AGRODEP
Household survey data course
Dakar, 8-10 October 2012

Sample calculation

IFPRI®

HarvestChoice
BETTER CHOICES, BETTER LIVES

THE WORLD BANK

LSMS
Living Standards Measurement Study

Sample Size

Maize yield (kg/hect)

- Mean: 802.6 kg/hect
- St. Dev. : 1027.79
- Deff: 1.28 (ρ=0.03)

Fertilizer Usage

- Mean: 0.553
- Deff: 3.63 (ρ=0.29)

Irrigation Usage

- Mean: 0.15
- Deff: 5.37 (ρ=0.49)

T values for two-tailed test

50%	80%	90%	95%	98%	99%
0.67449	1.281552	1.644854	1.95996	2.32635	2.57583

Sample Size

- What is the required sample size to detect a 10% change in maize yields?
- Note this is the same question as if you wanted to see a 10% difference between related questions in the same dataset (though the variance would probably be lower).

$$n = \left[\frac{4\sigma^2 (z_{1-\alpha/2} + z_{1-\beta})^2}{D^2} \right] \left[1 + \rho(m-1) \right]$$

```
sampsi 802.6 882.6, sd1(1027.79)

Estimated sample size for two-sample comparison of
  means

Test Ho: m1 = m2, where m1 is the mean in population
  1 and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =    802.6
            m2 =    882.6
           sd1 =  1027.79
           sd2 =  1027.79
        n2/n1 =     1.00

Estimated required sample sizes:

            n1 =     3469
            n2 =     3469
```

```
sampsi 802.6 882.6, sd1(1027.79) onesided

Estimated sample size for two-sample comparison of
   means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2

Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.9000
            m1 =    802.6
            m2 =    882.6
           sd1 =  1027.79
           sd2 =  1027.79
        n2/n1 =     1.00

Estimated required sample sizes:

            n1 =     2828
            n2 =     2828
```

```
sampsi 802.6 882.6, sd1(1027.79) power (0.8) onesided

Estimated sample size for two-sample comparison of
  means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.8000
            m1 =    802.6
            m2 =    882.6
           sd1 =  1027.79
           sd2 =  1027.79
        n2/n1 =     1.00

Estimated required sample sizes:

            n1 =     2041
            n2 =     2041
```

```
sampsi 802.6 882.6, ratio(2) sd1(1027.79)
Estimated sample size for two-sample comparison of
  means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.9000
            m1 =    802.6
            m2 =    882.6
           sd1 =  1027.79
           sd2 =  1027.79
        n2/n1 =     2.00

Estimated required sample sizes:

            n1 =     2602
            n2 =     5204
```

```
sampsi 802.6 1043.38, sd1(1027.79) onesided

Estimated sample size for two-sample comparison of
  means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.9000
            m1 =    802.6
            m2 =  1043.38
           sd1 =  1027.79
           sd2 =  1027.79
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      313
            n2 =      313
```

```
sampsi 802.6 882.6, sd1(1027.79) onesided
  method(change) pre(1) post(1) r01(0.7)

Estimated sample size for two samples with repeated measures
Assumptions:
                                alpha =   0.0500  (one-sided)
                                power =   0.9000
                                   m1 =    802.6
                                   m2 =    882.6
                                  sd1 =  1027.79
                                  sd2 =  1027.79
                                 n2/n1 =     1.00
              number of follow-up measurements =        1
               number of baseline measurements =        1
     correlation between baseline & follow-up =    0.700

Method: CHANGE                          Estimated required sample
 relative efficiency =     1.667       sizes:
    adjustment to sd =     0.775                    n1 =      1697
        adjusted sd1 =   796.123                    n2 =      1697
```

```
sampsi 802.6 907.6, sd1(1027.79) sd2(1456.85) onesided
  n1(1500) n2(1246)

Estimated power for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (one-sided)
            m1 =    802.6
            m2 =    907.6
           sd1 =  1027.79
           sd2 =  1456.85
sample size n1 =     1500
            n2 =     1246
         n2/n1 =     0.83

Estimated power:

         power =   0.6897
```

```
sampsi 802.6 907.6, sd1(1027.79) sd2(1456.85) onesided
  method(change) pre(1) post(1) r01(.7) n1(1500) n2(1246)

Estimated power for two samples with repeated measures
Assumptions:                alpha =   0.0500  (one-sided)
                            m1 =    802.6
                            m2 =    907.6
                            sd1 =  1027.79
                            sd2 =  1456.85
                            sample size n1 =     1500
                            n2 =     1246
                            n2/n1 =     0.83
           number of follow-up measurements =        1
            number of baseline measurements =        1
   correlation between baseline & follow-up =    0.700

  Method: CHANGE                                Estimated power:
   relative efficiency =    1.667                     power =    0.868
      adjustment to sd =    0.775
          adjusted sd1 =  796.123
          adjusted sd2 = 1128.471
```

Can't forget the Deff…

No design effects	With design effects
2,828	3,620
2,041	2,612
313	401
849	1,087

…where deff = 1.28

```
sampsi 802.6 1043.38, sd1(1027.79) onesided
            n1 =       313
            n2 =       313

sampclus, obsclus(10) rho(0.03)

Sample Size Adjusted for Cluster Design

   n1 (uncorrected) = 313
   n2 (uncorrected) = 313

   Intraclass correlation     = .03

   Average obs. per cluster   = 10
   Minimum number of clusters = 80


   Estimated sample size per group:

      n1 (corrected) = 398
      n2 (corrected) = 398
```

Exercise

- What is the sample size necessary to detect a 10 percent change in the percent of households using an improved water source?
- Improved water source is defined as piped water inside the dwelling, private outside standpipe/tap, public standpipe/tap, neighboring household, water vendor, water truck / tanker service.
- Hint: Calculate rho, then use *sampsi* and *sampclus* commands