STA305/1004 - Class 6

January 27, 2016

Today's Class

- 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.

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

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

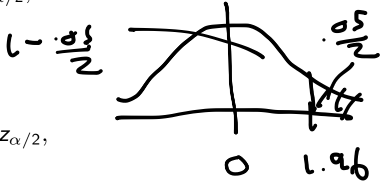
$\alpha = 0.05$

$\alpha/2 = .025$

$z_{\alpha/2} = 1.96$

will reject at level α if and only if

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



or

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

$\alpha = .05$

97.5% area below 1.96

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

Power of the one sample z-test

e.g.,

$\mu_1 = 1$

or $\mu_1 = 2$

or $\mu_1 = 1/2$

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

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

$N(\mu_1, \sigma^2/n)$

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

depends on α, n, μ_1, μ_0, σ, n

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

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

$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$

Cumulative distn. function

Power of the one sample z-test

as $\mu_1 \to \mu_0$ $\to$ power $1-\Phi(z_{\alpha/2})+\Phi(-z_{\alpha/2})$

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

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

as $\sigma \to 0$ power $\to 1$.

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

these two param. can be controlled

What is the limit of the power function as:

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

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

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

$n \to \infty$

$\Phi(\cdot) \to 0$

$x \to -\infty$ $\Phi(x) \to 0$

$\therefore$ power $\to 1$

Power of the one-sample z-test

arg1

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

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

arg2.

```
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)
}
```

CDF of $N(\mu, \sigma^2)$

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
```

$\mu = 0.15$

What does this mean?

$\mu = 0.02$

$H_0: \mu = 0$ vs $H_A: \mu = 0.2$

What does power=0.98 mean?

$P(\text{Reject } H_0 \mid \mu = 0.2)$

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**	6 X
17	The statistical test would reject the null hypothesis when the true mean is 0.2 in 98% of studies.	**B**	20
	The statistical test would fail reject the null hypothesis when the true mean is 0.2 in 98% of studies.	**C**	5 X
3	The statistical test would fail reject the null hypothesis when the true mean is 0 in 2% of studies.	**D**	10

Total Results: 0

Concept of Sampling distn. of $\bar{X}$.

- If one obtains different random samples from the pop. of interest then the distribution of $\bar{X}$ is called the sampling distn.

- Imagine the study could be repeated 100 times then 98 of the 100 studies would reject Ho when $\mu = 0.2$ at $\alpha = .05$, $n = 30$.

Power of the one-sample t-test

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

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

One Sample z-test

will reject at level α if and only if

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

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

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

t-test does not assume that σ is known.

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{is} \quad N\left(\mu_0, \sigma^2/n\right)$$

but μ_0 σ is replaced by S then

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad \text{is} \quad t_{n-1}$$

as n gets large $(n \to \infty)$

then t_{n-1} approaches the Normal.

approx. $n \geq 30^{\text{th}} - 40$

Power of the one-sample t-test

It can be shown that

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

where,

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

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

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

under H_0
$\gamma = 0$
under H_1
$\gamma \neq 0$

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

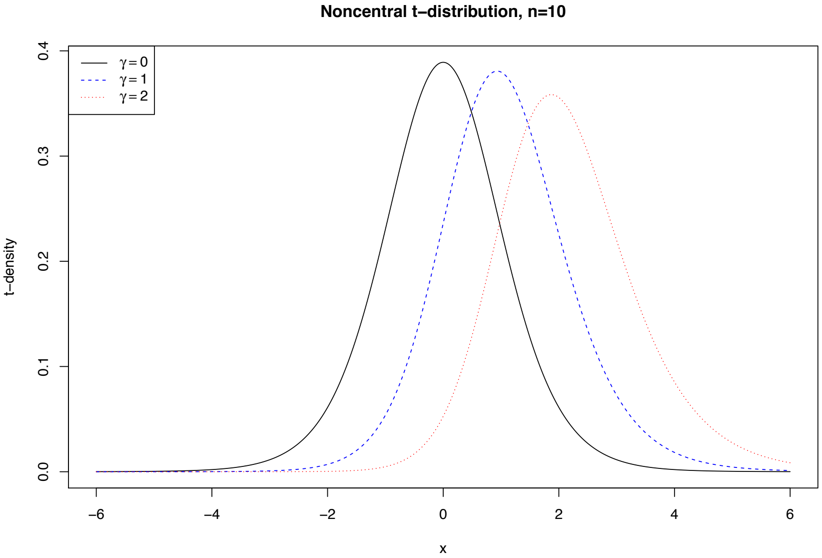
Power of the one-sample t-test

- 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.

occurs under H_1.

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

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

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

Power of the one-sample t-test

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

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

```
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)
}
```

non-centrality parameter.

pt(·) — CDF for $t_{n,\gamma}$

qt(·) gives the t-quantiles. df   non. cent. Param

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 $n = 10, \sigma = 0.2, \alpha = 0.05$ can be calculated by calling the above function is

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

```
[1] 0.5619339
```

power under this Scenario

Power of the one-sample t-test

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

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

$\mu_1 - \mu_0$

```
     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 $i = 1, \ldots, n_k$ 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,

diff. Sample size in each group

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

$K = 1, 2$ two treatments

- 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

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

Power of the two-sample t-test

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

The two-sample t statistic is given by

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

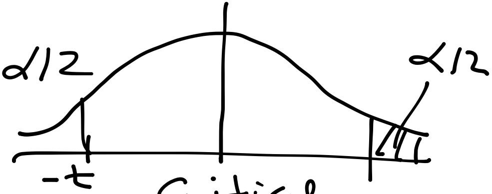
- $T_n \sim t_{n_1+n_2-2}$ under H_0 $\mu_1 = \mu_2$
- $T_n \sim t_{n_1+n_2-2,\gamma}$ with noncentrality parameter $\rightarrow \mu_1 - \mu_2 \neq 0$

$\underbrace{n_1+n_2-2}_{= \text{df}}$, γ = non. cent. param.

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

under H_1.

Power of the two-sample t-test



H_0 is rejected if

Critical Value.

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

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

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 (α) and type II (β) error rates,
- the standard deviation, σ
- the difference in treatment means that the clinical trial aims to detect.

$\mu_1 - \mu_2$

Power of the two-sample t-test

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

```
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)
}
```

Power of the two-sample t-test

$H_0: \delta = 0$

$H_A: \delta \neq 0$

two-sided ∴ question specifies "detect difference"!

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?
- What are the null and alternative hypotheses?
- How can the power of the study be calculated?

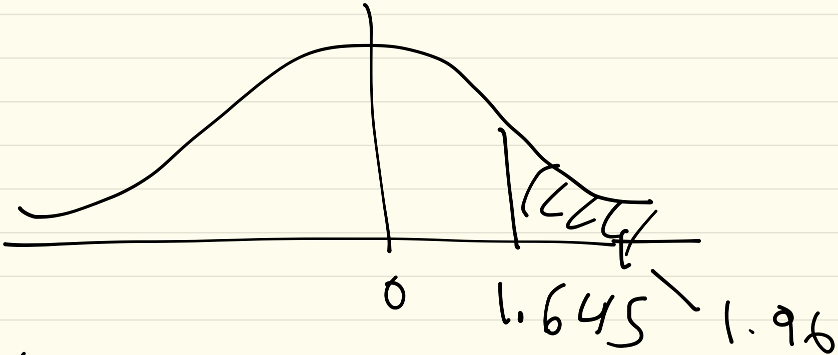
δ = diff. in tumour growth.

$= \mu_N - \mu_S$

μ_N = Mean tumour growth for New trt.

μ_S = mean tumour growth for standard trt.

$\alpha = 0.05$



If the test is one-sided e.g., $H_A: \delta > 0$ then power would ↑ for the same value of n. (power)

For a fixed value of power then n ↓ (Sample Size)

$H_0: \mu = 0$

$H_A: \mu \neq 0$

Calc. power to detect $\mu = 1\,cm$.

Power of the two-sample t-test

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

```
[1] 0.3785749
```

38% power.

If the study was repeated 100 times and the difference is at least 1 then 38 of the 100 studies will reject Ho.

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)
```

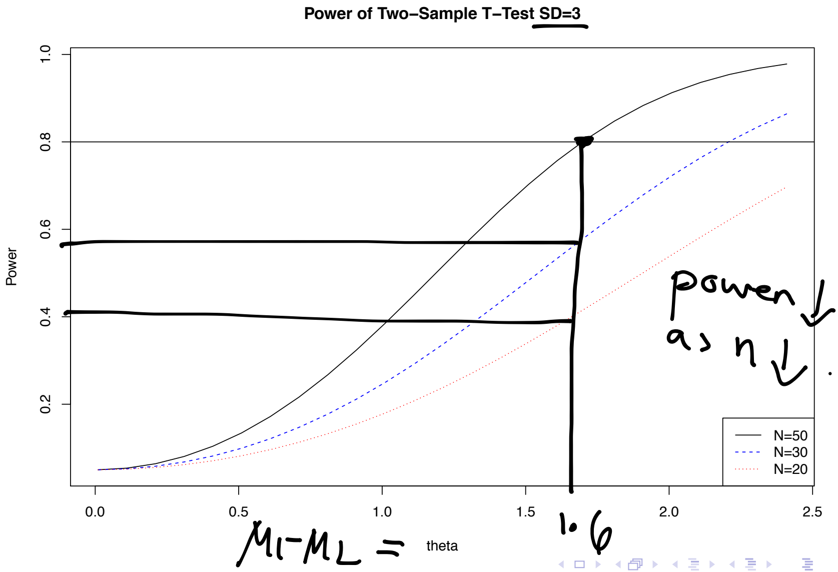
$1-\beta$ $\delta = \mu_1 - 1$ σ α

```
     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
```

Total # Subjects. is 142 x 2.

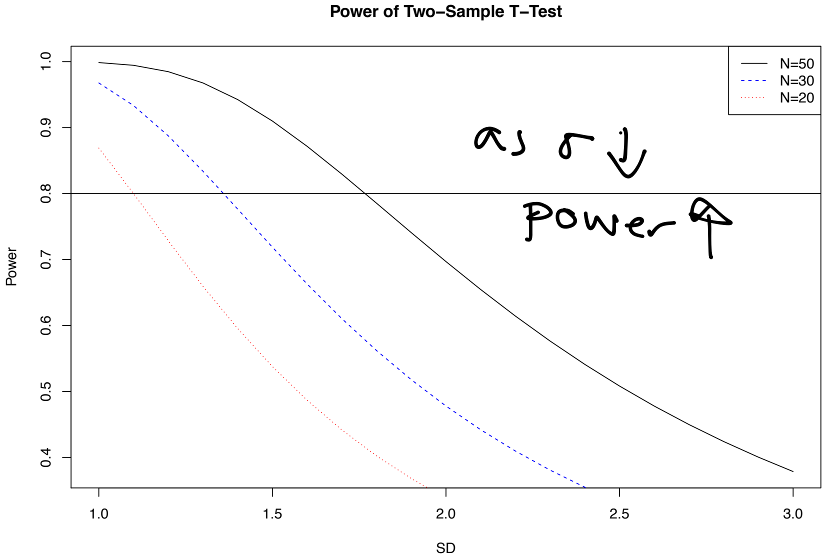
Power of the two-sample t-test

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 σ 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

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

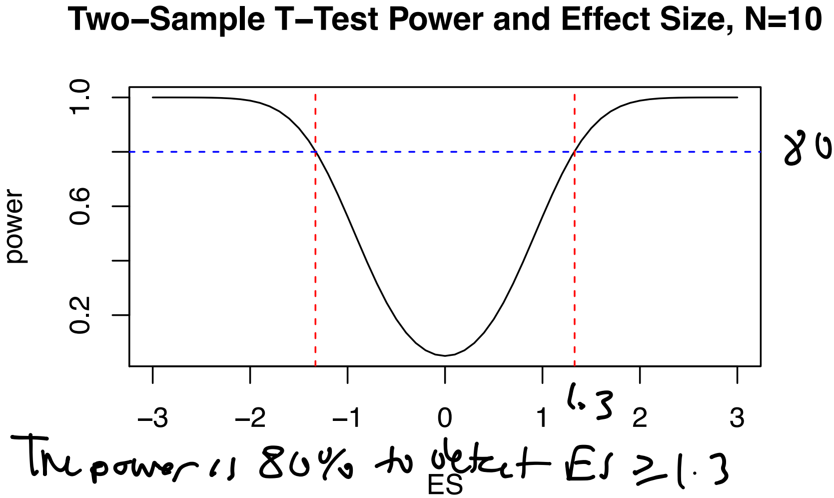
can be specified.

- 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

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.



80%

The power is 80% to detect ES ≥ 1.3

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

We assume that σ is known!

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

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

$\Phi^{-1}(\beta) = \Phi^{-1}\Phi(\ )$

$z_\beta = \Delta$

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

Sample size - known variance and equal allocation

The sample size is obtained by solving

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

(2)

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

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

$\sigma, \beta, \alpha, \theta = \mu_1 - \mu_2$

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

New trt. #subjects in Standard trt group.

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

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

allocation ratio of 2 : 1

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]
```

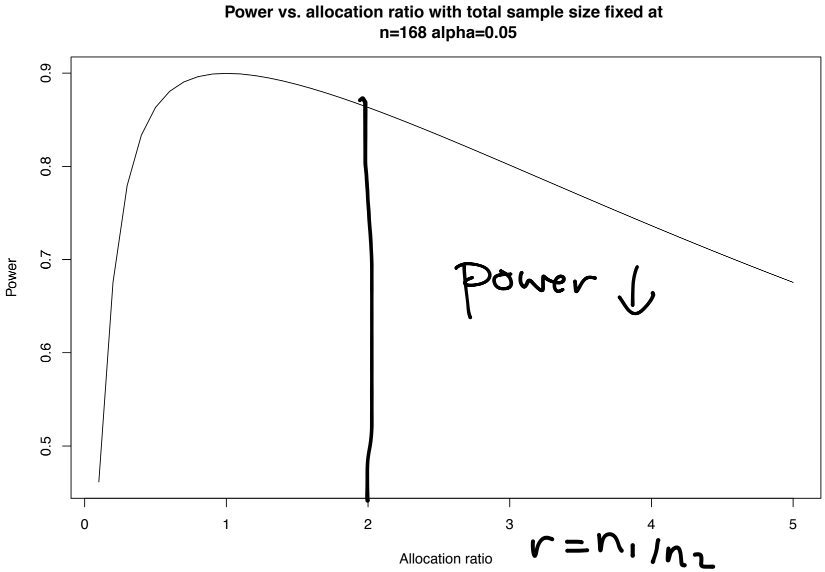
```
[1] 63.04454
```

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.



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 $\theta = p_1{}^{-}p_2$.

$$= p_1 - p_2$$

Comparing Proportions for Binary Outcomes

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

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

Bern(p_k).

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:

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

The test rejects at level α 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 β

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

P(type II)

Comparing Proportions for Binary Outcomes

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

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

α, θ, p_1, p_2

$.8 - .5 = \delta$

$.4 - .1 = \delta$

leads to diff. power.

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 = 0.25,power = 0.8)
```

p_1 p_2 $1-\beta$

```
     Two-sample comparison of proportions power calculation

              n = 1093.739
             p1 = 0.2
             p2 = 0.25
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

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

$n_1 = n_2 = 1094$

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. when $\mu_1 - \mu_2 = 0.5$

Calculating Power by Simulation

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

$\mu=3.5$ $\mu=3$

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

generates random sample of 30 obs. $N(3.5, 1)$

```
	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
```

Reject p-value ≤ 0.05

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
```

Estmate of power.

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
```