Analysis of Variance

Preliminaries

Project is due today at 4 p.m.

Office hours

– Today 3-5:30

Assignment #3

– Postponed until Thursday, 4 p.m.

– Group project, turn in one for the team (as on assignment #1)

Final exam review sessions

– In the coming weekend, with details to be announced.

Conjoint Analysis

We'll do one of these interesting and popular marketing experiments in class…

– Goal of a conjoint analysis is to determine how combinations of product attributes (a "bundle") appeal to a consumer.

– Interactions of the attributes are key:
Which combination offers the most appealing bundle to the consumer?

– Manipulate several factors at once rather than have consumer rate each attribute separately. Latter would not get at the interactions of the attributes.

Methods for conjoint analysis

– Remaining notes cover the statistical methods that one uses to analyze data from a conjoint analysis.

– Collectively, these methods are known as the "analysis of variance".
In spite of the name, these methods focus on the comparisons of the average response under different conditions (combinations of categorical factors).

Comparing Several Averages

Key question of interest

– Are the averages of different groups of observations significantly different, and if so, which are different?

– Extension of the t-test to more than two groups. We saw these ideas back in Statistics 603 in August.

Design

– The groups of observations are defined by one or more categorical factors.

– Groups may be defined in two ways…

• *Unstructured*, as levels of a single categorical factor with no pattern to how the groups are defined. These analyses give what is known as a one-way analysis of variance.

• *Structured*, as combinations of several categorical factors that generate a table-like pattern to the definitions of the categories, as in the following 3 by 3 table that defines a two-way analysis of variance…

Color\Font	Bold	Italic	Plain
Red			
Green			
Blue			

Experiments

– Data collection strategy focuses on getting data that facilitates comparison of the groups.

– Randomization and "balance" (i.e., equal numbers of subjects for each combination) are critical.

– Statistical analysis is "routine" – good data lead to a very straightforward analysis in comparison to the collinearity and subjectivity of regression as used in the project.

What is the process for performing an analysis of variance?

– For a *one-way* anova, the key steps (as seen in Stat 603) are …
(0) Are the groups comparable, or is confounding likely/possible?
(1) Check the overall F statistic.
Be very cautious going forward if it is not significant.
(2) Compare the mean values using the Tukey-Kramer procedure.
(3) Check assumptions, particularly the assumption of equal variance.

– For *two-way* anova (and higher, though we do not go there), the key steps are quite similar and are those we have seen from regression…
(0) As above, think about the presence of confounding.
(1) Check the partial F-test (Effect Test) for the interaction term.
(2) If partial the F-test for the interaction is NOT significant,
Consider marginal comparisons of the levels (e.g., color, font)
If the partial F test for the interaction IS significant,
Beware interpreting other effects. Check profile plots.
(3) Check assumptions

– Review examples from regression casebook
"Class 9" for one-way comparisons
"Class 10" for two-way comparisons (summarized below)

Casebook Examples

Package design experiment Design.jmp, page 255
(NO interaction)

"What is the right combination of color and lettering style?"

Data

12 stores each of Red/Green and Block/Script

Are stores really comparable? Possible confounding?

Analysis

No interaction

Two-way analysis of variance has no significant interaction. (page 256)
Only type of lettering style has significant effect.

Collinearity

VIF terms (not shown in text) exactly equal to zero for table of slopes on page 256. In a "balanced" experiment such as this, no collinearity clouds the issue. *Good data makes the analysis straightforward.* (Compare to the situation with Rush and Detail).

Profile plots

Profile plot shows parallel lines for the groups, little interaction. (p 258)
Obtain these from the leverage plot view of the factor (plot "LS Means")

Diagnostics

Residual diagnostics indicate no problems (page 260).

Conclude

Red script gained the highest sales, but only significantly better than green block so that you might choose another combination. Perhaps the best course of action would be to run further tests, avoiding green block.

The lack of interaction implies a simplified decision-making process. One group can pick the color, another can pick the type style. (Noting that color has only a very "modest" effect).

Evaluating employee time schedules (Lots of interaction) Flextime.jmp, page 262

"What type of scheduling is best for clerical staff in various depts?"

Data

Productivity gains of 33 staff members from each department.

Independent observations?

Does any confounding contaminate the experiment? For example, are staff aware/bothered by scheduling of others?

Analysis

Marginal analyses fail

Initial one-way analyses show no differences for the effect of schedule (p 263) unless we restrict our attention within departments (p 264, 265)

Significant, clear interaction

Two-way anova with interaction shows clear differences and a significant interaction. That is, the effects of the different office schedules vary by department. The DP group, e.g., prefers the flextime whereas the investment group prefers the 4-day schedule.

Profile plots

Profile plots are striking (p 267). The lack of parallel lines shows the effect of the interaction.

Conclude

The best schedule depends on the department (i.e., there's a lot of interaction). Flextime is best in DP, with a four-day arrangement being best in Investments. The schedule has little effect on productivity in the claims group.

Key Take-Away Points

Conjoint analysis

– Method of collecting data for an analysis of variance with several factors.

– Subjects in an experiment rate "bundles" of attributes.

– Anova is used to figure out which attributes are preferred and how they are related to one another.

Two-way anova

– Key use for a two-way anova: recognize interaction

– Interaction represents a synergy of the two factors.

– Profile plots reveal the interaction visually.

– Table organization as a metaphor that simplifies

– Analysis is straightforward
(after all, it is regression with just two uncorrelated categorical predictors)