

Mathematics | High School—Statistics and Probability★

Decisions or predictions are often based on data—numbers in context. These decisions or predictions would be easy if the data always sent a clear message, but the message is often obscured by variability. Statistics provides tools for describing variability in data and for making informed decisions that take it into account.

Data are gathered, displayed, summarized, examined, and interpreted to discover patterns and deviations from patterns. Quantitative data can be described in terms of key characteristics: measures of shape, center, and spread. The shape of a data distribution might be described as symmetric, skewed, flat, or bell shaped, and it might be summarized by a statistic measuring center (such as mean or median) and a statistic measuring spread (such as standard deviation or interquartile range). Different distributions can be compared numerically using these statistics or compared visually using plots. Knowledge of center and spread are not enough to describe a distribution. Which statistics to compare, which plots to use, and what the results of a comparison might mean, depend on the question to be investigated and the real-life actions to be taken.

Randomization has two important uses in drawing statistical conclusions. First, collecting data from a random sample of a population makes it possible to draw valid conclusions about the whole population, taking variability into account. Second, randomly assigning individuals to different treatments allows a fair comparison of the effectiveness of those treatments. A statistically significant outcome is one that is unlikely to be due to chance alone, and this can be evaluated only under the condition of randomness. The conditions under which data are collected are important in drawing conclusions from the data; in critically reviewing uses of statistics in public media and other reports, it is important to consider the study design, how the data were gathered, and the analyses employed as well as the data summaries and the conclusions drawn.

Random processes can be described mathematically by using a probability model: a list or description of the possible outcomes (the sample space), each of which is assigned a probability. In situations such as flipping a coin, rolling a number cube, or drawing a card, it might be reasonable to assume various outcomes are equally likely. In a probability model, sample points represent outcomes and combine to make up events; probabilities of events can be computed by applying the Addition and Multiplication Rules. Interpreting these probabilities relies on an understanding of independence and conditional probability, which can be approached through the analysis of two-way tables.

Technology plays an important role in statistics and probability by making it possible to generate plots, regression functions, and correlation coefficients, and to simulate many possible outcomes in a short amount of time.

Connections to Functions and Modeling. Functions may be used to describe data; if the data suggest a linear relationship, the relationship can be modeled with a regression line, and its strength and direction can be expressed through a correlation coefficient.