# POL SCI 231b: Power Review

GSI Natalia Garbiras-Díaz

University of California, Berkeley

Spring 2017

Before starting, here are some useful references:

- Gerber and Green (2012, p.93)
- EGAP 10 things you need to know about statistical power
- EGAP Power analysis for the standard design

## 1 What is power?

The *statistical power* of a test is the probability that it will reject the null hypothesis, given that the null hypothesis is false.

Calculating the statistical power of an experiment (or a test, in general) involves some guesswork. We need to make assumptions about the distributions of potential outcomes, and about the expected treatment effect. We often assume that our potential outcomes come from a normal distribution (or rely on the CLT).

## 2 Analytic formula

Gerber and Green (2012, p.93) provide a simple asymptotic approximation for the power of an experiment (where $N/2$ units are assigned to treatment) for a two-tailed hypothesis test:

$$\beta = \Phi\Big(\frac{\mid \mu_T - \mu_C \mid \sqrt{N}}{2\sigma} - \Phi^1(1 - \frac{\alpha}{2})\Big)$$

where:

- $\beta$ is the statistical power of our experiment
- $\Phi(\cdot)$ is the normal cumulative distribution function (CDF)
- $\mu_T - \mu_C$ is our expected/hypothesized $\tau$ (i.e., treatment effect)
- $N$ is the number of units in our experiment
- $\sigma$ is the expected noise in our experiment (i.e., the standard deviation of outcomes)
- $\alpha$ is our significance level. We usually set this to 0.05

# 3 Which parameters can you vary?

We can make power calculations varying some of the parameters of the experiment:

- N
- Noise ($\sigma$)
- Effect size ($\mid \mu_T - \mu_C \mid$)

# 4 Beyond the analytic formula: power calculations with simulations

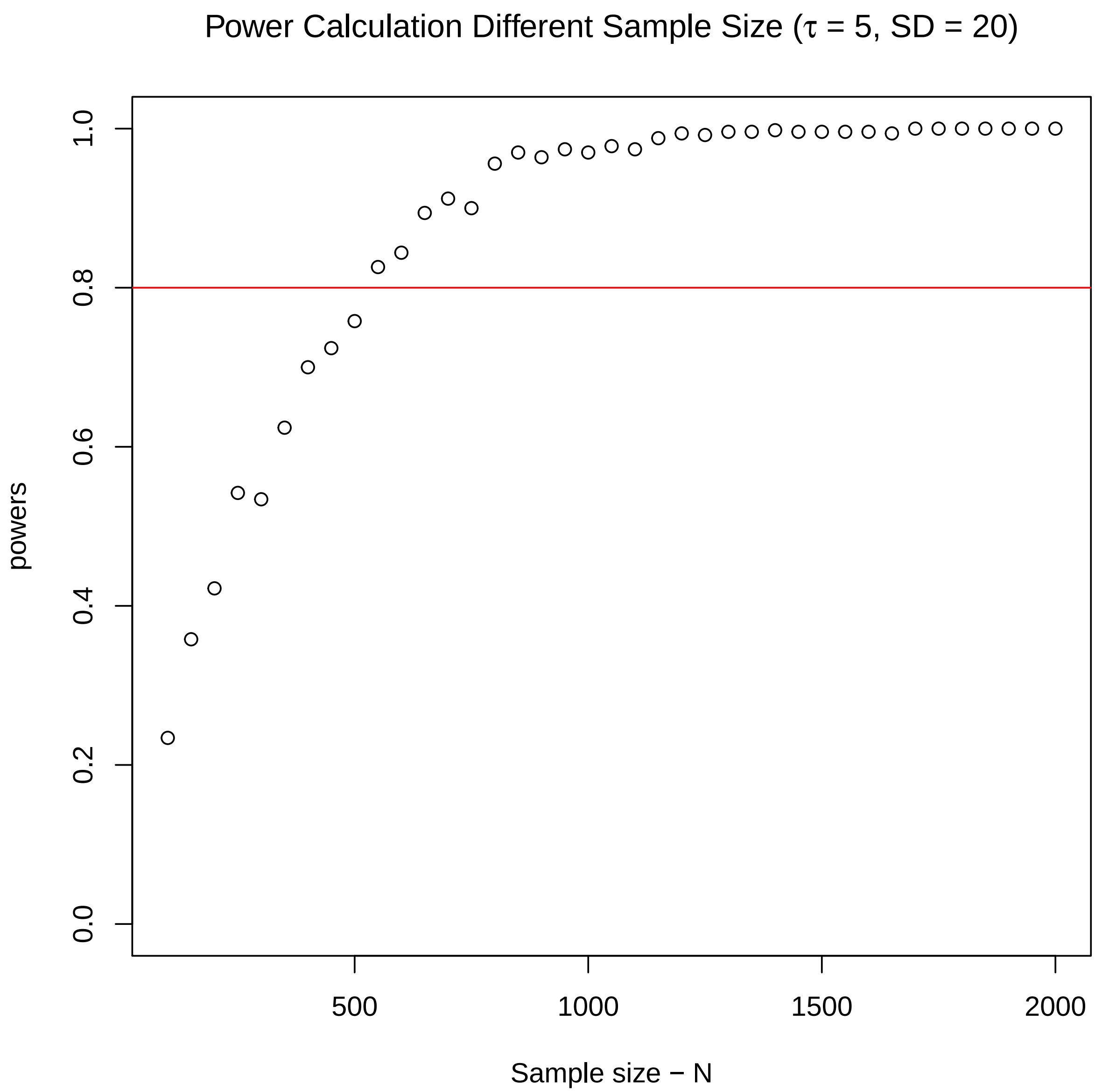
Imagine that we were able to conduct our experiment thousands of times. In this context, power is a measure of how often, given assumptions, we would obtain statistically significant results. So, instead of relying on the analytic formula, we can calculate the power of our experiment by simulating many many experiments on R.

## 4.1 An example of power calculations for different sample size

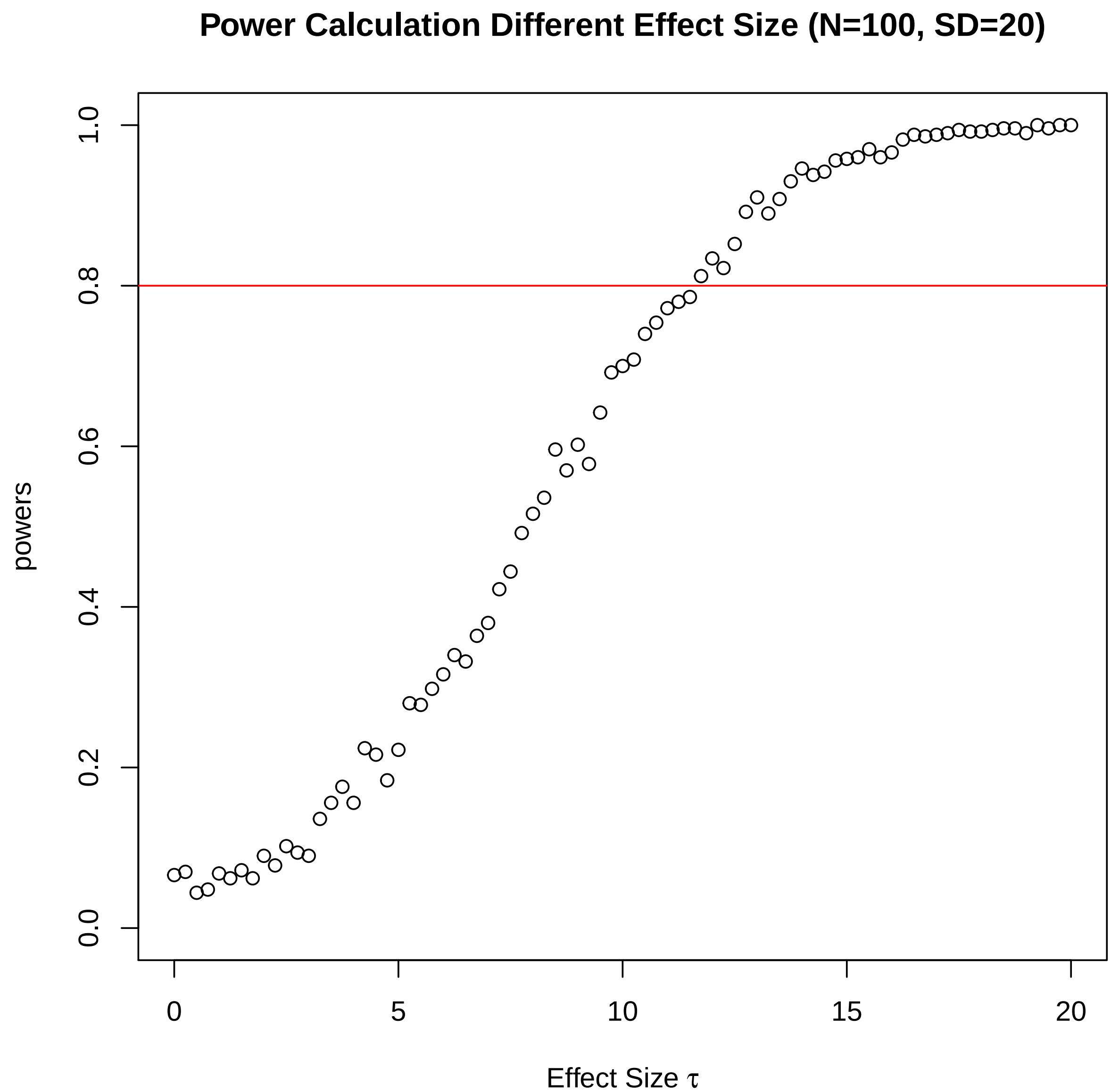
```
possible.ns <- seq(from = 100, to = 2000, by = 50)
powers <- rep(NA, length(possible.ns))
for (j in 1:length(possible.ns)) {
    N <- possible.ns[j]
    significant.experiments <- rep(NA, 500)
    for (i in 1:500) {
        Y0 <- rnorm(n = N, mean = 60, sd = 20)
        tau <- 5
        Y1 <- Y0 + tau
        Z.sim <- rbinom(n = N, size = 1, prob = 0.5)
```

```
        Y.sim <- Y1 * Z.sim + Y0 * (1 - Z.sim)
        fit.sim <- lm(Y.sim ~ Z.sim)
        p.value <- summary(fit.sim)$coefficients[2,
            4]
        significant.experiments[i] <- (p.value <=
            0.05)
    }
    powers[j] <- mean(significant.experiments)
}
```

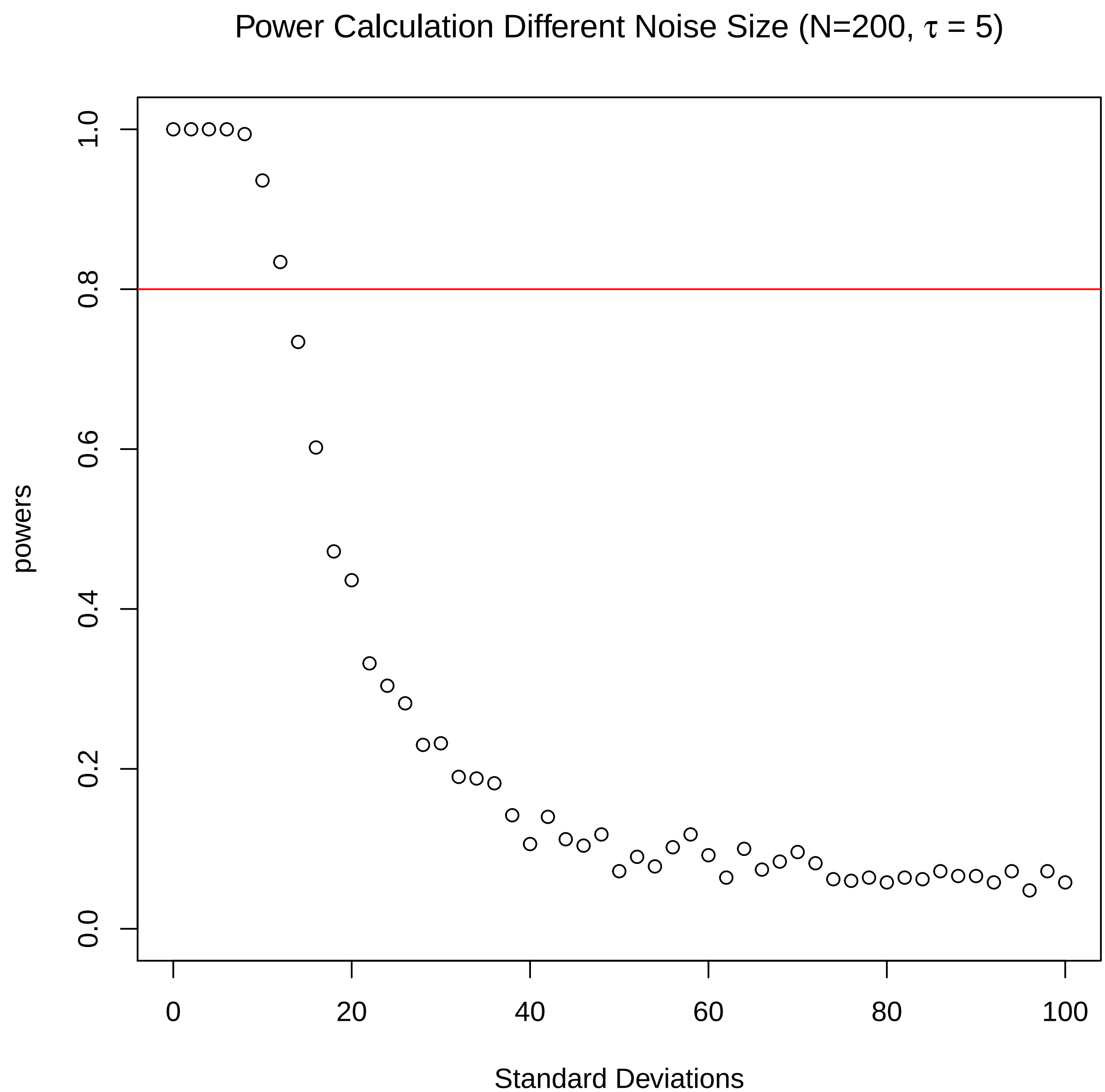
Let's see how this looks:



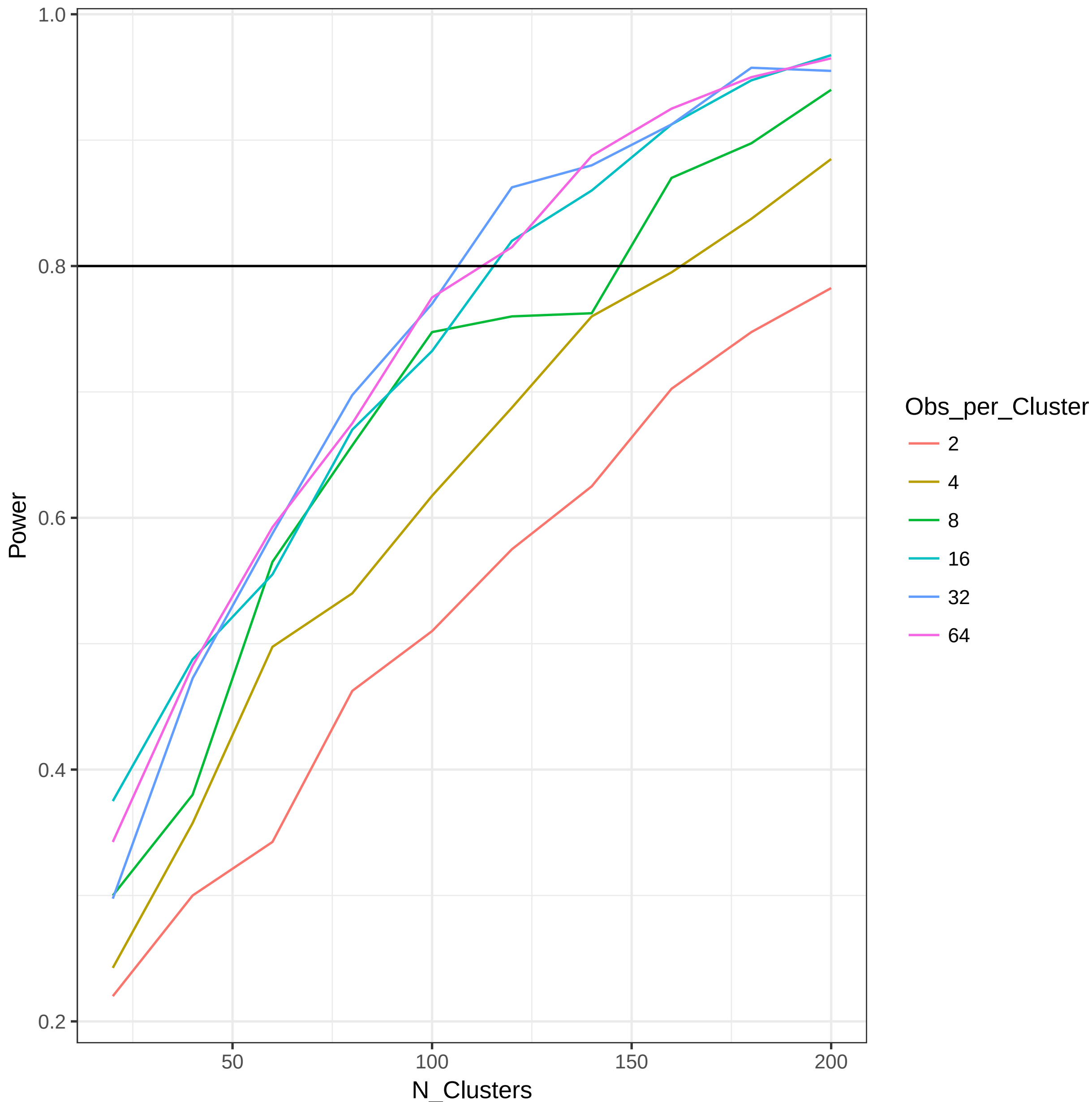

Let’s see what happens with different **effect sizes**:




Let's see what happens with **different noise**:




## 4.2 Power Analysis for clustered randomized experiments



1.0
0.8
0.6
0.4
0.2
Power
50
100
150
200
N_Clusters
Obs_per_Cluster
2
4
8
16
32
64