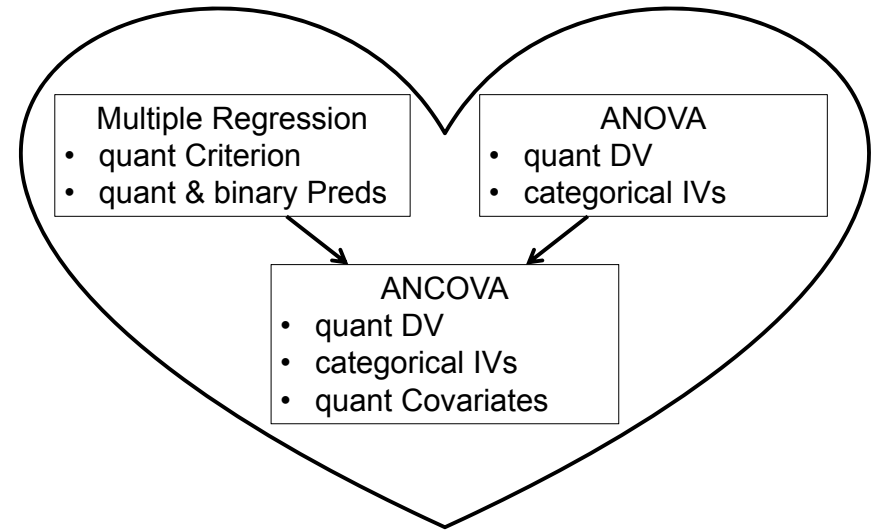# General Linear Model

- What is the General Linear Model??
- A bit of history & our goals
- Kinds of variables & effects
- It's all about the model (not the data)!!
  - plotting models & interpreting regression weights
  - model fit
  - "expected values"
  - main/simple/interaction & unconditional/conditional effects
- Coding & Centering

A common way to characterize the **GLM** is shown below

Multiple Regression
• quant Criterion
• quant & binary Preds

ANOVA
• quant DV
• categorical IVs

| ANCOVA |
|---|
| • quant DV |
| • categorical IVs |
| • quant Covariates |

We will use "GLM" to mean a model which has a quantitative criterion/dependent variable and any mix of several kinds of predictors/IVs/covariates/controls…

- Linear quantitative (raw & "re-centered")
- Non-linear quantitative (quadratic, cubic, etc)
- Binary categorical (raw, unit-, dummy-, effect
  & comparison-coded)
- Multiple categorical (raw, unit-, dummy-, effect-, trend-
  & comparison-coded)

- Categorical interactions
- Linear interactions
- Non-linear interactions

There is a lot of "historical baggage" captured this picture of the GLM!!
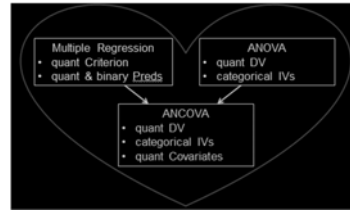
One way to think about it is that three different models each picked a subset of things in the GLM… for example…

Multiple regression is usually a "linear quant main effects model"
• lots of predictors
• usually linear-only & no interactions
ANOVA is "categorical interaction model"
• minimum number of IVs
• always look at interactions & sometimes looks at non-linear
ANCOVA is a "select combo of ANOVA & Regression"
• Minimum categorical IVs & quantitative covariates
• Looks at interactions among IVs – main effects of covariates

Our goal is to be able to put…
• any sort of variables
    • predictor, IV, covariate, control, etc
    • linear or non-linear quantitative variables…
    • binary, multiple-category, ordered-category variables…
    • categorical, quantitative, mixed & non-linear interactions

• coded or re-centered so that the associated regression weights
    • can be plotted to reveal the model
    • give as-direct-as-possible tests of RH: & RQ:

• using either "regression" "ANOVA/ANCOVA" or "GLM"
    • different analysts/packages call different collections of input and output different things
    • have to know how to control input and understand output
    • some are well-documented & understood… some…

What were those kinds of variables again ??

Up until now we have limited the kinds of predictor variables in our models to quantitative and binary (usually coded 1 & 2). In this section we will add several new variable types to our repertoire…

dummy & effect codes for binary qualitative variables
        -- different coding strategies allow us to test specific RH:

dummy, effect & comparison coding for k-category variables
        -- lots of categorical or qualitative variables aren't binary

        -- k-group variables can be important predictors, covariates, controls, etc – we have to be able to accommodate them in MR & GLM models

        -- different coding strategies allow us to test specific RH:
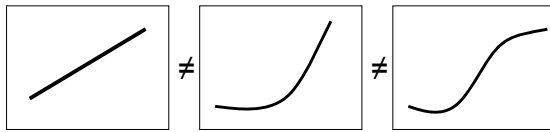
Non-linear relationships

-- not all the interesting relationships are "linear" or "just linear"
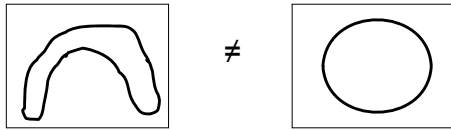
$r = 0.0$ for both → no linear relationship.

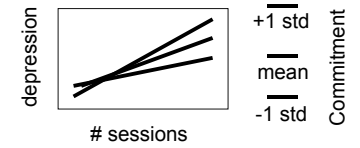But the plot on the left shows a strong relationship – just not a linear relationship
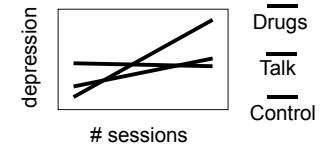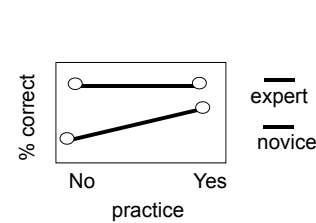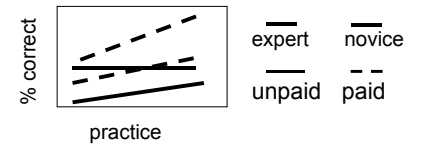
"not-just linear"

Each has a "linear component"
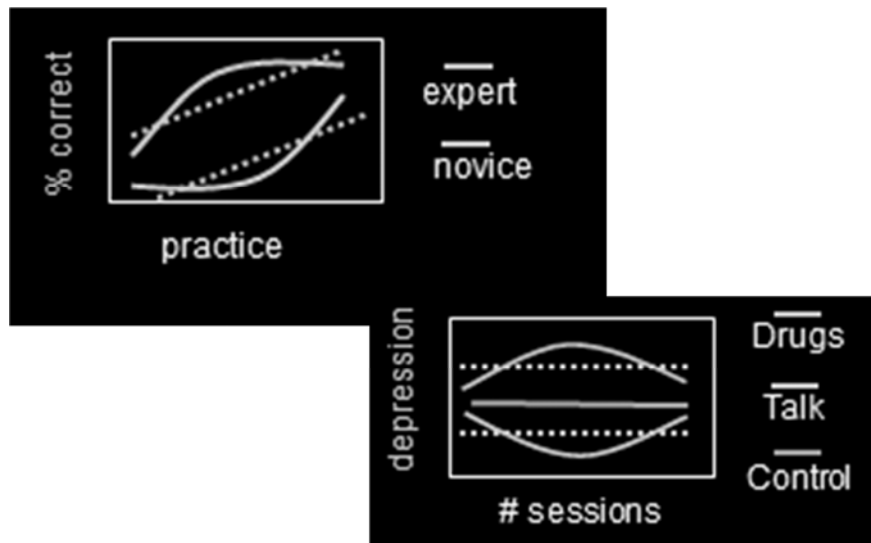
But not the "same shape of relationship"

2-way interactions - any mix of binary, k-group & quant predictor vars

3-way interactions - any mix of binary, k-group & quant predictor vars

---

non-linear interactions – you can miss interactions if only for linear interactions

Plotting Models!!!

So far we have emphasized the clear interpretation of regression weights for each type of predictor.

Many of the "more complicated" regression models, such as ANCOVA or those with non-linear or interaction terms are plotted, because, as you know, "a picture is worth a lot of words".

Each term in a multiple regression model has an explicit representation in a regression plot as well as an explicit interpretation – usually with multiple parallel phrasing versions.

So, we'll start simple with simple models and learn the correspondence between:

• interpretation of each regression model term

• the graphical representation of that term

Very important things to remember…

1) We plot and interpret *the model of the data* -- not the data
   • if the model fits the data poorly, then we're carefully describing and interpreting nonsense

2) The regression weights tell us the "**expected values of population parameters**"
   • sometimes "expected values" from the model match "descriptive values" from sample, but sometimes not…
   • One you know… descriptive marginal means match the expected marginal means from an orthogonal factorial, but not from a non-orthogonal factorial (which have been corrected for other effects in the model)
   • One you'll learn… descriptive grand mean across groups will match the expected grand mean only if the groups have equal-n (because the descriptive, but not expected value is influenced by "n")

One you'll hear a lot, in every model from here on…

3) The **interpretation** of regression weights in a main effects model **is different** than in a model including interactions
   • in a main effects model (without interactions) regression weights reflect "main effects"
   • in a model including interactions regression weights reflect "simple effects"

A bit of new language to go with this…

Unconditional effect – the effect is not dependent upon
higher-order effects
- effect in a main effect model (with no interaction terms)
- effect in a model with non-significant interaction(s)
- "descriptive" lower-order effect in a model with significant interaction(s), but effect matches corresponding "simple effects"

Conditional effect – the effect dependent upon higher-order
effects
- can't happen in a main effects model
- can't happen in a model with non-significant interaction(s)
- "misleading" lower-order effect in a model with significant interaction(s) and effect doesn't match corresponding "simple effects"

Coding & Centering predictors for plotting & interpreting

- Categorical predictors will be converted to dummy codes
  - comparison/control group coded 0
  - @ other group a "target group" of one dummy code, coded 1

- Quantitative predictors will be centered, usually to the mean
  - centered = score – mean   (like for quadratic terms)
  - so, mean = 0

Why?

Mathematically – 0s (as control group & mean) simplify the math
& minimize collinearity complications

Interpretively – the "controlling for" included in multiple regression
weight interpretations is really "controlling for all
other variables in the model at the value 0"

– "0" as the comparison group & mean will make
b interpretations simpler and more meaningful