

# SAS: PROC GLIMMIX

## Why are your students using it?

A. Michelle Edwards, Ph.D., MLIS

OAC Stats Consultant

# QUICK REVIEW – How we learned ANOVA

## ASSUMPTIONS

1. Normal Distribution
  - ▶ “robust”
2. Homogeneity of Variance
  - ▶ Levene’s test
3. Independence
  - ▶ Measures
  - ▶ Experimental units

# QUICK REVIEW – How we learned ANOVA

► Null Hypothesis:

$$H_o: \mu_i = \mu_j$$

$$H_a: \mu_i \neq \mu_j$$

► If  $p < 0.05$  then we got excited and said there were differences between our treatment groups

# Data Types and our Research

- ▶ Continuous data
  - ▶ We tend towards this type of data
- ▶ Categorical or Non-continuous data
  - ▶ Scores – Body Condition scores, Disease scores
  - ▶ Counts
  - ▶ Proportions

# Non-continuous data and ANOVA

- ▶ Transformations were our best friend!
- ▶ Log transformation was magic!
- ▶ There were others
  - ▶ Arcsin-square root
  - ▶ Box-cox

**Forcing that round peg into that square hole!!**

# Non-continuous data and ANOVA

- ▶ If that STILL failed....

## Non-parametric statistics

- ▶ Nothing wrong with this approach!!
- ▶ Conclusions are not as "firm"
- ▶ We CANNOT talk about means
- ▶ Maybe, just maybe, we are not analyzing the same model – no more RANDOM effects as an example

# Back to ANOVA

ANOVA = **A**nalysis of **V**ariance

Partitioning of variance (variation)

How can we explain or partition the variation seen in our outcome variable,  $Y$ , dependent variable?  
Regardless of what data type we collected!!!

Experimental design – known sources of variation

# Back to ANOVA

Total variation





# Back to ANOVA

- ▶ It was ALWAYS about the errors – how do we break apart of partition the variation in our outcome measures in a way that minimizes the random error that remains.

# Back to ANOVA

- First of all “Nothing has changed!!!” Statistics have NOT changed!!!
- The estimation methods on how to more “accurately” or better way to “define” our variation have improved
- Gone are the days of calculating SS by hand!

# Generalized Linear Mixed Models

## GLMM

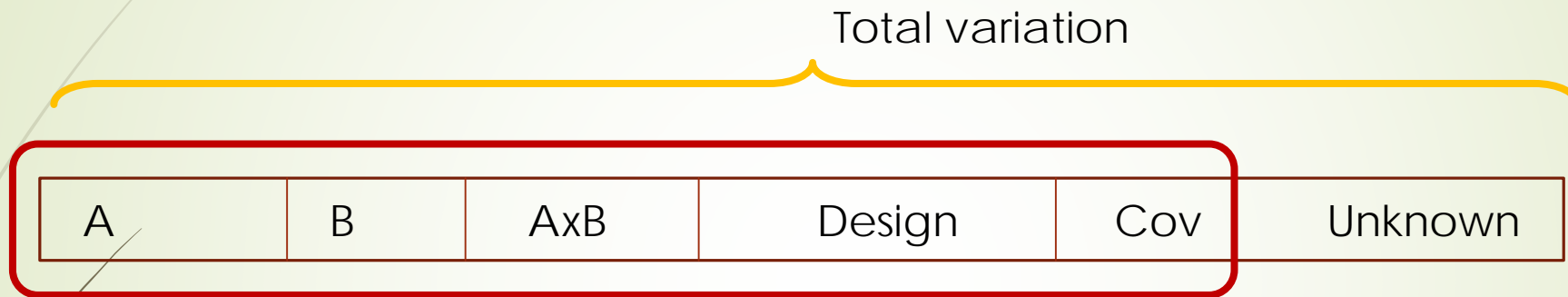
- ▶ Linear Mixed Models – we've been doing this for YEARS now
  - ▶ Are you including a RANDOM effect in your model?
  - ▶ Then you are using a Linear Mixed Model
- ▶ "Generalized" means you no longer have "normal" data going into the analysis – you are generalizing your model

Whoa!!!! What about the assumptions that were drilled into us???

### **Model assumptions – not data assumptions!**

1. Residuals are random
2. No dependencies on treatments – no patterns in our residuals
3. Homogeneity of our residuals across treatment groups
4. Residuals have a normal distribution
5. Residuals have a mean and sum =0

# ANOVA



- Errors (unknown parts)
  - Random
  - Independent of treatment and design effects
  - Common covariance (homogenous)
  - Normally distributed
  - Mean of 0

# Checking assumptions?

- Various plots of residuals
- Normality test

# What does this REALLY mean?

Data types we can use:

- Continuous (normal/Gaussian)
- Lognormal distribution
- Counts (Poisson or Negative Binomial)
- Yes/No (Binary)
- Proportions (Beta or Binomial)
- Time to event (Gamma)
- Scores (Multinomial nominal or ordinal)

# What does this REALLY mean?

For the most part – there will always be an exception

**NO MORE TRANSFORMATIONS!!!**

Analyse different types of data collected in a trial  
with the same model!!



# SAS – A little history

## ➤ PROC ANOVA

- 1966

- Balanced and FIXED effects ONLY!

## ➤ PROC GLM

- 1976

- Balanced and unbalanced (Type I and Type III SS)

- RANDOM statement – Expected Mean Square

# SAS – A little history

## ➤ PROC MIXED

➤ 1992

➤ REML – moving past the SS (OLS)

➤ RANDOM – that is incorporated – no more EMS

➤ REPEATED

# SAS – A little history

- ▶ Data MUST be from a normal distribution for:

**PROC ANOVA**

**PROC GLM**

**PROC MIXED**

# SAS – A little history

- PROC GLIMMIX
  - First introduced as %GLIMMIX – a macro
  - 2005
  - DISTRIBUTIONs

# SAS – PROC GLIMMIX

Let's work through 3 examples:

1. RCBD using PROC GLM, MIXED and GLIMMIX with continuous measure
  - To show similarities and differences between the 2 PROCs

# SAS – PROC GLIMMIX

Let's work through 3 examples:

2. An example with data that was collected as a count and a proportion
  - ▶ To show how to use a different distribution than the default Gaussian

# SAS – PROC GLIMMIX

Let's work through 3 examples:

## 3. An example with score data

- ▶ To show how to use a multinomial ordinal distribution

# SAS: PROC GLIMMIX

## Why are your students using it?

- Still an ANOVA or Partitioning of Variation
- Uses upto date estimation methods
- Allows us to use the same model for MOST datatypes collected during a trial



If you have any Questions or  
workshop/class requests

Email: [oadstats@uoguelph.ca](mailto:oadstats@uoguelph.ca)

Appointments: <http://oadstats.youcanbook.me>

Blog: oadstats.blog