

Statistics and Probability – UNIT 1
Interpreting Categorical and Quantitative Data

Introduction: Instructional time will focus on creating and interpreting visual displays of data. Students will choose appropriate visual displays, measures of center/spread for a specific data set. They will compare data sets using summary statistics.

CLUSTER	COMMON CORE STATE STANDARDS
Summarize, represent, and interpret data on a single count or measurement variable.	<p><u>HSS.ID.A.1*</u></p> <p>Represent data with plots on the real number line (dot plots, histograms, and box plots).</p> <p><u>HSS.ID.A.2*</u></p> <p>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.</p> <p><u>HSS.ID.A.3*</u></p> <p>Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).</p> <p><u>HSS.ID.A.4</u></p> <p>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.</p>
Summarize, represent, and interpret data on two categorical and quantitative variables.	<p><u>HSS.ID.B.5</u></p> <p>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.</p> <p><u>HSS.ID.B.6</u></p> <p>Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.</p> <p><u>HSS.ID.B.6.A</u></p> <p>Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or</p>

CLUSTER	COMMON CORE STATE STANDARDS
Interpret linear models.	<p>choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.</p> <p><u>HSS.ID.B.6.B</u></p> <p>Informally assess the fit of a function by plotting and analyzing residuals.</p> <p><u>HSS.ID.B.6.C</u></p> <p>Fit a linear function for a scatter plot that suggests a linear association.</p> <p><u>HSS.ID.C.7</u></p> <p>Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.</p> <p><u>HSS.ID.C.8</u></p> <p>Compute (using technology) and interpret the correlation coefficient of a linear fit.</p> <p><u>HSS.ID.C.9</u></p> <p>Distinguish between correlation and causation.</p>
MATHEMATICAL PRACTICES	LEARNING PROGRESSIONS
<p>1. Make sense of problems and persevere in solving them.</p> <p>2. Reason abstractly and quantitatively.</p> <p>3. Construct viable arguments and critique the reasoning of others.</p> <p>4. Model with mathematics.</p> <p>5. Use appropriate tools strategically.</p> <p>6. Attend to precision.</p> <p>7. Look for and make use of structure.</p> <p>8. Look for and express regularity in repeated reasoning.</p>	<p>Statistics and Probability Progression</p> <p>http://commoncoretools.me/wp-content/uploads/2012/06/ccss_progression_sp_hs_2012_04_21_bis.pdf</p>

ENDURING UNDERSTANDINGS	ESSENTIAL QUESTIONS	KEY VOCABULARY
<ul style="list-style-type: none"> Students create and interpret appropriate models of data (dotplots, stemplots, boxplots, histograms, two-way tables, scatterplots). Students compare shape, center, and spread for distributions of quantitative data. Students understand how measures of center and spread change in response to extreme observations/outliers. Students standardize data using z-scores and percentiles. Students understand the Normal distribution and can use it to estimate population percentages. <i>Ex. Empirical Rule (68-95-99.7)</i> Students compute marginal and conditional distributions and can mathematically justify whether two categorical variables are associated. Students can describe associations between two quantitative variables (form, direction, strength). Students can fit a function to the data (linear, quadratic, exponential, etc.). Students interpret components of a line of best fit (slope and y-intercept) and the correlation coefficient, r. Students understand the distinction between correlation and causation. 	<p>How can data be expressed so that their accurate meaning is concisely presented to a specific audience?</p> <p>How can you compare two distributions of quantitative data utilizing shape, center, spread, and outliers?</p> <p>How do extreme observations affect summary statistics?</p> <p>In what ways does knowing the outcome of one event affect subsequent events?</p> <p>What are the patterns in data collected and how are they useful?</p> <p>What are the statistical tools that can be used to create models, interpret data, and make predictions?</p> <p>How can you explain the association between two categorical/quantitative variables?</p> <p>Does correlation imply causation? Explain why or why not.</p>	<p>Average</p> <p>Bar graph</p> <p>Bimodal distribution</p> <p>Box plot</p> <p>Correlation (r)</p> <p>Coefficient of determination (r^2)</p> <p>Cumulative frequency distribution</p> <p>Dotplot</p> <p>Empirical rule</p> <p>Explanatory variable</p> <p>Extrapolation</p> <p>Five-number summary</p> <p>Frequency</p> <p>Histogram</p> <p>Interpolation</p> <p>Interquartile range (IQR)</p> <p>Linear correlation</p> <p>Mean</p> <p>Median</p> <p>Mode</p> <p>Normal curves</p> <p>Normal distributions</p> <p>Outlier</p> <p>Parameter</p> <p>Percentile</p> <p>Pie chart</p> <p>Population</p> <p>Qualitative variable</p> <p>Quantitative variable</p> <p>Quartile</p> <p>Range</p> <p>Relative-frequency histogram</p> <p>Residual</p> <p>Residual plot</p> <p>Resistant measure</p> <p>Response variable</p> <p>Sample</p> <p>Scatter plot</p> <p>Skewed left/right</p>

		Slope Standard Deviation Standard Normal distribution Standard units Statistic Stemplot Symmetric distribution Uniform distribution z-score
--	--	---

RESOURCES	INSTRUCTIONAL STRATEGIES	ASSESSMENT
<ul style="list-style-type: none"> Graphing calculator: Create graphs of quantitative data Statistical Software (Fathom Dynamic Software, IBM SPSS Statistics, etc.) Desmos.com: See regression tutorial StatKey: Free resource for creating graphs: http://lock5stat.com/statkey/ Microsoft Excel/Google sheets: Create bar graphs, pie charts, segmented bar graphs Rossmann Chance Applet collection (http://www.rossmanchance.com/applets/index.html): Guess the correlation applet https://www.amstat.org/censusatschool/: Data resource from students around the world <i>Against All Odds</i>: https://www.learner.org/resources/series65.html http://mathbits.com/: Tips to help with graphing calculators Gapminder: https://www.gapminder.org/for-teachers/ AP Stats Monkey: This site includes a wonderful collection of resources written by teachers and collected by Jason Molesky. http://apstatsmonkey.com/StatsMonkey/Statsmonkey.html 	<ul style="list-style-type: none"> When asking students to describe a distribution, students must describe the shape, outliers, center, and spread (SOCS) of the distribution. The acronym SOCS may help students remember to describe each of these features. Often times, students forget to describe data in context of a scenario. Statistics problems are almost always about some form of real world data, which should be emphasized as the students describe SOCS. <ul style="list-style-type: none"> E.g., when describing the center of the distribution of games won by all NBA teams last year, students should mention that the center of the distribution is _____ number of games won, rather than solely the center is _____. The effect of outliers on a set of data is very important in the real world. <ul style="list-style-type: none"> An example that can be used would be to provide students with 4 different test scores: 75, 80, 70, and 95. Ask the students to find the student's average test score, as well as the median, standard deviation, and Interquartile range. Then ask what would happen to each summary statistic if the student cheated on his/her next test and received a 0. This should help them 	<p>Formative Assessment SBAC - http://www.smarterbalanced.org/</p> <hr/> <p>LAUSD Interim Assessment District assessments can be accessed through: http://achieve.lausd.net/math http://achieve.lausd.net/ccss</p> <p>Use your Single Sign On to access the Interim Assessments</p> <hr/> <p>California will be administering the Smarter Balanced Assessment as the end of course for grades 3-8 and 11. The 11th grade assessment will include items from all High School Common Core strands, including Statistics and Probability. For examples, visit the Smarter Balanced Assessment at: http://www.smarterbalanced.org/</p>

RESOURCES	INSTRUCTIONAL STRATEGIES	ASSESSMENT
	<p>understand which types of statistics are resistant to outliers.</p> <ul style="list-style-type: none"> • Provide students with several different sets of data. Ask them when it would be appropriate to use each type of graphical display. Emphasize the difference between graphs for categorical vs. quantitative data. <ul style="list-style-type: none"> ○ Use statistical software to create a set of data and display a dotplot, boxplot, and histogram. Ask students to critique each type of graph (advantages/disadvantages) and justify their reasoning. • The Normal distribution is applied throughout many of the inference procedures used in AP Statistics. This distribution should be addressed, but does not need to be as emphasized as much as in the AP course. • Help students understand the purpose of a z-score in terms of location in a distribution. <ul style="list-style-type: none"> ○ For example, being 10 units higher than the mean is significant if you are talking about height in centimeters, but not very significant if you are talking about a person's salary. • To introduce two-variable quantitative data, a variety of in-class activities may be used. You may have students collect data on their hand spans and height and analyze the correlation. • Many of the interpretations that are required for two-variable quantitative data are challenging for students (slope, y-intercept, correlation coefficient, etc.). It may help to provide the students with sentence stems for each interpretation and have them practice interpreting the meanings in context. 	

LANGUAGE GOALS for **low achieving, high achieving, students with disabilities and English Language Learners**

- Students will use comparative language when comparing the center and spread for distributions of data in context.

Example: The center for distribution 1 is **higher/lower** than the center for distribution 2. There is **more/less** variability in distribution 2 than in distribution 1.

- Students will collaboratively decide which graphical display would be best suited for a specific set of data and *justify* orally or in writing why this graphical display is preferred over another.

Example: I would choose (graphical display) because...

- Students will interpret the slope and y-intercept of a regression equation in context.

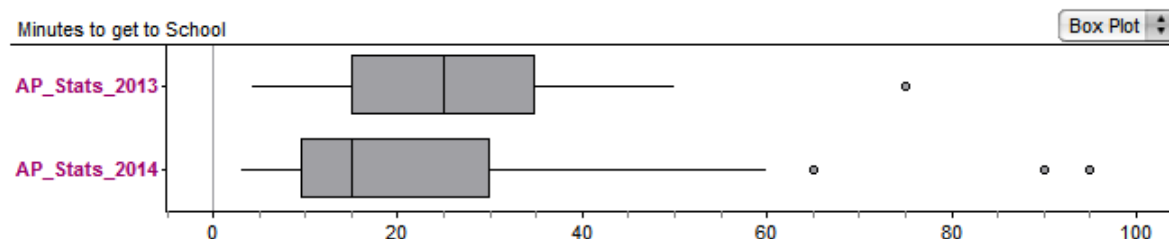
Example: Suppose we have a data set where x = the age of a plant in months, y = the height of a plant, and the regression equation is $\hat{y} = 1.2 + 2.3x$.

Slope: For every additional month that passes, the **predicted** height of the plant will increase by 2.3 cm.

Y-intercept: For a plant that is 0 months old, the **predicted** height is 1.2 cm.

PERFORMANCE TASK

An AP Statistics teacher asked two of his previous classes the question, “How long does it take you to get to school?” Below are parallel boxplots summarizing the students’ answers to this question. Compare the time it takes to get to school for the AP Stats classes in 2013 and 2014.



When asked to describe or compare distributions of data, students will be expected to describe the **Shape, Outliers, Center, and Spread (SOCS) in context of the scenario**. Here is a sample answer to the question above.

Shape: The distribution of minutes to get to school for the AP stats class in 2013 is *roughly symmetrical*, while the distribution of minutes to get to school for the AP stats class in 2014 is *skewed to the right*.

Outliers: For the class of 2013, there was one student who took 75 minutes to get to school, while in the class of 2014, there were 3 students whose times are considered outliers (65, 90, and 95 minutes to get to school). All four of these students took unusually longer to get to school than their classmates.

**** The 1.5(IQR) rule can be used to test for outliers if you are given the actual data.**

Center: The students in the class of 2013 take longer to get to school, on average, than the students in the class of 2014 (median = 25 minutes in 2013 vs. 15 in 2014).

**Measures of center include the mean and median.

Spread: There is more variability in the distribution of minutes to get to school for the class of 2014 than for the class of 2013. Although the middle 50% of the distributions are both spread over 20 minutes (definition of IQR), the distribution of data in the class of 2014 has a larger range ($95 - 3 = 92$ minutes) than the class of 2013 ($75 - 4 = 71$ minutes).

**Measures of spread include IQR ($Q3 - Q1$), standard deviation, and range ($\text{max} - \text{min}$).

Here are some additional questions to ask your students (strictly about quantitative data):

What was the cost of your last _____ (e.g. haircut)?

How many siblings/cousins/relatives do you have?


How many texts do you send per day? How many can you send per minute? (provide students with a list of messages and time them)

What is your GPA?

How old are you?

How many hours of sleep do you get? How long do you spend on homework/video games/watching Netflix/eating?

****Have the students create graphs of these topics and describe the distributions or compare different class sets of data using SOCS.**

DIFFERENTIATION 		
UDL/ FRONT LOADING	ACCELERATION	INTERVENTION
Statistics and Probability: <ul style="list-style-type: none"> As an opening activity, conduct an activity in which students collect a quick set of data from the class and analyze the mean, median, mode, and range. Survey questions could include: the number of letters in your name, number of cousins versus aunts and uncles, typical number of minutes to get to school, primary transportation method to get to school, shoe size. Model a simple activity involving collecting a quick set of data from the class and analyzing (calculate) the mean, median, mode, and range (e.g. the number of letters in your first and last name). Students could create a graphical display of the data collected from the activity above. Create a KWL (Know, Want to know, Learned) chart for unit vocabulary. 	Acceleration for high achieving students: <ul style="list-style-type: none"> While teaching two-variable quantitative data, encourage students to visit the site http://www.tylervigen.com to demonstrate that correlation does not imply causation. Have students research other publications that involve correlation. Present students with Anscombe's Quartet to help students understand the value of graphing data prior to calculating summary statistics. Present an example of Simpson's Paradox (2016 FRQ #6 from the AP Statistics Exam). 	<ul style="list-style-type: none"> Students who are having difficulty understanding standard deviation may need to be introduced to the idea of mean average deviation (MAD) first. This may help students understand the idea of average variability from the mean more clearly. Many of the interpretations that are required for two-variable quantitative data are challenging for students (slope, y-intercept, correlation coefficient, etc.). It may help to provide the students with sentence stems for each interpretation and have them practice interpreting the meanings in context.