Simulation-Based Power Analysis

Tutorial material from IQSS

1. Introduction

1.1 Basic Concepts

It is an important step to calculate statistical power in a research design. In a research design, we use statistical power to measure the probability that a null hypothesis is correctly rejected. Usually, researchers need to know the needed sample size to reject the null hypothesis at a given power level, while in other cases, people calculates the power when the sample size is fixed.

More often, in a randomized controlled trial with two groups, we can use a formula to calculate the needed sample size to reject the null hypothesis. We will use an example to show how we do this. For instance, when we plan to perform a test of a hypothesis comparing the proportions of successes of tossing coins of faces in two independent populations, we would list the following null and alternative hypothesis respectively:

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

where $p_1 = p_2$ are the proportions in the two populations for comparison. In order to make sure the test has a specific power, we can use the following formula to determine the sample sizes:

$$N = 2(\frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{ES})^2$$

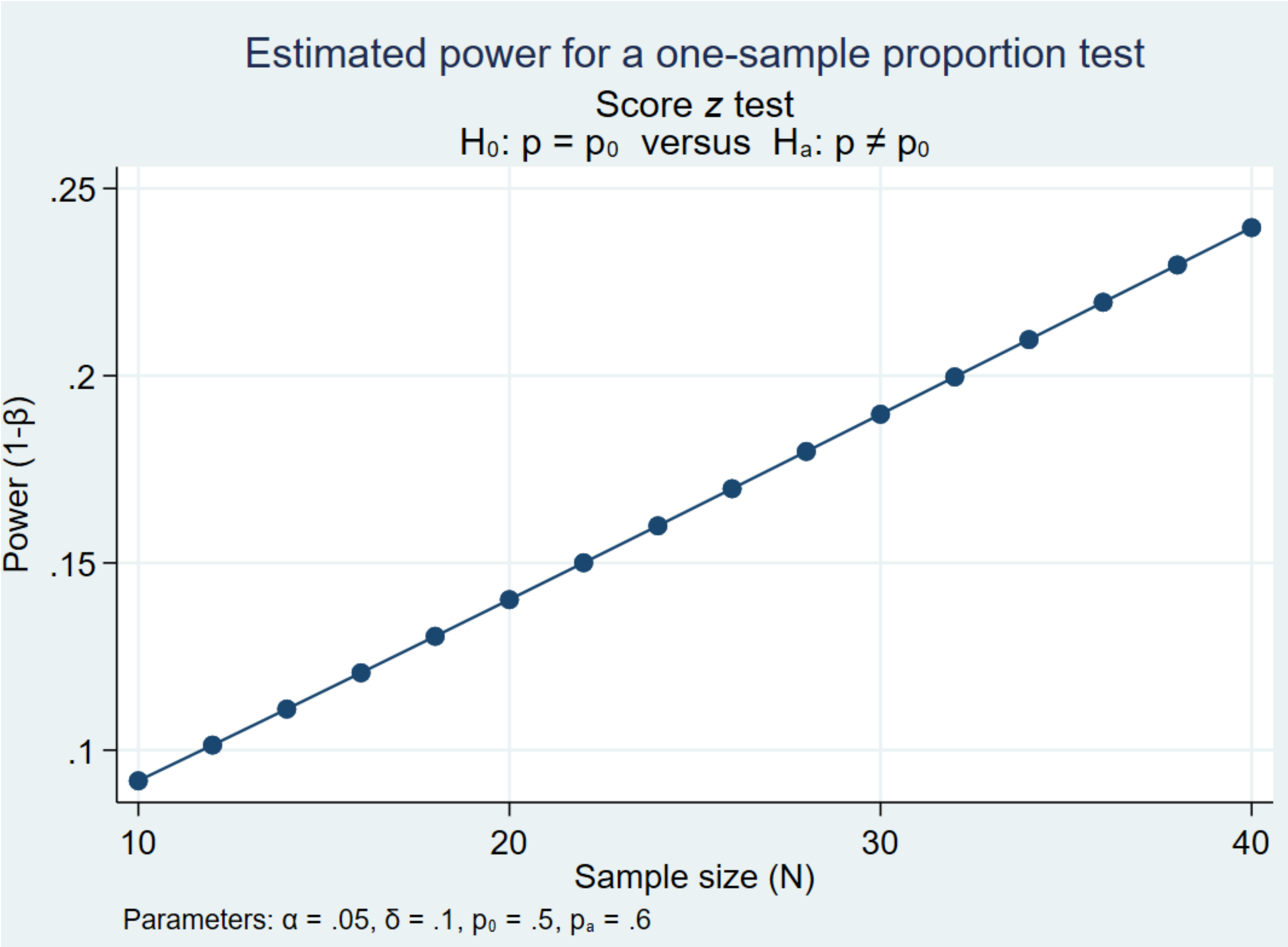
Where n_i is the sample size required in each group (i=1,2), α is the specific level of significance and $z_{1-\frac{\alpha}{2}}$ is the critical value corresponding to the significance level. $1 - \beta$ is the selected power and $z_{1-\beta}$ is the value from the standard normal distribution holding $1 - \beta$ below it. ES is the effect size, defined as follows:

$$ES = \frac{|p_1 = p_2|}{\sqrt{p(1-p)}}$$

where $|p_1 = p_2|$ is the absolute value of the proportions difference between the two groups holding under the alternative hypothesis, H_1, and p is the proportion by pooling the observations from the two comparison groups.

In Stata, we use the following code to calculate the sample size needed to reject the null hypothesis that $H_0 : p = 0.5$ ($H_1 : p = 0.6$) on different fixed power levels:

```
power oneproportion 0.5 0.6, n(10(2)40) graph
```



1.2 Procedures to perform power analysis

1. Specify a hypothesis test.

Usually, there are several hypotheses in a research design, but for sample size calculation, make explicit a null and alternative hypothesis.

2. Specify the significance level of the test.

It is usually α = .05, but other values could be taken too.

3. Get the values of the parameters necessary to compute the power function.

To solve for sample size n, we need a value for standard deviation and other parameters. Need to note, sometimes we need to use a pilot dataset to get these values.

4. Specify the intended power of the test.

The power of a test is the probability of finding significance if the alternative hypothesis is true.

5. Calculate the needed sample size for a fixed power level.

2. Power Analysis with simulation

Nevertheless, formulas don't always work out to calculate the needed sample size such as in complex study designs. In these cases, simulation based power analysis stand out. The basic idea is to simulate running the experiment many times and calculate the proportion of times we reject the null hypothesis. This proportion provides an estimate of power. Generating a dataset and running an analysis for the hypothesis test is part of the simulation. One thing to mention is that randomness is usually introduced into the process through the dataset generation.

For example, say, the fixed power level is 95%, and you want to calculate the sample size on this level. You can take a "guess and check" method. With this method, firstly, you choose a sample size n_1 and run the simulation to estimate your power. If power is estimated to be lower than 95%, you need to select a new value n_2 that is larger than n_1 running the simulation again. Multiple procedures are repeated until the estimated power is roughly 95%.

As the example shows in the introduction part, for multiple commonly used statistical tests, we can use Stata's power commands to calculate power and needed sample size. However, for complex models, such as multilevel or mixed effect models, we need to use simulations to calculate power and the needed sample size. In these scenarios, we usually use the following procedures to perform power analysis:

1.Write down the regression model of interest, including all parameters.

2.Specify the details of the covariates, such as the range of age or the proportion of females.

3.Locate or think about reasonable values for the parameters in your model.

4.Simulate a single dataset assuming the alternative hypothesis, and fit the model.

5.Write a program to create the datasets, fit the models, and use simulate to test the program.

6.Write a program to allows you to run your simulations with power.

7.Write a program that you can use numlists for all parameters.

3. Simulation-based Power Analysis in Stata

3.1 Simple linear regression

3.1.1 Write down the regression model of interest, including all parameters.

$$bpsystol = \beta_0 + \beta_1(age) + \beta_2(sex) + \beta_3(age * sex) + \epsilon$$

where the variables of interest are age, sex and the interaction of age and sex. Also, you need to estimate the coefficients for β_0, β_1, β_2, β_3.

3.1.2 Specify the details of the covariates.

You plan a study of systolic blood pressure (SBP) and you believe that there is an interaction between age and sex.

3.1.3 Locate or think about reasonable values for the parameters in your model.

```
webuse nhanes2

regress bpsystol c.age##ib1.sex

      Source |       SS           df       MS      Number of obs   =    10,351
-------------+----------------------------------   F(3, 10347)     =   1180.87
       Model |     1437147         3      479049   Prob > F        =    0.0000
    Residual |  4197523.03    10,347  405.675367   R-squared       =    0.2551
-------------+----------------------------------   Adj R-squared   =    0.2548
       Total |  5634670.03    10,350  544.412563   Root MSE        =    20.141

------------------------------------------------------------------------------
    bpsystol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |     .47062   .0167357    28.12   0.000     .4378147    .5034252
             |
         sex |
     Female  |  -20.45813   1.165263   -17.56   0.000    -22.74227   -18.17398
             |
   sex#c.age |
     Female  |   .3457346   .0230373    15.01   0.000      .300577    .3908923
             |
       _cons |   110.5691   .8440692   131.00   0.000     108.9146    112.2236
------------------------------------------------------------------------------
```

Using the data data from the National Health and Nutrition Examination Survey (NHANES), we can estimate β_0=110.6, β_1=0.47, β_2=-20.46, β_3=0.35.

3.1.4 Simulate a single dataset assuming the alternative hypothesis, and fit the model.

Next, we create a simulated dataset based on our assumptions about the model under the alternative hypothesis.

```
clear
set seed 15
set obs 100
generate age = runiformint(18,65)
generate female = rbinomial(1,0.5)
generate interact = age*female
generate e = rnormal(0,20)
generate sbp = 110 + 0.5*age + (-20)*female + 0.35*interact  + e
```

We can then test the null hypothesis that the interaction term equals zero using a likelihood-ratio test.

```
regress sbp age i.female c.age#i.female
estimates store full
regress sbp age i.female
estimates store reduced
```

```
Likelihood-ratio test                                 LR chi2(1)  =      0.68
(Assumption: reduced nested in full)                  Prob > chi2 =    0.4089
```

The test yields a p-value of 0.4089.

```
return list

scalars:
r(p) =  .4089399864598747
r(chi2) =  .6818803412616035
r(df) =  1
local reject = (r(p)<0.05)
```

3.1.5 Write a program to create the datasets, fit the models, and use simulate to test the program.

Next, let's write a program that creates datasets under the alternative hypothesis.

```
capture program drop simregress
program simregress, rclass
    version 16
    // DEFINE THE INPUT PARAMETERS AND THEIR DEFAULT VALUES
    syntax, n(integer)          /// Sample size
          [ alpha(real 0.05)    /// Alpha level
            intercept(real 110) /// Intercept parameter
            age(real 0.5)       /// Age parameter
            female(real -20)    /// Female parameter
            interact(real 0.35) /// Interaction parameter
            esd(real 20) ]      //  Standard deviation of the error
    quietly {
        // GENERATE THE RANDOM DATA
        clear
        set obs `n'
        generate age = runiformint(18,65)
        generate female = rbinomial(1,0.5)
        generate interact = age*female
        generate e = rnormal(0,`esd')
        generate sbp = `intercept' + `age'*age + `female'*female + ///
           `interact'*interact  + e
        // TEST THE NULL HYPOTHESIS
        regress sbp age i.female c.age#i.female
        estimates store full
        regress sbp age i.female
        estimates store reduced
        lrtest full reduced
    }
    // RETURN RESULTS
```

```
    return scalar reject = (r(p)<`alpha')
end
```

Below, we use simulate to run simregress 100 times and summarize the variable reject. The results indicate that we would have 16% power to detect an interaction parameter of 0.35 given a sample of 100 participants and the other assumptions about the model.

```
simulate reject=r(reject), reps(100):
>         simregress, n(100) age(0.5) female(20) interact(0.35)
>            esd(20) alpha(0.05)

      command:  simregress, n(100) age(0.5) female(20) interact(0.35)
> esd(20) alpha(0.05)
       reject:  r(reject)

Simulations (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

. summarize reject

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      reject |        100         .16    .3684529          0          1
```

3.1.6 Write a program called power_cmd_simregress, which allows you to run your simulations with power.

Next, let's write a program called power_cmd_simregress so that we can integrate simregress into Stata's power command.

```
capture program drop power_cmd_simregress
program power_cmd_simregress, rclass
    version 16
    // DEFINE THE INPUT PARAMETERS AND THEIR DEFAULT VALUES
    syntax, n(integer)          /// Sample size
          [ alpha(real 0.05)    /// Alpha level
            intercept(real 110) /// Intercept parameter
            age(real 0.5)       /// Age parameter
            female(real -20)    /// Female parameter
            interact(real 0.35) /// Interaction parameter
            esd(real 20)        /// Standard deviation of the error
            reps(integer 100)]  //   Number of repetitions

    // GENERATE THE RANDOM DATA AND TEST THE NULL HYPOTHESIS
    quietly {
        simulate reject=r(reject), reps(`reps'):              ///
             simregress, n(`n') age(`age') female(`female')    ///
```

```
                            interact(`interact') esd(`esd') alpha(`alpha')
        summarize reject
    }
    // RETURN RESULTS
    return scalar power = r(mean)
    return scalar N = `n'
    return scalar alpha = `alpha'
    return scalar intercept = `intercept'
    return scalar age = `age'
    return scalar female = `female'
    return scalar interact = `interact'
    return scalar esd = `esd'
end
```

3.1.7 Write a program called power_cmd_simregress_init.

Run power simregress for a range of input parameter values, including the parameters listed in double quotes.

```
capture program drop power_cmd_simregress_init
program power_cmd_simregress_init, sclass
    sreturn local pss_colnames "intercept age female interact esd"
    sreturn local pss_numopts  "intercept age female interact esd"
end
```

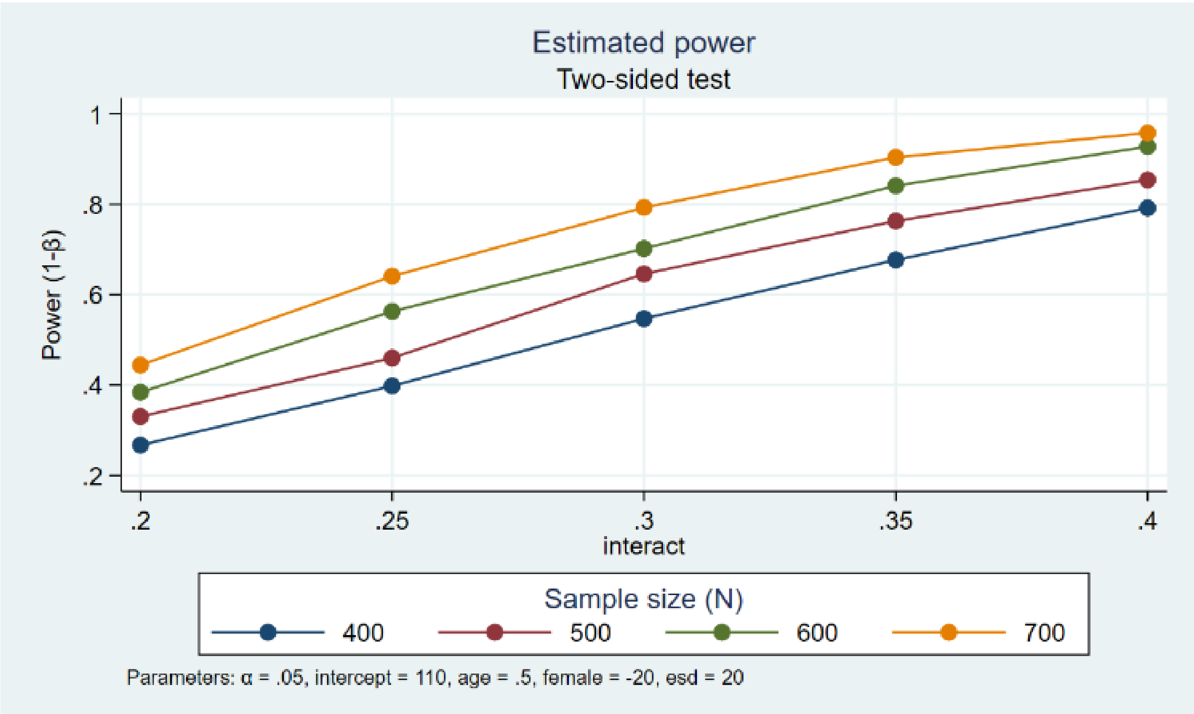
Now, we're ready to use power simregress! The output below shows the simulated power when the interaction parameter equals 0.2 to 0.4 in increments of 0.05 for samples of size 400, 500, 600, and 700.

```
power simregress, n(400(100)700) intercept(110)                 ///
                   age(0.5) female(-20) interact(0.2(0.05)0.4)   ///
                   reps(1000) table graph(xdimension(interact)   ///
                   legend(rows(1)))

Estimated power
Two-sided test

  +-------------------------------------------------------------------+
  |   alpha    power       N intercept     age   female interact     esd |
  |-------------------------------------------------------------------|
  |     .05     .267     400       110      .5      -20       .2      20 |
  |     .05     .398     400       110      .5      -20      .25      20 |
  |     .05     .547     400       110      .5      -20       .3      20 |
  |     .05     .677     400       110      .5      -20      .35      20 |
  |     .05     .792     400       110      .5      -20       .4      20 |
  |     .05      .33     500       110      .5      -20       .2      20 |
  |     .05      .46     500       110      .5      -20      .25      20 |
  |     .05     .646     500       110      .5      -20       .3      20 |
  |     .05     .763     500       110      .5      -20      .35      20 |
  |     .05     .854     500       110      .5      -20       .4      20 |
  |     .05     .384     600       110      .5      -20       .2      20 |
```

```
  |       .05     .563     600         110       .5      -20        .25        20 |
  |       .05     .702     600         110       .5      -20         .3        20 |
  |       .05     .841     600         110       .5      -20        .35        20 |
  |       .05     .928     600         110       .5      -20         .4        20 |
  |       .05     .444     700         110       .5      -20         .2        20 |
  |       .05     .641     700         110       .5      -20        .25        20 |
  |       .05     .793     700         110       .5      -20         .3        20 |
  |       .05     .904     700         110       .5      -20        .35        20 |
  |       .05     .958     700         110       .5      -20         .4        20 |
  +-------------------------------------------------------------------------------+
```



3.2 Fixed Effect models

3.2.1 Write down the regression model of interest, including all parameters.

$$weight_{it} = \beta_0 + \beta_1(age_{it}) + \beta_2(female_i) + \beta_3(age_{it} * female_i) + \mu_{0i} + \mu_{1i}(age) + \epsilon_{it}$$

where i stands for children, t for age, and we assume $\mu_{0i} \sim N(0, \tau_0)$, $\mu_{1i} \sim N(0, \tau_1)$, $\epsilon_{it} \sim N(0, \sigma)$, and $cov(\tau_0, \tau_1) = 0$.

The covariates are weight, age, female, and the interaction term age*female. Also, we need to estimate the coefficients for β_0, β_1, β_2, β_3,τ_0,τ_1 and σ.

3.2.2 Specify the details of the covariates, such as the range of age or the proportion of females.

Let's assume that we will measure the children's weight every 4 months for 4 years beginning at age 10. Also, in the sample, the proportion of female is equal to that of male. It's not difficult to calculate the iteration term when generate the variable for age and female.

3.2.3 Locate or think about reasonable values for the parameters in your model.

In this step, we use an external data set measuring Asian kid's data to estimate the coefficients for the above regression model, and we get β_0=5.35, β_1=3.59, β_2=-0.47, β_3=-0.24,τ_0=0.24,τ_1 -0.57and σ=1.17.

```
mixed weight c.age##i.girl || id: age, stddev

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -339.79909
Iteration 1:   log likelihood = -339.41232
Iteration 2:   log likelihood = -339.41033
Iteration 3:   log likelihood = -339.41032

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =        198
Group variable: id                              Number of groups  =         68

                                                Obs per group:
                                                              min =          1
                                                              avg =        2.9
                                                              max =          5

                                                Wald chi2(3)      =     680.51
Log likelihood = -339.41032                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   3.588488   .1892333    18.96   0.000     3.217598    3.959379
             |
        girl |
        girl |   -.468204   .2954284    -1.58   0.113    -1.047233     .110825
             |
  girl#c.age |
        girl |  -.2397793    .267607    -0.90   0.370    -.7642793    .2847208
             |
       _cons |   5.351429   .2076606    25.77   0.000     4.944421    5.758436
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Independent              |
                     sd(age) |   .5662947   .1219764      .3712777    .8637461
                   sd(_cons) |   .2384136   .4034962      .0086445    6.575403
-----------------------------+------------------------------------------------
               sd(Residual) |   1.165676   .0762599      1.025395    1.325148
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 20.38                 Prob > chi2 = 0.0000
```

```
Note: LR test is conservative and provided only for reference.
```

3.2.4 Simulate a single dataset assuming the alternative hypothesis, and fit the model.

Next, we create a simulated dataset based on our assumptions about the model under the alternative hypothesis. We will simulate 5 observations at 4-month increments for 200 children.

```
set seed 16
clear
set obs 200
generate child = _n
generate female = rbinomial(1,0.5)
generate u_0i = rnormal(0,0.25)
generate u_1i = rnormal(0,0.60)
expand 5
bysort child: generate age = (_n-1)*0.5
generate interaction = age*female
generate e_ij = rnormal(0,1.2)
generate weight = 5.35 + 3.6*age + (-0.5)*female + (-0.25)*interaction ///
    + u_0i + age*u_1i + e_ij
```

Our dataset includes the random deviations that we would not observe in a real dataset. We can then use mixed to fit a model to our simulated data.

```
mixed weight age i.female c.age#i.female || child: age , stddev nolog noheader
estimates store full
mixed weight age i.female || child: age , stddev nolog noheader
estimates store reduced
lrtest full reduced
```

We can then test the null hypothesis that the interaction term equals zero using a likelihood-ratio test.

```
 lrtest full reduced

Likelihood-ratio test                                 LR chi2(1)  =      8.23
(Assumption: reduced nested in full)                  Prob > chi2 =    0.0041
```

The p-value for our test is 0.0041, so we would reject the null hypothesis that the interaction term equals zero.

3.2.5 Write a program to create the datasets, fit the models, and use simulate to test the program.

Next, let's write a program that creates datasets under the alternative hypothesis, fits mixed models, tests the null hypothesis of interest, and uses simulate to run many iterations of the program.

```
capture program drop simmixed
program simmixed, rclass
    version 16
    // PARSE INPUT
    syntax, n1(integer)                 ///
            n(integer)                  ///
          [ alpha(real 0.05)            ///
            intercept(real 5.35)        ///
            age(real 3.6)               ///
            female(real -0.5)           ///
            interact(real -0.25)        ///
            u0i(real 0.25)              ///
            u1i(real 0.60)              ///
            eij(real 1.2) ]

    // COMPUTE POWER
    quietly {
        drop _all
        set obs `n'
        generate child = _n
        generate female = rbinomial(1,0.5)
        generate u_0i = rnormal(0,`u0i')
        generate u_1i = rnormal(0,`u1i')
        expand `n1'
        bysort child: generate age = (_n-1)*0.5
        generate interaction = age*female
        generate e_ij = rnormal(0,`eij')
        generate weight = `intercept' + `age'*age + `female'*female + ///
           `interact'*interaction  + u_0i + age*u_1i + e_ij

        mixed weight age i.female c.age#i.female || child: age, iter(200)
        local conv1 = e(converged)
        estimates store full
        mixed weight age i.female || child: age, iter(200)
        local conv2 = e(converged)
        estimates store reduced
        lrtest full reduced
        local reject = cond(`conv1' + `conv2'==2, (r(p)<`alpha'), .)
    }
    // RETURN RESULTS
    return scalar reject = `reject'
    return scalar conv = `conv1'+`conv2'
end
```

We then use simulate to run simmixed 10 times using the default parameter values for 5 observations on each of 200 children.

```
simulate reject=r(reject) converged=r(conv), reps(10) seed(12345):
             simmixed, n1(5) n(200)
```

```
        command:  simmixed, n1(5) n(200)
         reject:  r(reject)
      converged:  r(conv)

Simulations (10)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
```

simulate saved the results of the hypothesis tests to a variable named reject. The mean of reject is our estimate of the power to test the null hypothesis that the age×sex interaction term equals zero, assuming that the weight of 200 children is measured 5 times.

3.2.6 Write a program called power_cmd_mymethod, which allows you to run your simulations with power.

We could stop with our quick simulation if we were interested only in a specific set of assumptions. But it's easy to write an additional program named power_cmd_simmixed that will allow us to use Stata's power command to create tables and graphs for a range of sample sizes.

```
capture program drop power_cmd_simmixed
program power_cmd_simmixed, rclass
    version 16
    // PARSE INPUT
    syntax, n1(integer)               ///
            n(integer)                ///
          [ alpha(real 0.05)          ///
            intercept(real 5.35)      ///
            age(real 3.6)             ///
            female(real -0.5)         ///
            interact(real -0.25)      ///
            u0i(real 0.25)            ///
            u1i(real 0.60)            ///
            eij(real 1.2)             ///
            reps(integer 1000) ]

    // COMPUTE POWER
    quietly {
        simulate reject=r(reject), reps(`reps'):                                  ///
        simmixed, n1(`n1') n(`n') alpha(`alpha') intercept(`intercept')           ///
                  age(`age') female(`female') interact(`interact')                ///
                  u0i(`u0i') u1i(`u1i') eij(`eij')
        summarize reject
    }

    // RETURN RESULTS
    return scalar power = r(mean)
    return scalar n1 = `n1'
    return scalar N = `n'
    return scalar alpha = `alpha'
    return scalar intercept = `intercept'
    return scalar age = `age'
    return scalar female = `female'
```

```
        return scalar interact = `interact'
        return scalar u0i = `u0i'
        return scalar u1i = `u1i'
        return scalar eij = `eij'
end
```

3.2.7 Write a program called power_cmd_mymethod_init so that you can use numlists for all parameters.

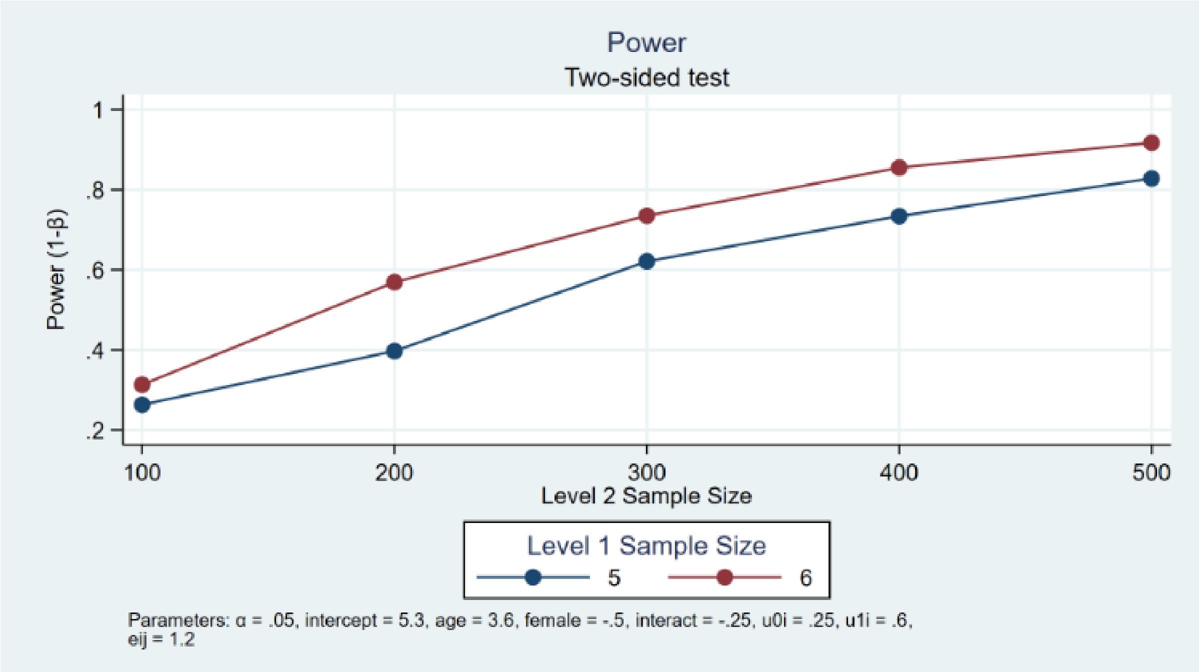
It's also easy to write a program named power_cmd_simmixed_init that will allow us to simulate power for a range of values for the parameters in our model.

```
capture program drop power_cmd_simmixed_init
program power_cmd_simmixed_init, sclass
      version 16
      sreturn clear
      // ADD COLUMNS TO THE OUTPUT TABLE
      sreturn local pss_colnames "n1 intercept age female interact u0i u1i eij"
      // ALLOW NUMLISTS FOR ALL PARAMETERS
      sreturn local pss_numopts  "n1 intercept age female interact u0i u1i eij"
end
```

Now, we can use power simmixed to simulate power for a variety of assumptions. The example below simulates power for a range of sample sizes at both levels 1 and 2. Level 2 sample sizes range from 100 to 500 children in increments of 100. At level 1, we consider 5 and 6 observations per child.

```
power simmixed, n1(5 6) n(100(100)500) reps(1000)
               table(n1 N power)
               graph(ydimension(power) xdimension(N) plotdimension(n1)
               xtitle(Level 2 Sample Size) legend(title(Level 1 Sample Size)))
xxxxxxxxxxxxxxxxxxxxxxxxxxx
Estimated power
Two-sided test

  +--------------------------+
  |      n1        N    power |
  |--------------------------|
  |       5      100   .2629 |
  |       6      100     .313 |
  |       5      200     .397 |
  |       6      200     .569 |
  |       5      300     .621 |
  |       6      300     .735 |
  |       5      400     .734 |
  |       6      400     .855 |
  |       5      500     .828 |
  |       6      500     .917 |
  +--------------------------+
```



4. Power Analysis in Python

4.1 Linear Model power analysis

```
import os
```

4.2 Multi-Variable Model power analysis

5. Power Analysis in R

5.1 Linear Model power analysis

5.2 Multi-Variable Model power analysis

```
```