A Re-Introduction to General Linear Models

- Today's Class:
  - Big picture overview
  - Why we are using restricted maximum likelihood within MIXED instead of least squares within GLM
  - Linear model interpretation
  - Example of main effects in general linear models

Lego-Based Quantitative Methods

The Big Picture:
If you understand the elemental building blocks of statistical models, then you can build **anything**!

The 4 Lego Building Blocks

1. **Linear models** (for **answering questions** of prediction)

2. **Estimation** (for iterative ways of **finding the answers**)

3. **Link functions** (for predicting **any type of outcome**)

4. (a) **Random effects** / (b) **Latent variables**

    (a) for modeling multivariate **"correlation/dependency"**
    (we will cover random intercepts this semester)

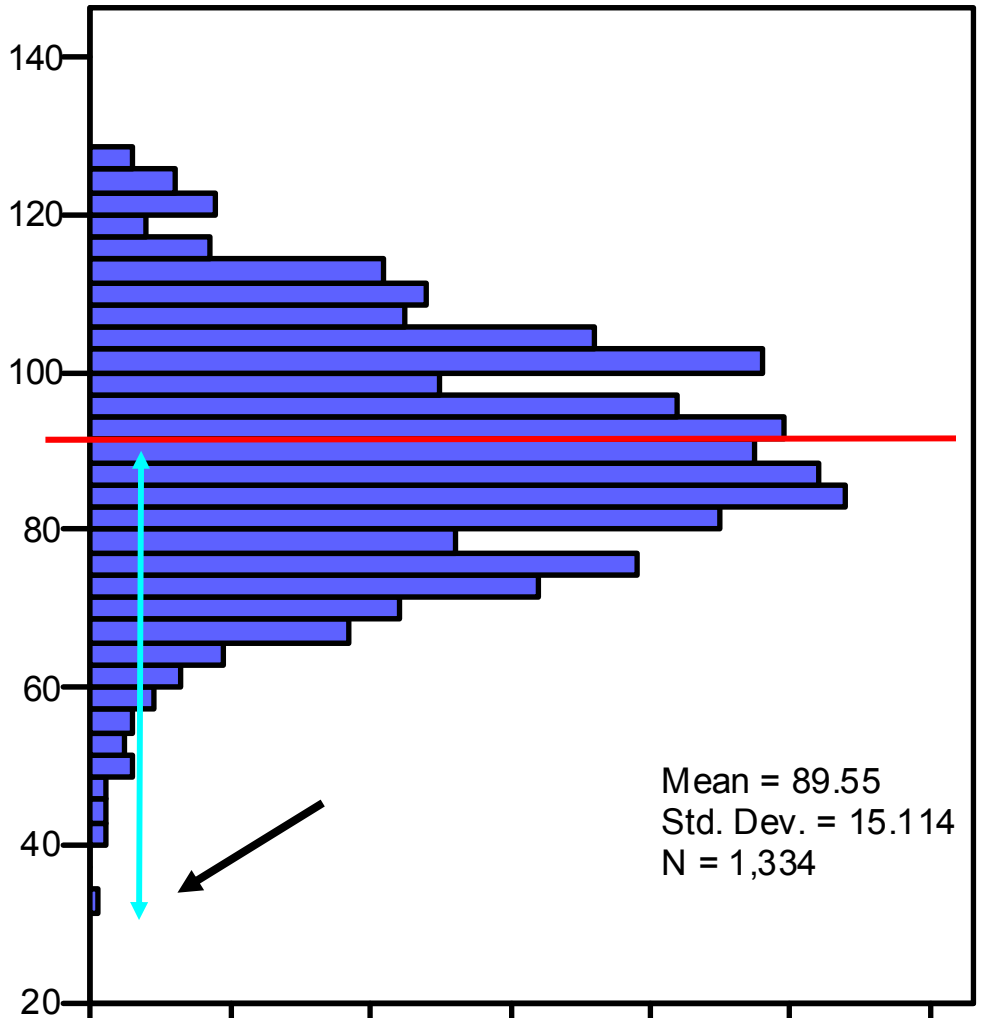
    (b) for modeling relations of **"unobserved constructs"**
    (you will need to take SEM for this; see PSYC 948)

The Two Sides of Any Model

- **Model for the Means:**
  - *Aka* **Fixed Effects**, Structural Part of Model
  - What you are used to **caring about for testing hypotheses**
  - How the expected outcome for a given observation varies as a function of values on predictor variables

- **Model for the Variance:**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
  - How residuals are distributed and related across observations
  - What you are used to **making assumptions about** instead…
  - For regression, that residuals come from a **normal** distribution, are **independent** across persons, and have **constant variance** across persons and predictors ("identically distributed")

The Simplest Possible Model: The "Empty" Model



$$y_i = \beta_0 + e_i$$

Filling in values:

$$32 = \underbrace{90}_{y_{pred}} + -58$$

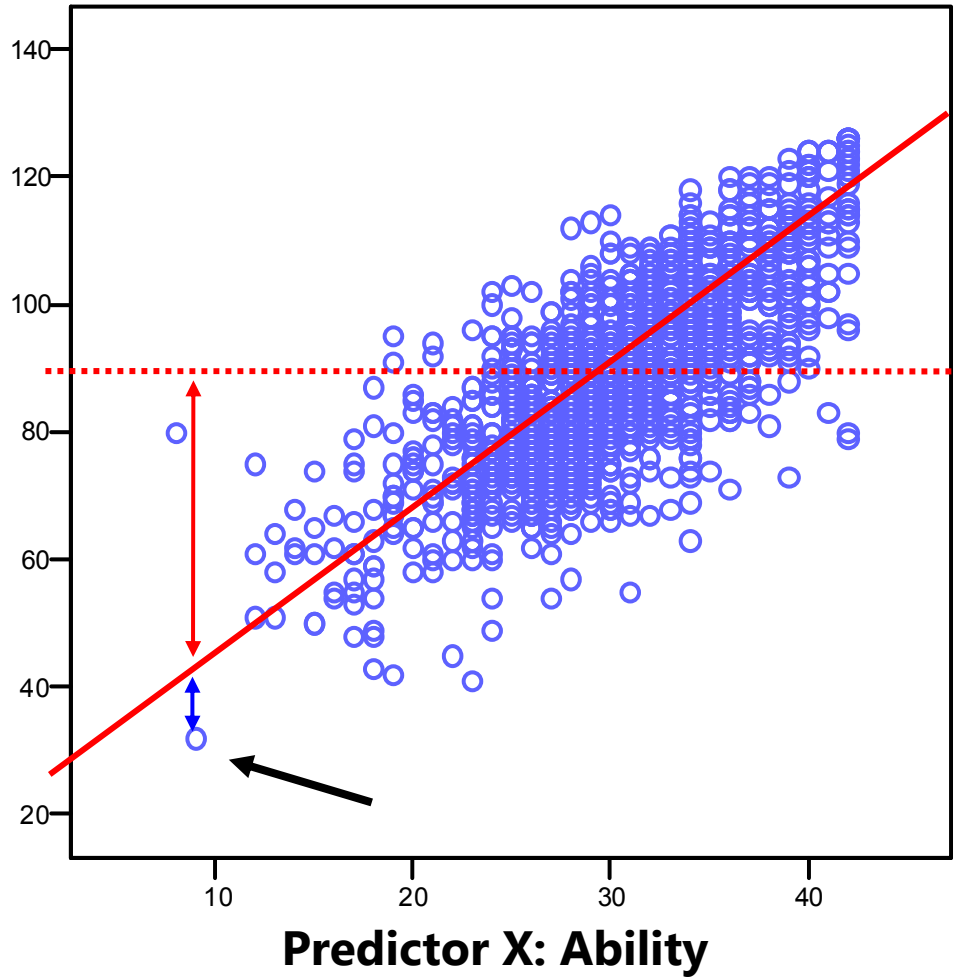
Model for the Means

y_{pred} is also called $\hat{y}$ ("y-hat")

y_i error variance:

$$\frac{\sum(y_i - y_{pred})^2}{N - 1}$$

"Linear Regression" with a Continuous Predictor (Ability)



$$y_i = \beta_0 + \beta_1 X_i + e_i$$

Empty Model:

$32 = 90 + -58$

Ability (X) Model:

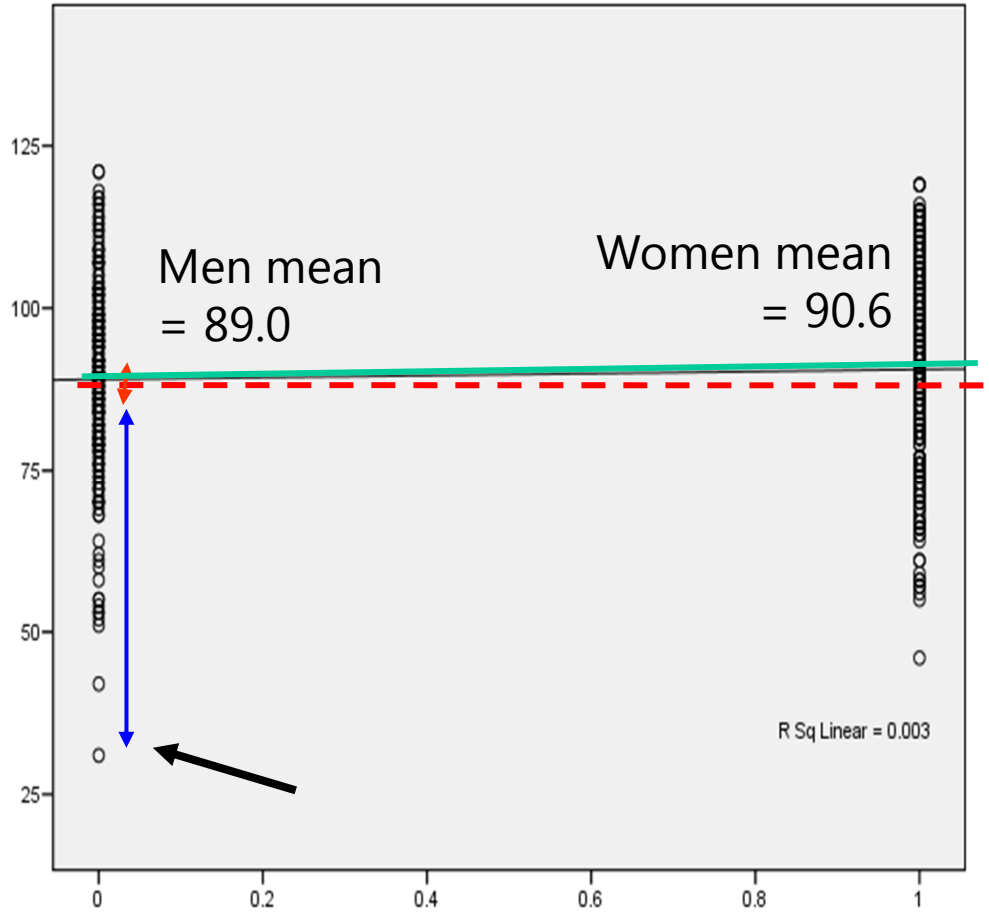
$32 = \underbrace{29 + 2*9}_{y_{pred}} + -15$

Model for the Means

y_i error variance:

$$\frac{\sum(y_i - y_{pred})^2}{N - 2}$$

Between-Person Model: One Categorical Predictor (Sex)



Predictor Z: Sex (0=M, 1=W)

$$y_i = \beta_0 + \beta_2 Z_i + e_i$$

Empty Model:

32 = 90 + −58

Sex (Z) Model:

32 = 89 + 1.6*0 + −57

y_{pred}

Model for the Means

y_i error variance:

$$\frac{\sum(y_i - y_{pred})^2}{N - 2}$$

The Two Sides of Each Model: Practice with Interpretation

	Model for the Means	Model for the Variance
"Empty"	$\beta_0 \rightarrow$ grand mean of y_i	$e_i \rightarrow$ diff between y_i and mean of y_i
"Regression"	$\beta_0 \rightarrow y_{pred}$ when X=0 $\beta_1 \rightarrow \Delta y_{pred}$ for a 1-unit ΔX_i	$e_i \rightarrow$ diff between y_i and y_i predicted from X_i
"ANOVA"	$\beta_0 \rightarrow y_{pred}$ when Z=0 $\beta_2 \rightarrow \Delta y_{pred}$ for a 1-unit ΔZ_i	$e_i \rightarrow$ diff between y_i and y_i predicted from Z_i
"ANCOVA"	$\beta_0 \rightarrow y_{pred}$ when X=0 & Z=0 $\beta_1 \rightarrow \Delta y_{pred}$ for 1-unit ΔX_i $\beta_2 \rightarrow \Delta y_{pred}$ for 1-unit ΔZ_i	$e \rightarrow$ diff between y_i and y_i predicted from X_i and Z_i

The Two Sides of a General Linear Model

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \cdots + e_i$$

Our focus today

- **Model for the Means (Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on X and Z (and any other predictors), each measured once per person (i.e., this is a univariate model)
  - **Estimated parameters are called fixed effects** (here, β_0, β_1, and β_2)
  - The number of fixed effects will show up in formulas as ***k*** (so ***k* = 3** here)

- **Model for the Variance:**
  - $e_i \sim N(0, \sigma_e^2)$ → ONE source of residual (unexplained) deviation
  - e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to X and Z, and is unrelated across people (across all observations, just people here)
  - **Estimated parameter is the residual variance only** (not each e_i)

The General Linear Model

- The general linear model incorporates many different labels of related analyses under one unifying umbrella term:

	Categorical X's	Continuous X's	Both Types of X's
Univariate (one outcome)	"ANOVA"	"Regression"	"ANCOVA"
Multivariate (2+ outcomes)	"MANOVA"	"Multivariate Regression"	"MANCOVA"

- What these models all have in common is the use of a **normal conditional distribution** (for the residuals that remain after creating conditional outcomes from the model predictors)
- The use of these words almost always means **estimation** using "**least squares**" (LS), *aka* "**ordinary least squares**" (OLS)

How Estimation Works (More or Less)

- Most statistical estimation routines do one of three things:

- Minimize Something: Typically found with names that have "least" in the title. Forms of **least squares** include "Generalized", "Ordinary", "Weighted", "Diagonally Weighted", "WLSMV", and "Iteratively Reweighted." **Typically the estimator of last resort**…

- Maximize Something: Typically found with names that have "maximum" in the title. Forms include "**Maximum likelihood**", "ML", "Residual Maximum Likelihood" (REML), "Robust ML". Typically the gold standard of estimators

- Use Simulation to Sample from Something: more recent advances in simulation use resampling techniques. Names include "**Bayesian** Markov Chain Monte Carlo", "Gibbs Sampling", "Metropolis Hastings", "Metropolis Algorithm", and "Monte Carlo". Used for complex models in which ML is not available or for methods where prior values are needed

Least Squares (LS) Estimation

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Square (MS)	F-ratio
Model (from predictor model)	$SS_{model} = \sum(\beta_0 - y_{pred})^2$	DF_{num} = #fixed effects – 1 (for β_0)	$MS_{model} = SS_{model} / DF_{num}$	F-ratio = MS_{model} / MS_{error}
Error (from empty model)	$SS_{error} = \sum(y_i - y_{pred})^2$	DF_{denom} = #people – #fixed effects – 1 (for β_0)	$MS_{model} = SS_{model} / DF_{num}$	
Total	$SS_{total} = SS_{model} + SS_{error}$	$DF_{total} = DF_{num} + DF_{denom}$		

- MS_{model} = how much error you reduced per added fixed effect
- MS_{error} = how much error is left, per possible new fixed effect (otherwise known as residual/error variance)
- Compare F-ratio to critical value given DF_{num} and DF_{denom} to get *p*-value for model R^2 (proportion error variance reduced)

Least Squares (LS) Estimation

- Uses fixed effect estimates that minimize: $\sum_{i=1}^{N}(e_i^2)$
  - (Sum of squared residuals across persons)
  - Invented c. 1840, can be done via matrix algebra, so will always work

- Has "closed form" solution (easy formula) when used for general linear models (GLM) for single outcomes
  - $e_i \sim N(0, \sigma_e^2)$ → normal, independent, constant variance

- For GLM for multiple outcomes, LS quickly becomes useless…
  - Cannot handle missing outcomes (listwise-deletes entire person instead)
  - Only two options for modeling covariance between outcomes
  - Then why do it this way? Dogma + lack of awareness of alternatives…

- For non-normal outcomes, LS can't be used at all…

Dimensions for Organizing Models

- Outcome type: General (normal) vs. General*ized* (not normal)
- Outcomes per person: One (so one variance term per outcome) vs. **Many** (need covariance among observations from same source)

- **General Linear Models:** conditionally normal outcome distribution, **fixed effects** (identity link; only one dimension of sampling)

Note: Least Squares is only for GLM

- **General*ized* Linear Models: any conditional outcome distribution**, **fixed** effects through **link functions**, no random effects (one dimension)
- **General Linear Mixed Models:** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **General*ized* Linear Mixed Models: any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)

- "Linear" means the fixed effects predict the *link-transformed* conditional mean of the DV in a linear combination of (effect*predictor) + (effect*predictor)…

Maximum Likelihood to the Rescue

- Maximum likelihood estimation is better way of finding the model estimates using all the data, and it comes in 2 flavors:
- **"Restricted (or residual) maximum likelihood"**
  - Only available for general linear models or general linear mixed models (that assume normally distributed residuals)
  - Is same as LS given complete outcomes, but it doesn't require them
  - Estimates variances the same way as in LS (accurate) → $\frac{\sum(y_i - y_{pred})^2}{N - k}$
- **"Maximum likelihood" (ML; also called FIML*)**
  - Is more general, is available for the above plus for non-normal outcomes and latent variable models (CFA/SEM/IRT)
  - Is NOT the same as LS: it under-estimates variances by not accounting for the # of estimated fixed effects → $\frac{\sum(y_i - y_{pred})^2}{N}$
- **FI = Full information → it uses all original data (they both do)*

Maximum Likelihood to the Rescue

- Even though REML = LS for complete outcomes, we will begin by using software based in REML instead of LS
  - In SPSS, SAS, or STATA: one routine called **"MIXED"** instead of separate routines for GLM, REGRESSION, or ANOVA (or t-tests)
  - So "sums of squares" and "mean squares" are no longer relevant

- Why?
  - **Big-time convenience**: MIXED has options to produce fixed effects that are model-implied, but not directly given (e.g., pairwise comparisons, simple slopes of interactions)
  - **Model comparisons** (F-ratio for change in R^2 from new effects) can be requested in a single step for any combination of effects
  - **Generalizability**: We can estimate **univariate or multivariate** models for normal outcomes using the same MIXED routine
  - For **non-normal outcomes**, there are parallel routines in SAS (GLIMMIX) and STATA (several), but not in SPSS ("pseudo-ML")

End Goals of Maximum Likelihood Estimation

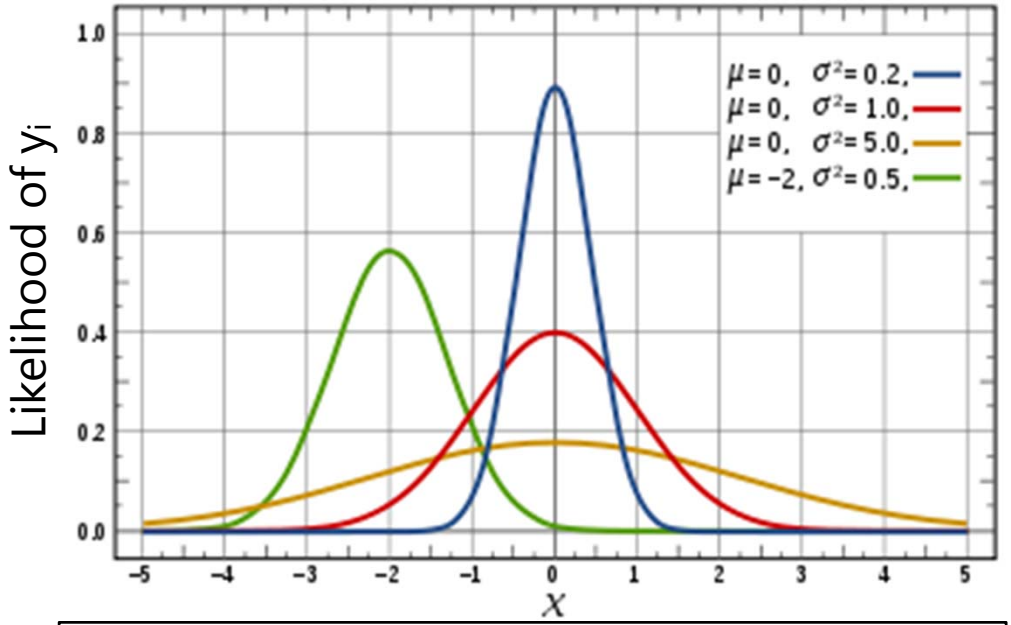
1. Obtain "most likely" values for each unknown model parameter (fixed effects, variances of residuals, and any other sources of variance and covariance) → **the estimates**

2. Obtain an index as to how likely each parameter value actually is (i.e., "really likely" or pretty much just a guess?) → **the standard error (SE) of the estimates**

3. Obtain an index as to how well the model we've specified actually describes the data → **the model fit indices**

How does all of this happen? Probability distributions!
(i.e., probability density functions, or PDFs)

Univariate Normal Distribution



Univariate Normal PDF:

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2} \right]$$

Sum over persons for log of f(yi)= Model Log-Likelihood → Model Fit

- This PDF tells us how **likely** any value of y_i is given two pieces of info:
  - Conditional mean $\hat{y}_i$
  - residual variance σ_e^2
- We can see this work using the NORMDIST function in excel!
  - Easiest for empty model:

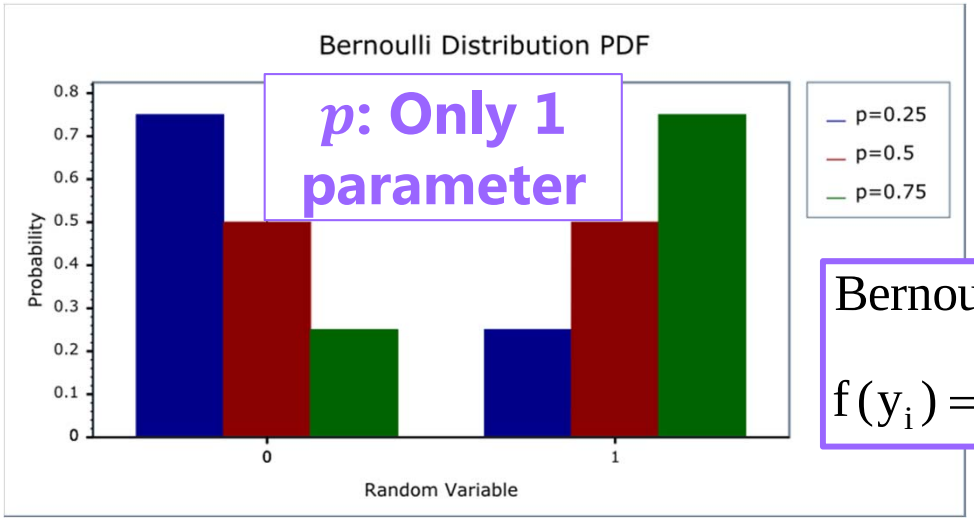
$$y_i = \beta_0 + e_i$$

- We can check our math via SAS/SPSS MIXED!

What if our outcome isn't normal?

- Pick a new probability distribution with which to find the height of each outcome given the model…

- Bernoulli for binary outcomes ("logistic regression"):



***p*: Only 1 parameter**

Bernoulli PDF:

$$f(y_i) = (p_i)^{y_i} (1-p_i)^{1-y_i}$$

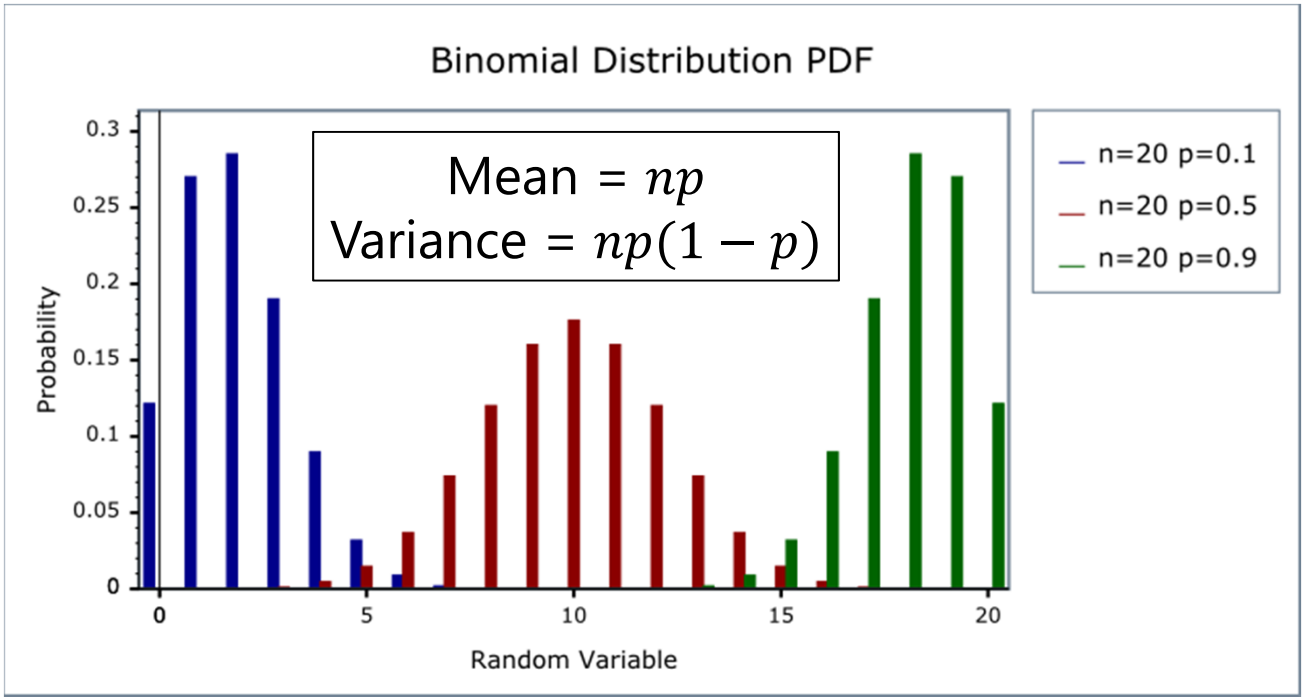
= *p*(1) if 1, *p*(0) if 0

- Here is a sampling of other distributions to use…

Binomial Distribution for Proportions

- The discrete **binomial** distribution can be used to predict $\boldsymbol{c}$ correct responses given $\boldsymbol{n}$ trials
  - Bernoulli for binary = special case of binomial when n=1
  - $Prob(y = c) = \frac{n!}{c!(n-c!)} p^c (1-p)^{n-c}$

p = probability of 1



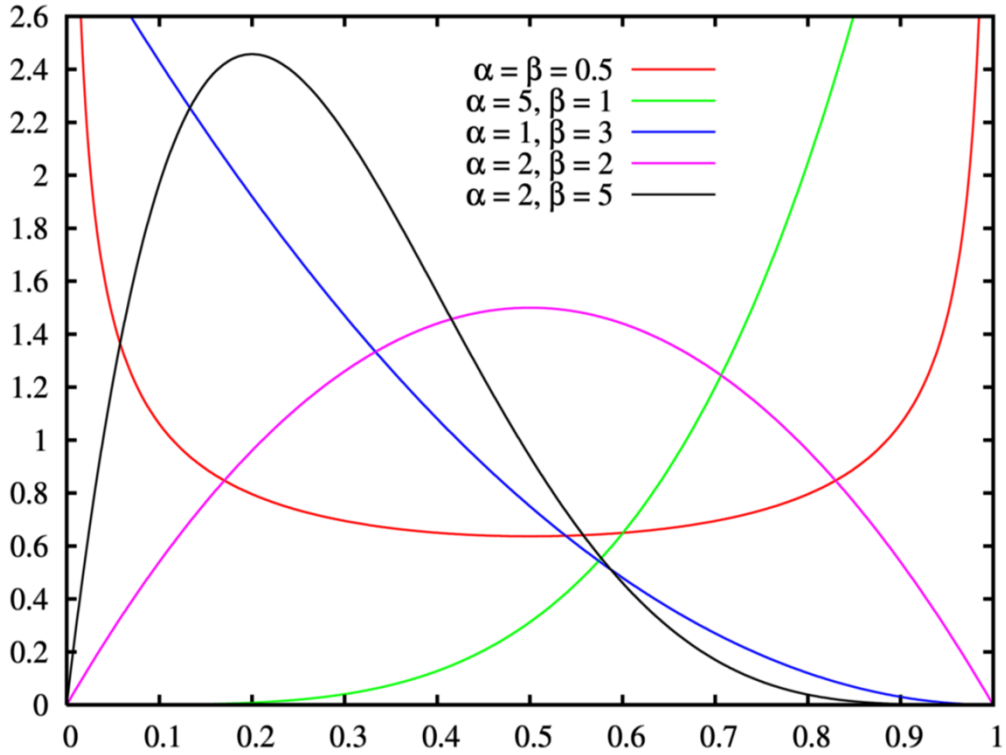
As p gets closer to .5 and n gets larger, the binomial pdf will look more like a normal distribution.

But if many people show floor/ceiling effects, a normal distribution is not likely to work well… so use a binomial!

Beta Distribution for Proportions

- The continuous **beta** distribution (LINK=LOGIT, DIST=BETA) can predict <u>percentage</u> correct $\boldsymbol{p}$ (must be $0 < \boldsymbol{p} < 1$)
  - $F(y|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \mathrm{y}^{\alpha-1}(1-\mathrm{y})^{\beta-1}$

α and β are "shape" parameters (> 0)



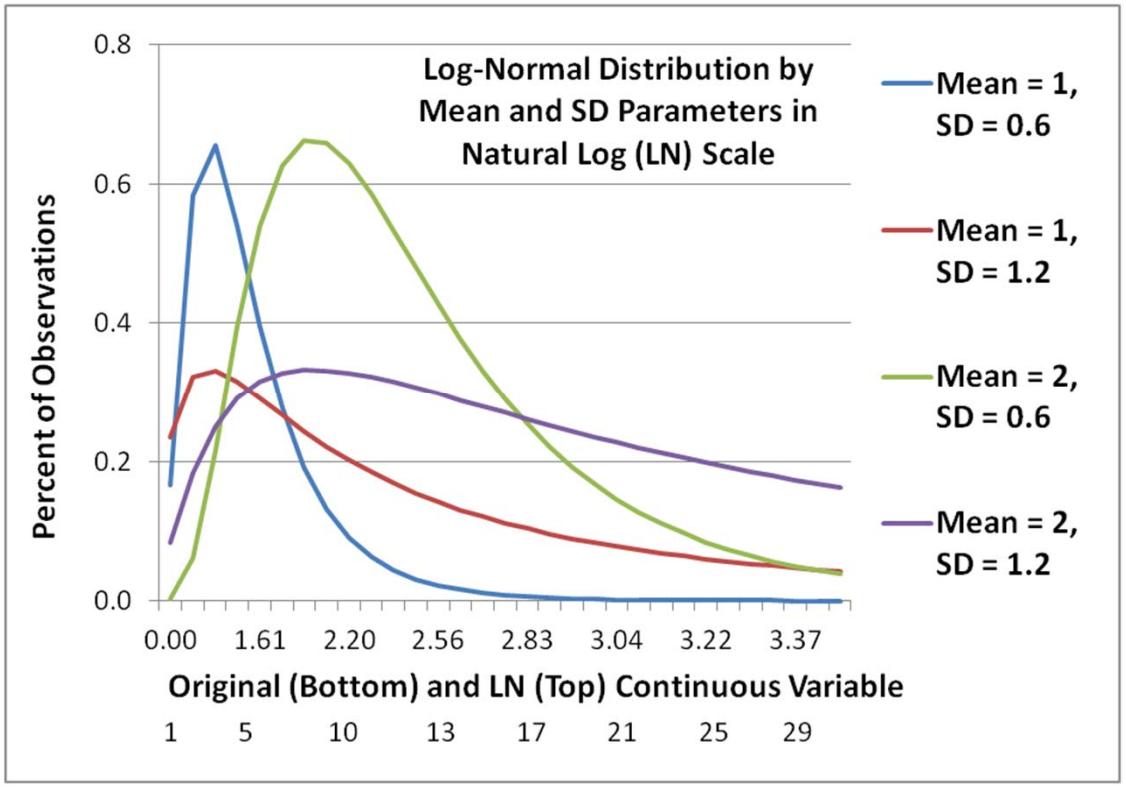
Mean = $\mu = \frac{\alpha}{\alpha+\beta}$

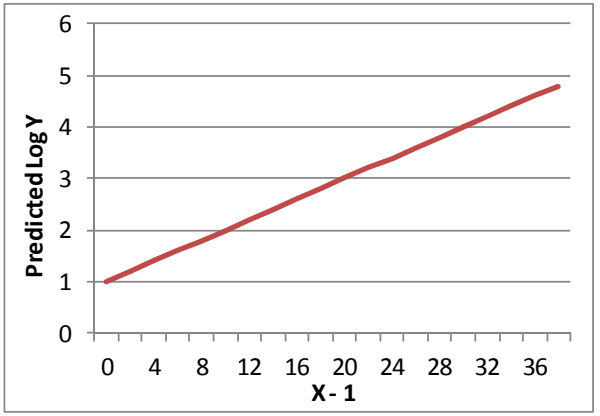
"Scale" = $\phi = \alpha + \beta$

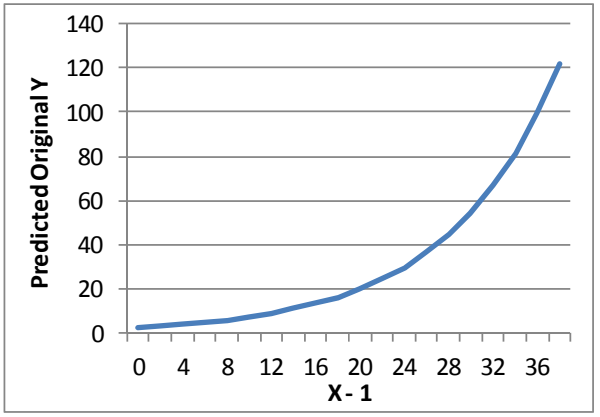
Variance = $\frac{\mu(1-\mu)}{1+\phi}$

SAS GLIMMIX will provide a fixed intercept as logit(μ) and the "scale" ϕ

Log-Normal Distribution for Skewed Outcomes







- GLIMMIX parameterization gives μ (= intercept) and $scale$ = (variance) to convert back into original data as follows:
  - $\text{Mean(Y)} = \exp(\mu) * \sqrt{\exp(scale)}$
  - $\text{Variance(Y)} = \exp(2\mu) * \exp(scale) * [\exp(scale) - 1]$

Zero-Inflated Distributions

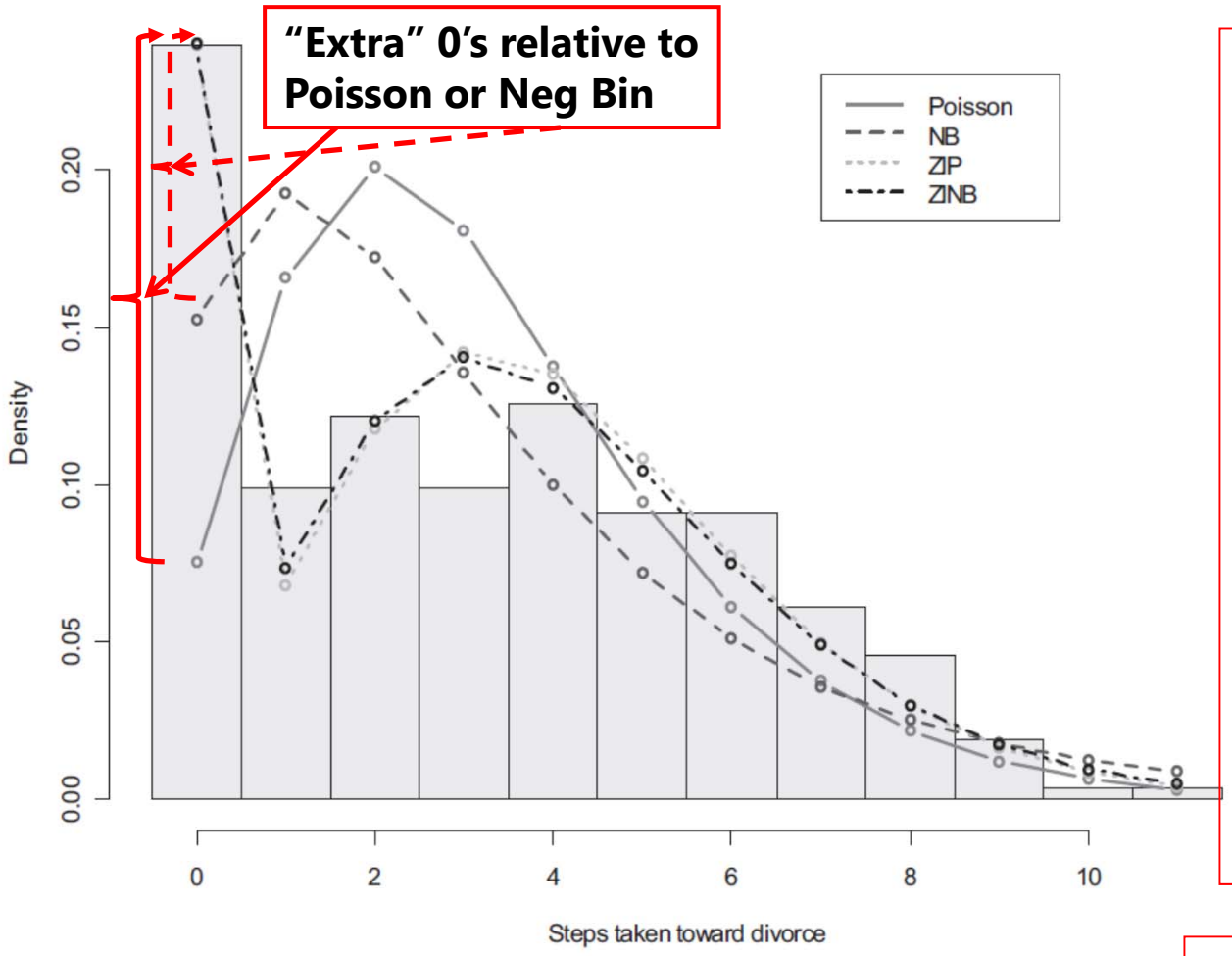


Figure 1. Histogram of Marital Status Inventory with predicted probabilities from regressions. NB = negative binomial; ZIP = zero-inflated Poisson; ZINB = zero-inflated negative binomial.

Zero-inflated distributions have extra "structural zeros" not expected from Poisson or NB ("stretched Poisson") distributions.

This can be tricky to estimate and interpret because the model distinguishes between *kinds of zeros* rather than zero or not, but "hurdle" models can be used instead to distinguish "if" and "how much" (or "two-part" for continuous "how much" distributions).

Image borrowed from Atkins & Gallop, 2007

Intermediate Summary

- **What is not new:**
  - We will be starting with the same kind of general linear univariate models for single outcomes per person (regression, ANOVA, ANCOVA) you already know…
  - We will examine main effects (today) and interaction terms (later) among all kinds of predictors

- **What is new:**
  - Rather than finding the fixed effects and residual variance through least squares (which yields sums of squares, mean squares, and so forth), we will find them using restricted maximum likelihood, of which least squares is a special case with limited applicability to general linear models…

Testing Significance of Fixed Effects in the Model for the Means

- Any estimated **fixed effect** has 4-5 relevant pieces of output:
  - **Estimate** = best guess for the fixed effect from our data
  - **Standard Error** = precision of fixed effect estimate (quality of most likely estimate)
  - ***t*-value or *z*-value** = Estimate / Standard Error → **Wald test**
  - ***p*-value** = probability that fixed effect estimate is ≠ 0
  - **95% Confidence Interval** = Estimate ± 1.96*SE = range in which true (population) value of estimate is expected to fall 95% of the time

- Compare Wald test statistic to critical value at chosen level of significance (known as alpha)
- Whether the *p*-value is based on *t* or *z* varies by program…

Evaluating Significance of Fixed Effects

Fixed effects can be tested via **Wald** tests: the ratio of its estimate/SE forms a statistic we compare to a distribution

	Denominator DF is assumed infinite	Denominator DF is estimated instead
Numerator DF = 1 *(test one effect)*	use ***z*** distribution (Mplus, STATA)	use ***t*** distribution (SAS, SPSS)
Numerator DF > 1 *(test 2+ effects at once)*	use χ^2 distribution (Mplus, STATA)	use ***F*** distribution (SAS, SPSS)
Denominator DF (DDFM) options	not applicable, so DDF is not given	SAS: BW and KR SAS and SPSS: Satterthwaite

Standard Errors for Fixed Effects

- Standard Error (SE) for estimate β_X in a one-predictor model (remember, SE is like the SD of the estimated parameter):

$$SE_{\beta X} = \sqrt{\frac{\text{residual variance of Y}}{\text{variance of X} * (N-k)}}$$

- When more than one predictor is included, SE turns into:

$$SE_{\beta X} = \sqrt{\frac{\text{residual variance of Y}}{Var(X) * (1-R_X^2) * (N-k)}}$$

R_X^2 = X variance accounted for by other predictors, so $1-R_X^2$ = unique X variance

- So all things being equal, SE is smaller when:
  - More of the outcome variance has been reduced (better model)
    - So fixed effects can become significant later if R^2 is higher then
  - The predictor has less covariance with other predictors (less collinearity)
    - Best case scenario: X is uncorrelated with all other predictors

- If SE is smaller → *t*-value or *z*-value is bigger→ *p*-value is smaller

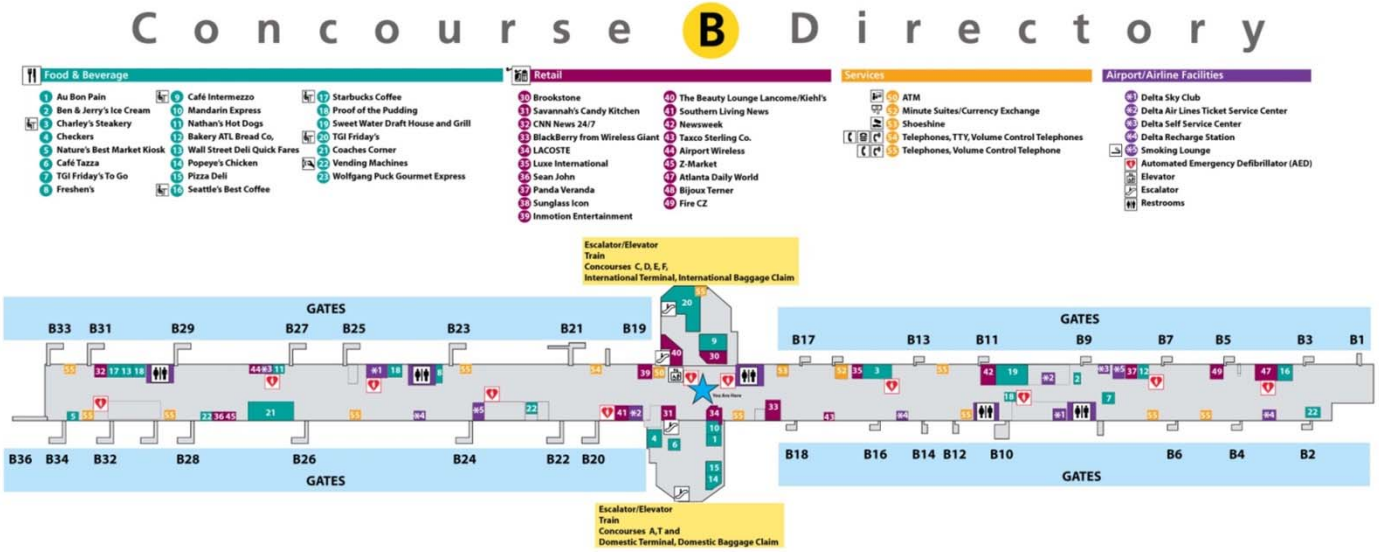
Representing the Effects of Predictors

- From now on, we will think carefully about exactly **how** the **predictor variables** are entered into the **model for the means** (i.e., by which a predicted outcome is created for each person)

- Why don't people always care? Because the scale of predictors:
  - Does NOT affect the amount of outcome variance accounted for (R^2)
  - Does NOT affect the outcomes values predicted by the model for the means (so long as the same predictor fixed effects are included)

- Why should this matter to us?
  - ***Because the Intercept = expected outcome when all predictors = 0***
  - Can end up with nonsense values for intercept if X = 0 isn't in the data, so we need to change the scale of the predictors to include 0
  - Scaling becomes more important once interactions are included or once random intercepts are included (i.e., variability around fixed intercept)

Why the Intercept β_0 *Should* Be Meaningful…



This is a very detailed map…
But what do we need to know
to be able to use the map at all?

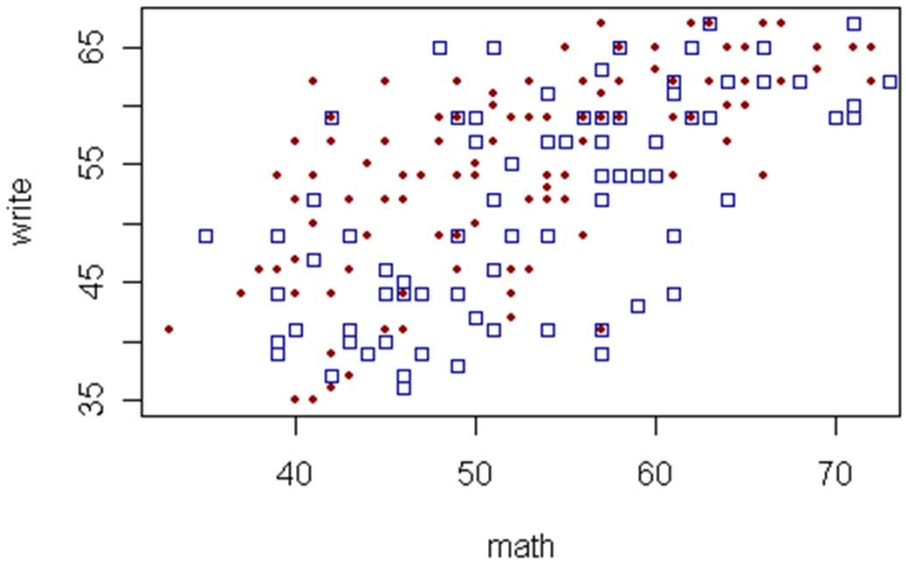
What the Intercept β_0 *Should* Mean to You…

The model for the means will describe what happens to the predicted outcome Y

"as X increases" or

"as Z increases"

and so forth…



But you won't know what the predicted outcome is supposed to be unless you know where the predictor variables are starting from!

Therefore, the **intercept** is the "**YOU ARE HERE**" sign in the map of your data… so it should be somewhere in the map*!

* There is no *wrong* way to center (or not), only *weird*…

What the Intercept *WILL* Mean to You…

Adjusting the Scale of Predictors

- For **continuous** (quantitative) predictors, **we** will make the intercept interpretable by **centering**:
  - **Centering** = subtract a constant from each person's variable value so that **the 0 value** falls within the range of the new centered predictor variable
  - Typical → Center around predictor's mean: $\text{Centered } X_1 = X_1 - \overline{X_1}$
    - Intercept is then expected outcome for "average X_1 person"
  - Better → Center around meaningful constant *C*: $\text{Centered } X_1 = X_1 - C$
    - Intercept is then expected outcome for person with that constant (even 0 may be ok)

- For **categorical** (grouping) predictors, **either we or the program** will make the intercept interpretable by **creating a reference group**:
  - **Reference group** is given a 0 value on all predictor variables created from the original grouping variable, such that the intercept is the expected outcome for that reference group specifically
  - Accomplished via "dummy coding" (aka, "reference group coding")
    → Two-group example using *Gender*: 0 = Men, 1 = Women
    (or 0 = Women, 1 = Men)

Adjusting the Scale of Predictors

- For more than two groups, need: ***dummy codes = #groups – 1***
  - "Treatgroup" variable: Control=0, Treat1=1, Treat2=2, Treat3=3
  - Variables:

    d1 = 0, 1, 0, 0 → difference between Control and T1
    d2 = 0, 0, 1, 0 → difference between Control and T2
    d3 = 0, 0, 0, 1 → difference between Control and T3

SAS CLASS or SPSS BY can do this for you ☺

- Potential pit-falls:
  - All predictors for the effect of group (e.g., d1, d2, d3) MUST be in the model at the same time for these specific interpretations to be correct!
  - Model parameters resulting from these dummy codes will not *directly* tell you about differences among non-reference groups (but they can)
- Other examples of things people do to categorical predictors:
  - "Contrast/effect coding"→ *Gender*: −0.5 = Men, 0.5 = Women
  - Test other contrasts among multiple groups → four-group example: *contrast1* = −3, 1, 1, 1 / DIVISOR=3 → Control vs. Any Treatment?

Categorical Predictors: Manual Coding

- Model: $y_i = \beta_0 + \beta_1 d1_i + \beta_2 d2_i + \beta_3 d3_i + e_i$
  - "Treatgroup" variable: Control=0, Treat1=1, Treat2=2, Treat3=3
  - New variables to be created for the model:
    d1= 0, 1, 0, 0 → difference between Control and T1
    d2= 0, 0, 1, 0 → difference between Control and T2
    d3= 0, 0, 0, 1 → difference between Control and T3
- How does the model give us **all possible group differences**? By determining each group's mean, and then the difference…

Control Mean (Reference)	Treatment 1 Mean	Treatment 2 Mean	Treatment 3 Mean
β_0	$\beta_0 + \beta_1 d1_i$	$\beta_0 + \beta_2 d2_i$	$\beta_0 + \beta_3 d3_i$

- The model for the 4 groups directly provides 3 differences (control vs. each treatment), and indirectly provides another 3 differences (differences between treatments)

Group Differences from Dummy Codes

- Model: $y_i = \beta_0 + \beta_1 d1_i + \beta_2 d2_i + \beta_3 d3_i + e_i$

Control Mean (Reference)	Treatment 1 Mean	Treatment 2 Mean	Treatment 3 Mean
$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_0 + \beta_1 d1_i}$	$\boldsymbol{\beta_0 + \beta_2 d2_i}$	$\boldsymbol{\beta_0 + \beta_3 d3_i}$

	Alt Group	Ref Group	Difference
• Control vs. T1 =	$(\beta_0+\beta_1)$	$-\ (\beta_0)$	$= \beta_1$
• Control vs. T2 =	$(\beta_0+\beta_2)$	$-\ (\beta_0)$	$= \beta_2$
• Control vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0)$	$= \beta_3$
• T1 vs. T2 =	$(\beta_0+\beta_2)$	$-\ (\beta_0+\beta_1)$	$= \beta_2 - \beta_1$
• T1 vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0+\beta_1)$	$= \beta_3 - \beta_1$
• T2 vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0+\beta_2)$	$= \beta_3 - \beta_2$

ESTIMATEs when using dummy codes

	Alt Group	Ref Group	Difference
• Control vs. T1 =	$(\beta_0+\beta_1)$	$-\ (\beta_0)$	$= \beta_1$
• Control vs. T2 =	$(\beta_0+\beta_2)$	$-\ (\beta_0)$	$= \beta_2$
• Control vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0)$	$= \beta_3$
• T1 vs. T2 =	$(\beta_0+\beta_2)$	$-\ (\beta_0+\beta_1)$	$= \beta_2 - \beta_1$
• T1 vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0+\beta_1)$	$= \beta_3 - \beta_1$
• T2 vs. T3 =	$(\beta_0+\beta_3)$	$-\ (\beta_0+\beta_2)$	$= \beta_3 - \beta_2$

Note the order of the equations: the reference group mean *is subtracted from* the alternative group mean.

In SAS ESTIMATE statements (or SPSS TEST or STATA LINCOM), the variables refer to their betas; the numbers refer to the operations of their betas.

```
TITLE "Manual Contrasts for 4-Group Diffs";
PROC MIXED DATA=dataname ITDETAILS METHOD=ML;
MODEL y = d1 d2 d3 / SOLUTION;
ESTIMATE "Control Mean" intercept 1 d1 0 d2 0 d3 0;
ESTIMATE "T1 Mean"      intercept 1 d1 1 d2 0 d3 0;
ESTIMATE "T2 Mean"      intercept 1 d1 0 d2 1 d3 0;
ESTIMATE "T3 Mean"      intercept 1 d1 0 d2 0 d3 1;
ESTIMATE "Control vs. T1"     d1  1 d2  0 d3 0;
ESTIMATE "Control vs. T2"     d1  0 d2  1 d3 0;
ESTIMATE "Control vs. T3"     d1  0 d2  0 d3 1;
ESTIMATE "T1 vs. T2"          d1 -1 d2  1 d3 0;
ESTIMATE "T1 vs. T3"          d1 -1 d2  0 d3 1;
ESTIMATE "T2 vs. T3"          d1  0 d2 -1 d3 1;
RUN;
```

Intercepts are used only in predicted values.

Positive values indicate addition; negative values indicate subtraction.

Creating Predicted Outcomes

- Three ways (in order of most to least painful):

1. In excel: input fixed effects, input variable values, write equation to create predicted outcomes
   - Good for pedagogy, but gets old quickly (and error-prone)

2. Via programming statements:
   - Per prediction: Use SAS ESTIMATE or SPSS TEST
   - For a range of predictor values: Use STATA MARGINS

3. Via "fake people"
   - Add cases to your data with desired predictor values
   - Ask program to save predicted outcomes for all cases
   - Fake cases won't contribute to model, but will get predictions