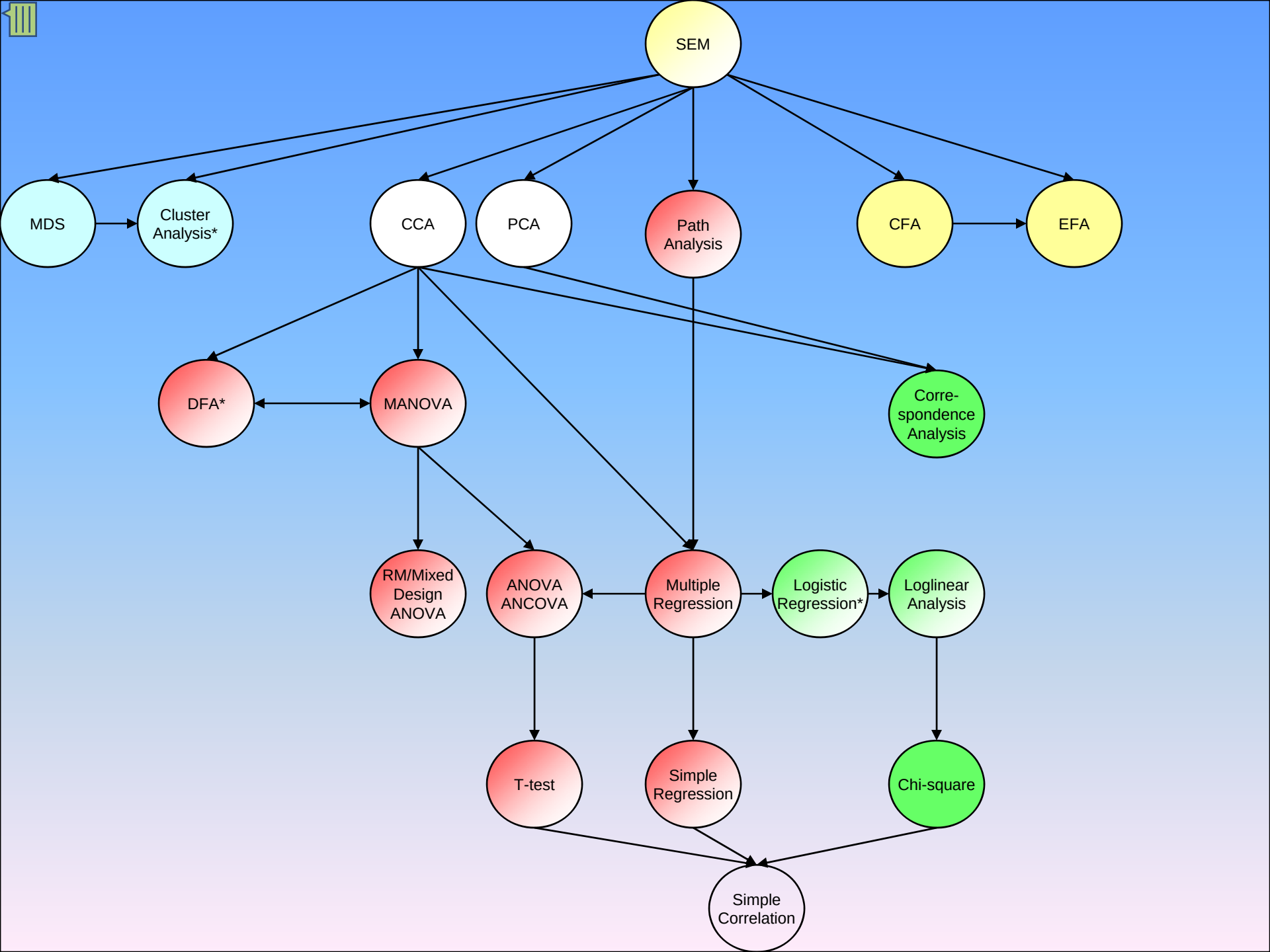
SCREW THE TREES, HERE'S THE FOREST

Relationships Between Modeling Techniques

- The following can be seen as a representation of the statistical galaxy of various techniques you've been exposed to or have/will have heard of when done with your statistical training
- It is not exact, this is just a perspective that allows one to see them relating to one another
- It's difficult to understand such a perspective when first starting out, as all the techniques are new and one struggles to see anything but a varied assortment that all seem to be doing different things and serving different purposes
- The reality, as we have noted time and time again, is that theory will determine a model which may lend itself to a host of available techniques because of their similarities
- As data type alone may suggest any number of techniques, it's theory that determines predictors/outcomes and so narrows down the choices, though possibly leaving many choices among techniques to be made in the end.
  - Variables = 3 continuous 1 categorical
    - Any number of techniques are available (MANOVA, ANCOVA, DFA, MR, Logistic regression etc.)
  - If categorical DV our options diminish, but there still may be choices to make
    - DFA or Logistic regression
- 'Thinking multivariately' allows one to expand their thinking about models/theories themselves, and will ultimately provide for a greater list of available options for analysis


SEM
MDS
Cluster Analysis*
CCA
PCA
Path Analysis
CFA
EFA
DFA*
MANOVA
Corre-spondence Analysis
RM/Mixed Design ANOVA
ANOVA ANCOVA
Multiple Regression
Logistic Regression*
Loglinear Analysis
T-test
Simple Regression
Chi-square
Simple Correlation

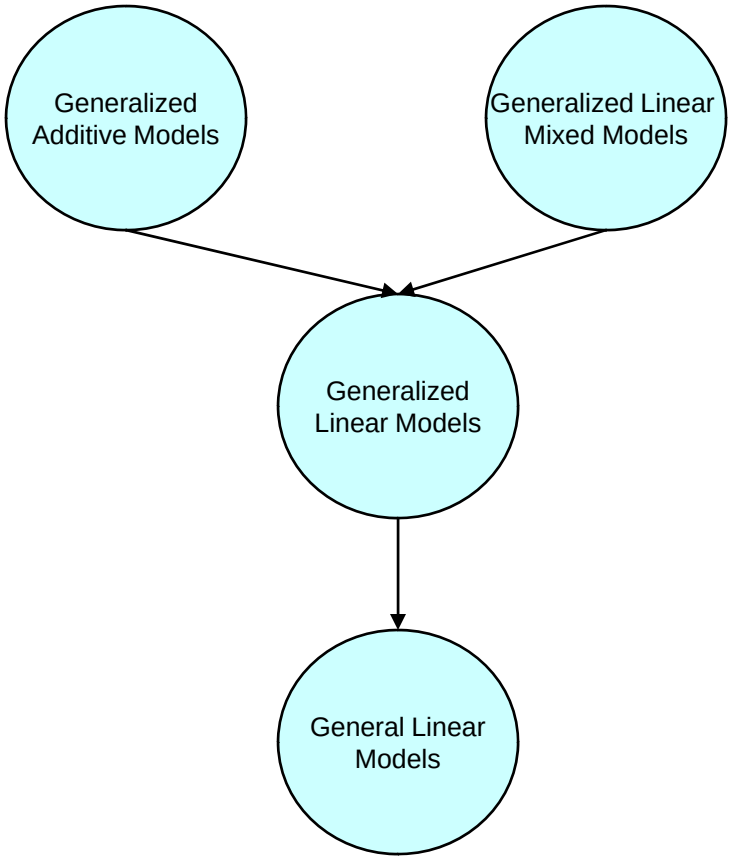
Examples

- A multiple regression can be seen as a structural model with no latent variables, a single outcome and no indirect effects
- A t-test is a special case of ANOVA, which itself tests a linear model that can be duplicated with appropriate coding in regression
- A simple correlation is a canonical correlation with only one member of each set that is to be correlated

Linear Combinations in Analysis

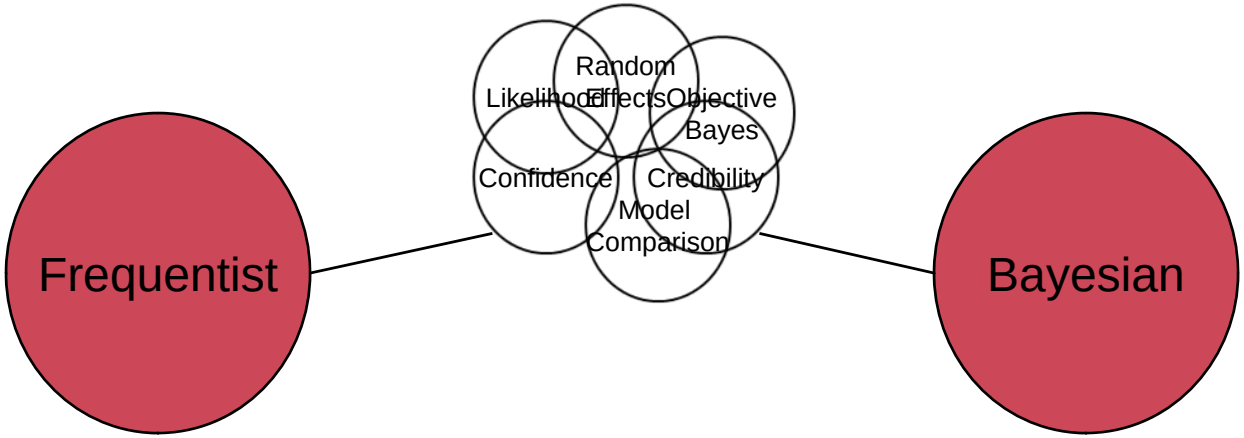
- Regression
  - Linear combination of predictors
- Contrast Analyses in ANOVA
  - Of Means
- PCA
  - Of observed variables
- Canonical Correlation
  - Of respective sets of variables
- Factor Analysis
  - Of factors

Models


Generalized Additive Models
Generalized Linear Mixed Models
Generalized Linear Models
General Linear Models

Models

- General Linear Models are the sort
- **Y = Xb + e**
- Where the contents may represent several variables (i.e. the usage of bold above implies we are dealing with vectors and matrices, hence b, the vector of coefficients, comes after X for matrix multiplication)
  - T-test:
    - $Y = \mu + bX + e$
      - Where the intercept is the grand mean, b is the mean difference of one of the group means from the grand mean and X is coded for group membership (-1,1)
  - Factor Analysis
    - $X = \lambda\xi + \delta$
      - Where X is the observed variable, λ the loading, ξ the common factor, δ the residual variance caused by unique factor(s)
- Generalized Linear models
  - Same but can take on non-continuous outcomes and/or different types of distributions
    - Logistic regression
      - **Y = Xb + e**
        - Where Y refers to log odds of being in a particular group or
- Generalized Additive Models
  - Can take nonlinear functions of predictors
    - $Y = b_0 + f_1(X_1) + f_2(X_2) \ldots + f_p(X_p)$
- Generalized Linear Mixed Models
  - Can take on random effects


Frequentist
Likelihood
Random Effects
Objective Bayes
Confidence
Credibility
Model Comparison
Bayesian

Frequentist vs. Bayesian

- Frequentist and Bayesian approaches more regard different mindsets than necessarily conflicting techniques
- As an example, when we did Bayesian model averaging before we ended up with a single regression model that could have had the usual t-tests for coefficients and F statistic for the model applied to it
  - Taking the Bayesian approach allowed for more to think about however as well as a more intuitively interpretable outcome
- Objective Bayesian approaches have the potential to produce similar results
  - For example, given certain design scenarios credible intervals may have similar values as confidence intervals
- Random effects modeling (MLM), for which frequentist approaches are often utilized for inference, have as their fundamental assumption that the true parameter in the population is random, not fixed, a notion that lies at the heart of the Bayesian mindset
- However, I'm still new to the Bayesian stuff and have a ways to go before sorting out all the differences/similarities, but some advantages of the Bayesian approach are:
  - Much more interpretable probabilities and intervals
  - Can incorporate prior information when appropriate
  - Built in Ockham's razor for model building

What analysis?

- Data type and theory limit but do not determine analysis
- There is *always* choice, there is no one right way to view the forest
  - Canopy
  - Root system underground
  - Something in between
- Some methods are equivalent or very nearly so
- Even with one analysis there might be an assortment of algorithms for it
  - E.g. various types of factor analyses, different robust estimation procedures etc.

Research

- When contemplating others' research, it's important the modeling approach fit the theory
- Many studies do not actually provide evidence that would favor one of competing theories because they do not choose an approach that allows that
- Trust your gut, if there seems to be theory-technique mismatch or missing evidence, there probably is
- You will also see other issues, e.g. redundancy or doing extra analyses in general, because of fundamental misunderstandings of how analyses relate to one another or how to obtain the desired information from the original analysis
- Examples
  - Claiming evidence for one theory when the only model comparison was to an untenable null model
  - Claiming evidence for a theory via acceptance of a null hypothesis of no effect
  - Following up MANOVA with univariate ANOVAs
  - Testing different orderings of the same variables in sequential regression
  - Testing multiple 3 variable mediation models instead of a single path analysis
- Such cases are the result of relying on simplistic texts or favoring limited software

Research

- The goal is to get the best answer to an appropriate question regarding the research problem
  - You can't change the theory to fit some analysis du jour, you have to express your theory and then choose analyses among appropriate options
- Standard general textbooks that cover many analyses are only a starting point (and are designed as introductions), you will almost always have to go (much) further than your initial training
- Undertaking psychological science is not easy, and there is no sense wasting all that time and effort with a poor and/or unenlightening analysis
- The goal is to reduce uncertainty/ignorance in some domain- sometimes that's accomplished and it may be in various ways. Sometimes new questions are uncovered.

- Be flexible enough and willing to do something new, rely on others' help and knowledge, and simply do your best considering the situation, and you'll be sure to have a satisfactory result.