Building Concepts: Statistics and Probability Overview

Building Concepts: Statistics and Probability consists of a sequential, coordinated series of interactive, dynamic TI-Nspire™ activities designed to support a developmental trajectory of concepts related to statistics and probability. The approach, described in the Common Core State Standards (CCSS), emphasizes variability as a central concept in statistics. In light of this variability, students formulate questions that can be answered with data, analyze and interpret distributions of data, collect data through random samples, interpret the results and use probability to draw conclusions from the data that relate to the questions posed.

The activities are created for those who are teaching or learning statistics and probability or reviewing the concepts, parents, or anyone interested in learning or helping others learn about statistics and probability. The activities are designed to develop students' expertise in the processes and proficiencies that serve as the foundation for the Standards for Mathematical Practice described in the Common Core State Standards.

Compatible TI Technologies

The TI-Nspire activities (TNS files) are compatible with the following technologies:

TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software

The activities can be adapted for different classroom environments. One teacher might use a whiteboard to demonstrate the activity to a large group of students, while another might have students working individually or in small groups using a TI-Nspire handheld. The delivery method used depends on the technology available in the classroom. The estimated activity time for students working at grade level is 50–90 minutes, depending on the delivery method of the activity.

Each TI-Nspire activity (TNS) is accompanied by Teacher Notes and a Student Activity sheet to help facilitate the use of the TI-Nspire activity in the classroom.

Teacher Notes

The Teacher Notes are intended to provide a starting point for how these activities can be used with students. It is important to note that the TI-Nspire activities themselves are typically robust enough to be adapted for learning goals not mentioned in the Notes.

Teacher Notes are designed to serve as a guide for implementation in classrooms and include:

- descriptions of the mathematics that underlies each lesson
- a description of the activity and how to use it
- mathematical goals for student learning
- discussion questions for teacher-led classroom exploration or individual student investigation
- assessment questions that vary in difficulty. These questions might be found on the National Assessment of Educational Progress mathematic assessments or a high stakes state test.

Instructors may choose to use all of the suggested questions or select a few from those presented.

Student Activity

The Student Activity sheet is an optional resource that may be used as a tool to keep students engaged and on track throughout the lesson. This resource is particularly valuable in a large-group instruction scenario.

Although teachers can successfully work through the lessons without using the Student Activity sheet, the resource can prove to be valuable. The statistics and probability concepts align with the math standards for grades 6–8 and the activity sheet has been created to match this range. The Student Activity is provided as an editable word document and therefore may be modified to use with middle-school students or higher, increasing its value.

Lesson Series

The series consists of the following 24 lessons. Each lesson should take 50–90 minutes to complete with students, though teachers may choose to extend lessons, as needed.

1. *Introduction to Data*

   The lesson introduces students to the notion of a statistical question, a question that anticipates variability in the response. In the context of maximum speeds and life spans of a selected set of animals, students investigate dot plots of distributions of sets of data and become familiar with different shapes: skew, symmetric, mound shaped, and uniform. Questions are designed to help students think about the distribution of speeds or life spans as a story about the data as a whole rather than thinking of data as individual elements (a common misconception). The lesson introduces range as a rough measure of spread and informally introduces the concept of outliers.

2. *Median and Interquartile Range*

   Measures of center and spread are both necessary to describe a distribution of data in a sense-making way, and thus, the lesson develops both concepts. The focus of this lesson is an approach to measures of center and spread based on ordering and relative position, which leads to the median as a measure of center and the range and interquartile range as measures of spread. Students find the median of a data set, and, by taking the median of each half of the ordered data, find the quartiles, and the interquartile range.

3. *Box Plots*

   Building on Lesson 2 in which students developed the concept of median and interquartile range, students develop a *box plot* as a way to display the range, center and spread of a data distribution. Students compare what can be observed from the box plot and from the dot plot for a given distribution of data and explore how different distributions of data can have box plots that are very similar.

4. *Mean as Fair Share*

   In this lesson, the mean as fair share is developed in two ways: 1) those with the most give something to those with the least until everyone has the same amount or 2) total all the contributions and divide the total equally among the contributors, which is the typical approach to finding a mean. Part of this lesson is designed to help students make the transition from bar graphs to representing data in a dot plot on a number line.

5. *Mean as Balance Point*

   Students calculate *deviations* from the mean, how far each value is above or below the mean. The average of the sum of the absolute mean deviations, the *mean absolute deviation*, or MAD, characterizes the typical distance of the data values from the mean. Exploring variation with the MAD sets the stage for introducing standard deviation in high school. Because the deviations "balance" around the mean, the activity highlights the mean as the balance point of a distribution, where the sum of the absolute deviations below the mean is equal to the sum of the absolute deviations above the mean.

6. *Tables and Measures of Center and Spread*

   The focus of this lesson is to develop an understanding of how to compute measures of center and spread when the data are given in a table, particularly when the data are tallied with a frequency count. The lesson is an informal introduction to a spreadsheet, where students are given choices for commands that will generate columns useful for finding summary statistics such as the mean or median. The lesson shows how the frequency is related to the actual counts of the data values, visually addressing a typical error students make of ignoring the frequency.

7. *Introduction to Histograms*

   In this lesson, a dot plot of a distribution of data "morphs" into a histogram. Students analyze the effect of bin widths on describing a distribution of data and consider measures of center and spread for histograms representing distributions of different shapes. They confront common misconceptions such as confusing histograms and bar graphs and judging variability by focusing on the varying heights of the bars. They add their own data to a given distribution and explore a histogram of the combined sets of data.

8. *Outliers*

   This lesson develops the concept of an outlier as any data point beyond 1.5 Interquartile ranges beyond the upper quartile and below the lower quartile. Students consider why outliers are important to think about and identify them visually in dot plots.

9. *Analyzing Distributions*

   Students return to the distributions of animal speeds and life span, this time using the summary measures, mean/MAD and median/IQR, to interpret the shape of the data. They investigate whether one set of summary measures conveys a different story than the other about the distribution, noting that in highly skewed distributions the mean/MAD can be misleading.

10*. Transforming Data*

Students investigate what happens to the summary measures for the distribution of a set of data when a constant is added to each data element and when each data element is multiplied by a constant.

11. *Comparing Distributions*

Students compare the life expectancies, income and literacy of the countries in different areas of the world, using a variety of graphs and different summary measures. The contexts are also used to examine bar graphs and how they differ from histograms, replicating typical displays of information found in many media sources.

12. *What Is Probability?*

Students play a game against Tinman to see who wins the most often and then analyze the sample space to figure out why. This leads to the definition of probability based on the theoretical model and shows how this plays out in an experimental setting. The second part of the lesson engages students in exploring the probability of an event using long run relative frequency.

13. *Probability, Diagrams, and Tables*

Students investigate the probability of compound events using tree diagrams and tables, in both two and three stage events. They organize the outcomes from contextual settings to make sense of the probabilities and relate the outcomes to an area model for probability.

14. *Unequally Likely Outcomes*

In this lesson, students contrast two simulated probability distributions of the sum of the faces when two dice are tossed, one assuming the outcomes are equally likely and the other recording the actual sum of the faces when tossing two dice. The first case produces a uniform or rectangular distribution, and the second produces a mound shaped symmetric distribution. Students confront the misconception that double sixes occur twice by selecting faces on two dice and relating them to a table of ordered pairs and sums.

15. *Probability and Simulation*

This lesson gives students a set of tools—coins, dice, and a spinner —to model probability situations. They simulate situations where the number of observations is known and the task is to identify the probability of successful outcomes. They also simulate situations where the number of observations depends on the sequence (i.e., “go until” five heads). The tools are used to solve a variety of contextual problems.

16*. Law of Large Numbers*

The lesson considers how samples can reflect the population, building on the long-run relative frequency lesson from *What Is Probability*? Observing how the shape of the distribution develops as the number of observations increases introduces the law of large numbers, i.e., the distributions of random processes stabilize as the sample size gets larger. Students distinguish between a frequency distribution and a relative frequency distribution noting that, as the sample size increases, the relative frequency for each outcome approaches the theoretical probability of that outcome.

17. *Why Random Sampling?*

Students select a sample that seems to be typical of the population by "judgment" and find the mean of their sample. They then compare the sampling distribution of their sample means to a simulated sampling distribution of the means of samples selected randomly.

18. *Choosing Random Samples*

Students explore four different methods for generating random samples: drawing names from a hat; assigning random numbers and using a random number generator; assigning random numbers to a block and then another set of random numbers to cells within the block; generating random pairs of points. The methods address a common misconception that random samples do not have streaks or clusters.

19. *Sample Proportions*

Students consider sampling variability within a sample and from sample to sample. Using the fact that sample proportions from the same populations vary, they simulate sampling distributions for known populations and develop a sense of what a simulated sampling distribution for a given sample size will look like with respect to shape, center and spread. They use the overlap in these distributions for given populations to begin reasoning about how to make an inference from a sample to a population. Students also examine how sample size can make a difference in the spread of the sampling distribution.

20. *Sample Means*

The focus of this lesson is to understand that medians and means computed from samples will also vary from sample to sample and that making informed decisions based on such sample statistics requires some knowledge of the amount of variation to expect. Students simulate random sampling from a known population to estimate variability in population means to get a sense of the variability around an estimate. They revisit the effect of sample size on the spread of a sampling distribution.

21. *Comparing Populations Using Sample Statistics*

Many of the practical problems dealing with measures of center involve comparing two or more groups, Students use random samples to compare differences between the means of two populations and use simulation to see if such differences would be likely to occur by chance.

22. *Scatter Plots*

Students interpret scatter plots of statistics associated with the all-time “greatest” professional basketball players, looking for patterns, gaps, clusters and outliers. They explore whether position affects certain statistics (steals, three-point shots) and revisit what percent means in comparative situations.

23. *Modeling Linear Relationships*

Students use a variety of contextual situations to model trends between two variables with linear equations, beginning with situations of the form $y = mx$. They investigate the residuals to think about the error in using a model to predict an outcome for a given input, observe how changes in the model is reflected in the sum of the absolute errors and use patterns in residuals to identify those that might be better modeled by a function that is not linear. They also consider the effect of outliers.

24. *Two-Way Tables and Categorical Data*

In this lesson, students consider relationships between paired *categorical* variables. Two-way tables provide a tool for investigating whether or not there is an association between being a member of a category of one type with being a member of a category of another type. Two-way tables are also used to introduce students to the idea and language of conditional probabilities.

Grade Level Alignment

It is recommended that the lessons be used in sequential order at any grade level. Each lesson builds upon a concept that has been presented in a previous lesson.

The following table shows how the lessons align with typical grade-level expectations.

	Grade 6		Grade 7		Grade 8	
	CCSS	TEKS	CCSS	TEKS	CCSS	TEKS
Lesson 1 *Introduction to Data*	6.SP.A.1 6.SP.A.2 6.SP.B.5b	6.12.B		7.12.A		8.12.A
Lesson 2 *Median and Interquartile Range*	6.SP.A.2 6.SP.B.5c	6.12.B		7.12.A		8.12.A
Lesson 3 *Box plots*	6.SP.B.4 6.SP.B.5a	6.12.B		7.12.A		
Lesson 4 *Mean as Fair Share*	6.SP.A.3 6.SP.B.5a 6.SP.B.5b	6.12.B		7.12.A		
Lesson 5 *Mean as Balance Point*	6.SP.B.5c	6.12.B		7.12.A		
Lesson 6 *Tables and Measures of Center and Spread*	6.SP.4 6.SP.5.b, c	6.12.A 6.12.B 6.12.C 6.13.A				
Lesson 7 *Introduction to Histograms*	6.SP.4, 6.SP.5. a and b	6.12.A 6.12.B 6.12.C 6.13.A				
Lesson 8 *Outliers*	6.SP.4 6.SP.5 a-d	6.12.A 6.12.B 6.12.C				
Lesson 9 *Analyzing Distributions*	6.SP.2 6.SP.3 6.SP.4 6.SP.5 a-d	6.12.A 6.12.B 6.12.C 6.13.A				

	Grade 6		Grade 7		Grade 8	
	CCSS	TEKS	CCSS	TEKS	CCSS	TEKS
Lesson 10 *Transforming Data*	6.SP.5 b, c	6.12.A 6.12.B 6.12.C 6.13.A				
Lesson 11 *Comparing Distributions*	6.SP.1 6.SP.2 6.SP.3 6.SP.4 6.SP.5 a-d	6.12.A 6.12.B 6.12.C 6.13.A				
Lesson 12 *What is Probability?*			7.SP.5 7.SP.6 7.SP.7	7.6.A 7.6.B 7.6.C 7.6.D 7.6.I		
Lesson 13 *Probability, Diagrams, and Tables*			7.SP.8 a and b	7.6.A 7.6.B 7.6.C 7.6.D 7.6.I		
Lesson 14 *Unequally Likely Outcomes*			7.SP.6 7.SP.7 a and b	7.6.A 7.6.B 7.6.C 7.6.D 7.6.I		
Lesson 15 *Probability and Simulations*			7.SP.8 c	7.6.A 7.6.B 7.6.C 7.6.D 7.6.I		
Lesson 16 *Law of Large Numbers*		6.13(B)	7.SP.5 7.SP.6 7.SP.7b	7.6.A 7.6.B 7.6.C 7.6.D 7.6.I		
Lesson 17 *Why Random Sampling?*			7.SP.1 7.SP.2	7.12.B		8.11.C

	Grade 6		Grade 7		Grade 8	
	CCSS	TEKS	CCSS	TEKS	CCSS	TEKS
Lesson 18 *Choosing Random Samples*			7.SP.1 7.SP.2	7.6.F 7.12.B		8.11.C
Lesson 19 *Sample Proportions*			7.SP.1 7.SP.2	7.12.B		8.11.C
Lesson 20 *Sample Means*			7.SP.1 7.SP.2	7.6.F 7.12.B		8.11.C
Lesson 21 *Comparing Populations Using Sample Statistics*			7.SP.3 7.SP.4	7.6.F 7.12.B		8.11.C
Lesson 22 *Scatter Plots*	6.EE.9				8.SP.1	8.11.A
Lesson 23 *Modeling Linear Relationships*					8.SP.1 8.SP.2 8.SP.3	8.11.A 8.5.C 8.5.
Lesson 24 *Two-WayTables and Categorical Data*			7.SP.5		8.SP.4	8.11.A

References

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics.* Washington, DC: Authors.

The Common Core Standards Writing Team. (2013, September 13). Progressions for Common Core State Standards in Mathematics 6-8 Statistics and Probability Retrieved from http://commoncoretools.me/wp-content/uploads/2011/12/ccss_progression_sp_68_2011_12_26_bis.pdf

Texas Essential Knowledge and Skills for Mathematics Subchapter A. Elementary, Chapter 111. (2012). Texas Education Agency

