

Usage:

Power Calculation for Within-Center Randomization Designs with Correlated Continuous Data

Yi LU 12/27/2011

Statement

The statistical program was originally developed by ??¹, edited by Yi Lu¹ under supervision of Drs. Mei-Cheng Wang¹ and Larry Wissow².

Sampling Design Background

The simulation and the power calculation are based on the sampling design below:

1. Suppose there are “N” clusters (e.g. hospitals/health centers), and within each cluster there are “2m” patients. In each cluster, “m” patients are randomized into the treatment group and the other “m” patients are in the control group.
2. Assume the “N” clusters are independent of each other. **Within** each cluster:
 - (a) Suppose for each patient there’s a continuous output which is normally distributed, with variance σ^2 , mean μ_0 for the control group, and μ_1 for the treatment group. We are interested in the “treatment effect”: $\Delta = \mu_1 - \mu_0$.
 - (b) The 2m patients (no matter group they are in) are correlated, and the correlation (coefficient) between any two patients is equal to “ ρ ” ($0 \leq \rho < 1$).

Subject to treatment-effect-size $\Delta = \mu_1 - \mu_0$, the standard deviation σ , the type I error “ α ” and correlation “ ρ ”, the power is to be calculated for the design above:

$$\text{Power}(N, m, \sigma, \Delta, \rho, \alpha) = \Pr(\hat{\Delta} \in R | \Delta),$$

where $\Delta \neq 0$, $\hat{\Delta}$ is the difference of the mean value of the $N \times m$ outputs from the patients of treatment groups and the mean value of the other $N \times m$ outputs from those in control groups, R is the rejection area determined by $\alpha = \Pr(\hat{\Delta} \in R | \Delta = 0)$.

[Remark]: The power calculation in R involves complicated simulations with the sample size as a necessary input parameter. Therefore, to estimate minimum sample size given a

¹ Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

² Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health

fixed power, it's suggested that the program be applied repeatedly with different combination of parameter settings (e.g. different sample sizes) so that the minimum sample size required could be approximated.

The .R Function: “power_normal”.

Based on the sampling design above, we have the following parameters to set:

N	Number of clusters.
m	Number of subjects (patients) in each cluster (within each group: treatment/control)
sigma	The standard deviation.
delta	The mean difference / treatment effect (treatment - control).
rho	The correlation between subjects (patients) in the same cluster.
alpha	Type I error (default value = 0.05).
Rep	Number of repeated sampling in simulation (default value = 2000).

We have the requirements for the parameter setting above:

$$0 \leq \rho < 1$$

Note that the above requirements are necessary due to the correlation structure of the normal data model. If the requirements are not satisfied, R will report errors without power calculations conducted.

We may follow the example below to illustrate the usage of the R function “power_normal”.

[Example 1]: Suppose we have the “power_normal.R” file downloaded at the directory: C:/Users/Documents. (Note that this directory can be changed up to your choice. If you do so, please make corresponding change in the “source” command below.)

Step 1: Download and install R.

Step 2: Open R. In the “R Console” window, type the following command:

```
source("C:/Users/Documents/power_normal.R")
```

Note that the directory part above is subject to change if you save the “power_normal.R” file at other directory. Each time you open R software, you may have to type the “source” command above to load the “power_normal.R” function in your R working space.

Step 3: Use the power_normal.R function to calculate power by setting the parameters in the following command (the 1st line below):

```
> power_normal(N=10,m=25,sigma=3,delta=0,rho=0.04)
```

The power is:

```
[1] 0.0533
```

The 2-3rd lines are the output from R. Note that “alpha” and “Rep” have default values as 0.05 and 2000, respectively, which can be omitted in the command. If needed, they can be set as:

```
power_normal(N=10,m=25,sigma=3,delta=1.5,rho=0.04,alpha=0.1)
```

```
power_normal(N=10,m=25,sigma=3,delta=1.5,rho=0.04,Rep=5000)
```

```
power_normal(N=10,m=25,sigma=3,delta=1.5,rho=0.04,alpha=0.01,Rep=500)
```

[Remark]: As noted by the footnote 3, the power calculation is an "estimate" (not the exact theoretical number) based on the simulation whose preciseness could be adjusted by the parameter "Rep" in the function. Generally speaking, by setting "Rep" large could make the estimate more precise and stable (i.e. converge to the true theoretical value). Therefore, it's suggested that the user start with a relatively small number of replications (e.g. the default number Rep = 1000) to see how fast it runs, and then increase the number of replications for stable results.

Moreover, the first 5 parameters are required to be set for the function to run. If you set them as the standard order of (N,m,sigma,delta,rho), you can omit these “symbols”. Otherwise, if you change this order, you have to keep these “symbols”. For example, the following commands are equivalent:

```
power_binary(N=10,m=25,sigma=3,delta=1.5,rho=0.4)
```

```
power_binary(10,25,3,1.5,0.4)
```

```
power_binary(rho=0.4,sigma=3,m=25,N=10,delta=1.5)
```

Finally, if your parameter setting misfit the parameter requirements mentioned above, you may obtain the error message like:

```
> power_normal(N=10,m=25,sigma=3,delta=1.5,rho=-0.1)
```

```
[1] "Errors in parameter setting! 'rho' should be between 0 and 1."
```

The power is:

```
[1] NaN
```

³ The calculation of power is through replications of simulation (with the replication number set by the parameter “Rep”) and therefore the calculated results **might (should) be different** each time the programs run even with the same parameter settings, although the difference is expected to be small when the number of replications (“Rep”) is large.