Chapter 8: Experiments with two factors

Introductory Statistics for Engineering Experimentation

Peter R. Nelson, Marie Coffin and Karen A.F. Copeland

Slides by Douglas Bates

Outline

8.1 Interaction

8.2 More than one observation per cell

8.3 Only one observation per cell

8.4 Blocking to reduce variability

Outline

8.1 Interaction

8.2 More than one observation per cell

8.3 Only one observation per cell

8.4 Blocking to reduce variability

Outline

8.1 Interaction

8.2 More than one observation per cell

8.3 Only one observation per cell

8.4 Blocking to reduce variability

Outline

Chapter 8: Experiments with two factors

- Recall from chapter 3 that when we have a response on a continuous scale and two categorical covariates, we can consider the observations as being in the *cells* of a *two-way layout* determined by the levels of the two factors.

- There will always be one experimental factor whose levels are of interest in themselves. The second factor may be a *blocking factor*: a known source of variability for which we are controlling.

- If the factors are A and B we write the number of levels of the factors as I and J, respectively, and the number of observations in the (i, j)th cell as $n_{ij}, i = 1, \ldots, I; j = 1, \ldots, J$.

- In a *balanced* experiment all the n_{ij} are equal so we simply write the number of observations per cell as n. The text provides methods for balanced designs only. The methods we will describe can be applied to unbalanced designs too.

Replicated *vs.* non-replicated designs

- If none of the n_{ij} are zero then we have a *complete layout*, which is a type of *full factorial design*.
- When there is at least one of the $n_{ij} > 1$ the design is said to be *replicated*.
- Observations in the same cell are called *replicates*.
- In chapter 3 we discussed fitting models of the form

  $$\mathcal{Y}_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}, \quad i = 1, \ldots, I; \ j = 1, \ldots, J; \ k = 1, \ldots$$

  when we have replicates. This model allows for interactions (the $\alpha\beta_{ij}$ terms).
- Our first test is on the null hypothesis that all the interactions are zero. If we do not reject this hypothesis then we simplify the model to the *additive model*

  $$\mathcal{Y}_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad i = 1, \ldots, I; \ j = 1, \ldots, J; \ k = 1, \ldots, n_{ij}$$

- We can fit a model with interactions to unreplicated data but cannot do further analysis. Thus we only fit the additive model.

Overview of techniques

- We use interaction plots to assess the changes due to the levels of the factors and due to possible interaction.

- If there are replicates we fit the model with interactions and test first for a significant interation term. If the interactions are not significant then we fit the additive model.

- For an additive model we can perform multiple comparisons on the levels of the factors separately (but only for experimental factors, not for blocking factors) For a model with significant interactions we cannot separate the effects of the factors.

- For an unreplicated design we can fit a model with interactions but cannot do any further analysis. We must fit the additive model even though we do not know if it is appropriate.

R functions used in this chapter

- Interaction plots are obtained with `dotplot`.
- We use `aov` to fit the models, `summary` or `anova` to obtain the analysis of variance table, and `model.tables` to obtain estimates of the cell means or the effects.
- We use `TukeyHSD` to perform multiple comparisons when appropriate.
- We assess residual plots obtained with `plot(fm, which = 1)` and `plot(fm, which = 2)`

Section 8.1, Interaction

- Recall that in an interaction plot a lack of interaction is indicated by more-or-less parallel lines joining the averages.
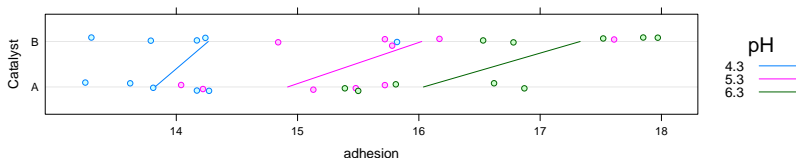

```
> summary(fm1 <- aov(adhesion ~ cat * factor(pH), adhesion2))

               Df Sum Sq Mean Sq F value    Pr(>F)
cat             1  6.712   6.712 11.4424   0.00246
factor(pH)      2 34.924  17.462 29.7694 3.161e-07
cat:factor(pH)  2  1.003   0.502  0.8553   0.43772
Residuals      24 14.078   0.587
```

Section 8.2 More than one observation per cell

- When we fit the model with interactions we always test first for significant interactions. The null hypothesis is

$$H_0 : \alpha\beta_{ij} = 0, \text{ for all } i, j$$

- If we do not reject this hypothesis then we fit the additive model.

- For the `adhesion2` anova (previous slide) the p-value for the interaction, labeled `cat:factor(pH)`, is 0.4377 so we do not reject H_0.

- The additive model fit is

```
> summary(fm2 <- aov(adhesion ~ cat + factor(pH), adhesion2))
           Df Sum Sq Mean Sq F value    Pr(>F)
cat         1  6.712   6.712  11.571  0.002176
factor(pH)  2 34.924  17.462  30.105 1.708e-07
Residuals  26 15.081   0.580
```

Multiple comparisons for an additive model

- With an additive model it makes sense to perform multiple comparisons on the levels of factors separately.
- In the previous example one of the factors (`cat`) is a two-level factor. We do not need to perform multiple comparisons for such a factor (when you have only two levels there is only one possible comparison of levels).
- We can use the optional argument `which` to `TukeyHSD` to restrict to only the nontrivial multiple comparisons.

```
> TukeyHSD(fm2, which = "factor(pH)")

  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = adhesion ~ cat + factor(pH), data = adhesion2)


$`factor(pH)`
          diff       lwr      upr      p adj
5.3-4.3 1.427 0.5806392 2.273361 0.0008057
6.3-4.3 2.640 1.7936392 3.486361 0.0000001
6.3-5.3 1.213 0.3666392 2.059361 0.0040071
```
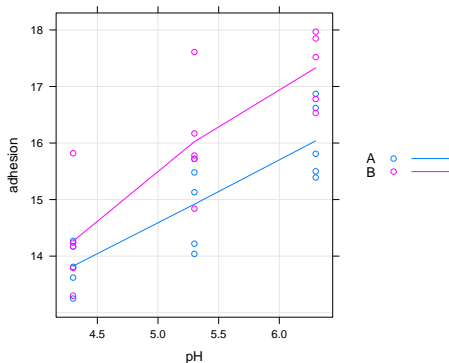
Alternative analysis of adhesion2

- This is jumping ahead a bit but for data like the `adhesion2` data we are losing power by treating the numeric `pH` variable as a factor. Over the observed range the adhesion is close to linear in `pH` for both catalysts.

Alternative analysis of adhesion2 (cont'd)

- We could fit a model that is linear in `pH` allowing for different slopes and intercepts for the two catalysts.

```
> summary(fm3 <- lm(adhesion ~ pH * cat, adhesion2))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.0596     1.2554   7.216 1.16e-07
pH            1.1070     0.2341   4.728 6.88e-05
catB         -1.3118     1.7755  -0.739    0.467
pH:catB       0.4260     0.3311   1.287    0.210
Residual standard error: 0.7403 on 26 degrees of freedom
Multiple R-squared: 0.7488, Adjusted R-squared: 0.7198
F-statistic: 25.83 on 3 and 26 DF,  p-value: 5.825e-08
> anova(fm3)

Analysis of Variance Table
Response: adhesion
          Df Sum Sq Mean Sq F value     Pr(>F)
pH         1 34.848  34.848 63.5812 1.881e-08
cat        1  6.712   6.712 12.2460 0.001699
pH:cat     1  0.907   0.907  1.6555 0.209545
Residuals 26 14.250   0.548
```

Alternative analysis of adhesion2 (cont'd)

- Notice that each the interaction term in this model (labeled `pH:catB` now) has only 1 degree of freedom so the `anova` results provide exactly the same information as the t-test in the summary.

- Because the interaction term is not significant we reduce to an additive model (corresponding to parallel lines w.r.t. `pH` for each of the two catalysts).

```
> summary(fm4 <- lm(adhesion ~ pH + cat, adhesion2))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.9307     0.9088   8.727 2.42e-09
pH            1.3200     0.1675   7.879 1.80e-08
catB          0.9460     0.2736   3.458  0.00182
Residual standard error: 0.7493 on 27 degrees of freedom
Multiple R-squared: 0.7328, Adjusted R-squared: 0.713
F-statistic: 37.01 on 2 and 27 DF,  p-value: 1.834e-08
```

Yet another analysis of adhesion2

- When we have a two-level factor (like the `cat` factor) or when we have a numeric factor observed only at a few, equally-spaced levels, it can be useful to convert such factors to `ordered` factors.

- For more than two levels the ordered factor produces coefficients for linear, quadratic, cubic, etc. terms. We can check if the higher order terms are not significant and reduce the model accordingly (always proceeding from the highest order down).

- For a two-level `ordered` factor the default parameterization of the model has certain advantages in the analysis of two-level factorial designs.

- The analysis of variance table is the same whether we use `factor` or `ordered`

Yet another analysis of adhesion2 (cont'd)

```
> ad2 <- within(adhesion2, {opH <- ordered(pH); ocat <- ordered(
> anova(fm5 <- lm(adhesion ~ ocat * opH, ad2))

Analysis of Variance Table
Response: adhesion
          Df Sum Sq Mean Sq F value    Pr(>F)
ocat       1  6.712   6.712 11.4424   0.00246
opH        2 34.924  17.462 29.7694 3.161e-07
ocat:opH   2  1.003   0.502  0.8553   0.43772
Residuals 24 14.078   0.587
> summary(fm6 <- lm(adhesion ~ ocat + opH, ad2))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.39967    0.13905 110.749  < 2e-16
ocat.L       0.66892    0.19665   3.402  0.00218
opH.L        1.86676    0.24084   7.751 3.18e-08
opH.Q       -0.08737    0.24084  -0.363  0.71973
Residual standard error: 0.7616 on 26 degrees of freedom
Multiple R-squared: 0.7341, Adjusted R-squared: 0.7034
F-statistic: 23.93 on 3 and 26 DF,  p-value: 1.206e-07
```

The quadratic term (`opH.Q`) is not significant so we reduce to
model `fm4`.

Example 8.2.9


```
> ## compare to anova table on p. 316
> summary(fm7 <- aov(time ~ brand * type, computer))
            Df Sum Sq Mean Sq F value  Pr(>F)
brand        2  99.41   49.71  2.4045 0.10939
type         2 219.34  109.67  5.3053 0.01139
brand:type   4 113.08   28.27  1.3676 0.27131
Residuals   27 558.15   20.67
```

Example 8.2.9 (cont'd)

```
> ## remove insignificant interaction term
> summary(fm8 <- aov(time ~ brand + type, computer))
           Df Sum Sq Mean Sq F value  Pr(>F)
brand       2  99.41   49.71  2.2956 0.11757
type        2 219.34  109.67  5.0651 0.01249
Residuals  31 671.23   21.65

> ## remove insignificant brand term
> summary(fm9 <- aov(time ~ type, computer))
           Df Sum Sq Mean Sq F value  Pr(>F)
type        2 219.35  109.67  4.6963 0.01604
Residuals  33 770.64   23.35
```

Example 8.2.9 (cont'd)

```
> TukeyHSD(fm9)

  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = time ~ type, data = computer)

$type
                                diff        lwr       upr       p adj
Expensive home-Inexpensive home 4.175 -0.6659665  9.015966 0.1019591
Business-Inexpensive home       5.875  1.0340335 10.715966 0.0145559
Business-Expensive home         1.700 -3.1409665  6.540966 0.6678201
```

Section 8.3 Only one observation per cell

- We must use an additive model when analyzing data from an unreplicated two-factor design, not because we know it is suitable but because a model with interactions does not provide an estimate of σ^2.

- The `particle` data, given in table 8.20, are an example of an unreplicated, two-factor experiment. The response, `amt`, is the amount of material left on a 40-mesh shaker table. The two continuous covariates, flow rate (`rate`) and `vacuum`, are observed at 3 levels each and are treated as factors. We store them as numeric values.

```
> str(particle)

'data.frame': 9 obs. of  3 variables:
 $ amt   : num  6.3 6.1 5.8 5.9 5.6 5.3 6.1 5.8 5.5
 $ rate  : num  85 90 95 85 90 95 85 90 95
 $ vacuum: num  20 20 20 22 22 22 24 24 24
```

Data plots for particle data
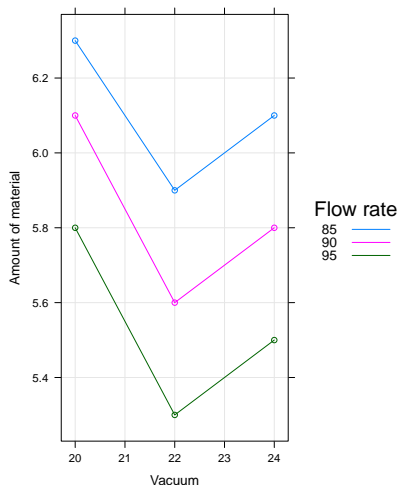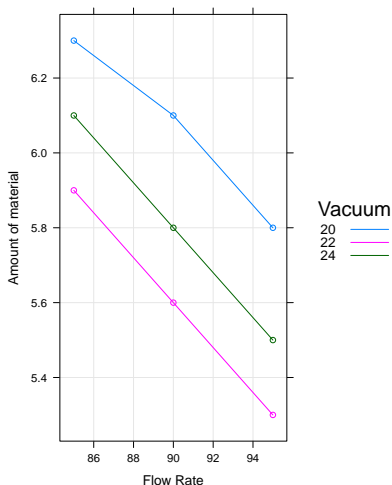
The data plots indicate a linear trend in flow rate but not in vacuum.

Additive model, both flow and vacuum as factors

```
> summary(fm10 <- aov(amt ~ factor(rate) + factor(vacuum), parti
                Df  Sum Sq Mean Sq F value     Pr(>F)
factor(rate)     2 0.48222 0.24111     217   8.34e-05
factor(vacuum)   2 0.32889 0.16444     148  0.0001778
Residuals        4 0.00444 0.00111
```

Additive model, both flow and vacuum as ordered

```
> part2 <- within(particle, {ora <- ordered(rate); ovac <- order
> anova(fm11 <- lm(amt ~ ora + ovac, part2))

Analysis of Variance Table
Response: amt
          Df  Sum Sq Mean Sq F value     Pr(>F)
ora        2 0.48222 0.24111     217   8.34e-05
ovac       2 0.32889 0.16444     148  0.0001778
Residuals  4 0.00444 0.00111
> summary(fm11)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.82222    0.01111 524.000 7.96e-11
ora.L       -0.40069    0.01925 -20.821 3.14e-05
ora.Q       -0.01361    0.01925  -0.707 0.518519
ovac.L      -0.18856    0.01925  -9.798 0.000608
ovac.Q       0.27217    0.01925  14.142 0.000145
Residual standard error: 0.03333 on 4 degrees of freedom
Multiple R-squared: 0.9946,  Adjusted R-squared: 0.9891
F-statistic: 182.5 on 4 and 4 DF,  p-value: 8.877e-05
```

Reduced model fit

```
> summary(fm12 <- aov(amt ~ rate + ovac, part2))
            Df  Sum Sq Mean Sq F value    Pr(>F)
rate         1 0.48167 0.48167  481.67 3.646e-06
ovac         2 0.32889 0.16444  164.44 2.744e-05
Residuals    5 0.00500 0.00100

> TukeyHSD(fm12, which = "ovac")

  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = amt ~ rate + ovac, data = part2)

$ovac
            diff        lwr        upr     p adj
22-20 -0.4666667 -0.5506823 -0.3826510 0.0000228
24-20 -0.2666667 -0.3506823 -0.1826510 0.0003450
24-22  0.2000000  0.1159844  0.2840156 0.0013440
```

Section 8.4: Blocking to reduce variability

- To reduce variability when comparing levels of an experimental factor, we sometimes run subgroups (blocks) under conditions where environmental factors are more homogeneous.

- We have already encountered an example of this as a *paired t-test*. The factor that determines the pairing (e.g. `Subject`) is called a *blocking factor*.

- For example, in the `uvcoatin` data, 10 subjects wore pairs of glasses in which one lens had the standard coating and one lens had the experimental coating. We create another data frame with the stacked responses and a blocking factor `subj`.

```
> str(uvstk <- within(stack(uvcoatin[,1:2]),
+                      subj <- gl(10,1,20,labels=LETTERS[1:10])))
```

```
'data.frame': 20 obs. of  3 variables:
 $ values: num  8.9 9.4 11.2 11.4 13 6.4 13.4 5.6 4.8 15.8 ...
 $ ind   : Factor w/ 2 levels "a","b": 1 1 1 1 1 1 1 1 1 1 ...
 $ subj  : Factor w/ 10 levels "A","B","C","D",..: 1 2 3 4 5 6 7 8 9 10
```

Paired t-test is equivalent to two-factor anova

```
> with(uvcoatin, t.test(a, b, paired = TRUE))

Paired t-test
data:  a and b
t = 2.8508, df = 9, p-value = 0.01907
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.04542425 0.39457575
sample estimates:
mean of the differences
               0.22

> summary(fm13 <- aov(values ~ subj + ind, uvstk))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
subj	9	241.802	26.867	902.2463	3.3e-12
ind	1	0.242	0.242	8.1269	0.01907
Residuals	9	0.268	0.030		

Notice that the p-values for the coating factor are the same in the two analysis. The F statistic for `ind` in the `aov` model is the square of the paired t-statistic.

Example 8.4.2

```
> str(deink2)

'data.frame': 15 obs. of  3 variables:
 $ formula : Factor w/ 3 levels "A","B","C": 1 1 1 1 1 2 2 2 2 2 ...
 $ newspaper: Factor w/ 5 levels "a","b","c","d",..: 1 2 3 4 5 1 2 3 4
 $ bright   : int  64 68 66 72 67 71 74 79 81 62 ...

> xtabs(bright ~ formula + newspaper, deink2)

        newspaper
formula  a  b  c  d  e
      A 64 68 66 72 67
      B 71 74 79 81 62
      C 85 87 81 76 81
```
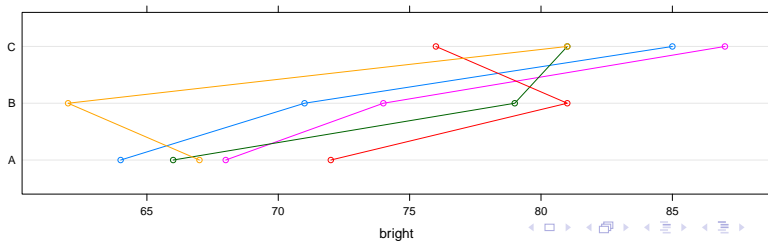
Model fit for example 8.4.2

- Whenever we have a blocking factor we list it first in the model formula. This doesn't affect the results for balanced designs but it will affect results from unbalanced data.
- We expect a blocking factor to be a significant source of variability (that's why we block on it) but we do not do any follow-up analysis for it.

```
> summary(fm14 <- aov(bright ~ newspaper + formula, deink2))
           Df Sum Sq Mean Sq F value   Pr(>F)
newspaper   4  86.27   21.57  0.7010 0.612724
formula     2 538.53  269.27  8.7519 0.009681
Residuals   8 246.13   30.77
```

Notice that the `newspaper` factor is *not* a significant source of variability here (or, more likely, the additive model is not suitable).
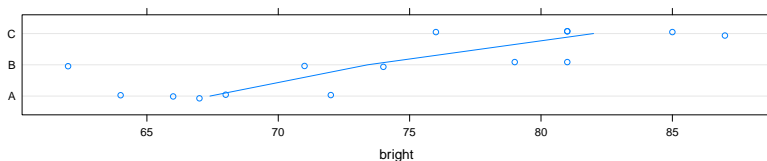
Re-fit
```
> summary(fm15 <- aov(bright ~ formula, deink2))
           Df Sum Sq Mean Sq F value   Pr(>F)
formula     2 538.53  269.27  9.7208 0.003091
Residuals  12 332.40   27.70
```

Multiple comparisons for de-ink methods

- We are now treating the `deink2` data as a one-factor design
- The significant difference is between levels A and C; B is intermediate and also shows greater variability.


```
> TukeyHSD(fm15)

  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = bright ~ formula, data = deink2)


$formula
     diff       lwr      upr    p adj
B-A   6.0 -2.8804201 14.88042 0.2102760
C-A  14.6  5.7195799 23.48042 0.0023594
C-B   8.6 -0.2804201 17.48042 0.0579518
```