Rigorous Science - Based on a probability value?

The linkage between Popperian science and statistical analysis

The Philosophy of science: the scientific Method - from a Popperian perspective

Philosophy → How we understand the world

↓

Science → How we expand that understanding

↓

Design → How we implement science

Arguments over how we understand and expand our understanding are the basis of debates over how science has been, is and should be done

The Philosophy of science: the scientific Method - from a Popperian perspective

Terms:

1. **Science** - A method for understanding rules of assembly or organization
   a) **Problem**: How do we, (should we) make progress in science
2. **Theory** - a set of ideas formulated to explain something
3. **Hypothesis** - supposition or conjecture (prediction) put forward to account for certain facts, used as a basis for further investigations
4. **Induction or inductive reasoning** - reasoning that general (universal) laws exist because particular cases that seem to be examples of those laws also exist
5. **Deduction or deductive reasoning** - reasoning that something must be true because it is a particular case of a general (universal) law

Induction →

← Deduction

Particular General

The Scientific Method - from a Popperian perspective

Extreme example

1. Induction
   "Every swan I have seen is white, therefore all swans are white"
2. Deduction
   "All swans are white, the next one I see will be white"

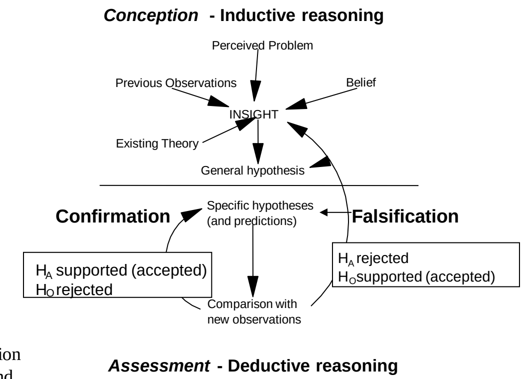
Compare these statements:

1. Which can be put into the form of a testable hypothesis? (eg. prediction, if - then statement)
2. Which is closer to how we operate in the world?
3. Which type of reasoning is most repeatable?

Is there a difference between ordinary understanding and scientific understanding (should there be?)

The Scientific Method - from a Popperian perspective
Hypothetico - deductive method

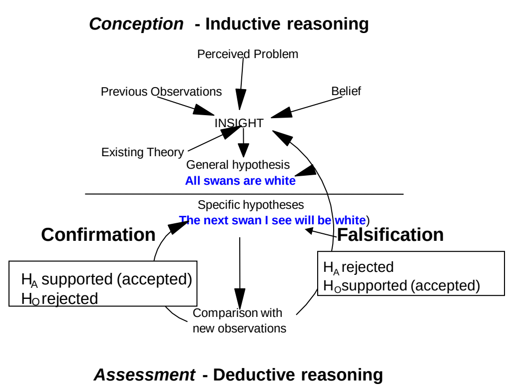
1. Conception - Inductive reasoning
   a. Observations
   b. Theory
   c. Problem
   d. Regulation
   e. Belief
2. Leads to Insight and a General Hypothesis
3. Assessment is done by
   a. Formulating Specific hypotheses
   b. Comparison with new observations
4. Which leads to:
   a. **Falsification** - and rejection of insight, and specific and general hypotheses, or
   b. **Confirmation** - and retesting of ***alternative hypotheses***



The Scientific Method - from a Popperian perspective
Hypothetico - deductive method

Questions and Notes

1. Is there any provision for accepting the insight or working hypothesis?
2. Propositions not subject to rejection by contrary observations are not "scientific"
3. Confirmation does not end hypothesis testing - new hypotheses should always be put forth for a particular observation, theory, belief…
4. In practice but rarely reported, alternatives are tested until only one (or a few) are left (not rejected). Then we say things like: *suggest, indicates, is evidence for*
5. Why is there no provision for accepting theory or working hypotheses?

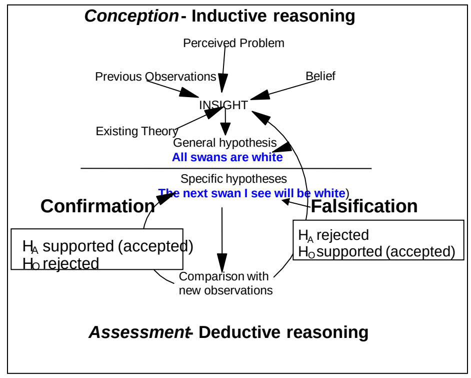


a) Because it is easy to find confirmatory observations for almost any hypothesis, but one negative result refutes it absolutely (**this assumes test was adequate - the quality of falsification is important**)

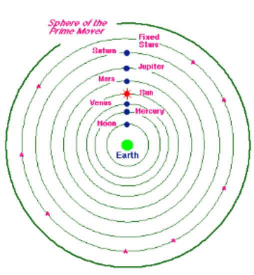
The Scientific Method - from a Popperian perspective
Hypothetico - deductive method

Considerations - problems with the Popperian hypothetico -deductive approach)

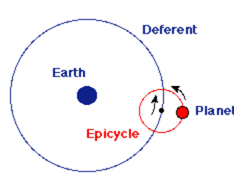
1) This type of normal science may rarely lead to revolutions in Science (Kuhn)
   - A) Falsification science leads to paradigms - essentially a way of doing and understanding science that has followers
   - B) Paradigms have momentum - mainly driven by tradition, infrastructure and psychology
   - C) Evidence against accepted theory is considered to be exceptions (that prove the rule)
   - D) Only major crises lead to scientific revolutions
     - 1) paradigms collapse from weight of exceptions - normal science - crisis - revolution - normal science



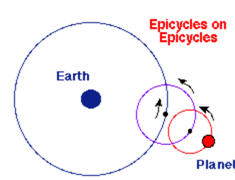
1. The paradigm: The earth must be the center of the universe – 350 BC
2. Exceptions are explained- Ptolemaic universe
   a) All motion in the heavens is uniform circular motion.
   b) The objects in the heavens are made from perfect material, and cannot change their intrinsic properties (e.g., their brightness).
   c) The Earth is at the center of the Universe.
3. Paradigm nears scientific collapse
4. Religion Intervenes – middle ages



1



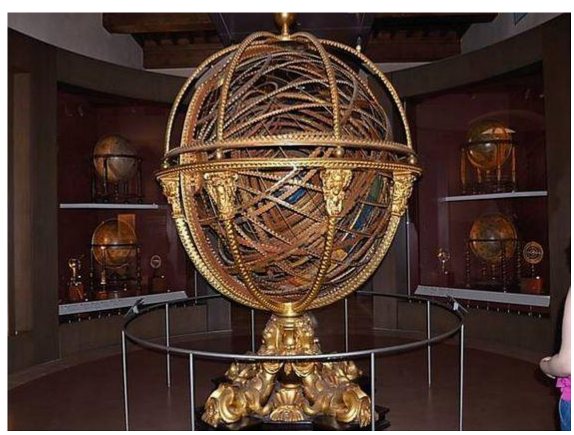
2



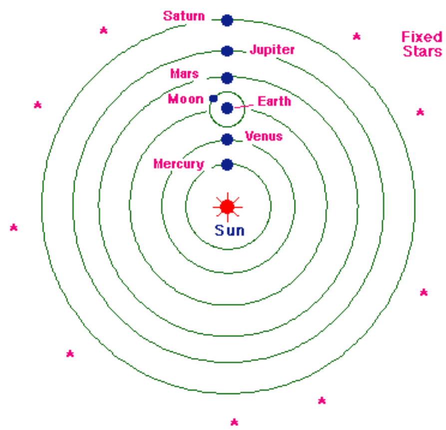
3

Ptolemaic armillary sphere

Ptolemaic armillary sphere



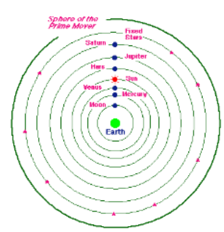
The Copernican Revolution
1543 AD



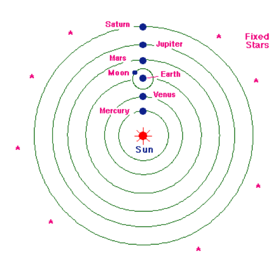
The Scientific Method - from a Popperian perspective
Hypothetico - deductive method

Considerations - problems with the Popperian hypothetico -deductive approach)

1) This type of normal science may rarely lead to revolutions in Science (Kuhn)
   - A) Falsification science leads to paradigms - essentially a way of doing and understanding science that has followers
   - B) Paradigms have momentum - mainly driven by tradition, infrastructure and psychology
   - C) Evidence against accepted theory is considered to be exceptions (that prove the rule)
   - D) Only major crises lead to scientific revolutions
     - 1) paradigms collapse from weight of exceptions - normal science - crisis - revolution - normal science



Aristotle - Ptolemian universe



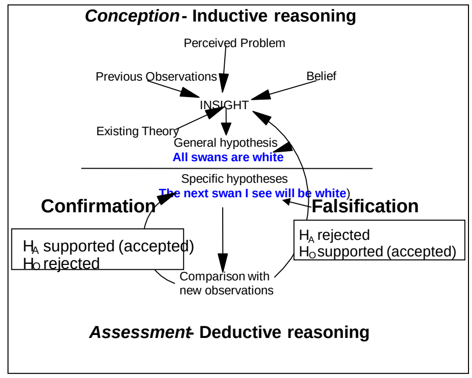
Copernican Universe

The Scientific Method - from a Popperian perspective

Hypothetico - deductive method

Considerations - problems with the Popperian hypothetico -deductive approach)

1) Choice of Method for doing science. Platt (1964) reviewed scientific discoveries and concluded that the most efficient way of doing science consisted of a method of formal hypothesis testing he called **Strong Inference**.
   - A) Apply the following steps to every problem in Science - **formally, explicitly and regularly**:
     - 1) Devise alternative hypotheses
     - 2) Devise critical experiments with alternative possible outcomes, each of which will exclude one or more of the hypotheses (rejection)
     - 3) Carry out procedure so as to get a clean result
     - 1') Recycle the procedure, making subhypotheses or sequential ones to define possibilities that remain

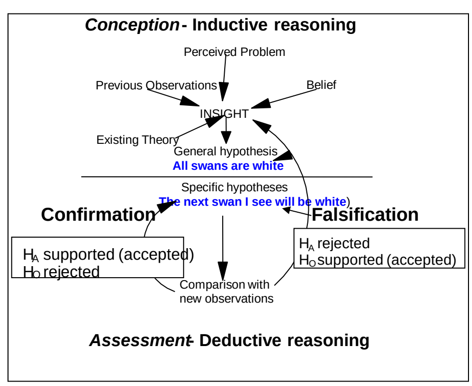


The Scientific Method - from a Popperian perspective

Hypothetico - deductive method

Considerations - problems with the Popperian hypothetico -deductive approach)

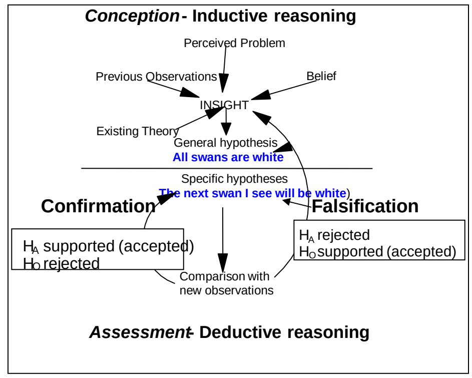
2) Philosophical opposition - (e.g. Roughgarden 1983)
   - A) Establishment of empirical fact is by building a convincing case for that fact.
   - B) We don't use formal rules in everyday life, instead we use native abilities and common sense in building and evaluating claims of fact
   - C) Even if we say we are using the hypothetico - deductive approach, we are not, instead we use intuition and make it appear to be deduction



The Scientific Method - from a Popperian perspective
Hypothetico - deductive method

Considerations - problems with the Popperian hypothetico -deductive approach)

3) Practical opposition - (e.g. Quinn and Dunham 1983)
- A) In practice ecology and evolution differ from Popperian science
  - 1) they are largely inductive
  - 2) although falsification works well in physical and some experimental areas of biology - it is difficult to apply in complex systems of multiple causality - e.g. Ecology and Evolution
  - 3) Hypothetico - deductive reasoning works well if potential cause is shown **not to work at all (falsified)** but this rarely occurs in Ecology or Evolution - usually effects are of degree.



This may be a potent criticism and it leads to the use of inferential statistics

Absolute vs. measured differences

A) Philosophical underpinnings of Popperian Method is based on absolute differences
- 1) E.g. All swans are white, therefore the next swan I see will be white - If the next swan is not white then the hypothesis is refuted absolutely.

B) Instead, most results are based on comparisons of measured variables
- 1) not really true vs. false but degree to which an effect exists

Example - **Specific hypothesis** – number of Oak seedlings is higher in areas outside impact sites than inside impact sites

Observation 1:

	Number inside	Number outside
	0	10
	0	15
	0	18
	0	12
	0	13
Mean	0	13

Observation 2:

	Number inside	Number outside
	3	10
	5	7
	2	9
	8	12
	7	8
Mean	5	9.2

What counts as a difference?
Are these different?

Almost all ordinary statistics are based on a null distribution

- If you understand a null distribution and what the correct null distribution is then statistical inference is straight-forward.
- If you don't, ordinary statistical inference is bewildering
- **A null distribution is the distribution of events that could occur if the null hypothesis is true**

A brief digression to re-sampling theory

	Number inside	Number outside
	3	10
	5	7
	2	9
	8	12
	7	8
Mean	5	9.2

Traditional evaluation would probably involve a t test: another approach is re-sampling.

Resampling to develop a null distribution

Treatment	Number
Inside	3
Inside	5
Inside	2
Inside	8
Inside	7
Outside	10
Outside	7
Outside	9
Outside	12
Outside	8

1) Assume both treatments come from the same distribution, that is, if sampled sufficiently we would find no difference between the values inside vs. outside.
   a. Usually we compare the means.
2) Resample groups of 5 observations (why 5?), with replacement, but irrespective of treatment

Resampling

Treatment	Number
Inside	3
Inside	5
Inside	2
Inside	8
Inside	7
Outside	10
Outside	7
Outside	9
Outside	12
Outside	8

1) Assume both treatments come from the same distribution
2) Resample groups of 5 observations, with replacement, but irrespective of treatment

Resampling

Treatment	Number
Inside	3
Inside	5
Inside	2
Inside	8
Inside	7
Outside	10
Outside	7
Outside	9
Outside	12
Outside	8

1) Assume both treatments come from the same distribution
2) Resample groups of 5 observations, with replacement, but irrespective of treatment
3) Calculate means for each group of 5

7.6

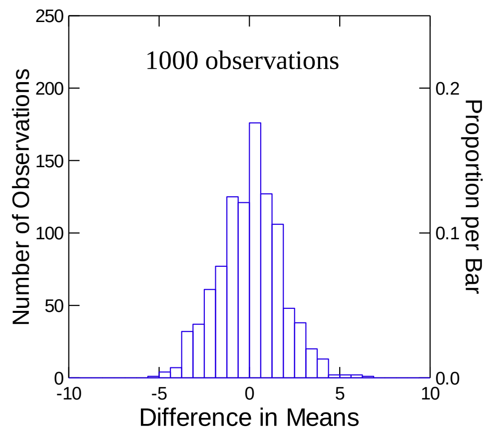
Resampling

Treatment	Number
Inside	3
Inside	5
Inside	2
Inside	8
Inside	7
Outside	10
Outside	7
Outside	9
Outside	12
Outside	8

1) Assume both treatments come from the same distribution
2) Resample groups of 5 observations, with replacement, but irrespective of treatment
3) Calculate mean for each group of 5
4) Repeat many times
5) Calculate differences between pairs of means (remember the null hypothesis is that there is no effect of treatment). This generates a distribution of differences.

Mean 1	Mean 2	Difference
8	7.8	0.2
5.6	8.2	-2.6
6	9	-3
8	5	3
6	6	0
7	8	-1
6	6.8	-0.8
8	7.2	0.8
8	6.6	1.4
7	8.4	-1.4
6	5.4	0.6
7	6.4	0.6
6.4	6.8	-0.4
5	3.4	1.6
6.8	4.8	2
6.4	7.2	-0.8
7.2	8	-0.8
6.4	4.6	1.8
8.4	6	2.4
7.4	6.6	0.8
5.6	8.4	-2.8
8.2	6.2	2
7.8	8.4	-0.6
8.6	6.6	2
6	10.2	-4.2
6.8	5.6	1.2
6.4	7.8	-1.4
7.2	4.8	2.4
6.6	7.2	-0.6
7	5.2	1.8
6.6	9.8	-3.2
8.4	7.8	0.6

Distribution of differences

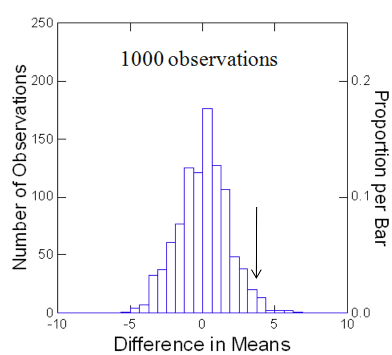


OK, now what?

Compare distribution of differences to real difference

	Number inside	Number outside
	3	10
	5	7
	2	9
	8	12
	7	8
Mean	5	9.2

Real difference = 4.2



Estimate likelihood that real difference comes from two similar distributions

Mean 1	Mean 2	Difference	Proportion of differences less than current
10.2	3.6	6.6	1
10	3.8	6.2	0.999
10.2	4.4	5.8	0.998
9.2	3.6	5.6	0.997
9.8	4.8	5	0.996
8.8	4.2	4.6	0.995
9.6	5.2	4.4	0.994
9.8	5.6	4.2	0.993
9.8	5.8	4	0.992
9.4	5.4	4	0.991

And on through 1000 differences

	Number inside	Number outside
	3	10
	5	7
	2	9
	8	12
	7	8
Mean	5	9.2

Likelihood is 0.007 that distributions are the same

What are constraints of this sort of approach?

These constraints and design complexity lead to more traditional approaches

Statistical analysis - cause, probability, and effect

Statistical tests - a set of rules whereby a decision about hypotheses is reached (accept, reject)

1) Associated with rules - some indication of the accuracy of the decisions - **that measure is a probability statement or p-value**
2) Statistical hypotheses:
   a) **do not become false** when a critical p-value is exceeded
   b) **do not become true** if bounds are not exceeded
   c) Instead p-values indicate a level of acceptable uncertainty
   d) **critical p-values are set by convention -** *what counts as acceptable uncertainty*
   e) Example - if critical p-value = 0.05 this means that we are unwilling to accept the posed **alternative hypothesis** unless:
      1) we 95% sure that it is correct, or equivalently that
      2) we are willing to accept an error rate of 5% or less that we are wrong when we accept the hypothesis

Statistical analysis - cause, probability, and effect

The logic of statistical tests - how they are performed

1. Assume the null hypothesis (H_o) is true: (e.g.) No difference in number of oak seedlings in impact an non-impact sites.
2. **Construct null distribution** (many forms). **Construction of correct null distribution is (in my opinion) the single most important step in inferential statistics)**
   a) Most null distributions use measures of central tendency (e.g. mean) and variability (e.g. standard error) from original data sets (e.g. number of oak seedlings in impact areas) in their construction.
3. Determine the probability the null hypothesis is true using null distribution
4. Compare that value to critical p-value to assign significance
5. Make a conclusion with respect to the null hypothesis

How to create a null distribution?

- One approach that approximates this is resampling, which uses measured observations to build a distribution of means.
  - Limited by…
- Other more traditional approach is to approximate distribution of means using a statistical distribution (for example using a t-test)
  - What is needed
    - Mean
    - Standard deviation

Types of statistical error – Type 1 and II

Type 1 and Type II error.

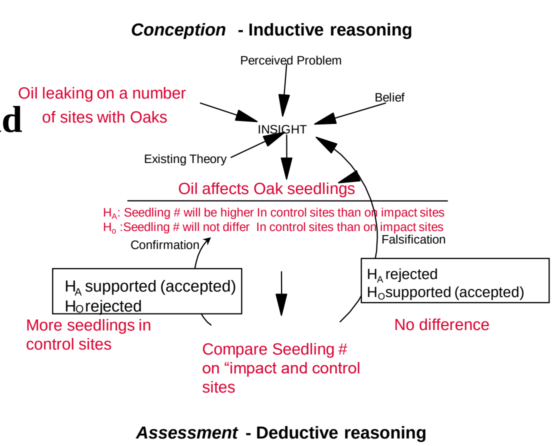
1) By convention, the hypothesis tested is the null hypothesis (no difference between)
   a) In statistics, assumption is made that a hypothesis is true (assume H_O true = assume H_A false)
   b) accepting H_O (saying it is likely to be true) is the same as rejecting H_A (falsification)
   c) Scientific method is to falsify competing alternative hypotheses (alternative H_A's)
2) Errors in decision making

	Decision	
Truth	Accept H_O	Reject H_O
H_O true	no error (1-α)	Type I error (α)
H_O false	Type II error (β)	no error (1- β)

Type I error - probability α that we mistakenly reject a true null hypothesis (H_O)
Type II error - probability β that we mistakenly fail to reject (accept) a false null hypothesis
Power of Test - probability (1-β) of not committing a Type II error - The more powerful the test the more likely you are to correctly conclude that an effect exists when it really does (reject H_O when H_O false = accept H_A when H_A true).

Scientific method and statistical errors

- case example



	Decision	
Truth	**Accept H_o**	**Reject H_o**
H_o true	no error (1-alpha)	Type I error (alpha)
H_o false	Type II error (beta)	no error (1-beta)

Error types and implications in basic and environmental science

	Monitoring Conclusion	
Biological Truth	**No Impact**	**Impact**
No Impact	Correct decision No impact detected	*Type 1 Error* False Alarm
Impact	*Type II Error* Failure to detect real impact; false sense of security	Correct decision Impact detected

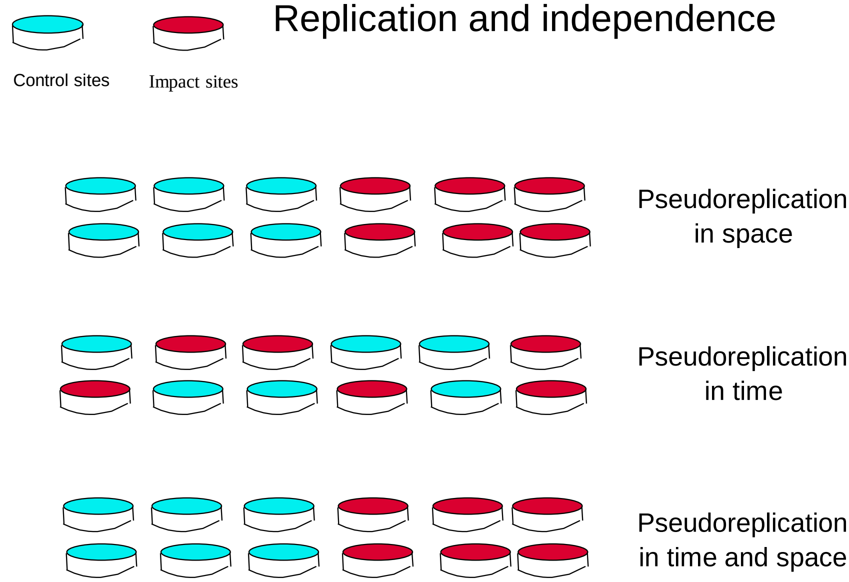
	Decision	
Truth	**Accept H_0**	**Reject H_0**
H_0 true	no error (1-alpha)	Type I error (alpha)
H_0 false	Type II error (beta)	no error (1-beta)

What type of error should we guard against?

Sampling Objectives

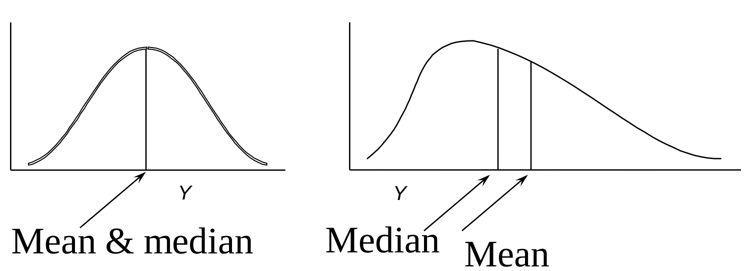
- To obtain an unbiased estimate of a population **mean**
  - Must ensure that replicate measurements are representative of the treatments of interest and the area and timeframe of inference
    - “representative” will be discuss much more later
- To assess the precision of the estimate (i.e. calculate the **standard error** of the mean)
- To obtain as precise an estimate of the parameters as possible for time, effort and money spent

Replication and independence



Measures of location

- Population mean (μ) - the average value
- Sample mean = $\overline{y}$ estimates μ (true mean)
- Population median - the middle value
- Sample median estimates population median
- In a normal distribution the mean=median (also the mode), this is ***not*** ensured in other distributions



Measures of dispersion

- Sample variance (s^2) estimates population variance

$$\frac{\Sigma(x_i - \bar{x})^2}{n - 1}$$

- Standard deviation (s)
  - square root of variance
  - same units as original variable

Measures (statistics) of Dispersion

Sample Sum of Squares $\quad SS = \Sigma(x_i - \bar{x})^2$

Sample variance $\quad s^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n - 1}$

- Note, units are squared
- Denominator is (n-1)

Sample standard deviation $\quad s = \sqrt{\dfrac{\Sigma(x_i - \bar{x})^2}{n - 1}}$

- Note, units are not squared

Standard error of the mean $\quad se = \sqrt{\dfrac{s^2}{n}} = \dfrac{s}{\sqrt{n}}$

Null distribution (example *t* distribution)

- The distribution of events if the null hypothesis is true.
- Ho: $$\bar{y}_1 = \bar{y}_2$$ $$\bar{y}_1 - \bar{y}_2 = 0$$
- Events come from the set of differences under null hypothesis, that is, set of $\bar{y}_1 - \bar{y}_2$ values that could exist if the null hypothesis is true

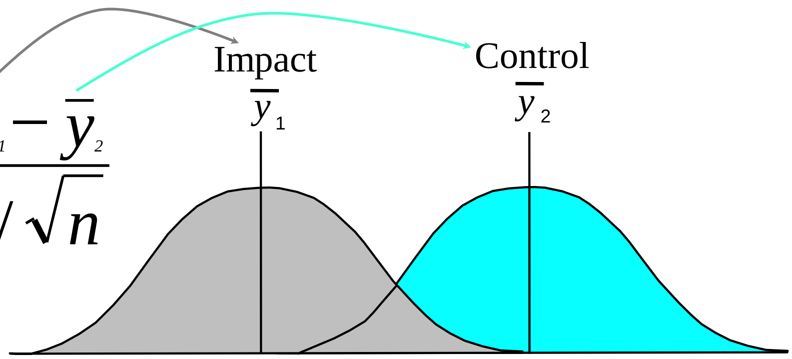
t statistic – interpretation and units

- The deviation between means is expressed in terms of Standard error (i.e. Standard deviations of the sampling distribution)
- Hence the value of t's are in standard errors
- For example t=2 indicates that the deviation ($\bar{y}_1$- $\bar{y}_2$) is equal to 2 x the standard error

$$\frac{\bar{y}_1 - \bar{y}_2}{s/\sqrt{n}}$$

Statistical Power, effect size, replication and alpha

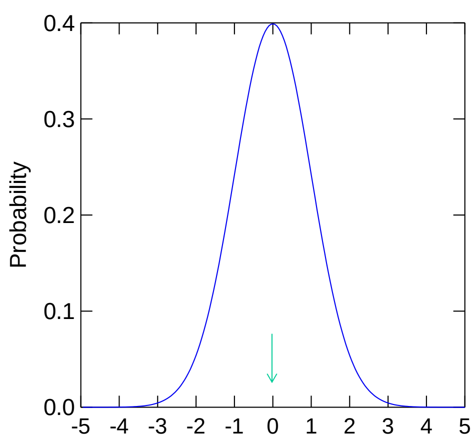
Statistical comparison of two distributions



$$\frac{\bar{y}_1 - \bar{y}_2}{s/\sqrt{n}}$$

Number of seedlings

Area under the curve = 1.00



$$\frac{\bar{y}_1 - \bar{y}_2}{s/\sqrt{n}}$$

Null distribution

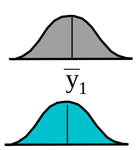
Ho:

$$\bar{y}_1 = \bar{y}_2$$

Components of the t equation-comparing two samples

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\text{Difference in the means of two samples}}{\text{Pooled standard error}}$$

1) Pooled standard error is an estimate of the error associated with the calculation of the difference between two sample means
2) Hence the t value increases with increasing difference between means and with decreasing standard error. The larger the value, the more confidence in concluding distributions (1 vs 2) are different



$\bar{y}_1$

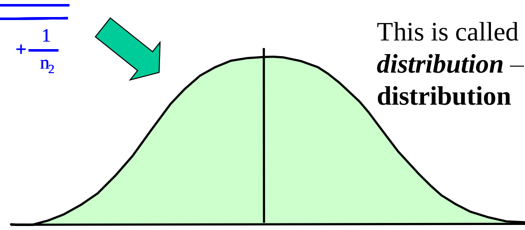
$\bar{y}_2$

Ho true: Distributions of means are truly the same

	Decision	
Truth	**Accept H_o**	**Reject H_o**
H_o true	**no error (1-alpha)**	**Type I error (alpha)**
H_o false	Type II error (beta)	no error (1-beta)

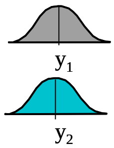
$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

This is called the ***central t distribution*** – it is a **null distribution**

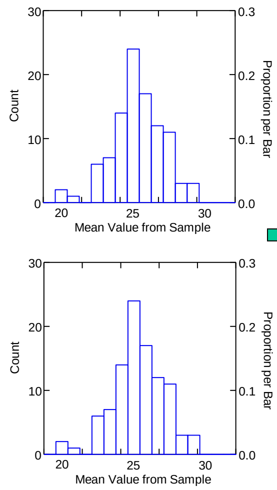


$\bar{t} = 0$

central t distribution (df)

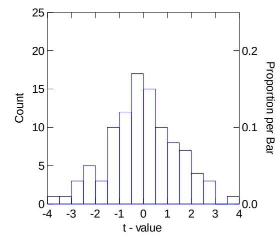


Ho true: Distributions of means are truly the same

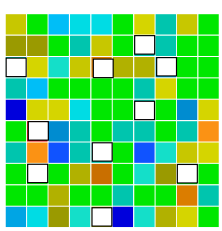


$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

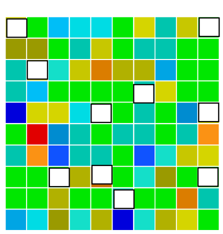
t distribution



True Mean = 25



Mean = 24.9

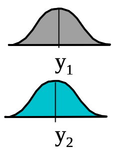


Mean = 25.1

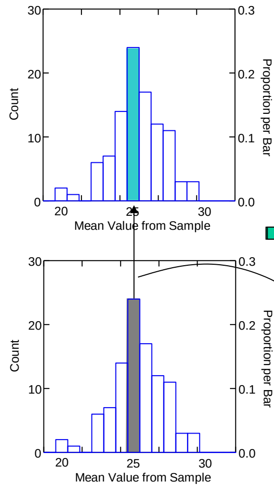
Sample Means

21.5
22.3
23.0
23.9
24.9
25.1
25.8
26.5
27.8
29.9

white = sample plots

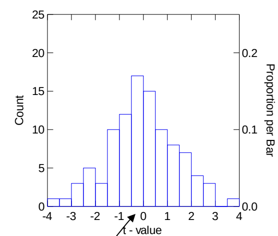


Ho true: Distributions of means are truly the same
1) Estimate of $\bar{y}_1$ = estimate of $\bar{y}_2$

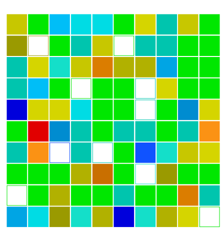


$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

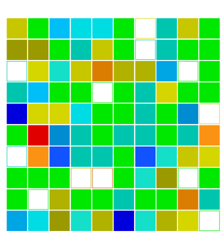
t distribution



True Mean = 25



Mean = 21.5

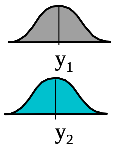


Mean = 29.9

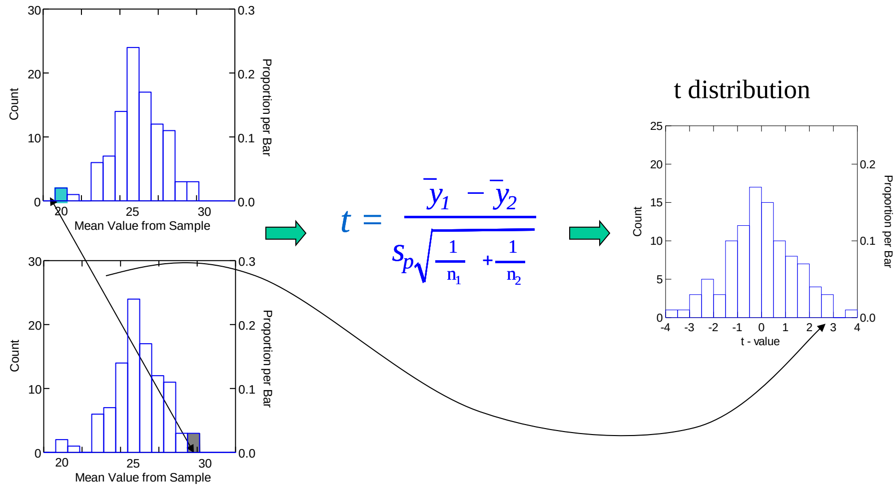
<u>Sample Means</u>

21.5
22.3
23.0
23.9
24.9
25.1
25.8
26.5
27.8
29.9

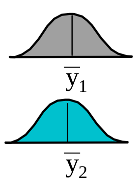
white = sample plots



Ho true: Distributions of means are truly the same
2) Estimate of $\bar{y}_1 \neq$ estimate of $\bar{y}_2$



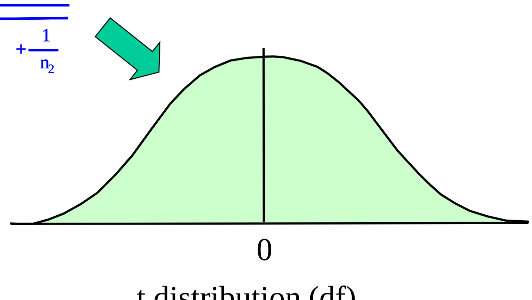
$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

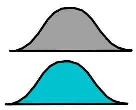


Ho true: Distributions of means are truly the same

	Decision	
Truth	Accept H0	Reject H0
H0 true	no error (1-alpha)	Type I error (alpha)
H0 false	Type II error (beta)	no error (1-beta)

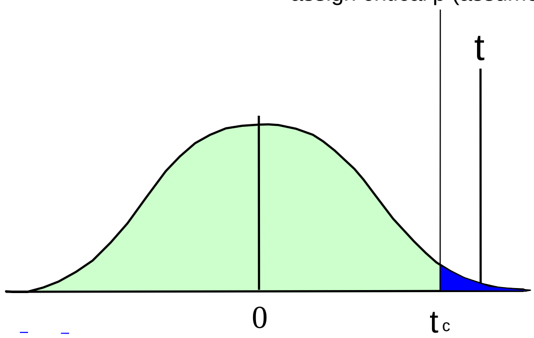
$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$





Ho true: Distributions of means are truly the same

assign critical p (assume = 0.05)



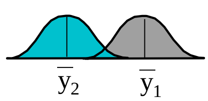
	Decision	
Truth	**Accept H₀**	**Reject H₀**
H₀ true	**no error (1-alpha)**	**Type I error (alpha)**
H₀ false	Type II error (beta)	no error (1-beta)

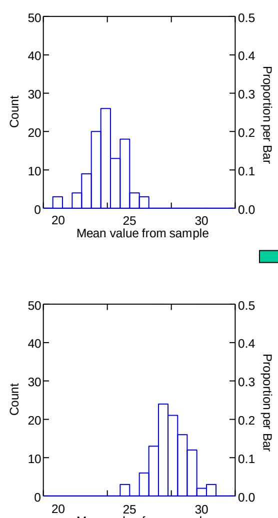
$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

t distribution (df)

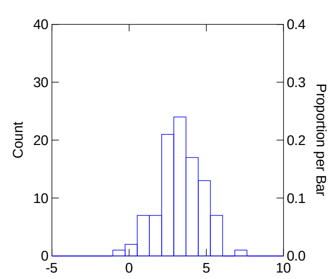
If distributions are truly the same then: area to right of critical t_c represents the Type 1 error rate (Blue); any calculated $t > t_c$ will cause incorrect conclusion that distributions are different

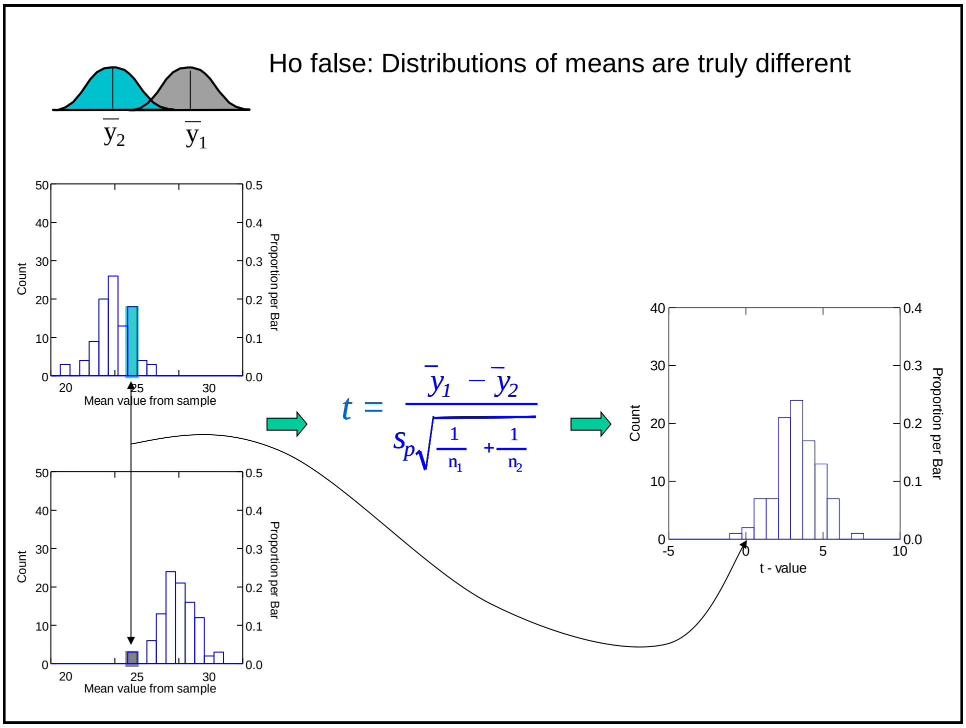
Ho false: Distributions of means are truly different

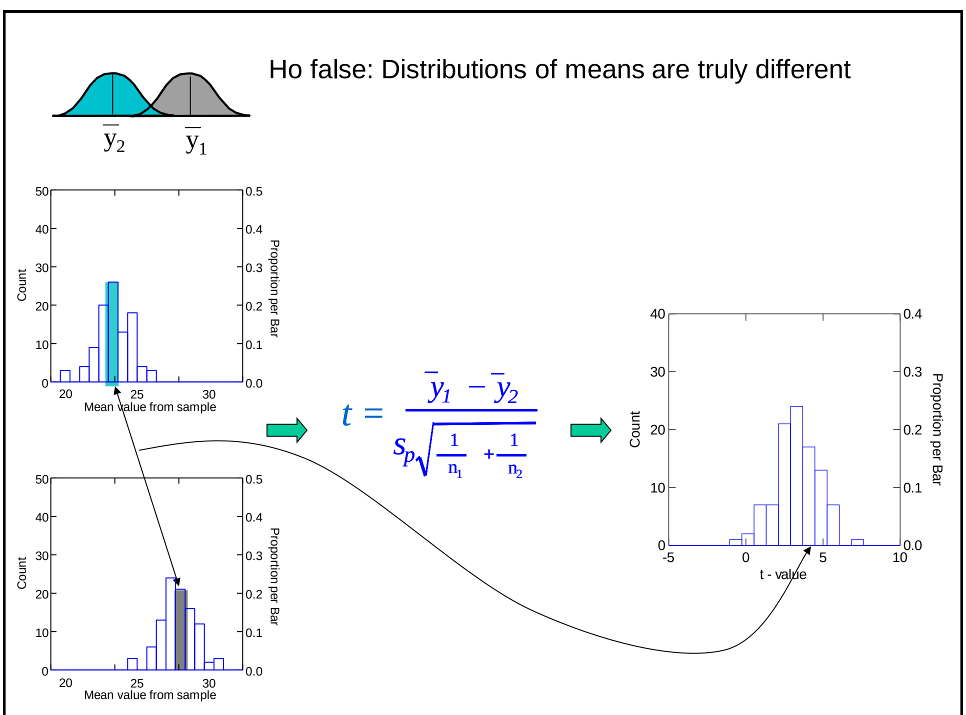




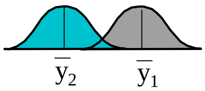
$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



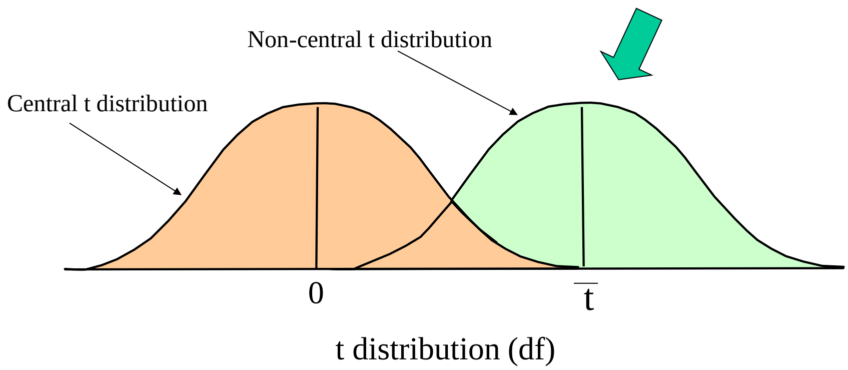

Ho false: Distributions of means are truly different
$\bar{y}_2$
$\bar{y}_1$
Count
Proportion per Bar
Mean value from sample
Count
Proportion per Bar
Mean value from sample
$t = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
Count
Proportion per Bar
t - value


Ho false: Distributions of means are truly different
$\bar{y}_2$
$\bar{y}_1$
Count
Proportion per Bar
Mean value from sample
Count
Proportion per Bar
Mean value from sample
$t = \dfrac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
Count
Proportion per Bar
t - value

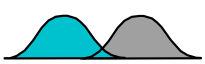
Ho false: Distributions of means are truly different

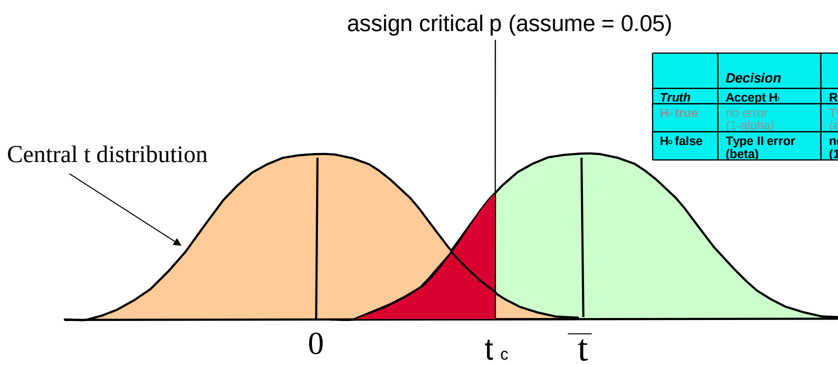


$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



Ho false: Distributions of means are truly different





	Decision	
Truth	Accept H_o	Reject H_o
H_o true	no error (1-alpha)	Type I error (alpha)
H_o false	Type II error (beta)	no error (1-beta)

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

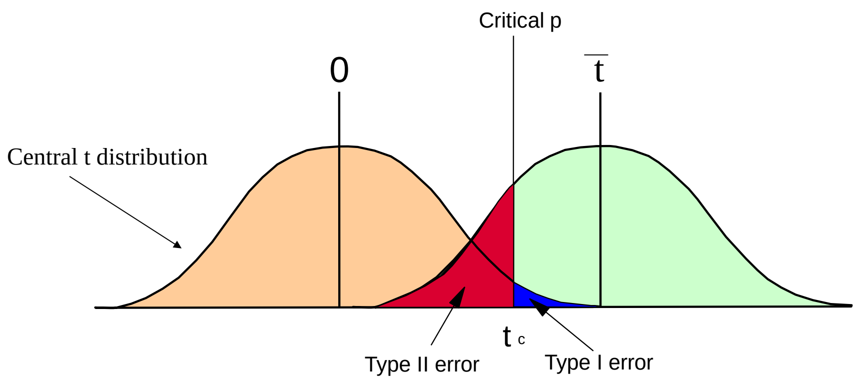
If distributions are truly different then: area to left of critical t_c represents the Type II error rate (Red); any Calculated $t < t_c$ will cause incorrect conclusion that distributions are same

How to control Type II error (distributions are truly different)

This will maximize statistical power to detect real impacts

1) Vary critical P-Values
2) Vary Magnitude of Effect
3) Vary replication

	Decision	
Truth	Accept H_0	Reject H_0
H_0 true	no error (1-alpha)	Type I error (alpha)
H_0 false	Type II error (beta)	no error (1-beta)



How to control Type II error (distributions are truly different)

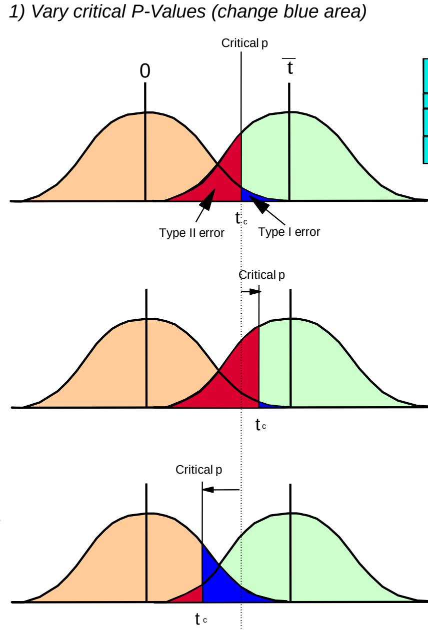
1) Vary critical P-Values (change blue area)

	Decision	
Truth	Accept H_0	Reject H_0
H_0 true	no error (1-alpha)	Type I error (alpha)
H_0 false	Type II error (beta)	no error (1-beta)

Reference

A) Make critical P more stringent (smaller)

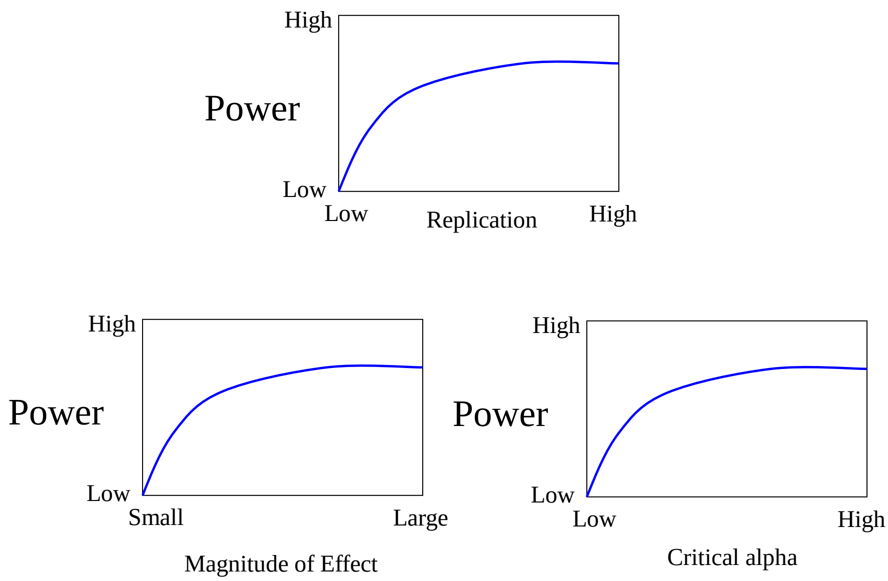
A) Relax critical P (larger values)



Type II error increases
Power decreases

Type II error decrease
Power increases

POWER


High
Power
Low
Low
Replication
High
High
Power
Low
Small
Large
Magnitude of Effect
High
Power
Low
Low
High
Critical alpha