STAT 3375Q: Introduction to Mathematical Statistics I

Lecture 1: Course Overview

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

January 17, 2024

Outline

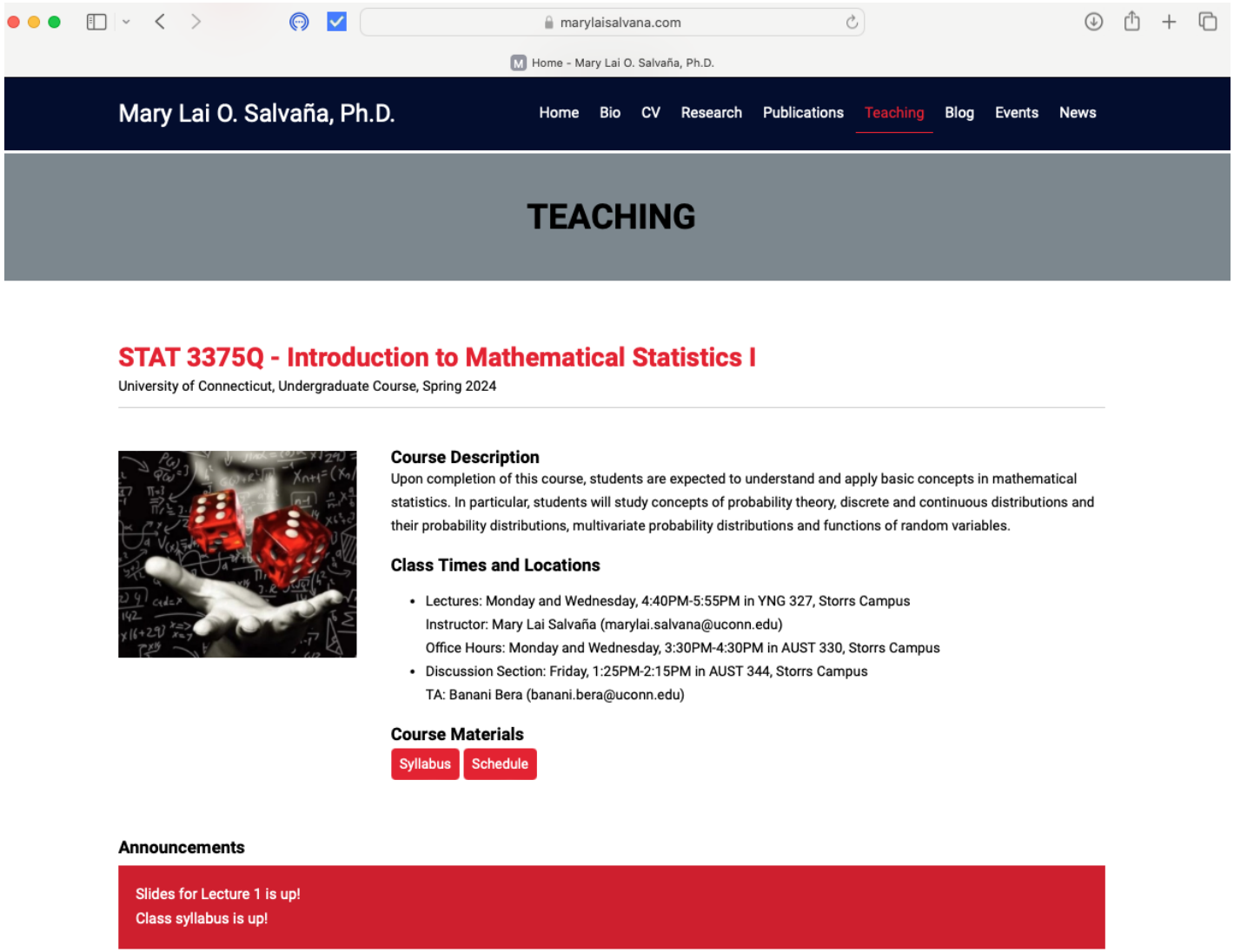
Welcome to STAT 3375Q!

About the Course

- Welcome!
- STAT3375Q: Introduction to Mathematical Statistics I
- Undergraduate Course
- Prerequisite: A grade of C+ or better in MATH 2110Q or 2130Q
- Lectures: MW, 4:40-5:55PM, YNG 327
- Discussions: F, 1:25-2:15PM, AUST 344
- Course Website: marylaisalvana.com/#teaching

About the Course

- Course Website: marylaisalvana.com/#teaching



About the Course

Course Goal

- to understand and apply basic concepts in mathematical statistics

Topics

- probability theory
- discrete distributions
- continuous distributions
- probability distributions (univariate and multivariate)
- functions of random variables

Instructor

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

Postdoctoral Researcher, Univ. of Houston
Ph.D. in Statistics, KAUST, Saudi Arabia
B.S. & M.S. in Applied Mathematics, Ateneo de Manila, Philippines

Expertise: spatial and spatio-temporal statistics, environmental statistics, computational statistics, high performance computing (supercomputers)

Website: marylaisalvana.com
Email: marylai.salvana@uconn.edu

Instructor

Mary Lai Salvaña, Ph.D.

Department of Statistics
University of Connecticut

Lectures: MW, 4:40-5:55PM, YNG 327
Office Hours: MW, 3:30-4:30PM, AUST 330

Teaching Assistants

Banani Bera

Department of Statistics
University of Connecticut

Ph.D. Student

Email: banani.bera@uconn.edu

Teaching Assistants

Manjun Yu

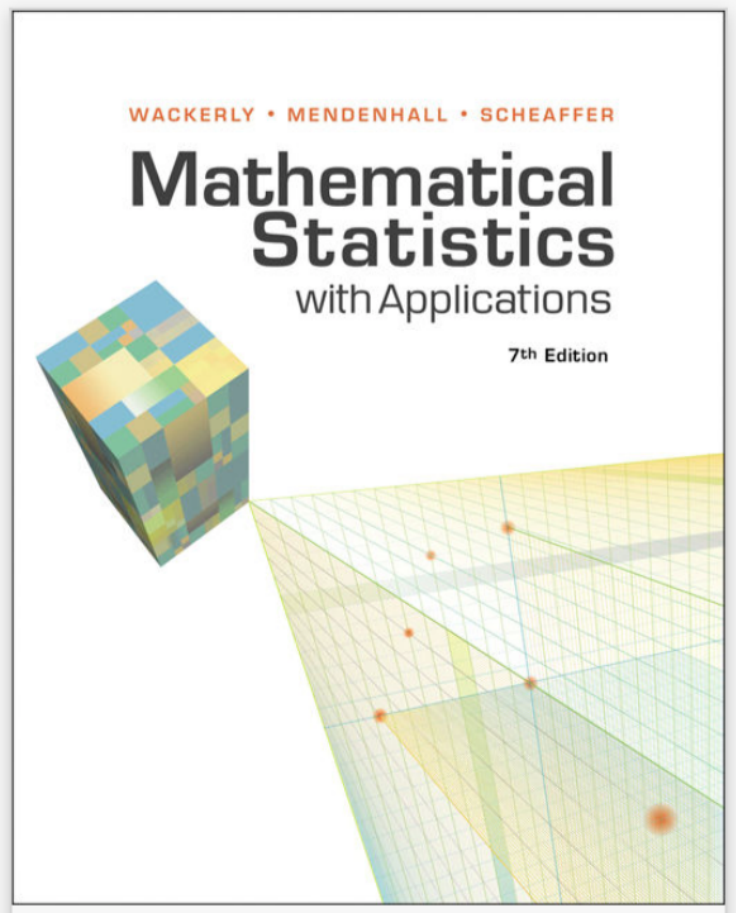
Department of Statistics
University of Connecticut

Ph.D. Student

Email: manjun.yu@uconn.edu

Logistics

Textbook



Mathematical Statistics with Applications, 7th Edition

Wackerly, Mendenhall, & Scheaffer

Cengage Learning

Grading

Weights of course components

Quizzes	25%
Midterm Exam 1	25%
Midterm Exam 2	25%
Final Exam	25%

Grade conversion

A	93-100
A−	90-92
B+	87-89
B	83-86
B−	80-82
C+	77-79
C	73-76
C−	70-72
D+	67-69
D	63-66
D−	60-62
F	0-59

Homeworks

- There will be recommended homework assignments every week, which will not be collected or graded.
- Many problems and questions on the quizzes and exams will be based on the homework problems and examples discussed in class.
- Recommended homework assignments will be discussed in the Discussion Section by the TA.
- The solutions to some challenging problems will be posted on HuskyCT.

Quizzes

- A 15-minute quiz will be given on five lecture sessions, spread across the semester.
- Quizzes are administered inside the classroom.
- Only the best four quizzes will count toward your grade.
- There will be no make-up quizzes.

Exams

- Midterm and final exams will cover both theory and application of the statistical concepts learned in class.
- Final exam is comprehensive.
- Exams will be administered online via HuskyCT, but you will be taking them inside the classroom.
- Students are required to be available for their exam during the scheduled time. Any conflict with the timing must be communicated with the Dean of Students Office.
- Missed exams cannot be made up unless with documentation of reasons required by University policy.

Attendance

- Attendance for the class is not mandatory.
- However, students are expected to attend the lectures and discussions and to engage in class activities.

Introduction to Statistics

Why do we need Statistics?

We need Statistics to answer questions that are important to us...

- *"Which stocks/investments should I buy or sell tomorrow?"*
  -stock traders, fund managers, investment bankers
- *"How much should I price these flood insurance?"*
  -actuaries, insurance companies (Travelers)
- *"Can we pay out all the claims?"*
  -reinsurance companies (Swiss Re, Munich Re)
- *"Should we cut interest rates in 2024?"*
  -U.S. Federal Reserve
- *"When should I get a new car?"*
  -Me

Why do we need Statistics?

We need Statistics to answer questions that are important to us...

- *"How can I store a 10 million x 10 million matrix in the computer memory?"*
  -computer scientists, statisticians, big data scientists
- *"How do we beat the Boston Celtics?"*
  -every NBA player, coach, and owner except those with the Celtics
- *"How do we beat the SF 49ers?"*
  -every NFL player, coach, and owner except those with the 49ers
- *"Can I trust my predictions?"*
  -any one whose wrong predictions will cause more harm/damage

What is Statistics?

The word 'Statistics' has been used in at least four different contexts as a term to refer to ...

1. a subject or discipline and everything that gets studied or practiced in its name;

   Example: Spatial Statistics, Environmental Statistics, Computational Statistics, Biostatistics
2. the *methods* to collect or process or interpret quantitative data;

   Example: Exploratory data analysis, Hypothesis Testing, Correlation Analysis, Regression Analysis, Causation Analysis
3. the *collections of data* gathered by those methods;

   Example: NBA/NFL/MLB/NHL stat sheets and box scores
4. a certain *number* that describes a collection of data;

   Example: Average, Field Goal %, Points/Rebounds/Assists per game

This course will focus on the 2nd definition.

Source: Rowntree, D. (2018). Statistics without Tears: An Introduction for Non-Mathematicians. Penguin Books.

What is Statistics?

- Textbook: Statistics is a theory of information, with inference making as its objective.

  2 key activities involved in Statistics:

  1. Gather information (data).
  2. Make an inference based on the information.

     What is Inference? It is a description of a phenomenon or an object of interest. It also means a conclusion that we make after seeing the evidence.

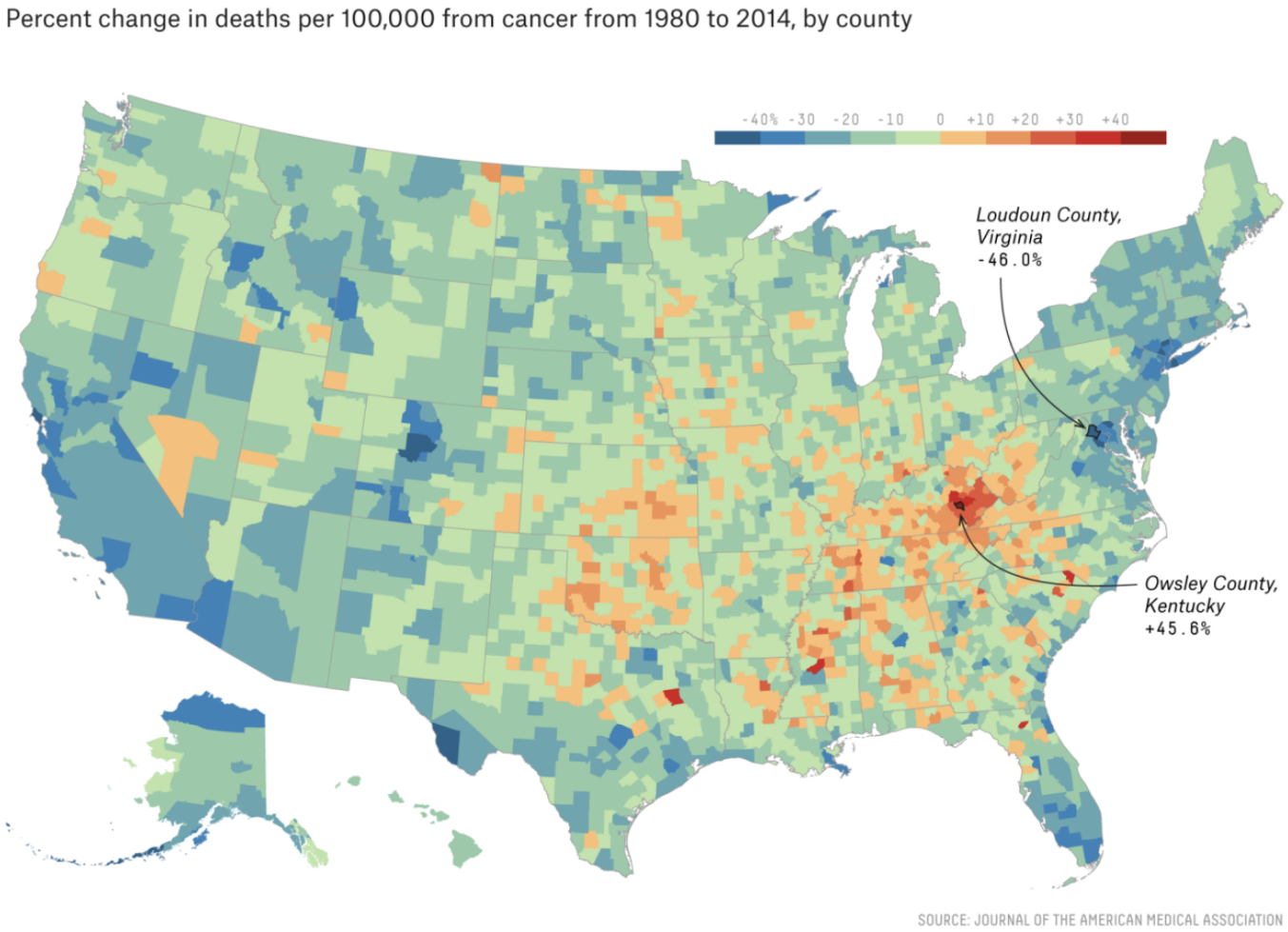
Statistics enables us to model a reality full of randomness in mathematical terms...

Source: Wackerly, D., Mendenhall, W., & Scheaffer, R. L. (2008). Mathematical statistics with applications, 7th edition.

Information $\rightarrow$ Inference

What information do we have?

What can we infer based on this information?

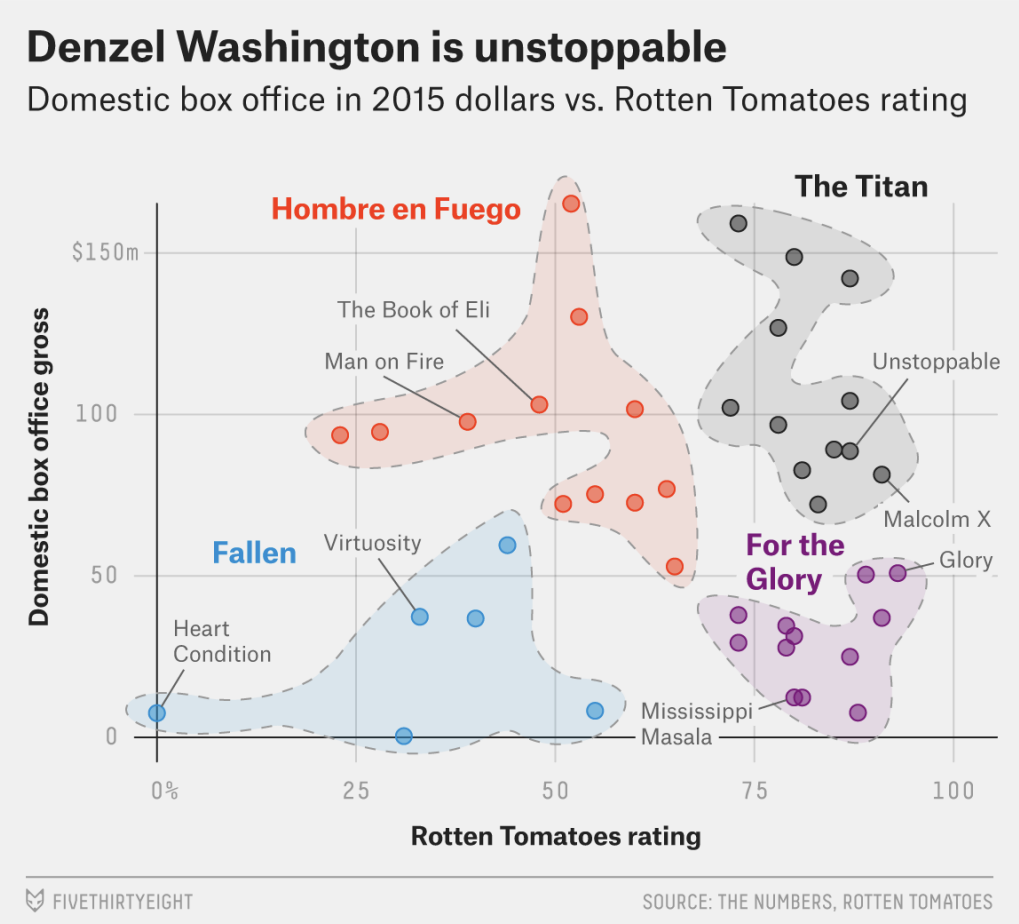


Source: https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/

Information → Inference

What information do we have?

What can we infer based on this information?

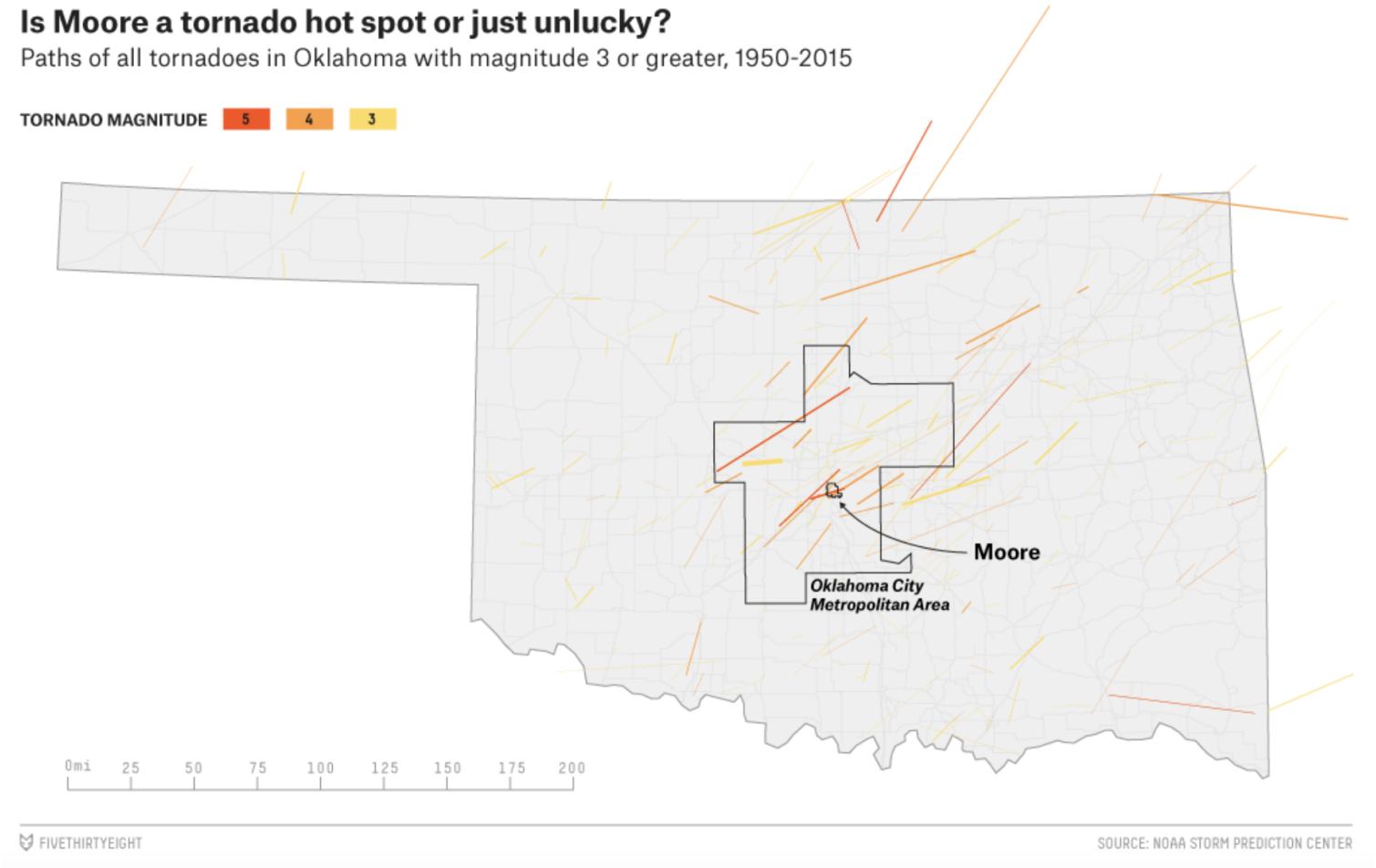


Source: https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/

Information $\rightarrow$ Inference

What information do we have?

What can we infer based on this information?

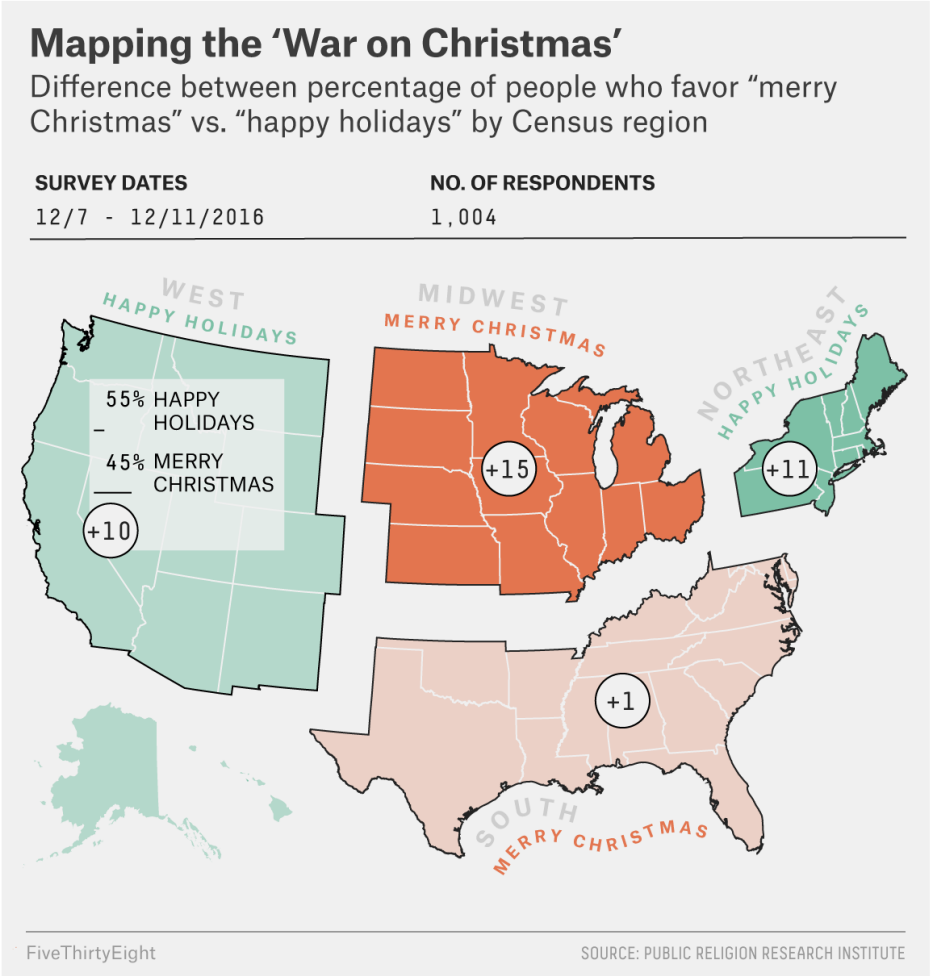


Source: https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/

Information $\rightarrow$ Inference

What information do we have?

What can we infer based on this information?

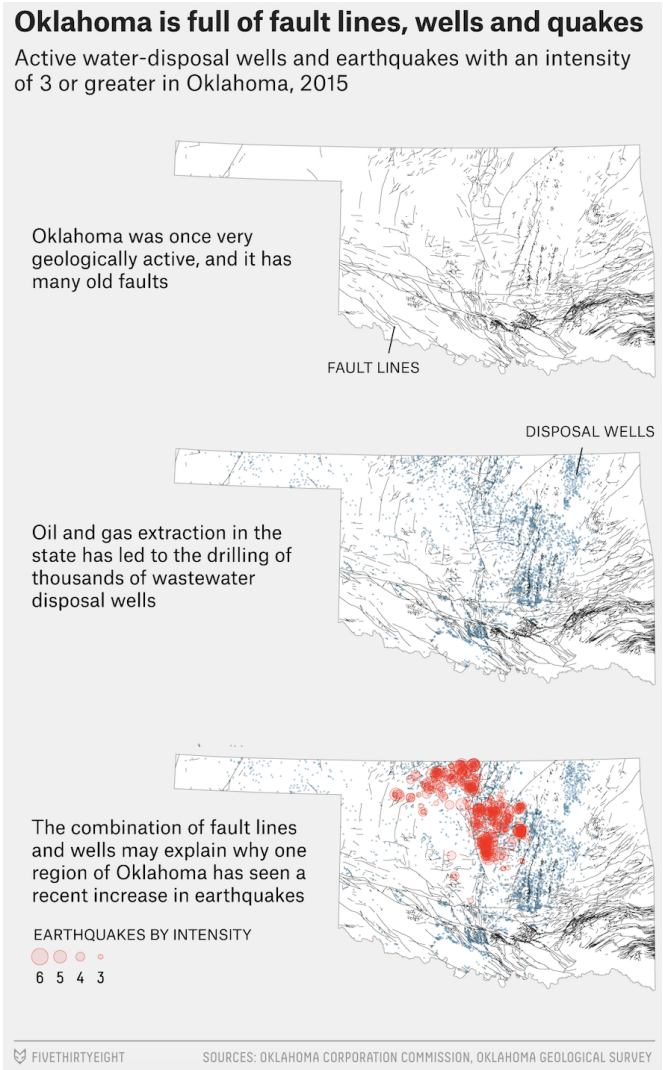


Source: https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/

Information → Inference

What information do we have?

What can we infer based on this information?



Source: https://fivethirtyeight.com/features/the-52-best-and-weirdest-charts-we-made-in-2016/

Fundamentals of Statistics

Week	Topics	Book Chapters
Jan 16-19, 2024	Probability: Introduction and definition Review of set notations Probabilistic model	2.1-2.4
Jan 22-26, 2024	Counting rules Conditional probability Independence Quiz 1 (Wed, Jan 24)	2.5-2.9
Jan 29-Feb 2, 2024	Bayes' rule Discrete random variables: Definition Expected value	2.10-3.3
Feb 5-9, 2024	Special discrete distributions Bernoulli, Binomial, Negative Binomial Quiz 2 (Wed, Feb 7)	3.4-3.6
Feb 12-16, 2024	Hypergeometric, Poisson Review problems for Midterm 1 Midterm Exam 1 (Wed, Feb 14)	3.7-3.8

Fundamentals of Statistics

Week	Topics	Book Chapters
Feb 19-23, 2024	Continuous random variables: Definition Expected value	4.1-4.3
Feb 26-Mar 1, 2024	Special continuous distributions Uniform, Normal, Student's-t Quiz 3 (Wed, Feb 28)	4.4-4.5
Mar 4-8, 2024	Exponential, Gamma, Beta	4.6-4.8
Mar 11-15, 2024	Spring Break	
Mar 18-22, 2024	Other expected values Multivariate probability distributions: Introduction Quiz 4 (Wed, Mar 20)	4.9, 5.1-5.2
Mar 25-29, 2024	Marginal and conditional probability distributions Independent random variables	5.3-5.4
Apr 1-5, 2024	Expectation and covariance Multinomial Distribution Review problems for Midterm 1 Midterm Exam 2 (Wed, Apr 3)	5.5-5.9

Fundamentals of Statistics

Week	Topics	Book Chapters
Apr 8-12, 2024	Function of random variables The method of distribution function	6.1-6.3
Apr 15-19, 2024	The method of transformation Quiz 5 (Wed, Apr 17)	6.4, 6.6
Apr 22-26, 2024	Order statistics Review problems for Final Exam	6.7
Apr 27-28, 2024	Reading Days	
Apr 29-May 4, 2024	Final Exam (TBD - See University Registrar schedule)	

Book Chapters 2.1-2.4

Probability

- commonly used to indicate one's belief in the occurrence of a future event
- can be indicated...
  - vaguely: "Chances are low it will rain today."
  - as percentages ranging from 0% to 100%.
- the foundation for statistical inference

Probability Theory

- branch of mathematics that deals with chance and uncertainty.
- provides us with tools to describe uncertain events mathematically
  - gambler: "Are the dice in this casino balanced?"
  - Geico: "What's the probability that this driver will get into an accident?"
  - NBA players guarding Steph Curry: "What are the chances that Steph Curry will go right and not left?"
  - stock traders: "Will the stock price of Saudi Aramco shares go higher than its IPO price?"

Sample Space and Outcomes

- Two ingredients in order to compute probability:
  1. sample space - the set of ALL possible outcomes for an activity or experiment
  2. outcome (also known as event) - a subset of the sample space whose probability we would like to compute
- Review of Set Theory:
  - set - a collection of objects.
    - integers: $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
    - NASDAQ stocks: {TESLA, AMD, Marathon Digital, NVIDIA, ... }
    - NYSE stocks: {Bank of America, Delta Air Lines, Ford, Nike, ... }
    - top 500 most powerful supercomputers the world: www.top500.org
    - natural disasters: {flood, wildfire, drought, hurricane, storm, ... }
  - subset - set that is entirely contained inside another set.
    Set A is a subset of Set B if all elements of A are elements of B.
    Mathematically, we write this as $A \subset B$.
    - Tech stocks in NASDAQ
    - Energy stocks in NASDAQ

Review: Set Operations

- Suppose two dice are tossed and the numbers on the upper faces are observed. Let S denote the set of all possible pairs that can be observed. (These pairs can be listed, for example, by letting $(5, 5)$ denote that a 5 was observed on the first die and a 5 on the second.)

- Define the following subsets of S:
  - A: The number on the second die is even.
  - B: The sum of the two numbers is even.
  - C: At least one number in the pair is odd.
- Question #1:
  - What is A?
    $A = \{(1,2),(1,4),(1,6),(2,2),(2,4),(2,6),(3,2),(3,4),(3,6),$
    $(4,2),(4,4),(4,6),(5,2),(5,4),(5,6)(6,2),(6,4),(6,6)\}$

Review: Set Operations

- Suppose two dice are tossed and the numbers on the upper faces are observed. Let S denote the set of all possible pairs that can be observed. (These pairs can be listed, for example, by letting $(5, 5)$ denote that a 5 was observed on the first die and a 5 on the second.)

- Define the following subsets of S:
  - A: The number on the second die is even.
  - B: The sum of the two numbers is even.
  - C: At least one number in the pair is odd.
- Question #2:
  - What is $\bar{C}$?
    - ★ $\bar{C}$ denotes the complement of C.
    - ★ $\bar{C}$ is the set of all elements in S that are NOT in C.
    - ★ $C = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$
      $(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),$
      $(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),$
      $(2,1),(2,3),(2,5),(4,1),(4,3),(4,5),(6,1),(6,3),(6,5)\}$
    - ★ $\bar{C} = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$

Review: Set Operations

- Suppose two dice are tossed and the numbers on the upper faces are observed. Let S denote the set of all possible pairs that can be observed. (These pairs can be listed, for example, by letting $(5, 5)$ denote that a 5 was observed on the first die and a 5 on the second.)

- Define the following subsets of S:
  - A: The number on the second die is even.
  - B: The sum of the two numbers is even.
  - C: At least one number in the pair is odd.
- Question #3:
  - What is $A \cap B$?
    - $\cap$ denotes the intersection of A and B.
    - $A \cap B$ is the set of all elements in both A AND B.
    - $A = \{(1,2),(1,4),(1,6),(2,2),(2,4),(2,6),(3,2),(3,4),(3,6),$ $(4,2),(4,4),(4,6),(5,2),(5,4),(5,6)(6,2),(6,4),(6,6)\}$
    - $B = \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),$ $(3,1),(3,3),(3,5),(4,2),(4,4),(4,6),$ $(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\}$
    - $A \cap B = \{(2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\}$

Review: Set Operations

- Suppose two dice are tossed and the numbers on the upper faces are observed. Let S denote the set of all possible pairs that can be observed. (These pairs can be listed, for example, by letting $(5, 5)$ denote that a 5 was observed on the first die and a 5 on the second.)

- Define the following subsets of S:
  - A: The number on the second die is even.
  - B: The sum of the two numbers is even.
  - C: At least one number in the pair is odd.
- Question #4:
  - What is $\bar{A} \cup B$?
    - ★ $\cup$ denotes the union of $\bar{A}$ and B.
    - ★ $\bar{A} \cup B$ is the set of all elements in either A OR B.
    - ★ $\bar{A} = \{(1,1),(1,3),(1,5),(2,1),(2,3),(2,5),(3,1),(3,3),(3,5),$ $(4,1),(4,3),(4,5),(5,1),(5,3),(5,5)(6,1),(6,3),(6,5)\}$
    - ★ $B = \{(1,1),(1,3),(1,5),(2,2),(2,4),(2,6),$ $(3,1),(3,3),(3,5),(4,2),(4,4),(4,6),$ $(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\}$
    - ★ $\bar{A} \cup B = S \backslash \{(1,2),(1,4),(1,6),(3,2),(3,4),(3,6),(5,2),(5,4),(5,6)\}$

Sample Space, Outcomes, and Probability

$$\begin{aligned} \text{probability} &= \frac{\text{number of desired outcomes}}{\text{total number of all possible outcomes}} \\ &= \frac{\text{number of elements inside outcome}}{\text{number of elements inside sample space}} \end{aligned}$$

Sample Space, Outcomes, and Probability

- Example 1: At the start of the season, what is the probability that Boston Celtics will win the 2024 NBA Finals?
  - sample space: (30 NBA teams)
    {Atlanta Hawks, Boston Celtics, Brooklyn Nets, Charlotte Hornets, Chicago Bulls, Cleveland Cavaliers, Dallas Mavericks, . . .}
  - outcome: {Boston Celtics}

$$
\begin{aligned}
P(\text{Boston Celtics win}) &= \frac{\text{number of elements inside outcome}}{\text{number of elements inside sample space}} \\
&= \frac{1}{30} \\
&= 3.33\%.
\end{aligned}
$$

Sample Space, Outcomes, and Probability

- Example 1: At the start of the season, what is the probability that Boston Celtics will win the 2024 NBA Finals?
  - sample space: (30 NBA teams)
    {Atlanta Hawks, Boston Celtics, Brooklyn Nets, Charlotte Hornets, Chicago Bulls, Cleveland Cavaliers, Dallas Mavericks, . . .}
  - outcome: {Boston Celtics}

$$
\begin{aligned}
P(\text{Boston Celtics win}) &= \frac{\text{number of elements inside outcome}}{\text{number of elements inside sample space}} \\
&= \frac{1}{30} \\
&= 3.33\%.
\end{aligned}
$$

Note: This formula assumes that each team in the sample space is equally capable of winning. Actual computations of win probabilities are more complicated than this.

Sample Space, Outcomes, and Probability

- Example 2: A vehicle arriving at an intersection can turn right, turn left, or continue straight ahead. Find the probability that the vehicle turns assuming the vehicle is equally likely to go right, left, or straight.

  Let $L = \{\text{turn left}\}$, $R = \{\text{turn right}\}$, & $C = \{\text{continuous straight}\}$.

  - sample space: $S = \{L, R, C\}$
  - outcome: $\{L, R\}$

$$\begin{aligned} P(\text{vehicle turns}) &= \frac{\text{number of elements inside outcome}}{\text{number of elements inside sample space}} \\ &= \frac{2}{3} \\ &= 66.6\% \end{aligned}$$

Questions?

Homework Exercises: 2.1, 2.3, 2.5, 2.7, 2.8

Solutions will be discussed this Friday by the TA.