Power Analysis via Simulation in SAS and Mplus

```
*ORIGINAL DATA FROM http://www.ats.ucla.edu/stat/sas/library/repeated_ut.htm;
OPTIONS NONUMBER NODATE NOCENTER PAGESIZE=MAX LINESIZE=120 formdlim='-';
* Kill default titles on output;
TITLE; ODS LISTING; ODS HTML CLOSE;

*****************************************************************************;
***** Make MVN data;
*****************************************************************************;
%MACRO mvn(varcov=,       /* dataset for variance-covariance matrix */
           means=,        /* dataset for mean vector */
           n=,            /* sample size */
           seed=,         /* seed for random number generator */
           sample= );     /* output dataset name */
* Get initial seed value. If seed<=0, then generate seed from system clock;
data _null_;
   if &seed. le 0 then do;
      seed = int(time());         /* get clock time in integer seconds */
      put seed=;
      call symput('seed',seed);   /* store seed as macro variable */
   end;
run;
* Generate the multivariate normal data in SAS/IML;
proc iml worksize=100;
   use &varcov.;                  /* read variance-covariance matrix */
   read all into cov;
   use &means.;                   /* read means */
   read all into mu;
   v=nrow(cov);                   /* calculate number of variables */
   n=&n.;
   seed = &seed;
   l=t(root(cov));                /* calculate cholesky root of cov matrix */
   z=normal(j(v,&n.,&seed.));     /* generate nvars*samplesize normals */
   x=l*z;                         /* premultiply by cholesky root */
   x=repeat(mu,1,&n.)+x;          /* add in the means */
   tx=t(x);
   create &sample. from tx;       /* write out sample data to sas dataset */
   append from tx;
quit;
%MEND mvn;

*VALUE LABELS FOR EACH LEVEL OF OUR CATEGORICAL VARIABLES;
PROC FORMAT;
VALUE exercises
    1 = "Aerobic Stair Climbing"
    2 = "Racquetball"
    3 = "Weight Training";
VALUE diets
    1 = "Meat Eaters"
    2 = "Vegetarians";
VALUE intensities
    1 = "Pulse: Warm Up"
    2 = "Pulse: Jogging"
    3 = "Pulse: Running";
RUN;

DATA work.dietwide;
INPUT exertype pulse1 pulse2 pulse3 diet personID;
*LABELING VARIABLES;
LABEL exertype = "Exercise Type"
      pulse1   = "Pulse After Warmup"
      pulse2   = "Pulse After Jogging"
      pulse3   = "Pulse After Running"
      diet     = "Diet Type";
```

```
*ADDING VALUE LABELS TO VARIABLES;
FORMAT exertype exercises. diet diets.;

*CREATING DUMMY CODED VARIABLES FOR EXERCISE TYPE:;
IF exertype = 1 THEN DO; dEXERCISE_ASC = 1; dEXERCISE_R = 0; dEXERCISE_WT = 0; END;
IF exertype = 2 THEN DO; dEXERCISE_ASC = 0; dEXERCISE_R = 1; dEXERCISE_WT = 0; END;
IF exertype = 3 THEN DO; dEXERCISE_ASC = 0; dEXERCISE_R = 0; dEXERCISE_WT = 1; END;

*CREATING DUMMY CODED VARIABLES FOR DIET TYPE;
IF diet = 1 THEN DO; dDIET_M = 1; dDIET_V = 0; END;
IF diet = 2 THEN DO; dDIET_M = 0; dDIET_V = 1; END;

*IMPORTING DATA;
DATALINES;
1          112        166        215        1        1
1          111        166        225        1        2
1          89         132        189        1        3
1          95         134        186        2        4
1          66         109        150        2        5
1          69         119        177        2        6
2          125        177        241        1        7
2          85         117        186        1        8
2          97         137        185        1        9
2          93         151        217        2       10
2          77         122        178        2       11
2          78         119        173        2       12
3          81         134        205        1       13
3          88         133        180        1       14
3          88         157        224        1       15
3          58          99        131        2       16
3          85         132        186        2       17
3          78         110        164        2       18
;
RUN;

*CONVERTING DATA TO STACKED FORM FOR PROC MIXED;
DATA WORK.dietstack;
    SET WORK.dietwide;
      FORMAT intensity intensities.; *ADDING A FORMAT STATEMENT FOR INTENSITY VARIABLE;
*FIRST OUTCOME: PULSE 1 (AFTER WARM UP);
      pulse = pulse1;
    intensity = 1;
      dINTENSITY_W = 1; dINTENSITY_J = 0; dINTENSITY_R = 0; *DUMMY CODED VARIABLES FOR ANALYSIS;
    OUTPUT; *OUTPUT MAKES THE LINE OF DATA GET WRITTEN TO THE NEW DATA SET;
*SECOND OUTCOME: PULSE 2 (AFTER JOGGING);
    pulse = pulse2;
    intensity = 2;
      dINTENSITY_W = 0; dINTENSITY_J = 1; dINTENSITY_R = 0;
    OUTPUT;
*THIRD OUTCOME: PULSE 3 (AFTER RUNNING);
    pulse = pulse3;
    intensity = 3;
      dINTENSITY_W = 0; dINTENSITY_J = 0; dINTENSITY_R = 1;
    OUTPUT;
RUN;

*ORIGINAL ANALYSIS - ADDING LSMEANS STATEMENT FOR PREDICTED MEANS FOR EACH CONDITION;
TITLE "MULTIVARIATE MODEL WITH DIET PREDICTOR ONLY - CLASS STATEMENT";
PROC MIXED DATA=WORK.dietstack METHOD=ML COVTEST NOPROFILE ITDETAILS IC NAMELEN=50;
CLASS intensity diet exertype;
MODEL pulse = exertype|intensity|diet / S DDFM=KENWARDROGER;
REPEATED intensity / SUBJECT=personID TYPE=UN R RCORR;
LSMEANS exertype|intensity|diet;
RUN;
TITLE;
```

```
*FOR DATA GENERATION LATER;
DATA predmeans;
INPUT diet exertype pulse1p pulse2p pulse3p;
DATALINES;
1 1 104.00 154.67 209.67
1 2 102.33 143.67 204.00
1 3 85.6667 141.33 203.00
2 1 76.6667 120.67 171.00
2 2 82.6667 130.67 189.33
2 3 73.6667 113.67 160.33
;
RUN;

*STEP #1: GENERATE ERROR TERMS FROM MULTIVARIATE NORMAL DISTRIBUTION USING MULTNORM MACRO:;
DATA rmatrix;
INPUT col1-col3;
DATALINES;
126.41      162.87      185.76
162.87      250.67      290.13
185.76      290.13      395.22
;
RUN;

DATA meanvec;
INPUT col1;
DATALINES;
0
0
0
;
RUN;

*BEGIN DATA MAKING MACRO FOR REPLICATIONS;
DATA mvnall;
RUN;

%MACRO makedata(reps);

%DO i=1 %TO &reps.;
      %mvn(varcov=rmatrix,means=meanvec,n=18,seed=&i.,sample=mvnerror);

      DATA mvnerror; SET mvnerror;
            error_warmup = col1;
            error_jogging = col2;
            error_running = col3;
            personID = _N_;
            DROP col1-col3;
            replication = &i.;
      RUN;

      DATA mvnall;
            SET mvnall mvnerror;
            IF personID=. THEN DELETE;
      RUN;
%END;

%MEND;

%MAKEDATA(100);
```

```
*STEP 2: MERGE MVN ERROR TERMS WITH ORIGINAL DATA;
PROC SORT DATA=mvnall; BY personID; RUN;

DATA merge1;
       MERGE mvnall dietwide;
       BY personID;
       IF personID=. THEN DELETE;
RUN;

PROC SORT DATA=merge1;
BY exertype diet;
RUN;

PROC SORT DATA=predmeans;
BY exertype diet;
RUN;

DATA merge2;
       MERGE merge1 predmeans;
       BY exertype diet;
       pulse1 = pulse1p+error_warmup;
       pulse2 = pulse2p+error_jogging;
       pulse3 = pulse3p+error_running;
RUN;

*CONVERTING DATA TO STACKED FORM FOR PROC MIXED;
DATA stackreplications;
    SET merge2;

*FIRST OUTCOME: PULSE 1 (AFTER WARM UP);
       pulse = pulse1;
    intensity = 1;
    OUTPUT;

*SECOND OUTCOME: PULSE 2 (AFTER JOGGING);
    pulse = pulse2;
    intensity = 2;
    OUTPUT;

*THIRD OUTCOME: PULSE 3 (AFTER RUNNING);
    pulse = pulse3;
    intensity = 3;
    OUTPUT;
RUN;

PROC SORT DATA=stackreplications;
BY replication;
RUN;

*ORIGINAL ANALYSIS - ADDING LSMEANS STATEMENT FOR PREDICTED MEANS FOR EACH CONDITION;
TITLE "MULTIVARIATE MODEL WITH DIET PREDICTOR ONLY - CLASS STATEMENT";
PROC MIXED DATA=WORK.stackreplications METHOD=ML COVTEST NOPROFILE ITDETAILS IC NAMELEN=50;
BY replication;
CLASS intensity diet exertype;
MODEL pulse = exertype|intensity|diet / S DDFM=KENWARDROGER;
REPEATED intensity / SUBJECT=personID TYPE=UN R RCORR;
ODS OUTPUT Tests3 = Tests3;
RUN;TITLE;

DATA Tests3; SET Tests3;
       IF ProbF LT .05 THEN Signif=1; ELSE Signif=0;
RUN;

PROC SORT DATA=Tests3; BY effect; RUN;
TITLE "Proportion Significant Omnibus F-test for Group";
PROC MEANS DATA=Tests3 MEAN; CLASS effect; VAR Signif; RUN;
```

Mplus also offers power analysis via simulation using the MONTECARLO command. Two examples are below.

1. Post-Hoc Power Analysis from Hoffman & McDowd (2010, *Psychology and Aging*)

The code and output for the original model (as presented in lecture 20) is given below, now explicitly referencing all estimated parameters. In addition to the path model direct effects (and indirect effect from before), we are also estimating (by default) intercepts and residual variances for continuous outcomes and thresholds for categorical outcomes. Given that we are using a logit link we also estimate a threshold for each binary outcome, which is the logit of the probability of a 0 response when all predictors are 0 (i.e., it is the opposite of an intercept, which would provide the logit of the probability of a 1 response). Thresholds are an Mplus convention that some other programs use as well. By definition, though, the binary outcomes do not have estimated residual variances (they are determined by the conditional probability of each response in binary outcomes).

```
TITLE:   Path Analysis for Dissertation Follow-up
DATA:    FILE = driver.dat;

VARIABLE:
! List of variables in data file
  NAMES = PartID sex age75 cs_1_5 cs_3 cs_6 cs_12 cs_18 far near
          zufov1 zufov2 zufov3 Dscan lane da_task crash stop speed time
          simfac part visfac attfac limit4 ticket2 speed2 follow attr
          nacc2 jacc2 jacc20 acc2;
! Variables to be analyzed in this model
  USEVARIABLE = sex age75 visfac zufov1 zufov2 zufov3 Dscan simfac
                limit4 speed2 acc2;
! Missing data identifier
  MISSING = .;
! Categorical outcomes
  CATEGORICAL = acc2 speed2;

ANALYSIS: LINK = LOGIT; ESTIMATOR = MLR; INTEGRATION = MONTECARLO;
OUTPUT:   STDYX;

MODEL:
! Path model direct effects (labels for each path in parentheses)
  simfac ON sex age75 visfac zufov1 zufov2 zufov3 Dscan               (sim1-sim7);
  limit4 ON sex age75 visfac zufov1 zufov2 zufov3 Dscan simfac        (lim1-lim8);
  acc2   ON sex age75 visfac zufov1 zufov2 zufov3 Dscan simfac limit4 (acc1-acc9);
  speed2 ON sex age75 visfac zufov1 zufov2 zufov3 Dscan simfac limit4 (spd1-spd9);


! Outcome intercepts (for continuous variables)
  [simfac limit4];
! Outcome thresholds (for categorical variables)
  [speed2$1 acc2$1];
! Estimated residual variances for continuous outcomes
  simfac limit4;

MODEL CONSTRAINT:       ! Like ESTIMATE in SAS
  NEW(DStoAcc);           ! List names of estimated effects on NEW
  DStoAcc = sim7 * acc8;  ! Indirect effect of Dscan --> Sim --> Acc
```

The model output has some of the necessary information we need to do a post-hoc power analysis. We also need the means and variances of the predictors (that are not estimated in the likelihood) from the original data as well.

	UNSTANDARDIZED				STANDARDIZED STDYX			
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SIMFAC ON								
SEX	0.008	0.121	0.067	0.946	0.005	0.078	0.067	0.946
AGE75	-0.002	0.014	-0.111	0.911	-0.009	0.083	-0.111	0.911
VISFAC	-0.021	0.074	-0.287	0.774	-0.025	0.087	-0.286	0.775
ZUFOV1	0.033	0.069	0.482	0.630	0.043	0.089	0.482	0.630
ZUFOV2	0.085	0.086	0.992	0.321	0.111	0.110	1.003	0.316
ZUFOV3	0.105	0.082	1.283	0.200	0.135	0.105	1.277	0.201
DSCAN	0.216	0.081	2.661	0.008	0.279	0.104	2.681	0.007
LIMIT4 ON								
SEX	0.304	0.200	1.522	0.128	0.154	0.102	1.511	0.131
AGE75	-0.001	0.024	-0.054	0.957	-0.006	0.114	-0.054	0.957
VISFAC	0.009	0.108	0.080	0.937	0.008	0.099	0.080	0.937
ZUFOV1	0.098	0.127	0.772	0.440	0.100	0.129	0.775	0.438
ZUFOV2	0.149	0.124	1.205	0.228	0.152	0.124	1.225	0.221
ZUFOV3	-0.163	0.139	-1.178	0.239	-0.166	0.141	-1.181	0.238
DSCAN	0.154	0.131	1.178	0.239	0.157	0.134	1.171	0.242
SIMFAC	0.026	0.155	0.169	0.866	0.021	0.122	0.169	0.866
ACC2 ON								
SEX	-0.770	0.560	-1.374	0.169	-0.162	0.121	-1.341	0.180
AGE75	0.058	0.083	0.700	0.484	0.114	0.167	0.680	0.496
VISFAC	0.066	0.309	0.213	0.831	0.025	0.117	0.215	0.830
ZUFOV1	-0.438	0.608	-0.720	0.472	-0.186	0.240	-0.775	0.438
ZUFOV2	-0.822	0.379	-2.171	0.030	-0.350	0.149	-2.342	0.019
ZUFOV3	-0.066	0.391	-0.169	0.866	-0.028	0.165	-0.169	0.865
DSCAN	-0.477	0.320	-1.491	0.136	-0.202	0.130	-1.552	0.121
SIMFAC	1.497	0.532	2.813	0.005	0.492	0.131	3.759	0.000
LIMIT4	-0.387	0.411	-0.943	0.346	-0.161	0.161	-1.001	0.317
SPEED2 ON								
SEX	-1.387	0.749	-1.852	0.064	-0.290	0.129	-2.245	0.025
AGE75	-0.065	0.051	-1.265	0.206	-0.126	0.097	-1.299	0.194
VISFAC	-0.180	0.343	-0.525	0.599	-0.068	0.125	-0.549	0.583
ZUFOV1	-0.475	0.483	-0.982	0.326	-0.200	0.177	-1.133	0.257
ZUFOV2	-0.289	0.610	-0.474	0.636	-0.122	0.242	-0.503	0.615
ZUFOV3	-1.136	0.370	-3.067	0.002	-0.477	0.149	-3.195	0.001
DSCAN	0.968	0.518	1.868	0.062	0.408	0.174	2.345	0.019
SIMFAC	-0.382	0.682	-0.561	0.575	-0.125	0.208	-0.599	0.549
LIMIT4	-0.343	0.480	-0.715	0.475	-0.141	0.205	-0.690	0.490
Intercepts								
SIMFAC	0.016	0.090	0.174	0.862	0.020	0.116	0.174	0.862
LIMIT4	0.062	0.160	0.389	0.697	0.064	0.163	0.392	0.695
Thresholds								
SPEED2$1	1.727	0.507	3.407	0.001	0.728	0.177	4.114	0.000
ACC2$1	2.059	0.546	3.772	0.000	0.875	0.197	4.440	0.000
Residual Variances								
SIMFAC	0.479	0.052	9.140	0.000	0.799	0.067	11.880	0.000
LIMIT4	0.876	0.122	7.169	0.000	0.914	0.059	15.467	0.000
New/Additional Parameters								
DSTOACC	0.323	0.160	2.026	0.043				

The code and output for the power analysis simulation model is given below.

```
TITLE:  Power Analysis for DriverScan2

MONTECARLO:
! Names of to-be-created variables in model
  NAMES = sex age75 visfac zufov1 zufov2 zufov3 Dscan simfac limit4 speed2 acc2;
! For categorical variables, how many thresholds (=#categories-1) and logit/probit
  GENERATE = speed2(1 L) acc2(1 L);
  CATEGORICAL = speed2 acc2;
! Number of people in each sample -- I used # complete cases for outcomes
  NOBSERVATIONS = 114;
! Number of samples to create
  NREPS = 1000;
! Random number seed to duplicate results
  SEED = 8675309;

ANALYSIS: LINK = LOGIT; ESTIMATOR = MLR; INTEGRATION = MONTECARLO;

MODEL POPULATION:  ! Tell Mplus how to make the data
! Using @ for parameters that are not going to be estimated in the likelihood
! Using * for parameters that are going to be estimated in the likelihood

! Predictor means (from original data)
    [sex@.579 age75@-.094 visfac@-.090];
    [zufov1@-.072 zufov2@-.025 zufov3@.059 dscan@-.051];
! Predictor variances (from original data)
    sex@.246 age75@22.455 visfac@.779;
    zufov1@.809 zufov2@.908 zufov3@1.030 dscan@.963;

! Outcome intercepts (estimated for continuous variables)
    [simfac*.016]; [limit4*.062];
! Outcome thresholds (estimated for categorical variables)
    [speed2$1*1.727]; [acc2$1*2.059];
! Residual variances (estimated for continuous outcomes)
    simfac*.479 limit4*.876;

! Path model direct effects (unstandardized effects estimated by model)
   simfac ON sex*.008 age75*-.002 visfac*-.021;
   simfac ON zufov1*.033 zufov2*.085 zufov3*.105 Dscan*.216;
   limit4 ON sex*.304 age75*-.001 visfac*.009;
   limit4 ON zufov1*.098 zufov2*.149 zufov3*-.163 Dscan*.154 simfac*.026;
   acc2   ON sex*-.770 age75*.058 visfac*.066 zufov1*-.438 zufov2*-.822;
   acc2   ON zufov3*-.066 Dscan*-.477 simfac*1.497 limit4*-.387;
   speed2 ON sex*-1.387 age75*-.065 visfac*-.180 zufov1*-.475 zufov2*-.289;
   speed2 ON zufov3*-1.136 Dscan*.968 simfac*-.382 limit4*-.343;

MODEL:  ! Tell Mplus to run the model on each created sample dataset

simfac ON sex*.008 age75*-.002 visfac*-.021                              (sim1-sim3);
simfac ON zufov1*.033 zufov2*.085 zufov3*.105 Dscan*.216                 (sim4-sim7);
limit4 ON sex*.304 age75*-.001 visfac*.009                               (lim1-lim3);
limit4 ON zufov1*.098 zufov2*.149 zufov3*-.163 Dscan*.154 simfac*.026    (lim4-lim8);
acc2   ON sex*-.770 age75*.058 visfac*.066 zufov1*-.438 zufov2*-.822     (acc1-acc5);
acc2   ON zufov3*-.066 Dscan*-.477 simfac*1.497 limit4*-.387             (acc6-acc9);
speed2 ON sex*-1.387 age75*-.065 visfac*-.180 zufov1*-.475 zufov2*-.289  (spd1-spd5);
speed2 ON zufov3*-1.136 Dscan*.968 simfac*-.382 limit4*-.343             (spd6-spd9);

MODEL CONSTRAINT:   ! Like ESTIMATE in SAS
  NEW(DStoAcc*.323);      ! List names of estimated effects on NEW
  DStoAcc = sim7 * acc8;  ! Indirect effect of Dscan --> Sim --> Acc
```

		SUMMARY OF ESTIMATES				SUMMARY OVER REPLICATIONS		
SIMFAC	ON	**Population**	**Average**	**Std. Dev.**	**S. E. Average**	**M.S.E.**	**95% Cover**	**% Sig Coeff**
SEX		0.008	0.013	0.133	0.129	0.018	0.943	0.054
AGE75		-0.002	-0.001	0.014	0.014	0.000	0.946	0.065
VISFAC		-0.021	-0.020	0.077	0.073	0.006	0.930	0.077
ZUFOV1		0.033	0.033	0.077	0.071	0.006	0.924	0.115
ZUFOV2		0.085	0.083	0.070	0.067	0.005	0.936	0.258
ZUFOV3		0.105	0.108	0.067	0.063	0.005	0.929	0.414
DSCAN		0.216	0.214	0.072	0.065	0.005	0.915	**0.860**
LIMIT4	ON							
SEX		0.304	0.309	0.184	0.174	0.034	0.932	0.433
AGE75		-0.001	-0.001	0.019	0.018	0.000	0.939	0.059
VISFAC		0.009	0.009	0.107	0.098	0.011	0.928	0.071
ZUFOV1		0.098	0.093	0.103	0.096	0.011	0.922	0.189
ZUFOV2		0.149	0.143	0.095	0.091	0.009	0.929	0.354
ZUFOV3		-0.163	-0.162	0.091	0.086	0.008	0.938	0.481
DSCAN		0.154	0.158	0.102	0.092	0.011	0.923	0.422
SIMFAC		0.026	0.019	0.135	0.125	0.018	0.915	0.087
ACC2	ON							
SEX		-0.770	-0.964	1.109	0.794	1.266	0.918	0.247
AGE75		0.058	0.069	0.103	0.081	0.011	0.913	0.151
VISFAC		0.066	0.105	0.549	0.424	0.303	0.915	0.098
ZUFOV1		-0.438	-0.560	0.550	0.436	0.317	0.919	0.273
ZUFOV2		-0.822	-1.051	0.591	0.453	0.402	0.927	**0.671**
ZUFOV3		-0.066	-0.105	0.481	0.377	0.232	0.914	0.087
DSCAN		-0.477	-0.603	0.528	0.420	0.294	0.917	0.299
SIMFAC		1.497	1.917	0.959	0.672	1.096	0.923	**0.864**
LIMIT4		-0.387	-0.498	0.542	0.410	0.306	0.909	0.244
SPEED2	ON							
SEX		-1.387	-1.800	1.740	0.895	3.193	0.916	**0.567**
AGE75		-0.065	-0.090	0.147	0.085	0.022	0.917	0.209
VISFAC		-0.180	-0.291	1.720	0.450	2.969	0.900	0.123
ZUFOV1		-0.475	-0.657	0.678	0.453	0.492	0.915	0.294
ZUFOV2		-0.289	-0.371	0.594	0.426	0.359	0.911	0.175
ZUFOV3		-1.136	-1.523	1.286	0.516	1.801	0.909	**0.894**
DSCAN		0.968	1.352	1.801	0.520	3.386	0.921	**0.745**
SIMFAC		-0.382	-0.442	1.250	0.577	1.564	0.916	0.145
LIMIT4		-0.343	-0.454	0.754	0.428	0.580	0.905	0.212
Intercepts								
SIMFAC		0.000	0.011	0.101	0.099	0.010	0.936	0.064
LIMIT4		0.000	0.056	0.147	0.133	0.025	0.905	0.095
Thresholds								
SPEED2$1		0.000	2.390	4.328	0.680	24.429	0.018	**0.982**
ACC2$1		0.000	2.611	0.994	0.685	7.802	0.007	**0.993**
Residual Variances								
SIMFAC		0.500	0.445	0.062	0.058	0.007	0.766	**1.000**
LIMIT4		0.500	0.802	0.113	0.104	0.104	0.121	**1.000**
New/Additional Parameters								
DSTOACC		0.323	0.408	0.253	0.196	0.071	0.931	**0.541**

BOLD VALUES INDICATE EFFECTS THAT WERE ~SIGNIFICANT IN ORIGINAL MODEL.

2. A Priori Power Analysis for Path Model (from grant analysis plan for Scott Stoltenberg)

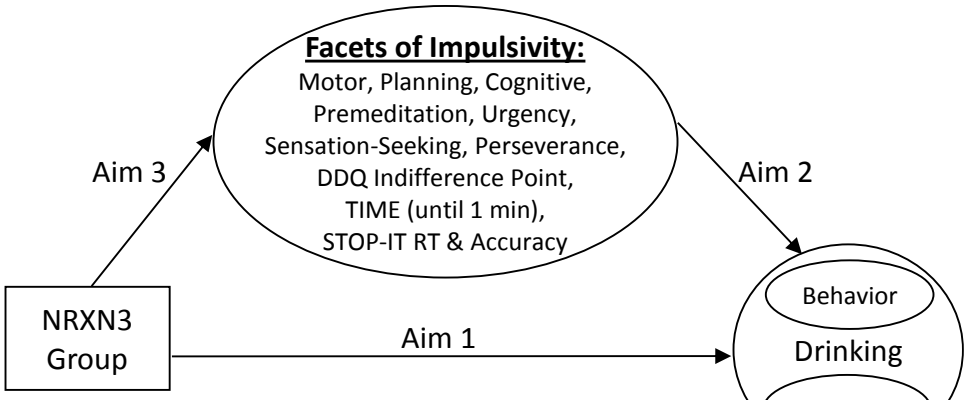

Figure 3. Path diagram specifying the hypotheses to be tested in Specific Aims 1-3. NRXN3 represents genotype at NRXN3 SNPs. Latent traits will be estimated for the facets of impulsivity and high-risk drinking.

This conceptual model illustrates the basic premise: the kind of genes you have "causes" impulsivity which then "causes" behavioral outcomes. Because this is a proposal, we have little information with which to make guesses about model parameters, but we'll use the equations that underlie the mediation model. Here's the end result:

```
TITLE: Figure 3 model a priori power analysis
MONTECARLO:
! Names of to-be-created variables in model
  NAMES = drink impulse gene;
! For categorical predictors, where to cut a continuous variable to make groups
  CUTPOINTS = gene(0.44);
! Number of people proposed in each sample
  NOBSERVATIONS = 300;
! Number of samples to create
  NREPS = 1000;
! Random number seed to duplicate results
  SEED = 8675309;
ANALYSIS:   ESTIMATOR = MLR;
MODEL POPULATION:   ! Tell Mplus how to make the data
! Using @ for parameters that are not going to be estimated in the likelihood
! Using * for parameters that are going to be estimated in the likelihood
! Mean and variance for original gene variable to be cut up into two groups
    [gene@0]; gene@1;
! Intercepts and residual variances for impulse mediator and drink outcome
    [impulse*0 drink*0];
    impulse*.910  drink*.938;
! Mediation direct effects
    impulse ON gene*.638    (a);
    drink   ON impulse*.150 (b);
    drink   ON gene*.425    (c);

MODEL:  ! Tell Mplus to estimate our model
! Intercepts and residual variances for impulse mediator and drink outcome
    [impulse*0 drink*0];
    impulse*.910  drink*.938;
! Mediation direct effects
    impulse ON gene*.638    (a);
    drink   ON impulse*.150 (b);
    drink   ON gene*.425    (c);

MODEL CONSTRAINT:   ! Like ESTIMATE in SAS
  NEW(XtoMtoY*.096);    ! List names of estimated effects on NEW
  XtoMtoY = a * b;      ! Indirect effect of Gene --> Impulse --> Drink
```

So where did these numbers come from???
Start with a standardized version of each variable (mean = 0, variance = 1).

1. Choose **size of standardized paths** for each unique direct effect:
   X → M = $a = .30$ M → Y = $b = .15$ X→ Y = $c = .20$

2. **Calculate variance and Z-score for X based on chosen proportion of ~33% with gene:**
   $Mean(X) = .33$, so $Var(X) = .33 * (1 - .33) = .\mathbf{221}$
   If $probability(gene = 1) = .33$ → $Zscore$ for area to the right $= .\mathbf{44}$

3. **Calculate model-predicted residual variance for M:**
   if $M = aX + e^m$, then $Var(M) = a^2Var(X) + Var(e^m)$
   $Var(e^m) = Var(M) - a^2Var(X)$, or $Var(e^m) = 1 - .30^2 = .\mathbf{910}$

4. **Calculate model-predicted residual variance for Y:**
   if $Y = cX + bM + e^y$, then $Y = cX + b(aX + e^m) + e^y$, then
   $Var(Y) = c^2Var(X) + b^2a^2Var(X) + b^2Var(e^m) + Var(e^y)$
   $Var(e^y) = Var(Y) - c^2Var(X) - b^2a^2Var(X) - b^2Var(e^m)$
   $Var(e^y) = 1 - (.20^2) - (.15^2 * .30^2) - (.15^2 .91) = .\mathbf{938}$

5. **Calculate unstandardized paths from results of steps 1–4 to use in MODEL POPULATION:**

$$a^{uns} = \frac{a\sqrt{Var(M)}}{\sqrt{Var(X)}} = \frac{.30\sqrt{1}}{\sqrt{.221}} = .638$$

$$b^{uns} = \frac{b\sqrt{Var(Y)}}{\sqrt{Var(M)}} = \frac{.15\sqrt{1}}{\sqrt{1}} = .150$$

$$c^{uns} = \frac{c\sqrt{Var(Y)}}{\sqrt{Var(X)}} = \frac{.20\sqrt{1}}{\sqrt{.221}} = .425$$

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
IMPULSE ON							
GENE	0.638	0.6333	0.1220	0.1167	0.0149	0.942	**1.000**
DRINK ON							
IMPULSE	0.150	0.1528	0.0585	0.0586	0.0034	0.952	**0.747**
GENE	0.425	0.4205	0.1228	0.1244	0.0151	0.958	**0.929**
Intercepts							
DRINK	0.000	-0.0023	0.0653	0.0680	0.0043	0.954	0.046
IMPULSE	0.000	-0.0004	0.0653	0.0668	0.0043	0.951	0.049
Residual Variances							
DRINK	0.938	0.9284	0.0783	0.0753	0.0062	0.935	1.000
IMPULSE	0.910	0.8999	0.0712	0.0728	0.0052	0.936	1.000
New/Additional Parameters							
XTOMTOY	0.096	0.0969	0.0421	0.0418	0.0018	0.938	**0.686**