# STA305/1004 - Class 7

September 26, 2019

# Today's Class

- Sample size formulae two-samples means/proportions
- Calculating power via simulation

# Problem

You are working as a statistician on the design of a two-arm clinical trial. The trial will compare means, $\mu_1, \mu_2$ of a continuous outcome (e.g., mean tumour size).

$$H_0 : \theta = 0 \text{ vs. } H_a : \theta \neq 0,$$

$\theta = \mu_1 - \mu_2$.

Develop a formula for estimating sample size.

# The problem

The New York Times

## *A Cancer Conundrum: Too Many Drug Trials, Too Few Patients*

Dr. Wassim Abida, a medical oncologist at Memorial Sloan Kettering Cancer Center, examined Bruce Fenstermacher, a patient taking part in a clinical trial.

George Etheredge for The New York Times

By Gina Kolata

Aug. 12, 2017

83

With the arrival of two revolutionary treatment strategies, immunotherapy and personalized medicine, cancer researchers have found new hope — and a problem that is perhaps unprecedented in medical research.

There are too many experimental cancer drugs in too many clinical

# Sample size - known variance and equal allocation

Allocation refers to: a clinical trial design strategy used to assign participants to an arm of a study.

If the variance is known then the test statistic is

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sigma\sqrt{(1/n_1 + 1/n_2)}} \sim N(0,1).$$

This is the test statistic of the two-sample z-test.

The power at $\theta = \theta_1$ is given by

$$1-\beta = P\left(Z \geq z_{\alpha/2} - \frac{\theta_1}{\sigma\sqrt{1/n_1 + 1/n_2}}\right) + P\left(Z < -z_{\alpha/2} - \frac{\theta_1}{\sigma\sqrt{1/n_1 + 1/n_2}}\right).$$

Ignoring terms smaller than $\alpha/2$ and combining positive and negative $\theta$

$$\beta \approx \Phi\left(z_{\alpha/2} - \frac{|\theta_1|}{\sigma\sqrt{1/n_1 + 1/n_2}}\right).$$

# Sample size - known variance and equal allocation

The sample size is obtained by solving

$$z_\beta + z_{\alpha/2} = \left( \frac{|\theta_1|}{\sigma\sqrt{1/n_1 + 1/n_2}} \right).$$

If we assume that there will be an equal allocation of subjects to each group then $n_1 = n_2 = n/2$, the total sample size for the phase III trial is

$$n = \frac{4\sigma^2 \left( z_\beta + z_{\alpha/2} \right)^2}{\theta^2}.$$

# Sample size - known variance and unequal allocation

- In many trials it is desirable to put more patients into the experimental group to learn more about this treatment.
- If the patient allocation between the two groups is $r = n_1/n_2$ then $n_1 = r \cdot n_2$ then

$$n_2 = \frac{(1 + 1/r)\sigma^2 \left(z_\beta + z_{\alpha/2}\right)^2}{\theta^2}.$$

An R function to compute the sample size in groups 1 and 2 for unequal allocation is

```
size2z.uneq.test <- function(theta,alpha,beta,sigma,r)
{ zalpha <- qnorm(1-alpha/2)
  zbeta <- qnorm(1-beta)
  n2 <- (1+1/r)*(sigma*(zalpha+zbeta)/theta)^2
  n1 <- r*n2
  return(c(n1,n2))}
```

# Sample size - known variance and unequal allocation

What is the sample size required for 90% power to detect $\theta = 1$ with $\sigma = 2$ at the 5% level in a trial where two patients will be enrolled in the experimental arm for every patient enrolled in the control arm?

```
# sample size for theta =1, alpha=0.05,
# beta=0.1, sigma=2, r=2
# group 1 sample size (experimental group)
size2z.uneq.test(theta = 1,alpha = .05,beta = .1,sigma = 2,r = 2)[1]
```

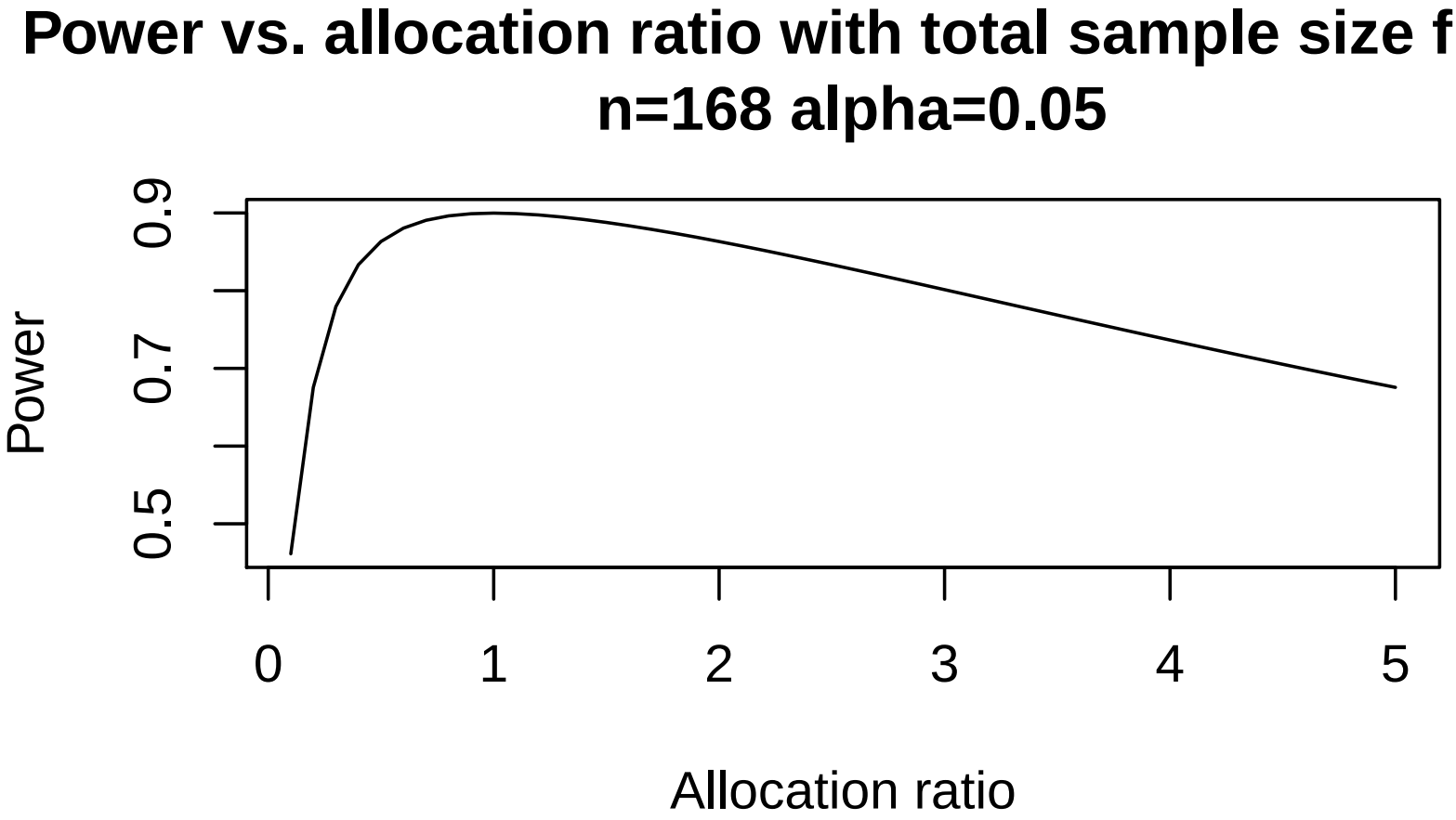
```
[1] 126.0891
```

```
# group 2 sample size (control group)
size2z.uneq.test(theta = 1,alpha = .05,beta = .1,sigma = 2,r = 2)[2]
```

```
[1] 63.04454
```

## Sample size - known variance and unequal allocation

The power of the two-sample z-test can be studied as a function of the allocation ratio $r$.




The plot shows that imbalance typically leads to loss of power.

# Comparing Proportions for Binary Outcomes

- In many clinical trials, the primary endpoint is dichotomous, for example, whether a patient has responded to the treatment, or whether a patient has experienced toxicity.
- Consider a two-arm randomized trial with binary outcomes. Let $p_1$ denote the response rate of the experimental drug, $p_2$ as that of the standard drug, and the difference is $\theta = p_1 — p_2$.
- In case:

$$H_0 : \theta = 0 \text{ vs. } H_a : \theta \neq 0,$$

where $\theta = p_1 - p_2$.

# Comparing Proportions for Binary Outcomes

Let $Y_{ik}$ be the binary outcome for subject $i$ in arm $k$; that is,

$$Y_{ik} = \begin{cases} 1 & \text{with probability } p_k \\ 0 & \text{with probability } 1 - p_k, \end{cases}$$

for $i = 1, ..., n_k$ and $k = 1, 2$. The sum of indendent and identically distributed Bernoulli random varaibles has a binomial distribution,

$$\sum_{i=1}^{n_k} Y_{ik} \sim Bin(n_k, p_k),\ k = 1, 2.$$

(Yin, pg. 173-174)

# Comparing Proportions for Binary Outcomes

The sample proportion for group $k$ is

$$\hat{p}_k = \bar{Y}_k = (1/n_k) \sum_{i=1}^{n_k} Y_{ik},\ k = 1, 2,$$

and $E\left(\bar{Y}_k\right) = p_k$ and $Var\left(\bar{Y}_k\right) = \frac{p_k(1-p_k)}{n_k}$.

The goal of the clinical trial is to determine if there is a difference between the two groups using a binary endpoint. That is we want to test $H_0 : \theta = 0$ versus $H_0 : \theta \neq 0$.

The test statistic (assuming that $H_0$ is true) is:

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \sim N(0, 1),$$

# Comparing Proportions for Binary Outcomes

The test rejects at level $\alpha$ if and only if

$$|T| \geq z_{\alpha/2}.$$

Using the same argument as the case with continuous endpoints and ignoring terms smaller than $\alpha/2$ we can solve for $\beta$

$$\beta \approx \Phi\left(z_{\alpha/2} - \frac{|\theta_1|}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}\right).$$

# Comparing Proportions for Binary Outcomes

Using this formula to solve for sample size. If $n_1 = r \cdot n_2$ then

$$n_2 = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\theta^2} \left(p_1(1-p_1)/r + p_2(1-p_2)\right).$$

# Comparing Proportions for Binary Outcomes

- The built-in R function `power.prop.test()` can be used to calculate sample size or power.
- For example suppose that the standard treatment for a disease has a response rate of 20%, and an experimental treatment is anticipated to have a response rate of 28%.
- The researchers want both arms to have an equal number of subjects. How many patients should be enrolled if the study will conduct a two-sided test at the 5% level with 80% power?

```
power.prop.test(p1 = 0.2, p2 = 0.28, power = 0.8)
```

```
     Two-sample comparison of proportions power calculation

              n = 446.2054
             p1 = 0.2
             p2 = 0.28
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

# Calculating Power by Simulation

- If the test statistic and distribution of the test statistic are known then the power of the test can be calculated via simulation.
- Consider a two-sample t-test with 30 subjects per group and the standard deviation of the clinical outcome is known to be 1.
- What is the power of the test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_0 : \mu_1 - \mu_2 = 0.5$, at the 5% significance level?
- The power is the proportion of times that the test correctly rejects the null hypothesis in repeated sampling.

# Calculating Power by Simulation

We can simulate a single study using the `rnorm()` command. Let's assume that $n_1 = n_2 = 30, \mu_1 = 3.5, \mu_2 = 3, \sigma = 1, \alpha = 0.05$.

```
set.seed(2301)
t.test(rnorm(30,mean=3.5,sd=1),rnorm(30,mean=3,sd=1),var.equal = T)
```

```
    Two Sample t-test

data:  rnorm(30, mean = 3.5, sd = 1) and rnorm(30, mean = 3, sd = 1)
t = 2.1462, df = 58, p-value = 0.03605
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.03458122 0.99248595
sample estimates:
mean of x mean of y
 3.339362  2.825828
```

Should you reject $H_0$?

# Calculating Power by Simulation

- Suppose that 10 studies are simulated.
- What proportion of these 10 studies will reject the null hypothesis at the 5% level?
- To investigate how many times the two-sample t-test will reject at the 5% level the `replicate()` command will be used to generate 10 studies and calculate the p-value in each study.
- It will still be assumed that
  $n_1 = n_2 = 30, \mu_1 = 3.5, \mu_2 = 3, \sigma = 1, \alpha = 0.05.$

```
set.seed(2301)
pvals <- replicate(10,t.test(rnorm(30,mean=3.5,sd=1),
                             rnorm(30,mean=3,sd=1),
                             var.equal = T)$p.value)
pvals # print out 10 p-values
```

```
 [1] 0.03604893 0.15477655 0.01777959 0.40851999 0.34580930 0.11131007
 [7] 0.14788381 0.00317709 0.09452230 0.39173723
```

```
#power is the number of times the test rejects at the 5% level
sum(pvals<=0.05)/10
```

```
[1] 0.3
```

# Calculating Power by Simulation

But, since we only simulated 10 studies the estimate of power will have a large standard error. So let's try simulating 10,000 studies so that we can obtain a more precise estimate of power.

```
set.seed(2301)
pvals <- replicate(10000,t.test(rnorm(30,mean=3.5,sd=1),
                                rnorm(30,mean=3,sd=1),
                                var.equal = T)$p.value)
sum(pvals<=0.05)/10000
```

```
[1] 0.4881
```

# Calculating Power by Simulation

This is much closer to the theoretical power obtained from `power.t.test()`.

```
power.t.test(n = 30,delta = 0.5,sd = 1,sig.level = 0.05)
```

```
     Two-sample t test power calculation

              n = 30
          delta = 0.5
             sd = 1
      sig.level = 0.05
          power = 0.477841
    alternative = two.sided

NOTE: n is number in *each* group
```

# Calculating Power by Simulation

- The built-in R functions `power.t.test()` and `power.prop.test()` don't have an option for calculating power where the there is unequal allocation of subjects between groups.
- These built-in functions don't have an option to investigate power if other assumptions don't hold (e.g., normality).
- One option is to simulate power for the scenarios that are of interest. Another option is to write your own function using the formula derived above.

# Calculating Power by Simulation

- Suppose the standard treatment for a disease has a response rate of 20%, and an experimental treatment is anticipated to have a response rate of 28%.
- The researchers want both arms to have an equal number of subjects.
- A power calculation above revealed that the study will require $446 \times 2 = 892$ patients for 80% power.
- What would happen to the power if the researchers put more patients in the experimental arm compared to the control arm?

# Calculating Power by Simulation

- The number of subjects in the experimental arm that have a positive response to treatment will be an observation from a $Bin(1500, 0.28)$.
- The number of subjects that have a positive response to the standard treatment will be an observation from a $Bin(500, 0.2)$.
- We can obtain simulated responses from these distributions using the `rbinom()` command in R.

```
set.seed(2301)
rbinom(1,1500,0.28)
```

```
[1] 403
```

```
rbinom(1,500,0.20)
```

```
[1] 89
```

# Calculating Power by Simulation

- The p-value for this simulated study can be obtained using `prop.test()`.

```
set.seed(2301)
prop.test(x=c(rbinom(1,1500,0.28),rbinom(1,500,0.20)),
          n=c(1500,500),correct = F)
```

```
    2-sample test for equality of proportions without continuity
    correction

data:  c(rbinom(1, 1500, 0.28), rbinom(1, 500, 0.2)) out of c(1500, 500)
X-squared = 16.62, df = 1, p-value = 4.568e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 0.05032654 0.13100680
sample estimates:
   prop 1    prop 2
0.2686667 0.1780000
```

# Calculating Power by Simulation

Assume the standard treatment for a disease has a response rate of 20%, and an experimental treatment is anticipated to have a response rate of 28%.

```
set.seed(2301)
n1 <- 300
n2 <- 100
pvals <- replicate(10000,
prop.test(x=c(rbinom(n = 1,size = n1,prob = 0.28),
              rbinom(n = 1,size = n2,prob = 0.20)),
          n=c(n1, n2),correct = F)$p.value)
sum(pvals <= 0.01)
```

```
## [1] 1434
```

If the researchers enrol _____ subjects in the experimental arm, and _____ subjects in the standard arm then the power is _____, at the _____ significance level. Power was calculated by simualting _____ hypothetical studies.

# Question

Respond at **PollEv.com/nathantaback**
Text **NATHANTABACK** to **37607** once to join, then **1, 2, 3, 4, or 5**

**If the researchers enrol A subjects in the experimental arm, and B subjects in the standard arm then the power is C, at the D significance level. Power was calculated by simulating E hypothetical studies. The values A, B, C, D, E are:**

| | |
|---|---|
| A = 100, B = 300, C = 0.1434, D = 0.01, E = 400 | **1** |
| A = 300, B = 100, C = 1434, D = 0.01, E = 10000 | **2** |
| A = 300, B = 100, C = 0.1434, D = 0.01, E = 10000 | **3** |
| A = 300, B = 100, C = 0.1434, D = 0.05, E = 10000 | **4** |
| A = 100, B = 300, C = 0.1434, D = 0.05, E = 10000 | **5** |