Data Analytics Models

Introduction

Data analytics models are mathematical or computational frameworks that help us understand, predict, and optimize real-world phenomena using data. The three main types of models in analytics are:

1. **Statistical Models**
2. **Machine Learning Models**
3. **Simulation Models**

Each serves a unique purpose, has its own mathematical intuition, and is widely used in real-life scenarios.

---

1. Statistical Models

Definition

A statistical model uses mathematical equations to describe relationships between variables in data, often for explanation or prediction.

How It Works

- Identify **dependent** (output) and **independent** (input) variables.
- Use historical data to estimate model parameters.
- The model quantifies how changes in inputs affect outputs.

Classic Example: Linear Regression

Scenario:
Predicting a student's exam score based on study hours.

Model Equation:

$$Exam\ Score = \beta 0 + \beta 1 \times Study\ Hours + \epsilon$$

- $\beta 0$: Intercept (baseline score)
- $\beta 1$: Change in score per study hour
- ϵ: Random error

Working Demo:

Student	Study Hours	Exam Score
A	2	65
B	4	75
C	6	85

Suppose the fitted model gives:

$$\beta_0 = 55, \beta_1 = 5$$

Prediction:

A student studying 5 hours:

$$PredictedScore = 55 + 5 \times 5 = 80$$

Real-Life Applications:

- **Business:** Forecasting sales based on advertising spend.
- **Healthcare:** Predicting patient wait times.
- **Marketing:** Assessing the impact of campaigns.

---

2. Machine Learning Models

Definition

Machine learning (ML) models are algorithms that automatically learn patterns from data, often used for prediction, classification, or clustering.

How It Works

- Data is split into **training** (to learn) and **testing** (to validate).
- The algorithm “learns” from the training data to generalize to new, unseen data.

Example: Decision Tree (Classification)

Scenario:
Predict if a patient has diabetes based on BMI and age.

Demo Data:

Patient	BMI	Age	Diabetes?
X	22	30	No
Y	28	40	Yes
Z	35	50	Yes
W	20	25	No

How a Decision Tree Works:

- If BMI > 25 → likely "Yes"
- If BMI ≤ 25 and Age > 35 → check further, etc.

Prediction:
A new patient with BMI 30 and age 45 predicts “Yes” for diabetes.

Real-Life Applications:

- **E-commerce:** Product recommendations.
- **Banking:** Fraud detection.
- **Healthcare:** Automated risk predictions.

---

3. Simulation Models

Definition

Simulation models use mathematics or computer programs to mimic complex real-life processes and assess possible outcomes under uncertainty.

How It Works

- Define possible scenarios and inputs (including randomness).
- Run multiple “simulations” to see the range and likelihood of results.

Example: Hospital Queue Simulation

Scenario:
A hospital wants to predict patient waiting times based on arrival and service rates.

- **Inputs:** 10 patients/hour arrive, service rate is 12 patients/hour.
- Run the simulation 1,000 times to estimate average queue length and waiting time.

Outcome:
Hospital management learns that at busy hours, wait times jump—so they plan for extra staff during these rushes.

Real-Life Applications:

- **Airlines:** Simulate boarding processes.
- **Logistics:** Optimize delivery routes.
- **Call Centers:** Manage peak hours.

Summary Table

Model Type	What It Does	Classic Example	Common Application
Statistical	Mathematical relationship, explain	Linear Regression	Forecasting, KPIs
Machine Learning	Learns from data to predict/classify	Decision Tree	Recommendations, risk
Simulation	Mimics reality for planning/estimates	Queue Simulation	Optimization, planning

---

Visuals (Conceptual)

- **Linear Regression:**
  Scatter plot of Study Hours vs. Exam Score with a best-fit line.
- **Decision Tree:**
  Tree diagram showing how yes/no questions predict outcomes.
- **Simulation:**
  Histogram showing simulated waiting times in hospital queues.

Research & Theory References

- **Regression:** Francis Galton (1880s), Ronald Fisher (1930s)
- **Decision Trees:** J. Ross Quinlan, “ID3” (1986)
- **Simulation:** Queuing theory, A.K. Erlang (early 1900s)

Conclusion

- **Statistical models** describe and forecast based on known relationships.
- **Machine learning models** adaptively learn complex patterns for advanced predictions.
- **Simulation models** help visualize “what-if” scenarios in uncertain, dynamic environments.

Each model type underpins real-world analytics solutions: from healthcare to e-commerce, these frameworks turn data into smart, actionable decision-making tools.

References:

- 315326-DATA-ANALYTICS.pdf
- Data-Analytics-Lifecycle-Data-Attributes.pptx