Measures of Spread

Introduction to Probability

[POLS 4150] Intro. to Probability Theory, Discrete and Continuous Distributions

Professor Jason Anastasopoulos
ljanastas@uga.edu

University of Georgia

January 31, 2017
From last time...

- Measures of the spread of a distribution.
- The 68 – 95 – 99% rule.
- Outliers.
Three measures of spread

What were they?
Three measures of spread

What were they?

1. Range.
2. Standard deviation.
3. Percentiles.
Three measures of spread

What were they?

1. Range.
2. Standard deviation.
3. Percentiles.
Three measures of spread

What were they?
1. Range.
2. Standard deviation.
3. Percentiles.
Three measures of spread

What were they?

1. Range.
2. Standard deviation.
3. Percentiles.
Standard deviation

\[ s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \]

- Standard deviation is a measure of the average deviation of the observations \( x \) from the mean \( \bar{x} \).
Standard deviation

What if we swapped $\sum_{i=1}^{N} (x_i - \bar{x})^2$ with $\sum_{i=1}^{N} (\bar{x} - x_i)^2$

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (\bar{x} - x_i)^2}$$

Would the answer change?
Empirical Rule (68 – 95 – 99.7 Rule)
Unemployment in Georgia Counties

\[ \bar{x} = 7\% \] – Average unemployment rate of GA counties.

\[ s = 1\% \] – Standard deviation of unemployment rate.
Unemployment in Georgia Counties

- $\bar{x} = 7\%$ – Average unemployment rate of GA counties.
- $s = 1\%$ – Standard deviation of unemployment rate.

**Question 1:** What % of counties in GA have between 5% and 9% unemployment?

**Answer:** 95% of counties in GA.

**Question 2:** Is it unusual for a GA county to have a 10% unemployment rate? If yes how unusual is it?

**Answer:** Yes. Less than 1% of GA counties have an unemployment rate this high or higher.
Unemployment in Georgia Counties

- $\bar{x} = 7\%$ – Average unemployment rate of GA counties.
- $s = 1\%$ – Standard deviation of unemployment rate.

**Question 1:** What % of counties in GA have between 5% and 9% unemployment?

**Answer:** 95% of counties in GA.

**Question 2:** Is it unusual for a GA county to have a 10% unemployment rate? If yes how unusual is it?

**Answer:** Yes. Less than 1% of GA counties have a unemployment rate this high or higher.
Unemployment in Georgia Counties

- $\bar{x} = 7\%$ – Average unemployment rate of GA counties.
- $s = 1\%$ – Standard deviation of unemployment rate.

**Question 1:** What % of counties in GA have between 5\% and 9\% unemployment?

**Answer:** 95\% of counties in GA.

**Question 2:** Is it unusual for a GA county to have a 10\% unemployment rate? If yes how unusual is it?

**Answer:** Yes. Less than 1\% of GA counties have a unemployment rate this high or higher.
Unemployment in Georgia Counties

- \( \bar{x} = 7\% \) – Average unemployment rate of GA counties.
- \( s = 1\% \) – Standard deviation of unemployment rate.

**Question 1:** What % of counties in GA have between 5% and 9% unemployment?

**Answer:** 95% of counties in GA.

**Question 2:** Is it unusual for a GA county to have a 10% unemployment rate? If yes how unusual is it?

**Answer:** Yes. Less than 1% of GA counties have a unemployment rate this high or higher.
Percentiles

Point such that $p\%$ of observations fall below or at that point and $(100 - p)\%$ fall above it.

- $90^{th}\%ile \rightarrow 90\%$ of observations below, $10\% above$ etc.
- Eg) An LSAT score of 164 is the $90^{th}$ percentile.
Percentiles and Outliers

\[ \text{Outlier} = \left\{ \langle 25\% \text{ile} - 1.5 \times IQR, > 75\% \text{ile} + 1.5 \times IQR \right\} \]

- An outlier is typically defined as an observation \( 1.5 \times IQR \) below the \( 25\% \text{ile} \) or \( 1.5 \times IQR \) above the \( 75\% \text{ile} \).
Boxplots and Percentiles

- Boxplots display a percentile distribution of a variable.
- They show: median, IQR, 25\textsuperscript{th} percentile, 75\textsuperscript{th} percentile, outliers.
Distribution of Races/Ethnicity in White U.S. House Members Facebook Images
Distribution of Races/Ethnicity in White U.S. Senators
Facebook Images

% of Race in Facebook Photos, White Senate Members

Professor Jason Anastasopoulos ljanastas@uga.edu
University of Georgia

[POLS 4150] Intro. to Probability Theory, Discrete and Continuous Distributions
Philosophical foundations of probability

- If someone asked you: “What is the probability that there will be a major terrorist attack on the U.S. in the next year?”
- What would be your numerical estimate?
- How did you arrive at that estimate?
Philosophical foundations of probability

- If someone asked you: “What is the probability that there will be a major terrorist attack on the U.S. in the next year?”
- What would be your numerical estimate?
- How did you arrive at that estimate?
Philosophical foundations of probability

- If someone asked you: “What is the probability that there will be a major terrorist attack on the U.S. in the next year?”
- What would be your numerical estimate?
- How did you arrive at that estimate?
Two views of probability: frequentist and Bayesian

- **Frequentist probability** – assumes that the likelihood of an outcome can be based on the proportion of times it occurred over a long sequence.

- **Bayesian probability** – assumes the existence of prior beliefs about an outcome and updates these prior beliefs according to new data.
Two views of probability: frequentist and Bayesian

- **Frequentist probability** – assumes that the likelihood of an outcome can be based on the proportion of times it occurred over a long sequence.

- **Bayesian probability** – assumes the existence of prior beliefs about an outcome and updates these prior beliefs according to new data.
Bayesian Probability

- Bayesian probability was based on the ideas of an 18th century English Presbyterian minister.
- Developed an interest in probability in his 50’s when he published his famous piece “An Essay towards solving a Problem in the Doctrine of Chances.”
Bayes Theorem

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

- **\(P(A|B)\)** – Probability of an outcome \(A\) given some information \(B\)
- **\(P(B|A)\)** – Probability of observing some information given the outcome.
- **\(P(A)\)** – Prior beliefs about the outcome.
- **\(P(B)\)** – Probability of observing the information under all possible outcomes.
Bayes Theorem and Terrorism

\[ P(T_{t+1}^+ | T_{t-1}^+) = \frac{P(T_{t-1}^+ | T_{t+1}^+) P(T_{t+1}^+)}{P(T_{t-1}^+)} \]

- \( P(T_{t+1}^+ | T_{t-1}^+) \) – Probability of a terrorist attack next year given Terrorist attacks last year.
- \( P(T_{t-1}^+ | T_{t+1}^+) \) – Probability of terrorist attack last year given beliefs about this year.
- \( P(T_{t+1}^+ \) – Prior beliefs about a terrorist attack next year.
- \( P(T_{t-1}^+ \) – Probability of a terrorist attack last year.
Bayes Theorem and Terrorism

\[
P(T_{t+1}^- | T_{t-1}^-) = \frac{P(T_{t-1}^+ | T_{t+1}^-)P(T_{t+1}^-)}{P(T_{t+1}^-)}
\]

\[\begin{align*}
\blacklozenge & \quad P(T_{t+1}^- | T_{t-1}^-) = \text{Unknown.} \\
\blacklozenge & \quad P(T_{t-1}^+ | T_{t+1}^-) = \frac{4 \text{ Days with attacks}}{365 \text{ Days in the year}} = 0.01 \\
\blacklozenge & \quad P(T_{t+1}^-) = 0.5. \\
\blacklozenge & \quad P(T_{t-1}^+) = P(T_{t-1}^+ | T_{t+1}^-)P(T_{t+1}^-) + P(T_{t-1}^+ | T_{t+1}^+)P(T_{t+1}^+) = 0.01 \times 0.5 + 0.99 \times 0.05 = 0.056
\end{align*}\]
Bayes Theorem and Terrorism

\[ P(T_{t+1}^+ \mid T_{t-1}^+) = \frac{0.01 \times 0.5}{0.056} = 0.089 \]

- \( P(T_{t+1}^+ \mid T_{t-1}^+) = \) Unknown.
- \( P(T_{t-1}^+ \mid T_{t+1}^+) = \frac{4 \text{ Days with attacks}}{365 \text{ Days in the year}} = 0.01 \)
- \( P(T_{t+1}^+) = 0.5 \).
- \( P(T_{t-1}^+) = P(T_{t-1}^+ \mid T_{t+1}^+)P(T_{t+1}^+) + P(T_{t-1}^+ \mid T_{t+1}^-)P(T_{t+1}^-) = 0.01 \times 0.5 + 0.99 \times 0.05 = 0.056 \)
What did we do here?

- **Original beliefs:** 50% probability of a terrorist event.
- **Data suggested:** 1% probability of a terrorist event.
- **Updated predictions:** 8.9% probability of a terrorist event.
What did we do here?

- **Original beliefs:** 50% probability of a terrorist event.
- **Data suggested:** 1% probability of a terrorist event.
- **Updated predictions:** 8.9% probability of a terrorist event.
What did we do here?

- **Original beliefs:** 50% probability of a terrorist event.
- **Data suggested:** 1% probability of a terrorist event.
- **Updated predictions:** 8.9% probability of a terrorist event.
What did we do here?

- Bayesian probability allowed us to incorporate our prior beliefs with new data to come up with new estimates.
Frequentist probability

- **Frequentist statistics**, unified in large part by Jerzy Neyman, a Polish mathematician, views probabilities as the proportion of outcomes that an event occurred over a long sequence of observations.
Terrorism from a frequentist perspective

- Might count the number of days in which a terrorist event occurred over the last year divided by the total number of days in the year.

  Eg) $\frac{4}{365} = 0.01$
What probability of observing a heads in a coin flip?

- Over 100 flips = $P(H) = \{0, 1, 0, \cdots\} = 0.46$
- Over 1000 flips = $P(H) = \{0, 1, 0, \cdots\} = 0.48$
- Over $\infty$ flips = $P(H) = \{0, 1, 0, \cdots\} = 0.50$
Frequentist probability

- Frequentist probability is concerned with the long-run or asymptotic frequencies of event.
- Uses mathematical models to infer probabilities when $N \to \infty$.
- Frequentism is the underlying philosophical foundation of all of the tools that we will learn in this class:
  1. Statistical sampling.
  2. Hypothesis testing.
  3. Confidence intervals.
  4. Linear regression etc etc.
Introduction to probability

\[ P(A) = \text{Probability of an event, } A, \text{ occurring} \]

- Before we start discussion distributions, let’s take a step back and talk about some basic rules of probability.
- Probability is fundamentally about assigning probabilities to events.
- An event can be pretty much anything for which there is an alternative outcome.
- Eg) \( A = \{ \text{sun rises tomorrow, Supreme Court nominee will be blocked, etc.} \} \)
Rules of probability

\[ P(A^C) = 1 - P(A) \]

- \( P(A^C) \) Probability of something not happening.
- \( P(A) \) Probability of something happening.
Complements of events

- If $P(\text{Terrorism}) = 0.01$.
- What is $P(\text{Terrorism}^c) =$?
Union of events

\[ P(A \text{ or } B) = P(A) + P(B) \]

\[ A \perp B \]

- Probability that one event or another event that are independent from each other is just their sum.
If $P(Terrorism) = 0.01$ and;
$P(\text{Falcons win superbowl}) = 0.05$
What is:
$P(\text{Terrorism or Falcons win superbowl}) =$?
Union of events

If \( P(\text{Terrorism}) = 0.01 \) and;
\[ P(\text{Falcons win superbowl}) = 0.05 \]

\[
\begin{align*}
P(\text{Terrorism or Falcons win superbowl}) &= P(\text{Terrorism}) + P(\text{Falcons win superbowl}) = \\
&= 0.01 + 0.05 = 0.06
\end{align*}
\]
Intersection of events

\[ P(A \text{ and } B) = P(A)P(B) \]
\[ A \perp B \]

- Probability of both independent events occurring is just their probabilities multiplied together.
Intersection of events

If $P(\text{Terrorism}) = 0.01$ and;
$P(\text{Falcons win superbowl}) = 0.05$
What is:
$P(\text{Terrorism and Falcons win superbowl}) =$?
Intersection of events

If \( P(\text{Terrorism}) = 0.01 \) and;
\[
P(\text{Falcons win superbowl}) = 0.05
\]

\[
P(\text{Terrorism and Falcons win superbowl}) = P(\text{Terrorism})P(\text{Falcons win superbowl}) = (0.01)(0.05) = 0.0005
\]
Probability distributions for discrete and continuous variables

- Probability distributions are full distributions of all possible outcomes and probability of those outcomes occurring.
- Recall that **discrete** variables and variables which take on a finite number of values.
- **Continuous** variables take on a theoretically infinite number of values.
Probability distributions for discrete and continuous variables

- **Discrete probability distributions** are probability distributions which assign a probability to each individual outcome.

- **Continuous probability distributions** are probability distributions which assign probabilities to intervals.

Professor Jason Anastasopoulos  ljanastas@uga.edu

University of Georgia

[POLS 4150] Intro. to Probability Theory, Discrete and Continuous Distributions
Probability distribution of a discrete variable

Children in families = \( y = \{4, 6, 2, 1, 1, 2\} \)

- The number of children in families is a good example of a discrete variable.
Probability distribution of a discrete variable

Children in families = $y = \{4, 6, 2, 1, 1, 2\}$

$$0 \leq P(y) \leq 1$$

$$\sum_{i=1}^{N} P(y) = 1$$

$P(y = 4) = 1/6, P(y = 6) = 1/6,$
$P(y = 2) = 2/6, P(y = 1) = 2/6$

This is the full probability distribution of $y$. 