Statistics

Marcio Augusto Diniz

Biostatistics Research Center
Samuel Oschin Comprehensive Cancer Institute
Cedars Sinai Medical Center

Summary

History

What is the origin of the word?

▶ Statistics is derived from the New Latin statisticum collegium (council of state) and the Italian word statista (statesman or politician);

▶ The German word Statistik, first introduced by Gottfried Achenwall (1749), originally designated the systematic collection of demographic and economic data by states, signifying the science of state (then called political arithmetic in English);

History

What is the origin of the word?

- ▶ Statistics is derived from the New Latin statisticum collegium (council of state) and the Italian word statista (statesman or politician);
- ▶ The German word Statistik, first introduced by Gottfried Achenwall (1749), originally designated the systematic collection of demographic and economic data by states, signifying the science of state (then called political arithmetic in English);
- ▶ Tabulations of human and material resources to be taxed or put to military use.

History

Origins

- The earliest writing on statistics was found in a 9th-century book entitled: "Manuscript on Deciphering Cryptographic Messages", written by Al-Kindi (801-873 CE), describing how to use statistics and frequency analysis to decipher encrypted messages.

History

Origins

- The earliest writing on statistics was found in a 9th-century book entitled: "Manuscript on Deciphering Cryptographic Messages", written by Al-Kindi (801-873 CE), describing how to use statistics and frequency analysis to decipher encrypted messages.

- It is often dated to 1662, William Petty developed early human statistical and census methods that provided a framework for modern demography. He produced the first life table, giving probabilities of survival to each age, and estimated the population of London.

History

Games of chance

▶ In the 16th century, Gerolamo Cardano, Pierre de Fermat and Blaise Pascal exchanged letters discussing games of chance. Jakob Bernoulli's Ars Conjectandi (posthumous, 1713) and Abraham de Moivre's The Doctrine of Chances (1718) treated the subject as a branch of mathematics.

The meeting of two worlds

▶ Adrien-Marie Legendre (1805), Robert Adrain (1808), and Carl Friedrich Gauss (1809) used probability and statistics to estimate the position of planets.

History

Fundamentals of Modern Statistics

▶ At the end of 19th century, Francis Galton and Karl Pearson transformed statistics into a rigorous mathematical discipline used for analysis, not just in science;

History

## Fundamentals of Modern Statistics

- At the end of 19th century, Francis Galton and Karl Pearson transformed statistics into a rigorous mathematical discipline used for analysis, not just in science;

- In the 1910s and 20s, William Sealy Gosset and Ronald Fisher developed better design of experiments, significance tests and other methods for small sample sizes;

History

Fundamentals of Modern Statistics

▶ At the end of 19th century, Francis Galton and Karl Pearson transformed statistics into a rigorous mathematical discipline used for analysis, not just in science;

▶ In the 1910s and 20s, William Sealy Gosset and Ronald Fisher developed better design of experiments, significance tests and other methods for small sample sizes;

▶ In the 1930s, Egon Pearson and Jerzy Neyman introduced the concepts of confidence intervals, test of hypothesis and power.

History
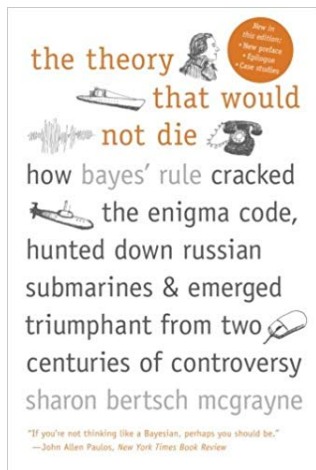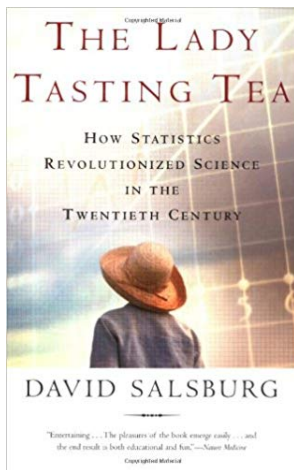
Which lake does have more fishes?
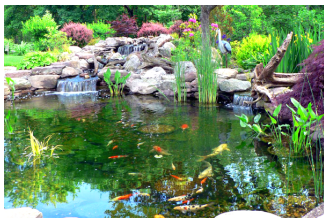
234 fishes

235 fishes

Which lake does have more fishes?

234 and 235 fishes sampled

Which lake does have more fishes?

▶ What is the difference between the first and second scenarios?

Which lake does have more fishes?

- ► What is the difference between the first and second scenarios?
  - ► In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.

Which lake does have more fishes?

▶ What is the difference between the first and second scenarios?
  ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
▶ What is the difference between census and sample?

Which lake does have more fishes?

▶ What is the difference between the first and second scenarios?
  ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
▶ What is the difference between census and sample?
  ▶ Uncertainty.

Which lake does have more fishes?

- ▶ What is the difference between the first and second scenarios?
  - ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
- ▶ What is the difference between census and sample?
  - ▶ Uncertainty.
- ▶ How do we handle uncertainty?

Which lake does have more fishes?

- ▶ What is the difference between the first and second scenarios?
  - ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
- ▶ What is the difference between census and sample?
  - ▶ Uncertainty.
- ▶ How do we handle uncertainty?
  - ▶ Identify sources of variability;

Which lake does have more fishes?

- ▶ What is the difference between the first and second scenarios?
  - ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
- ▶ What is the difference between census and sample?
  - ▶ Uncertainty.
- ▶ How do we handle uncertainty?
  - ▶ Identify sources of variability;
  - ▶ Minimize;

Which lake does have more fishes?

- ▶ What is the difference between the first and second scenarios?
  - ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
- ▶ What is the difference between census and sample?
  - ▶ Uncertainty.
- ▶ How do we handle uncertainty?
  - ▶ Identify sources of variability;
  - ▶ Minimize;
  - ▶ Measure;

Which lake does have more fishes?

▶ What is the difference between the first and second scenarios?
  ▶ In the first scenario, the total number of fishes is derived from a census, while it is from a sample in the second scenario.
▶ What is the difference between census and sample?
  ▶ Uncertainty.
▶ How do we handle uncertainty?
  ▶ Identify sources of variability;
  ▶ Minimize;
  ▶ Measure;
  ▶ Make decisions.

What is Statistics?

- American Statistical Association, funded in 1839: It is the science of learning from data, and of measuring, controlling and communicating uncertainty.
  - It is a **science**, therefore, statistical tools are invented and improved by statisticians;
  - It measures **uncertainty** allowing us to make inferences from a sample to a population.

Use of Statistics

► Does the new medication have any effect on tumor growth?

Use of Statistics

- ▶ Does the new medication have any effect on tumor growth?
- ▶ Is the treatment effect different in WT and PR-364 mouse model?

Use of Statistics

- Does the new medication have any effect on tumor growth?
- Is the treatment effect different in WT and PR-364 mouse model?
- Which genes are differently expressed between healthy and cancer patients?

Use of Statistics

- ▶ Does the new medication have any effect on tumor growth?
- ▶ Is the treatment effect different in WT and PR-364 mouse model?
- ▶ Which genes are differently expressed between healthy and cancer patients?
- ▶ What is the minimum effective dose of a drug combination to give to mice aiming to increase cancer survival?

Use of Statistics

- Does the new medication have any effect on tumor growth?
- Is the treatment effect different in WT and PR-364 mouse model?
- Which genes are differently expressed between healthy and cancer patients?
- What is the minimum effective dose of a drug combination to give to mice aiming to increase cancer survival?
- What is the normal range for a biomarker?

# Use of Statistics

- Does the new medication have any effect on tumor growth?
- Is the treatment effect different in WT and PR-364 mouse model?
- Which genes are differently expressed between healthy and cancer patients?
- What is the minimum effective dose of a drug combination to give to mice aiming to increase cancer survival?
- What is the normal range for a biomarker?
- Is the new biomarker useful to detect early stage cancer?

Use of Statistics

- ▶ Does the new medication have any effect on tumor growth?
- ▶ Is the treatment effect different in WT and PR-364 mouse model?
- ▶ Which genes are differently expressed between healthy and cancer patients?
- ▶ What is the minimum effective dose of a drug combination to give to mice aiming to increase cancer survival?
- ▶ What is the normal range for a biomarker?
- ▶ Is the new biomarker useful to detect early stage cancer?
- ▶ Which patients' baseline characteristics can predict disease free survival at 5 years?

When should we talk to a statistician?
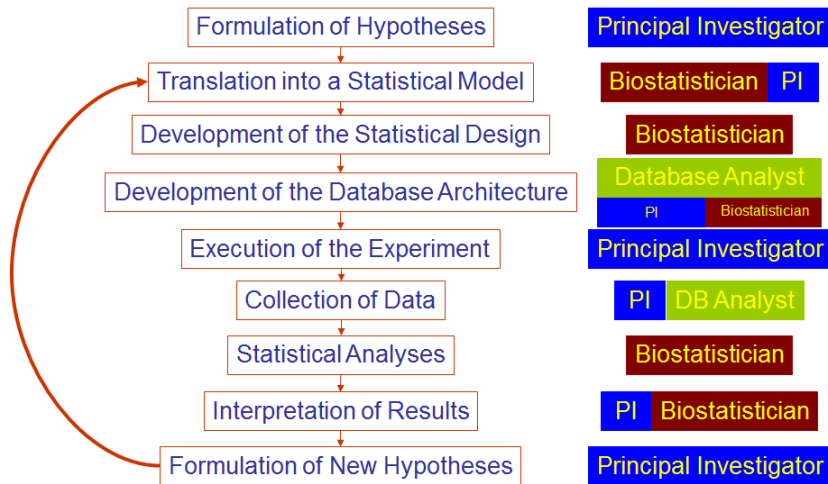
Steps of Research

Cedars Sinai Medical Center

Biostatistics Core

- ▶ Request Form: http://www.cedars-sinai.edu/forms/cancercenter/biostatrequest.asp

Members

- ▶ http://cedars-sinai.edu/Research/Research-Cores/Biostatistical-Sciences/Staff.aspx
- ▶ Marcio Diniz (marcio.diniz@cshs.org)
- ▶ Michael Luu (michael.luu@cshs.org)