Introduction to Probability and Statistics

Jerome Dumortier

26 August 2022

Risk and Uncertainty in Everyday Life

Grades:

- Uncertainty surrounding class grade during a semester
- Association of probabilities with each grade

Fire station calls:

- Number and location of calls
- Number of fire trucks and other vehicles required

Two outcomes does not mean a 50% chance for each to happen:

- Success of a free throw by Stephen Curry
- Flight delay due to fog

Do You Know When You Were Hooked? Netflix Does:

- Recording of all activities, e.g., browsing, pausing, rewinding
- 75% of viewings are based on recommendations by Netflix

Some Examples of Statistics in the News I

Election outcomes

- 2016 U.S. Presidential Election
    - FiveThirtyEight forecast of Donald Trump winning: 28.6%
    - Cognitive biases versus data as an explanation
- 2002 French Presidential Election
    - Two-stage election
    - Final round: Jacques Chirac (82.2%) and Jean-Marie Le Pen (17.8%)

Evolution of the stock market

- Importance of correlation among stocks and mutual funds

Path of hurricane Sandy

Scottish Ministers' Widows' Fund

Preceding work

- Edmond Halley's (same as comet) life tables for the city of Breslau (today Wrocław)
- Detailed work on birth and death by age

Insurance fund calculations in 1744 by Alexander Webster and Robert Wallace:

- Payments to widows and heirs after death of ministers
- Required information: Number of clergymen, deaths per year, life expectancy of surviving family, time of remarriage, etc.
- Calculation of annual payments into the fund

Fund balance (in pound sterling) in 1765:

- Estimated: 58,348
- Actual: 58,347

Probability, Statistics, and Regression Analysis

Probability:

- Providing means for modeling populations, experiments, and any other random phenomena
- Probability distributions: How do we model random outcomes?
- Foundation for statistics

Statistics:

- Learning something about the population based on a sample
- Confidence intervals and hypothesis testing

Regression analysis:

- Mathematical relationship among variables
- Example: Price of a used car as a function of mileage

Difference between probability and statistics: Bucket example

Difference between Population and Statistic

Population:

- A population is the collection of all possible individuals, entities, objects, or measurements of interest for a particular investigation. A sample is any portion or subset of the population. A *parameter* characterizes the population and is usually unknown (forever).

Sample:

- A statistic is any measurable characteristic of a sample. Statistical analysis utilizes statistics from representative samples to infer the parameters of an entire population.

Using a sample rather than the population:

- Cost considerations
- Possible destruction of observation units (e.g., mileage of tires)
- Unfeasible to study all units of observations

Variables

Qualitative variables:

- Non-numeric, e.g., gender, political affiliation, state of residence
- Can be transformed into numerical value, i.e., "dummy variables'' in regression analysis

Quantitative variables:

- Numeric, e.g, age, income, GPA, number of kids

Quantitative variables can be either:

- Discrete: Take two close values and there is no value in between, e.g., number of people in a class
- Continuous: Take two close values and there is always (!) a value in between, e.g., weight of a people

Levels of Variable Measurements

Nominal:

- Categories, e.g., eye color, gender, religious affiliation, mode of transportation to O'Neill IUPUI
- No natural ordering

Ordinal:

- Categories, e.g., level of happiness, Homeland Security Advisory System
- Natural ordering, i.e., data can be ordered

Interval:

- Intervals between levels are equally spaces and differences between variables have a meaning
- Examples: Income, GPA, etc.
- Most commonly used in this class.