STAT 350: Lecture 25

Interaction effects

- If a model contains terms $\beta_u U + \beta_v V$ then we define the UV interaction term to be $\beta_{uv} UV$, that is, we define a new column of the design matrix which is the product of the U column and the V column.
- Something analogous can be done if one of the variables (or both) is categorical.
- A factor at K levels will have $K-1$ columns in the design matrix.
- We create an interaction between this factor and a continuous variable X by adding $K-1$ new columns obtained by multiplying the X column by each of the $K-1$ columns.
- We create an interaction between a factor at K_1 levels and another at K_2 levels by multiplying each of the $K_1 - 1$ columns corresponding to the first factor by each of the $K_2 - 1$ columns of the second factor.
- Three way interactions are obtained by multiplying columns corresponding to 3 covariates, and so on for higher way interactions.
- An interaction term between a continuous covariate and a categorical factor amounts to allowing the slope corresponding to the continuous covariate to depend on the level of the categorical factor.

Examples

Two way analysis of variance

- Experiment involving 2 diets, 3 drugs and 2 experimental units per combination – a factorial experiment.
- $Y_{i,j,k}$ is response for Diet i, Drug j and replicate k.
- Additive model is

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \epsilon_{i,j,k}$$

- Impose restriction $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 + \beta_3 = 0$.
  DESIGN MATRIX

$$X_a = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

- First column corresponds to μ, the grand mean. Second column to α_1 and uses $\alpha_2 = -\alpha_1$. Third and fourth columns for β_1 and β_2 use $\beta_3 = -\beta_1 - \beta_2$.

- To put in interaction terms multiply column 2 by 3 and 2 by 4 to get

$$X_b = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

  This design matrix corresponds to a model equation

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \lambda_{i,j} + \epsilon_{i,j,k}$$

  where the $\lambda_{i,j}$ are interaction effects satisfying $\sum_i \lambda_{i,j} = \sum_j \lambda_{i,j} = 0$.

Analysis of Covariance

This is the name given to the analysis of models in which there are categorical factors and continuous covariates. In the car example we had the categorical factor VEHICLE and the continuous covariate MILEAGE. Earlier I gave the design matrix for the model in which there are different intercepts for the two cars but 1 common slope. thus this model is 2 parallel lines. If we use corner point coding and fit a model in which VEHICLE and MILEAGE interact then the design matrix for the small data set above is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1000 & 0 \\ 1 & 0 & 2000 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1100 & 1100 \end{bmatrix}$$

where the last column is the product of columns 2 and 3. The design matrix corresponds to a model equation with some slope β_1 for the first vehicle and a slope $\beta_1+\gamma_2$ for the second vehicle. That is, the coefficient of the last column of the design matrix is the difference in slopes between the 2 vehicles. Use of the alternative coding based on an average intercept leads now to the design matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1000 & 1000 \\ 1 & 1 & 2000 & 2000 \\ 1 & -\frac{3}{2} & 0 & 0 \\ 1 & -\frac{3}{2} & 1100 & -1650 \end{bmatrix}$$

Again the last column is the product of columns 2 and 3. the coefficient of column 3 is the average slope while the coefficient of the last column is the difference between the slope for vehicle 1 and this average slope.

You saw, in assignment 3, how to test the hypothesis of no interaction in this model.

Analysis of Models with Interaction Terms

- Usually: use F-tests to try to eliminate higher order interactions (i.e., interactions involving several variables).
- Then test for lower order interactions.

- To use SAS: put interaction terms in last. Use Type I or sequential SS.
- Do not consider models involving a two factor interaction unless individual (also called **main**) effects included.
- Similarly SS for 3 factor interaction normally adjusted for
  - all 3 main effects
  - all 3 two variable interactions of the 3 variables.

Examples

Two way ANOVA: influence of SCHOOL, REGION on STAY

```
options pagesize=60 linesize=80;
data scenic;
 infile 'scenic.dat' firstobs=2;
 input Stay  Age Risk Culture Chest Beds
          School Region Census Nurses Facil;
proc glm  data=scenic;
  class school region ;
  model Stay = School | Region / E
          SOLUTION SS1 SS2 SS3 SS4 XPX INVERSE;
  output out=scout P=Fitted PRESS=PRESS H=HAT
   RSTUDENT =EXTST R=RESID DFFITS=DFFITS COOKD=COOKD;
run ;
proc means data=scout;
  var stay;
  class school region;
run;
proc print data=scout;
```

EDITED SAS OUTPUT