STA305/1004 - Class 6

January 25, 2017

Today's Class

Reminder: Assignment is due Friday, Jan. 27 by 22:00. Submission is only via Crowdmark.

- Power of Statistical Tests
- Power of the one-sample z-test
- Power of the one-sample t-test
- Power of the two-sample t-test
- Sample size formulae two-samples means/proportions
- Calculating power via simulation

Why is Power and Sample Size Important in Phase III Clinical Trials?

- If a new treatment is to be used in patients then it should be compared to the standard treatment.
- Evidence is required that the new treatment is effective and safe.
- The form of the evidence is a hypothesis test.
- Will the hypothesis test reject if a difference between the treatments really exists?
- High power will ensure that if a difference exists then the hypothesis test will have a high probability of rejecting.
- The most practical way to ensure the test is powerful is to enrol enough patients in each arm of the trial.

$$\text{Power} = 1-\beta = 1 - P(\text{Type II})$$

$$= P(\text{Reject } H_0 \text{ when } H_A \text{ is true})$$

$H_0: \mu_A = \mu_B$

$H_A: \mu_A \neq \mu_B$

Sample Size and Power in Phase III Clinical Trials

- The sample size is calculated under the alternative hypothesis based on the type I error rate α and power $1-\beta$.
- Specify a clinically meaningful difference that is to be detected at the conclusion of the trial.
- Intuitively, if a small difference (effect size) is expected between the two treatments in comparison, a large sample size would be required, and vice versa. Why?
- Sample size also depends on the variance.
- The larger the variance, the harder it is to detect the difference and thus a larger sample size is needed.

Power of the one sample z-test

Let $X_1, ..., X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. A test of the hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_A : \mu \neq \mu_0$$

will reject at level α if and only if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2},$$

or

$$\left| \bar{X} - \mu_0 \right| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2},$$

where $z_{\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ percentile of the $N(0, 1)$.

Power of the one sample z-test

The power of the test at $\mu = \mu_1$ is

$\mu_1 \neq \mu_0$

H_0

$N(\mu_0, \frac{\sigma^2}{n})$

$$
\begin{aligned}
1 - \beta &= 1 - P\,(\text{type II error}) \\
&= P\,(\text{Reject } H_0, \text{ when } H_1 \text{ is true}) \\
&= P\,(\text{Reject } H_0, \text{ when } \mu = \mu_1) \\
&= P\left(\left|\bar{X} - \mu_0\right| \geq \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \text{ when } \mu = \mu_1 \right)
\end{aligned}
$$

H_1

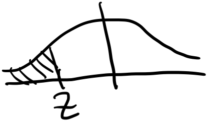
$N(\mu_1, \frac{\sigma^2}{n})$

Subtract the mean μ_1 and divide by $\sigma/\sqrt{n}$ to obtain (why?):

$$
1 - \beta = 1 - \Phi\left(z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right) \right) + \Phi\left(-z_{\alpha/2} - \left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right) \right),
$$

where $\Phi(\cdot)$ is the $N(0, 1)$ CDF.

Power of the one sample z-test



$P(Z \le z)$

The power function of the one-sample z-test is:

$$1-\beta = 1-\Phi\left(z_{\alpha/2}-\overset{\to 0}{\left(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)}\right)+\Phi\left(-z_{\alpha/2}-\overset{\to 0}{\left(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)}\right).$$

What is the limit of the power function as:

- $n \to \infty$
- $\mu_1 \to \mu_0$
- $\sigma \to 0$

↳ power → 1

as $\mu_1 \to \mu_0$ power → α

$1-(1-\alpha/2)+\alpha/2=\alpha$

- power → $\underline{1}$ as $n \to \infty$
- as sample size gets very large the power to detect the alternative → 1.

$z_{\alpha/2}$ is the $100(1-\alpha)$ percentile.

There are typically three ways to change power

1. Increase Sample size ($n \to \infty$)
2. Choose an alternative value μ_1 further away from μ_0 ($\mu_1 \to \mu_0$)
3. Decrease the variance ($\sigma \to 0$)

---

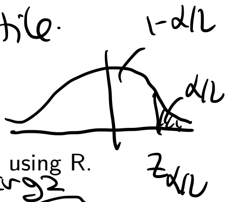
3. isn't practical
2. usually corresponds to the Scientific question.

$H_0: \mu = \mu_0$

$H_A: \mu \neq \mu_0$

Power of the one-sample z-test

$z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile.

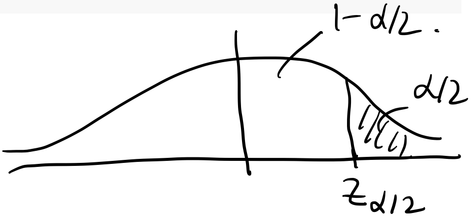


The power function for a one-sample z-test can be calculated using R.

$$1-\beta = 1-\Phi\left(\overbrace{z_{\alpha/2} - \left(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)}^{\text{arg1}}\right) + \Phi\left(\overbrace{-z_{\alpha/2} - \left(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}\right)}^{\text{arg2}}\right).$$

```
pow.z.test <- function(alpha,mu1,mu0,sigma,n){
  arg1 <- qnorm(1-alpha/2)-(mu1-mu0)/(sigma/sqrt(n))
  arg2 <- -1*qnorm(1-alpha/2)-(mu1-mu0)/(sigma/sqrt(n))
  1-pnorm(arg1)+pnorm(arg2)
}
```

Normal CDF



Power of the one-sample z-test

For example the power of the test

$$H_0 : \mu = 0 \text{ versus } H_0 : \mu = 0.2$$

with $n = 30, \sigma = 0.2, \alpha = 0.05$ can be calculated by calling the above function.

```
pow.z.test(.05,.15,0,.2,30)
```

```
[1] 0.9841413
```

What does this mean?

$H_0 : \mu = 0$ versus $H_0 : \mu = 0.2$ with $n = 30, \sigma = 0.2, \alpha = 0.05$

What does power=0.98 mean?

Respond at **PollEv.com/nathantaback**

Text **NATHANTABACK** to **37607** once to join, then **A, B, C, or D**

- The statistical test would reject the null hypothesis when the true mean is 0 in 2% of studies. **A** 4
- The statistical test would reject the null hypothesis when the true mean is 0.2 in 98% of studies. **B** 33
- The statistical test would ~~fail~~ reject the null hypothesis when the true mean is 0.2 in 98% of studies. **C** 4

  2% T
- The statistical test would ~~fail~~ reject the null hypothesis when the true mean is 0 in ~~2%~~ of studies. **D** 7

If the Study was replicated 100 times then on average the null hypothesis would be correctly rejected in 98 Studies.

Power of the one-sample t-test

Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. A test of the hypothesis

$$H_0 : \mu = \mu_0 \text{ versus } H_0 : \mu \neq \mu_0$$

will reject at level α if and only if

replace σ by S

$$\left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right| \geq t_{n-1,\alpha/2},$$

where $t_{n-1,\alpha/2}$ is the $100(1-\alpha/2)^{th}$ percentile of the t_{n-1}.

Power of the one-sample t-test

It can be shown that

$$\sqrt{n}\,\frac{(\bar{X}-\mu_1+\mu_1-\mu_0)\frac{\sqrt{n-1}}{\sigma}}{\frac{\sqrt{n-1}}{\sigma}\,S}$$

$$\sqrt{n}\left[\frac{\bar{X}-\mu_0}{S}\right] = \frac{Z+\gamma}{\sqrt{V/(n-1)}},$$

exercise for the class to show that

where,

$$Z = \frac{\sqrt{n}(\bar{X}-\mu_1)}{\sigma}$$

$$\gamma = \frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}$$

$$V = \frac{(n-1)}{\sigma^2}S^2.$$

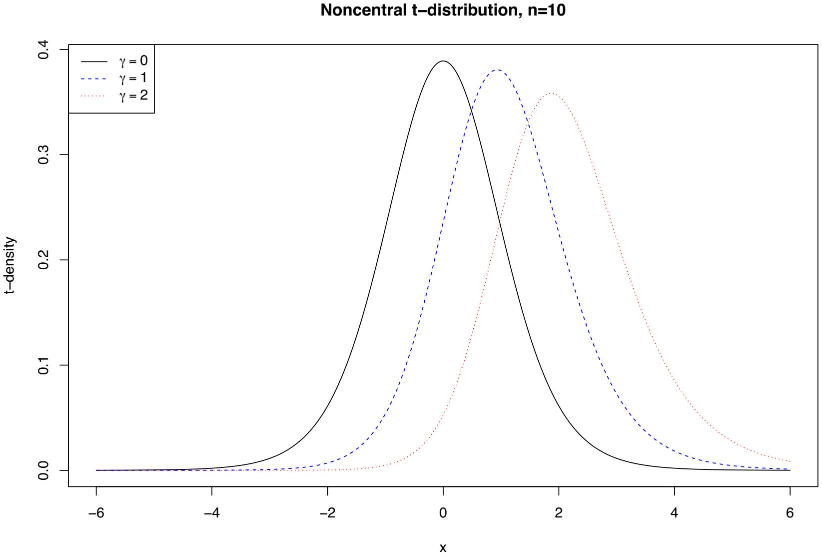
$Z \sim N(0,1)$ and $V \sim \chi^2_{n-1}$ and Z is independent of V.

Power of the one-sample t-test

- If $\gamma = 0$ then then $\sqrt{n}\left[\frac{\bar{X}-\mu_0}{S}\right] \sim t_{n-1}$. This is sometimes called the **central t-distribution**.
- If $\gamma \neq 0$ then $\sqrt{n}\left[\frac{\bar{X}-\mu_0}{S}\right] \sim t_{n-1,\gamma}$, where $t_{n-1,\gamma}$ is the **non-central t-distribution** with non-centrality parameter γ.

Power of the one-sample t-test

A plot of the central ($\gamma = 0$) and non-central t ($\gamma = 1, 2$) are shown in the plot below.



Power of the one-sample t-test

The power of the test at $\mu = \mu_1$ is

$$
\begin{aligned}
1 - \beta &= 1 - P\,(\text{type II error}) \\
&= P\,(\text{Reject } H_0 | H_1 \text{ is true}) \\
&= P\,(\text{Reject } H_0 | \mu = \mu_1) \\
&= P\left(\left| \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right| \geq t_{n-1,\alpha/2} | \mu = \mu_1 \right) \\
&= P(t_{n-1,\gamma} \geq t_{n-1,\alpha/2}) + P(t_{n-1,\gamma} < -t_{n-1,\alpha/2})
\end{aligned}
$$

Power of the one-sample t-test

Power $=$ $P(t_{n-1,\gamma} \geq t_{n-1,\alpha/2}) + P(t_{n-1,\gamma} < -t_{n-1,\alpha/2})$

The following function calculates the power function for the one-sample t-test in R:

type I   Sample Size   SD.

```
onesampttestpow <- function(alpha,n, mu0, mu1,sigma)
{delta <- mu1-mu0
t.crit <-qt(1-alpha/2,n-1)
t.gamma <- sqrt(n)*(delta/sigma)
t.power <- 1-pt(t.crit,n-1,ncp=t.gamma)
            +pt(-t.crit,n-1,ncp=t.gamma)
return(t.power)
}
```

H_0 value H_1 value

$$\gamma = \frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}$$

Power of the one-sample t-test

The power of the t-test for testing

$$H_0 : \mu = 0 \text{ versus } H_0 : \mu = 0.15$$

with $n = 10, \overset{S}{\cancel{\sigma}} = 0.2, \alpha = 0.05$ can be calculated by calling the above function is

```
onesampttestpow(.05,10,0,.15,0.2)
```

```
[1] 0.5619339
```

56% power.

Power of the one-sample t-test

Use the built-in function in R to calculate the power of t-test `power.t.test()`.

$\mu_1 - \mu_0$

```
power.t.test(n = 10,delta = 0.15,sd = 0.2,
             sig.level = 0.05,type = "one.sample" )
```

```
     One-sample t test power calculation

              n = 10
          delta = 0.15
             sd = 0.2
      sig.level = 0.05
          power = 0.5619339
    alternative = two.sided
```

Power of the two-sample t-test

- Consider a two-sample comparison with continuous outcomes. Let Y_{ik} be the observed outcome for the i^{th} subject in the kth treatment group, for $i = 1, ..., n_k$, and $k = 1, 2$. The outcomes in the two groups are assumed to be independent and normally distributed with different means but an equal variance σ^2,

two Samples

$N(\mu_1, \sigma^2)$
$N(\mu_2, \sigma^2)$

$$Y_{ik} \sim N(\mu_k, \sigma^2).$$

- Let $\theta = \mu_1 - \mu_2$, the difference in the mean between treatment 1 (the new therapy) and treatment 2 (the standard of care).
- To test whether the effects of the two treatments are the same, we formulate the null and alternative hypotheses as

$\mu_1 = \mu_2$

$\Leftrightarrow \underbrace{\mu_1 - \mu_2}_{= \Theta} = 0$

$$H_0 : \theta = 0 \text{ versus } H_0 : \theta \neq 0.$$

Data from both distributions want to test if $\mu_1 = \mu_2$

Power of the two-sample t-test

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-2)S_2^2}{(n_1-1)+(n_2-1)}}$$

The two-sample t statistic is given by

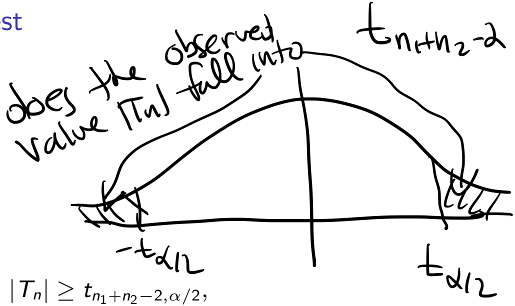
$$T_n = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{(1/n_1 + 1/n_2)}} \sim t_{n_1+n_2-2}.$$

- $T_n \sim t_{n_1+n_2-2}$ under H_0
- $T_n \sim t_{n_1+n_2-2,\gamma}$ with noncentrality parameter

$$\gamma = \frac{\mu_1 - \mu_2}{\sigma\sqrt{1/n_1 + 1/n_2}},$$

under H_1.

Power of the two-sample t-test



H_0 is rejected if

$$|T_n| \geq t_{n_1+n_2-2,\alpha/2},$$

where $t_{df,\alpha/2}$ is the $100(1-\alpha/2)th$ percentile of the central t-distribution with df degrees of freedom. (Yin, pg. 164-165)

- use `t.test()` to do the calculations.

Power of the two-sample t-test

The power of the test is

$$1 - \beta = 1 - P(t_{n_1+n_2-2,\gamma} \geq t_{n_1+n_2-2,\alpha/2}) + P(t_{n_1+n_2-2,\gamma} < -t_{n_1+n_2-2,\alpha/2})$$

The sample size can be solved from this equation which does not have a closed form.

The sample size can be determined by specifying:

- type I and type II error rates,
- the standard deviation,
- the difference in treatment means that the clinical trial aims to detect.

Power of the two-sample t-test

quantile from $t_{n_1+n_2-2}$

$$1-\beta = P(t_{n_1+n_2-2,\gamma} \geq t_{n_1+n_2-2,\alpha/2}) + P(t_{n_1+n_2-2,\gamma} < -t_{n_1+n_2-2,\alpha/2})$$

```
twosampttestpow <- function(alpha,n1,n2, mu1, mu2,sigma){
delta <- mu1-mu2
t.crit <-qt(1-alpha/2,n1+n2-2)
t.gamma <- delta/(sigma*sqrt(1/n1+1/n2))
t.power <- 1-pt(t.crit,n1+n2-2,ncp=t.gamma)+
             pt(-t.crit,n1+n2-2,ncp=t.gamma)
return(t.power)
}
```

non-centrality parameter

pt is the CDF of the $t_{n,\gamma}$

non-centrality parameter.

Power of the two-sample t-test

use this info. in the power calc.

A clinical trial to test a new treatment against the standard treatment for colon cancer is being designed. The investigators feel that the smallest meaningful difference in tumour growth is 1cm. The standard deviation of tumour growth is 3cm. The investigators feel that they can enrol 50 subjects per arm. Will this clinical trial have adequate power to detect a difference between the treatments?

- What are the parameters of interest? $\longrightarrow$ $\mu_S - \mu_N = \theta$
- What are the null and alternative hypotheses? $\longrightarrow$ $H_0: \theta = 0$, $H_A: \theta \neq 0$ (diff. exists.)
- How can the power of the study be calculated?

↳ use the function on previous slide with $\alpha = .05$, $n_1 = n_2 = 50$, $\delta = \mu_1 - \mu_2 = 1$, $s = 3$ cm.

Power of the two-sample t-test

$=\alpha \quad = n_1 \quad = n_2 \quad = \mu_1$

```
twosampttestpow(.05,50,50,1,2,3)
```

$\mu_2 = \quad = \sigma$

```
[1] 0.3785749
```

Power of the two-sample t-test

- `power.t.test()` can calculate the number of subjects required to achieve a certain power.
- Suppose the investigators want to know how many subjects per group would have to be enrolled in each group to achieve 80% power under the same conditions?

```
power.t.test(power = 0.8,delta = 1,sd = 3,sig.level = 0.05)
```

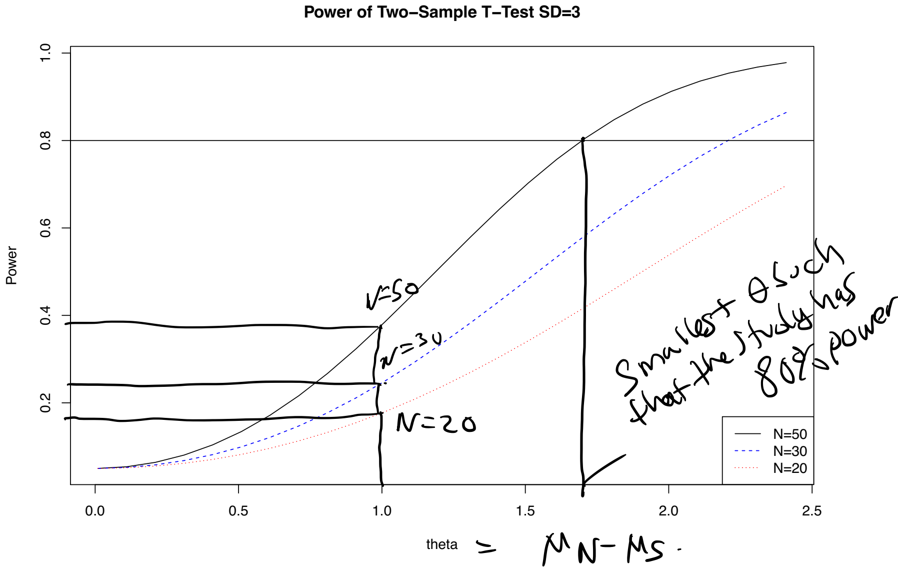
$\mu_{NT} - \mu_S$

```
     Two-sample t test power calculation

              n = 142.2466
          delta = 1
             sd = 3
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

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

This study requires 142 patients in each group (142x2 total patients) in order to detect a difference of at least 1 with 80% at the 5% sig. level.

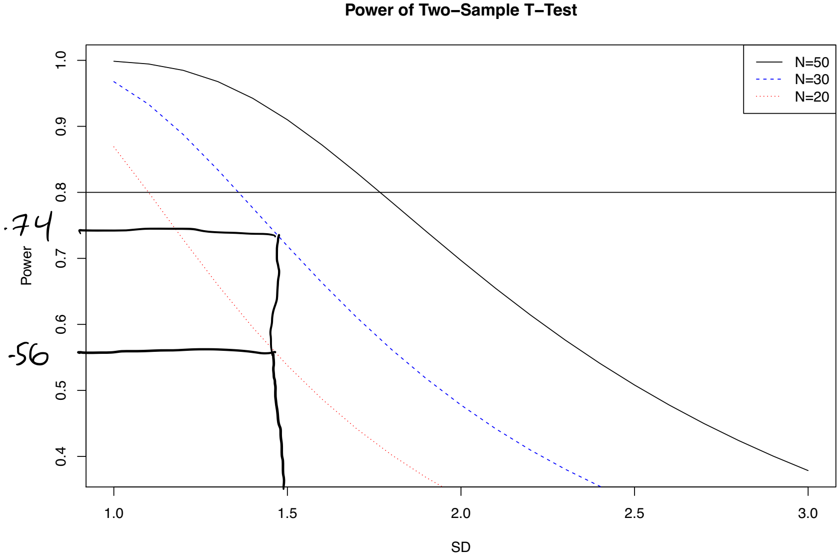
Power of the two-sample t-test

The following plot shows power of the two-sample t-test as a function of the difference $\theta = \mu_1 - \mu_2$ to be detected and equal sample size per group.



Power of the two-sample t-test

This plot shows power as a function of $\cancel{\alpha}$ S and sample size per group.



Power of the two-sample t-test

In some studies instead of specifying the difference in treatment means and standard deviation separately the ratio

$$ES = \frac{\mu_1 - \mu_2}{\sigma}$$

can be specified.

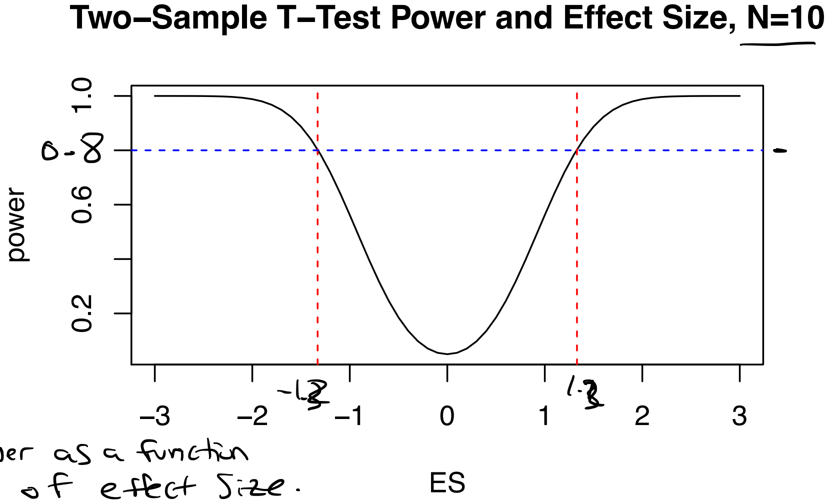
- ES is called the scaled effect size.
- Cohen (1992) suggests that effect sizes of 0.2, 0.5, 0.8 correspond to small, medium , and large effects respectively.

Power of the two-sample t-test

$ES = \frac{\mu_1 - \mu_2}{\sigma}$

Power as a function of effect size can be investigated.

The plot shows that for $n_1 = n_2 = 10$ the two-sample t-test has at least 80% power for detecting effect sizes that are at least 1.3.



Power as a function of effect size.

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

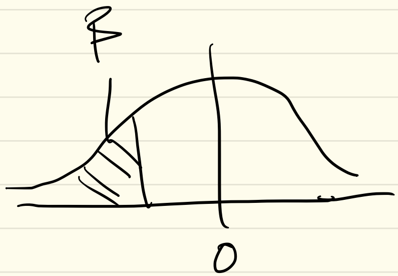
$$1-\beta = P\left(Z \geq z_{\alpha/2} - \frac{\theta_1}{\sigma\sqrt{1/n_1 + 1/n_2}}\right) + P\left(Z < -z_{\alpha/2} - \frac{\theta_1}{\sigma\sqrt{1/n_1 + 1/n_2}}\right).$$

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

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

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

$$\Phi^{-1}(\beta) \simeq z_{\alpha/2} - \frac{|\theta_1|}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$



$\parallel$?

$$-z_\beta = z_{\alpha/2} - \frac{|\theta_1|}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$\parallel$

z_β is the 100(1−β) percentile.

$n_1 = n_2 = \frac{n}{2}$ then Solve for n.

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

this can be shown from previous formula.

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

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))}
```

$r = \frac{n_1}{n_2}.$

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(1,.05,.1,2,2)[1]
```

```
[1] 126.0891
```

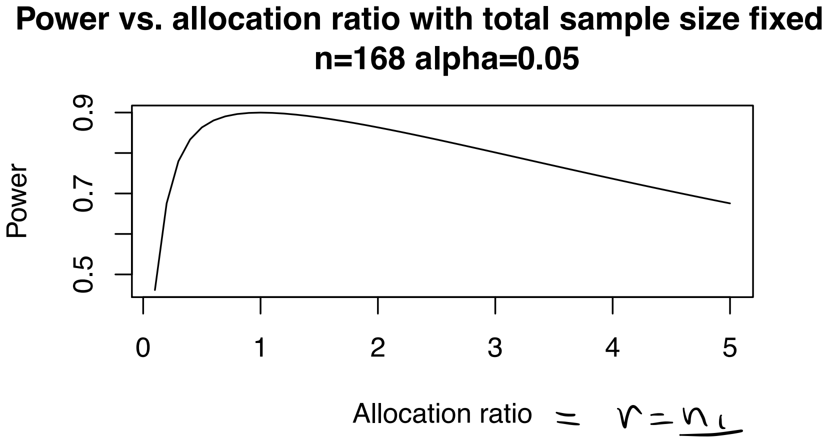
```
# group 2 sample size (control group)
size2z.uneq.test(1,.05,.1,2,2)[2]
```

```
[1] 63.04454
```

total Sample Size
= 126+63

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.



$= r = \frac{n_1}{n_2}$

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