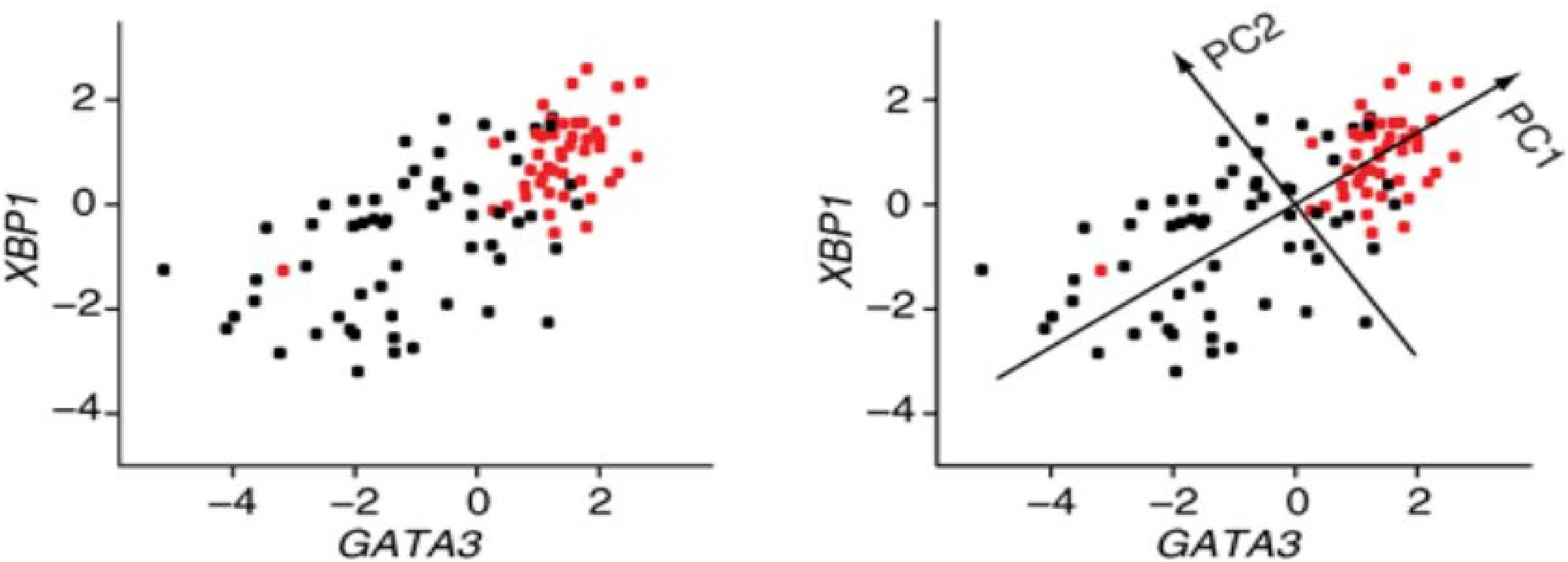
DAY 4 – LECTURE OUTLINE

- ESEMPIO METABOANALIST
- Machine learning algorithms
  - Unsupervised
    - Example: PCA
- Elements of Statistical Power analsys
  - Underlying principles of statistical power
  - Power calculations for basic study designs
  - Use power and sample size calculations as the basis of argument in support of study design, feasibility, and testing

PCA

Principal Component Analysis

Principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. In this the dimension of the data is reduced to make the computations faster and easier. It is used to explain the variance-covariance structure of a set of variables through linear combinations. It is often used as a dimensionality-reduction technique.



The way we find the principal components is as follows:
Given a dataset with *p* predictors: X_1, X_2, … , $X_{p,}$, calculate Z_1, … , Z_M to be the *M* linear combinations of the original *p* predictors where:

- $Z_m = \Sigma \Phi_{jm} X_j$ for some constants Φ_{1m}, Φ_{2m}, Φ_{pm}, m = 1, …, M.
- Z_1 is the linear combination of the predictors that captures the most variance possible.
- Z_2 is the next linear combination of the predictors that captures the most variance while being *orthogonal* (i.e. uncorrelated) to Z_1.
- Z_3 is then the next linear combination of the predictors that captures the most variance while being orthogonal to Z_2.
- And so on.

In practice, we use the following steps to calculate the linear combinations of the original predictors:

1. Scale each of the variables to have a mean of 0 and a standard deviation of 1.

2. Calculate the covariance matrix for the scaled variables.

3. Calculate the eigenvalues of the covariance matrix.

Using linear algebra, it can be shown that the eigenvector that corresponds to the largest eigenvalue is the first principal component. In other words, this particular combination of the predictors explains the most variance in the data.

The eigenvector corresponding to the second largest eigenvalue is the second principal component, and so on

POWER ANALYSIS

BIOSTATISTICS TRAINING COURSE

by Maria Chiara Mimmi, Ph.D. in Chemistry,
Post-Lauream Degree in Biostatistics and Epidemiology

DAY 5
Power analysis

Decision Errors

Two types of errors can result from a hypothesis test.

Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by α.

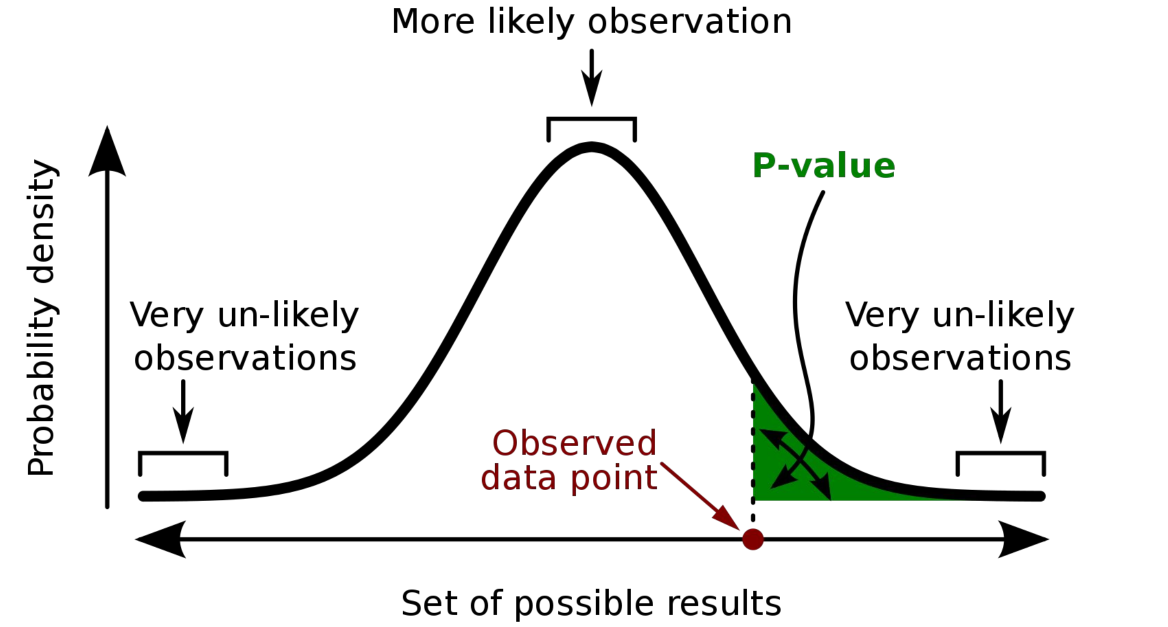
Type II error occurs when the researcher fails to reject a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by β. The probability of not committing a Type II error is called the **Power** of the test.

The Table Summarizing Type I and Type II Errors

	decide H_0	decide H_1
true H_0 probability	Correct action $1-\alpha$	Type I error α
true H_1 probability	Type II error β	Correct action power $= 1-\beta$

$$\alpha = P(H_1|H_0)$$
$$\beta = P(H_0|H_1)$$

P-value



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

- The P-value corresponds to the answer the question: "What is the probability of the observed test statistic or one more extreme when H_0 is true?"

A very small p-value means that such an extreme observed outcome would be very unlikely under the null hypothesis.

Usually the researcher fix the type I error (α) he can tolerate before experiment and then compare the p value and takes a decision.

Controlling Type I and Type II Errors

- α, β, and n are related
- when two of the three are chosen, the third is determined
- α and n are usually chosen
- try to use the largest α you can tolerate
- if Type I error is serious, select a smaller α value and a larger n value

Recall:

- $\alpha = P(H_1|H_0)$
- $\beta = P(H_0|H_1) = Pr(\text{retain } H_0|H_0 \text{ is false}) \equiv$ probability of a Type II error
- $1-\beta$ is called power of hypothesis test;
- $1-\beta = P(\text{reject } H_0|H_0 \text{ is false}) \equiv$ probability of avoiding a Type II error

Controlling Type I and Type II Errors

- α, β, and n are related
- when two of the three are chosen, the third is determined
- α and n are usually chosen
- try to use the largest α you can tolerate
- if Type I error is serious, select a smaller α value and a larger n value

Recall:

- $\alpha = P(H_1|H_0)$
- $\beta = P(H_0|H_1) = Pr(\text{retain } H_0|H_0 \text{ is false}) \equiv$ probability of a Type II error
- $1-\beta$ is called power of hypothesis test;
- $1-\beta = P(\text{reject } H_0|H_0 \text{ is false}) \equiv$ probability of avoiding a Type II error

Power of a z test

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

where

- $\Phi(z)$ represent the cumulative probability of Standard Normal Z
- μ_0 represent the population mean under the null hypothesis
- μ_1 represents the population mean under the alternative hypothesis

Power of a z test

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

where

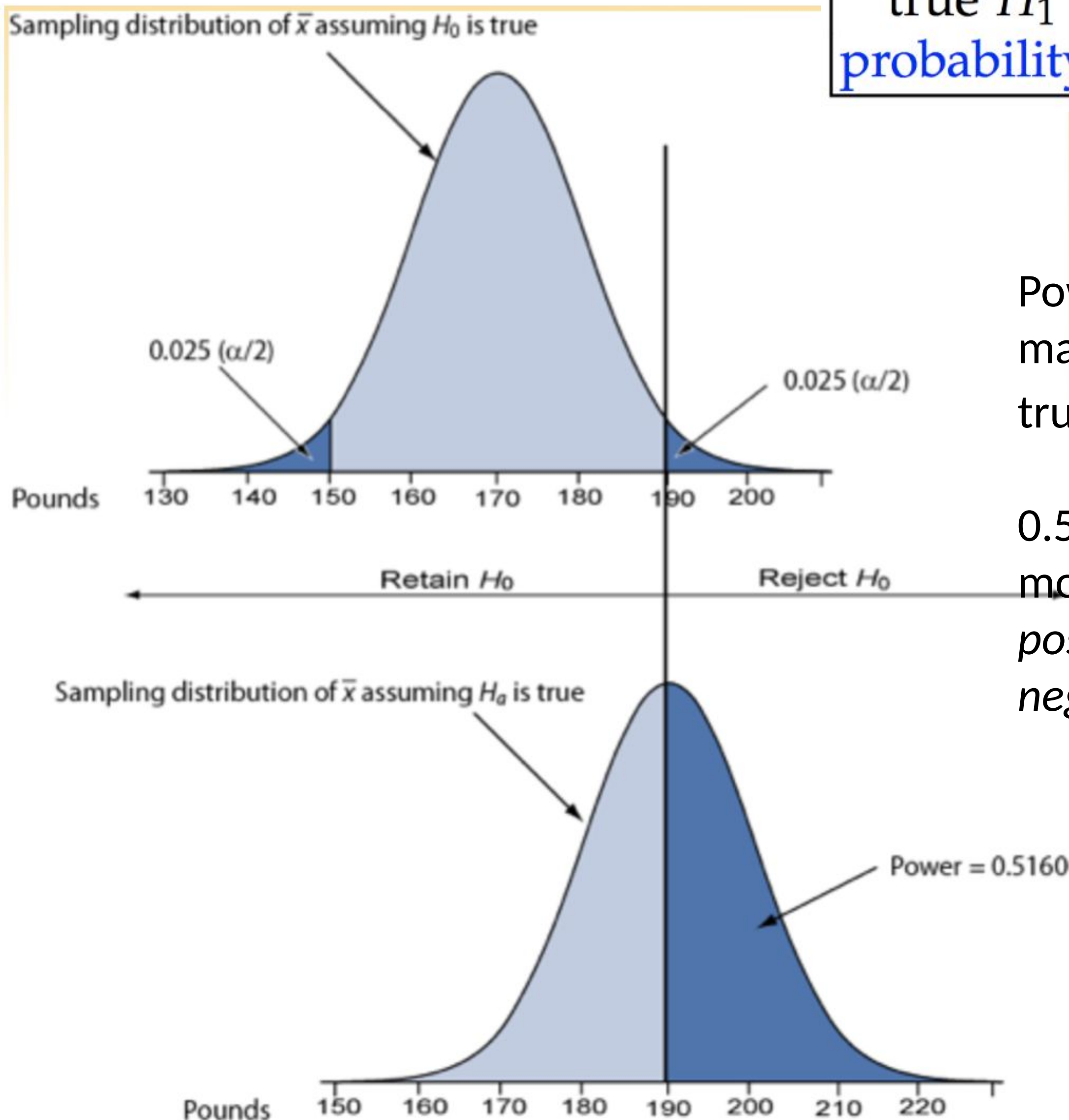
- $\Phi(z)$ represent the cumulative probability of Standard Normal Z
- μ_0 represent the population mean under the null hypothesis
- μ_1 represents the population mean under the alternative hypothesis

Example: Calculating Power

A study of $n = 16$ retains $H_0 : \mu = 170$ at $\alpha = 0.05$ (two-sided); $\sigma = 40$. What was the power of test's conditions to identify a population mean of 190?

$$\begin{aligned} 1-\beta &= \Phi\left(-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_1|\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(-1.96 + \frac{|170 - 190|\sqrt{16}}{40}\right) \\ &= \Phi(0.04) = 0.5160 \end{aligned}$$

	decide H_0	decide H_1
true H_0 probability	Correct action $1-\alpha$	Type I error α
true H_1 probability	Type II error β	Correct action power $= 1-\beta$



Power = 1-β = probability to make the correct decision in case H_1 is true.

0.51 is a low power, but normally we pay more attention to Type I error (*false positive*) thann to Type II error (*false negative*).

Sample Size Required

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

where

- $1 - \beta \equiv$ desired power
- $\alpha \equiv$ desired significance level (two-sided)
- $\sigma \equiv$ population standard deviation
- $\Delta = \mu_0 - \mu_1 =$ difference worth decision

	decide H_0	decide H_1
true H_0 probability	Correct action $1-\alpha$	Type I error α
true H_1 probability	Type II error β	Correct action power $= 1-\beta$

Example: Sample Size

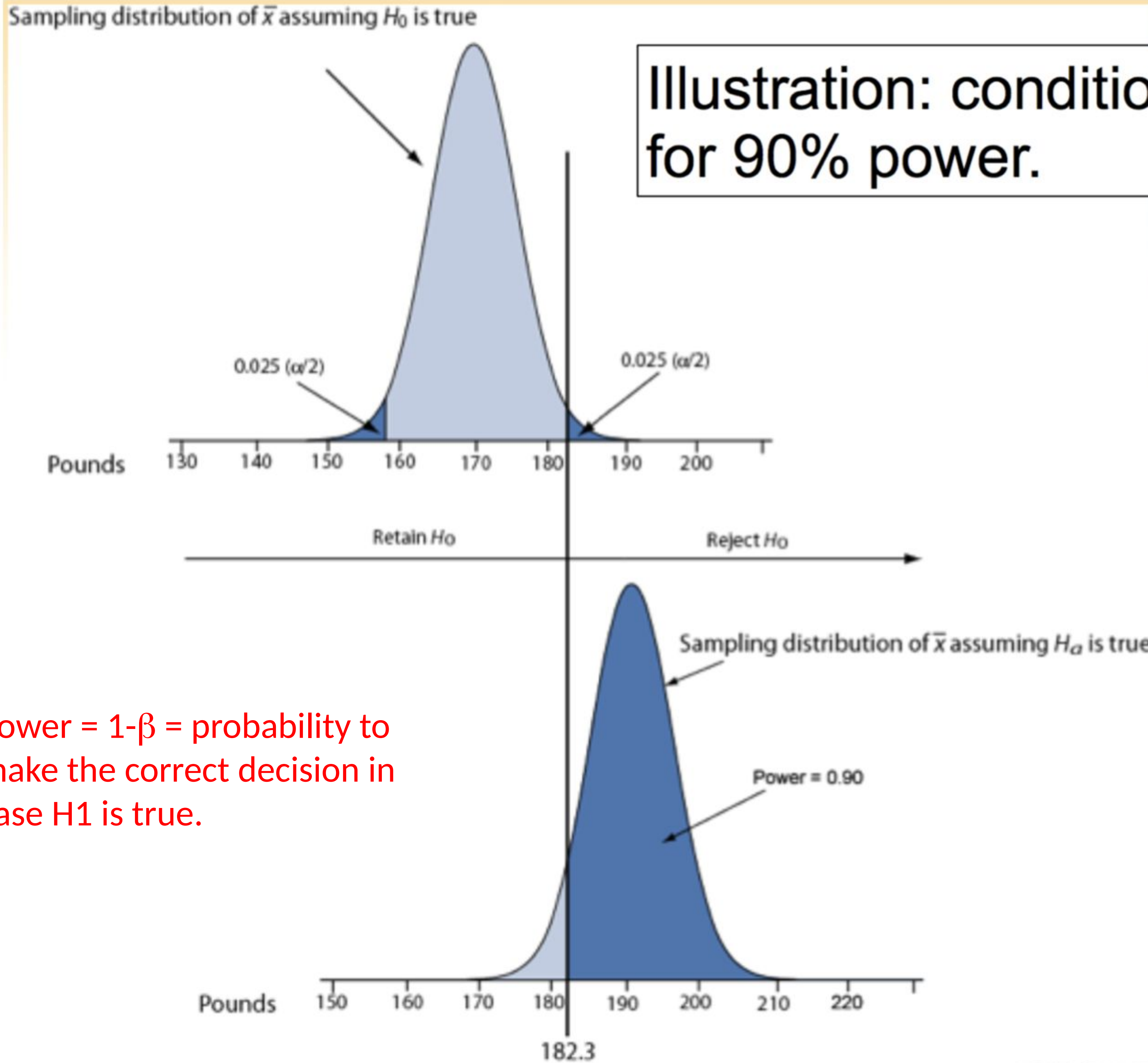
How large a sample is needed for a one-sample z test with 90% power and $\alpha = 0.05$ (two-tailed) when $\sigma = 40$? Let $H_0 : \mu = 170$ and $H_1 : \mu = 190$ (thus, $\Delta = \mu_0 - \mu_1 = 170 - /190 = -20$)

$$n = \frac{\sigma^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{2} = \frac{40^2 (1.28 + 1.96)^2}{(-20)^2} = 41.99$$

Round up to 42 to ensure adequate power.

to be compared with the power in case of a sample n=16

Illustration: conditions for 90% power.



Power = 1-β = probability to make the correct decision in case H1 is true.

Factors Affecting Power

Suppose a math achievement test were known to be normally distributed with a mean of 75 and a *standard deviation* of σ.

A researcher is interested in whether a new method of teaching results in a higher mean. Assume that **the population mean μ for the new method is larger than 75**.

The researcher plans to sample N subjects and do a one-tailed test of whether the sample mean is significantly higher than 75.

We consider factors that affect the probability that the researcher will **correctly reject the false *null hypothesis* that the population mean is 75.**

In other words, factors that affect power.

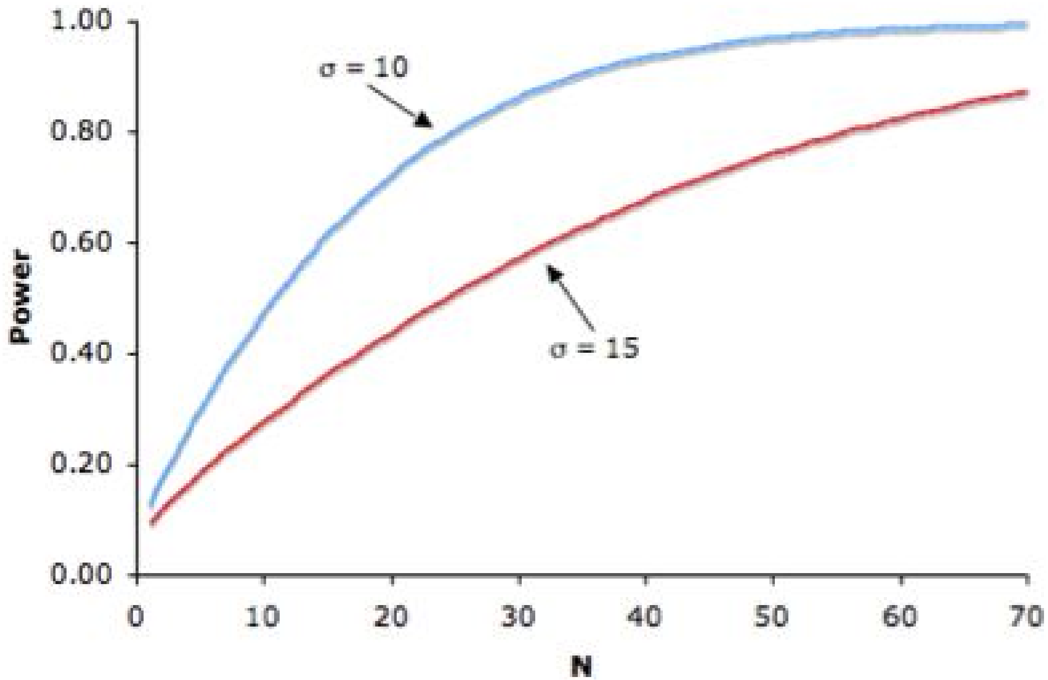


Figure 1. The relationship between sample size and power for H0: μ = 75, real μ = 80, one-tailed α = 0.05, for σ's of 10 and 15.

Sample Size

Figure 1 shows that the larger the sample size, the higher the power. Since sample size is typically under an experimenter's control, increasing sample size is one way to increase power. However, it is sometimes difficult and/or expensive to use a large sample size.

Standard Deviation

Figure 1 also shows that power is higher when the standard deviation is small than than for the standard deviation of 15 (except, of course, when N = 0). Experimenter can reduce the standard deviation by sampling by a homogenoeous population of subjects.

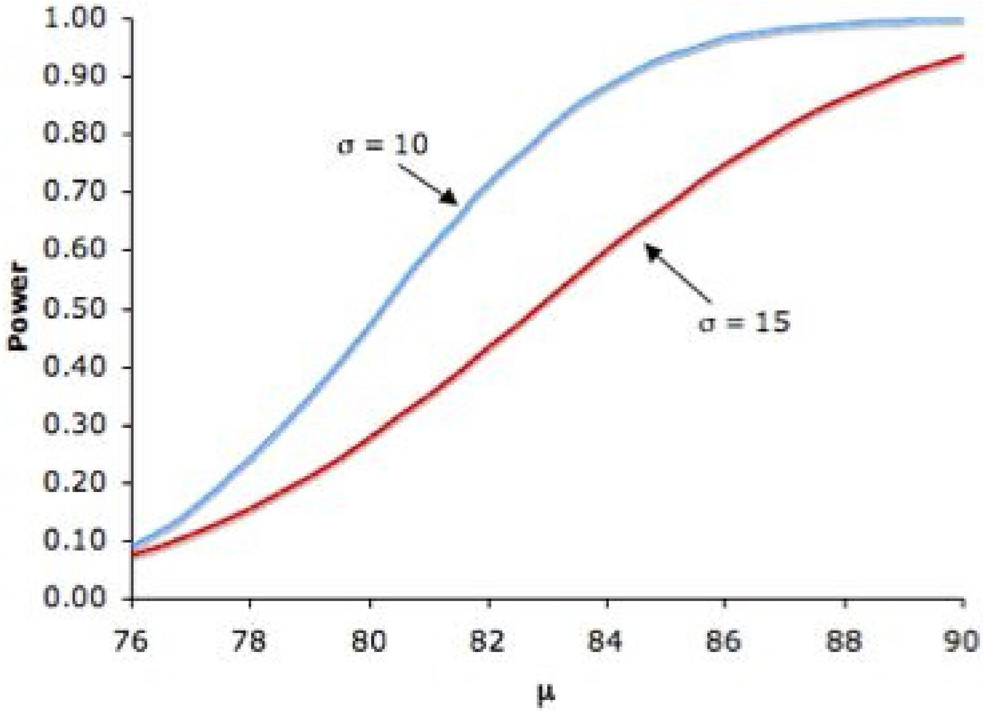


Figure 2. The relationship between μ and power for H_0: μ = 75, one-tailed α = 0.05, for σ's of 10 and 15.

Difference between Hypothesized and true mean

Naturally, the larger the effect size, the more likely it is that an experiment would find a significant effect. Figure 2 shows the effect of increasing the difference between the mean specified by the null hypothesis (75) and the population mean μ for standard deviations of 10 and 15.

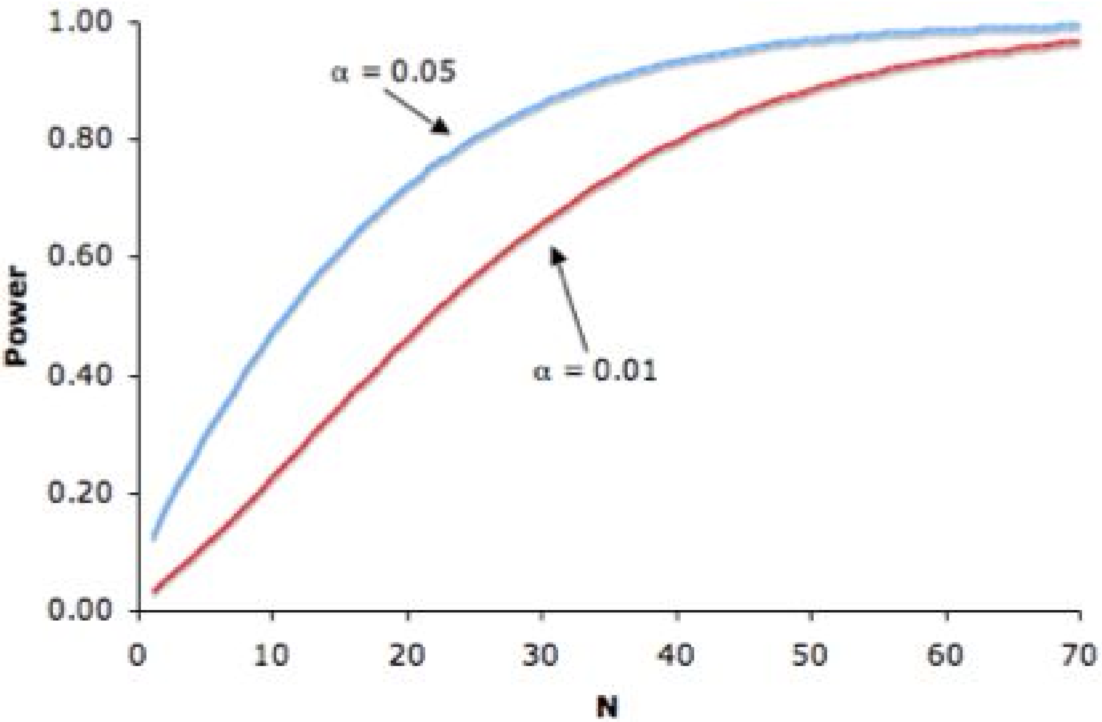


Figure 3. The relationship between significance level and power with one-tailed tests: $\mu = 75$, real $\mu = 80$, and $\sigma = 10$.

Significance Level

There is a trade-off between the *significance level* and power: the more stringent (lower) the significance level, the lower the power.

Figure 3 shows that power is lower for the 0.01 level than it is for the 0.05 level. Naturally, the stronger the evidence needed to reject the null hypothesis, the lower the chance that the null hypothesis will be rejected.

One- versus Two-Tailed Tests

Power is higher with a *one-tailed* test than with a *two-tailed* test as long as the hypothesized direction is correct. A one-tailed test at the 0.05 level has the same power as a two-tailed test at the 0.10 level. A one-tailed test, in effect, raises the significance level.

Threats for your experiments

Type 1 error (α)

Maintaining Type 1 error means doing all we can to assure that the **false positive** rate really is set to whatever nominal level (usually 5%) we have chosen.

It basically involves choosing an appropriate statistical procedure and assuring that the assumptions of our chosen procedure are reasonably met.

Using models with broken assumptions tend to result in an increased chance of making false claims in the presence of ineffective treatments.

Power (1-β)

If some particular true difference in outcomes is caused by the active treatment, and you have low power to detect that difference, you will probably make a Type 2 error (have a "**false negative**" result) in which you conclude that the treatment was ineffective, when it really was effective

Studying sufficient numbers of subjects is the most well known way to assure sufficient power.

POWER ANALYSIS with R

Overview

Power analysis is an important aspect of experimental design. It allows us to determine the sample size required to detect an effect of a given size with a given degree of confidence.
Conversely, it allows us to determine the probability of detecting an effect of a given size with a given level of confidence, under sample size constraints.
If the probability is unacceptably low, we would be wise to alter or abandon the experiment.
The following **four quantities** have an intimate relationship:

1. sample size
2. effect size
3. significance level = P(Type I error) = probability of finding an effect that is not there
4. power = 1 - P(Type II error) = probability of finding an effect that is there

Given any three, we can determine the fourth.

Power Analysis in R

The pwr package develped by Stéphane Champely, impliments power analysis as outlined by Cohen (1988).
For each of these functions, you enter three of the four quantites (effect size, sample size, significance level, power) and the fourth is calculated.
The significance level defaults to 0.05. Therefore, to calculate the significance level, given an effect size, sample size, and power, use the option "sig.level=NULL".

function	power calculations for
pwr.2p.test	two proportions (equal n)
pwr.2p2n.test	two proportions (unequal n)
pwr.anova.test	balanced one way ANOVA
pwr.chisq.test	chi-square test
pwr.f2.test	general linear model
pwr.p.test	proportion (one sample)
pwr.r.test	correlation
pwr.t.test	t-tests (one sample, 2 sample, paired)
pwr.t2n.test	t-test (two samples with unequal n)

Power Analysis in R

Specifying an effect size can be a daunting task.
ES formulas and Cohen's suggestions (based on social science research) are provided below.
Cohen's suggestions should only be seen as very rough guidelines. Your own subject matter experience should be brought to bear.

Test	Effect size	Small	Medium	Large
Difference between two means	d	0.20	0.50	0.80
Difference between many means	*f*	0.10	0.25	0.40
Chi-squared test	*w*	0.10	0.30	0.50
Pearson's correlation coefficient	*ρ*	0.10	0.30	0.50

Table 1 Small, medium and large effect sizes as defined by Cohen

t-tests

For t-tests, use the following functions:

pwr.t.test(n = , d = , sig.level = , power = , type = c("two.sample", "one.sample", "paired"))

where n is the sample size, d is the effect size, and type indicates a two-sample t-test, one-sample t-test or paired t-test.

If you have unequal sample sizes, use

pwr.t2n.test(n1 = , n2= , d = , sig.level =, power =)

where n1 and n2 are the sample sizes.

For t-tests, the effect size is assessed as

Cohen suggests that d values of 0.2, 0.5, and 0.8 represent small, medium, and large effect sizes respectively.

You can specify alternative="two.sided", "less", or "greater" to indicate a two-tailed, or one-tailed test. A two tailed test is the default.

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

where μ_1 = mean of group 1
μ_2 = mean of group 2
σ^2 = common error variance

pwr.t.test(n = , d = , sig.level = , power = , type = c("two.sample", "one.sample", "paired"))

1) Calculate power

```
library(pwr)
pwr.t.test(n=25,d=0.75,sig.level=.01,alternative="greater")
```

Output

```
     Two-sample t test power calculation

              n = 25
              d = 0.75
      sig.level = 0.01
          power = 0.5988572
    alternative = greater
```

NOTE: n is number in *each* group

pwr.t.test(n = , d = , sig.level = , power = , type = c("two.sample", "one.sample", "paired"))

2)

```
# Using a two-tailed test proportions, and assuming a
# significance level of 0.01 and a common sample size of
# 30 for each proportion, what effect size can be
detected
# with a power of .75?
```

pwr.t.test(n=30, power = 0.75 ,sig.level=.01,alternative="two.sided")

Output

Two-sample t test power calculation

```
          n = 30
          d = 0.8640596
  sig.level = 0.01
      power = 0.75
alternative = two.sided
```

NOTE: n is number in *each* group

```
# For a one-way ANOVA comparing 5 groups, calculate the
# sample size needed in each group to obtain a power of
# 0.80, when the effect size is moderate (0.25) and a
# significance level of 0.05 is employed.

pwr.anova.test(k=5,f=.25,sig.level=.05,power=.8)

     Balanced one-way analysis of variance power calculation

              k = 5
              n = 39.1534
              f = 0.25
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

pwr.chisq.test - Goodness of fit test

(*From Cohen*, *example 7.1*) A market researcher is seeking to determine preference among 4 package designs. He arranges to have a panel of 100 consumers rate their favorite package design. He wants to perform a chi-square goodness of fit test against the null of equal preference (25% for each design) with a significance level of 0.05. What's the power of the test if 3/8 of the population actually prefers one of the designs and the remaining 5/8 are split over the other 3 designs?

We use the ES.w1 function to calculate effect size. To do so, we need to create vectors of null and alternative proportions:

pwr.chisq.test - Goodness of fit test

```
> null <- rep(0.25, 4)
> alt <- c(3/8, rep((5/8)/3, 3))
> ES.w1(null,alt)
[1] 0.2886751

> pwr.chisq.test(w=ES.w1(null,alt), N=100,
df=(4-1), sig.level=0.05)

     Chi squared power calculation

              w = 0.2886751
              N = 100
             df = 3
      sig.level = 0.05
          power = 0.6739834
```

NOTE: N is the number of observations

If our estimated effect size is correct, we only have about a 67% chance of finding it (i.e., rejecting the null hypothesis of equal preference). How many subjects do we need to achieve 80% power?

```
pwr.chisq.test(w=ES.w1(null,alt), df=(4-1), power=0.8, sig.level = 0.05)
## ## Chi squared power calculation
## ## w = 0.2886751
## N = 130.8308
## df = 3 ## sig.level = 0.05
## power = 0.8
## ## NOTE: N is the number of observations
```

If our alternative hypothesis is correct then we need to survey at least 131 people to detect it with 80% power.

pwr.chisq.test - test of association

We want to see if there's an association between gender and flossing teeth among college students. We randomly sample 100 students (male and female) and ask whether or not they floss daily. We want to carry out a chi-square test of association to determine if there's an association between these two variables.
We set our significance level to 0.01.
To determine effect size we need to propose an alternative hypothesis, which in this case is a table of proportions. We propose the following:

We use the `ES.w2` function to calculate effect size for chi-square tests of association

```
prob <- matrix(c(0.1,0.2,0.4,0.3), ncol=2,
               dimnames = list(c("M","F"),c("Floss","No Floss")))
prob
```

```
##   Floss No Floss
## M   0.1      0.4
## F   0.2      0.3
```

This says we sample even proportions of male and females, but believe 10% more females floss.

Now use the matrix to calculate effect size:

```
ES.w2(prob)
```

```
## [1] 0.2182179
```

We also need degrees of freedom. `df` = (2 - 1) * (2 - 1) = 1

And now to calculate power:

```
pwr.chisq.test(w = ES.w2(prob), N = 100, df = 1, sig.level = 0.01)
```

```
##
##      Chi squared power calculation
##
##               w = 0.2182179
##               N = 100
##              df = 1
##       sig.level = 0.01
##           power = 0.3469206
##
## NOTE: N is the number of observations
```

https://www.statmethods.net/stats/power.html