

Today's example is from Walpole, Myers & Myers, "Probability and Statistics for Engineers and Scientists," 6th ed. Problem 13.16.3. A 1983 Virginia Polytechnic Institute study wished to test how different protein levels in feed rations for cows affects the daily milk production. Five kinds of feed were given to five cows which which were milked at five different lactation periods. To account for the differences in cow and period, the experiment was conducted as a 5 x 5 Latin Square, where rows were periods, columns were cows and the treatments were symbols in the table. The following Latin square was chosen randomly from all 5 x 5 squares for treatments (protein)

A	B	C	D	E
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D
B	C	D	E	A

The rows and columns (periods and cows) may be viewed as randomized "blocks," that contribute variability to the experiment which may obscure the influence of the treatment (feed). By partitioning the contributions in the variance from rows, columns, treatments and the rest lets us remove the variability contributed by the two blocks (row and column). Also, this is a fractional experiment: only 25 observations are taken from a possible $5^3 = 125$. But, since in the Latin square each treatment occurs exactly once in each row and each column, every row and every column receives each treatment exactly once.

If we propose the fixed effects model

$$X_{ij(k)} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ij(k)}$$

where $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \tau_k = 0$ and $\epsilon_{ij(k)}$ are i.i.d. normal variables with mean zero and variance σ^2 . The index " $ij(k)$ " means i, j, k run through N levels of the three factors, but that any two of the i, j, k determine the third. Thus cell sums are as usual. For $N = 5$ and the Latin square above,

$$\begin{aligned} X_{1.(.)} &= X_{11(A)} + X_{12(B)} + X_{13(C)} + X_{14(D)} + X_{15(E)} \\ X_{.1(.)} &= X_{11(A)} + X_{21(C)} + X_{31(D)} + X_{41(E)} + X_{51(B)} \end{aligned}$$

but also the sum is over all cells that get the same treatment

$$X_{..(A)} = X_{11(A)} + X_{24(A)} + X_{33(A)} + X_{42(A)} + X_{55(A)}$$

so $\bar{X}_{..(k)} = \frac{1}{N} X_{..(k)}$. These sums are orthogonal and represent independent contrasts of the data.

That implies also that the corresponding sums may be partitioned off in the sum square identity. Thus, if we have SSA and SSB as usual, we also have

$$SSC = \sum_{i,j(k)} (X_{..(k)} - X_{i.(.)})^2$$

which is a sum over all N^2 observations. Not all triples ijk occur in the sum, only the ones in the Latin square: we may name two indices and the third is determined. Thus we may run

through all pairs ij and then k is determined or just as well we may run through all ik and j is determined. Hence we get the partitioning of the total sum of squares:

$$SST = SSA + SSB + SSC + SSE$$

where

$$SST = \sum_{i,j(k)} (X_{ij(k)} - \bar{X}_{..(\cdot)})^2$$

and

$$SSE = \sum_{i,j(k)} (X_{ij(k)} - \hat{X}_{ij(k)})^2$$

with

$$\begin{aligned} \hat{X}_{ij(k)} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\tau}_k \\ &= \bar{X}_{..(\cdot)} + (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) + (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)}) + (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) \\ &= \bar{X}_{i(\cdot)} + \bar{X}_{\cdot j(\cdot)} + \bar{X}_{..(k)} - 2\bar{X}_{..(\cdot)} \end{aligned}$$

The proof of the sum square identity is as usual: add and subtract the effects and then verify that the cross terms vanish. $SST =$

$$\begin{aligned} &= \sum_{i,j(k)} [X_{ij(k)} - \bar{X}_{..(\cdot)}]^2 \\ &= \sum_{i,j(k)} [(X_{ij(k)} - \bar{X}_{i(\cdot)} - \bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(k)} + 2\bar{X}_{..(\cdot)}) + (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) + (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)}) + (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})]^2 \\ &= \sum_{i,j(k)} [(X_{ij(k)} - \bar{X}_{i(\cdot)} - \bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(k)} + 2\bar{X}_{..(\cdot)})^2 + (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)})^2 + (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)})^2 + (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})^2 \\ &\quad + 2(X_{ij(k)} - \bar{X}_{i(\cdot)} - \bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(k)} + 2\bar{X}_{..(\cdot)})(\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) + \dots + 2(\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)})(\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})] \end{aligned}$$

The cross terms vanish. For example the first is

$$\begin{aligned} &\sum_{i,j(k)} (X_{ij(k)} - \bar{X}_{i(\cdot)} - \bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(k)} + 2\bar{X}_{..(\cdot)})(\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) \\ &= \sum_{i,j(k)} (X_{ij(k)} - \bar{X}_{i(\cdot)})(\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) - (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)})(\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) - (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})(\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) \\ &= \sum_i (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) \sum_j (X_{ij(k)} - \bar{X}_{i(\cdot)}) - \sum_j (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)}) \sum_i (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) \\ &\quad - \sum_k (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) \sum_i (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)}) \\ &= \sum_i (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)})(N\bar{X}_{i(\cdot)} - N\bar{X}_{..(\cdot)}) - \sum_j (\bar{X}_{\cdot j(\cdot)} - \bar{X}_{..(\cdot)})(N\bar{X}_{..(\cdot)} - N\bar{X}_{..(\cdot)}) \\ &\quad - \sum_k (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})(N\bar{X}_{..(\cdot)} - N\bar{X}_{..(\cdot)}) = 0 + 0 + 0. \end{aligned}$$

The last is

$$\begin{aligned}
 & \sum_{i,j(k)} (\bar{X}_{.j(\cdot)} - \bar{X}_{..(\cdot)}) (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) \\
 &= \sum_j \sum_k (\bar{X}_{.j(\cdot)} - \bar{X}_{..(\cdot)}) (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) \\
 &= \sum_k (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) \sum_j (\bar{X}_{.j(\cdot)} - \bar{X}_{..(\cdot)}) \\
 &= \sum_k (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)}) (N\bar{X}_{..(\cdot)} - N\bar{X}_{..(\cdot)}) = 0.
 \end{aligned}$$

It follows that

$$\begin{aligned}
 SST &= \sum_{i,j(k)} (X_{ij(k)} - \bar{X}_{i(\cdot)} - \bar{X}_{.j(\cdot)} - \bar{X}_{..(k)} + 2\bar{X}_{..(\cdot)})^2 + (\bar{X}_{i(\cdot)} - \bar{X}_{..(\cdot)})^2 + (\bar{X}_{.j(\cdot)} - \bar{X}_{..(\cdot)})^2 + (\bar{X}_{..(k)} - \bar{X}_{..(\cdot)})^2 \\
 &= SSE + SSA + SSB + SSC.
 \end{aligned}$$

Data Set Used in this Analysis :

```

# Math 3080           Milking Example           Feb. 8, 2014
# Treibergs
#
# Data from Walpole, Myers & Myers, "Probability and Statistics for
# Engineers and Scientists," 6th ed. Prob 13.16.3
# A 1983 Virginia Polytechnic Institute study wished to test how different
# protein levels in feed rations for cows affects the daily milk
# production. Five kinds of feed were given to five cows which which were
# milked at five different lactation periods. To account for the
# differences in cow and period, the experiment was conducted as a 5 x 5
# Latin Square, where rows were periods, columns were cows and the
# treatments were symbols in the table. The following Latin square was
# chosen randomly from all 5 x 5 squares for treatments (protein)
#
#   A B C D E
#   C D E A B
#   D E A B C
#   E A B C D
#   B C D E A
#
#
# For the 25 combinations given by the square, milk production (kg) is
# recorded. Is there a significant difference due to feed?
"Period" "Cow" "Feed" "Milk"
1 1 A 33.1
1 2 B 34.4
1 3 C 26.4
1 4 D 34.6
1 5 E 33.9
2 1 C 30.7

```

2 2 D 28.7
2 3 E 24.9
2 4 A 28.8
2 5 B 28.0
3 1 D 28.7
3 2 E 28.8
3 3 A 20.0
3 4 B 31.9
3 5 C 22.7
4 1 E 31.4
4 2 A 22.3
4 3 B 18.7
4 4 C 30.1
4 5 D 21.3
5 1 B 28.9
5 2 C 22.3
5 3 D 15.8
5 4 E 30.9
5 5 A 19.0

R Session:

R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrestreibergs/.Rapp.history]

```
> tt=read.table("M3082DataMilking.txt",header=T)
> attach(tt)
> tt
  Period Cow Feed Milk
1      1   1   A 33.1
2      1   2   B 34.4
3      1   3   C 26.4
```

```

4      1  4  D 34.6
5      1  5  E 33.9
6      2  1  C 30.7
7      2  2  D 28.7
8      2  3  E 24.9
9      2  4  A 28.8
10     2  5  B 28.0
11     3  1  D 28.7
12     3  2  E 28.8
13     3  3  A 20.0
14     3  4  B 31.9
15     3  5  C 22.7
16     4  1  E 31.4
17     4  2  A 22.3
18     4  3  B 18.7
19     4  4  C 30.1
20     4  5  D 21.3
21     5  1  B 28.9
22     5  2  C 22.3
23     5  3  D 15.8
24     5  4  E 30.9
25     5  5  A 19.0
> Period=ordered(Period)
> Cow=ordered(Cow)
> Feed=ordered(Feed)
>
> #####PRINT THE DATA AND LATIN SQUARE MATRIX #####
> matrix(Milk,ncol=5,byrow=T)

      [,1] [,2] [,3] [,4] [,5]
[1,] 33.1 34.4 26.4 34.6 33.9
[2,] 30.7 28.7 24.9 28.8 28.0
[3,] 28.7 28.8 20.0 31.9 22.7
[4,] 31.4 22.3 18.7 30.1 21.3
[5,] 28.9 22.3 15.8 30.9 19.0

> matrix(Feed,ncol=5,byrow=T)

      [,1] [,2] [,3] [,4] [,5]
[1,] "A"  "B"  "C"  "D"  "E"
[2,] "C"  "D"  "E"  "A"  "B"
[3,] "D"  "E"  "A"  "B"  "C"
[4,] "E"  "A"  "B"  "C"  "D"
[5,] "B"  "C"  "D"  "E"  "A"

```

```

> ##### ANOVA #####

> a1=aov(Milk~Period+Cow+Feed)
> summary(a1)
      Df Sum Sq Mean Sq F value    Pr(>F)
Period    4  249.82   62.455  42.078 5.735e-07 ***
Cow       4  345.42   86.355  58.180 9.322e-08 ***
Feed      4   90.23   22.559  15.198 0.0001206 ***
Residuals 12   17.81    1.484
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> ##### MODEL TABLE OF EFFECTS #####
> model.tables(a1)
Tables of effects

Period
Period
  1    2    3    4    5
5.428 1.168 -0.632 -2.292 -3.672

Cow
Cow
  1    2    3    4    5
3.508 0.248 -5.892 4.208 -2.072

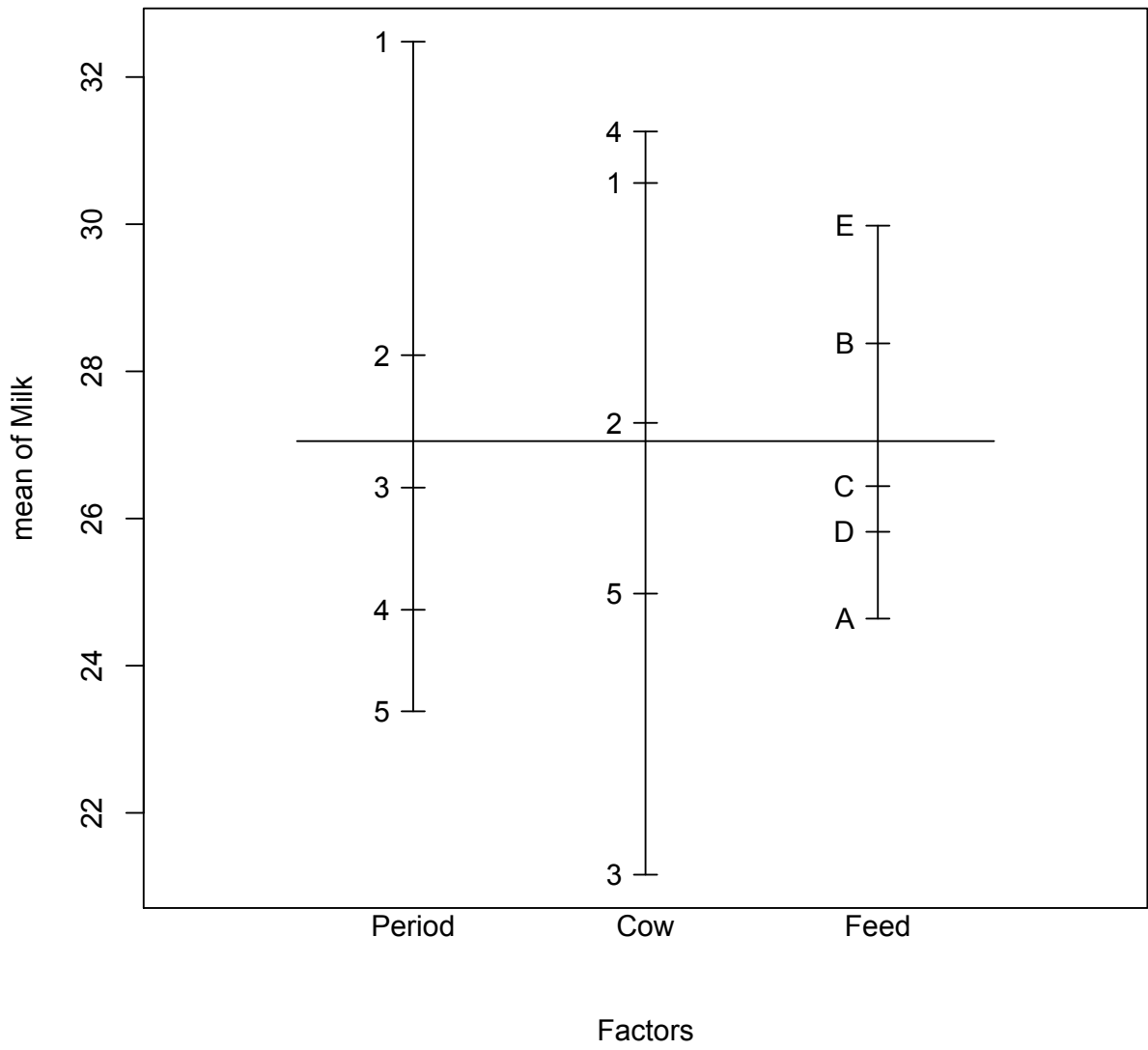
Feed
Feed
  A    B    C    D    E
-2.412 1.328 -0.612 -1.232 2.928

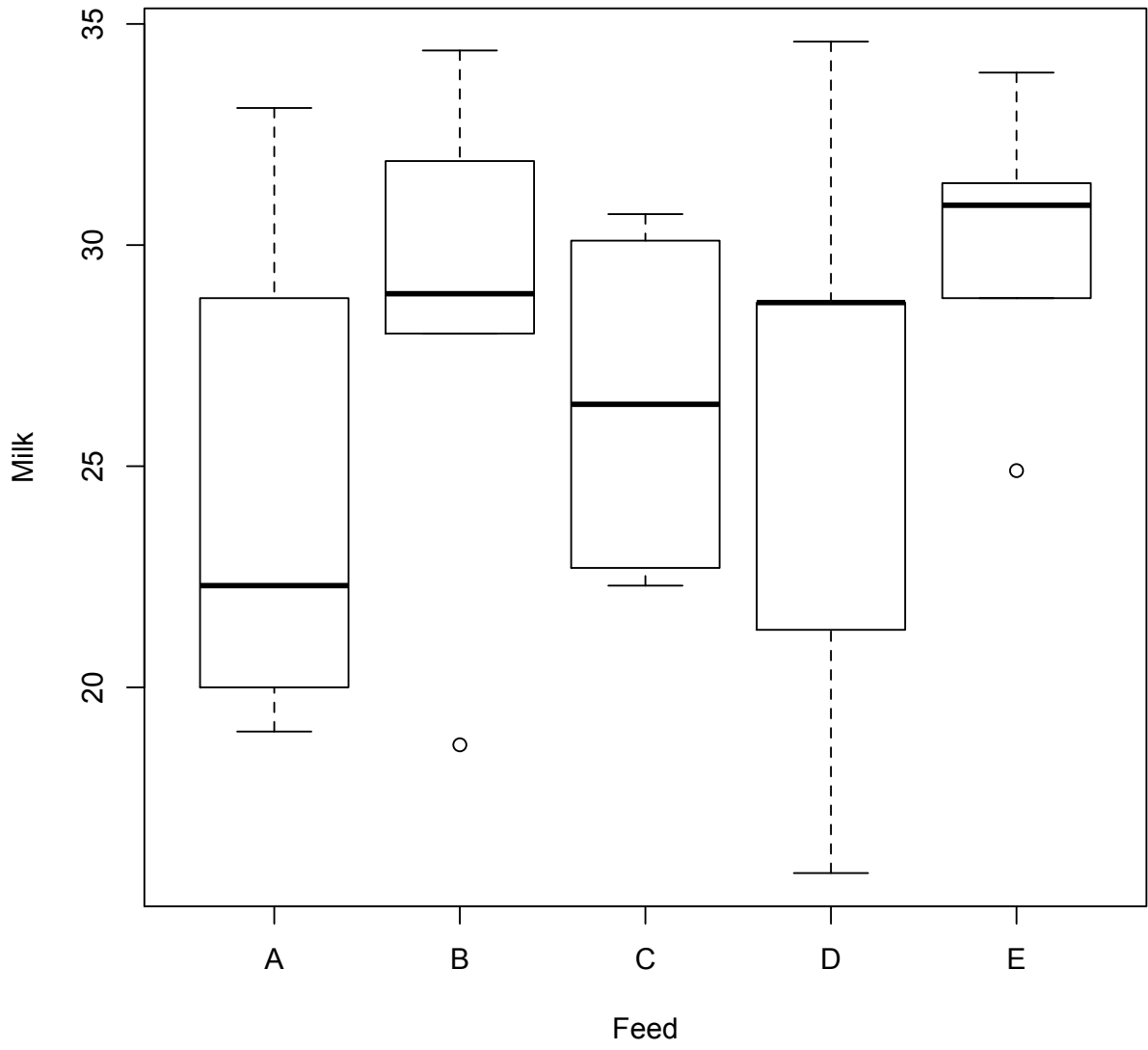
>
> ##### DESIGN, BOX AND INTERACTION PLOTS #####
> plot.design(Milk~Period+Cow+Feed)
> plot(Milk~Feed)
> interaction.plot(Feed,Cow,Milk)
> interaction.plot(Feed,Period,Milk)

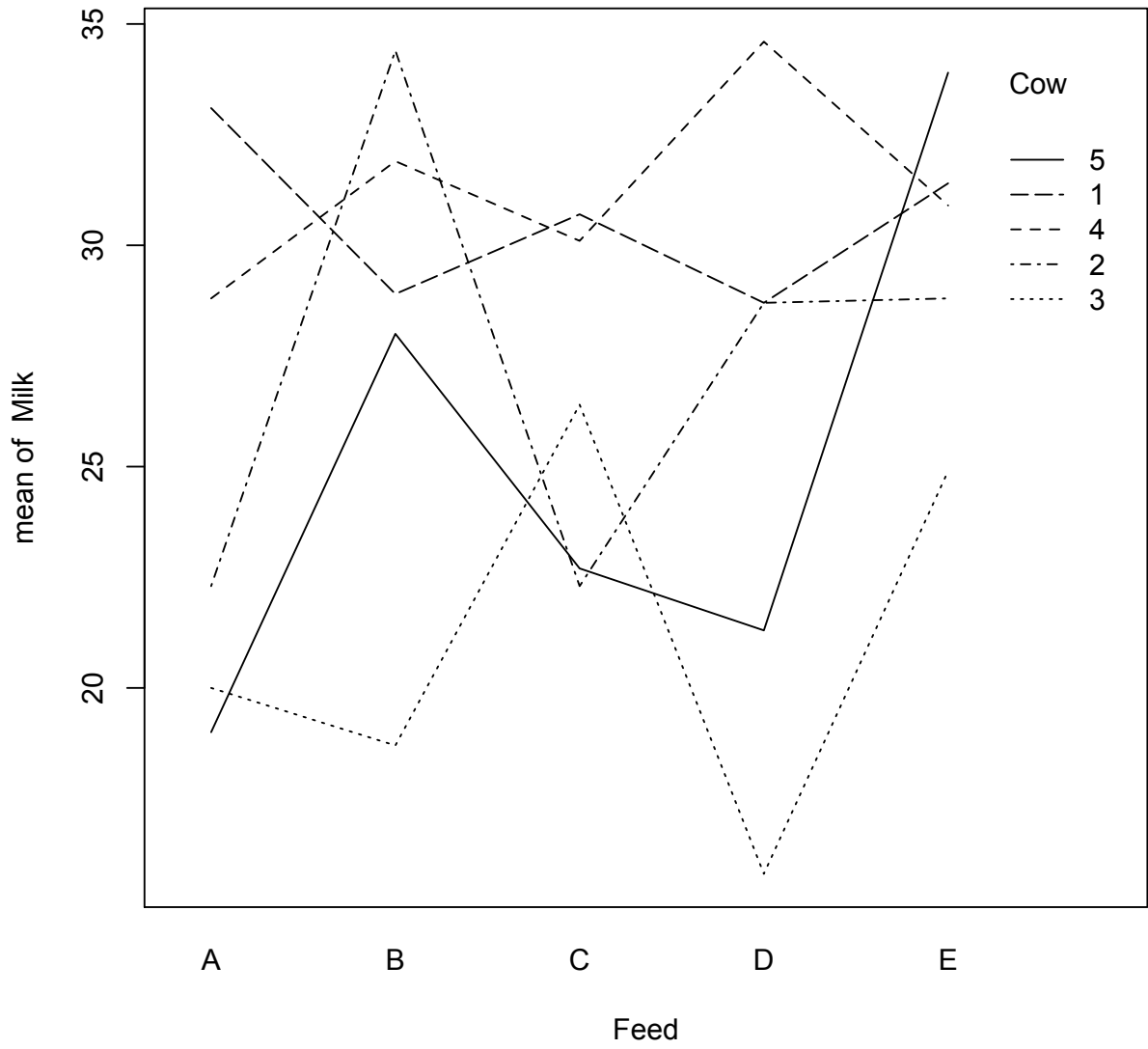
>
> ##### FOR COMPARISON: TREAT AS SINGLE FACTOR EXPERIMENT #####

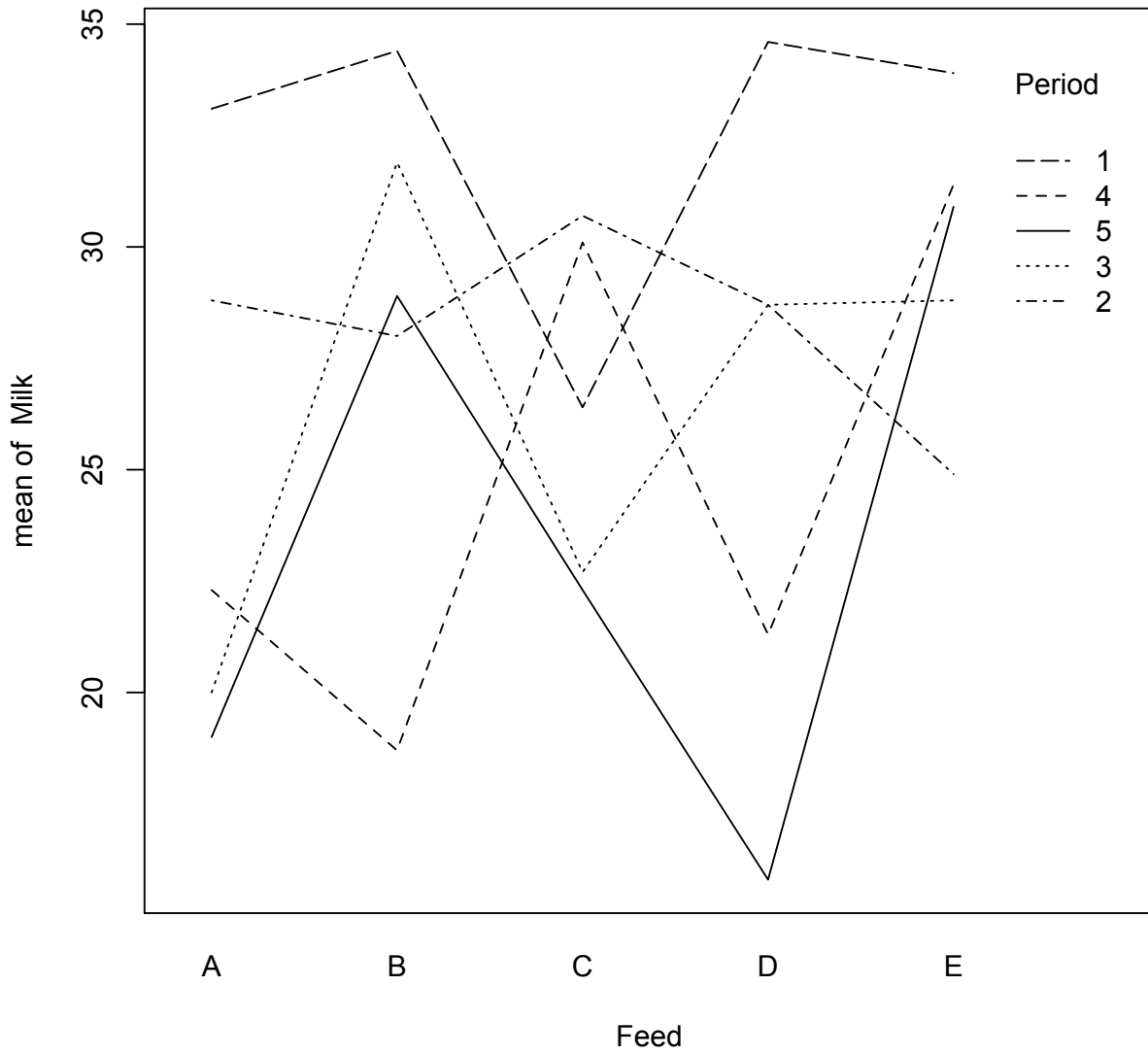
> a3=aov(Milk~Feed)
> summary(a3)
      Df Sum Sq Mean Sq F value    Pr(>F)
Feed      4   90.23   22.559   0.7359 0.5783
Residuals 20 613.05   30.652
>

```









```

> ## DIAGNOSTIC PLOTS: RESID. VS. FITTED & NORMAL QQ PLOT #####
> plot(a1)
Hit <Return> to see next plot:
Hit <Return> to see next plot:
>

```

