**Key Differences Between Statistics and Machine Learning**

| **Aspect** | **Statistics** | **Machine Learning** |
| --- | --- | --- |
| **Primary Goal** | *Inference and explanation* | *Prediction and generalization* |
| **Approach** | Model-driven (assumes data follows a known distribution) | Data-driven (learns patterns from data without assumptions) |
| **Model Interpretability** | High (e.g., coefficients in regression are explainable) | Often low (e.g., deep learning models are black-box) |
| **Assumptions** | Strong (e.g., normality, independence, linearity) | Minimal or none (flexible to data structure) |
| **Data Requirements** | Works well with small, clean datasets | Excels with large, complex datasets |
| **Validation Techniques** | Hypothesis testing, confidence intervals | Cross-validation, train/test splits |
| **Examples** | Linear regression, ANOVA, logistic regression | Decision trees, SVMs, neural networks |
| **Uncertainty Quantification** | Central (e.g., p-values, standard errors) | Often secondary or implicit (e.g., ensemble variance) |
| **Use Cases** | Scientific research, surveys, clinical trials | Image recognition, recommendation systems, fraud detection |

**How They Complement Each Other**

* **Feature Selection**: Statistical tests help identify relevant variables for ML models.
* **Model Diagnostics**: Statistical tools assess residuals, bias, and variance.
* **Hybrid Models**: Techniques like *Bayesian machine learning* and *regularized regression* blend both paradigms.
* **Interpretability**: Statistics adds transparency to ML predictions, especially in regulated domains like healthcare and finance.