ST430: Introduction to Regression Analysis, Chapters 11-12, Case Study 7

Luo Xiao

November 30, 2015
Design and Analysis of Experiments
Design of experiments

The linear regression model relates the expected value of a dependent variable, or *response*, $E(Y)$, to some independent variables, or *factors*, $X_1, X_2, \ldots$

- In an *observational* study, we have no control over the values of the factors.
- In an *experimental* study, we can control them.

*Designing* the experiment includes:

- Choosing the *levels* of each factor;
- Deciding how many trials to carry out;
- Choosing the *treatment* (levels of the factors) in each trial;
- Blocking and randomizing the trials.
Example 11.1

A marketing study on weekly coffee sales:

1. Select the factors: brand and shelf location;
2. Choose levels of each factor:
   - Two brands (A and B);
   - Three shelf locations (bottom, middle, and top);
3. Choose the treatments:
   - *Complete factorial* design uses all 6 combinations of factor levels;
4. Decide sample size for each treatment (3, in the example);
5. Assign treatments to *experimental units* (weeks 1 to 18).
Completely random design (CRD)

To avoid the possibility that small time trends might be confused with factor effects, randomize the order of the treatments (see "output1.txt"):

```r
design = expand.grid(Brand = c("A", "B"), 
                      Shelf = c("bottom", "middle", "top"), 
                      Rep = 1:3)

design[sample(1:18),1:2]
```
Blocking
Suppose that weeks 1–6 were at one store, weeks 7–12 at another, and weeks 13–18 at a third.

Complete randomization might put brand A more often at store 1, which might have higher overall coffee sales.

Randomized complete block design (RCBD)
Assign all 6 treatments once to each block (store), and randomize the order within blocks (see "output2.txt"):

\[
\text{design} = \text{expand.grid}(\text{Brand} = \text{c("A", "B")}, \\
\text{Shelf} = \text{c("bottom", "middle", "top")}, \\
\text{Rep} = 1:3) \\
\text{design[c(sample(1:6), sample(7:12), sample(13:18))],}
\]

Fractional factorial designs

Sometimes many candidate factors need to be evaluated.

If \( k \) factors have been identified, and each is used at only 2 levels, there are \( 2^k \) treatments.

If we do not have the resources to carry out even a single replicate of the complete \( 2^k \) design, we use a carefully chosen \( fraction \) of the treatments.
One quarter fraction of the $2^5$ factorial ($2^{5-2}$)

Output on next slide:

```r
design = expand.grid(A = c(-1,1), B = c(-1,1), C = c(-1,1))
design$D = design$A * design$B
design$E = design$A * design$C
design
```
# Design table from previous slide

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Analysis of experimental data

In a designed experiment, we may have both qualitative and quantitative factors.

Usually all factors have relatively few levels, so treating them all as qualitative is feasible.

We can use the same tools to analyze the data as for more general regression models.
Example 12.4

Grade point averages for students from different socioeconomic groups; a single factor.

See “output3.txt”:

```r
setwd("~/Dropbox/teaching/2015Fall/R_datasets/Exercises&Examples")
load("GPA3.RData")

plot(GPA ~ CLASS, GPA3) #scatter plot
anova(lm(GPA ~ CLASS, GPA3)) #usual approach
summary(aov(GPA ~ CLASS, GPA3)) #another approach
#adjust for multiple comparisons of means
TukeyHSD(aov(GPA ~ CLASS, GPA3))
```
Box plot
Case Study 7
Reluctance to transmit bad news: the MUM effect

Psychologists found that people are reluctant to transmit bad news “to peers in a nonprofessional setting.”

They called it the MUM effect.
An experiment

Forty subjects (Duke undergrads) each administered an IQ test to a test-taker, and informed the test-taker of the percentile score.

Two experimental factors:

subject visibility: Some subjects were told the test-taker could see them, some were told the test-taker could not see them.

confederate success: The test-taker was given one of two answer keys; one placed the test-taker in the top 20% of Duke undergrads, the other in the bottom 20%.

The response is the latency to feedback: time between the end of the test and delivery of the result.
Completely randomized $2 \times 2$ factorial design

The subjects were randomly assigned to 4 groups of 10 subjects.

```r
setwd("~/Dropbox/teaching/2015Fall/R_datasets/Cases")
load("MUM.RData")

with(MUM, interaction.plot(CONFED, SUBJECT, FEEDBACK))
```

The interaction plot (next slide) suggest that the factors interact, so we will fit the complete model, which includes their interaction, to test whether this graphical indication is supported.
Interaction plot

![Interaction plot diagram](image)

- **CONFED**: x-axis
- **mean of FEEDBACK**: y-axis
- **SUBJECT**: NotVis (dashed line), Visible (solid line)
- **Failure**
- **Success**

Case Study 7
The complete model (two qualitative factors):

\[ E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2, \]

where \( X_1 \) and \( X_2 \) are indicator variables for subject visibility and confederate outcome.

Analysis of variance:

\[ \text{summary}(\text{aov}(\text{FEEDBACK} \sim \text{SUBJECT} \times \text{CONFED}, \text{MUM})) \]
## ANOVA table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUBJECT</td>
<td>1</td>
<td>8381</td>
<td>8381</td>
<td>25.5</td>
<td>1.29e-05 ***</td>
</tr>
<tr>
<td>CONFED</td>
<td>1</td>
<td>8151</td>
<td>8151</td>
<td>24.8</td>
<td>1.60e-05 ***</td>
</tr>
<tr>
<td>SUBJECT:CONFED</td>
<td>1</td>
<td>20839</td>
<td>20839</td>
<td>63.4</td>
<td>1.87e-09 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>36</td>
<td>11834</td>
<td>329</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
The output confirms that the interaction term is important, supporting the indication from the interaction plots.

The interaction plots also suggested that three of the treatment means are about equal, and only the fourth is different.

Use Tukey’s Honest Significant Difference (HSD) for comparing multiple means (see "output4.txt"):

```R
TukeyHSD(aov(FEEDBACK ~ SUBJECT*CONFED, MUM))
```
Note that "Visible:Failure" is significantly higher than "NotVis:Failure", "NotVis:Success", and "Visible:Success".

Also, there are no significant differences among "NotVis:Failure", "NotVis:Success", and "Visible:Success".

See the treatment means:

```
with(MUM, tapply(FEEDBACK, SUBJECT:CONFED, mean))
```
Conclusion

The subjects took essentially the same amount of time to pass the result to the test-taker in all situations, except that they took significantly longer, roughly twice as long, to inform the test-taker about a poor performance when visible to the test-taker.