Unbalanced Two-Way ANOVA

Applied Regression and Other Multivariable Methods
Sections 20-1 - 20-4

Introduction

- Topics 21 & 22 assumed complete, balanced design
  - **Balanced** - equal # of observation per cell
  - **Complete** - at least one observation per cell
  - This results in orthogonal model factors
  - Notice Type I and Type III SS are the same
- In practice, will rarely end up with such a design
- For this topic, still assume complete design
- Will allow design to become unbalanced
  - Will lose orthogonality
  - Type I and Type III SS will differ

Example I - KKMN 20-1

In a study concerning the effects of rapid cultural change, an experimenter collected the blood pressure on 30 randomly sampled males over 40 years of age. Each person also has their social status ranked based on the modern and traditional cultures.

	Traditional Rank		
Modern Rank	HIGH	MED	LOW
HIGH	130, 140, 135	150, 145	175, 160, 170, 165, 155
MED	145, 140, 150	150, 160, 155	165, 155, 165, 170, 160
LOW	180, 160, 145	155, 140, 135	125, 130, 110

Because the people were randomly sampled, it would be very unlikely to have an equal number of subjects per cell.

In regression, each variable is assessed by fitting it last in the model. When a two-way ANOVA is unbalanced, we will assess each factor similarly.

First compute F(Interaction | Factor 1 and 2)

If interaction is not significant, will perform tests

F(Factor 1 | Factor 2, Interaction)

F(Factor 2 | Factor 1, Interaction)


```
options nocenter ls=75;

data problem1;
 infile 'i:\.www\datasets502\EX2001.DAT' firstobs=2 dlm='09'x;
 input modern $ trad $ y;
 mod = 3; if modern = 'HI' then mod = 1; if modern = 'MED' then mod = 2;
 trd = 3; if trad = 'HI' then trd = 1; if trad = 'MED' then trd = 2;

proc print;

run;
quit;
```

```
                                 Sum of
Source                DF       Squares   Mean Square  F Value  Pr > F
Model                  8   5744.166667    718.020833     9.47  <.0001
Error                 21   1592.500000     75.833333
Corrected Total       29   7336.666667

R-Square     Coeff Var      Root MSE        y Mean
0.782940      5.779801      8.708234      150.6667

Source                DF     Type I SS   Mean Square  F Value  Pr > F
mod                    2    977.702020    488.851010     6.45  0.0066
trd                    2    195.522423     97.761211     1.29  0.2964
mod*trd                4   4570.942224   1142.735556    15.07  <.0001

Source                DF   Type III SS   Mean Square  F Value  Pr > F
mod                    2    708.453895    354.226948     4.67  0.0210
trd                    2     37.049731     18.524866     0.24  0.7855
mod*trd                4   4570.942224   1142.735556    15.07  <.0001
```