MS 302 In-class Problems

May 8, 2025

Table of Contents

Chapter 1 Section 1

Suppose that you produce structural metal bars by heat treating them using induction heating and that they must withstand 10,000 lbs.

You set as your goal a breaking point of 15,000 lbs.

Will the breaking point be exactly 15,000 lbs?

Almost every procedure we deal with has some form of variation built into it.

1. The failure point of structral metal bars
2. How long it takes to brush your teeth in the morning
3. The location that a basketball hits as you practice your free throw shot.

In order to understand and maybe even control the *uncertainty* and *variation* we use

Definition 1 (statistical inference)

The drawing of inferences about a population based on data taken from a sample of that population.*

*OED

Definition 1 (statistical inference)

The drawing of inferences about a population based on data taken from a sample of that population.*

Two *factors* that contribute to the strength of a bar in our process are temperature and time in the heat treatment.

Definition 2 (factors)

are properties or characteristics of a population.

*OED

To determine how much weight our bars can support we could put them into a machine that destructively tests them.

To determine how much weight our bars can support we could put them into a machine that destructively tests them. But...?

To determine how much weight our bars can support we could put them into a machine that destructively tests them. But...?

Definition 3 (Population)

collection of all individuals in a particular scientific system or process.

To determine how much weight our bars can support we could put them into a machine that destructively tests them. But...?

Definition 3 (Population)

collection of all individuals in a particular scientific system or process.

Definition 4 (Sample)

collection of observations from the population.

To determine how much weight our bars can support we could put them into a machine that destructively tests them. But...?

Definition 3 (Population)

collection of all individuals in a particular scientific system or process.

Definition 4 (Sample)

collection of observations from the population.

Definition 5 (Observation)

individual or item of a particular type involved in the scientific system or process.

To determine how much weight our bars can support we could put them into a machine that destructively tests them. But...?

Definition 3 (Population)

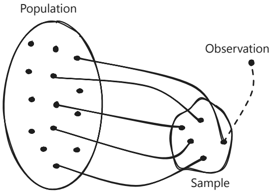
collection of all individuals in a particular scientific system or process.

Definition 4 (Sample)

collection of observations from the population.

Definition 5 (Observation)

individual or item of a particular type involved in the scientific system or process.



In addition to *Inferential statistics* there is *Descriptive statistics*.

Definition 6 (Descriptive statistics)

are used when seeking only to gain some summary of a set of data represented by a sample (single-number statistics that provide a sense of center, variability, or general nature of the distribution of the sample data).

An understanding of probability provides a basis to understand statistical inference.

An understanding of probability provides a basis to understand statistical inference.

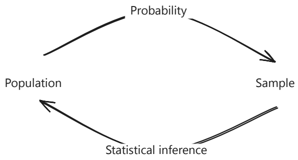
Ex The average height (in inches) of a male JSU student is $\mu_M = 68$ and the standard deviation is $\sigma_M = 3$ and for women $\mu_F = 63$ and $\sigma_F = 2.7$. Now suppose that someone believes that the average male JSU student height is $\mu_M = 63$. How can we show that this is not true without measuring almost every male JSU student? We can show that it is *probably* not true using statistical inference. Take a sample of heights

$$69.8, 68.0, 70.2, 70.6, 65.4, 72.0, 73.3, 63.1, 73.5, 69.1, 66.8, 64.8,$$

$$64.0, 66.4, 65.1, 70.8, 75.4, 71.5, 67.1, 66.0, 64.1, 64.9, 69.4, 68.2, 69.0$$

This $n = 25$ sample data set has a sample mean $\bar{x} = 68.4$ and sample standard deviation $s = 3.3$, but let's use $\sigma_M = 3$.

Population with know features + Probability: allow us to draw conclusions about characteristics of hypothetical data taken from the population.



Sample + Inferential Statistics: allow us to draw conclusions about the population.

Chapter 1 Section 2

Note to myself: Open mpg.csv

Definition 7 (Simple Random Sample)

Given a specified sample size, then every sample is as likely to be chosen as any other sample.

Definition 7 (Simple Random Sample)

Given a specified sample size, then every sample is as likely to be chosen as any other sample.

Ex Suppose that a city has two malls and you want to determine on average how many people visit each store per day. You could chose to use a sample of size of 10 and then

1. chose a mall at random
2. pick 10 stores at random.

Definition 7 (Simple Random Sample)

Given a specified sample size, then every sample is as likely to be chosen as any other sample.

Ex Suppose that a city has two malls and you want to determine on average how many people visit each store per day. You could chose to use a sample of size of 10 and then

1. chose a mall at random
2. pick 10 stores at random.

While this is a *random sample* it is not a simple random sample. Why?

Definition 7 (Simple Random Sample)

Given a specified sample size, then every sample is as likely to be chosen as any other sample.

Ex Suppose that a city has two malls and you want to determine on average how many people visit each store per day. You could chose to use a sample of size of 10 and then

1. chose a mall at random
2. pick 10 stores at random.

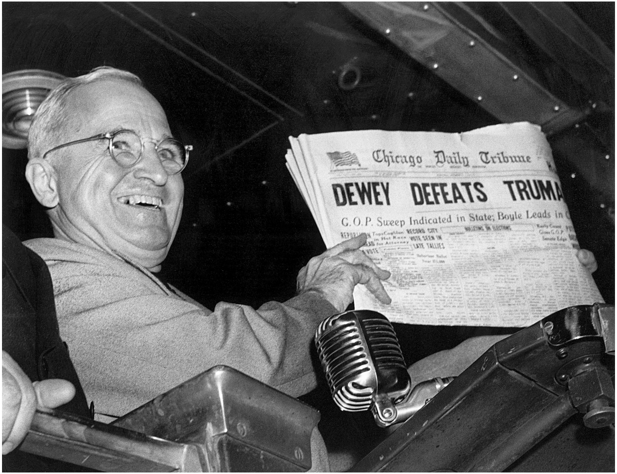
While this is a *random sample* it is not a simple random sample. Why?
Because while every store has an equally likely chance of being chosen, there is no way to get a store from mall 1 and mall 2 in a sample at the same time. Therefore there are samples that could never be chosen.

Definition 8 (Biased Sample)

is a sample that does not accurately represent the population (it over/under-represents some segment of the population).

Definition 8 (Biased Sample)

is a sample that does not accurately represent the population (it over/under-represents some segment of the population).



Chapter 1 Section 3
Measures of Location

Definition 9 (Sample mean)

If the observations in a sample are $x_1, x_2, \ldots, x_n$*, then the sample mean is*

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Definition 9 (Sample mean)

If the observations in a sample are $x_1, x_2, \ldots, x_n$, then the sample mean is

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Problem is that it is influenced by outliers. Note that as a single value gets larger so does the mean.

Ex Find the sample mean of

$$50000, 30000, 45000, 33000, 47000, 51000, 6744000$$

Definition 10 (Sample median)

If the observations in a sample are $x_1, x_2, \ldots, x_n$ *are arranged in* **increasing** *order, then the sample median is*

$$\tilde{x} = \begin{cases} x_{\frac{n+1}{2}}, & \text{if } n \text{ odd} \\ \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & \text{if } n \text{ even} \end{cases}$$

Ex Find the sample median of

$$50000, 30000, 45000, 33000, 47000, 51000, 6744000$$

Chapter 1 Section 4
Measures of Variability

Both data sets

$$1, 2, 3, 7, 8, 9$$

and

$$1, 1, 1, 9, 9, 9$$

have a mean of 5.

Both data sets

$$1, 2, 3, 7, 8, 9$$

and

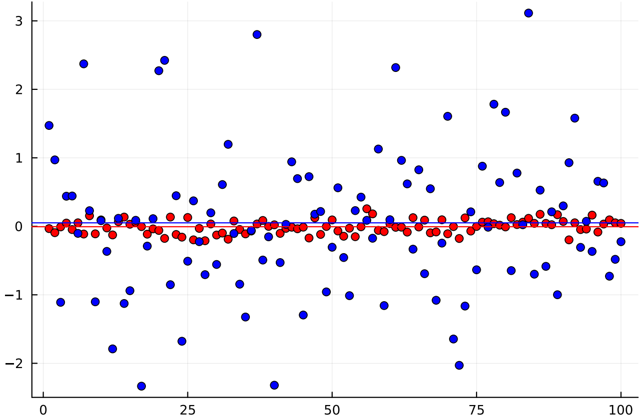
$$1, 1, 1, 9, 9, 9$$

have a mean of 5.
But the data sets are clearly different.

These two data sets of 100 points each have slightly different centers (means)

$$\mu_{red} = -0.006, \quad \mu_{blue} = 0.051$$

but there spreads are very different.



How should we measure this variation in the data?

How should we measure this variation in the data?

Definition 11

If the observations in a sample are $x_1, x_2, \ldots, x_n$*, then*

sample range $x_{\mathit{max}} - x_{\mathit{min}}$

How should we measure this variation in the data?

Definition 11

If the observations in a sample are $x_1, x_2, \ldots, x_n$, then

sample range $x_{max} - x_{min}$

sample variance

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

How should we measure this variation in the data?

Definition 11

If the observations in a sample are $x_1, x_2, \ldots, x_n$, then

sample range $x_{max} - x_{min}$

sample variance

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

sample standard deviation

$$s = \sqrt{s^2}$$

Ex Find the sample range, variance, and standard deviation for

$$1.7, 2.2, 3.9, 3.11, 14.7$$

Chapter 1 Section 5
Discrete and Continuous Distributions

Definition 12 (Discrete data)

can only take on certain values.

Definition 12 (Discrete data)

can only take on certain values.

Example 5.1

1. *Number of students in my class, e.g.* $0, 1, 2, \ldots, 35$
2. *FM radio station frequencies can be from 88.1 MHz to 108.1 (or 107.9) MHz with a step size of 0.2 MHz, e.g. 106.9, 100.7, 103.5, 95.3, 95.1*

Definition 13 (Continuous data)

can take on any value in a range.

Definition 13 (Continuous data)

can take on any value in a range.

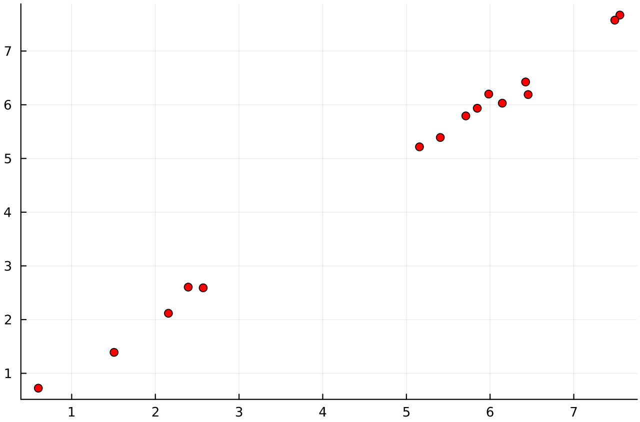
Example 5.2

Heights of adult humans can be anything from 20 *inches to* 110 *inches.*

Chapter 1 Section 6

Scatter plot

x	y
0.601	0.723
1.507	1.390
2.156	2.118
2.393	2.605
2.572	2.592
5.156	5.216
5.405	5.391
5.710	5.793
5.847	5.935
5.984	6.199
6.146	6.030
6.425	6.424
6.455	6.190
7.490	7.575
7.551	7.670



Steam and leaf

Ordered data:
$20, 20, 20, 21, 22, 23, 27, 33, 33, 35, 38, 39, 39, 45, 46, 47, 47, 48, 48,$
$50, 52, 52, 53, 53, 54, 54, 56, 56, 56, 58$

Steam and leaf

Ordered data:
$20, 20, 20, 21, 22, 23, 27, 33, 33, 35, 38, 39, 39, 45, 46, 47, 47, 48, 48, 50, 52, 52, 53, 53, 54, 54, 56, 56, 56, 58$
The stem are the digits excluding the least significant digit and they lie on the left of the plot.

Steam and leaf

Ordered data:
$20, 20, 20, 21, 22, 23, 27, 33, 33, 35, 38, 39, 39, 45, 46, 47, 47, 48, 48, 50, 52, 52, 53, 53, 54, 54, 56, 56, 56, 58$
The stem are the digits excluding the least significant digit and they lie on the left of the plot.
The leaves are what remains after removing the stem and are written to the right and are vertically aligned.

Steam and leaf

Ordered data:
$20, 20, 20, 21, 22, 23, 27, 33, 33, 35, 38, 39, 39, 45, 46, 47, 47, 48, 48,$
$50, 52, 52, 53, 53, 54, 54, 56, 56, 56, 58$

The stem are the digits excluding the least significant digit and they lie on the left of the plot.

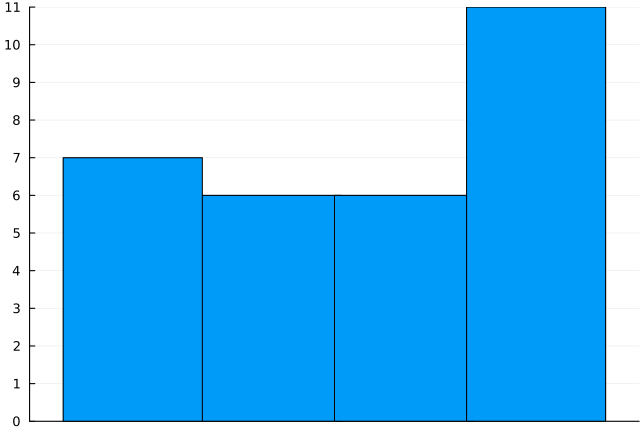
The leaves are what remains after removing the stem and are written to the right and are vertically aligned.

2	0001237
3	335899
4	567788
5	02233446668

Key $3|5 = 35$

Histogram

Sam data as before:
$20, 20, 20, 21, 22, 23, 27, 33, 33, 35, 38, 39, 39, 45, 46, 47, 47, 48, 48,$
$50, 52, 52, 53, 53, 54, 54, 56, 56, 56, 58$



Chapter 2 Section 1

Definition 14 (Experiment)

Definition 15 (Sample space)

Ex Flip a coin. What is the sample space as a set and a tree.

Ex Flip a coin twice. What is the sample space as a set and a tree.

Ex Flip a coin, if H roll a 4-sided die, else flip the coin again. What is the sample space as a set and a tree.

Ex Flip a coin until you get a H. What is the sample space as a set and a tree.

Ex Consider the set S to be all of the points that make up a circle with radius 5 that is centered at the origin.
List the sample space.

Chapter 2 Section 2

Definition 16 (Event)

Definition 17 (Null (Empty) Set)

Denoted by $\emptyset$.

Ex Flip a coin. List all events.

Ex Flip a coin twice.

1. $E_1 \equiv$ flip exactly one H.

Ex Flip a coin twice.

1. $E_1 \equiv$ flip exactly one H.
2. $E_2 \equiv$ flip at least one H.

Ex Flip a coin twice.

1. $E_1 \equiv$ flip exactly one H.
2. $E_2 \equiv$ flip at least one H.
3. $E_3 \equiv$ the second flip is the same as the first.

Ex Flip a coin twice.

1. $E_1 \equiv$ flip exactly one H.
2. $E_2 \equiv$ flip at least one H.
3. $E_3 \equiv$ the second flip is the same as the first.
4. $E_4 \equiv$ the coin lands on it's edge.

Ex Let $S = \{x : x \geq 0\}$ be the distance Forrest Gump ran before he tripped...