STAT 141 **11/02/04**

POWER and SAMPLE SIZE

Rejection & Acceptance Regions

Type I and Type II Errors (S&W Sec 7.8)

Power

Sample Size Needed for One Sample z-tests.

Using R to compute power for t.tests

For Thurs: read the Chapter 7.10 and chapter 8

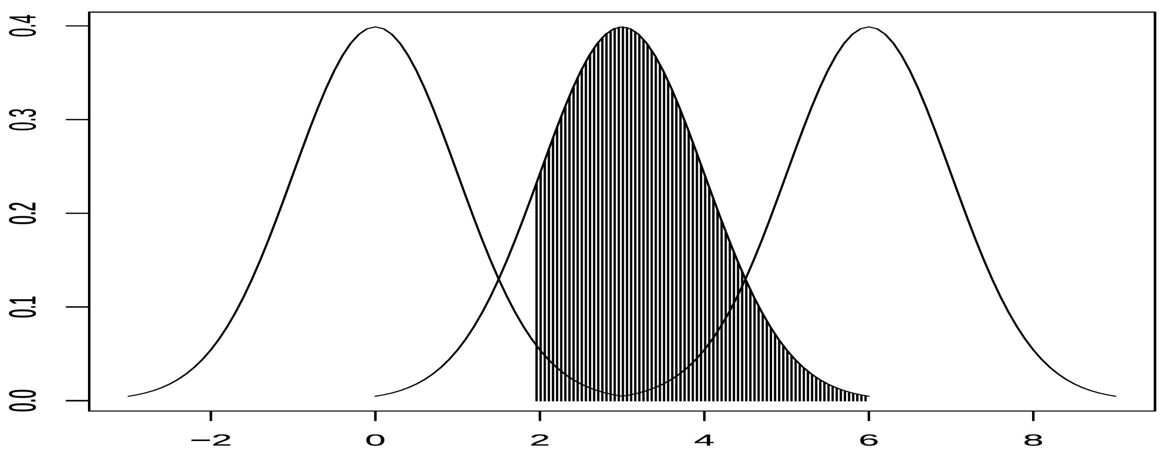
A typical study design question: A new drug regimen has been developed to (hopefully) reduce weight in obese teenagers. Weight reduction over the one year course of treatment is measured by change X in body mass index (BMI). Formally we will test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$. Previous work shows that $\sigma_x = 2$. A change in BMI of 1.5 is considered important to detect (if the true effect size is 1.5 or higher we need the study to have a high probability of rejecting H_0. How many patients should be enrolled in the study?

The testing example we use below is the simplest one: if $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$, test $H_0 : \mu = \mu_0$ against the two-sided alternative $H_1 : \mu \neq \mu_0$ However the concepts apply much more generally.

A test at level α has both:

$$\textit{Rejection region} : R = \{\bar{x} > \mu_0 + z_{\alpha/2}\sigma_{\bar{x}}\} \cup \{\bar{x} < \mu_0 - z_{\alpha/2}\sigma_{\bar{x}}\}$$

$$\text{“Acceptance” region} : A = \{|\bar{x} - \mu_0| < z_{\alpha/2}\sigma_{\bar{x}}\}$$



Two kinds of errors:
Type I error is the error made when the null hypothesis is rejected when in fact the null hypothesis is true.
Alpha (α) is the probability of rejecting a true null hypothesis.
Type II error is the error made when the null hypothesis is not rejected when in fact the alternative hypothesis is true.

Beta (β) is the probability of not rejecting a false null hypothesis   Power $= 1 - \beta$

The probability of rejecting false null hypothesis. The power of a test tells us how likely we are to find a significant difference given that the alternative hypothesis is true (the true mean is different from the mean under the null hypothesis).

	$\bar{x} \in A$ "accept H_0"	$\bar{x} \in R$ reject H_0"
H_0 true	OK	Type I error $\alpha = P($ Type I $\|H_0)$ false alarm
H_A true	Type II error $\beta = P($ Type II $\|H_A)$ Alarm doesn't go off with fire	OK

Fact: A level α test controls type I error.

What about $\beta = P(\text{Type II error})$, we want this to be small. But this is not guaranteed by controlling α: the two types of error do not play a symmetric role.

Note from the figure that

$$Power = 1 - \beta = 1 - P(\text{type II error}) = P(\text{ reject } | H_0 false) = P(\bar{x} \in R | \mu \neq \mu_0)$$

- depends on μ
- increases as $\mu - \mu_0$ increases.

In our example, the z-test, we can be more explicit and derive a formula which shows how the power depends on n, $\mu - \mu_0$, α and σ.

First, define the *effect size* $\Delta = \delta = \frac{\mu - \mu_0}{\sigma}$, the number of standard deviations the true mean is away from the tested one.

Also, recall that we denoted $P(Z \leq z) = \Phi(z)$, the area to the left of z under the standard Normal curve.

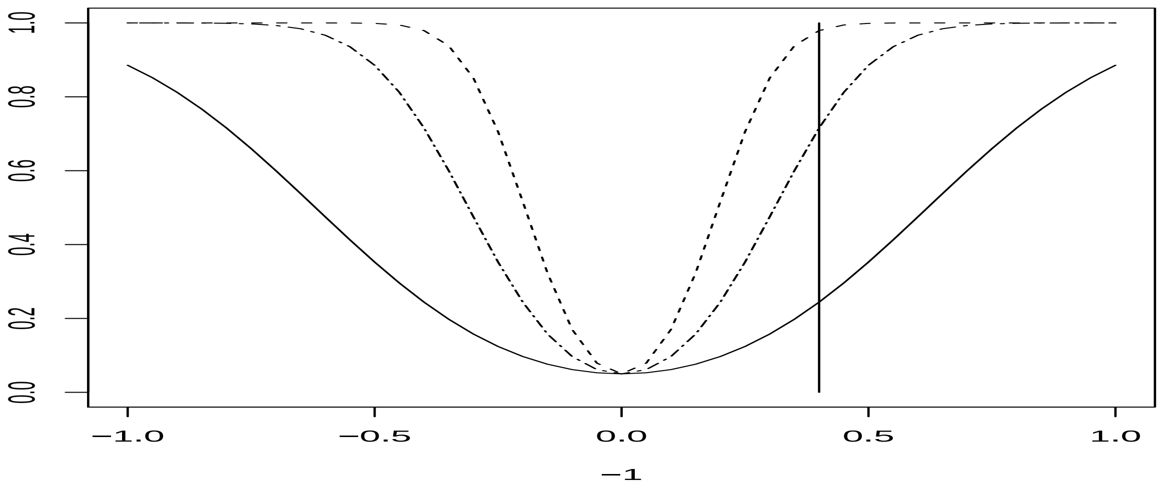
Fact: If $\bar{x} \sim \mathcal{N}(\mu, \sigma_{\bar{x}}^2)$ then the power of the two sided z-test at level α is given by

$$\begin{aligned}
\text{Power} &= P_\mu(\bar{x} > \mu_0 + z_{1-\alpha/2}\sigma_{\bar{x}}) + P_\mu(\bar{x} < \mu_0 - z_{1-\alpha/2}\sigma_{\bar{x}}) \\
&= P(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} > \frac{\mu_0 - \mu}{\sigma_{\bar{x}}} + z_{1-\alpha/2}) + P(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} < \frac{\mu_0 - \mu}{\sigma_{\bar{x}}} - z_{1-\alpha/2}) \\
&= \Phi(\sqrt{n}\Delta - z_{1-\alpha/2}) + \Phi(-\sqrt{n}\Delta - z_{1-\alpha/2})
\end{aligned}$$

(The approximation is $\Phi(-\sqrt{n}\Delta - z_{1-\alpha/2}) \simeq 0$ o.k. if $\sqrt{n}\Delta \geq 1$)

Power curve plots the power as a function of the effect size for several values of n.

```
plot(delta,pnorm(sqrt(10)*delta-qnorm(0.975))+
pnorm(-sqrt(10)*delta-qnorm(0.975)),xlab=delta,type='l',
ylim=c(0,1) ,ylab='' )
 par(new=TRUE)
 lines(delta,pnorm(sqrt(40)*delta-qnorm(0.975))+
pnorm(-sqrt(40)*delta-qnorm(0.975)),lty=6)
 lines(delta,pnorm(sqrt(100)*delta-qnorm(0.975))+
pnorm(-sqrt(100)*delta-qnorm(0.975)),lty=2)
lines(c(0.4,0.4),c(0,1))
```



Hence: Ways to increase power:
♠ larger *n*
♢ larger $\mu - \mu_0$
♡ larger α
♣ smaller σ

Sample size needed to achieve a desired power: single sample

Suppose we want power = $1-\beta$ (e.g. .90 or .95 say) to detect an effect of size δ.
Solve the equation $\Phi(\sqrt{n}\Delta - z_{1-\alpha/2}) = 1-\beta$ to yield the formula for the **necessary sample size** as

$$n = (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{1}{\Delta^2}$$

Table of multipliers $(z_{1-\alpha/2} + z_{1-\beta})^2$

Power/Alpha	.01	.05	.10
.80	11.7	7.9	6.2
.90	14.9	10.5	8.6
.95	17.8	13.0	10.8

Example: $\alpha = 0.05$, Power=0.95 $\longrightarrow \beta = 0.05$, $z_{0.975} = 1.96$, $z_{0.95} = 1.65$, $(z_{0.975} + z_{0.95})^2 = 3.6^2 = 13$
BMI Example:
$\Delta = \frac{1.5}{2}$, $\frac{1}{\Delta^2} = \frac{4^2}{3^2}$, $n = 13 \times \frac{16}{9} = 23.11 \longrightarrow n = 24$ patients are needed.
Summary: To calculate the necessary sample size, we have to specify

1. α the level of the test
2. the desired power : 1-β.
3. the SD of a single observation σ
4. the magnitude of the difference you want to detect $\mu - \mu_0$

Remarks:

1. A Type I error can only occur when a null hypothesis is true. (You incorrectly reject a true null hypothesis.)

2. A Type II error can only occur when a null hypothesis is false. (You incorrectly fail to reject a false null hypothesis.)

3. The Power of a test is 1 - probability (Type II error). (This is the probability that you correctly reject a false null hypothesis.)

4. One needs an alternative to the null hypothesis in order to calculate a Type II error. Without an alternative hypothesis, the question "what is the probability of a Type II error?" is meaningless.

Computing power with R:

```
power.t.test(n=10,delta=0.4,type="one.sample")
     One-sample t test power calculation
              n = 10
          delta = 0.4
             sd = 1
      sig.level = 0.05
          power = 0.2041945
    alternative = two.sided
######################################

power.t.test(n=40,delta=0.4,type="one.sample")
     One-sample t test power calculation
              n = 40
          delta = 0.4
             sd = 1
      sig.level = 0.05
          power = 0.6939817
    alternative = two.sided
###################################################
power.t.test(delta=.75,type="one.sample",alternative="t",power=.95)
     One-sample t test power calculation
              n = 25.11093
          delta = 0.75
             sd = 1
      sig.level = 0.05
          power = 0.95
    alternative = two.sided


#delta is the true difference in means, not
#the number of standard deviations the means are apart
#in the traditional notation, the default is for sd=1,
#then of course it has the same meaning.
```

Two sample tests

The best use of $2n$ observations is to make two equal sample sizes.

```
power.t.test(n=NULL, delta=NULL, sd=1, sig.level=0.05, power=NULL,
            type=c("two.sample", "one.sample", "paired"),
            alternative=c("two.sided", "one.sided"), strict=FALSE)
```

Example: Influence of milk on growth. We want to know th sample size needed, for a power of 0.9 or 90% using a two-sided test at the 1% level. The minimum detectable difference should be 0.5cm and the sd of the distribution is 2cm.

```
> power.t.test(delta=0.5,sd=2,sig.level=0.01,power=0.9)

     Two-sample t test power calculation

              n = 477.8021
          delta = 0.5
             sd = 2
      sig.level = 0.01
          power = 0.9
    alternative = two.sided

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

Actually, a sample size of 450 was used, what is the power if only n=450 is used in each sample.

```
> power.t.test(n=450,delta=0.5,sd=2,sig.level=0.01)

     Two-sample t test power calculation

              n = 450
          delta = 0.5
             sd = 2
      sig.level = 0.01
          power = 0.8784433
    alternative = two.sided

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

Power for proportion tests

```
> power.prop.test(power=.85,p1=.48,p2=.52,sig.level=0.01)

     Two-sample comparison of proportions power calculation

              n = 4075.766
             p1 = 0.48
             p2 = 0.52
      sig.level = 0.01
          power = 0.85
    alternative = two.sided

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

Only two sample problems are considered as yet.