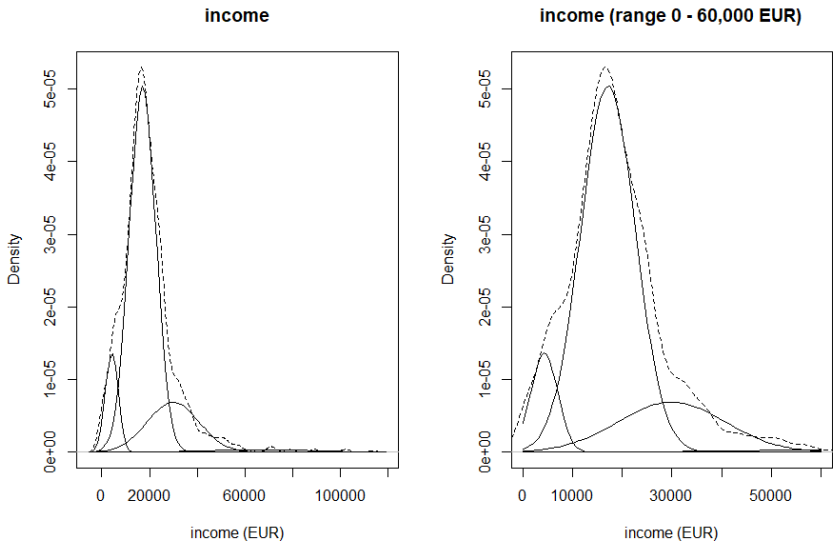
What is Statistics and what is it used for?

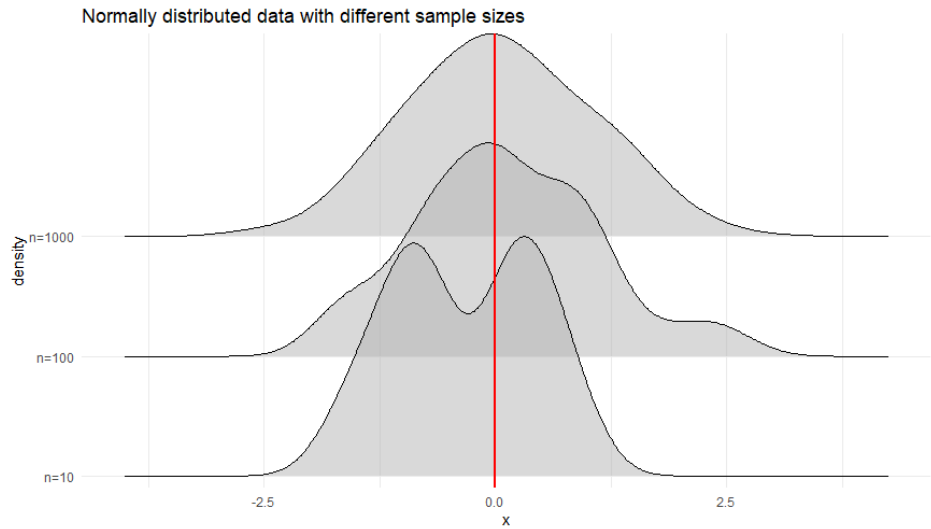
Statistics could be best understood as a **branch of mathematics** based on measure theory and probability theory which primarily studies the **properties of estimators** and the **distributions of errors**.

Statistics helps us, through the use of **assumptions and models**, to transform **raw information** from data into **knowledge** to guide decision-making.

Statistics is the natural language of **research** and the grammar of **science**.

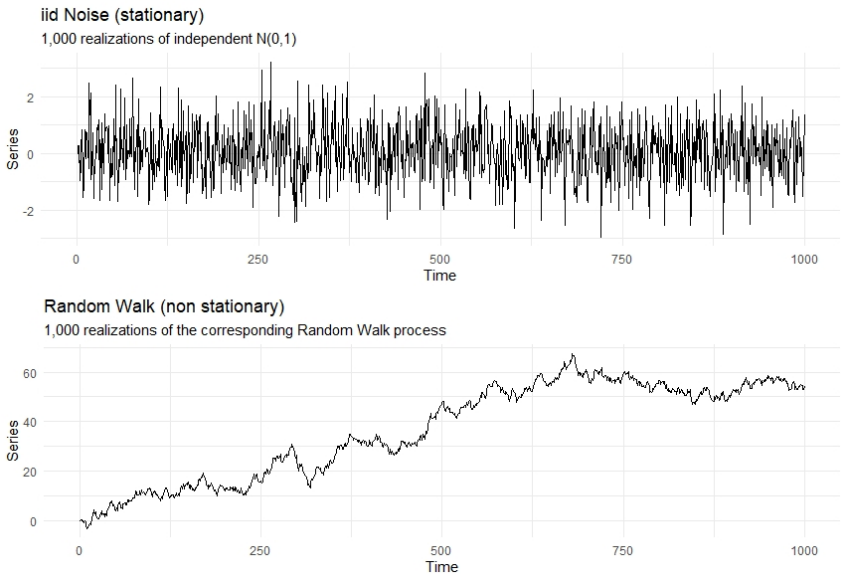
Inferential Statistics

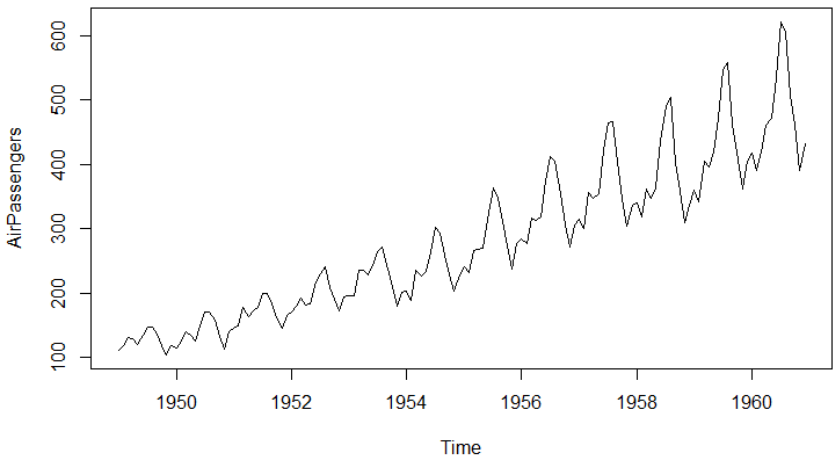




While **Descriptive Statistics** summarize the characteristics of a dataset, **Inferential Statistics** allows us to **generalize results** computed on a sample to a broader population.

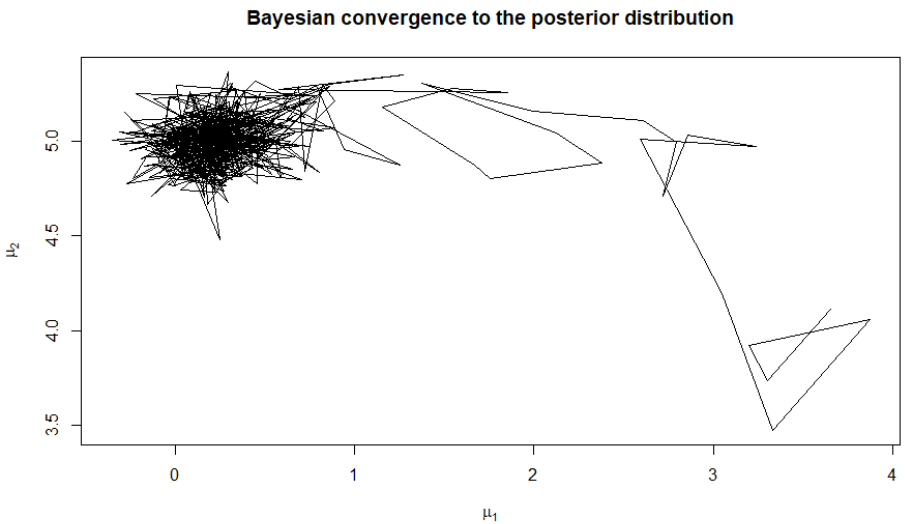
Time Series Analysis

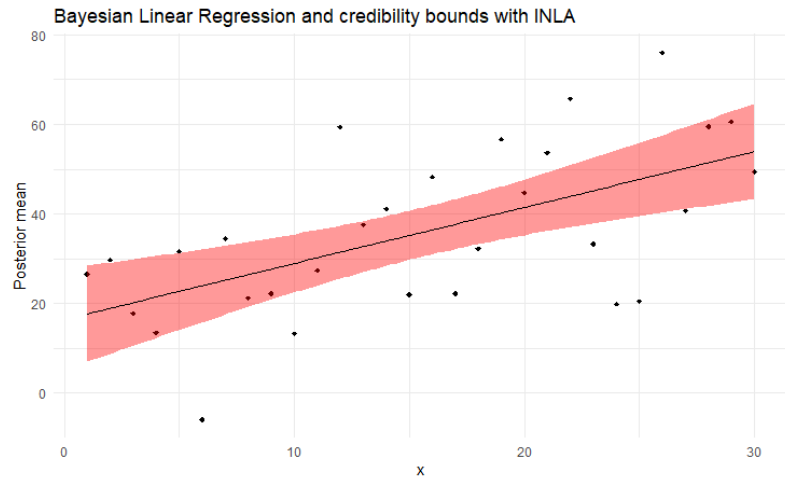




Time Series Analysis is a specific way of analyzing a sequence of data points collected over an interval of time. We use **Time Series** models when we have **temporal data**.

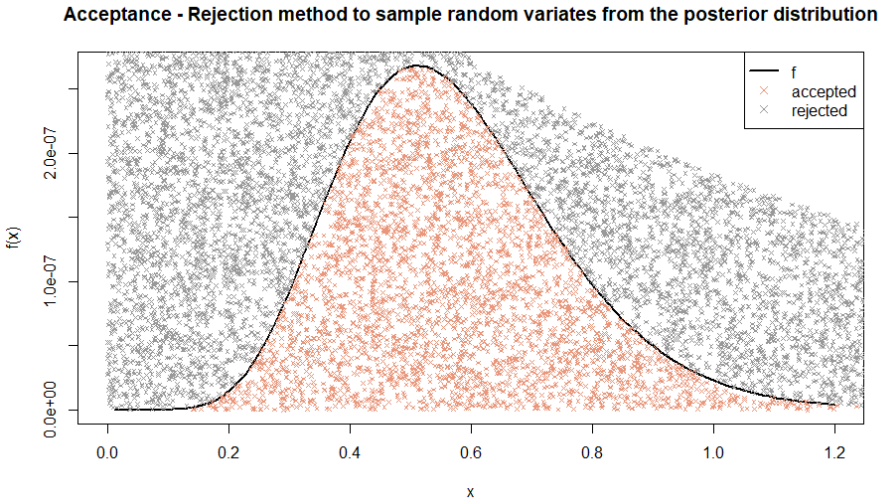
Bayesian Statistics

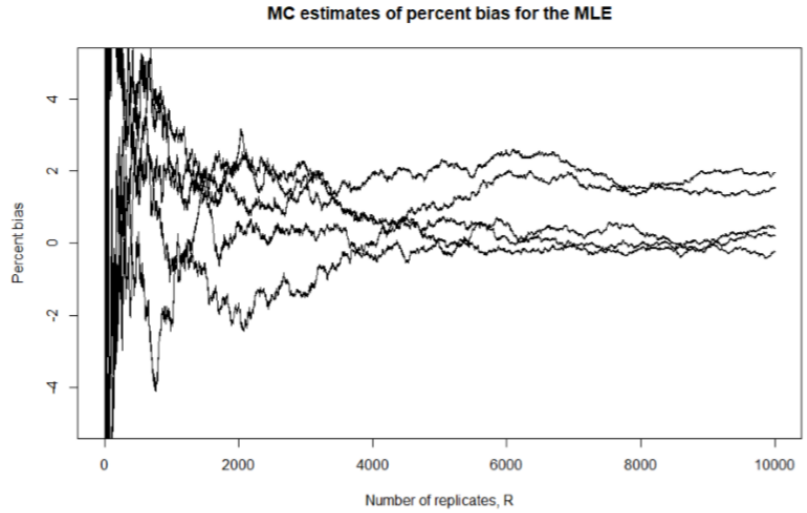




Unlike in the traditional **frequentist approach to Statistics**, the **bayesian approach to Statistics** views parameters as random variables rather than fixed, unknown quantities. We can incorporate subjective expert opinions in the **Bayesian approach to Statistics** through the use of prior distributions. The **Bayesian approach to Statistics** requires computationally intensive algorithms such as **Monte Carlo Markov Chains (MCMC) sampling**.

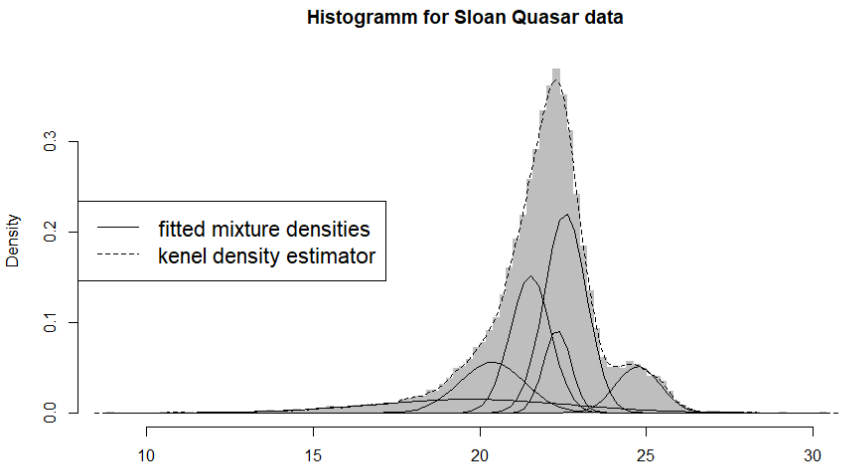
Computational Statistics

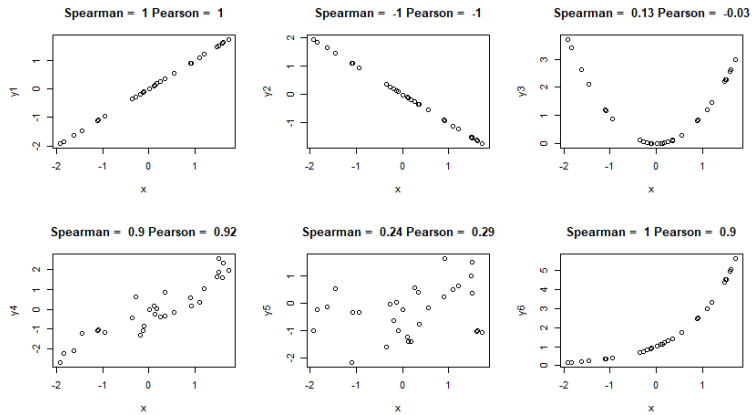




Computational Statistics, or statistical computing, is the bond between Statistics and Computer Science. It means statistical methods that are enabled by using **computational methods**. Some famous families of methods in Computational Statistics include **Monte Carlo simulations** and the **Bootstrap ressampling methods**.

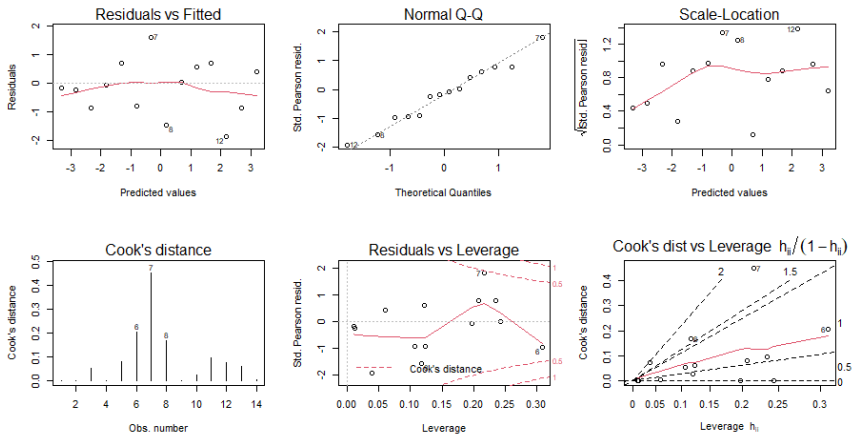
Nonparametric Statistics

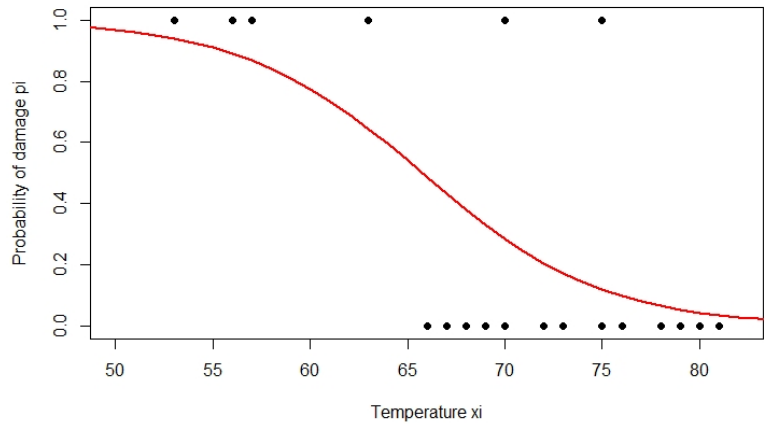




There is no unique definition of the field encompassed by the term **Nonparametric Statistics**. It is better seen as a set of techniques applied to statistical observations whereby a minimal modelization is assumed. These techniques are, for this reason, often also called **distribution-free**. We use **Nonparametric Statistics** when we don't want to assume a parametric model (examples: Normal model, Gamma model). These methods works well on small samples.

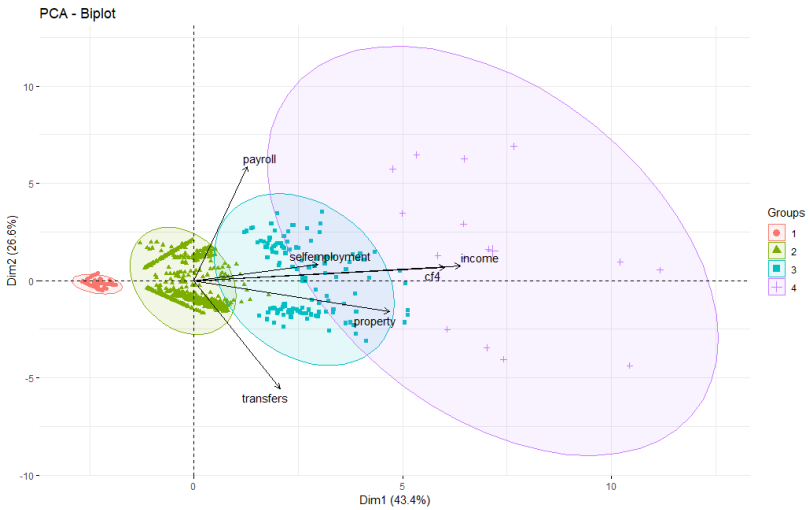
Generalized Linear Models

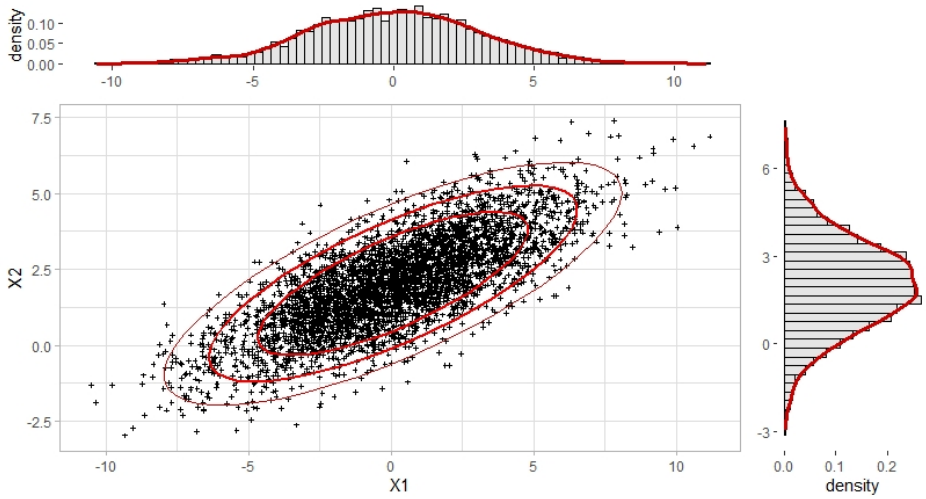




In statistics, a **Generalized Linear Model** is a flexible generalization of ordinary linear regression. The **Generalized Linear Model** generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. Some famous **Generalized Linear Model** include the **Logistic regression** and the **Poisson regression**.

Multivariate Statistics





Multivariate Statistics is a subdivision of Statistics encompassing the simultaneous observation and analysis of **more than one variable**. Some famous methods include the **Principal Component Analysis** (PCA) and the Multiple Linear Regression.

There is more

Of course, there exist many more areas of **Statistics**, for example **Spatial Statistics**, **Spatio-temporal Statistics**, **Survey Sampling**, **Small Area Estimation**...

Statistics can be applied to all sciences, for example Environmental science, Medical science, Economics, Finance, Biology, Physics or Agricultural science.

Statistics is the foundation of **Data Science** and **Machine Learning** for example.

Bayesian Questionnaire: Question 1

What is the field of Statistics which models temporal data using a Bayesian approach ?

A: Frequentist Statistics
B: Bayesian Nonparametric Statistics
C: Bayesian Time Series Analysis
D: Bayesian Spatial Statistics