STAT 5200 Handout #21

Split Split Plot Example (Ch. 16)

<u>Build off the Split-Plot Design</u>: (see Handouts #20 and #20a)

We can add a third treatment factor and split the split-plot units into split-split-plot units, and add a third randomization at this level.

In principle we could add even more treatment factors and further splits, with randomizations at these deeper levels.

Why do this? Same reason as for split-plot design: Some trt. factors are easier to vary than others

<u>Example: Management / Grass / Legume</u>

An experiment was carried out to evaluate the effects of 3 factors on the total dry matter yield of animal feed. The experimental factors were:

- **Management**
  2 levels (<u>Early</u> first cutting and <u>Late</u> first cutting). This is a FIXED factor.

- **Grass**
  6 levels (KBG, MB, OG, PRB, PRM, and TF). This is a FIXED factor.

- **Legume**
  5 levels (None, ALF, BFT, CM, and WC). This is a FIXED factor.
  The factor level "None" is the "Control" for this factor.

One purpose of the experiment was to determine if growing legumes and grass on the same plots could <u>increase the total yield</u> over just growing grass.

The design of the experiment was as follows. Five fields were available for the experiment. Each field was subdivided into 2 large plots and the levels of **Management** randomly assigned to the 2 plots. (So 5 randomizations done here – one for each field.)

Next, each large plot in each field was subdivided into 6 small plots and the levels of **Grass** were randomly assigned to the small plots. Ten distinct randomizations took place at this stage: one for each large plot in each field ($2 \times 5 = 10$).

Finally, the small plots in each field were further subdivided into 5 smaller plots and the levels of **Legume** randomly assigned to these smaller plots. At this level of the experiment $6 \times 2 \times 5 = 60$ different randomizations were carried out (one for each small plot in each large plot in each field).

The response is the total dry matter yield of the smaller plots, one measurement per plot.

large plot #1 large plot #2

One field:

KBG	PRB	PRB	MB
OG	MB	TF	KBG
TF	SRM	OG	SRM

early management late management

"Zoom in" to see the splits in one small plot:

ALF	None	WC	BFT	CM

This is a **Split-Split-Plot Design**:

- One **Blocking Factor**: R = Rep = Field
- Three treatment factors (each with their own size of unit, and a different randomization):
  - M = Management (**Whole-Plot Factor**)
    - Randomized to large plots in field

      whole plot units

      in a Randomized complete block design
  - G = Grass (**Split-Plot Factor**)
    - Randomized to small plots (within large plots in field)

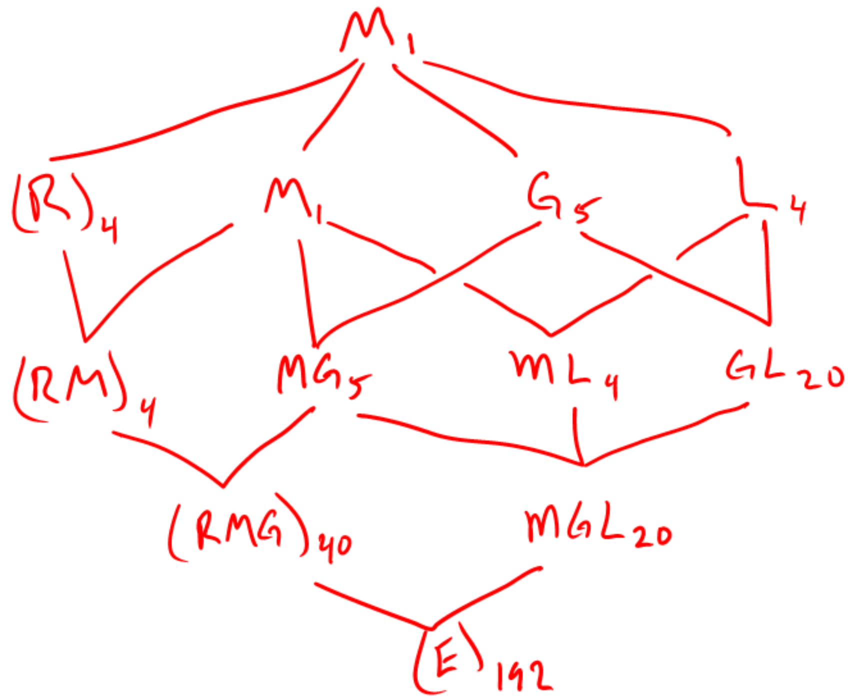
      split-plot units
  - L = Legume (**Split-Split-Plot Factor**)
    - Randomized to smaller plots (within small plots within large plots in field)

      split-split plot units

easier to vary

The analysis is similar to that for split plot designs but with an additional factor. Although the randomization of the **Grass** and **Legume** factors are nested, the factors themselves are not. The treatment structure is a full factorial in the three treatment factors and the blocks.

Hasse Diagram for Management/Grass/Legume Example: Factors: R, M, G, L



- SPE term (RMG) is actually R × G(M) due to design's nested randomization; its DF account for this nesting:

  $DF_R \times DF_G \times$ (# levels M) = 4 × 5 × 2 = 40

- SSPE term (E) is actually R × L(M × G) due to design's [double] nested randomization; its DF account for this nesting:

  $DF_R \times DF_L \times$ (# levels M) × (# levels G) = 4 × 4 × 2 × 6 = 192

Split-Split-Plot Model for Management/Grass/Legume Example:

$$Y_{ijkl} = \mu + R_i + M_j + RM_{ij}$$ } whole-plot

$i = 1..5$ $\quad j = 1,2$

$$+ G_k + MG_{jk} + RMG_{ijk}$$ } split-plot

$k = 1..6$

$$+ L_l + ML_{jl} + GL_{kl} + MGL_{jkl} + \varepsilon_{ijkl}$$ } split-split plot

$l = 1..5$

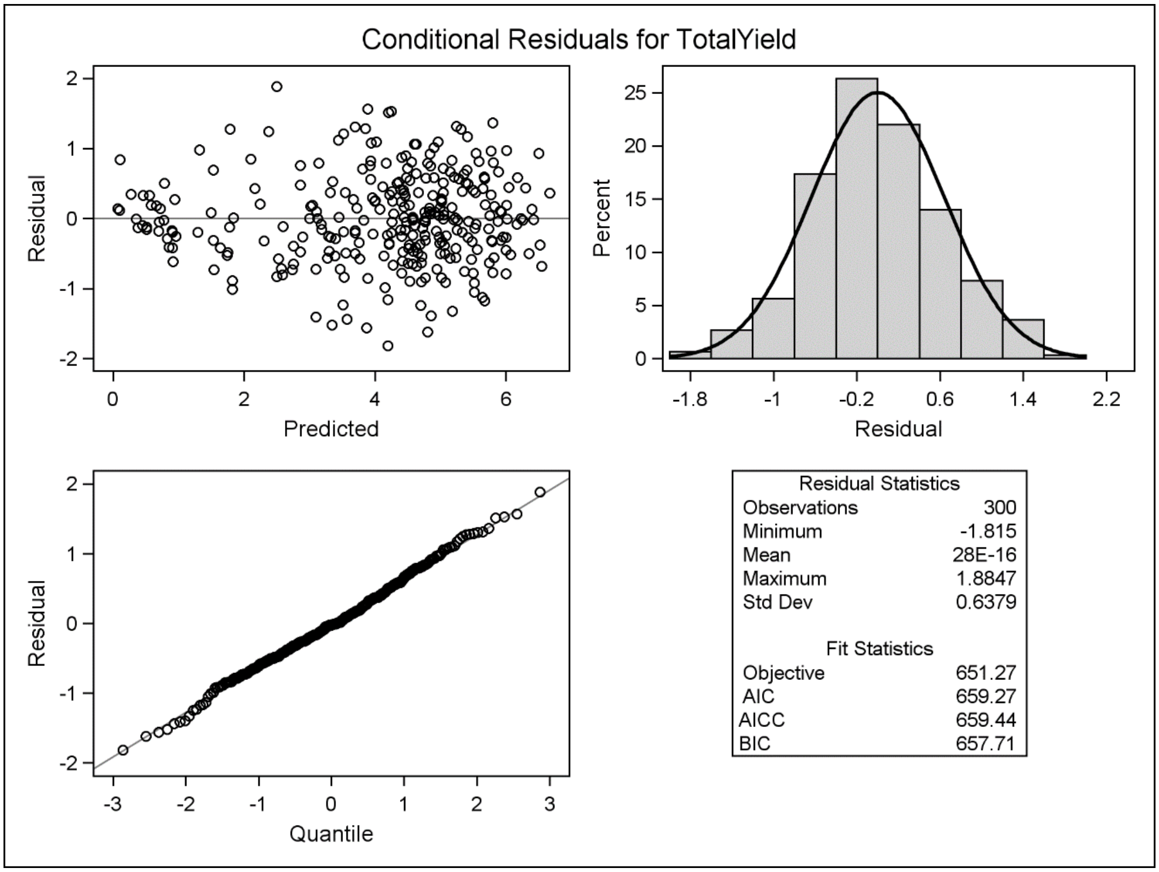
Data in a comma-delimited (.csv) file; can open in Excel:

	Mixture	Management	Rep	TotalYield	Grass	Legume
1	Mixture	Management	Rep	TotalYield	Grass	Legume
2	KBG	earl	1	1.291336	KBG	None
3	KBG	earl	2	0.298822	KBG	None
4	KBG	earl	3	0.789743	KBG	None
5	KBG	earl	4	0.725923	KBG	None
6	KBG	earl	5	0.654633	KBG	None
7	KBG+ALF	earl	1	4.703025	KBG	ALF
8	KBG+ALF	earl	2	4.15917	KBG	ALF

```
/* STAT 5200
   split-split plot design
   management grass legume data
*/

/* Define options */
ods html image_dpi=300 style=journal;

/* Read in data */
proc import out=work.MGL
datafile="C:\...\ManagementGrassLegumeAll.csv"
    dbms=csv replace;
 getnames=yes;
run;

/* Fit split-split plot model & generate interaction plots.
   Recall: SLICEBY tells which variable to use
           as plotting characters,
   and PLOTBY tells which variable to separate on
           for three-way interaction plots */
proc glimmix data = MGL plot=residualpanel;
  class Grass Legume Management Rep;
  model TotalYield = Management | Grass | Legume /
        ddfm=satterthwaite;
  random Rep Rep*Management Rep*Management*Grass;
  covtest / wald; /* requests tests for var components */
  lsmeans Management Grass Legume /
          pdiff=all adjust=Tukey lines;
  lsmeans Grass*Legume /
          plot=mean(sliceby=Legume join);
  lsmeans Management*Grass /
          plot=mean(sliceby=Management join);
  lsmeans Management*Grass*Legume /
       plot=mean(sliceby=Legume join plotby=Management);
  title1 'Split-Split Plot Model';
run;
```

Split-Split Plot Model

The GLIMMIX Procedure



Class Level Information		
Class	**Levels**	**Values**
Grass	6	KBG MB OG PRB PRM TF
Legume	5	ALF BFT CM None WC
Management	2	earl late
Rep	5	1 2 3 4 5

Convergence criterion (ABSGCONV=0.00001) satisfied.

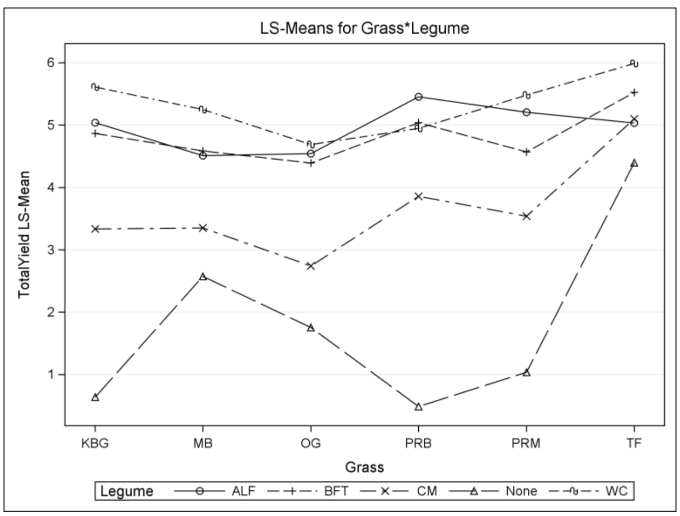
Covariance Parameter Estimates				
Cov Parm	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**
Rep	0.05070	0.04951	1.02	0.1529
Management*Rep	0.01081	0.02486	0.43	0.3318
Grass*Management*Rep	0.03325	0.03346	0.99	0.1602
Residual	0.5402	0.05514	9.80	<.0001

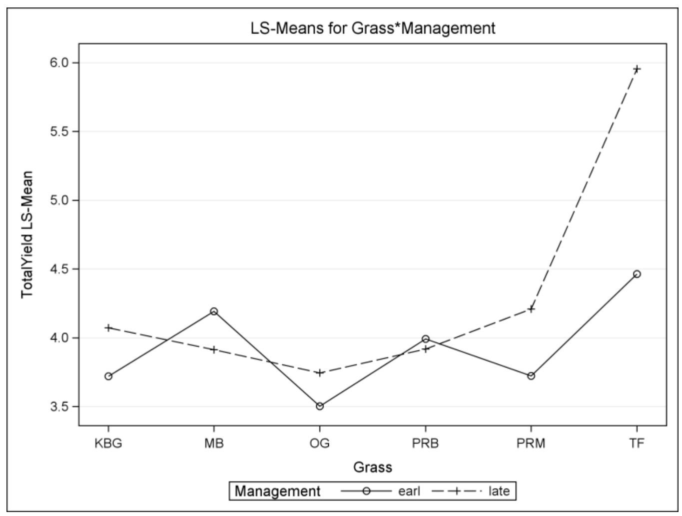
Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Management	1	4	9.98	0.0342
Grass	5	40	21.77	<.0001
Grass*Management	5	40	6.74	0.0001
Legume	4	192	227.93	<.0001
Legume*Management	4	192	1.86	0.1188
Grass*Legume	20	192	8.46	<.0001
Grass*Legume*Managem	20	192	1.14	0.3155

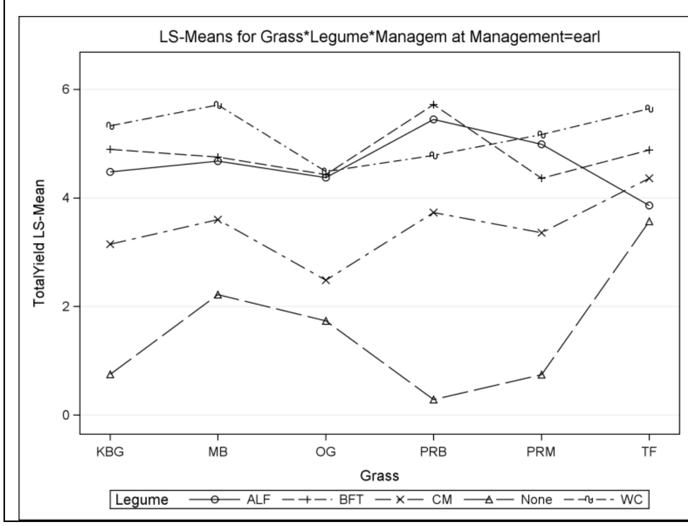
Tukey-Kramer Grouping for Management Least Squares Means (Alpha=0.05)		
LS-means with the same letter are not significantly different.		
Management	**Estimate**	
late	4.3034	A
earl	3.9330	B

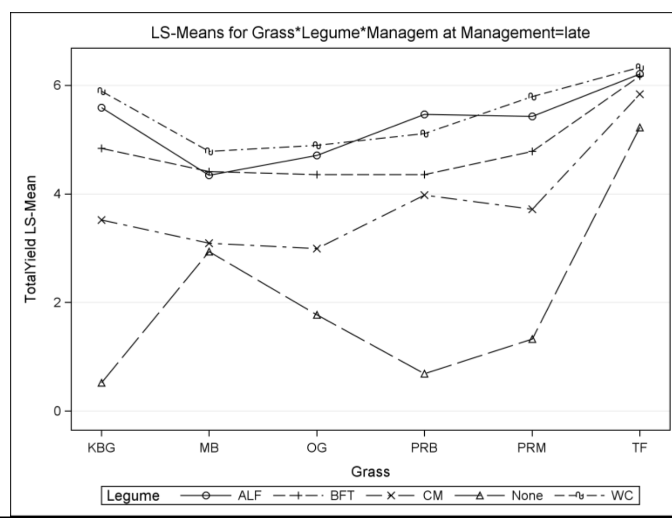
Tukey-Kramer Grouping for Grass Least Squares Means (Alpha=0.05)		
LS-means with the same letter are not significantly different.		
Grass	**Estimate**	
TF	5.2101	A
MB	4.0534	B
PRM	3.9673	B
PRB	3.9565	B
KBG	3.8977	B
OG	3.6242	B

Tukey-Kramer Grouping for Legume Least Squares Means (Alpha=0.05)			
LS-means with the same letter are not significantly different.			
Legume	**Estimate**		
WC	5.3277		A
ALF	4.9650	B	A
BFT	4.8297	B	
CM	3.6539		C
None	1.8146		D









Some Discussion Points for this Example:

1. Note that the variance component for the Residual is HUGE (.5402) compared to the variance component for the Reps (.0507), and the other two variance components. (So there is minimal variation between fields.)

2. The three-way interaction is highly non-significant (p-value .3155; reflected by the striking similarity in the last two interaction plots above).

3. The Legume-Management interaction is non-significant (p-value .1188), but the Grass-Legume and Grass-Management interactions are highly statistically significant (both p-values < .001), and so we must characterize them with interaction plots (the first two interaction plots above).

4. The Grass-Management interaction is easy to understand and explain. Its statistical significance is driven by the unusually high yield for the TF—Late First Cutting combination, which is a good 33% higher ($6/4.5 \approx 1.33$) than any other combination. (That's big!)

5. The Grass-Legume interaction is, well, a mess! Some things are immediately apparent on the interaction plots. For example, Legume = "None" seems to be the worst option (in terms of yield) across the board. Legume = "CM" is the next lowest for most species of grass. The remaining types of Legume result in similar yields, but it looks like Legume = "WC" has the highest yield, or close to the highest yield, for all the Grass species.

   With regards to Grass types, it looks like "TF" ("Tall Fescue", perhaps?) has the highest yield or close to the highest yield for all types of Legume. So, a good combination of factors levels is Grass = "TF", Legume = "WC", and Management = "Late First Cutting".

   There may be other combinations that have similar yields and are not statistically significantly different from this combination. We could check this in 3-way LSMEANS statement (not included here just to avoid lengthy output):

```
lsmeans Management*Grass*Legume /
                      pdiff=all adjust=Tukey lines;
```