# STAT 423/523 Statistical Methods for Engineers and Scientists

## Lecture 0: Overview

Chencheng Cai

Washington State University

# Course Information

**Lecture** Tu/Th 10:35 – 11:50 AM @FULM 201

**Office Hours** Tu/Wed 3:00 – 5:00 PM @Neill 405 or by appointment

**Spring Break** Mar 10 – Mar 14

**Contact**
- email: chencheng.cai@wsu.edu

**Notes**
- Canvas
- https://chenchengcai.com/teaching/Stat523

# Course Requirements

- Prerequisites:
  - One 3-credit 300-level STAT course (for example, STAT 360 or STAT 370)
  - Basic programming in any script language (R, Python, Matlab, etc.)
- Textbook (optional):
  **Probability and Statistics for Engineering and the Sciences**, 9th Edition. Jay L. Devore. 2015.
  - Not required to purchase.
  - Both 8th and 9th editions are fine.
  - The course notes will be self-contained.

# Assessment

- **Homework: 30%**
  - Around five homework.
  - May contains both theoretical and computational problems.
  - Have two weeks to finish.
  - Submit on Canvas.

# Assessment

- **Quizzes: 30%**
  - Around three quizzes.
  - 30 – 40 minutes in classes on Thursdays.
  - Open book and open notes.
  - No programming problems.
  - **Dates: Jan 30, Mar 20, Apr 17**.

# Assessment

- **Midterm Exam: 20%**
  - 60 – 75 minutes in class.
  - Closed book and closed notes.
  - One-page cheating sheet allowed.
  - No programming problems.
  - **Date: Feb 20**.

# Assessment

- **Project: 20%**
  - Group project (2 – 3 students).
  - Apply statistical methods to a real dataset.
  - 1 page proposal (problem and dataset to study).
  - 3 – 5 pages report (motivation, problem, method, results).
  - 10 – 15 minutes presentation (depending on the number of groups).
  - **Important Dates:**
    - Form groups before **Feb 10**.
    - Proposal due on **Mar 7**.
    - Presentations on **Apr 22 and Apr 24**.
    - Report due on **Apr 28**.

# Tenative Schedule

- (1 weeks): Course overview and review on proabability and distributions.
- (2 weeks): Point Estimation and Confidence Intervals.
- (1 weeks): Programming Foundation.
- (2 weeks): Hypothesis Testing
- (3 weeks): Analysis of Variance.
- (4 weeks): Linear Regression.
- (extra weeks): advanced topics.

# Programming

- Programming is needed for the homework and project.
- You may choose any script programming language you are comfortable with.
- I will use R or Python in the class:
  - **Please finish the programming survey on Canvas by Jan 13**.
- The textbook is programming-free with sets of distributional tables in the appendix. We will use R or Python for the computation instead.

# Before we start: Accurate Statistical Statement

Besides the technical skills, a good statistical statement should also be:

- **Accurate**:
  - The statement should be correct and precise.
- **Clear**:
  - The statement should be easy to understand.
- **Concise**:
  - The statement should be brief and to the point.
- **Relevant**:
  - The statement should be related to the problem.

Examples:

- A Carbon Output Example
- Bertrand's Paradox

# A Carbon Output Example

11 organizations calculated the carbon output for a flight from New York to Los Angeles. The results are as follows:

| Carbon Calculator | $CO_2$ (lb) |
|---|---|
| Terra Pass | 1924 |
| Conservation International | 3000 |
| Cool It | 3049 |
| World Resources Institute/Safe Climate | 3163 |
| National Wildlife Federation | 3465 |
| Sustainable Travel International | 3577 |
| Native Energy | 3960 |
| Environmental Defense | 4000 |
| Carbonfund.org | 4820 |
| The Climate Trust/CarbonCounter.org | 5860 |
| Bonneville Environmental Foundation | 6732 |

# A Carbon Output Example

We may compute the summary statistics:

$$\bar{X} = \frac{1}{11}\sum_{i=1}^{11} X_i = 3959.1$$

$$s = \sqrt{\frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2} = 1376.9$$

Statements on the carbon output:

- The carbon output for the flight is 3959.1.

# A Carbon Output Example

We may compute the summary statistics:

$$\bar{X} = \frac{1}{11}\sum_{i=1}^{11} X_i = 3959.1$$

$$s = \sqrt{\frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2} = 1376.9$$

Statements on the carbon output:

- The carbon output for the flight is 3959.1. (✗)
- The mean carbon output for the flight is 3959.1.

# A Carbon Output Example

We may compute the summary statistics:

$$\bar{X} = \frac{1}{11}\sum_{i=1}^{11} X_i = 3959.1$$

$$s = \sqrt{\frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2} = 1376.9$$

Statements on the carbon output:

- The carbon output for the flight is 3959.1. (✗)
- The mean carbon output for the flight is 3959.1. (✗)
- The sample mean for the carbon output estimations is 3959.1.

# A Carbon Output Example

We may compute the summary statistics:

$$\bar{X} = \frac{1}{11}\sum_{i=1}^{11} X_i = 3959.1$$

$$s = \sqrt{\frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2} = 1376.9$$

Statements on the carbon output:

- The carbon output for the flight is 3959.1. (✗)
- The mean carbon output for the flight is 3959.1. (✗)
- The sample mean for the carbon output estimations is 3959.1. (✓)
- We estimate the carbon output for the flight as 3959.1.

# A Carbon Output Example

We may compute the summary statistics:

$$\bar{X} = \frac{1}{11}\sum_{i=1}^{11} X_i = 3959.1$$

$$s = \sqrt{\frac{1}{10}\sum_{i=1}^{11}(X_i - \bar{X})^2} = 1376.9$$

Statements on the carbon output:

- The carbon output for the flight is 3959.1. (✗)
- The mean carbon output for the flight is 3959.1. (✗)
- The sample mean for the carbon output estimations is 3959.1. (✓)
- We estimate the carbon output for the flight as 3959.1. (✓)

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657).

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657). (✗)
- The approximate 95% CI for the carbon output is (1260, 6657).

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657). (✗)
- The approximate 95% CI for the carbon output is (1260, 6657). (✓)
- Under the normality assumption, the 95% CI for the carbon output is (1260, 6657).

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657). (✗)
- The approximate 95% CI for the carbon output is (1260, 6657). (✓)
- Under the normality assumption, the 95% CI for the carbon output is (1260, 6657). (✓)
- The carbon output is between 1260 and 6657.

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657). (✗)
- The approximate 95% CI for the carbon output is (1260, 6657). (✓)
- Under the normality assumption, the 95% CI for the carbon output is (1260, 6657). (✓)
- The carbon output is between 1260 and 6657. (✗)
- The carbon output is between 1260 and 6657 with 95% confidence.

# A Carbon Output Example

We may compute the following numbers:

$$\bar{X} - 1.96s = 1260$$
$$\bar{X} + 1.96s = 6657$$

Statements on the confidence interval (CI):

- The 95% CI for the carbon output is (1260, 6657). (✗)
- The approximate 95% CI for the carbon output is (1260, 6657). (✓)
- Under the normality assumption, the 95% CI for the carbon output is (1260, 6657). (✓)
- The carbon output is between 1260 and 6657. (✗)
- The carbon output is between 1260 and 6657 with 95% confidence. (✓)

# Bertrand's Paradox

Suppose we have a circle with radius 1. If we **randomly** draw a chord, what is the probability that the chord is longer than $\sqrt{3}$?

Clarifications:

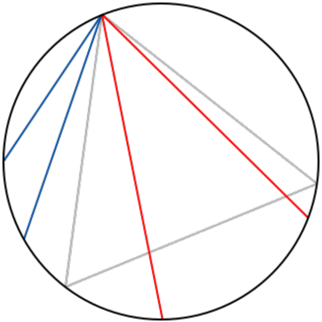
- A **chord** is a straight line segment whose endpoints both lie on the circle.
- $\sqrt{3}$ is the length of the side of an equilateral triangle inscribed in a circle.

# Bertrand's Paradox

The random endpoints method:

- Randomly select two points uniformly on the circumference of the circle.
- Draw the chord connecting these two points.

Using this method, the probability that the chord is longer than $\sqrt{3}$ is $\frac{1}{3}$.

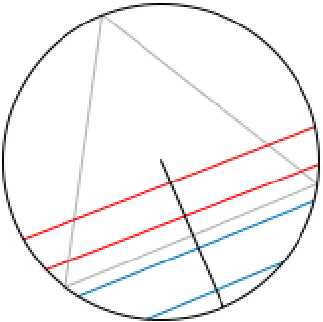


# Bertrand's Paradox

The random radius method:

- Choose a radius of the circle uniformly.
- Randomly select a point uniformly on the radius.
- Draw the chord perpendicular to the radius at the selected point.

Using this method, the probability that the chord is longer than $\sqrt{3}$ is $\frac{1}{2}$.

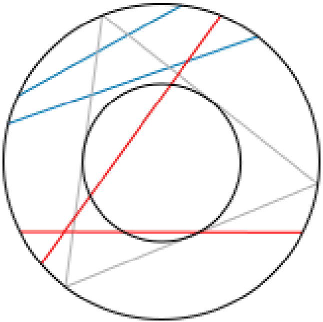


# Bertrand's Paradox

The random midpoint method:

- Randomly select a point uniformly in the circle as the midpoint.
- Construct a chord with the chosen point as its midpoint.

Using this method, the probability that the chord is longer than $\sqrt{3}$ is $\frac{1}{4}$.



# Bertrand's Paradox

- The paradox is that the probability of the event has different values depending on the method of random selection.
- The statement should be clear and acurate to avoid ambiguity.

Chords generated by the three methods: