Sample size calculations

Optional lecture

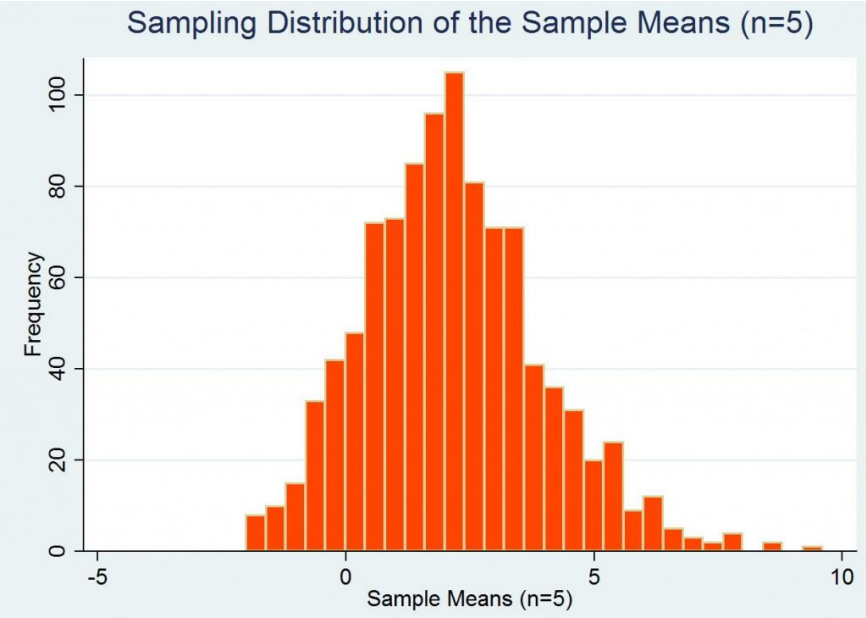
Law of big numbers

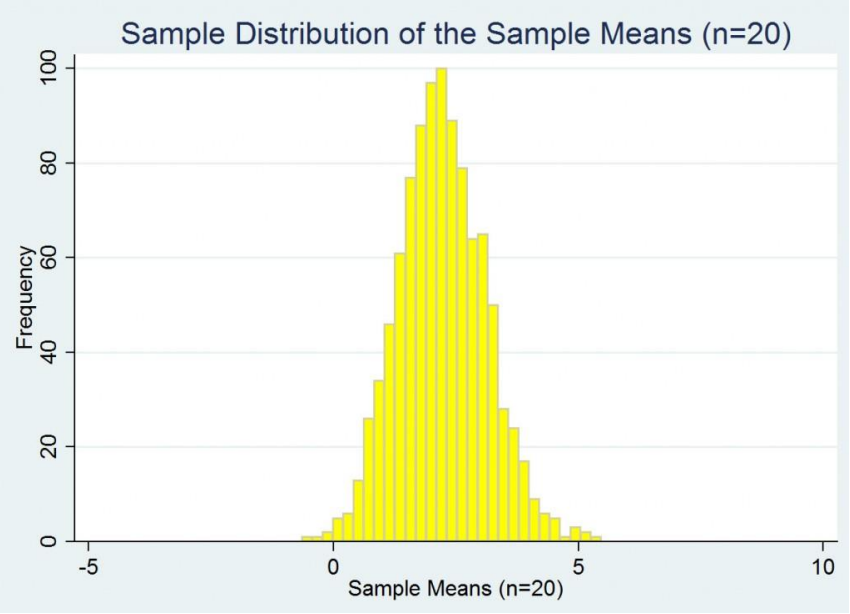
- Bigger and bigger numbers of observation/experiments = results closer and closer to the theoretical law

- Ex. we flip a coin 10 time, we got 4 tails
- flip the coin 100 time, we got 52 tails
- flip the coin 1000 time, we got 501 tails

Law of big numbers

- bigger and bigger the sample is (more and more people taken into the study) = closer and closer to the frequency of the disease in the whole population





Albright EA. Sampling and the Central Limit Theorem. Nicholas School of the Environment, Duke University

- small sample = speed, low cost, low measurement errors, etc.
- big sample = closer to the true mean/frequency

- What is the minimum number of individuals needed to closely approximate the population's mean frequency?

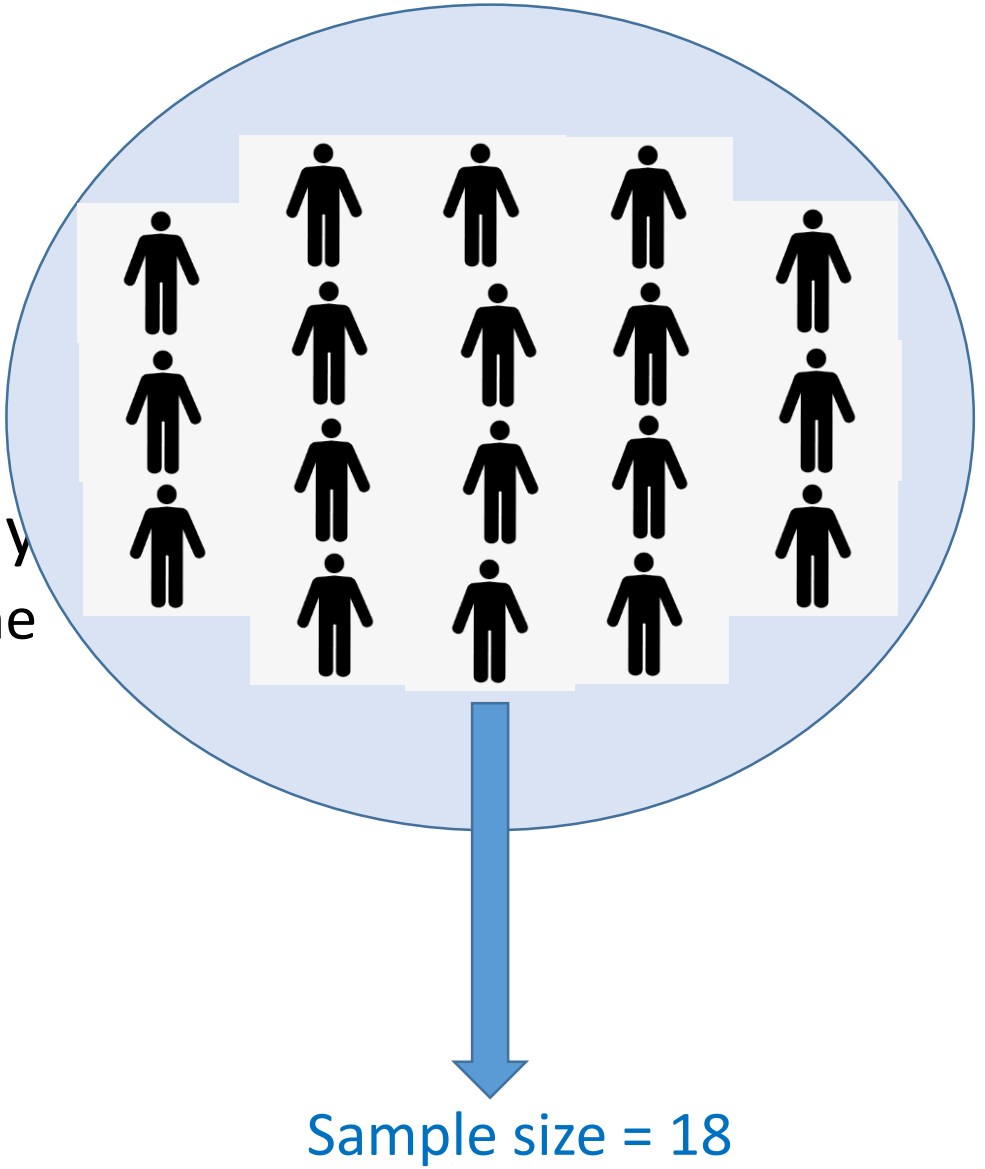
this kind of sample size can be calculated

different formulas for different goals

! a statistician is needed

Sample size

- number of subjects
  - also called volume
- it is calculated before the start of the study
  - different statistical formulas depending on the objective of the study



- How many individuals we should take into the study to get a good approximation of the frequency of the disease in the population?
- How many individuals we should take into the study to get a good approximation of the mean of... in the population?

can be computed..., you need a statistician for this task (different situations, different formulas)

Compute the sample size

- error level α, k length of the interval = precision:

[f - k , f + k]

- starting from this formula

$$\left[f - Z_\alpha \cdot \sqrt{\frac{f(1-f)}{n}},\ f + Z_\alpha \cdot \sqrt{\frac{f(1-f)}{n}} \right]$$

$$k = Z_\alpha \cdot \sqrt{\frac{f(1-f)}{n}}$$

$$n = \frac{Z_\alpha^2 \cdot f(1-f)}{k^2}$$

EXAMPLE

- Objective: to evaluate the prevalence of diabetes type 2 in the population over 50 years old.
- we know from other studies that is higher than 30%.
- precision k=5% (the length of 10%) error level $\alpha = 0.05$ ($Z_{\alpha} = 1.96$)
- Sample size

$$\frac{(1.96)^2 \times 0.3 \times 0.7}{(0.05)^2}$$

- we found n = 323.

EXEMPLE

- Objective: to evaluate the prevalence of diabetes type 2 in the population over 50 years old.
- we know from other studies that is higher than 30%.
- precision k=1% (the length of 2%) error level $\alpha = 0.05$ ($Z_\alpha = 1.96$)
- Sample size

$$\frac{(1.96)^2 \times 0.3 \times 0.7}{(0.01)^2}$$

- we found n = 8067.

EXEMPLE

- Objective: to evaluate the prevalence of diabetes type 2 in the population over 50 years old.
- we know from other studies that is higher than 50%.
- precision k=5% (the length of 10%) error level $\alpha = 0.05$ ($Z_{\alpha} = 1.96$)
- Sample size

$$\frac{(1.96)^2 \times 0.5 \times 0.5}{(0.05)^2}$$

- we found n = 384> 323 (f=0.3).

Sample size calculation exemple

Error 5%

we want to compare two frequencies http://statpages.org/proppowr.html

we need two samples

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.30	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

the expected difference between the frequencies =20%

Compute

equal sample sizes

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	93	93	186
With Continuity Correction:	103	103	206

Sample size calculation for two frequencies

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

difference = 10%

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	387	387	775
With Continuity Correction:	407	407	814

Sample size calculation for two frequencies

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.47	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

difference = 3%

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	4355	4355	8711
With Continuity Correction:	4422	4422	8844

Sample size calculation for two frequencies

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.60	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Power = 80%

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	97	97	194
With Continuity Correction:	107	107	213

Sample size calculation for two frequencies

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	85	(Usually 80)
Group 1 Population Proportion:	.40	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.60	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

Power = 85%

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	111	111	221
With Continuity Correction:	120	120	241

Sample size calculation for two arithmetic mean in the case of normal distribution

http://sampsize.sourceforge.net/iface/s2.html#nm

expected difference between the arithmetic mean = 230-210=20

```
Assumptions:

     alpha =          5  (two-sided)
     power =         90
        m1 =        230
        m2 =        210
       sd1 =         26
       sd2 =         33
     n2/n1 =          1

Estimated sample size:

        n1 =         47
        n2 =         47
```

Sample size calculation for two arithmetic mean in the case of normal distribution

difference = 230-220=10

```
Assumptions:

      alpha =          5  (two-sided)
      power =         90
         m1 =        230
         m2 =        220
        sd1 =         26
        sd2 =         33
      n2/n1 =          1

Estimated sample size:

         n1 =        186
         n2 =        186
```

Sample size calculation for two arithmetic mean in the case of normal distribution

difference = 230-225=5

```
Assumptions:

      alpha =          5 (two-sided)
      power =         90
         m1 =        230
         m2 =        225
        sd1 =         26
        sd2 =         33
      n2/n1 =          1

Estimated sample size:

         n1 =        742
         n2 =        742
```