

6–8 Statistics and Probability

Overview

In Grade 6, students build on the knowledge and experiences in data analysis developed in earlier grades (see K–3 Categorical Data Progression and Grades 2–5 Measurement Progression). They develop a deeper understanding of variability and more precise descriptions of data distributions, using numerical measures of center and spread, and terms such as cluster, peak, gap, symmetry, skew, and outlier. They begin to use histograms and box plots to represent and analyze data distributions. As in earlier grades, students view statistical reasoning as a four-step investigative process:

- Formulate questions that can be answered with data
- Design and use a plan to collect relevant data
- Analyze the data with appropriate methods
- Interpret results and draw valid conclusions from the data that relate to the questions posed.

Such investigations involve making sense of practical problems by turning them into statistical investigations (MP1); moving from context to abstraction and back to context (MP2); repeating the process of statistical reasoning in a variety of contexts (MP8).

In Grade 7, students move from concentrating on analysis of data to production of data, understanding that good answers to statistical questions depend upon a good plan for collecting data relevant to the questions of interest. Because statistically sound data production is based on random sampling, a probabilistic concept, students must develop some knowledge of probability before launching into sampling. Their introduction to probability is based on seeing probabilities of chance events as long-run relative frequencies of their occurrence, and many opportunities to develop the connection between theoretical probability models and empirical probability approximations. This connection forms the basis of statistical inference.

With random sampling as the key to collecting good data, students begin to differentiate between the variability in a sample and

the variability inherent in a statistic computed from a sample when samples are repeatedly selected from the same population. This understanding of variability allows them to make rational decisions, say, about how different a proportion of "successes" in a sample is likely to be from the proportion of "successes" in the population or whether medians of samples from two populations provide convincing evidence that the medians of the two populations also differ.

Until Grade 8, almost all of students' statistical topics and investigations have dealt with univariate data, e.g., collections of counts or measurements of one characteristic. Eighth graders apply their experience with the coordinate plane and linear functions in the study of association between two variables related to a question of interest. As in the univariate case, analysis of bivariate measurement data graphed on a scatterplot proceeds by describing shape, center, and spread. But now "shape" refers to a cloud of points on a plane, "center" refers to a line drawn through the cloud that captures the essence of its shape, and "spread" refers to how far the data points stray from this central line. Students extend their understanding of "cluster" and "outlier" from univariate data to bivariate data. They summarize bivariate categorical data using two-way tables of counts and/or proportions, and examine these for patterns of association.