Displays for Statistics 5303

Lecture 20

October 21, 2002

Christopher Bingham, Instructor

612-625-7023 (St. Paul)
612-625-1024 (Minneapolis)

Class Web Page
http://www.stat.umn.edu/~kb/classes/5303

Distribution of Mid term exam grades

```
n=23, Min=19, Q1=27.5, M=38, Q3=52.5, Max=69, ybar=40.3, s=15.7
    1     1|9
    8     2|0266789
  ( 4)    3|1148
   11     4|1247
    7     5|235
    4     6|0579
```

Grades were lower than I had hoped.

It was partly due to it being slightly too long, but mainly due to some students having apparent gaps in their understanding.

You did best on problem 4 (average 8.1/10) and worst on 3 (average 7.4/20). On problems 1 and 2 the averages were 10.5/25 and 14.3/25.

These averages include 0's for those who did not do some or all of a problem.

Although I reviewed most of the exam in class I am not including that here.

See the posted solutions to the exam

Contrasts and multiple comparisons for factorial data

With factorial experiments you have richer classes of contrasts to consider.

Main effect contrasts are contrasts among the main effect coefficients $\{\alpha_i\text{'s}\}$, $\{\beta_i\text{'s}\}$, $\{\gamma_i\text{'s}\}$, ...

For main effects, the analysis of contrasts and effects is almost the same as in the single factor case.

A **main effect contrast** for, say, factor A is $\sum_i w_i \alpha_i$, where $\{w_1,\ldots,w_a\}$ are weights with $\sum w_i = 0$.

You use **contrast()** to estimate and test $\sum_i w_i \alpha_i$ as for the single factor case.

And you can use, say **pairwise(a,.05, hsd:T)** to apply a multiple comparison method to compare factor levels of factor A.

For a two-way interaction contrast, you need a matrix (2 dimensional array) $\{w_{ij}\}$ that sums to 0 across rows and down columns ($\sum_i w_{ij} = 0$, $\sum_j w_{ij} = 0$). The value of an A×B interaction, say is $\sum_i \sum_j w_{ij} \alpha\beta_{ij}$.

For an k-way interaction contrast, you need a k dimensional array $\{w_{ij\ldots\ell}\}$ that sums to 0 when you sum along each subscript

$$(\sum_i w_{ij\ldots\ell} = \sum_j w_{ij\ldots\ell} = \ldots = \sum_\ell w_{ij\ldots\ell} = 0).$$

The value is $\sum_i \sum_j \ldots \sum_\ell w_{ij\ldots\ell} \alpha\beta\ldots\delta_{ij\ldots\ell}$ where $\{\alpha\beta\ldots\delta_{ij\ldots\ell}\}$ are the k-way interaction effects.

Polynomial contrasts with carbon wire data

```
Cmd> data <- read("")
carbonwire      72     4
) These are the carbon wire data from Anderson and Bancroft
) Col. 1: degassing treatment (2 levels)
) Col. 2: temperature treatment (2 levels, 2200F, 2350F, 2500F)
) Col. 3: Diffusion time treatment (3 levels, 4h, 8h and 12h)
) Col. 4: Response
) There were four replications per factor level combination.
) The response is the average resistivity of carbon wire in
) micro-ohms per cubic centimeter.
Read from file "TP1:Stat5303:Data:carbonwire.dat"

Cmd> makecols(data,degas,temp,time,y)

Cmd> degas <- factor(degas);time <- factor(time)

Cmd> temp <- factor(temp)

Cmd> anova("y=temp*time*degas",pvals:T)
Model used is y=temp*time*degas
                    DF          SS          MS     P-value
CONSTANT             1       36154       36154           0
temp                 2      410.69      205.35           0
time                 2      80.541       40.27           0
temp.time            4      14.814      3.7035  2.5659e-07
degas                1     0.46722     0.46722     0.21249
temp.degas           2     0.31361     0.15681     0.58919
time.degas           2     0.70194     0.35097     0.31036
temp.time.degas      4     0.87472     0.21868     0.56559
ERROR1              54       15.85     0.29352
```

The "shortcut" **"y=temp*time*degas"** "expands" to the terms in the order in the table, mixing main effects and interactions. Probably better is the shortcut **"y=(temp+time+degas)^3"**. This yields the same terms without such mixing.

Both time and temperature factors are quantitative and equally spaced, so it's natural to examine the linear and quadratic effects.

You can use the contrasts in D.6 because

- The levels (4h, 8h, 12h) of time are equally spaced as are the levels (2200F, 2350F, 2500F) of temperature
- The design is balanced so that all the means in a marginal table are calculated from the same number of cases.

You need to run **anova()** before you can use **contrast()**.

Table D.6 gives -1, 0, 1 and 1, -2, 1 as the linear and quadratic contrasts when there are three levels.

```
Cmd> contrast(temp,vector(-1,0,1)) #temp main effect linear
component: estimate
(1)        5.85
component: ss
(1)      410.67
component: se
(1)      0.1564
```

This is the linear effect of temperature averaged across 6 combinations of levels of the other two factors. The t-statistic t = 5.85/.1564 = 37.4 is extremely significant. The SS accounts for nearly all the temperature main effect sum of squares in the ANOVA table (410.69).

```
Cmd> contrast(temp,vector(1,-2,1)) #temp main effect quadratic
component: estimate
(1)      -0.075
component: ss
(1)      0.0225
component: se
(1)     0.27089
```

There is essentially no quadratic effect of temperature. The t-statistic t = -0.075/0.27089 = -0.277 is certainly not significant.

Note $SS_{lin} + SS_{quad}$ = 410.67 + 0.0225 = 410.69 = SS_{temp}, in the ANOVA table.

Contrasts in **time**:

```
Cmd> contrast(time,vector(-1,0,1)) # linear in time
component: estimate
(1)      2.4417
component: ss
(1)      71.541      Most of SS_time = 80.541
component: se
(1)      0.1564

Cmd> contrast(time,vector(1,-2,1)) # quadratic in time
component: estimate
(1)        -1.5
component: ss
(1)           9
component: se
(1)     0.27089
```

Again $SS_{lin} + SS_{quad}$ = 71.541 + 9.00 = 80.541 = SS_{time} in ANOVA table.

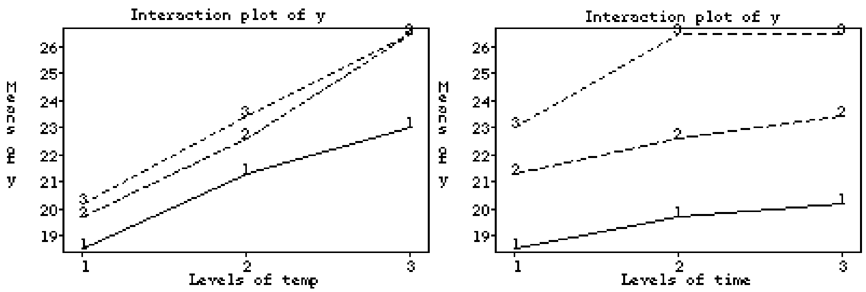
Both the linear and quadratic effects are significant (2.4417/0.1564 = 15.612, -1.5/0.27089 = -5.5373), so you have strong evidence that the effect of diffusion time is nonlinear.

Reminder: These main effect contrasts are averaging over all the combinations of levels of the other factors. This makes sense only if the value of the contrast doesn't depend on the levels of other factors.

Look at time by temperature interaction effects.

```
Cmd> interactplot(y,temp,time)

Cmd> interactplot(y,time,temp)
```



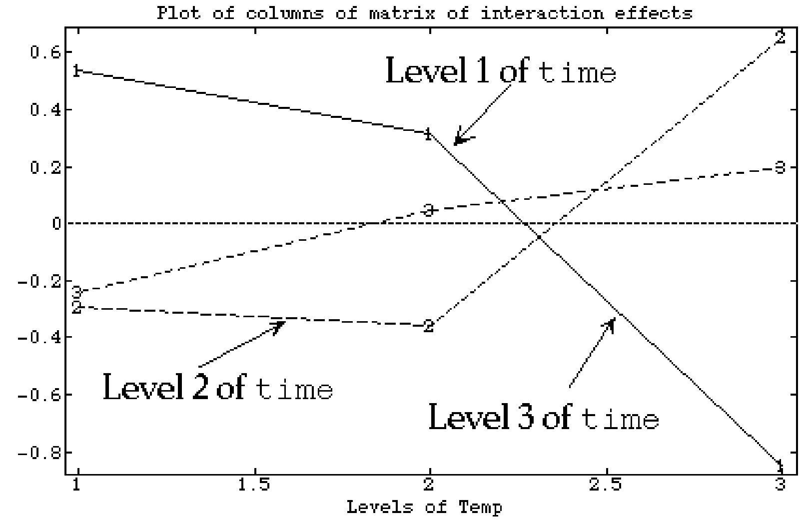
These are plots of means for each combination of levels of **temp** and **time**. It should be fairly clear why we found significant linear main effect contrasts, and why there was an apparent quadratic dependence on time. The lines are not very parallel, suggesting interaction of **time** and **temp**.

It's somewhat clearer when you look at the estimated interaction effects $\widehat{\alpha\beta}_{ij}$

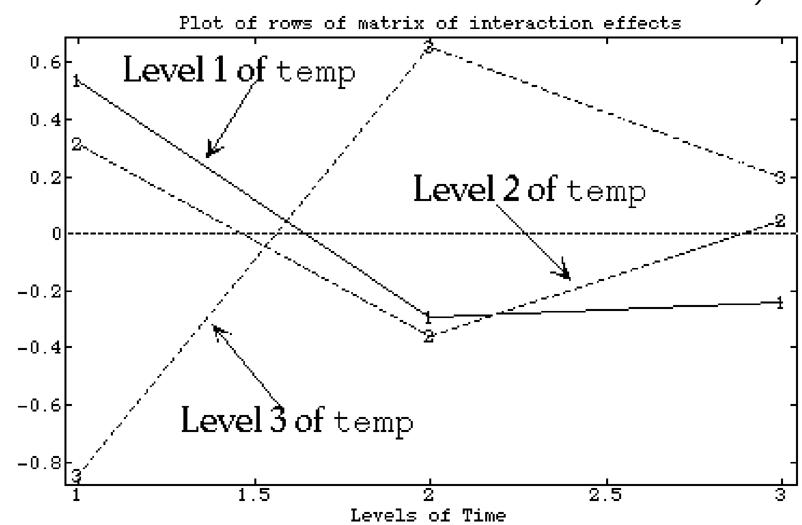
```
Cmd> coefs("temp.time") # interaction effect
(1,1)      0.5375    -0.29583    -0.24167
(2,1)      0.3125    -0.35833    0.045833
(3,1)       -0.85     0.65417     0.19583
```



```
Cmd> colplot(coefs("temp.time"),xlab:"Levels of Temp",\
 title:"Plot of columns of matrix of interaction effects")
```




```
Cmd> rowplot(coefs("temp.time"),xlab:"Levels of Time",\
 title:"Plot of rows of matrix of interaction effects")
```



To do interaction contrasts you need a *matrix* of interaction coefficients. One way to get them is to form the products of a row contrast and a column contrast.

You use **outer()** which generalizes to three and four factor interaction contrasts.

Linear by Linear

```
Cmd> outer(vector(-1,0,1),vector(-1,0,1)) # same thing
(1,1)          1          0         -1
(2,1)          0          0          0
(3,1)         -1          0          1

Cmd> contrast("temp.time",outer(vector(-1,0,1),vector(-1,0,1)))
component: estimate
(1)      1.825
component: ss
(1)     6.6612
component: se
(1)    0.38309
```

This is the linear by linear interaction contrast and it is highly significant (t = 1.8254/.38309 = 4.76.

But the SS = 6.6612 is much less than the over $SS_{time.temp}$ = 14.814, so there is a lot more interaction to "explain".

Linear in temp by quadratic in time

```
Cmd> outer(vector(-1,0,1),vector(1,-2,1)) # linear by quadratic
(1,1)         -1          2         -1
(2,1)          0          0          0
(3,1)          1         -2          1

Cmd> contrast("temp.time",outer(vector(-1,0,1),vector(1,-2,1)))
component: estimate
(1)      -2.85
component: ss
(1)      5.415
component: se
(1)    0.66353
```

t = -2.85/0.66353 = -4.30 is significant

Quadratic in temp by linear in time

```
Cmd> contrast("temp.time",outer(vector(1,-2,1),vector(-1,0,1)))
component: estimate
(1)        0.8
component: ss
(1)     0.42667
component: se
(1)     0.66353
```

t = 0.8/0.66353 = 1.21 is not significant

Quadratic by quadratic

```
Cmd> contrast("temp.time",outer(vector(1,-2,1),vector(1,-2,1)))
component: estimate
(1)     -3.225
component: ss
(1)     2.3112
component: se
(1)     1.1493
```

t = -3.225/1.1493 = -2.81, significant at the 1% level.