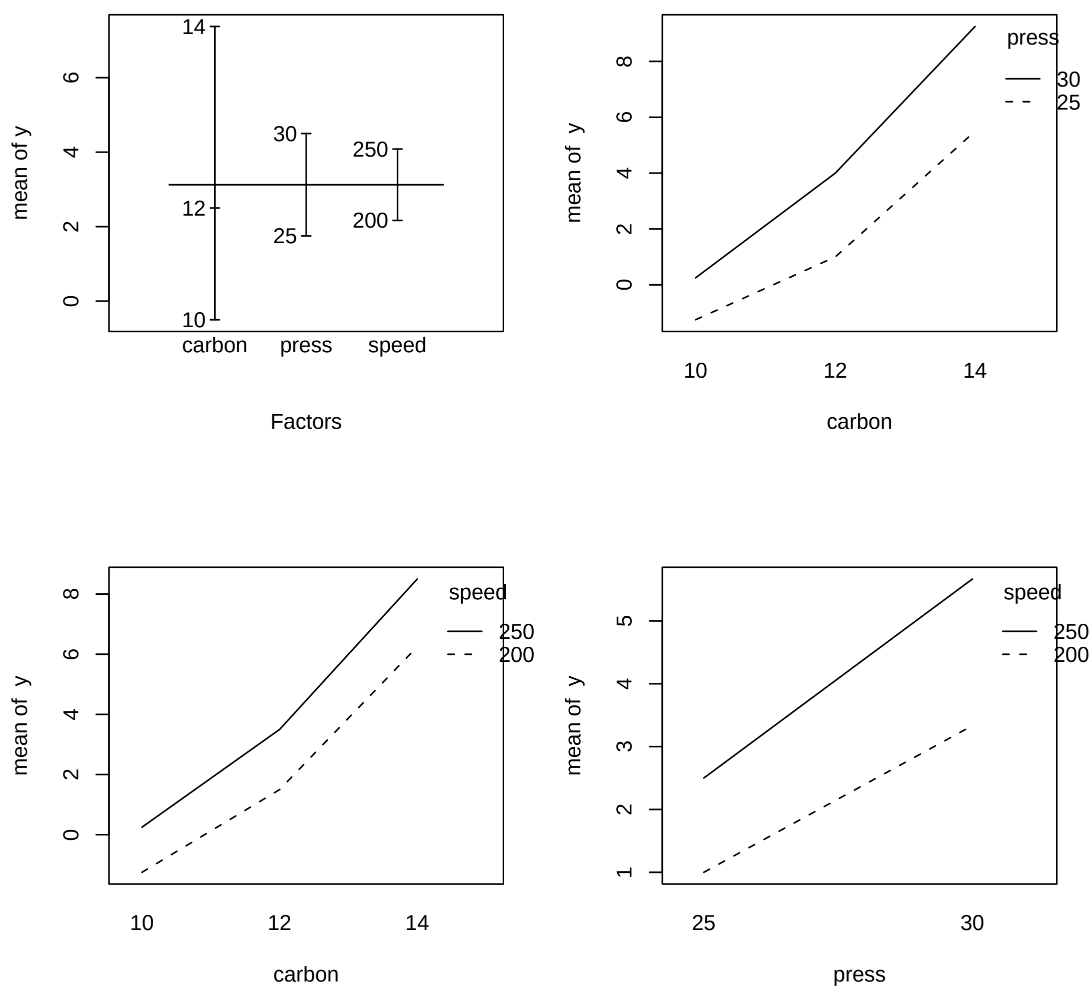


Fig. 5.5.

Full 3 factor model:

$$\begin{aligned} y_{ijkl} &= \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} \\ &\quad + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}. \end{aligned}$$

```
> g <- lm(y ~carbon + press + speed + carbon*press
 + carbon*speed + press*speed + carbon*press*speed)
> anova(g)
Analysis of Variance Table

Response: y
       Df  Sum Sq Mean Sq  F value    Pr(>F)
C       2 252.750 126.375 178.4118 1.186e-09
P       1  45.375  45.375  64.0588 3.742e-06
S       1  22.042  22.042  31.1176 0.0001202
C:P     2   5.250   2.625   3.7059 0.0558081
C:S     2   0.583   0.292   0.4118 0.6714939
P:S     1   1.042   1.042   1.4706 0.2485867
C:P:S  2   1.083   0.542   0.7647 0.4868711
Resid 12   8.500   0.708
```

It seems that interactions are largely absent, and that all three main effects are significant. In particular, the low level of pressure results in smaller mean deviations from the target. A CI on $\beta_2 - \beta_1 = E\left[\bar{y}_{.2.} - \bar{y}_{.1.}\right]$ is ($\alpha = .05$)

$$\begin{aligned} & \bar{y}_{.2.} - \bar{y}_{.1.} \pm t_{\alpha/2,12}\sqrt{MS_E\left(\frac{1}{12}+\frac{1}{12}\right)} \\ = \; & 1.75 - 4.5 \pm 2.1788\sqrt{\frac{.708}{6}} \\ = \; & -2.75 \pm .75 \end{aligned}$$

or $[-3.5, -2]$.

17. 2^2 factorials

- We'll start with a basic 2^2 design, where it is easy to see what is going on. Also, these are very widely used in industrial experiments.

- Two factors (A and B), each at 2 levels - low ('$-$') and high ('$+$'). # of replicates $= n$.

- Example - investigate yield (y) of a chemical process when the concentration of a reactant (the primary substance producing the yield) - factor A - and amount of a catalyst (to speed up the reaction) - factor B - are changed. E.g. nickel is used as a 'catalyst', or a carrier of hydrogen in the hydrogenation of oils (the reactants) for use in the manufacture of margarine.

Factor		$n = 3$ replicates				
A	B	I	II	III	Total	Label
$-$	$-$	28	25	27	$80 =$	(1)
$+$	$-$	36	32	32	$100 =$	a
$-$	$+$	18	19	23	$60 =$	b
$+$	$+$	31	30	29	$90 =$	ab

- Notation

$$\begin{aligned} (1) &= \text{sum of obs'ns at low levels of both factors,} \\ a &= \text{sum of obs'ns with A high and B low,} \\ b &= \text{sum of obs'ns with B high and A low,} \\ ab &= \text{sum of obs'ns with both high.} \end{aligned}$$

- Effects model. Use a more suggestive notation:

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk} \; (i, j = 1, 2, \; k = 1, ..., n)$$

- E.g. $A_1 =$ main effect of low level of A, $A_2 =$ main effect of high level of A. But since $A_1 + A_2 = 0$, we have $A_1 = -A_2$.

- We define the 'main effect of Factor A' to be

$$A = A_2 - A_1.$$

- What is the LSE of A? Since A is the effect of changing factor A from high to low, we expect

$$\begin{aligned}\hat{A} &= \text{ average } y \text{ at high A - average } y \text{ at low A} \\ &= \frac{a+ab}{2n} - \frac{(1)+b}{2n} \\ &= \frac{a+ab-(1)-b}{2n}.\end{aligned}$$

This is the LSE.
Reason: We know that the LSE of A_2 is

$$\hat{A}_2 = \text{average } y \text{ at high } A - \text{overall average } y,$$

and that of A_1 is

$$\hat{A}_1 = \text{average } y \text{ at low } A - \text{overall average } y,$$

so that

$$\begin{aligned}\hat{A} &= \hat{A}_2 - \hat{A}_1 \\ &= \text{ average } y \text{ at high A - average } y \text{ at low A.}\end{aligned}$$

- Often the 'hats' are omitted (as in the text). Similarly,

$$
\begin{aligned}
B &= \frac{b + ab - a - (1)}{2n} \\
AB &= \text{difference between effect of A at high B,} \\
&\quad \text{and effect of A at low B} \\
&= \frac{ab - b}{2n} - \frac{a - (1)}{2n} \\
&= \frac{ab - b - a + (1)}{2n}.
\end{aligned}
$$

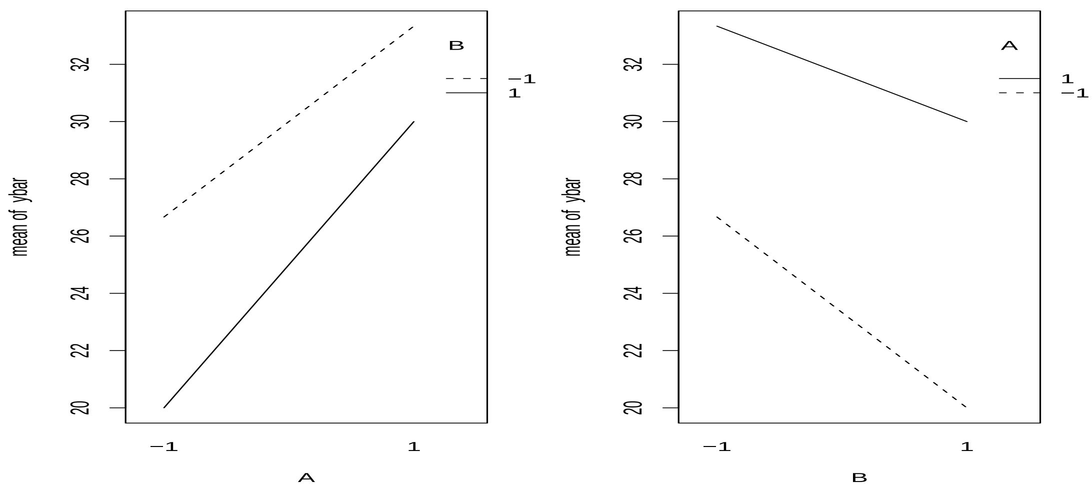
With $(1) = 80, a = 100, b = 60, ab = 90$ we find

$$
\begin{aligned}
A &= 8.33, \\
B &= -5.0 \\
AB &= 1.67.
\end{aligned}
$$

- It appears that increasing the level of A results in an increase in yield; that the opposite is true of B, and that there isn't much interaction effect. To confirm this we would do an ANOVA.

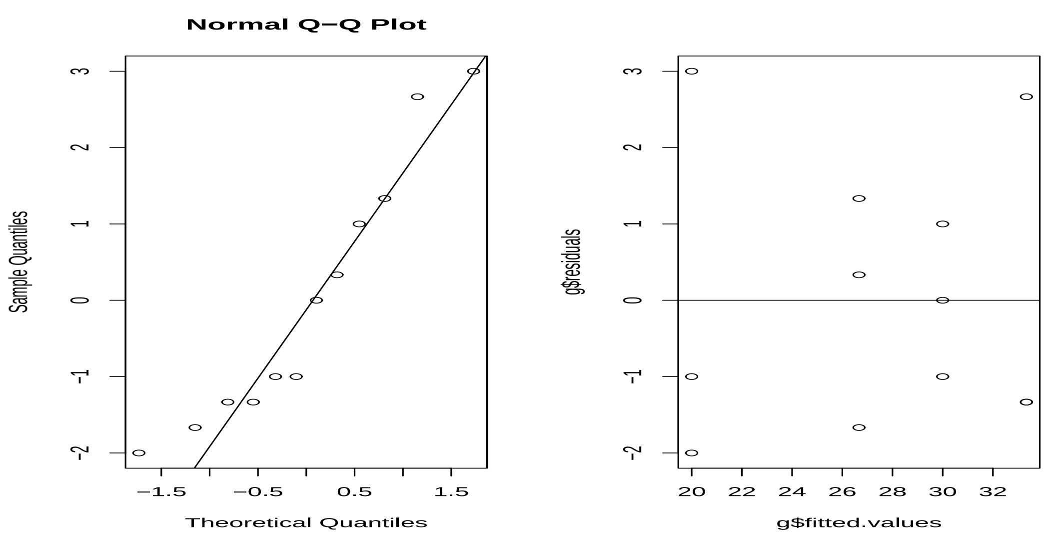
```
> A <- c(-1, 1, -1,1)
> B <- c(-1, -1, 1, 1)
> I <- c(28, 36, 18, 31)
> II <- c(25, 32, 19, 30)
> III <- c(27, 32, 23, 29)
>
> data <- data.frame(A, B, I, II, III)
> data
   A  B  I II III
1 -1 -1 28 25  27
2  1 -1 26 32  32
3 -1  1 18 19  23
4  1  1 31 30  29


# compute sums for each combination
> sums <- apply(data[,3:5], 1, sum)
> names(sums) <- c("(1)", "(a)", "(b)", "(ab)")
> sums
 (1)  (a)   (b) (ab)
  80  100   60   90
```

```
# Interaction plots
> ybar <- sums/3
> par(mfrow=c(1,2))
> interaction.plot(A, B, ybar)
> interaction.plot(B, A, ybar)

# Build ANOVA table

> y <- c(I, II, III)
> factorA <- as.factor(rep(A,3))
> factorB <- as.factor(rep(B,3))
> g <- lm(y ~factorA + factorB + factorA*factorB)
> anova(g)

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
factorA    1 208.333 208.333 53.1915 8.444e-05
factorB    1  75.000  75.000 19.1489  0.002362
AB         1   8.333   8.333  2.1277  0.182776
Residuals  8  31.333   3.917
```



mean of ybar
A
B
-1
1
mean of ybar
B
A
1
-1

Fig. 6.1



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
g$residuals
g$fitted.values

Fig. 6.2

Contrasts. The estimates of the effects have used only the terms ab, a, b and (1), each of which is the sum of $n = 3$ independent terms. Then

$$\begin{aligned} A &= \frac{ab + a - b - (1)}{2n} = \frac{C_A}{2n}, \\ B &= \frac{ab - a + b - (1)}{2n} = \frac{C_B}{2n}, \\ AB &= \frac{ab - a - b + (1)}{2n} = \frac{C_{AB}}{2n}, \end{aligned}$$

where C_A, C_B, C_{AB} are orthogonal contracts (why?) in ab, a, b and (1). In our previous notation, the SS for Factor A (we might have written it as $bn \sum \hat{A}_i^2$) is

$$\begin{aligned} SS_A &= 2n\left(\hat{A}_1^2 + \hat{A}_2^2\right) = 4n\hat{A}_2^2 = nA^2 = \frac{C_A^2}{4n}, \\ &\quad \text{and similarly} \\ SS_B &= \frac{C_B^2}{4n},\ SS_{AB} = \frac{C_{AB}^2}{4n}. \\ SS_E &= SS_T - SS_A - SS_B - SS_{AB}. \end{aligned}$$

In this way $SS_A = [90 + 100 - 60 - 80]^2 / 12 = 208.33$.