The Scientific Method

http://www.pelagicos.net/classes_biometry_fa16.htm

What is Science ?

Science - *the attempt to come up with systematic and coherent descriptions of how the world works*

Scientific Method – the approach whereby scientists decide among possible competing explanations on the basis of observations and predictions

The Scientific Method

1. Assemble a model of how a particular phenomenon works
2. Develop predictions based on that conceptual model
3. Test predictions by collecting observations (data) through correlational or manipulative experiments

Scientific Reasoning

DEDUCTION: From general to specific

1. All of the turtles in Kailua Bay are green sea turtles
2. I sampled this particular turtle in Kailua Bay
3. This particular turtle is a green sea turtle

INDUCTION: From specific to general

1. All 100 of these turtles are green sea turtles
2. All 100 of these turtles were sampled in Kailua Bay
3. All of the turtles in Kailua Bay are green sea turtles

Two Ways of "Finding Out" Nature

- **Induction (14th century)**

• Inference of generalized conclusions from compilation of specific instances (observations)

• A inductive conclusion is based on enumeration

- For example:

Mathematical Induction:

Demonstrating validity of a law concerning all the positive integers by proving that it holds for the integer 1 and that, if it holds for an arbitrarily chosen positive integer k, it must hold for integer k+1

Two Ways of "Finding Out" Nature

- **Deduction (15th century)**

- The derivation of a conclusion by reasoning
- Inference in which conclusions about particulars follow necessarily from general or universal premises
- Deductive conclusions require the formulation and testing of a model describing the general premises

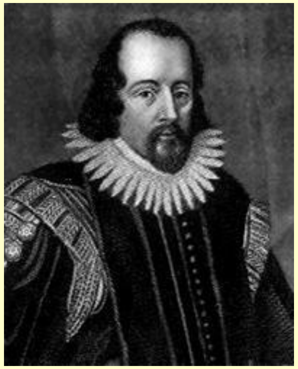
- For example:

Isaac Newton's Law of Universal Gravitation

$$Fg = \frac{(G\ m_1\ m_2)}{r^2}$$

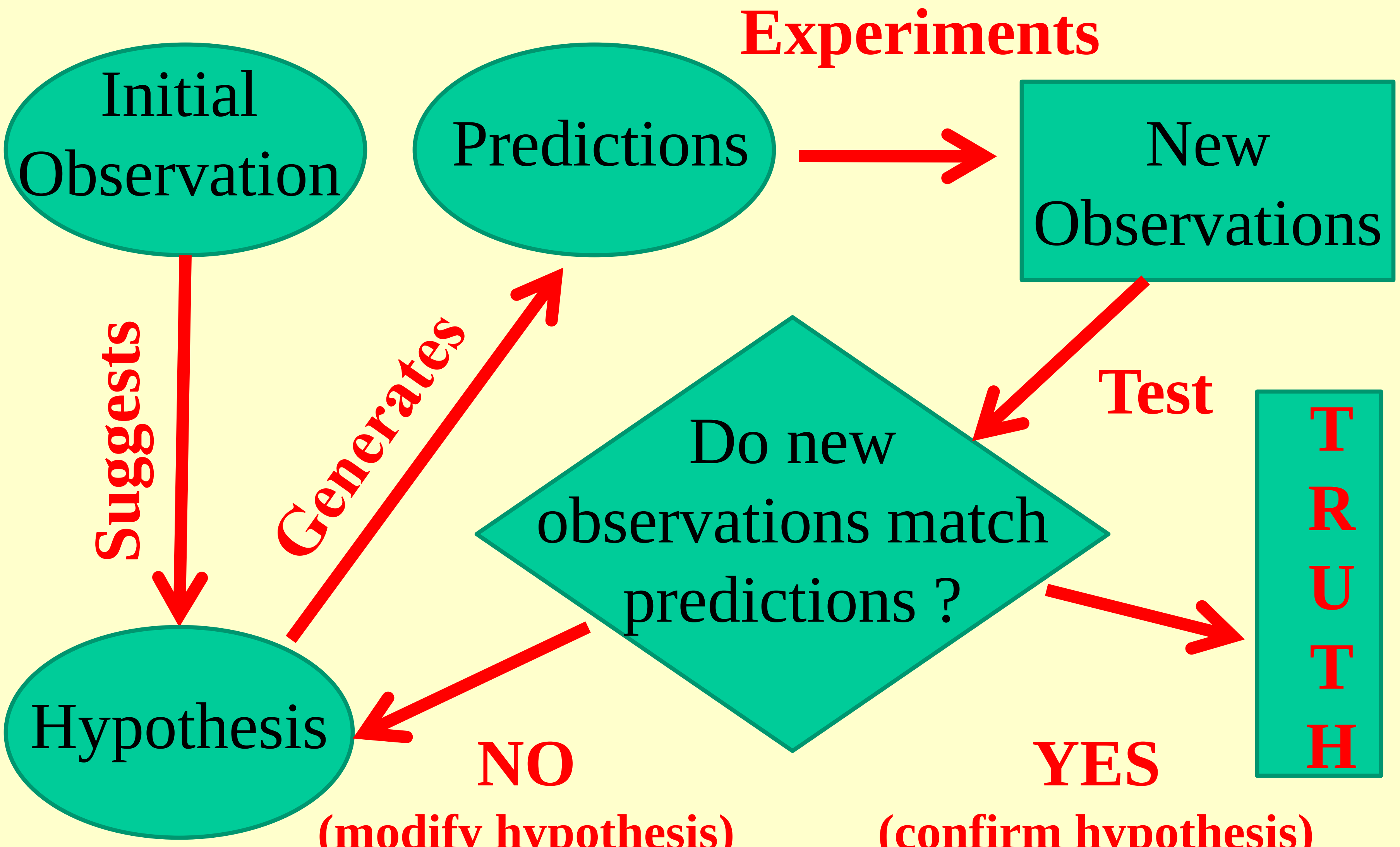
Development of Strong Inference

Francis Bacon (1561-1626)



- Urged an experimental approach
- Showed fruitfulness of interconnecting theory and experiments, so they checked each other
- Suggested conditional inductive tree, which:
  - Starts with alternative hypothesis ("causes")
  - Proceeds through crucial experiments
  - Leads to exclusion of alternatives
  - Promotes adoption of what is left
- Science establishes axioms – "accepted truths"

The Deductive Method



Strong Inference Defined

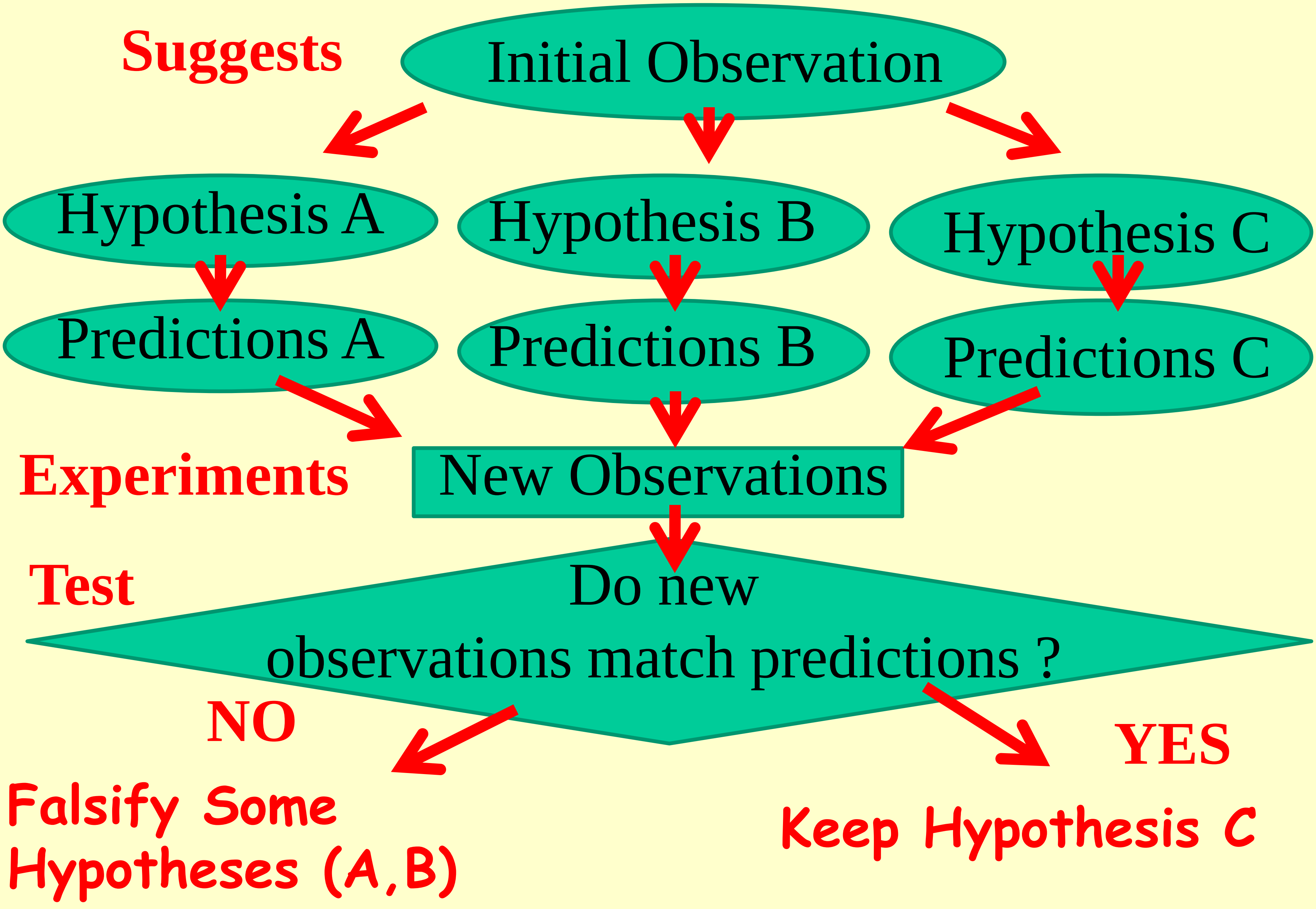
- Strong Inference is a highly systematic approach designed to exclude possible explanations using conclusive experiments
- Strong Inference takes advantage of the relationships among possible hypotheses and the body of existing knowledge
  - Sequential hypotheses
  - "Standing on the back of giants"

Strong Inference in Practice

➢ Strong inference consists of applying the following steps to answer scientific questions, formally and explicitly:

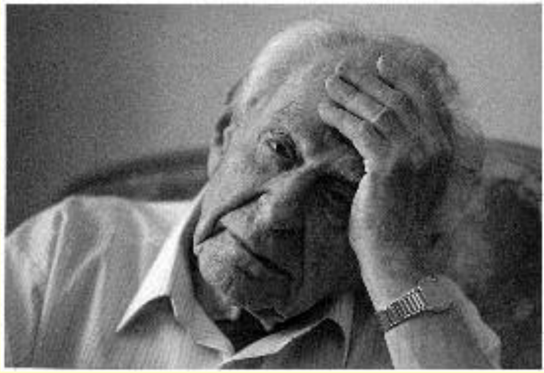
- 1. Consider alternative hypotheses
- 2. Devise crucial experiments, with alternative possible outcomes, which will exclude one or several of the possible hypotheses
- 3. Carry out the experiment(s)
- 4. Recycle the procedure, devise sub-hypotheses or sequential hypotheses to refine remaining options
- 5. And so on …

The Hypothetico-Deductive Method



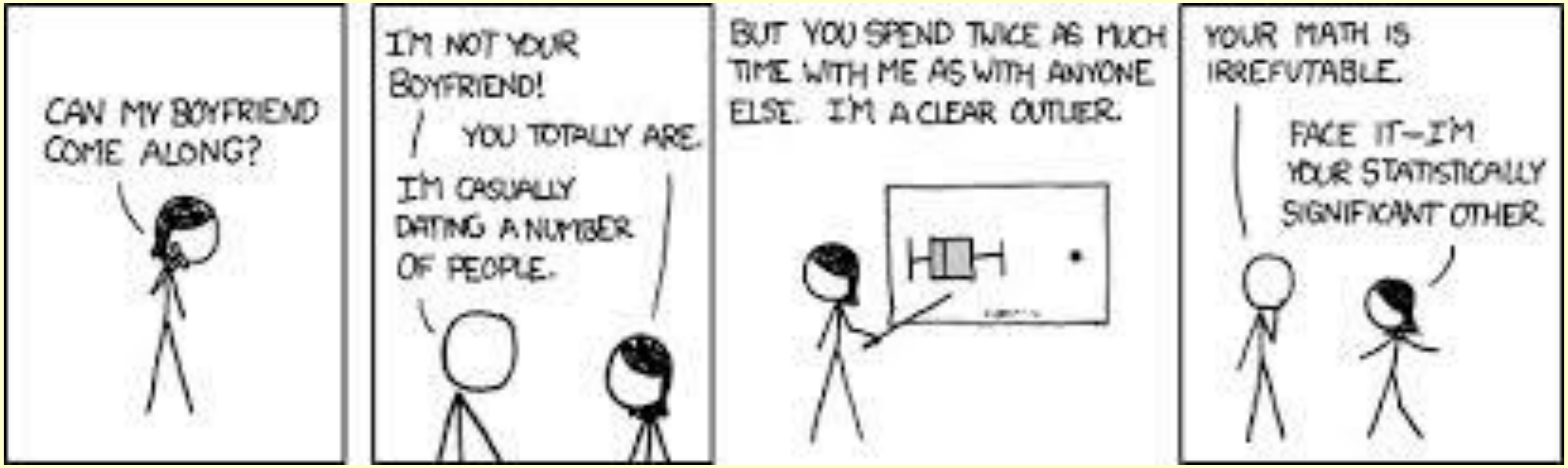
Refinement of Strong Inference

Karl Popper (1902 – 1994)

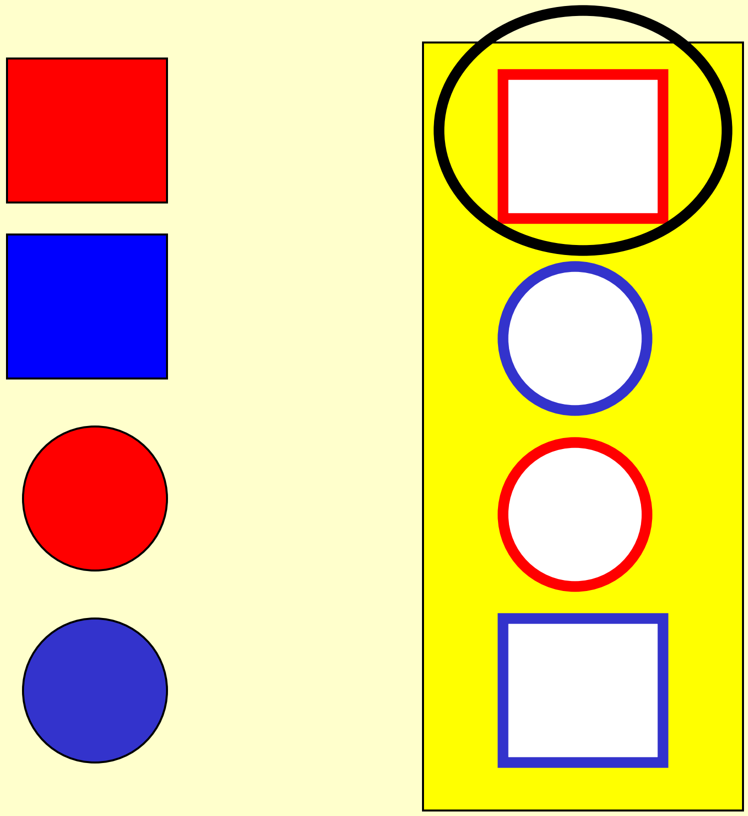


- There are no proofs in science because later explanations may be as good or better than current explanation (technology, progress)
- Therefore, science advances only by disproofs
- There is no point in making hypotheses that are not falsifiable because such hypotheses say nothing about how the world works
- It must be possible for all empirical scientific systems to be refuted by experience

Hypotheses Testing in Every-Day Life



Hypotheses Testing in Every-Day Life



Variables

Shape

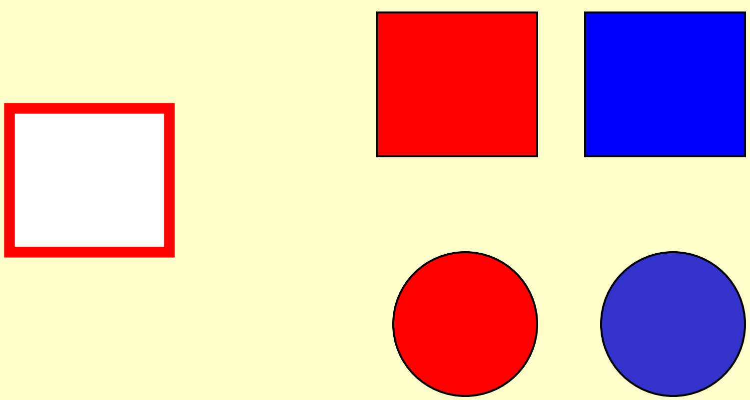
Color

Hypotheses Testing in Every-Day Life

- Develop mutually exclusive and exhaustive hypotheses and rule out all but one of them

Hypothesis 1

Shape = Square

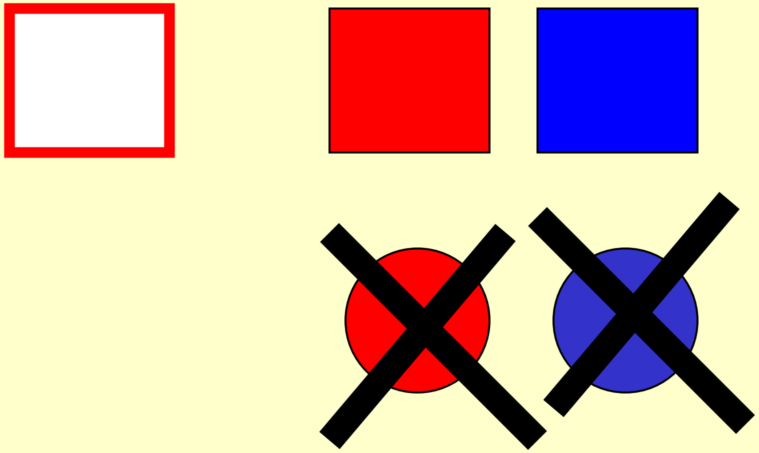


Hypotheses Testing in Every-Day Life

- Hypothesis 1: Allows us to exclude two cases where: Shape = Circle

Hypothesis 1

Shape = Square

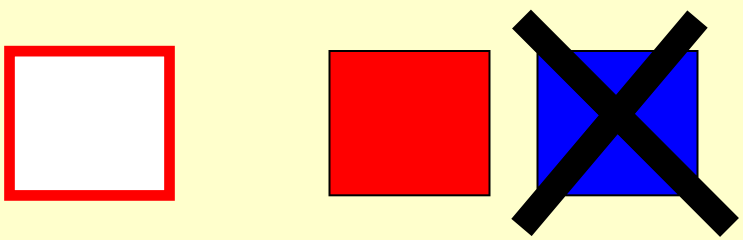


Hypotheses Testing in Every-Day Life

- Hypothesis 2: Allows us to exclude one case where: Color = Blue

Hypothesis 2

Color = Red



Hypotheses Testing in Every-Day Life

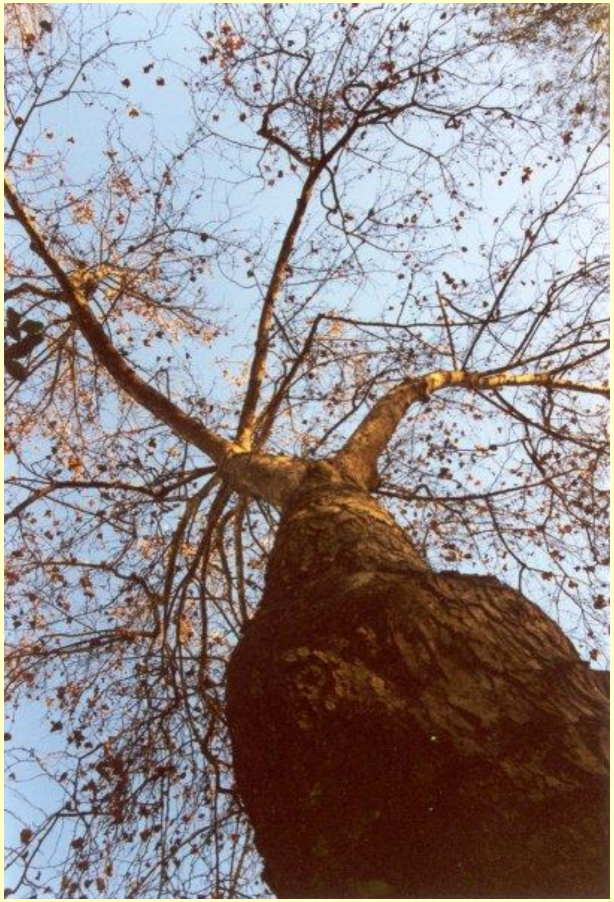
- Three Take-Home Lessons:

  - Phenomena must be testable
  (e.g., undetectable aliens living amongst us)
  (e.g., Galapagos finch related to Peru species)

  - Hypotheses must be mutually exclusive
  (focus on critical discerning predictions)

  - Testing sequence does not affect outcome

Application of Strong Inference



Application of Strong Inference

Logical Trees

- If the answer to a research problem is a leaf, Platt would argue that it is a poor strategy to just go browsing , because they are usually innumerable

- Rather, climb the tree from ground level up

- Then narrow things down to a few major branches by clean experimental exclusions

- Continue moving out secondary and tertiary branches, … until arriving to where answer lies

Application of Strong Inference

Recipe for Efficient Hypothesis Testing

- Sets of models can be broken down into subsets (family lineages) that share particular features
- If one can do a test that falsifies that common feature, then a whole lineage of possible models can be eliminated (pruned) in one swift blow
- This is an important key to research efficiency
- Effective investigators entertain multiple models simultaneously
- Strong tests eliminate some of the models by addressing the critical discerning predictions

Application of Strong Inference

What happens with the unfalsified hypothesis?

- Unfalsified hypotheses gain strength
- Eventually, they become axioms
- Axiom:
  - A maxim widely accepted on its intrinsic merit
  - A statement accepted as true. A postulate
  - An established rule, principle or self-evident truth
- Example: "Ghost of Competition Past" invoked to explain the structure of biotic communities

Derailing Strong Inference

http://www.lbl.gov/Science-Articles/Archive/elements-116-118.html

June 7, 1999 BERKELEY, CA — Discovery of two new "superheavy" elements has been announced by scientists at the U.S. Department of Energy's Lawrence Berkeley National Laboratory. Element 118 and its immediate decay product, element 116, were discovered at Berkeley Lab's 88-Inch Cyclotron by bombarding targets of lead with an intense beam of high-energy krypton ions.

Although both new elements almost instantly decay into other elements, the sequence of decay events is consistent with theories that have long predicted an "island of stability“ for nuclei with approximately 114 protons and 184 neutrons.

http://www.lbl.gov/Science-Articles/Archive/118-retraction.html

On July 27, 2001, the results reported below were retracted through a correspondence with Physical Review Letters

"In 1999, we reported the synthesis of element 118 in the lead-krypton reaction based upon the observation of three decay chains, each consisting of an implanted heavy atom and six sequential high-energy alpha decays, correlated in time and position.

Prompted by the absence of similar decay chains in subsequent experiments, we (along with independent experts) re-analyzed the primary data files from our 1999 experiments. Based on these re-analyses, we conclude that the three reported chains are not in the 1999 data. We retract our published claim for the synthesis of element 118."

Derailing Strong Inference

Axioms

- An established rule or principle
  - ➢ Climbing up wrong tree ("biases")
  - ➢ Picking Specific Branch ("pet hypotheses")

Dogma

- Something held as an
  established opinion
- A tenet put forth as authoritative
  without adequate grounds

Working with Hypotheses

Formulating Hypotheses

Stated as the existence / absence of statistical associations between different processes. (variables in the conceptual model we are testing)

The null hypothesis is the starting point of a scientific investigation.

It attempts to account for patterns observed in the data (observations) in simplest way possible.

This means that the observed variation in the data is due to randomness and measurement error.

Working with Hypotheses

The Null Hypothesis

States that there is no real pattern
(e.g., no association or response).

If the simplest possible explanation is rejected, we can entertain more complex explanations.

The Alternative Hypothesis

States that there is a real pattern
(e.g., a non-random association or response).

Cannot be accepted.

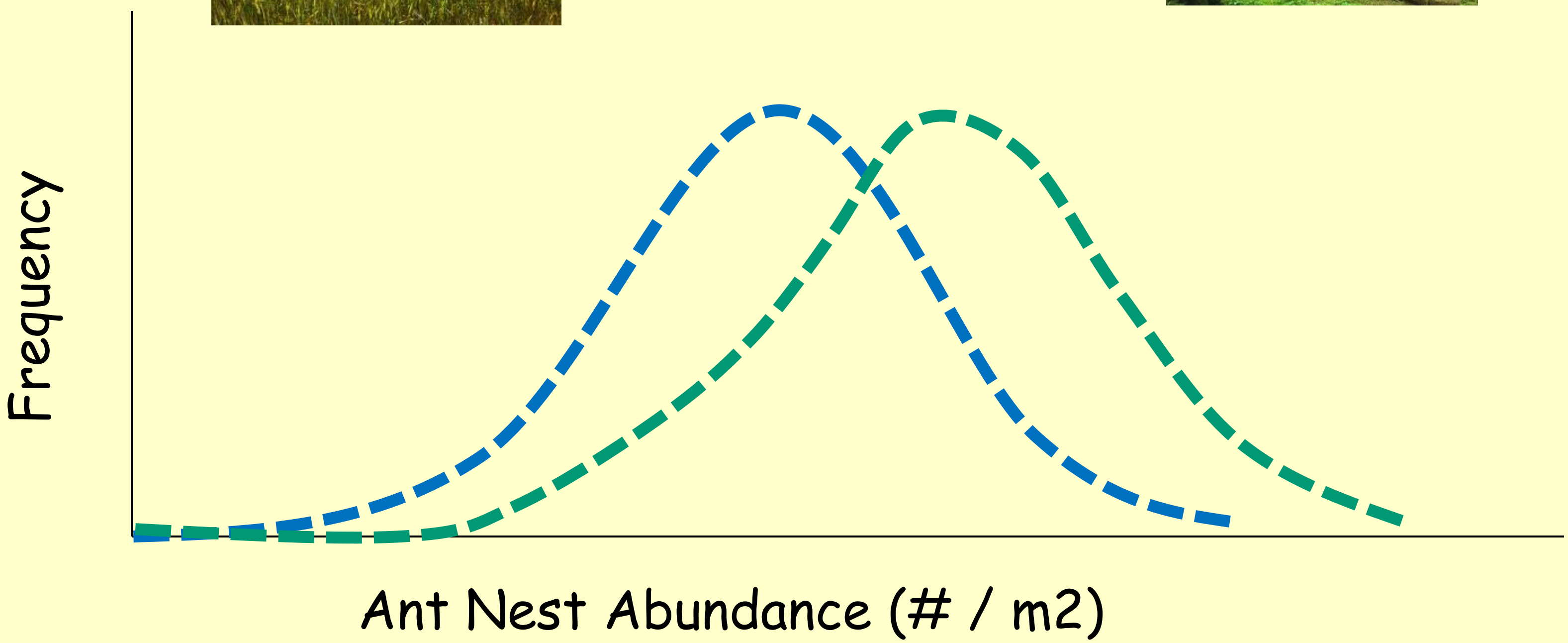
Best explanation, when null hypothesis is rejected.

Parametric Statistics Example

Based on the assumption that these data were sampled from two Normal Distributions (Fisherian Approach):

Field: green curve

Forest: blue curve

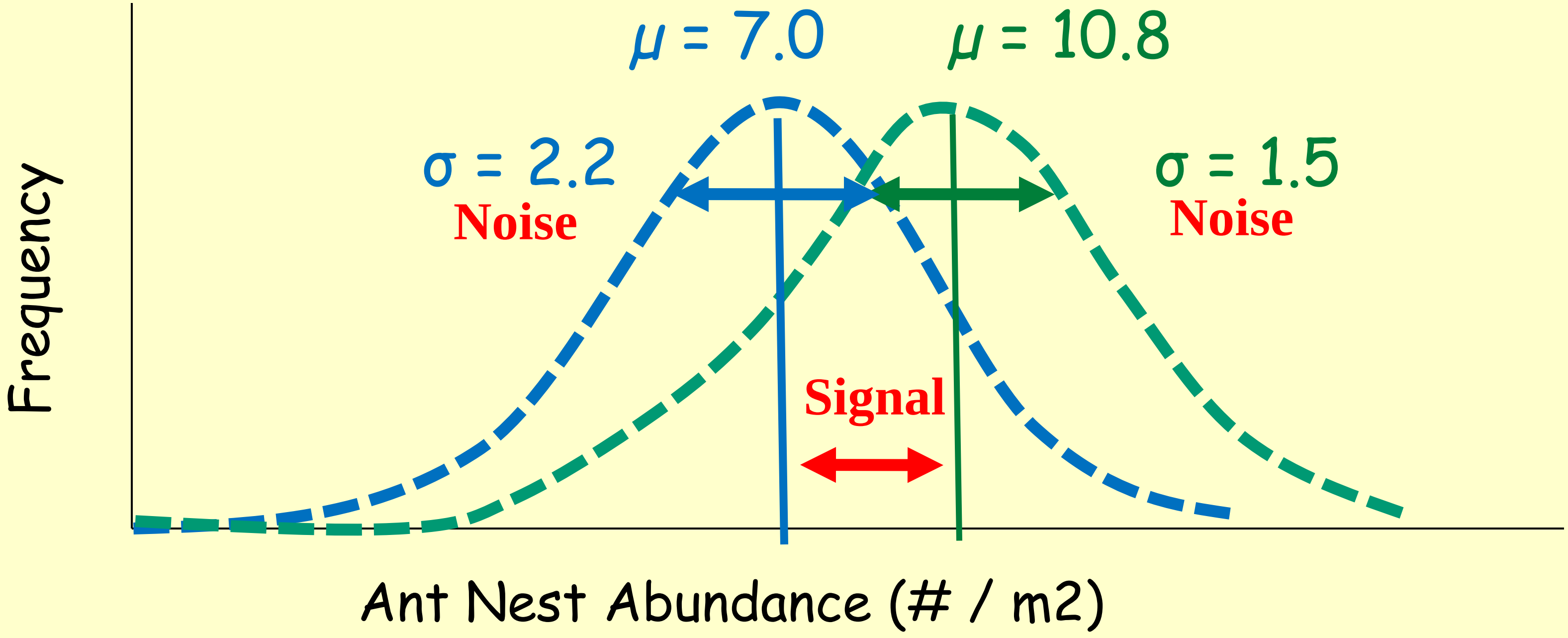


Working with Hypotheses

Once we have made the assumption of normality, we can proceed with the statistical comparison of the two sites:

Specifying the test statistic

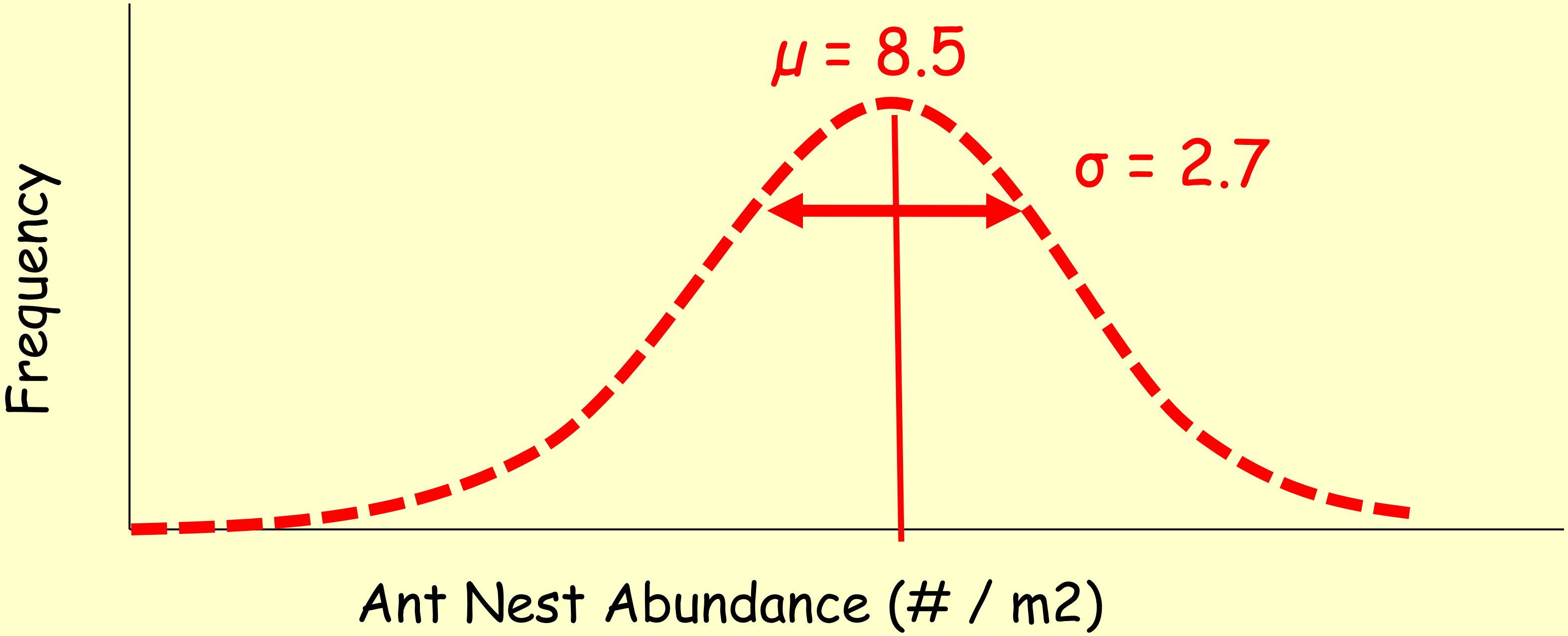
- Compare means (central tendency)
- Relate to their S.D.s (variability)



Working with Hypotheses – Ant Example

The Null Hypothesis ("there is no real pattern")

Ant nest abundance in the forest and the field is the same



Prediction

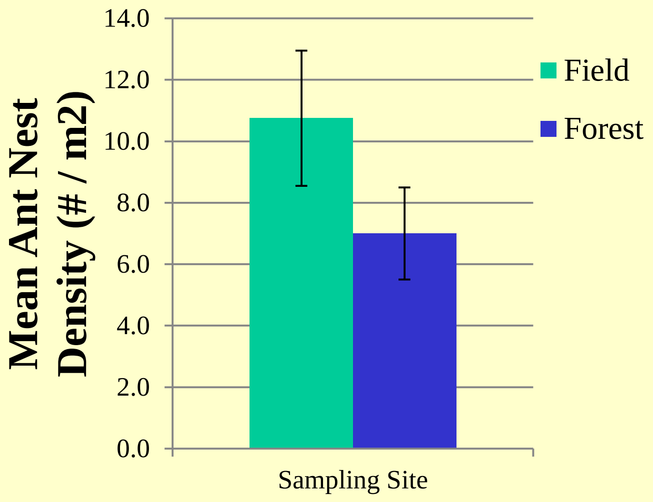
Mean ant nest abundance is the same in both sites

Working with Hypotheses

Spell out the Working Hypotheses:

Null Hypothesis	Alternative Hypothesis
$\mu = \mu$	$\mu \neq \mu$

Perform the Test: Compare Data with Predictions



Anova:

F	P-value	F crit
8.78	0.018	5.32

Working with Hypotheses

The Statistical Null Hypothesis

States that the differences between groups are not greater than would be expected due to random variation alone (e.g., due to sampling).

Thus, this hypothesis states that there is no other specific mechanism or force – other than random variation – causing the observed differences.

So…
how great can these random differences be ?

Testing Statistical Hypotheses

The P-value

The probability of obtaining the observed data (pattern) due to random variation (sampling).

Basically, viewing the data through the lens of the Null Hypothesis. The p - value is the probability of obtaining the data , given than Ho is, in fact, true.

NOTE:
p – value defined using arbitrary alpha level. Usually set at 0.05 (1 in 20).

Testing Statistical Hypotheses

Calculating the test statistic

$$F \text{ ratio: } \frac{\text{(variance between groups)}}{\text{(variance within groups)}}$$

Note: the larger the difference among the two groups, the larger the F

For this example, the F-test yields an F-ratio of:

8.78 = 33.75 / 3.84

The F-ratio is used to quantify the differences among the means from the different groups.

This test bears in mind the variability within and across the groups being compared.

Testing Statistical Hypotheses

Calculating the Probability (P value)

The p-value is defined as the probability of obtaining the observed pattern (a test statistic equal to or larger than the one calculated by the test), *given that* the null hypothesis is, in fact, true.

The P-value is the probability of obtaining an F-ratio $\geq$ 8.78, given a true null hypothesis

(NOTE: THE P VALUE IS BASED ON NOTION THAT THERE REALLY IS NO PATTERN IN THE DATA … BUT ONE APPEARS DUE TO SAMPLING – by chance)

Testing Statistical Hypotheses

Anova: Single Factor

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	33.75	1	33.75	8.78	0.018	5.32
Within Groups	30.75	8	3.84			
Total	64.5	9				

Critical value:
Test statistic that, given the sample size, yields a p value of 0.05 (1 / 20)

Significance:
When p value $\leq$ 0.05

Note: alpha (0.05) is an arbitrary level

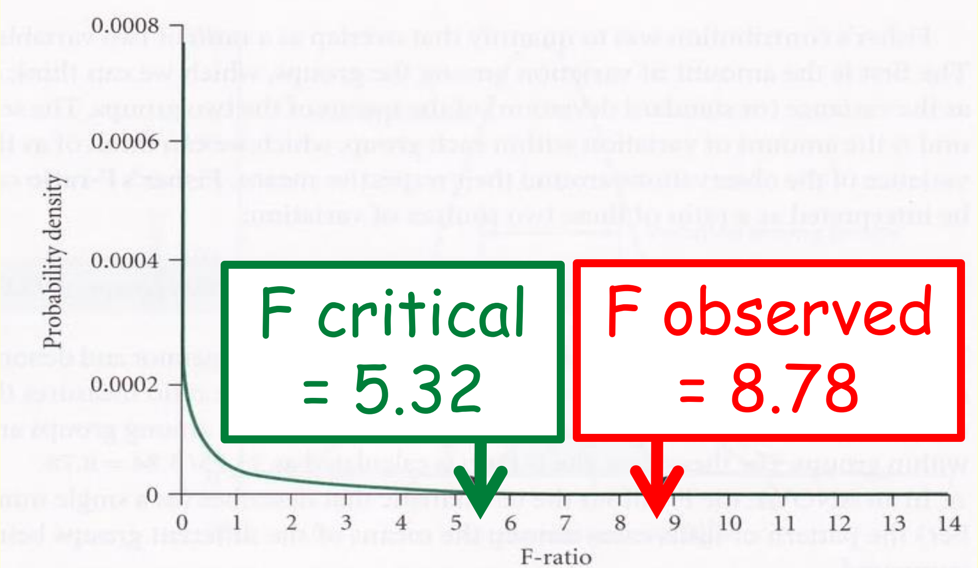


Figure 5.5 Theoretical F-distribution. The larger the observed F-ratio, the more unlikely it would be if the null hypothesis were true. The critical value for this distribution equals 5.32; the area under the curve beyond this point is equal to 5% of the area under the entire curve. The observed F-ratio of 8.78 lies beyond the critical value, so $P(\text{ant data} \mid \text{null hypothesis}) \leq 0.05$. In fact, the actual probability of $P(\text{ant data} \mid \text{null hypothesis}) = 0.018$, because the area under the curve to the right of the observed F-ratio represents 1.8% of the total area under the curve.

The P Value

What Factors Determine the P-value

Sample Size:

The larger the sample size, the easier to find a significant difference (lower p-value)

Signal Strength:

The larger the difference between the two means, the easier to find a significant difference (lower p-value)

Noise:

The larger the variation within the two groups, the harder to find a significant difference (higher p-value)

Testing Statistical Hypotheses

Decision about Null Hypothesis		Null Hypothesis: True	Null Hypothesis: False
	Reject	**False Positive**	correct decision
	Accept	correct decision	**False Negative**

Type I error: rejection of a true null hypothesis

- occurs at rate chosen alpha rate (0.05; 1 in 20)
- Often caused by violation of test assumptions

Type II error: acceptance of a false null hypothesis

- occurs at rate of Beta = (1 – power)
- low power due to small sample size, measurement error

Trade-Offs in Statistical Testing

Decision about Null Hypothesis		Null Hypothesis: True	Null Hypothesis: False
	Reject	**False Positive**	correct decision
	Accept	correct decision	**False Negative**

False Positive: rejection of a true null hypothesis

Error of Falsity: Support more complex relationship than random association. Mis-direct future researchers.

False Negative: acceptance of a false null hypothesis

Error of Ignorance: Support notion of random association. Fail to find difference / response – management implications

Trade-offs in Statistical Testing

Trade-off Between Type I and Type II Errors

Ideally, we want to minimize the risk of making both errors. We want to always make the right decision.

The probability of a type I error (rejecting a true null) is directly proportional on the alpha level.

The probability of a type II error (accepting false null) is inversely proportional on the alpha level.

However, there is no simple mathematical relationship between these two rates… the trade-off depends on the sample size and the magnitude of the signal / noise.

Trade-offs in Statistical Testing

Trade-off Between Type I and Type II Errors

Statistical Power = (1 – β)

Probability of rejecting a false Ho

For Higher Statistical Power:

- Large sample size
- Higher P-value
- Larger pattern (signal/noise ratio)

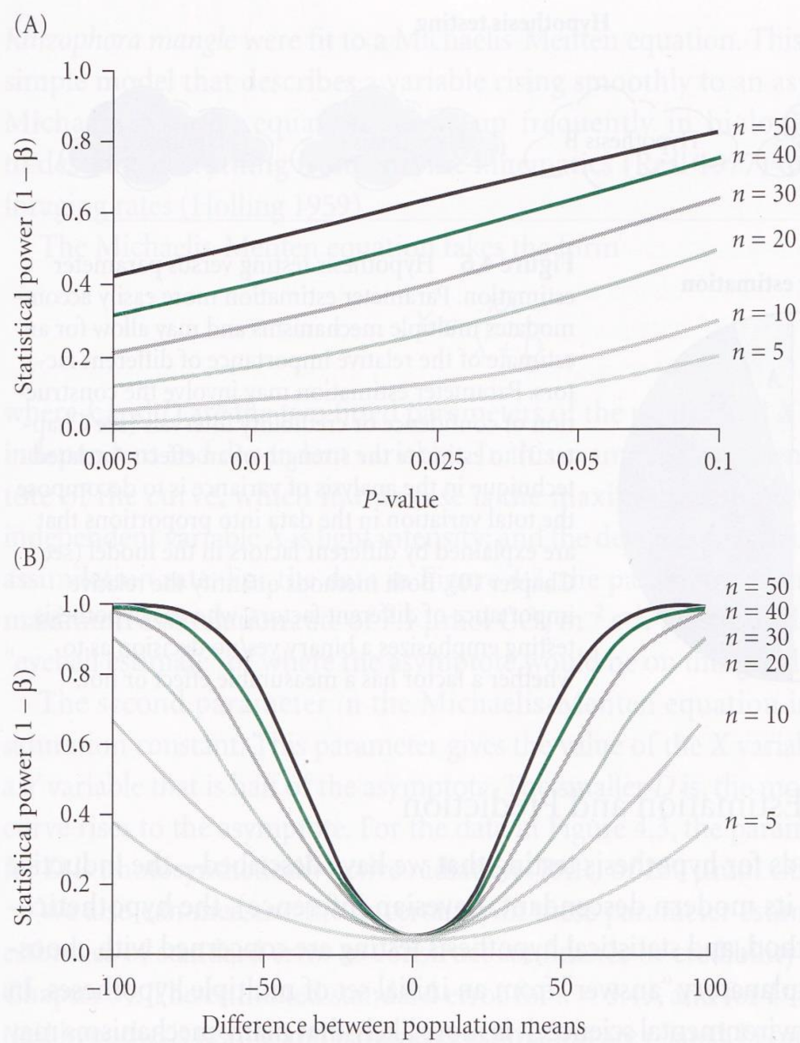


Figure 4.5 The relationship between statistical power, *P*-values, and observable effect sizes as a function of sample size. (A) The *P*-value is the probability of incorrectly rejecting a true null hypothesis, whereas statistical power is the probability of correctly rejecting a false null hypothesis. The general result is that the lower the *P*-value used for rejection of the null hypothesis, the lower the statistical power of correctly detecting a treatment effect. At a given *P*-value, statistical power is greater when the sample size is larger. (B) The smaller the observable effect of the treatment (i.e., the smaller the difference between the treatment group and the control group), the larger the sample size necessary for good statistical power to detect a treatment effect.[21]

Summary

The Rules of Strong Inference

- Strong inference looks at scientific problems by rejecting some hypotheses and accepting those remaining hypotheses that cannot be rejected.

- Using Strong Inference entails these steps:
    - Devise multiple, exhaustive hypotheses
    - Design experiment(s) to eliminate hypotheses
    - Carry out experiment(s) to get clean results
    - Repeat the cycle as many times as necessary
    - Refine hypotheses, devise sub-hypotheses

Summary

- Strong inference requires:

formulation of mutually exclusive and exhaustive hypotheses with quantifiable predictions

and experiments to rule out all but one of them

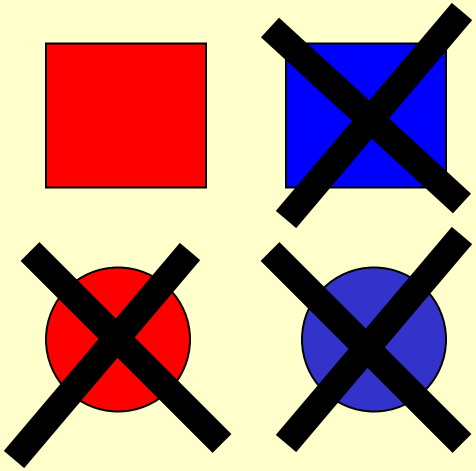
POSSIBLE HYPOTHESES

- Conclusive experiments are based on testing clear predictions from alternative hypotheses

Summary

- Conceptual models help us perform experiments by matching predictions with observations

HYPOTHESES



EXPERIMENTS

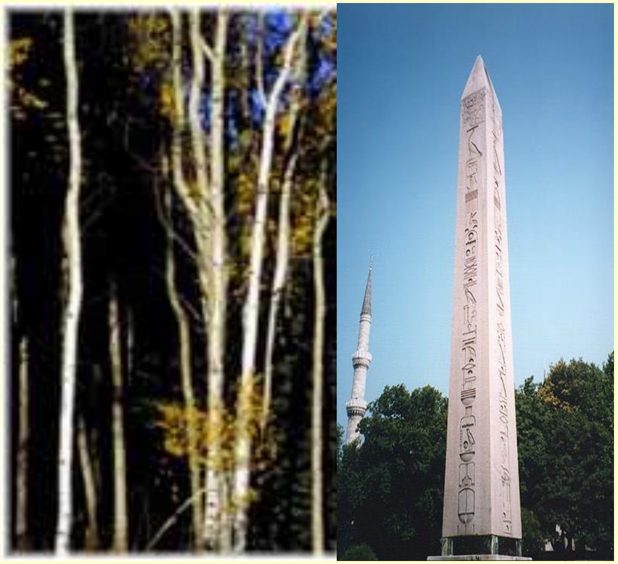
Shape = Square

Color = Red

Summary

How Science Is Done

- The experimental enterprise is cumulative
- New experiments build on old experiments; research is conducted as part of larger investigation program
- Beware of "pet branches"
- Do not grow fond of any hypotheses / models
- Do not be scared to challenge scientific paradigms



Summary

Working Hypotheses

Null Hypothesis (Ho) states observed variation due to sampling. (All samples belong to same population)

Alternative Hypothesis (Ha) states there is a significant pattern: non-random association or response. (Samples do not belong to same population)

Testing Hypotheses

Hypotheses generate predictions

Hypotheses tested with observations / experiments (repeatable, reproducible, controlled)

Summary

Assessing Significance

P-value is the probability that the observed pattern is the result of chance ; merely due to sampling

(Probability of obtaining data if Ho is, in fact, true)

Significance level (alpha) : set at 0.05 (1/20)

Type I Error: reject true Ho

Type II error: accept false Ho

Power affects the balance between these errors

Summary

Hypothesis Testing and the Scientific Method

If Ho accepted, it may not explain the phenomenon under study. There may be a better explanation.

If Ho rejected by test of available data, we are left with the Alternative Hypothesis.

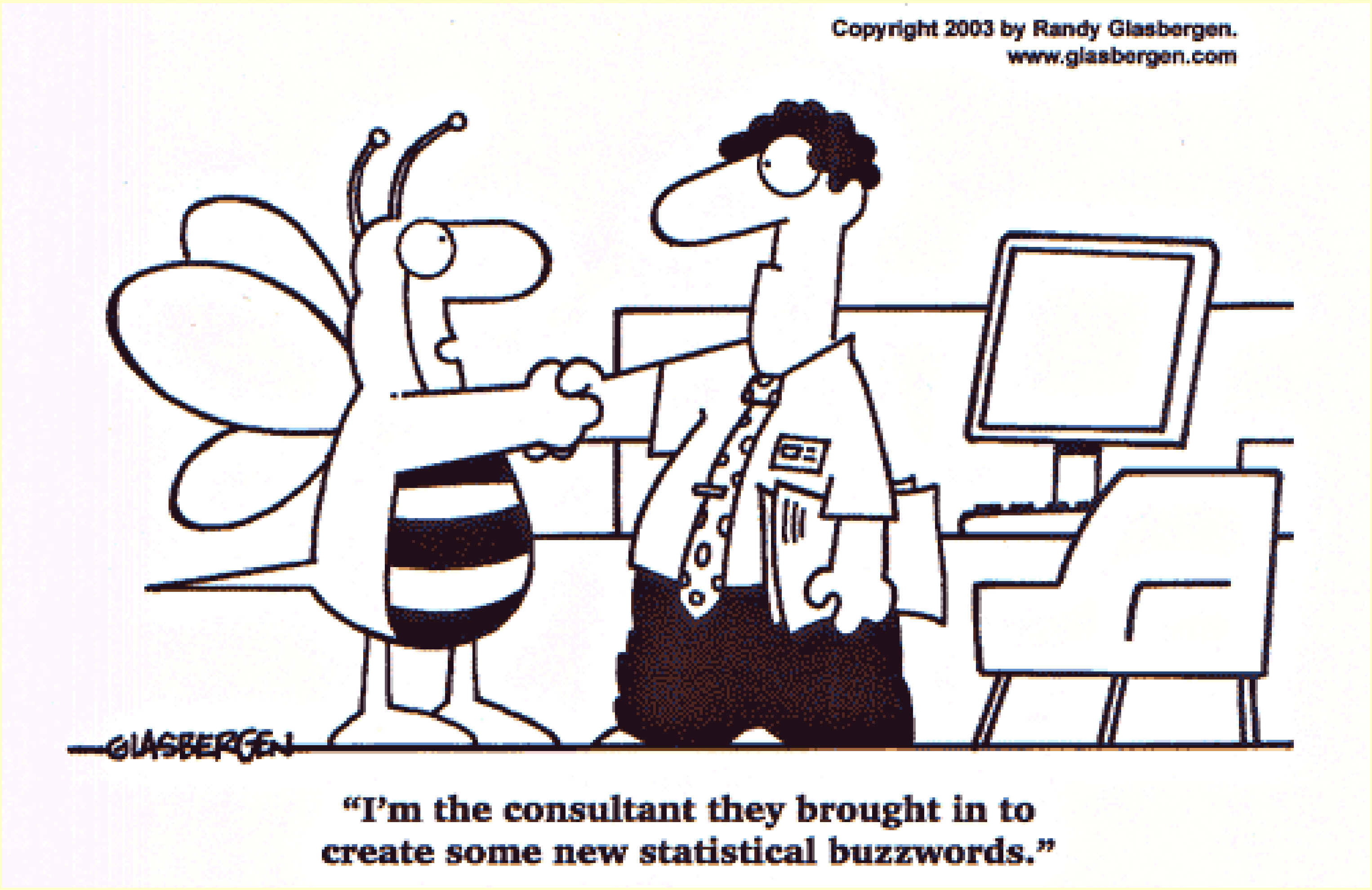
So, basically, the Null Hypothesis cannot be proved. It can only be disproved or rejected by the evidence.

Unrefuted hypotheses are strengthened.

Eventually, they become "laws" of science.

Summary

Expect you to Understand and Define Key Terms



Readings

Required

Platt, J.R. 1964. Strong Inference. Science 146: 347-353

Quinn J.F., Dunham A.E. 1983. On hypothesis testing in ecology and evolution. The American Naturalist 122: 602-617

Optional

O'Donohue, W., & Buchanan, J.A. 2001. The weaknesses of strong inference. Behavior & Philosophy 29: 1-20