Establishing a Foundational Understanding of Model Construction in an Intermediate Statistics Course

Julie Garai: *University of the South* | **Marina Ptukhina:** *Whitman College* | **Walt Stroup:** *University of Nebraska-Lincoln*

Lost in Translation

Goals of What Would Fisher Do (WWFD)

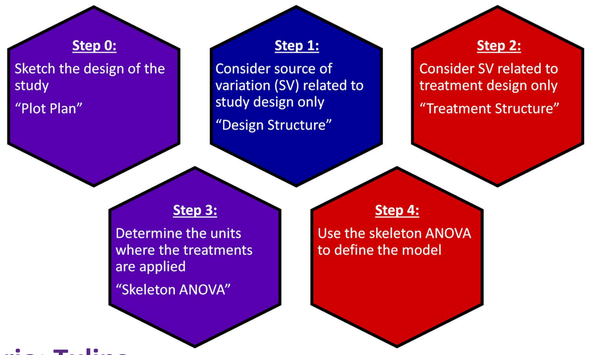
- Help students differentiate between scenario components
  - Design Structure
  - Treatment Structure
- Help students relate these structures together to properly partition variability
- Help students to model data as it naturally arises

General Form of Model

Current practice[1]: *Data = Model + Error*

Proposed alternative[2]: *Data = Treatment Structure + Design Structure + Error*

WWFD Process



Scenario: Tulips

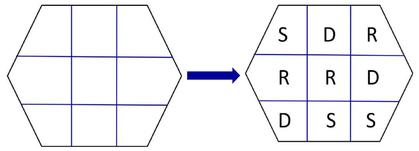
We are interested in creating a tulip display for a competition next year. One of the primary categories for victory is the height of the tulips, thus we need to determine which tulip variety to plant in order to win! We have:

- 3 tulip varieties (Double, Rembrandt and Single)- i.e. 3 levels of 1 treatment
- A plot of land divided into 9 sections- i.e. 9 experimental units

We randomize the tulip varieties to each section, thus we have 3 replications of each variety.

Step 0: Plot Plan

- Begin by sketching the study design
  - Plot of land split into 9 sections
- Randomly assign tulip variety to each section
  - Double (D), Rembrandt (R), Single (S)



Step 1: Design Structure

- Identify sources of variation related to study design
  - 9 Sections = 8 degrees of freedom

Design Structure	
Source	**df**
Sections	9-1=**8**
Total	9-1=8

Step 2: Treatment Structure

- Identify sources of variation related to treatment
  - 3 levels of Variety = 2 degrees of freedom
  - "parallels"[3] = leftover degrees of freedom

Treatment Structure	
Source	**df**
Variety	3-1=**2**
"parallels"	8-2=6
Total	9-1=8

Step 3: Skeleton ANOVA

- Identify the experimental unit where treatment levels are applied (represent with an arrow)
  - Variety is applied to each section shown with an arrow assigning Variety to Sections in the skeleton ANOVA table below
- Combine design and treatment structures to complete skeleton ANOVA table
  - Degrees of freedom for Variety carry over (**2**)
  - Degrees of freedom for Sections have to account for the variation due to Variety (**8**-**2**=**6**)
    - Sections after adjusting for Variety (denoted Sections|Variety) represents the unit level variation, e.g. residual/error

Design Structure		Treatment Structure		Combined	
Source	**df**	**Source**	**df**	**Source**	**df**
		Variety	3-1=**2**	Variety	**2**
Sections	9-1=**8**	"parallels"	8-2=6	Sections|Variety	**8**-**2**=**6**
Total	9-1=8	Total	9-1=8	Total	8

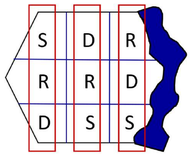
Step 4: Define Model

- Using our new notation, we have…

Data = Variety + Error

Scenario: Tulips Revisited

The planter notices there is a creek next to the plot of land, making the soil closer to the creek more damp than the soil further away. To account for this nuisance factor (soil moisture), the planter ensures that each treatment has exposure equally to the water source.



Step 0: Plot Plan

- Plot of land split into 9 sections
- Sections are grouped into blocks based on proximity to creek
- Randomly assign tulip variety to a section within each block

Step 1: Design Structure

- 3 Blocks = 2 degrees of freedom
- 3 Sections per Block = 2 degrees of freedom per block ✕ 3 blocks

Step 2: Treatment Structure

- 3 levels of Variety = 2 degrees of freedom
- "parallels"[3] = leftover degrees of freedom

Step 3: Skeleton ANOVA

- No treatments are applied to blocks.
- Variety is applied to Sections(Block) shown with an arrow below
- Combine structures to complete skeleton ANOVA table
  - Degrees of freedom for Block carry over (**2**)
  - Degrees of freedom for Variety carry over (**2**)
  - Degrees of freedom for Sections(Block) account for variation due to Variety (**6**-**2**=**4**), i.e. Sections(Blocks) after adjusting for Variety represents the unit within block variation after adjusting for treatment level, e.g. residual/error

Design Structure		Treatment Structure		Combined	
Source	**df**	**Source**	**df**	**Source**	**df**
Block	**2**			Block	**2**
		Variety	3-1=**2**	Variety	**2**
Sections(Block)	(3-1)*3=**6**	"parallels"	8-2=6	Sections(Block)|Variety	**6**-**2**=**4**
Total	8	Total	8	Total	8

Step 4: Define Model

Data = Variety + Block + Error

Checkpoint

Use degrees of freedom from skeleton ANOVA to verify model is correctly implemented.

[1] Cannon, Ann R, et al. *Stat 2: Modeling with Regression and ANOVA*. 2nd ed, 2018.

[2] Miliken, George A., and D. E. Johnson. *Analysis of Messy Data, Volume I: Designed Experiments*. 2nd ed, 2004.

[3] Stroup, Walter W. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. 2016.