

Biostatistics 140.623

Third Term, 2017-2018

Problem Set 1 (with R)

Vitamin A Supplementation to Prevent Children's Mortality in Nepal

Study Design - Sample Size Estimation

Learning Objectives:

Students who successfully complete this project section will be able to:

- Calculate the sample size necessary for estimating mortality in children < 3 years of age with a desired level of precision.
- Estimate the sample size for a new study of vitamin A and mortality in children < 3 years of age.

Data Set:

The Nepal data set is located in the `.csv` data file named `nepal621.csv`.

Methods:

- 1) Suppose you are interested in choosing an appropriate sample size for estimating the 16-month mortality rate for children younger than 3 years of age in a developing country in which vitamin A supplementation is not currently available.
 - a. Use the available information from the Nepal data set to choose a sample size so that you estimate this rate to within $\pm 0.5\%$.
 - b. Now, suppose **no information** is available from this Nepal study. Determine what sample size would be required for each of a range of plausible values of the mortality rate. Summarize your sample size findings in a **table**.
- 2) Now suppose you have a chance to investigate the effect of vitamin A supplementation on the mortality of children under 3 years of age. The `power.prop.test()` command in R can be used with the results of the Nepal trial to choose the size of the vitamin A and control groups (assuming equal sample sizes for both groups) for the new study. Confirm from the data set that the 16-month mortality in the placebo group is 0.0294 and the 16-month mortality in the Vitamin A group is 0.0245 for the Nepal study. The estimated relative risk of death in the placebo group as compared to the Vitamin A group is $0.029/0.0245 = 1.2$. Assuming a significance level of 0.05 and power of 80%, the sample size needed in the new study to detect a relative risk of 1.2 is **17,144 children per group** according to the results on the next page. A total sample size of 34,288 children would be required.

```
> power.prop.test(n=NULL, p1=0.0294, p2=0.0245, sig.level=0.05, power=0.8,
                  alternative="two.sided")
```

Two-sample comparison of proportions power calculation

```
      n = 17143.9
      p1 = 0.0294
      p2 = 0.0245
sig.level = 0.05
  power = 0.8
alternative = two.sided
```

NOTE: n is number in *each* group

- 3) Verify R's calculations for part 2) by hand using the method learned in class. Expect your answer to be close in value to, but not exactly the same as, that provided by R, due to rounding in hand calculations. (Stata uses a continuity correction that R doesn't, so the value calculated from Stata will also be different than the one you calculated by hand and in R.)

- 4) Construct a **table** that displays the total sample sizes required under various assumptions about the mortality rate in the control group and the relative risk of interest. Assume a significance level of 0.05 and 80% power. Comment on what you observe.

Vary the assumptions by:

- a. Assuming that the control group mortality rate (risk) is:
 1. the same as that observed in Nepal placebo group of children < 3 years of age
 2. or .5% lower
 3. or .5% higher
 - b. Assuming that the relative risk of death for children in the control group as compared to children receiving vitamin A is hypothesized to be:
 1. 1.2 (the same as the relative risk that was estimated for Nepali children in this age group)
 2. or 1.5
 3. or 1.75.
- 5) Construct another **table** that displays the total sample sizes required under the same varying assumptions of the mortality rate in the control group and the relative risk of interest. This time, assume a significance level of 0.05 and 90% power. Comment on what you observe.
- 6) Select a design based upon your findings from parts 4 and 5 above. Write a brief paragraph that presents and justifies your choice. Be numerate.