

Develop understanding of statistical variability.

1. Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, “How old am I?” is not a statistical question, but “How old are the students in my school?” is a statistical question because one anticipates variability in students’ ages.*
2. Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
3. Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

6.SP 3
variance (Ben)

Summarize and describe distributions.

4. Display numerical data in plots on a number line, including dot plots, histograms, and box plots.
5. Summarize numerical data sets in relation to their context, such as by:
 - a. Reporting the number of observations.
 - b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.
 - c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
 - d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

6.SP 5d
card shuffling (Laurent)

Use random sampling to draw inferences about a population.

1. Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
2. Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions. *For example, estimate the mean word length in a book by randomly sampling words from the book; predict the winner of a school election based on randomly sampled survey data. Gauge how far off the estimate or prediction might be.*

Draw informal comparative inferences about two populations.

3. Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable.*
4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book.*

7.SP 4
variance (Ben)

Investigate chance processes and develop, use, and evaluate probability models.

5. Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.

7.SP 5
card shuffling (Laurent)

7.SP 5
notation & relative frequency (Tasia)

6. Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability. *For example, when rolling a number cube 600 times, predict that a 3 or 6 would be rolled roughly 200 times, but probably not exactly 200 times.*
7. Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy.
- Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events. *For example, if a student is selected at random from a class, find the probability that Jane will be selected and the probability that a girl will be selected.*
 - Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process. *For example, find the approximate probability that a spinning penny will land heads up or that a tossed paper cup will land open-end down. Do the outcomes for the spinning penny appear to be equally likely based on the observed frequencies?*
8. Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation.
- Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.
 - Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language (e.g., “rolling double sixes”), identify the outcomes in the sample space which compose the event.
 - Design and use a simulation to generate frequencies for compound events. *For example, use random digits as a simulation tool to approximate the answer to the question: If 40% of donors have type A blood, what is the probability that it will take at least 4 donors to find one with type A blood?*

7.SP 6
notation & relative frequency (Tasia)

7.SP 7
rock-paper-scissors game (Tasia)
card shuffling (Laurent)

7.SP 8
rock-paper-scissors game (Tasia)

Investigate patterns of association in bivariate data.

1. Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.
2. Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
3. Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept. *For example, in a linear model for a biology experiment, interpret a slope of 1.5 cm/hr as meaning that an additional hour of sunlight each day is associated with an additional 1.5 cm in mature plant height.*
4. Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables. *For example, collect data from students in your class on whether or not they have a curfew on school nights and whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?*

8.SP 1
variance (Ben)

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.

Statistics and Probability Overview

Interpreting Categorical and Quantitative Data

- Summarize, represent, and interpret data on a single count or measurement variable
- Summarize, represent, and interpret data on two categorical and quantitative variables
- Interpret linear models

Making Inferences and Justifying Conclusions

- Understand and evaluate random processes underlying statistical experiments
- Make inferences and justify conclusions from sample surveys, experiments and observational studies

Conditional Probability and the Rules of Probability

- Understand independence and conditional probability and use them to interpret data
- Use the rules of probability to compute probabilities of compound events in a uniform probability model

Using Probability to Make Decisions

- Calculate expected values and use them to solve problems
- Use probability to evaluate outcomes of decisions

Mathematical Practices

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.
6. Attend to precision.
7. Look for and make use of structure.
8. Look for and express regularity in repeated reasoning.

Interpreting Categorical and Quantitative Data**S-ID****Summarize, represent, and interpret data on a single count or measurement variable**

1. Represent data with plots on the real number line (dot plots, histograms, and box plots).
2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
3. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
4. Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

S-ID #1-4
variance (Ben)**Summarize, represent, and interpret data on two categorical and quantitative variables**

5. Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.
6. Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
 - a. Fit a function to the data; use functions fitted to data to solve problems in the context of the data. *Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.*
 - b. Informally assess the fit of a function by plotting and analyzing residuals.
 - c. Fit a linear function for a scatter plot that suggests a linear association.

Interpret linear models

7. Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.
8. Compute (using technology) and interpret the correlation coefficient of a linear fit.
9. Distinguish between correlation and causation.

Making Inferences and Justifying Conclusions**S-IC****S-IC**
coin toss experiment (Lionel)**Understand and evaluate random processes underlying statistical experiments**

1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
2. Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. *For example, a model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model?*

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

S-IC 3
pitfalls in statistics (Russ)**S-IC 3**
Simpson's Paradox
(Mark, Tianyi, Mathav)

- Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
- Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.
- Evaluate reports based on data.

S-IC 4-6Simpson's Paradox
(Mark, Tianyi, Mathav)**Conditional Probability and the Rules of Probability**

S-CP

Understand independence and conditional probability and use them to interpret data

- Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events (“or,” “and,” “not”).
- Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.
- Understand the conditional probability of A given B as $P(A \text{ and } B)/P(B)$, and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A , and the conditional probability of B given A is the same as the probability of B .
- Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. *For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.*
- Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. *For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.*

S-CP 1

notation & relative frequency (Tasia)

S-CP 4-5Simpson's Paradox
(Mark, Tianyi, Mathav)**Use the rules of probability to compute probabilities of compound events in a uniform probability model**

- Find the conditional probability of A given B as the fraction of B 's outcomes that also belong to A , and interpret the answer in terms of the model.
- Apply the Addition Rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, and interpret the answer in terms of the model.
- (+) Apply the general Multiplication Rule in a uniform probability model, $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$, and interpret the answer in terms of the model.
- (+) Use permutations and combinations to compute probabilities of compound events and solve problems.

Using Probability to Make Decisions

S-MD

Calculate expected values and use them to solve problems

- (+) Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.
- (+) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.

S-MDcoupon collector problem (Tasia)
hat matching problem (Tasia)

3. (+) Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value. *For example, find the theoretical probability distribution for the number of correct answers obtained by guessing on all five questions of a multiple-choice test where each question has four choices, and find the expected grade under various grading schemes.*
4. (+) Develop a probability distribution for a random variable defined for a sample space in which probabilities are assigned empirically; find the expected value. *For example, find a current data distribution on the number of TV sets per household in the United States, and calculate the expected number of sets per household. How many TV sets would you expect to find in 100 randomly selected households?*

Use probability to evaluate outcomes of decisions

5. (+) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values.
 - a. Find the expected payoff for a game of chance. *For example, find the expected winnings from a state lottery ticket or a game at a fast-food restaurant.*
 - b. Evaluate and compare strategies on the basis of expected values. *For example, compare a high-deductible versus a low-deductible automobile insurance policy using various, but reasonable, chances of having a minor or a major accident.*
6. (+) Use probabilities to make fair decisions (e.g., drawing by lots, using a random number generator).
7. (+) Analyze decisions and strategies using probability concepts (e.g., product testing, medical testing, pulling a hockey goalie at the end of a game).