# Week 3 - Concepts in Mathematical Modelling

**Introduction:**

Mathematical modeling is a critical tool for understanding, analyzing, and solving complex real-world problems across various disciplines. In this module, we will delve into the core concepts of mathematical modeling and how it is applied in practice.

**Learning objectives**

- Understand what models and mathematical models are.
- Think about a real-world problem from a modelling perspective
- Framework of modelling
- Have a broad understanding of mathematical models and the scope of its application

**What is a model?**

There are three components of a model.

- Represents a phenomena
- Simple
- Flexible

**An example of a Physical Model** - Normally a prototype (constructed copy) of a real object - Examples: model building, bridges, airplane, solar system

**An example of a Schematic Model** - Normally used to visualize a system structure - Examples: Organisation charts, Maps, Diagrams, Process charts

**An example of a Mathematical Model** - A simple and flexible representation of a real object or phenomena using mathematical language - Examples: Algebraic equations, Differential equation, Algorithm - For data science: descriptive and inferential statistics

## Why Mathematics?

1. Mathematics is a precise language

- Forces us to formulate concrete ideas and assumptions in an unambiguous way.

2. Mathematics is a concise language

- One equation says more than 1000 words

3. Mathematics is a universal language

- Same mathematical techniques can be applied over a range of scales.

4. Mathematics is the language that computers understand best.

So the objective of mathematical and statistical modelling is to analyse, make predictions and provide insights of real-world phenomena and the applications of mathematical modelling stretch far and wide.

Examples: - Epidemiology - Analysis of financial markets - Strategies for cooperative and noncooperative interactions - Software development - Economics (term of trades, rates, surplus analysis) - Electoral studies - Astronomy - Organisational behavious

However, there are some limitations of mathematical modelling. Mathematical modelling is only as good as the data and assumptions you provide it and even then you will never be able to model all the variables required for a model to be 100% accurate.

## Definition of Mathematical Modeling:

Mathematical modeling is the process of using mathematical equations, algorithms, and simulations to represent, analyze, and make predictions about real-world phenomena. It involves creating simplified, abstract representations of complex systems to gain insights and solve practical problems.

## Definition of Statistical Modeling

Statistical modeling is a process used to describe and understand the relationships between variables in a dataset. It involves the use of statistical techniques to analyze data, make predictions, and draw conclusions about the underlying patterns and structures within the data.

**Key Concepts in Mathematical Modeling:**

1. Abstraction:

Abstraction is the process of simplifying a complex real-world problem by focusing on its essential features while ignoring irrelevant details. Models are abstractions that capture the critical aspects of a system.

2. Variables:

Variables are quantities that can change in a model. They represent important parameters or properties of the system. Variables can be categorized as independent (inputs), dependent (outputs), or control variables.

3. Equations:

Mathematical models are typically represented by mathematical equations that describe the relationships between variables. Equations can be algebraic, differential, integral, or a combination of these.

4. Assumptions:

Assumptions are simplifications and constraints applied to the model to make it tractable. Assumptions should be justified and clearly stated in the modeling process.

5. Parameters:

Parameters are constants in the model that affect its behavior. Determining parameter values often requires data collection or expert knowledge.

6. Validation and Verification:

Validation is the process of testing a model's accuracy by comparing its predictions to real-world observations. Verification ensures that the model is implemented correctly and behaves as intended.

7. Simulation and Analysis:

Models are used to perform simulations to study the system's behavior under different conditions. Analysis of model results helps in drawing conclusions and making decisions.

8. Sensitivity Analysis:

Sensitivity analysis assesses how changes in parameters or assumptions affect the model's predictions. It helps identify the most critical factors influencing the system.

9. Model Selection:

Depending on the problem, different modeling techniques and approaches may be suitable, such as deterministic or stochastic models, continuous or discrete models, etc.

10. Interpretation:

The interpretation of model results is crucial for drawing meaningful conclusions and making decisions. It involves translating mathematical findings into actionable insights.

11. Communication:

Effective communication of modeling results to stakeholders is essential. Reports, presentations, and visualizations help convey the model's findings.

## Key Concepts in Statistical Modelling

1. Variables:

In statistical modeling, variables are characteristics or attributes that can take different values. They can be classified into different types, such as categorical (e.g., gender, color) or continuous (e.g., age, temperature).

2. Data Distribution:

Understanding the distribution of the data is crucial in statistical modeling. This includes assessing the shape, center, spread, and any potential outliers or anomalies in the data.

3. Probability Distributions:

Probability distributions describe the likelihood of observing different outcomes in a dataset. Common probability distributions used in statistical modeling include the normal (Gaussian), binomial, Poisson, and exponential distributions.

4. Parameters:

Parameters are the unknown quantities in statistical models that need to be estimated from the data. They represent the characteristics of the population being studied and are typically denoted by Greek letters (e.g.,  for population mean,  for population standard deviation).

5. Estimation:

Estimation involves the process of determining the values of unknown parameters based on observed data. Estimators are statistical techniques used to estimate parameters, and they can be point estimators (providing a single value estimate) or interval estimators (providing a range of plausible values).

6. Hypothesis Testing:

Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data. It involves formulating null and alternative hypotheses, calculating test statistics, and determining whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis.

7. Regression Analysis:

Regression analysis is a statistical technique used to model the relationship between one or more independent variables (predictors) and a dependent variable (outcome). It is widely used for prediction and inference in various fields.

8. Model Assumptions:

Every statistical model is based on certain assumptions about the data. These assumptions need to be carefully assessed to ensure the validity of the model and the reliability of the inferences drawn from it.

9. Model Selection and Validation:

Model selection involves choosing the most appropriate model among competing alternatives based on criteria such as goodness-of-fit, simplicity, and interpretability. Model validation involves assessing the performance of the selected model using techniques such as cross-validation or independent data validation.

10. Uncertainty and Confidence Intervals:

Uncertainty quantifies the degree of variability or randomness in statistical estimates. Confidence intervals provide a range of plausible values for a parameter estimate, along with a level of confidence that the true parameter value lies within that range.

## Four stages of mathematical modelling

**Stage 1:** Building the model

1. Define the model objectives

- Be clear about what you want your model to do

2. Determine the appropriate level & key model components

- What level of simplification is required?
- Apply the principle of Ockham's razor

3. Define your assumptions

- Arguably the most important step in building a model

- Assumptions reflect our understanding how the phenomena occurs
- Remember: the model results are only as valid as the asssumptions!
- Different assumptions can lead to fundamental differences

4. Create a flow diagram

- Useful especially if your objective is to understand the relationship between variables
- A visual tool to simplify an otherwise complex model

5. Develop your model

- Depends on the objectives and your modelling approach – Statistical models normally use a single linear or non-linear function – Deterministic models of dynamical system normally use differential equations – Stochastic models require expression for the probability of events

**Stage 2:** Analysing the model

1. Analytical method (using mathematical principle)

- Precise! Provide an exact solution
- Usually only possible for simple systems or problems with strict assymptions
- In pure mathematics, the analysis of a pattern may lead to a proposition!

2. Numerical method (using computers)

- Applies to moth mathematical model: Widely use in many fields
- Requires the use of numerical algorithms implemented in computational systems (e.g. euler methods, monte-carlo simulations)
- Provide an approximate solution
- Use establish code. Avoid writing your own numerical solver as much as possible. You will be able to use existing libraries to create mathematical models. This might save a lot of time and cost especially in the experimental study.

3. Uncertainty analysis

- Normally found in stochastic (probabiliistic) model.
- Analysing the variable in output from the variablility in input e.g. How confident are we on the model output?

4. Sensitivity analysis

- Analyse how sensitive the output of the model when certain parameters (input) are changed. e.g. When we change our input, by how much does the output change?

**Stage 3.** Validating the model

1. Pure mathematics

- The process of writing out the proof of our proposition
- There are many techniques in proving a theorem: direct proof, proof by cases, proof by contradiction, proof by induction, etc.)

2. Applied mathematics (Predominantly statistical model)

- Comparison of model prediction to the real data (observations).
- In most cases, these statistical techniques are useful to check the model's robustness.
- Hypothesis testing
- Analysis of standard deviation and Mean-Squared Error (MSE)
- Identify why the model prediction may differ with the real data: natural variablility (2008 financial crash), important variables ignored in the model, etc

**Stage 4.** Applying the model

1. Pure mathematics

- Decide on the term of a mathematical statement. Does your solution consitute as a Theorem, Lemma, Corollary, or simply a generalization?
- Define the possible applications and extension to your model and findings.

2. Applied mathematics (Predominantly statistical model)

- Valuable decision support tools!
- In Economics and Finance, a model may offer several options for public decision-making
- Mathematical models such as linear programming and game theory are very popular because they have a wide range of applications.

## What makes a good model?

The key attributes of a good model.

1. Fit for purpose

- Precisely and concisely solves the real-world problem for the end users
- Appropriate balance between accuracy, simplicity and flexibility

2. For statistical predictive model: How easy is it to use your model for different types of data (replicability)

## Drawbacks of Modelling

Assumptions: Many statistical and mathematical models rely on strict assumptions about the data, such as linearity, normality, independence, and homoscedasticity. Violations of these assumptions can lead to biased estimates and incorrect inferences.

Overfitting: Overfitting occurs when a model captures noise or random fluctuations in the data rather than the underlying patterns. This can result in a model that performs well on the training data but fails to generalize to new, unseen data.

Underfitting: Underfitting occurs when a model is too simplistic to capture the underlying structure of the data. This can result in poor predictive performance and biased parameter estimates.

Complexity and Interpretability: Some models, especially those based on machine learning algorithms, can be highly complex and difficult to interpret. This lack of interpretability can make it challenging to understand the underlying mechanisms driving the model's predictions.

Data Requirements: Many statistical and mathematical models require large amounts of data to produce reliable estimates and predictions. In situations where data are scarce or incomplete, these models may perform poorly or fail altogether.

Data Quality: The quality of the data used to train and validate statistical models can significantly impact their performance. Missing values, measurement errors, and outliers can all distort the results and lead to incorrect conclusions.

Computational Complexity: Some models, particularly those involving large datasets or complex algorithms, can be computationally intensive and require substantial computational resources and time to train and evaluate.

Assumption of Causality: Statistical models often establish associations between variables but may not necessarily imply causality. Inferring causal relationships from observational data can be challenging and may require additional methods such as experimental design or causal inference techniques.

Model Uncertainty: Statistical models provide point estimates or predictions along with measures of uncertainty, such as confidence intervals or prediction intervals. However, these measures may not fully capture all sources of uncertainty, leading to overly optimistic or conservative estimates.

Subjectivity and Bias: The choice of model, variables, and analysis techniques can be influenced by subjective judgments and biases. Researchers may inadvertently introduce bias into the modeling process, leading to erroneous conclusions.

## Definitions

Validity and Generalizability are crucial concepts that assess the quality and applicability of model findings:

Validity: Validity refers to the extent to which a research study accurately measures or assesses the phenomenon it claims to investigate. It assesses whether the study design, methods, and instruments effectively capture the intended constructs or variables under study. Validity encompasses various aspects, including:

Internal Validity: Internal validity concerns the degree to which the observed effects or relationships within a study can be attributed to the manipulated variables or interventions, rather than confounding factors or biases.

External Validity: External validity, also known as ecological validity, refers to the extent to which the findings of a study can be generalized or applied to populations, settings, or conditions beyond those directly examined in the study. It assesses the relevance and representativeness of the study's results beyond the specific context in which they were obtained.

Generalizability: Generalizability pertains to the extent to which the findings of a research study can be extended or applied to populations, contexts, or situations beyond the specific sample or conditions studied. It reflects the degree of confidence researchers can have in extrapolating the study results to broader populations or real-world scenarios. Factors influencing generalizability include:

Sample Representativeness: The extent to which the characteristics of the study sample accurately reflect those of the target population of interest. A representative sample enhances the generalizability of study findings.

Contextual Factors: Consideration of contextual variables, such as cultural, social, or environmental factors, that may influence the applicability of study findings across different settings or populations.

Study Design and Methodology: The rigor and appropriateness of the study design, sampling methods, and data collection procedures can impact the generalizability of the findings. Well-designed studies with robust methodologies tend to yield more generalizable results.

External Validity: As mentioned earlier, the external validity of a study is closely related to its generalizability. Ensuring that study conditions and settings are representative of real-world contexts enhances the external validity and, consequently, the generalizability of the findings.

In summary, validity ensures that a research study accurately measures what it intends to measure, while generalizability assesses the extent to which the study findings can be applied to broader populations or contexts beyond the specific conditions of the study. Both concepts are essential for establishing t

## Biases in mathematical modelling

Biases in mathematical modeling can significantly affect the accuracy and fairness of the models' outcomes. These biases can stem from various sources, including the data used, the assumptions made by the modelers, and the interpretation of the results. Understanding and addressing these biases is crucial for developing models that provide reliable and equitable predictions or analyses. Here are some common forms of biases in mathematical modeling:

**Data Bias:** This occurs when the dataset used to train or inform the model does not accurately represent the real-world scenario it aims to simulate or predict. Data can be biased due to underrepresentation of certain groups, overrepresentation of others, or simply because it contains errors or inaccuracies. For instance, if a model designed to predict creditworthiness is trained on data that lacks diversity, it may unfairly disadvantage certain groups of people.

An example of data bias can be found in the context of hiring practices. Imagine a company uses an algorithm to screen job applicants based on their resumes. However, historical hiring data used to train the algorithm shows a bias towards selecting candidates from certain demographic groups, such as white males, over others.

This bias could stem from various sources, such as:

Historical Hiring Practices: If the company has historically hired predominantly from one demographic group due to implicit biases or structural inequalities in the hiring process, the algorithm may inadvertently learn and perpetuate this bias.

Biased Training Data: If the training data used to develop the algorithm is not representative of the diverse applicant pool, it may result in biased outcomes. For example, if resumes from certain demographic groups are underrepresented or systematically overlooked, the algorithm may not accurately assess candidates from those groups.

Proxy Variables: The algorithm may inadvertently use proxy variables that correlate with demographic factors, such as zip code or educational background, leading to discriminatory outcomes. For instance, if the algorithm favors candidates from affluent neighborhoods or prestigious universities, it may disadvantage applicants from marginalized communities.

In this scenario, the algorithm, despite its intention to streamline the hiring process, perpetuates systemic biases and discrimination against certain demographic groups. This highlights the importance of carefully examining training data, identifying potential sources of bias, and implementing measures to mitigate bias in algorithmic decision-making processes.

**Selection Bias:** Selection bias happens when the process used to collect data introduces an inherent bias. For example, if a health study only includes volunteers who are inherently healthier, more active, or more health-conscious than the general population, the results can't be accurately generalized to all demographics.

An example of selection bias can be observed in a study examining the effectiveness of a new medication for a particular medical condition. Suppose researchers recruit participants for the

study by advertising in a health magazine. However, the magazine's readership predominantly consists of individuals who are highly health-conscious and proactive about seeking medical treatment.

In this scenario:

Selection Process: By recruiting participants solely from the readership of the health magazine, the researchers unintentionally select a sample that may not be representative of the broader population of individuals with the medical condition.

Exclusion of Certain Groups: Individuals who do not read the health magazine or who are less proactive about seeking medical treatment may be underrepresented in the study. This could include individuals from lower socioeconomic backgrounds, ethnic minorities, or those with limited access to healthcare.

Impact on Study Results: As a result of the selection bias, the study's findings may not accurately reflect how the medication performs in the general population. The efficacy and safety outcomes observed in the study may be exaggerated or misrepresented due to the skewed sample composition.

Generalizability: The study's findings may not be applicable to populations that were not adequately represented in the sample. Therefore, healthcare professionals may struggle to make informed decisions about prescribing the medication to diverse patient populations.

In this example, selection bias compromises the validity and generalizability of the study results, highlighting the importance of recruiting a representative sample and minimizing biases in participant selection to ensure the reliability of research findings.

**Confirmation Bias:** This is a type of cognitive bias that can affect the way modelers interpret data or results, leading them to favor outcomes that confirm their preexisting beliefs or hypotheses. This can inadvertently influence the model's development and its conclusions.

Imagine a financial analyst tasked with forecasting the performance of a particular stock. This analyst has a bullish outlook on the stock due to prior research, personal beliefs, or pressure from stakeholders. As they construct their financial model, they may unintentionally exhibit confirmation bias in the following ways:

Data Selection: The analyst may selectively choose historical data or performance metrics that support their bullish thesis while disregarding or downplaying data points that suggest a more cautious or bearish outlook. For instance, they might focus on periods of strong growth in the stock's price while ignoring instances of volatility or decline.

Parameter Assumptions: In building the model, the analyst may make assumptions about key parameters, such as revenue growth rates, profit margins, or market share, that align with their bullish perspective. They might rely on optimistic forecasts or industry projections that reinforce their existing beliefs about the stock's potential.

Interpretation of Results: When running simulations or analyzing the outputs of the model, the analyst may disproportionately emphasize findings that confirm their bullish outlook while downplaying or rationalizing any contradictory results. They might focus on scenarios where the stock performs exceptionally well while dismissing scenarios of underperformance as outliers or anomalies.

Feedback Loop: As the analyst iterates on the model and refines their assumptions based on preliminary results, confirmation bias can create a feedback loop wherein each iteration reinforces their initial bullish perspective. They may adjust parameters or tweak the model structure in ways that further align with their preconceived beliefs, rather than objectively evaluating alternative scenarios.

In this example, confirmation bias can distort the financial modeling process, leading to over-optimistic projections and potentially flawed investment decisions. By selectively seeking and interpreting information that confirms their existing beliefs, the analyst risks overlooking potential risks or downside scenarios, ultimately undermining the reliability and usefulness of the model for decision-making.

**Algorithmic Bias:** Algorithmic bias occurs when the algorithms that underpin models systematically produce outcomes that are biased towards certain groups or individuals. This can be due to the way the algorithms are designed, the data they're trained on, or the objectives they're set to optimize.

Suppose a tech company develops a facial recognition algorithm designed to identify individuals in images or videos for various applications, such as security systems or automated identity verification processes. The algorithm is trained on a large dataset of facial images collected from diverse sources.

Here's how algorithmic bias can manifest in this scenario:

Data Collection Bias: The dataset used to train the facial recognition algorithm may contain an overrepresentation of certain demographic groups, such as individuals of predominantly white or male ethnicity. If the dataset lacks diversity in terms of race, ethnicity, gender, age, or other factors, the algorithm may perform poorly when attempting to recognize faces from underrepresented groups.

Feature Representation Bias: The algorithm may prioritize certain facial features or characteristics that are more prevalent in the training data, leading to biases in face detection and recognition. For example, if the training data primarily consists of images of individuals with lighter skin tones, the algorithm may struggle to accurately detect and identify faces of individuals with darker skin tones due to differences in feature representation.

Algorithm Design Bias: The underlying design and architecture of the facial recognition algorithm may inadvertently encode biases present in the training data. For instance, if the algorithm relies on predefined templates or templates generated from biased training data, it may produce biased or inaccurate results, particularly for faces that deviate from the dominant characteristics in the dataset.

Performance Disparities: Biases in facial recognition algorithms can result in performance disparities across different demographic groups. For example, studies have shown that facial recognition systems tend to have higher error rates when identifying faces of people with darker skin tones or female faces, compared to faces of lighter-skinned individuals or male faces.

Ethical Implications: Biased facial recognition algorithms can have serious ethical implications, including privacy violations, wrongful identifications, and exacerbation of social inequalities. Misidentifications and false positives can lead to wrongful arrests or discriminatory treatment, particularly for individuals from marginalized communities who are already disproportionately impacted by biased law enforcement practices.

In this example, algorithmic bias in facial recognition technology highlights the importance of addressing biases in training data, algorithm design, and deployment to ensure fair and equitable performance across diverse populations. Efforts to mitigate algorithmic bias in facial recognition systems should prioritize data diversity, transparency, accountability, and stakeholder engagement to promote ethical and responsible use of the technology.

**Simplicity Bias:** In an effort to make models tractable and understandable, modelers might oversimplify complex systems. This can lead to models that fail to capture critical dynamics of the systems they aim to represent, potentially leading to inaccurate predictions or analyses.

Simplicity bias refers to the tendency to favor simple explanations or solutions over more complex ones, often leading to oversimplification of complex phenomena. One example of simplicity bias can be seen in medical diagnoses.

Consider a patient presenting with a variety of symptoms that could indicate several potential medical conditions. A doctor, influenced by simplicity bias, might be inclined to jump to a straightforward diagnosis based on the most obvious symptoms without fully considering other possibilities. For instance, if the patient exhibits fatigue and weight loss, the doctor might quickly diagnose them with a common condition like depression or a thyroid disorder, without thoroughly investigating less obvious but potentially serious conditions like cancer or autoimmune diseases.

In this scenario:

Simplistic Diagnosis: The doctor's tendency towards simplicity bias leads them to favor a diagnosis that aligns with the most common or easily recognizable symptoms, potentially overlooking less common or more complex underlying conditions.

Failure to Consider Alternatives: By focusing on a simple diagnosis, the doctor may neglect to explore alternative explanations for the patient's symptoms, missing important clues that could lead to a more accurate diagnosis.

Delayed or Missed Diagnoses: Simplicity bias can result in delayed or missed diagnoses of serious medical conditions, as the doctor may dismiss symptoms that do not fit the simplistic explanation or fail to consider less obvious but potentially life-threatening possibilities.

Patient Harm: Patients may suffer harm due to misdiagnosis or delayed diagnosis resulting from simplicity bias, as their underlying medical conditions go untreated or are inadequately managed.

In this example, simplicity bias in medical diagnosis underscores the importance of thorough and systematic evaluation, considering a wide range of potential explanations for symptoms, and avoiding the trap of prematurely settling on a simplistic diagnosis. Effective medical decision-making requires a balanced approach that acknowledges the complexity of individual cases and prioritizes accuracy and patient safety over simplicity.

**Social and Ethical Bias:** These biases arise when models, intentionally or not, reinforce or perpetuate social inequalities or ethical issues. For example, models used in criminal justice or hiring processes might inadvertently encode societal biases against certain demographic groups.

An example of social and ethical bias can be observed in the use of predictive algorithms in the criminal justice system, particularly in the context of risk assessment for pretrial detention or sentencing decisions.

Suppose a court system adopts a predictive algorithm designed to assess the likelihood of a defendant reoffending or failing to appear for future court dates. The algorithm utilizes historical data on previous criminal cases, including demographic information, criminal history, and socioeconomic factors, to generate risk scores that inform judges' decisions on pretrial release conditions or sentencing recommendations.

Here's how social and ethical bias can manifest in this scenario:

Data Bias: The historical data used to train the predictive algorithm may reflect systemic biases in the criminal justice system, such as racial disparities in arrest rates or disproportionate representation of certain demographic groups in the criminal justice system. If the algorithm learns from biased data, it may perpetuate or exacerbate existing disparities in decision-making.

Risk Assessment Bias: The algorithm may inadvertently encode social biases into its risk assessment process, leading to unfairly harsh treatment or overrepresentation of marginalized communities in pretrial detention or sentencing decisions. For example, if the algorithm considers factors like zip code or neighborhood crime rates, it may penalize individuals from low-income or predominantly minority neighborhoods, regardless of their individual circumstances or risk factors.

Transparency and Accountability: Predictive algorithms used in the criminal justice system are often proprietary or opaque, making it difficult for stakeholders, including defendants and their legal representatives, to understand how decisions are made or challenge biased outcomes. Lack of transparency and accountability can exacerbate social and ethical biases by perpetuating unfair or discriminatory practices without meaningful oversight or recourse.

Feedback Loop: Biased outcomes generated by the predictive algorithm can reinforce existing social inequalities and discriminatory practices within the criminal justice system. For instance, if the algorithm recommends pretrial detention or harsher sentences for individuals from marginalized communities at higher rates, it may contribute to cycles of incarceration and further marginalization.

In this example, social and ethical bias in predictive algorithms used in the criminal justice system raises concerns about fairness, transparency, and accountability in decision-making processes. Addressing these biases requires careful consideration of data collection practices, algorithmic design, and the broader social and ethical implications of using predictive analytics in legal contexts. Efforts to mitigate social and ethical bias should prioritize fairness, equity, and respect for individual rights and dignity within the criminal justice system.

## Case Study: Predictive Policing

### Problem Setup:

Predictive policing models use historical crime data, such as types of crimes committed, locations, times, and sometimes demographic information about known offenders. The objective is to predict future crime hotspots or individuals who might be at a higher risk of committing crimes.

### Biases Identified:

**Data Bias:** Historical crime data is not a perfect representation of all crimes committed, only those that have been reported and recorded. If certain communities are over-policed, the data will show higher crime rates in those areas, not necessarily because more crime occurs there, but because more crime is detected and reported.

**Algorithmic Bias:** The models may inadvertently learn to associate crime with specific neighborhoods or demographic groups, based on the biased historical data. This can lead to a feedback loop where police resources are increasingly allocated to these areas or groups, leading to more arrests and reinforcing the data bias.

**Ethical and Social Bias:** The use of such models can exacerbate social inequalities. Communities that are already disadvantaged may face increased surveillance and policing, leading to higher rates of arrests and convictions. This perpetuates a cycle of disadvantage and can erode trust in law enforcement.

**Consequences**

**Inequitable Policing:** Certain neighborhoods, often those with higher populations of minority groups, may experience disproportionate policing efforts, leading to an unfair distribution of law enforcement resources and negative impacts on community relations. **Self-Perpetuating Biases:** The feedback loop created by relying on biased data can make it difficult to break the cycle of over-policing in certain areas, as the model continues to predict high crime rates based on skewed arrest records.

**Mitigation Strategies:**

**Diverse Data Sources:** Incorporating a wider range of data sources can help mitigate biases in the dataset. This might include community reports, socioeconomic data, and other indicators that provide a more balanced view of risk factors. **Algorithmic Fairness Techniques:** Applying techniques designed to detect and correct biases in algorithms can help ensure that predictions do not disproportionately impact certain groups. **Transparency and Oversight:** Making the models' workings transparent and subject to community oversight can help build trust and ensure that they are used ethically and fairly. **Continuous Evaluation:** Regularly evaluating the models against real-world outcomes and updating them based on current data and ethical considerations can help prevent biases from becoming entrenched.

This case study underscores the importance of carefully considering the sources of data, the design of algorithms, and the broader social impact of mathematical models. It highlights the need for ongoing vigilance, ethical considerations, and community engagement in the development and application of predictive technologies.

**Links**

https://daily.jstor.org/what-happens-when-police-use-ai-to-predict-and-prevent-crime/
https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

**Exercise**

Last week, we asked you to think about a phenomena you would like to model. Now you have been given some formal steps and some ideas on how to model your phenomena, think now about how you may formalise this model?

Think about what specific measureable phenomena you would like to measure Think about what these key factors are and how factors interact with one another? Think about how you might collect data for your model? Think about any biases your model may have and design some mitigation strategies for these

**Conclusion:**

Both Mathematical modeling and Statistical modelling provide powerful and versatile approaches for tackling complex real-world problems. It involves abstraction, equations, assumptions, parameter estimation, validation, simulation, and interpretation. Whilst Mathematical models provide insights, support decision-making, and drive innovation in various fields. Statistical modelling also does this but relies on real world data to create and validate models.