

Stat 201: Introduction to Statistics

Standard 1: Reasons for Using Statistical Methods

Chapter One

Summaries

From *Naked Statistics*: Introduction

- “Believe me, the technical details are crucial (and interesting) – but it’s just Greek if you don’t understand the intuition.”
- “Statistics can be overly accessible in the sense that anyone with data and a computer can do sophisticated statistical procedures with a few keystrokes. The problem is that if the data are poor, or if the statistical techniques are used improperly, the conclusions can be wildly misleading and even potentially dangerous.”

From *Naked Statistics*: Chapter One

- “Statistics rarely offers a single ‘right’ way of doing anything. Does it provide meaningful information in an easily accessible way? Absolutely”
- “Hal Varian, chief economist at Google, told the New York Times that being a statistician will be ‘the sexy job’ over the next decade.”

From *Naked Statistics*: Chapter One

- “One key function of statistics is to use the data we have to make informed conjectures about larger questions for which we do not have full information.”
- “A statistical anomaly does not prove wrongdoing... Probability is one weapon in an arsenal that requires good judgement”

From *Naked Statistics*: Chapter One

- “The scientific method dictates that if we are testing a scientific hypothesis, we should conduct a controlled experiment in which the variable of interest (e.g., smoking) is the only thing that differs between the experimental group and the control group.”
- “What is the point? The point is not to do math, or to dazzle friends and colleagues with advanced statistical techniques. The point is to learn things that inform our lives.”

From *Naked Statistics*: Chapter One

- “Even in the best of circumstances, statistical analysis rarely unveils ‘the truth.’ We are usually building a circumstantial case based on imperfect data.”
- “The approach is not like addition or long division, in which the correct technique yields the ‘right’ answer and a computer is always more precise and less fallible than a human. Statistical analysis is like good detective work.”

Why Do We Use Statistics?

- We use statistics to make the leap from historical data to story telling – we call this **summary or descriptive statistics**.
- We use statistics to make **inference** on things that are too big to measure by measuring a smaller group.
- With **summaries and inference** we can apply statistics to answer questions, find patterns or make decisions.

Walkthrough

Why Do We Use Statistics?

- 1) We use statistics to make the leap from historical data to story telling – we call this **summary or descriptive statistics**.
 - 2) We use statistics to make **inference** on things that are too big to measure by measuring a smaller group and extending our findings.
- With **summaries and inference** we can apply statistics to answer questions, find patterns or make decisions.

Why Do We Use Statistics?

3) We use statistics to measure things that are far too big to measure exactly

- Height of US citizens
- Average age of everyone in the world
 - Think about measuring these things – by the time you finished measuring everyone people would have died and been born; **the measuring would never end!**

Why Do We Use Statistics?

- **Descriptive Statistics:** This is when we use statistics to make the leap from massive datasets to what they tell us.
 - Sometimes scientists, politicians, computer engineers etc. have thousands or millions of rows of data in Excel and they want to draw conclusions
 - Here, they could use **descriptive statistics** and charts to summarize thousands of rows with just a few numbers

Why Do We Use Statistics?

- **Descriptive Statistics:**
 - We have a dataset for a **sample** or **population**
 - We have one or more **variables** measured for each **experimental unit (or subject)** that we would like to summarize
 - We use **descriptive statistics** and charts to summarize the data set to identify patterns, trends etc.

Why Do We Use Statistics?

- **Inferential Statistics:** This is when we use the descriptive statistics of a sample to make estimates or predictions about a population
 - Sometimes scientists, politicians, computer engineers etc. have a small **sample** and want to draw conclusions about the larger **population**
 - Here, they could use **descriptive statistics** and statistical methodology to estimate the **population parameters** based off of the **sample statistics** which requires a **measure of reliability** which quantifies the uncertainty of our estimate. Think “plus or minus.”

Real World Careers

- Actuaries use historical data to predict lifespans, driving habits, etc (**inference**)
- Econometricians use statistics to understand economic theories and relationships (**patterns**)
- QA/Reliability use statistics to aid in accepting batches of products using sampling (**summary**)
- Psychologists/Pharmacists/Biologists use statistics to prove theories with experimental data(**inference**)

Watch These Videos!

- Why Statistics*
 - <https://www.youtube.com/watch?v=yxXsPc0bphQ>
- Target uses statistics
 - <https://www.youtube.com/watch?v=jgsdQxTv5kY>
- Changes in America
 - <https://www.youtube.com/watch?v=ke52L1k9VzQ>
- Wealth Distribution*
 - <https://www.youtube.com/watch?v=krwdJ6DyafQ>

A Great Example of Why we Use Statistics for Watching at Home

- Cool graphics and statistics about the world:
 - <https://www.youtube.com/watch?v=hVimVzgtD6w>

Why Do We Use Statistics?

- **Inferential Statistics:**

- There's a **population** of interest often too big or too expensive to measure
- We have a **sample**, a subset of the **population**
- We have one or more **variables** measured for each **experimental unit (or subject)** that we would like to summarize
- We use the **descriptive statistics** of the **sample** to make estimates about the **population**
- We report a **measure of reliability** of our inference, which is generally dependent on how large our sample is