



THE UNIVERSITY OF
ALABAMA AT BIRMINGHAM



Design of Experiment in Metabolomics

Hemant K. Tiwari, Ph.D.
Professor

Department of Biostatistics
School of Public Health

Key Stages of a Metabolomics Study



Design of Experiments/ Experimental Design

- A controlled experiment to either test a hypothesis or generate hypotheses
- In the design of experiments, the experimenter is usually interested in the effect of some process or intervention on some subjects

History of Experimental Design

- In 1747, James Lind (a Scottish Physician) developed the theory that citrus fruits cured **scurvy** (Symptoms include bleeding sores, tooth loss, anemia, and a reduced rate of healing for injuries), while serving as surgeon on HMS Salisbury Ship of the Royal Navy. This was the first ever clinical trial conducted.
- It can be fatal if left untreated.
- Scurvy is disease resulting from a deficiency of Vitamin C.

Lind's Experiment

- Lind selected 12 men from the ship suffering from scurvy. He divided them into six pairs, giving each pair different supplements to their basic diet for two weeks. The treatments were all remedies that had been proposed:
 - A quart of cider every day
 - Twenty five drops of *elixir vitriol* (sulphuric acid) three times a day upon an empty stomach
 - One half-pint of seawater every day
 - A mixture of garlic, mustard, and horseradish in a lump the size of a nutmeg
 - Two spoonful of vinegar three times a day
 - Two oranges and one lemon every day

Result: The men given citrus fruits recovered dramatically within a week.

Lading Tasting Tea Experiment

Design of experiments was born as a result of an unlikely, but true anecdote: A lady claimed before R.A. Fisher that she was able to ascertain whether milk was poured before or after tea in her cup of tea. Fisher devised a study to verify her claim and, in turn, this gave birth to Experimental Design.



There are 70 different outcomes: $\frac{8!}{4!4!} = 70$

Her answers:

		True order		Total
		Tea First	Milk First	
Lady's Guesses	Tea First	a=3	b=1	a+b=4
	Milk First	c=1	d=3	c+d=4
Total		a+c=4	b+d=4	N=8

Experimental Design: Define the Problem

- What is the topic?
- What is the good question for an experiment?
- Is your question testable with the materials in your hand?
- Need to know hypothesis to guide your experiment?
- Design your experiment that will test your hypothesis

Experimental Design

- It is essential to make sure that samples reflect and represent the biological question under investigation
- Examine the most influential factors that are relevant to your hypothesis under investigation
- External factors have to be eliminated or identified, so they can be modeled during the analysis

Aims of experimental design

- To provide answers to research hypothesis or generate hypothesis:
 - Minimize the biological variance and technical or experimental variance since metabolome can change very rapidly in response to subtle changes in environment

Main Aims of the Experimental Design

- Maximize the Systematic/ experimental variance of the variable(s) of the research hypothesis (i.e. maximize the difference in the dependent variable (outcome) caused by maximizing the differences in the independent variable (treatment).
- Control the variance of extraneous (unwanted) variables that may affect the outcome other than treatment that could be causing differences in the outcome.
- Minimize the random variance/error due to unreliable measurement instruments that have high error of measurement.

Control for Extraneous Variable

- Eliminate the variable (for example if sex effect exist, then include only males or females, i.e. stratify).
- Randomization
- Build it into design
- Match subjects

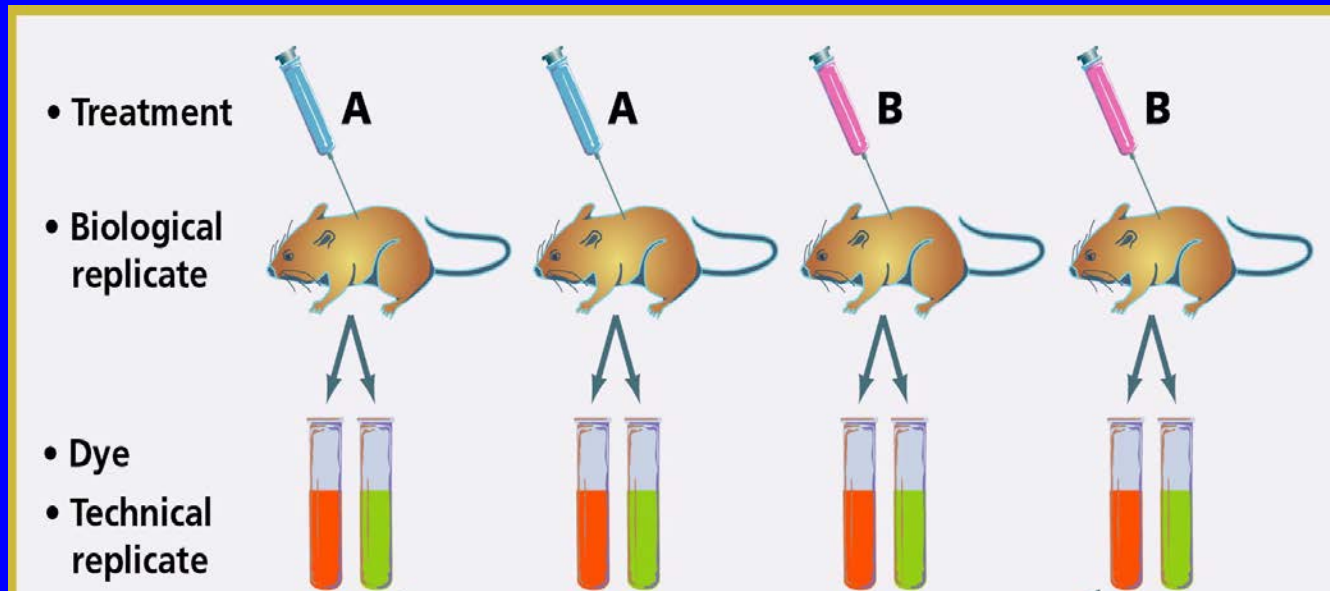
General Statistical Principles of Experimental Design

- Replication
- Randomization
- Blocking (Stratification)
- Use of factorial experiments instead of the one-factor-at-a-time methods

Replication

- Replication – repetition of a basic experiment without changing any factor settings, allows the experimenter to estimate the experimental error in the system used to determine whether observed differences in the data are “real” or “just noise”, allows the experimenter to obtain more statistical power (ability to identify small effects)
- Replications should not be confused with repeated measurements which refer to taking several measurements of a single occurrence of a phenomenon (single experiment).

Replications should not be confused with repeated measurements.



Replicates

Repeated measure

Another Example

- Number of replicates matter in power of the analysis
- Experiment: one mouse per group (treatment group vs. untreated group)- you can only measure the difference in metabolites, but no variance
- 5 or 10 mouse per group- you can measure both the difference in metabolites and the variance (very important for statistical testing)

More terms saying the same things

- What to replicate?
 - Biological replicates (replicates at the experimental unit level, e.g. mouse, plant, pot of plants...)
 - Experimental unit is the unit that the **experiment treatment or condition is directly applied** to, e.g. a plant if hormone is sprayed to individual plants; a pot of seedlings if different fertilizers are applied to different pots.
 - Technical replicates
 - Any replicates below the experimental unit, e.g. different leaves from the same plant sprayed with one hormone level; different seedlings from the same pot; Different aliquots of the same RNA extraction; multiple arrays hybridized to the same RNA; multiple spots on the same array.

Why Replicate?

- Reduce the effect of uncontrolled variation (i.e., increase precision).
- Quantify uncertainty.

Randomization

- Randomization – a statistical tool used to minimize potential uncontrollable factors “lurking variables” (which might vary over the length of the experiment) in the experiment by randomly assigning treatment to the experimental units.
- Results in “averaging out” the effects of the extraneous factors that may be present in order to minimize the risk of these factors affecting the experimental results.
- Randomization is essential for making causal inferences.

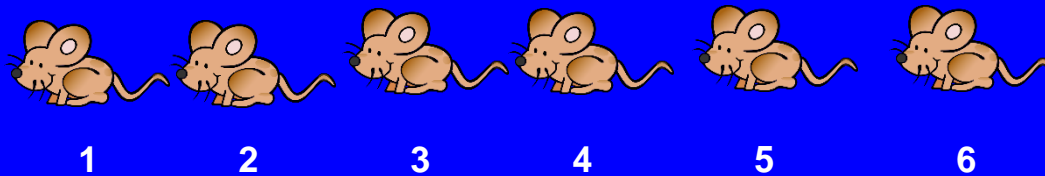
Randomization

- Experimental units (people, mouse, plant etc.) should be assigned to treatment groups **at random**.
- Can be done by using
 - Computer
 - Coins

Example

- Number the objects to be randomized and then randomly draw the numbers.

Example: Assign treatment/Special Diet and no treatment (control) to 6 mice (3 each)



Special Diet/treatment : 1, 3, 4
Control : 2, 5, 6

Blocking/ Stratification

- Blocking – technique used to increase the precision of an experiment by breaking the experiment into homogeneous segments (blocks) in order to control any potential block to block variability (e.g. measurement of metabolites in different days or shift, by different technicians, by different machines). Any effects on the experimental results as a result of the blocking factor will be identified and minimized.

Blocking

- If you anticipate a difference in measurement of metabolites in different days or shift, by different technicians, by different machines:
 - Ensure that within each period, there are equal numbers of subjects in each treatment group.
 - Take account of the difference between periods in your analysis.

Blocking

- Some of these identified uninteresting but varying factors can be controlled through blocking.
 - COMPLETELY RANDOMIZED DESIGN
 - COMPLETE BLOCK DESIGN
 - INCOMPLETELY BLOCK DESIGNS

Completely Randomized Design

There is no blocking

➔ Example

- ◆ Compare two hormone treatments (trt and control) using 6 Arabidopsis plants (or mice or human).



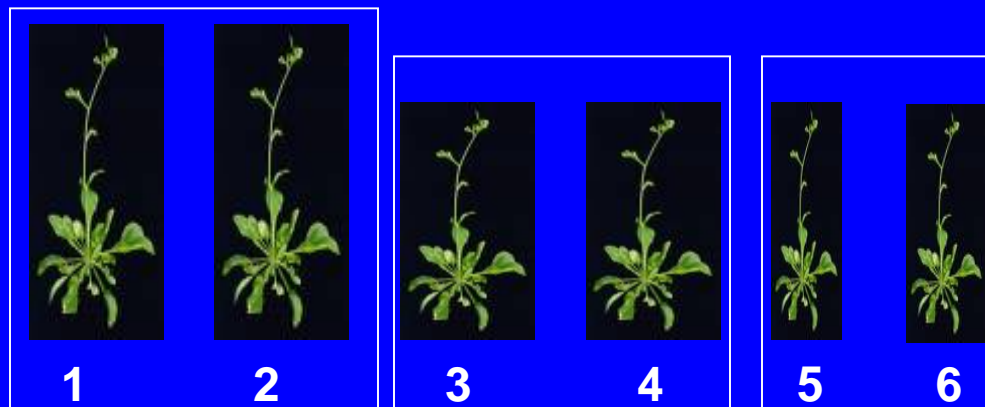
Hormone trt: (1,3,4); (1,2,6)
Control : (2,5,6); (3,4,5)

Complete Block Design

- ➔ There is blocking and the block size is equal to the number of treatments.

Example:

- ◆ Compare two hormone treatments (trt and control) using 6 Arabidopsis plants. For some reason plant 1 and 2 are taller, plant 5 and 6 are thinner.



Hormone treatment: (1,4,5) ; (1,3,6)

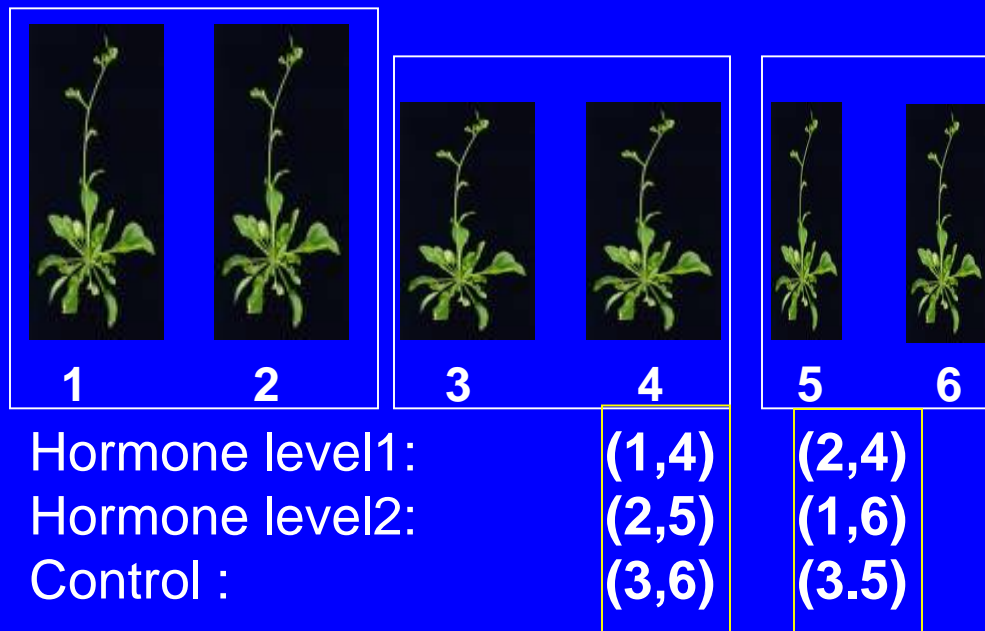
Control : (2,3,6) ; (2,4,5)

⇒ Randomization within blocks

Incomplete Block Design

➔ There is blocking and the block size is smaller than the number of treatments. You can assign all treatments in each block.

Example: Compare three hormone treatments (hormone level 1, hormone level 2, and control) using 6 Arabidopsis plants. For some reason plant 1 and 2 are taller, plant 5 and 6 are thinner.



⇒ Randomization within blocks

Another Example

Experiment:

- 32 mice (16 males and 16 females)
- Half to be treated and other half left untreated
- A technician can work only 4 mice per day and only on Monday through Thursday


Bad design (Cardinal Sin)

Week 1

Mon	Tue	Wed	Thr
Trt	Trt	Trt	Trt
Trt	Trt	Trt	Trt
Trt	Trt	Trt	Trt
Trt	Trt	Trt	Trt

Week 2

Mon	Tue	Wed	Thr
Cntl	Cntl	Cntl	Cntl
Cntl	Cntl	Cntl	Cntl
Cntl	Cntl	Cntl	Cntl
Cntl	Cntl	Cntl	Cntl

 = females

 = males


Randomization


Week 1

Mon	Tue	Wed	Thr
Trt	Trt	Trt	Trt
Cntl	Trt	Trt	Cntl
Trt	Cntl	Trt	Cntl
Cntl	Cntl	Cntl	Trt

Week 2

Mon	Tue	Wed	Thr
Cntl	Trt	Trt	Cntl
Cntl	Cntl	Cntl	Trt
Cntl	Trt	Cntl	Trt
Trt	Cntl	Trt	Cntl

 = females

 = males


Blocking


Week 1

Mon	Tue	Wed	Thr
Trt	Trt	Trt	Cntl
Cntl	Cntl	Cntl	Trt
Cntl	Cntl	Trt	Cntl
Trt	Trt	Cntl	Trt

Week 2

Mon	Tue	Wed	Thr
Trt	Cntl	Trt	Cntl
Cntl	Trt	Trt	Cntl
Trt	Cntl	Cntl	Trt
Cntl	Trt	Cntl	Trt

 = females

 = males

Confounding

- Confounding - A concept that basically means that multiple effects are tied together into one parent effect and cannot be separated. For example,
 - Two people flipping two different coins would result in the effect of the person and the effect of the coin to be confounded
 - As experiments get large, higher order interactions are confounded with lower order interactions or main effect.

Factorial Experiments

- A *factor* is a discrete variable used to classify experimental units.
- For example, "Gender" might be a factor with two levels "male" and "female" and "Diet" might be a factor with three levels "low", "medium" and "high" protein. The levels within each factor can be discrete, such as "Drug A" and "Drug B", or they may be quantitative such as 0, 5, 10, 15, 20, 25, and 30 mg/kg.

Factorial Experiments

- A *factorial design* is one involving two or more factors in a single experiment. Such designs are classified by the number of levels of each factor and the number of factors.
- Example: A 2x2 factorial will have two levels or two factors and a 2x3 factorial will have three factors each at two levels.

Factorial Experiments

- Typically, there are many factors such as sex, genotype, diet, blood pressure, socio economic status, and age which can influence the outcome of an experiment. These often need to be investigated in order to determine the *generality* of a response.
- It may be important to know whether a response is only seen in, say, females but not males or in particular strain of mice. One way to do this would be to do separate experiments in each sex. This “One Variable at A Time” approach is very wasteful of scientific resources. A much better alternative is to include both sexes or more than one strain etc. in a single “factorial” experiment. Such designs can include several factors without using excessive numbers of experimental subjects.
- Factorial designs are efficient and provide extra information (the interactions between the factors), which can not be obtained when using single factor designs.

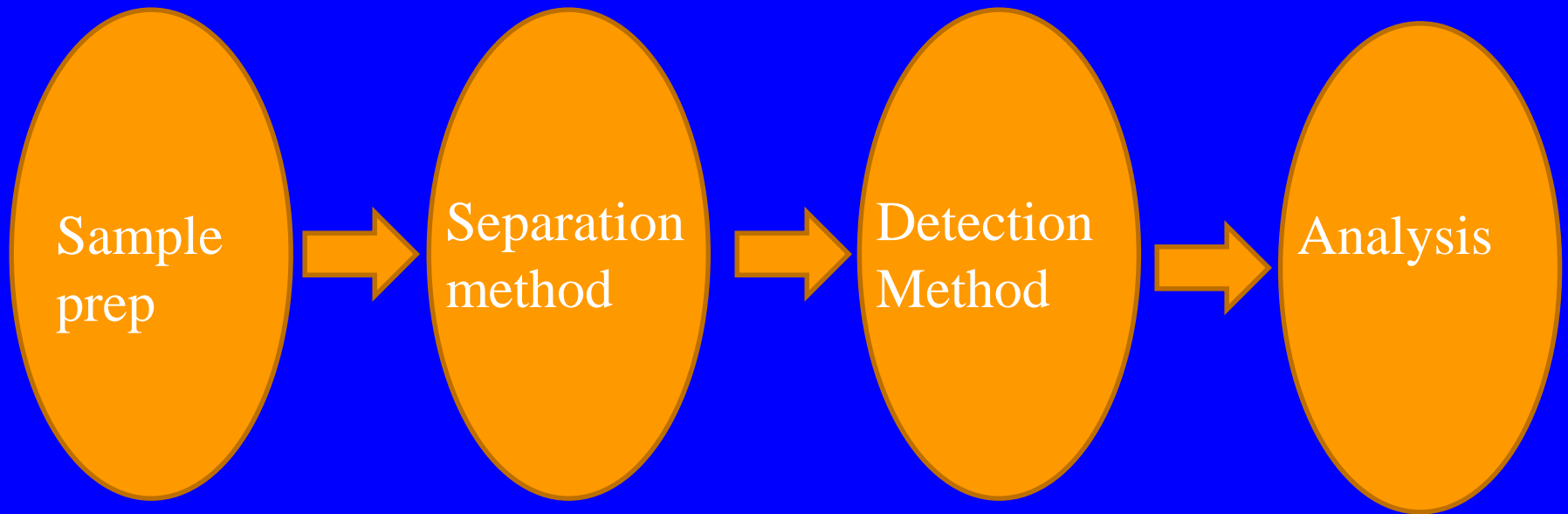
RA Fisher (1960)

“If the investigator confines his attention to any single factor we may infer either that he is the unfortunate victim of a doctrinaire theory as to how experimentation should proceed, or that the time, material or equipment at his disposal is too limited to allow him to give attention to more than one aspect of his problem.....

.... Indeed in a wide class of cases (by using factorial designs) an experimental investigation, at the same time as it is made more comprehensive, may also be made more efficient if by more efficient we mean that more knowledge and a higher degree of precision are obtainable by the same number of observations.”

(Fisher RA. 1960. The design of experiments. New York: Hafner Publishing Company, Inc. 248 p.)

Steps in metabolomics experiment



At each step, the errors can occur

Aim of the experiment is to minimize the errors

Aims of experimental design

- Minimize the variance of extraneous variable variance that may have impact on outcome variables
 - Examples: Metabolites are sensitive to environmental influences such as sample storage
 - Temperature
 - Aliquoting
 - Biobanking

Measurement Issues in metabolomics

- To get the sample ready
 - The sample is prepped and put onto wells on a silicon plate
 - Each well's aliquot is subjected to gas and/or liquid chromatography
 - After separation, the sample goes to a mass spectrometry

Measurement Issues

- Sources of errors at the prep stage
 - Within subject variation
 - Within tissue variation
 - Contamination by cleaning solvents
 - Calibrate uncertainty
 - Evaporation of volatiles

Errors at separation method

- Gas chromatography (GC) (GC creates ionized aerosol, each droplet evaporates to a single ion and this is separated by mass in the column, then ejected to the spectrometer):
 - Imperfect evaporation
 - Adhesion in the column
 - Ion fragmentation

Measurement Issues at Detection

- Mass spectrometry (MS) is used to identify and quantify metabolites after separation by GC, HPLC (LC-MS)
 - Baseline removal
 - denoising
 - Peak detection (the process of distinguishing peptide and noise peaks)
 - Normalization

Sample Size and Power

- Purpose
 - Planning a study: number of individuals to recruit or number of mice to test a research hypothesis
 - Understand sample size implications of alternative study designs
 - Sample was already collected and wants study using new technology
 - GWAS was done, but wants to do metabolomics
GWAS

Sample Size and Power Calculation

- Often the number of samples to be used for the experiments dictated by the reality of resources available, not science.
 - How much money is available for the experiment
 - What is the cost per sample
 - Thus, sample size = \$ available / cost per sample

Hypothesis Testing

- Power calculations are based on the principles of hypothesis testing
- A hypothesis is a statement about population parameter
- The two complementary hypotheses in a hypothesis testing problem are called the null hypothesis (H_0) and alternate hypothesis (H_1)

Two types of errors in hypothesis testing

$H_0: \theta=0$ versus $H_1: \theta \neq 0$

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct Decision	Type I error (α)
	H_1	Type II error (β)	Correct decision

Reject H_0 when we should not

Don't reject H_0
when we should

- Type I Error: Probability of finding a statistically significant effect when the truth is that there is no effect
- Type II Error: Probability of not finding a statistically significant effect when there is none.
- Power = $1 - \beta$, Significance level = α
- Goal is to minimize both types of errors

Power depends on ...

- Design
- The method of analyzing the data
- The effect size
- Standard deviation of the effect of interest
- Measurement variability
- The chosen significance level (α)
- The sample size

We usually use significance level of 5% and 80% power to estimate the sample size

Factors Affecting the sample size

Effect size	↑	Required sample size	↓
Variation of data	↑	Required sample size	↑
Type I error rate	↓	Required sample size	↑
Power	↑	Required sample size	↑

- Type I error rate (α) is kept fixed and becomes smaller as number of tests increase
- Effect size and variation of the data (σ^2) is either obtained through pilot study or vary to calculate different sample sizes.

To calculate Sample Size

- Need to know level of significance (α)
- Statistical power ($1 - \beta$)
- Effect size (expected difference)
- Standard deviation
- What statistical test we are going to use

Sample Size Formula for difference in means

- A sample size formula to test difference of means between two groups (two-tailed test)

$$- n_1 = ((r + 1)\sigma^2 (Z_{1-\beta} + Z_{\alpha/2})^2) / (r \Delta^2)$$

where,

- n_1 = size of the smaller group
- r = ratio of larger group to smaller group
- $Z_{1-\beta}$ = standard normal deviate corresponds to $1-\beta$
- $Z_{\alpha/2}$ = standard normal deviate corresponds to two-tailed significance level
- Δ = difference in means of the outcome
- σ^2 = Variance of the difference of the means

Common standard normal deviates Z_α and Z_β

α or β	One-sided test Z_α or Z_β	Two-sided test Z_α
0.001	3.09	3.29
0.005	2.58	2.81
0.01	2.33	2.58
0.025	1.96	2.24
0.05	1.64	1.96
0.10	1.28	1.64
0.20	0.84	1.28

Simple Example

- How many people would you need to sample in each group (assuming both groups of equal size) to achieve power of 80% if $SD = \sigma = 10$, difference in mean is 5 with fixed $\alpha = 0.05$. So $Z_{\alpha/2} = Z_{0.025} = 1.96$, $Z_{1-\beta} = 0.84$, $1-\beta = 0.80$, so $\beta = 0.20$ and $r = 1$, then

- $$\begin{aligned} n_1 &= ((r + 1)\sigma^2 (Z_{\beta} + Z_{\alpha/2})^2) / (r \Delta^2) \\ &= 2 (100) (.84 + 1.96)^2 / (5^2) \\ &= 62.72 \sim 63 \end{aligned}$$

63 per group implies 126 altogether.

Sample Size

- Two-tailed test: When the investigator is interested in determining whether a treatment A is different from a treatment B
- Usually a 2 tailed test is performed with the risk of making a Type-1 error set at $\alpha / 2$ in each tail.
- For a 2 tailed test at $\alpha = .05$ and equal allocation of type-1 error to each tail $Z_{\alpha/2} = 1.96$

Sample Size

Sometimes an investigator is only interested in a difference between treatments in one direction.

This is appropriate when The hypothesis is to test whether a treatment A is better than treatment B

For a 1 tailed test at $\alpha = .05$ $Z_{\alpha} = 1.65$

Conclusions

- Brainstorm with your colleagues and senior faculty to decide on the experiment
- Experiment should be designed with consultation with the statistician and metabolomics assays provider
- Good design and good analytic methods can lead to reduced sample size and also lead to valid meaningful results

Advice from Sir Ronald Aylmer Fisher (17 Feb 1890 - 29 Jul 1962)

- *“To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.”*

