- Understanding distributions
- Assessing a model
- Causality and correlations

Five-figure summary

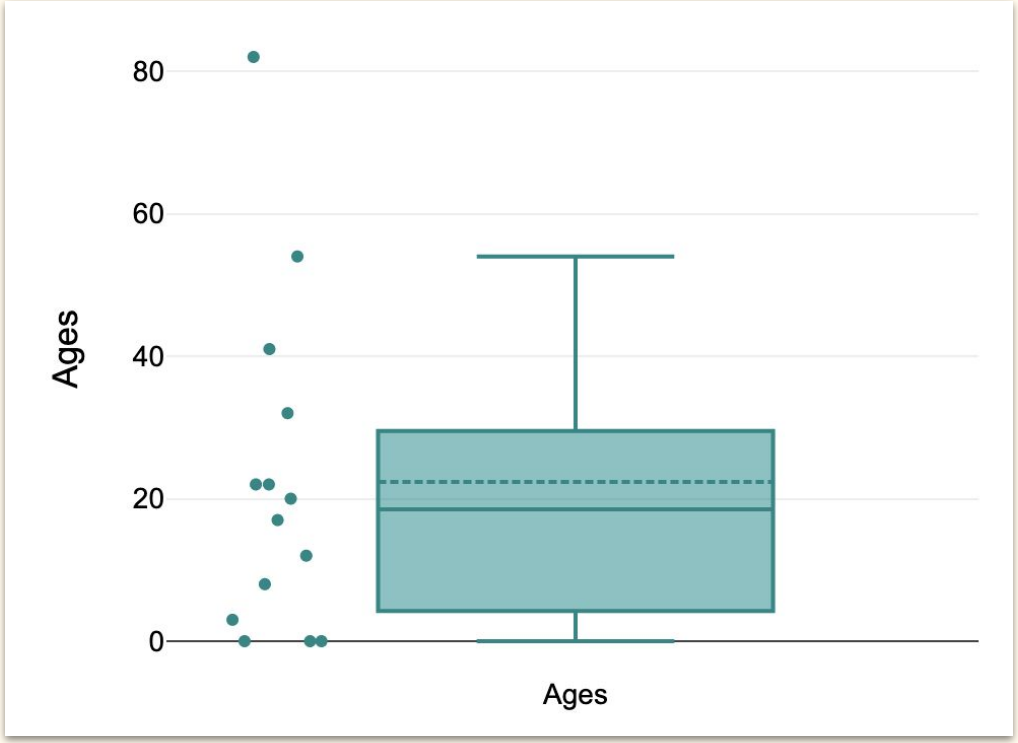
- Median
- Lower quartile
  - Marks the value that represents 25% of the data
- Upper quartile
  - Marks the value that represents 75% of the data
- Minimum observation
- Maximum observation

*mean is occasionally used and presented in a five-figure summary

Presenting the five figure summary

Ages:

8, 20, 82, 3, 17, 0, 0, 22, 12, 22, 54, 0, 32, 41

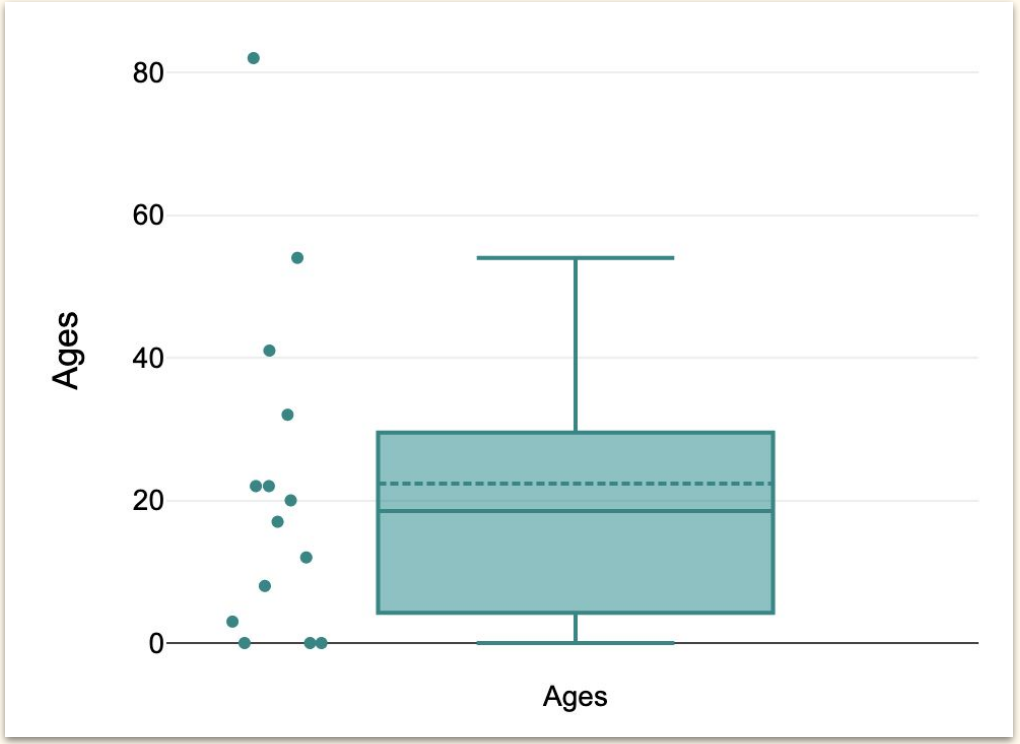


Presenting the five figure summary

Ages:

[0, 0, 0, ***3***, 8, 12, 17],

[20, 22, 22, ***32***, 41, 54, 82]



Practice: Five figure summary

Let’s plot our class’ ideal temperature for spring!

You will need to calculate the:

- Median and mean
- Upper quartile
- Lower quartile
- Minimum observation
- Maximum observation

Practice: Five figure summary

60	65	Around 70 degrees	maybe around the 60s, 74 the highest	65
70 degrees	75 Fahrenheit	67	65 degrees!	22
65-70 Degrees	50-70 degrees	75 degree Fahrenheit	70 degrees	60 degrees Fahrenheit
60* F	I like it cool	65	70	62
70 degrees	65 degrees F	70	Between 57-62	67

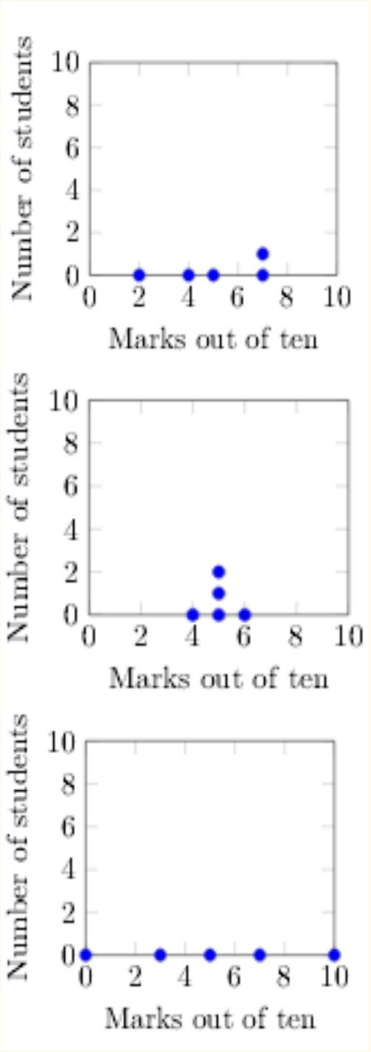
Practice: Five figure summary

67	74	70	70	55-60 with a nice breeze
65	35	20	65 degrees	65
65-70	72	70	70 winds	70
50	60	50 degrees	70 degrees	70

Spread and Distribution

Purpose of standard deviation

Subject	Marks out of ten	Mean Average	Median average
French	2, 4, 5, 7, 7	5	5
Religious Studies	0, 5, 10, 7, 3	5	5
History	5, 5, 4, 6, 5	5	5



Calculating the standard deviation (population)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

Calculating the standard deviation

- Calculate the mean
- Calculate the deviation
  - Difference between the observation and the mean
- Calculate the sum of the squared deviation

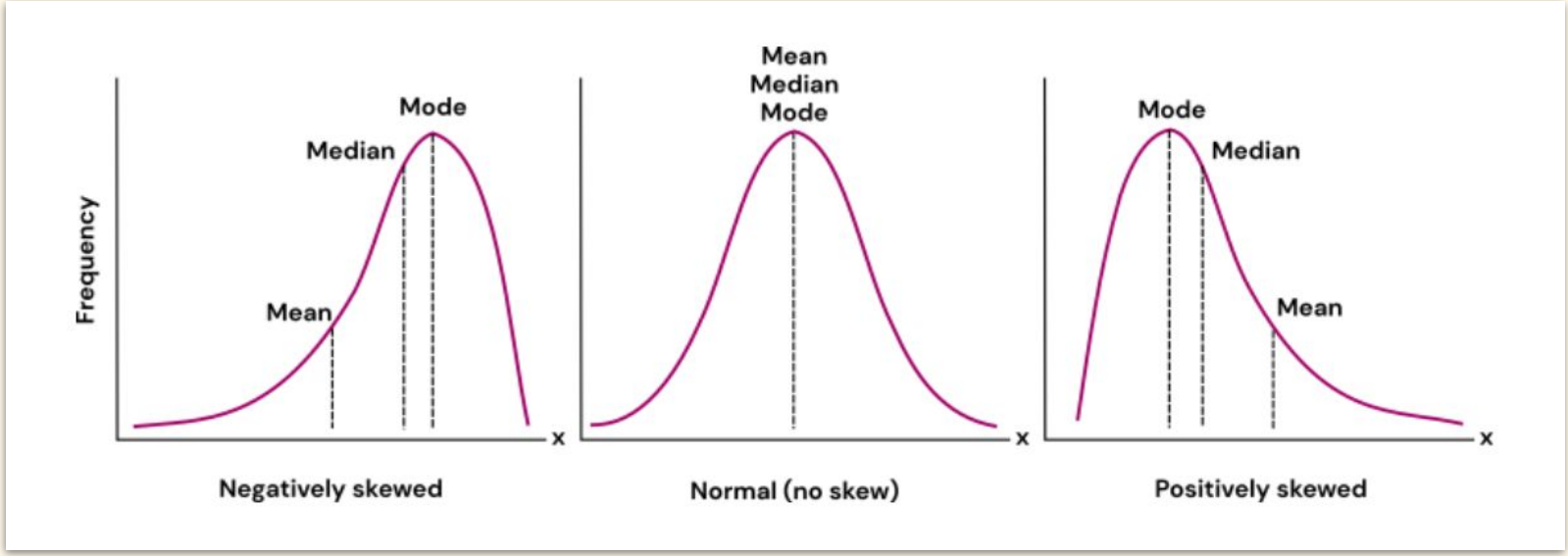
Calculating the standard deviation

- Calculate the variance
  - How spread out is the data
- Calculate the standard deviation
  - Square root of variance

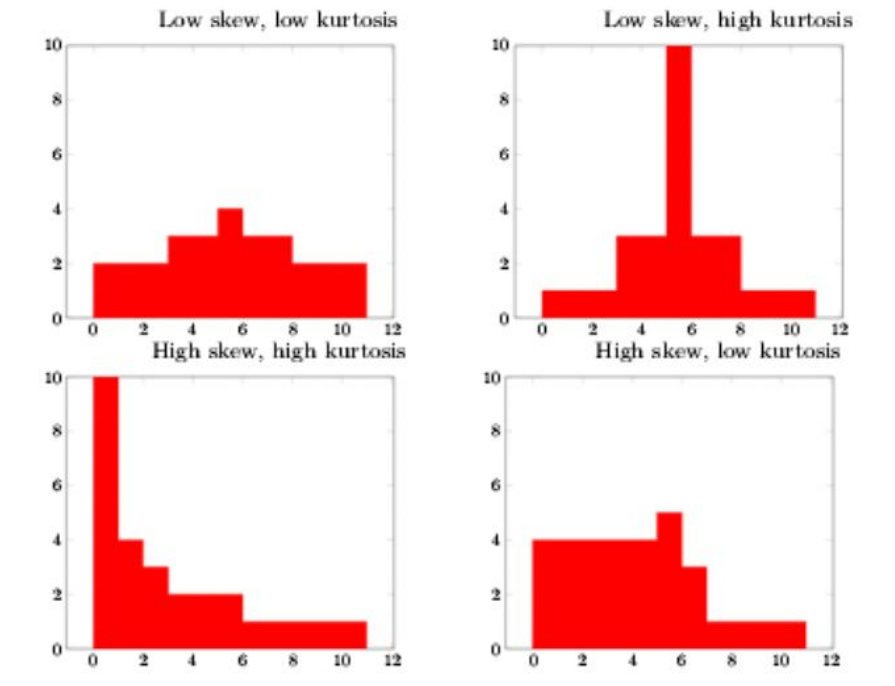
$$Variance = \frac{sum\ of\ the\ squared\ deviations}{number\ of\ observations}$$

Understanding skewness and kurtosis

Skewness


Frequency
Mode
Median
Mean
x
Negatively skewed
Mean
Median
Mode
x
Normal (no skew)
Mode
Median
Mean
x
Positively skewed

Distribution of skew and kurtosis



Assessing a model

Example: Jury selection

- An *impartial* jury should be representative of the population of the relevant region
  - Final trial jury selected from group of prospective jurors by deliberate inclusion/exclusion

Example: Jury selection

- Supreme Court case of Robert Swain
  - Black men convicted and sentenced to death by all white jury in Talladega County, Alabama, 1962
  - Alabama supreme court declared that that jury was constitutionally composed

Example: Jury selection

At the time of this trial

- Only men aged 21 and above and eligible
- 26% of the men in the population identified as Black
- Yet, only 8% of the representative population of eligible jurors were Black

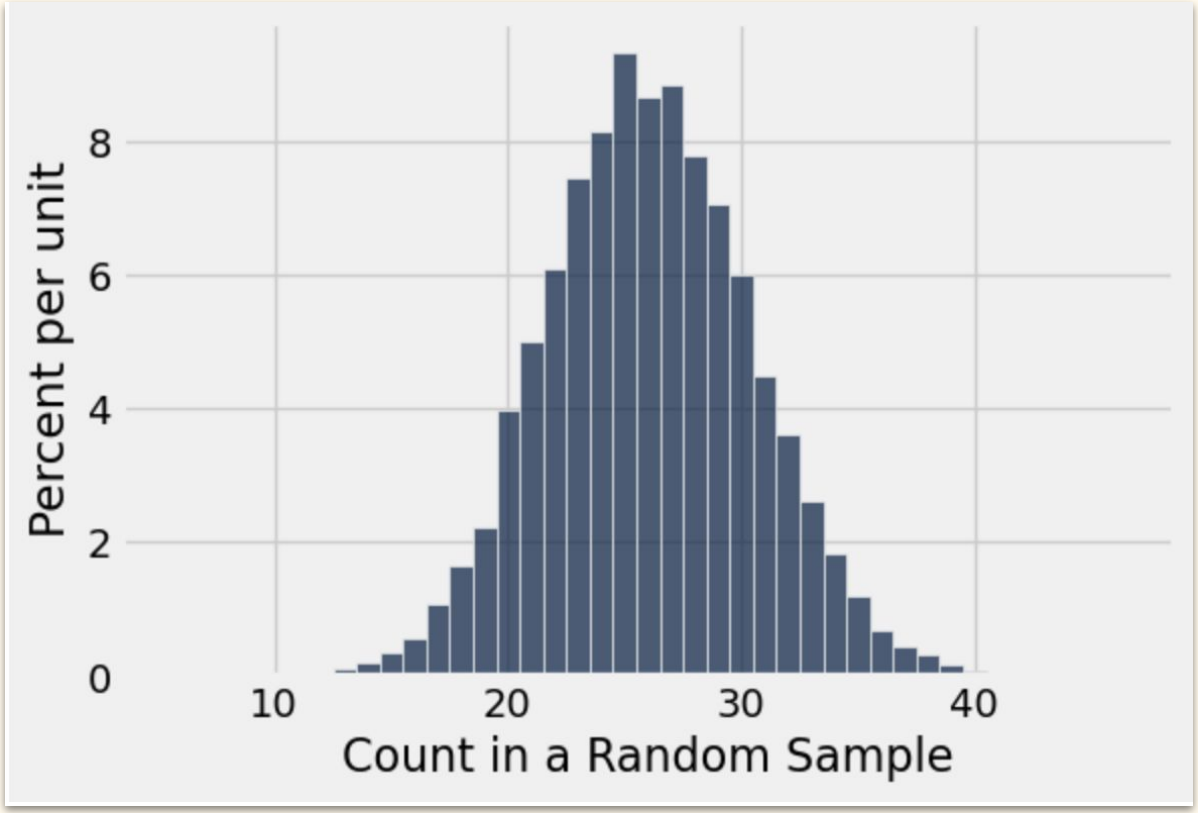
U.S. Supreme Court concluded ““the overall percentage disparity has been small.”

Example: Jury selection

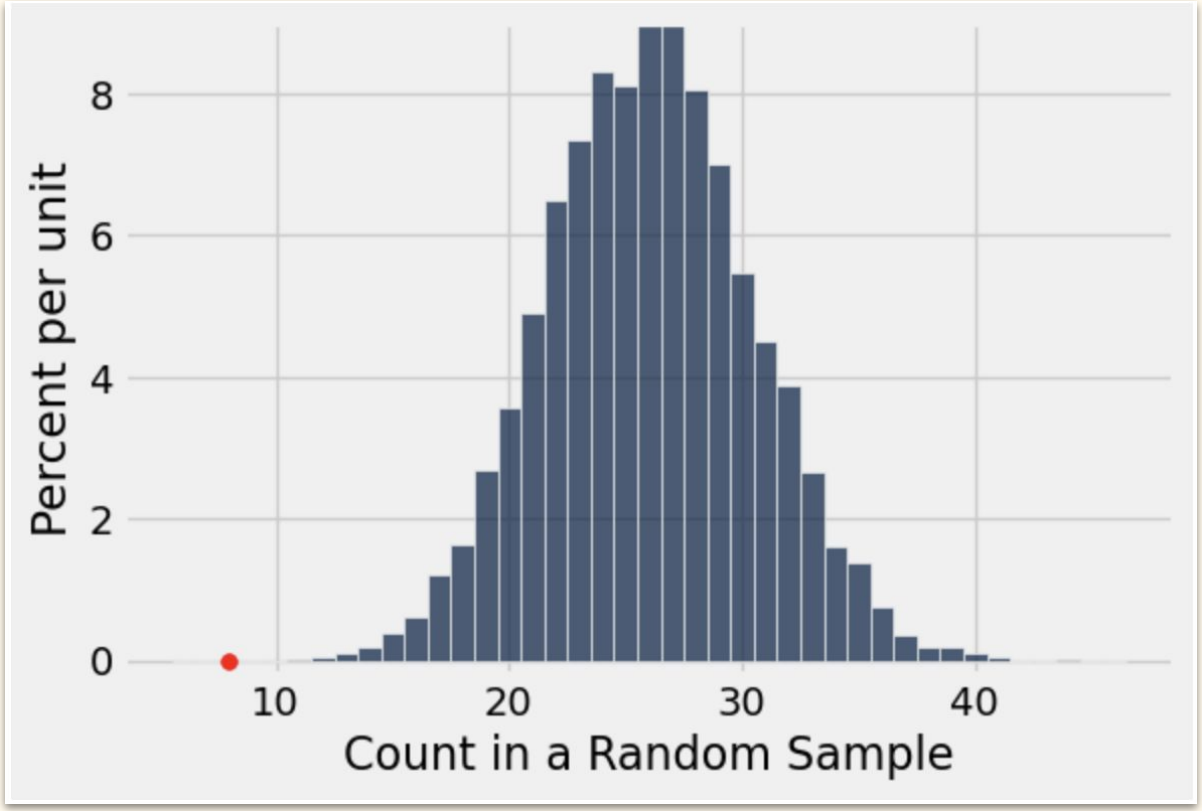
With a population of whom 26% are Black, how likely is it to draw a panel that Black folx only make up 8%?

1. Simulate data based on the model
    a. 26% Black, 74% White and others
2. Simulate drawing at random from this population
3. Demonstrate the chances of this panel (8%) being selected at random
    a. Is this panel likely to happen at random (therefore small disparity)?

Prediction under model of random selection



Comparing the prediction and data



So, what can we conclude?

- If we select a panel of size 100 at random it is *very unlikely* to get the counts that we saw at Robert Swain's trial
  - *Very unlikely* that this panel is drawn by chance, with the model of random selection we simulated
- We can reasonably assume that this panel was ***not selected*** by random sampling from eligible jurors
  - Difference between 26% and 8% is not so small as to be explained well by chance alone

Why did this happen?

- Jury panels selected from a jury roll of names that jury commissioners acquired
  - Often in favor of people in the commissioners' social and professional circles
- When there are Black panelists in the selection pool, most don't make it to the final jury panel

Causality and correlations

John Snow and the Broad Street pump

Correlations ≠ causation

- Observational studies can help us establish a link between 2 variables
  - Could be 2 phenomenons happening at the same time
  - Not always a situation where phenomenon A causes phenomenon B

Correlations ≠ causation

- Confounding factor(s)
  - Coffee and lung cancer
  - Ice cream sales and rate of drowning

Role of randomization

- Assignning individuals to the treatment and control groups at random
- Randomized controlled trial (RCT) v. observational study
  - Treatment groups
  - Control groups

Role of randomization

- To account—mathematically—for the possibility that randomization produces treatment and control groups that are quite different from each other
- To make precise mathematical statements about differences between the treatment and control groups
  - Statistically significant
  - Make justifiable conclusions about whether the treatment has any effect