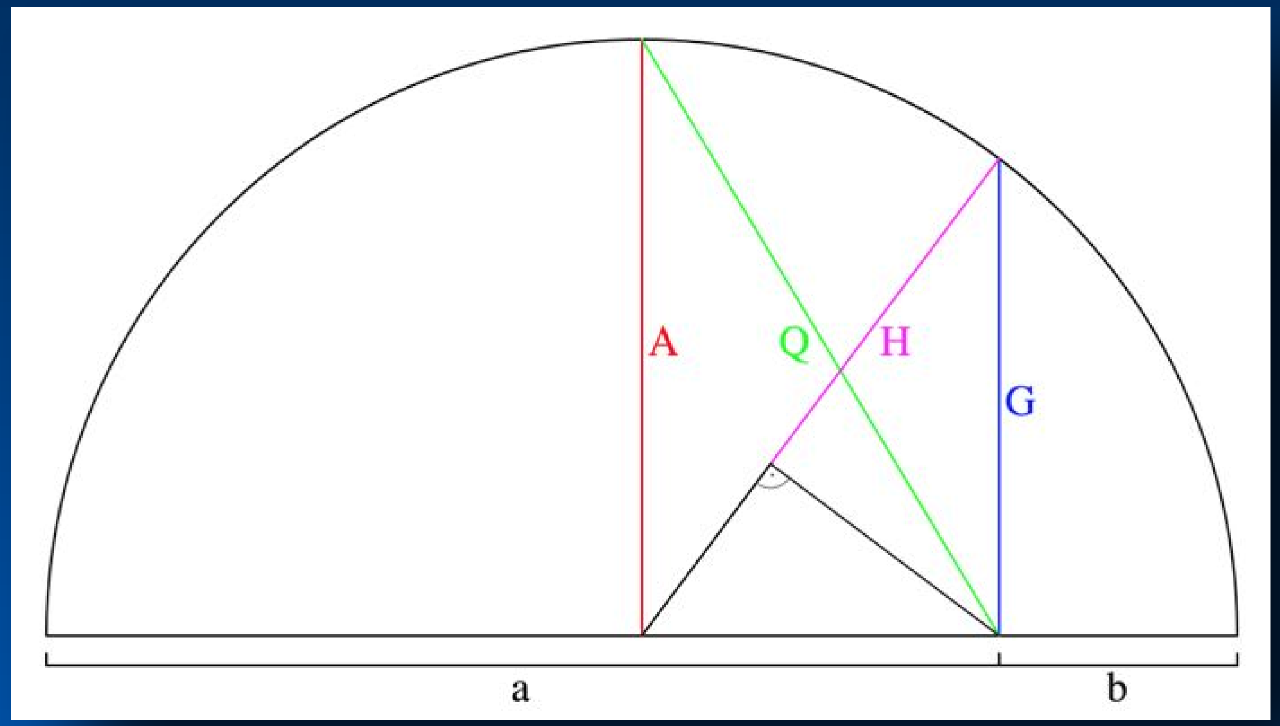
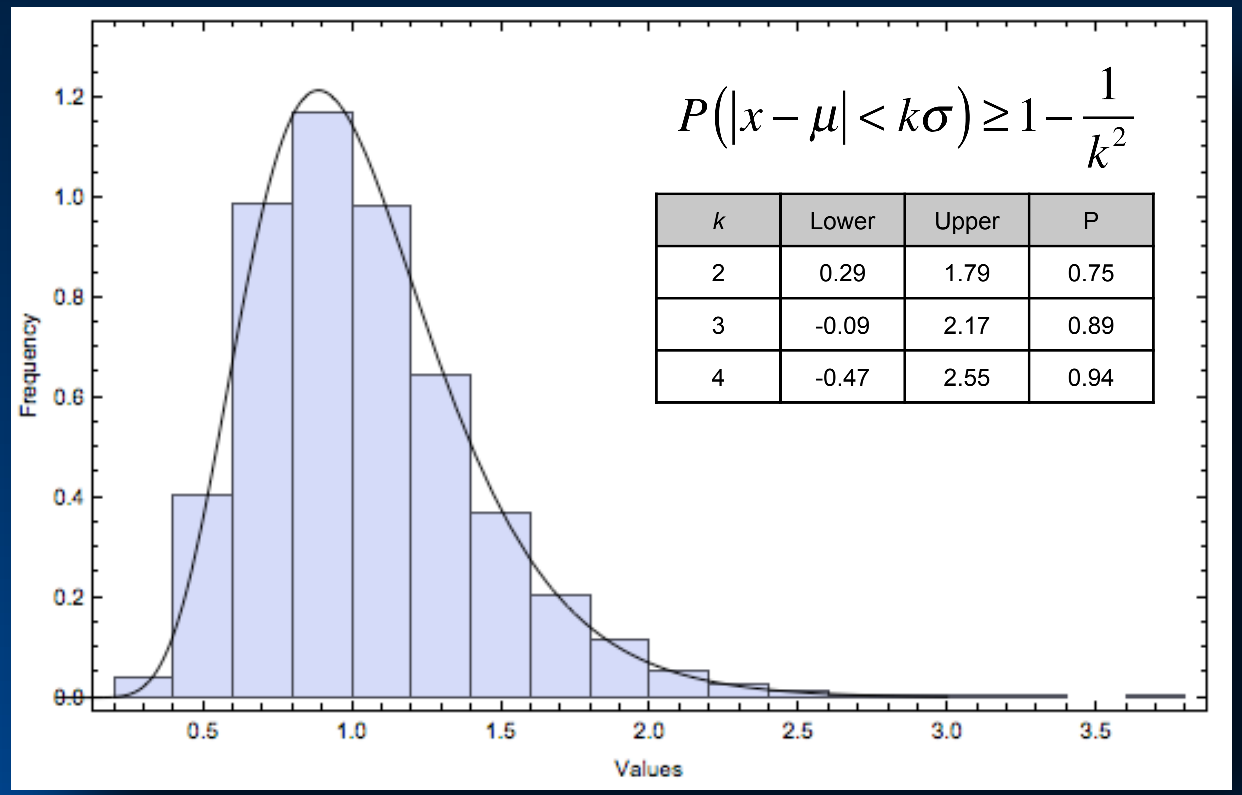
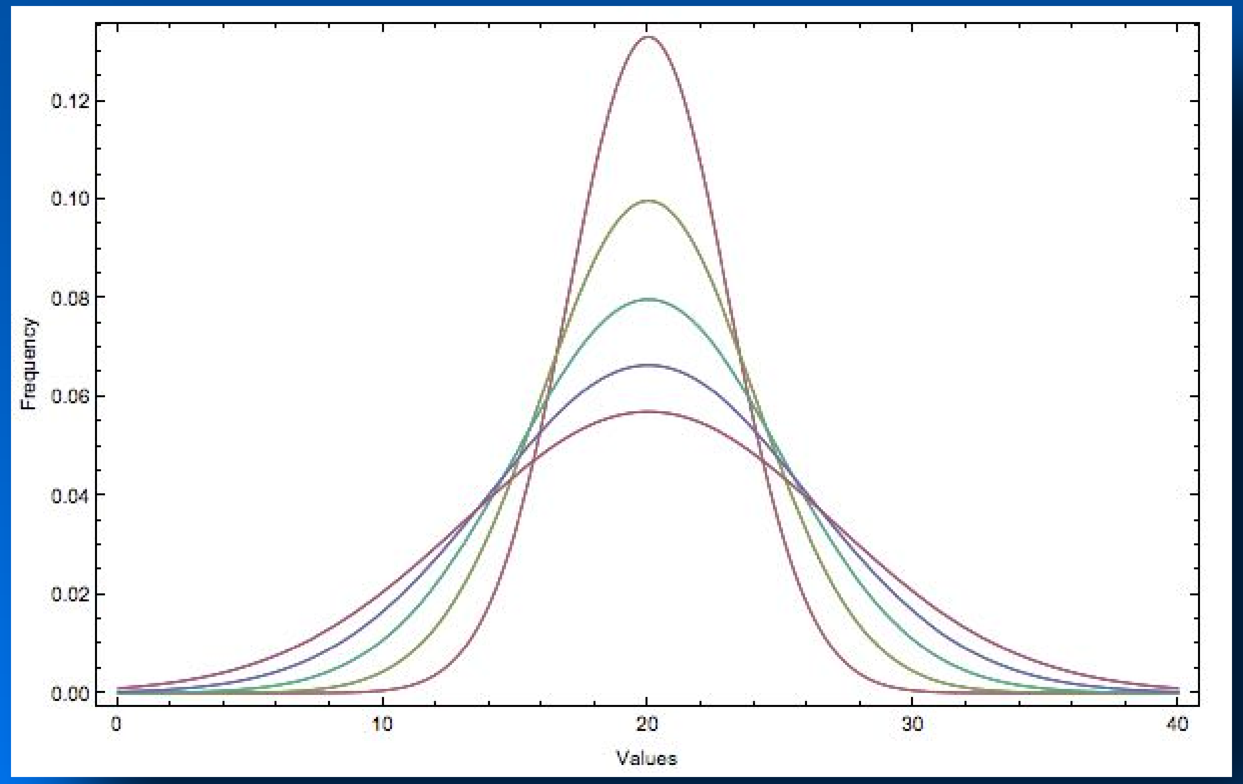
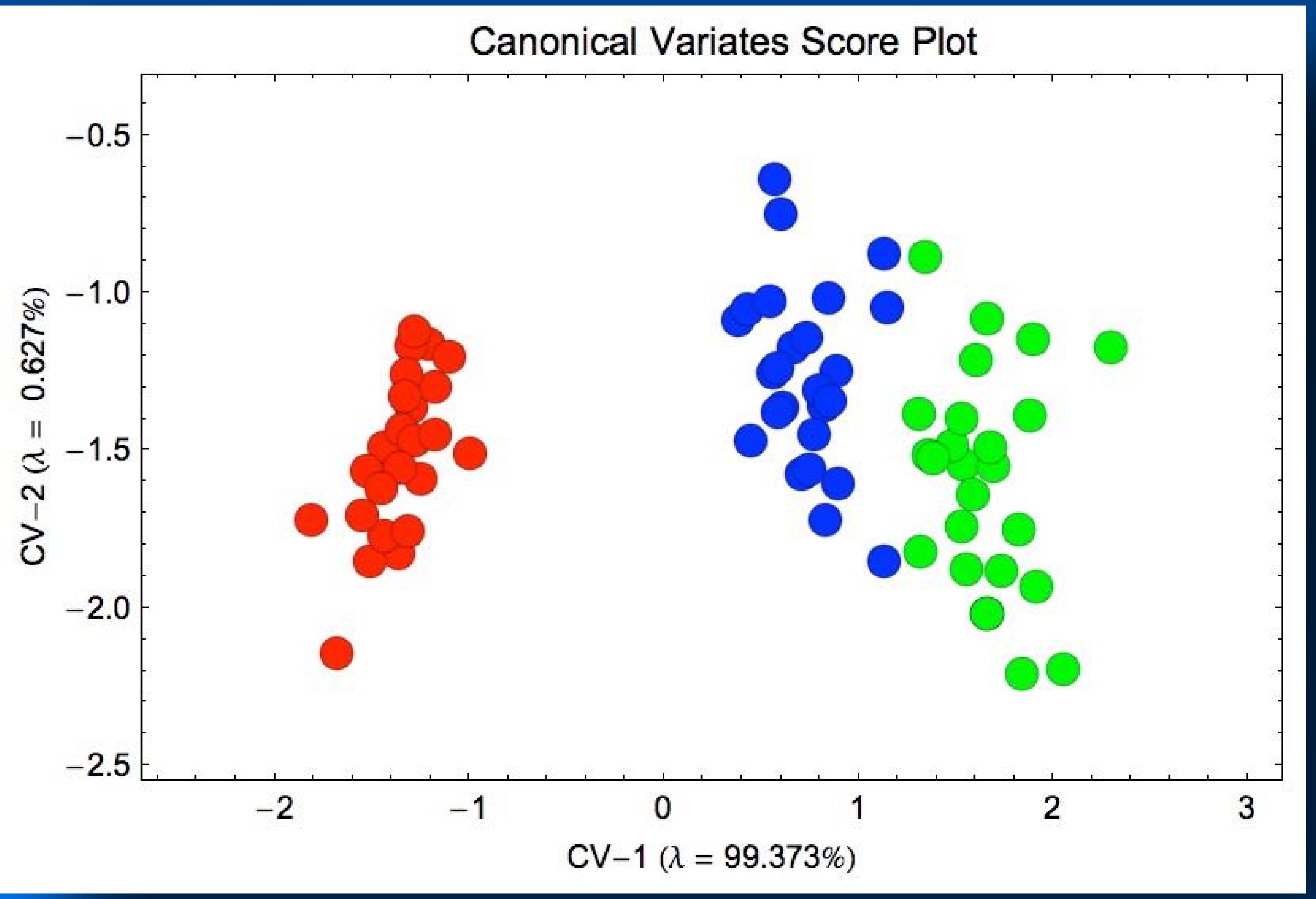
Statistics & Probability for Earth Scientists - A Review?

Prof. Norman MacLeod
School of Earth Sciences & Engineering, Nanjing University

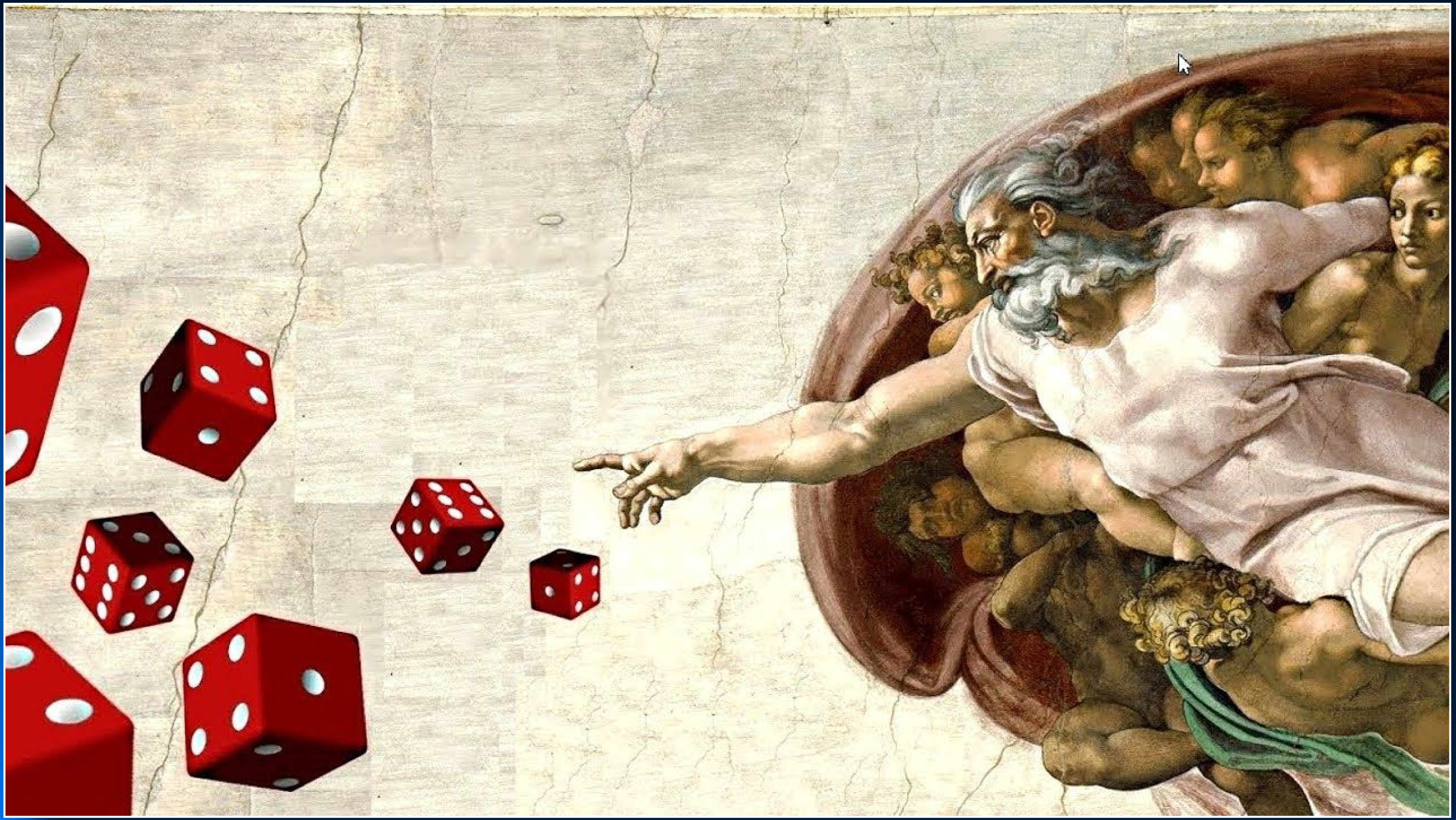




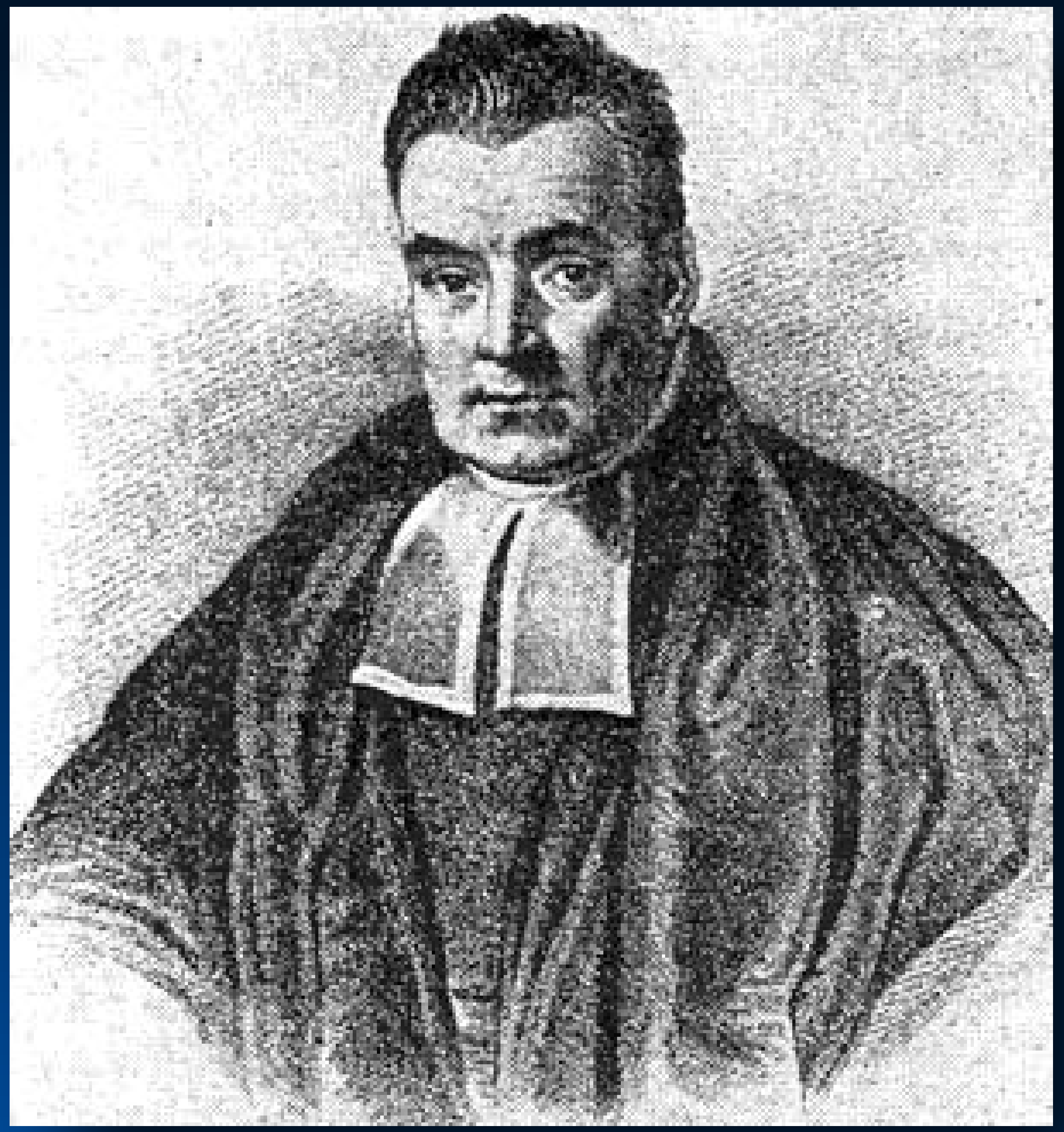




Bayesian Inference



Bayesian Inference



Rev. Thomas Bayes
(1701–1761)

In 1763 an essay written by Rev. Thomas Bayes, and entitled *An Essay Towards Solving a Problem in the Doctrine of Chances*, was read to the Royal Society which set out the principles of what has come to be known as Bayesian inference. This assumes evidence or observations can be used to calculate the probability that a hypothesis may be true, or to update its previously calculated (prior) probability. Computationally this result comes from a combination of the inherent likelihood (or prior probability) of the hypothesis and the compatibility of the observed evidence with the hypothesis.

The Bayesian approach to inference contrasts with the standard — or frequentist — approach insofar as it allows probabilities to be associated with unknown parameters such that they can sometimes have a frequency probability and interpretation of their own. These probabilities are usually interpreted as representing the analyst's prior belief that given values of the parameter are true. Unknown, or prior, probabilities are then updated via the examination of new evidence.

Statistical Inference

Frequentist Inference - measures the probability that a given observation or set of observations fits a statistical model.

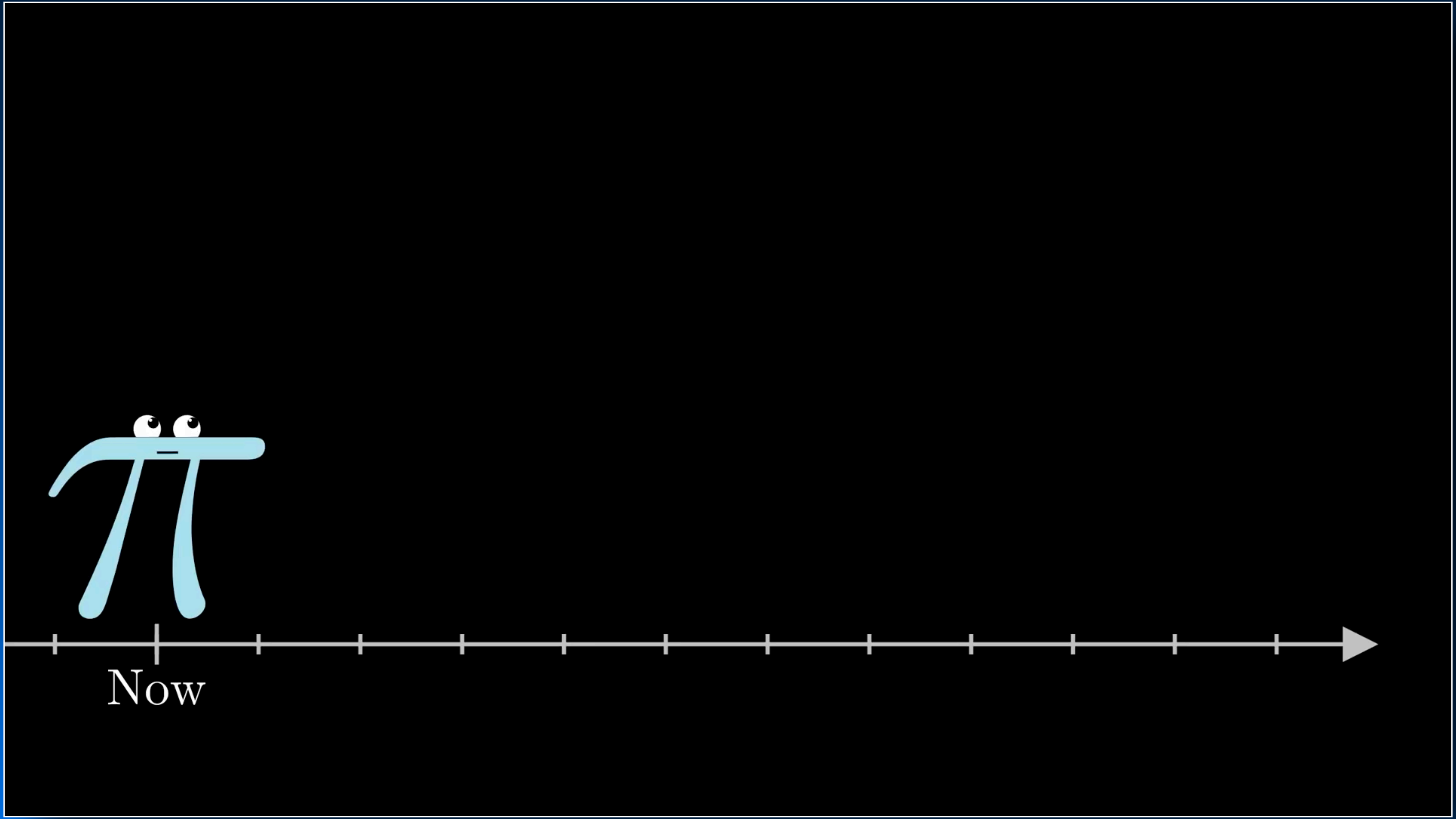
$$P(E|H)$$

Bayesian Inference - measures the likelihood that a given statistical model fits an observation or set of observations.

$$P(H|E)$$

Statistical Inference

Bayes Theorem



Bayes Theorem

$$P(H \mid E) = \frac{P(H) \cdot P(E \mid H)}{P(H) \cdot P(E \mid H) + P(\neg H) \cdot P(E \mid \neg H)}$$

$P(H)$ - Prior probability; probability the hypothesis *(H)* is true (before seeing any evidence *(E)*).

$P(E \mid H)$ - Likelihood; Probability of seeing the evidence *(E)* if the hypothesis *(H)* is true.

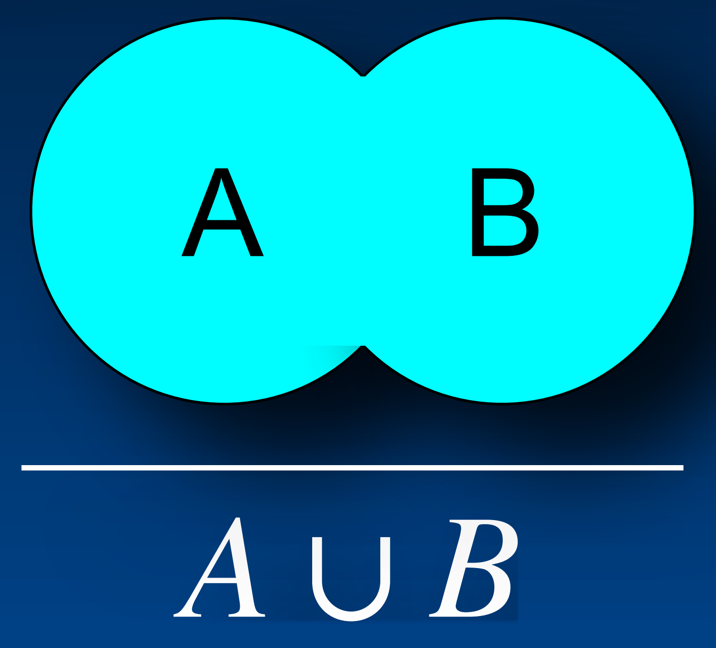
$P(\neg H))$ - Probability of the hypothesis *(H)* being false.

$P(E \mid \neg H)$ - Probability of seeing evidence the hypothesis *(H)* is false.

$P(H \mid E)$ - Posterior probability; probability the hypothesis *(H)* is true after examination of evidence *(E)*.

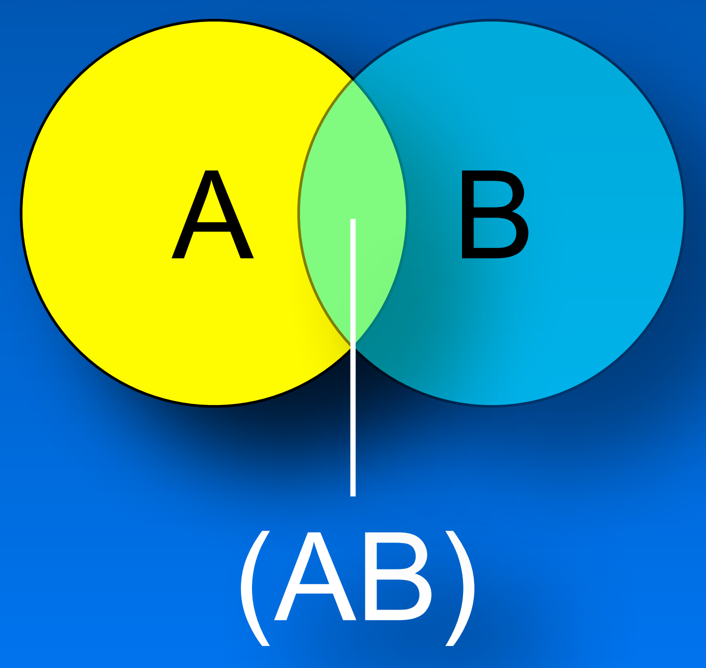
Combining Statistical Inferences

The Additive Law - given two probabilities drawn from the same population, their union is equal to their sum.



$$P(A \cup B) = P(A) + P(B)$$

The Multiplicative Law - given two probabilities drawn from the same population, their intersection is equal to their product.

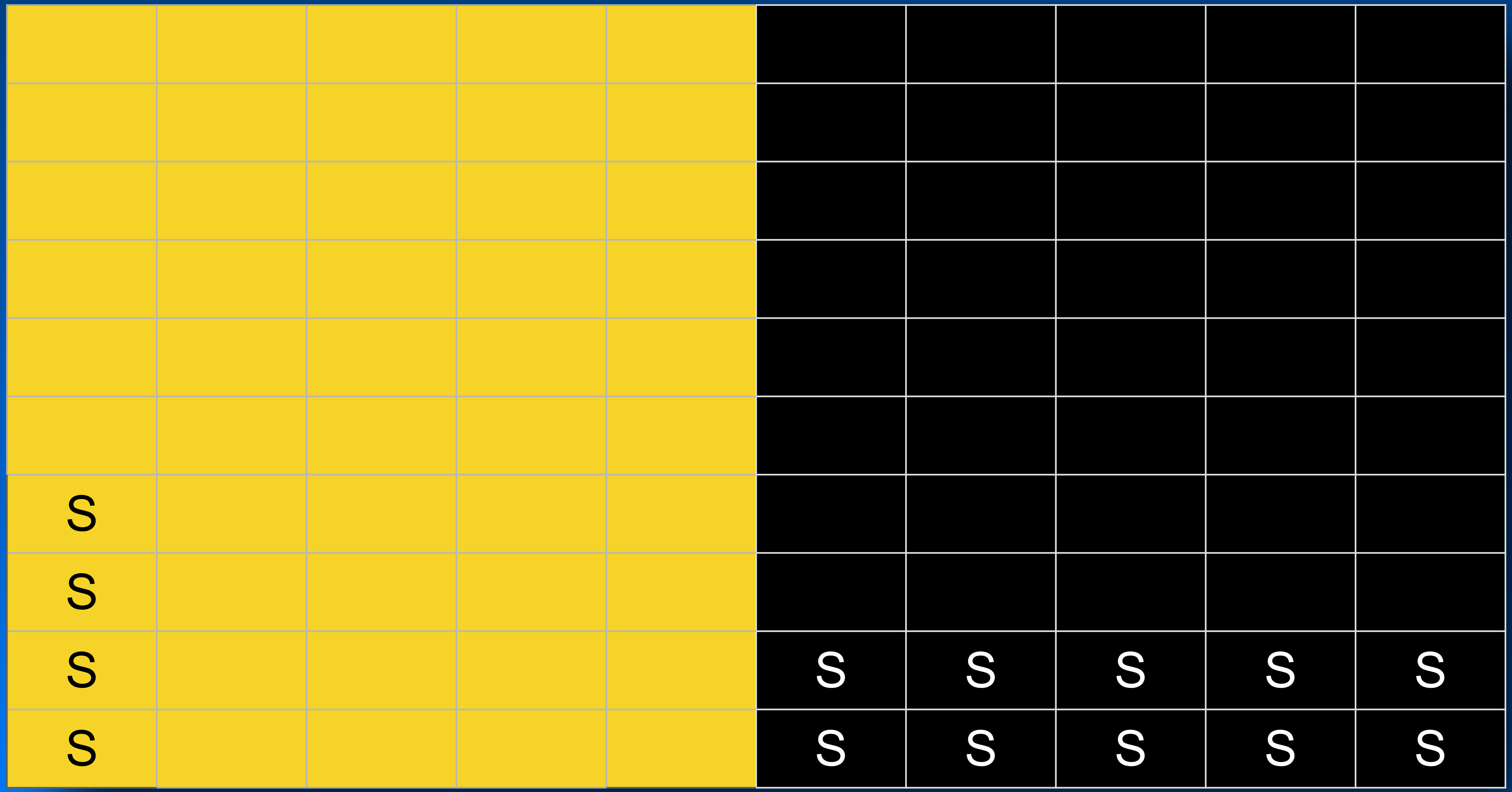


$$P(AB) = P(A) \cdot P(B)$$

Bayesian Inference

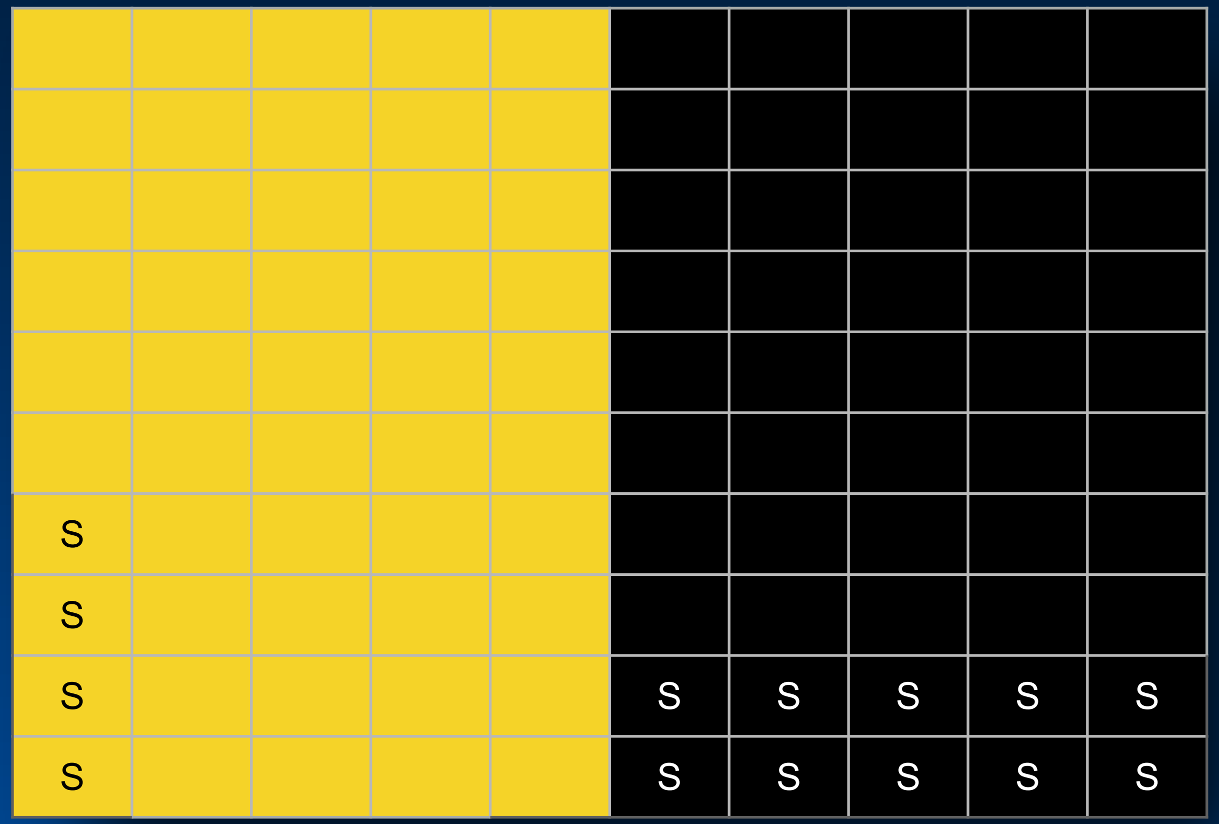
Bayes Theorem: Example

We are point-counting a petrographic thin section composed predominantly of yellow and black grains in approximately equal proportion. Based on a prior analysis of 100 grains we know 12 grains were spinel, but that 4 of these were yellow and 8 were black. What is the probability of a spinel grain being yellow?



Bayesian Inference

Bayes Theorem: Example



$P(H)$ = P(yellow) = 50/100 = 0.50

$P(E \mid H)$ = P(spinel|yellow) = 4/50 = 0.08

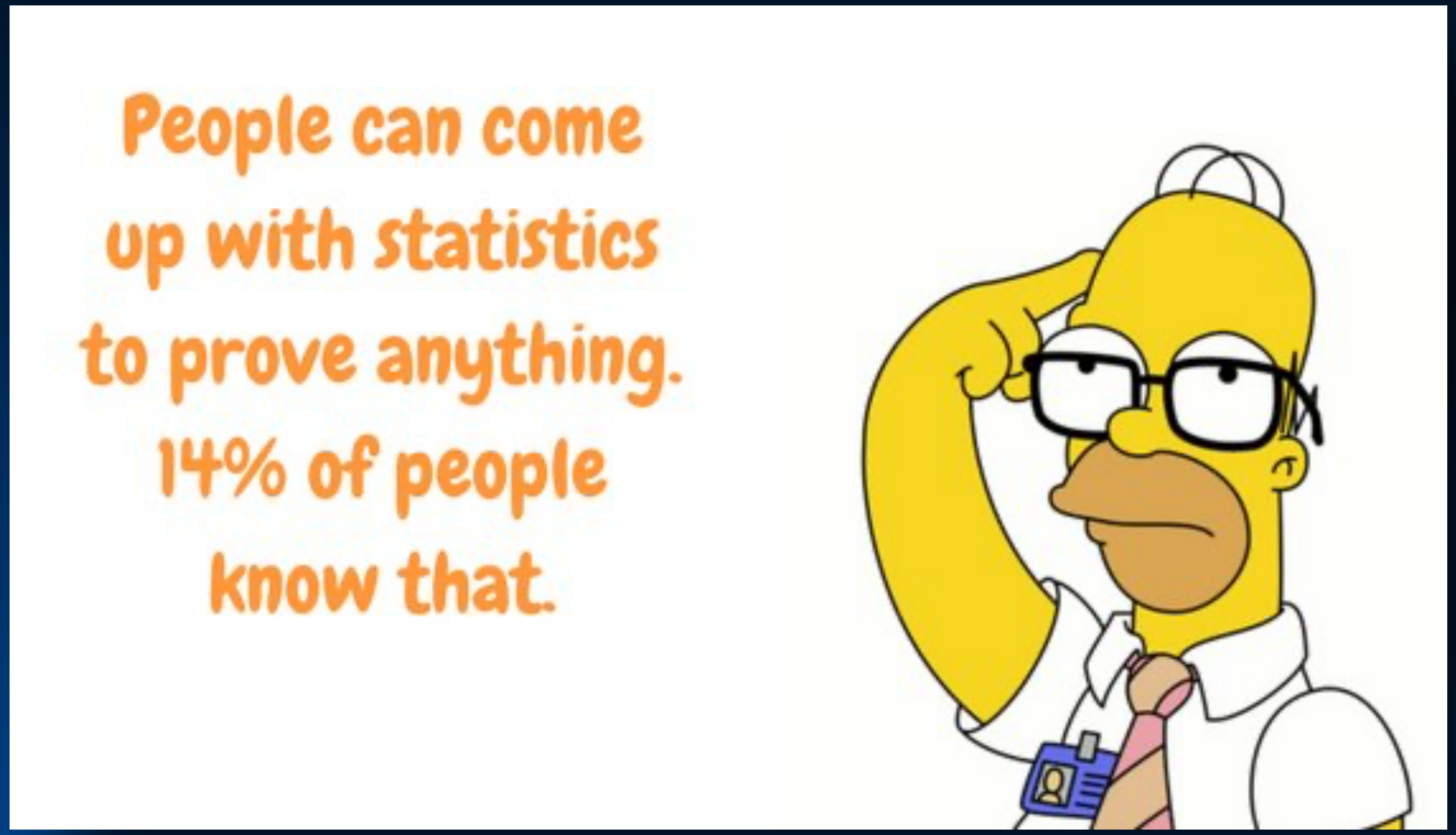
$P(\neg H)$ = P(black) = 50/100 = 0.50

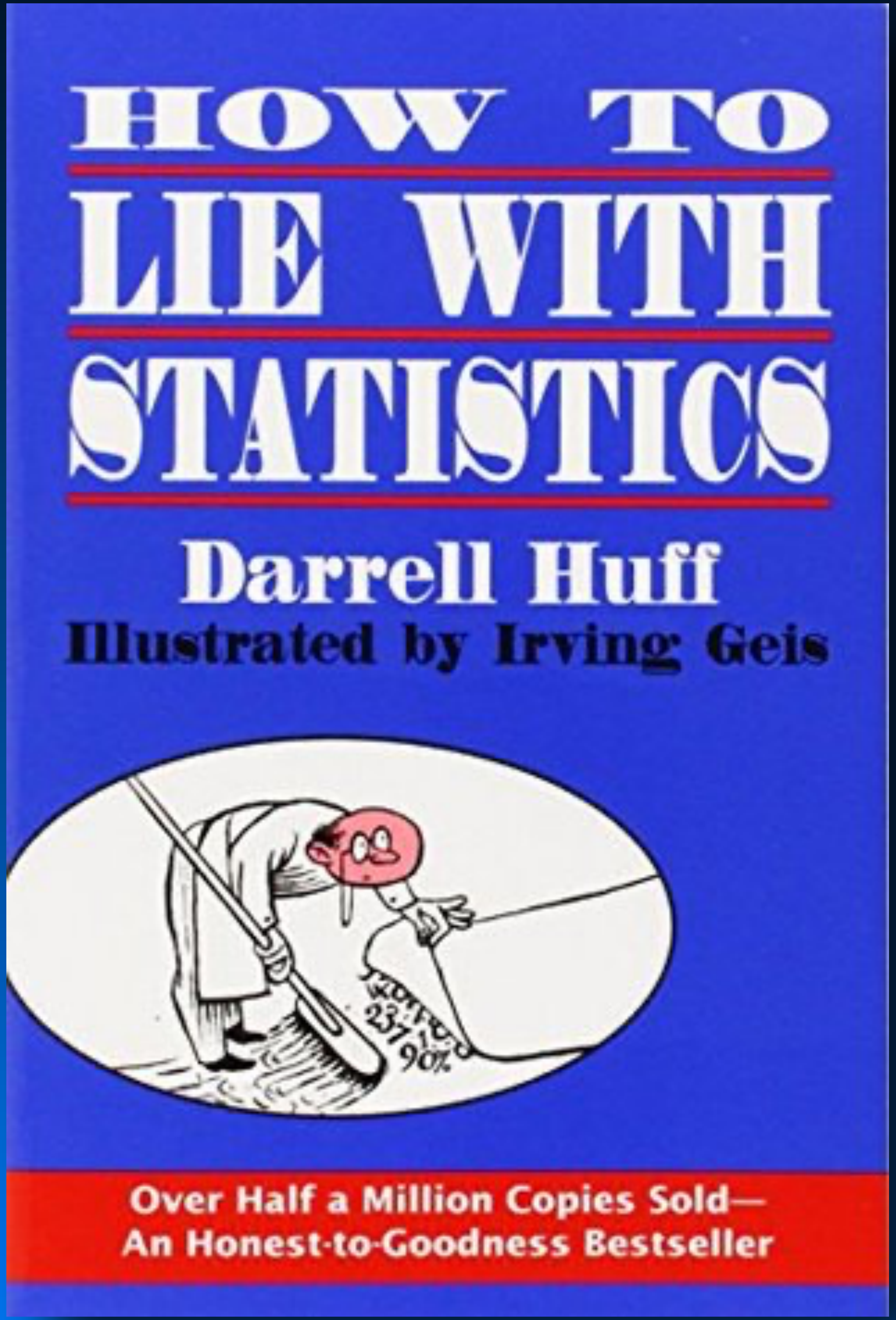
$P(E \mid \neg H)$ = P(spinel|black) = 10/50 = 0.20

$$P(H \mid E) = \frac{P(H) \cdot P(E \mid H)}{P(H) \cdot P(E \mid H) + P(\neg H) \cdot P(E \mid \neg H)}$$

$$P(H \mid E) = \frac{0.50 \cdot 0.08}{(0.50 \cdot 0.08) + (0.50 \cdot 0.20)} = \frac{0.04}{0.14} = 0.2857$$

Statistics as Hypothesis Testing





Statistical Inference

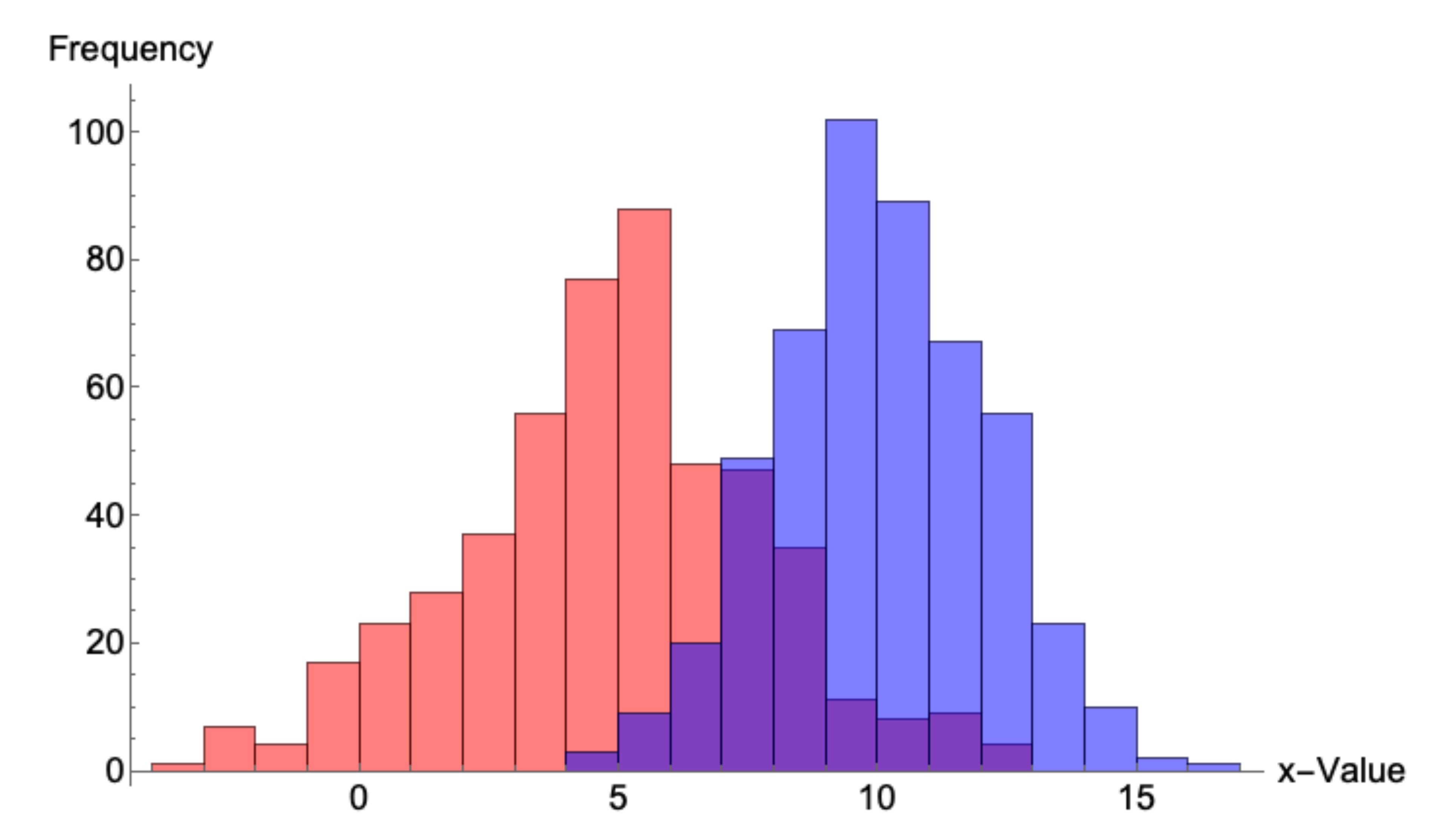
Frequentist vs Bayesian Interpretation

Frequentist Interpretation - Often regarded as the data-analysis standard, but can be too inflexible and polarized. If prior and new observations are drawn from samples of different populations it's difficult to know how to combine them.

Bayesian Interpretation - Very successful at focusing analyses on what really matters and emphasizing the importance of data. But some regard its results as somewhat speculative. It incorporates more information and allows inferences to be made from indirect data. However, Bayesian results are often counterintuitive.

Inferential Statistics

Quite aside from statistic's use as set of effective ways to describe sets of numbers and make certain inferences about probability, statistics represents one of the most effective means available for determining whether a wide range of statements, assumptions and/or assertions are likely to be true in the face of limited and/or uncertain data. Its ability to test hypotheses distinguishes this aspect of inferential statistics from data analysis.

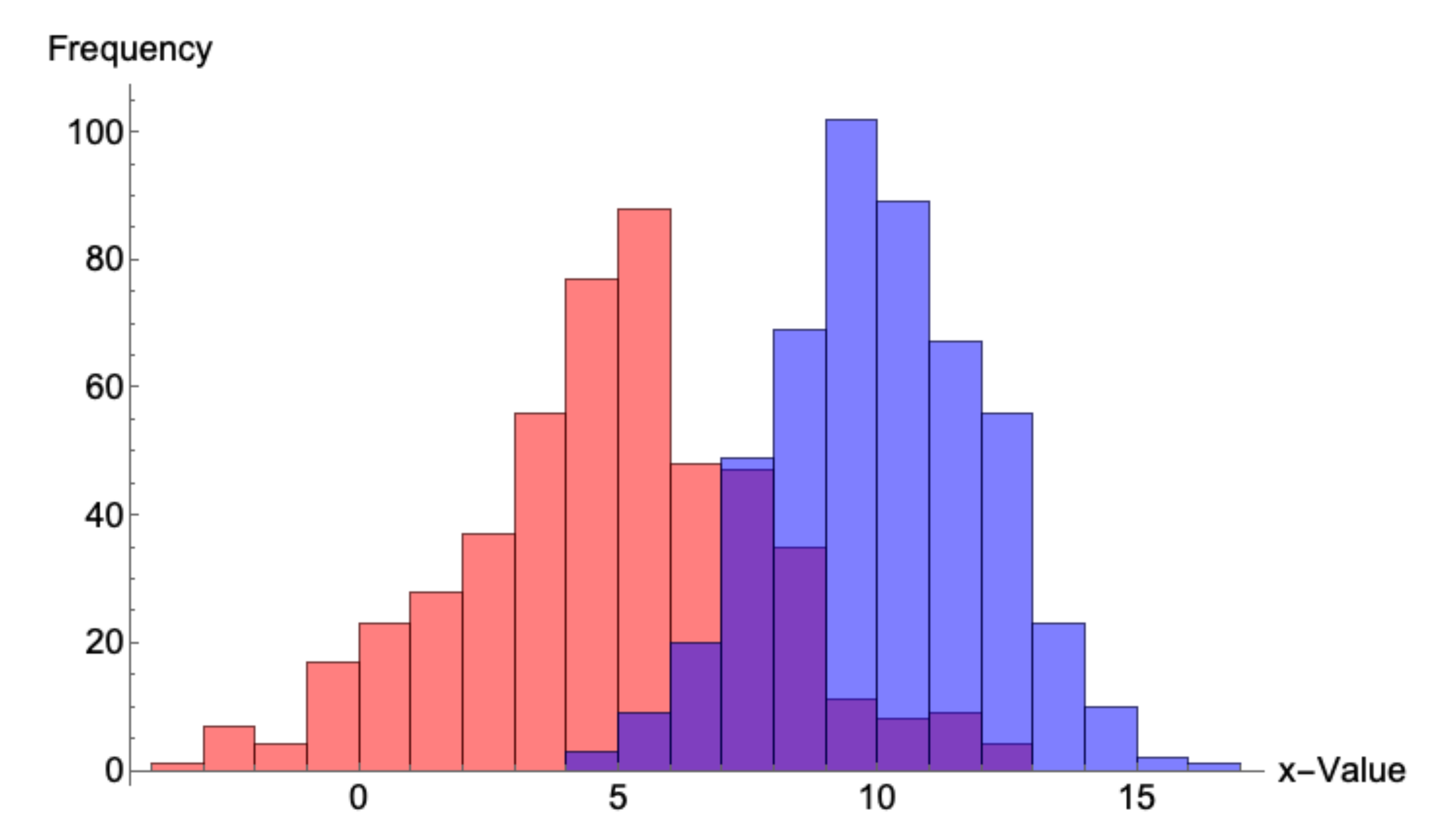


Data Analysis - Numerical procedures designed to summarize and detect patterns in data.

Inferential Statistics - Numerical procedures designed to determine the meaning of patterns detected in data.

Statistics as Hypothesis Testing

Two 500-value datasets



Parameter	Data Set 1 (Red)	Data Set 2 (Blue)
n	500	500
Mean	4.88	10.53
Variance	8.49	4.15

Are these datasets drawn from the same population?

Statistics as Hypothesis Testing

The Logic of Statistical Hypothesis Testing

- Hypotheses are considered in terms of acceptance of the hypothesis of no change, or the ‘null hypothesis’ (H_0).
- Acceptance of the null hypothesis means the observed difference(s) can be accounted for as a result of random sampling error.
- Rejection of the null hypotheses means the observed difference(s) cannot be accounted for by a simple appeal to random sampling error.
- However, whether this rejection implies acceptance of an alternative hypothesis (H_A) is controversial.

Statistics as Hypothesis Testing

The Logic of Statistical Hypothesis Testing

There exist two philosophical points-of-view with respect to the correct interpretation of inferential statistical results.

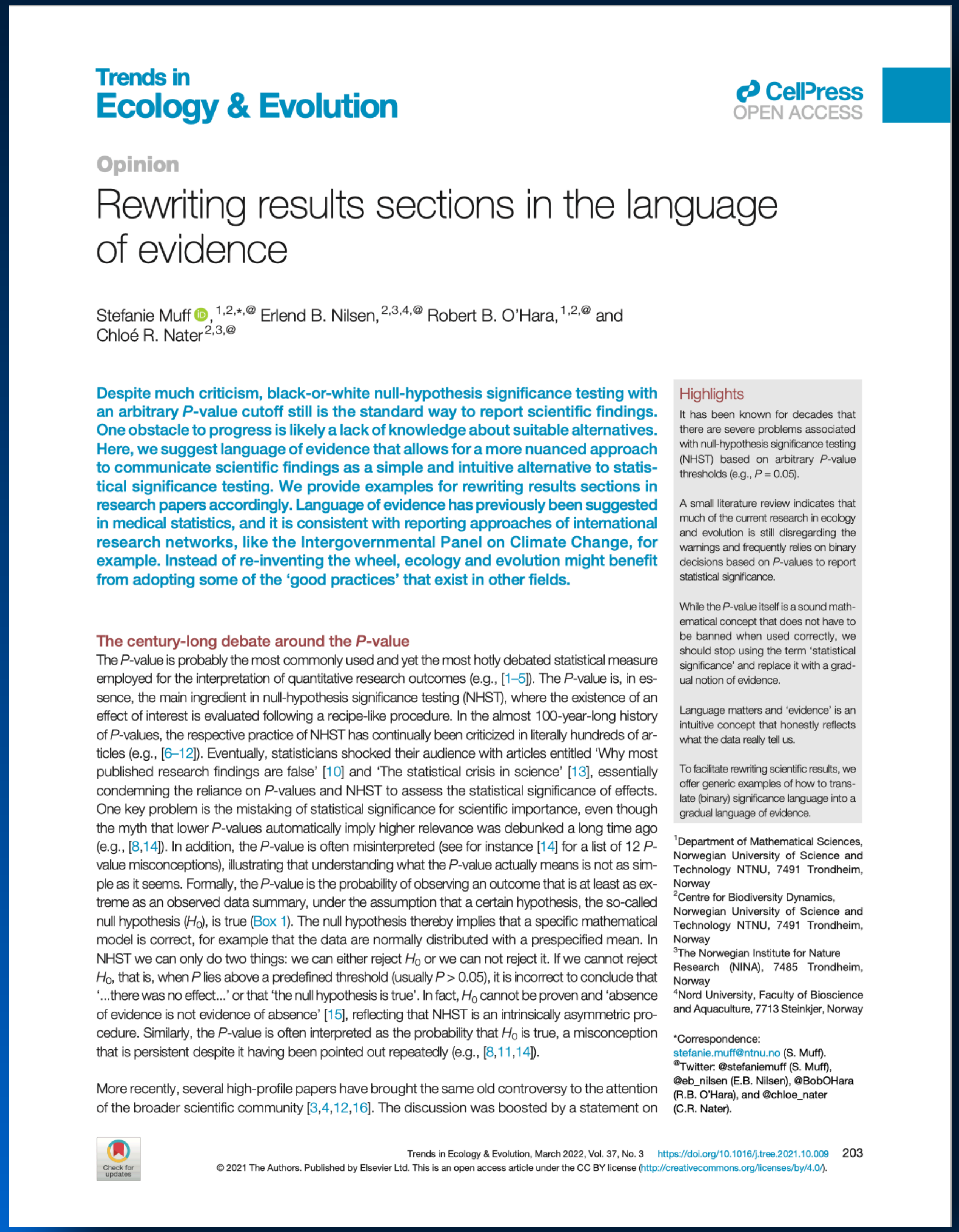
- Rejection of the null hypothesis implies acceptance of an appropriately stated alternative hypothesis, at least provisionally, pending the collection of additional data and the performance of additional statistical tests.
- The null hypothesis can never be rejected formally, only assigned an appropriate probability or likelihood value. It is always up to the researcher to make decisions in light of the results of inferential statistical tests.

In effect, this controversy turns on one's acceptance or rejection that a critical *p*-value exists.

Statistics as Hypothesis Testing

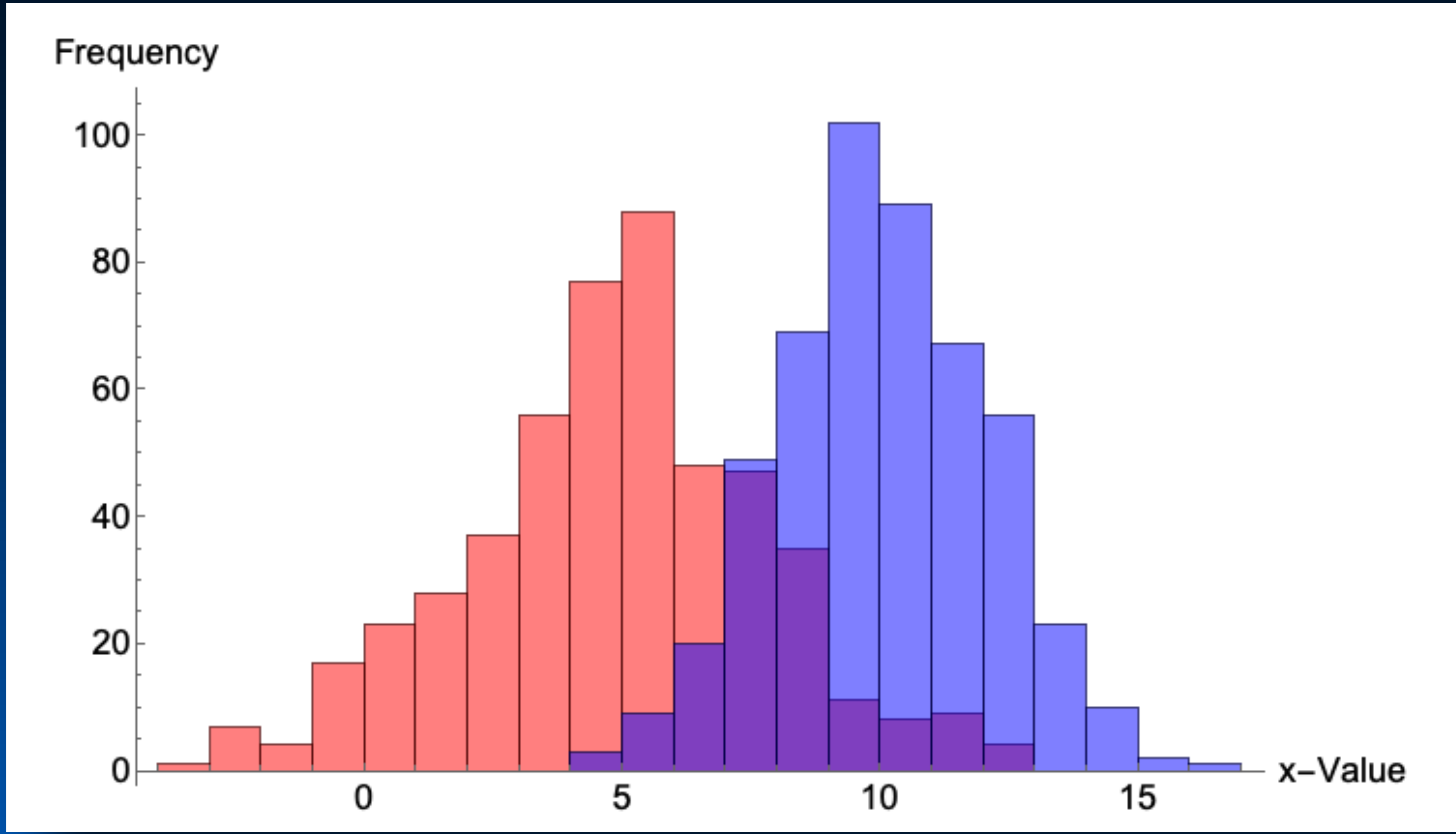
The Logic of Statistical Hypothesis Testing

Hartig, F., and Barraquand, F., 2022, The evidence contained in the P-value is context dependent: Trends in Ecology & Evolution, v. 37, p. 569–570, doi:10.1016/j.tree.2022.02.011.



Muff, S., Nilsen, E.B., Nater, C.R., and O'Hara, R.B., 2022, Joint reply to "Rewriting results in the language of compatibility" Trends in Ecology & Evolution, v. 37, p. 571–572, doi:10.1016/j.tree.2022.03.007.

Statistics as Hypothesis Testing



H_0: No significant difference between the locations of the means relative to the dispersion of data about the means.

H_A: Significant difference between the locations of the means relative to the dispersion of data about the means.

Statistics as Hypothesis Testing

Two-Sample Students *t*-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_e} \qquad s_e = s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$s_p^2 = \frac{((n_1 - 1) \cdot s_1^2) + ((n_2 - 1) \cdot s_2^2)}{n_1 + n_2 - 2}$$

Where: $\bar{x}$ = mean of samples 1 and 2;
s^2 = variances of samples 1 and 2;
n = sizes and samples 1 and 2;
s_e = standard error;
s_p^2 = pooled standard deviation.

Statistics as Hypothesis Testing

Two-Sample Students *t*-Test

$$t = \frac{10.05 - 4.88}{0.16} = 32.55$$

$$s_e = 2.51 \cdot \sqrt{\frac{1}{500} + \frac{1}{500}} = 0.16$$

$$s_p^2 = \frac{((500 - 1) \cdot 8.49) + ((500 - 1) \cdot 4.15)}{500 + 500 - 2} = 6.32$$

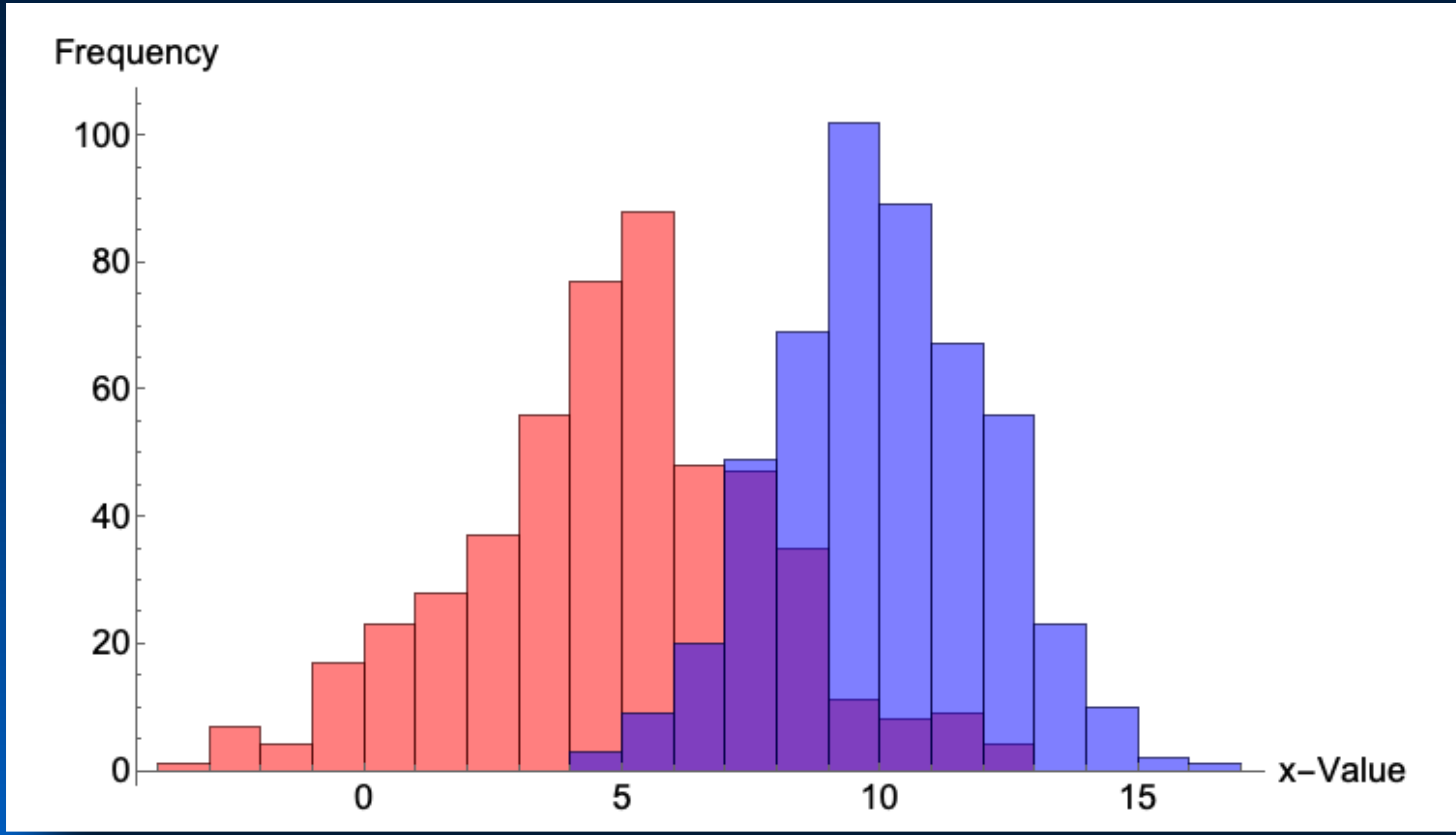
t = 32.55

$p = 7.91 \times 10^{-154}$

Very, very, very small probability of observing a difference this large.

Statistics as Hypothesis Testing

Two-Sample Students *t*-Test Assumptions



- Both samples are drawn from normally distributed populations.
- Samples have equivalent variances.*

Types of Statistical Tests

Parametric Statistical Tests - statistical tests which assume the data have come from a population whose frequency distribution is known. The term ‘parametric’ is used because the equation describing the shape of the distribution is assumed to have a fixed number of parameters.

Sampling Assumptions

Parametric Statistical Test Assumptions

- Specimens comprizing the sample must have been obtained independently and randomly (= equiprobable selection potential).
- Variables of interest must be normally distributed in the population(s) of interest.
- If multiple variables or multiple populations are being compared each must have equivalent variances (= homeoscedasticity).
- The means of the variables/populations must be linear combinations of effects on specimens (rows) and/or variables (columns).

Distributional Assumptions

- All parametric statistical tests (*z*-test, *t*-test, *F*-test) are only valid in a strict sense if the population variable(s) of interest is/are normally distributed in the population of interest.
- If the sample has been obtained correctly its distribution should reflect the distribution of the population.
- Whereas it is often claimed that certain parametric statistical tests are robust to deviations from distributional assumptions, no parametric test is robust to all deviations from its assumptions.
- A variety of alternative statistical tests are available for use with data known or suspected to deviate from normality.

What if Your Data Aren't Normal?

Parametric Statistical Tests - statistical tests which assume the data have come from a population whose frequency distribution is known. The term 'parametric' is used because the equation describing the shape of the distribution is assumed to have a fixed number of parameters.

Non-parametric Statistical Tests - statistical tests which do not assume that the data have come from a population whose frequency distribution is known. The term 'non-parametric' is used because the equation describing the shape of the distribution is assumed to have an unknown number of parameters.

Statistical Hypothesis Tests

- Parametric Statistical Tests - arrive at probability estimates via reference to a particular probability distribution (e.g., normal distribution, *t*-distribution, *F*-ratio distribution).
    - Significance estimates are accurate so long as distribution assumptions are correct.
- Non-parametric Statistical Tests - arrive at probability estimates via reference to any conceivable frequency distribution regardless of its shape.
    - Significance estimates may be less accurate, but are often better than parametric estimates based on incorrect assumptions.

Sampling Assumptions

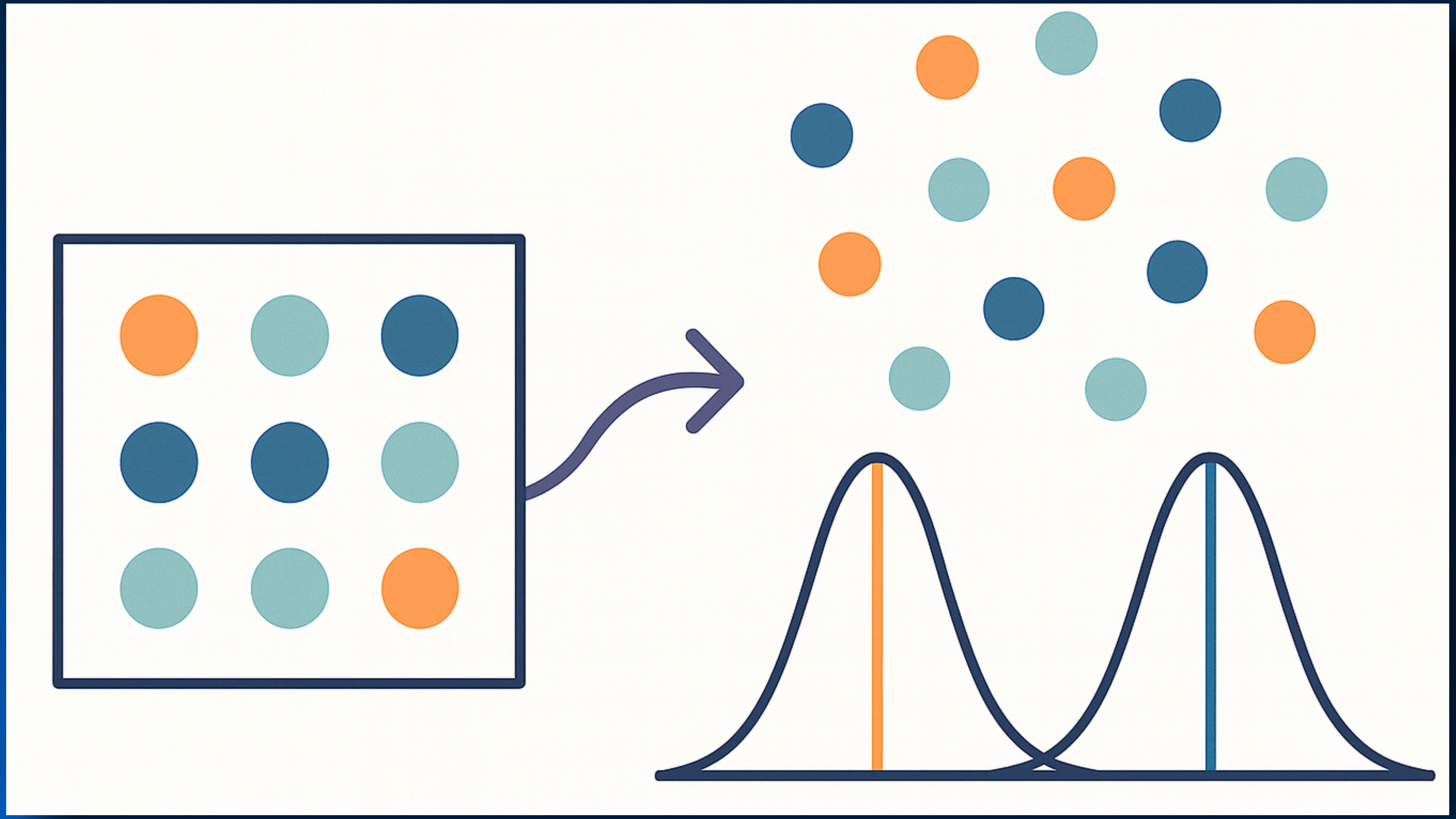
Non-Parametric Statistical Test Assumptions

- Specimens comprizing the sample must have been obtained independently and randomly (= equiprobable selection potential).
- Variables of interest must exhibit an underlying continuity.

Non-Parametric Statistical Tests

Non-parametric Forms of Parametric Tests

- Permutation Test - calculate all possible values of the test statistic under rearrangements of the labels of the observed data points. In other words, the method by which treatments are allocated to subjects in an experimental design is mirrored in the analysis of that design. If the labels are exchangeable under the null hypothesis, then the resulting tests yield exact significance levels.



Non-Parametric Statistical Tests

Non-parametric Forms of Parametric Tests

- **Monte Carlo Simulation** - generate the reference distribution by Monte Carlo sampling, which takes a small (relative to the total number of permutations) random sample of the possible replicates. This procedure is asymptotically equivalent to a permutation test and should be used when there are too many possible orderings of the data to allow complete enumeration in a convenient manner.

Non-Parametric Statistical Tests

Non-parametric Forms of Parametric Tests

- Jackknifing - systematically recomputing the statistic estimate leaving out one or more observations at a time from the sample set. From this new set of replicates of the statistic, an estimate for the bias and an estimate for the variance of the statistic can be calculated.

Non-Parametric Statistical Tests

Non-parametric Forms of Parametric Tests

- Bootstrapping - estimate the sampling distribution of statistic by sampling with replacement from the original sample.
  - Non-Parametric Bootstrapping - accept all pseudo-replicate samples assembled regardless of their distribution.
  - Parametric Bootstrapping - ensure each pseudo-replicate sample has the same mean and variance as the original sample.

How to Avoid Lying With Statistics

- Understand what question(s) you're trying to answer.
- Be suspicious of your sample. Is it representative? Do the data collected conform to a normal distribution?
- Think hard about what data you want/need to collect and why.
- Read about and understand the procedures and tests you are considering before you decide to use any. If you have questions, ask someone with more experience.
- Use robust tests that are designed to analyze the data you happen to have.
- Be suspicious of any results you obtain; use multiple tests, look for consistency.
- Do not succumb to the temptation to use statistics to try to 'produce' support for any particular hypothesis.

Statistics & Probability for Earth Scientists - A Review?

Prof. Norman MacLeod

School of Earth Sciences & Engineering, Nanjing University