|  |  |
| --- | --- |
| **Unit** | **S2** |
| **Title** | Statistics 2 |
| **Target Standards** | **S.ID.6, S.ID.6a, S.ID.6b, S.ID.6c, S.ID.7, S.ID.8, S.ID.9, S.ID.3, F.IF.6** |
| **Mathematical Goals** | **Students will…**   * Represent data on two quantitative variables on a scatter plot S.ID.6 * Describe how two quantitative variables on a scatter plot are related S.ID.6 * Interpret the slope and the intercept of a linear model in the context of the data S.ID.7 * Quantify the goodness of fit of a small data set by plotting and analyzing residuals S.ID.6b * Know that the correlation coefficient, r, is a measure of how linear the data are S.ID.8 * Know that r values closer to 1 imply a higher degree of correlation S.ID.8 * Learn how to compute correlation coefficient using the technology of your choice S.ID.8 * Practice steps to find the “best” line using technology of your choice S.ID.6a * Fit a linear function for a scatter plot that suggests a linear association S.ID.6c * Understand that two variables that are correlated are not necessarily causal S.ID.9 |
| ***The story before this unit (including prior knowledge)*** | From their experiences with linear functions in eighth grade, students are familiar with the ideas of slope, intercepts, and are able to write and interpret equations of lines. In eighth grade, students also learned to construct scatterplots and describe patterns that they saw. For scatterplots that suggested a linear association, they informally fit a straight line and judged its closeness to the data points. They wrote equations for these informal linear models, and interpreted their slope and intercepts in the context of the data. |
| ***The part of the story happening in this unit*** | In this unit, students will build on the statistical work they did in eighth grade. They’ll assess the fit of a function more precisely by plotting and analyzing residuals. They’ll be able to precisely compute the best possible fit line by using the regression function of a graphing calculator or other software. They’ll be able to compare the degree of correlation between various variables by computing and interpreting the correlation coefficient of a linear fit. Finally, they’ll be able to recognize situations with a causal relationship, versus situations that are correlated but not causal. |
| ***The story after this unit*** | The study of data and representing data will be a theme that is seen throughout the later units. Additionally, students will encounter nonlinear bivariate data and will be able to expand on the ideas of this unit in their later work. |

**UNIT FLOW SUMMARY**

|  |  |
| --- | --- |
| **UNIT S2** (10 days) | **Statistics 2** |
| **Section 0** (1 day) | **Diagnostic Pre-Unit Assessment** |
| **Section 1** (2 days) | **Preview (hook/umbrella activity)** |
| **Section 2** (3 days) | **Getting More Precise in Assessing Fit: Residuals and Regressions** |
| **Section 3** (2 days) | **Interpreting the Correlation Coefficient** |
| **Section 4** (1 day) | **Correlation vs Causation** |
| **Section 5** (2 - 4 days) | **Bringing It All Together** |
| **Section 6** (1 day) | **Summative Assessment** |

|  |  |
| --- | --- |
| **Section 0:** 1 day | **Diagnostic Pre-assessment** |
| **Pre-Unit Assessment Targets** | Diagnose students’ ability to create linear functions from data and interpret in a context 8-F.4   * Determine a rate of change and initial value given several data points that exhibit a linear relationship * Write an equation for a linear relationship * Use the equation to make predictions * Interpret the parameters of a linear equation in a context |
| **Sample Activity** | * [*Pre-Assessment*](https://docs.google.com/a/mathalicious.com/document/d/1GHRF5200P3IKSr6dGlnjWi02BFF1qxK_8zdd7tNizFA/edit) |



|  |  |
| --- | --- |
| **Section 1:** 2 days | **Preview (hook/umbrella activity)** |
| **Mathematical Goals** | Students will...   * Activate prior experience in eighth grade with informally fitting a straight line to a scatterplot and informally judging its goodness of fit 8.SP.2 * Model the relationship with an equation for a line and use it to make predictions 8.SP.3 * Represent data on two quantitative variables on a scatter plot S.ID.6 * Describe how two quantitative variables on a scatter plot are related S.ID.6 * Interpret the slope and the intercept of a linear model in the context of the data S.ID.7   (Note: since 8.SP.2 and 8.SP.3 are prerequisites, they are not target standards in this unit. However, they are standards involved in one of the suggested hook activities.) |
| **Narrative overview of section**  (and how the standards are achieved) | Students are presented with bivariate data that suggest a strong linear correlation. They will plot the points by hand or with technology, informally create a line of best fit, write an equation for the line, and use the equation to make predictions. Familiarity with these informal techniques is a good basis for the work of this unit: assessing fit more precisely by analyzing residuals S.ID.6 and computing the correlation coefficient of a linear fit, S.ID.8, as well as continuing the work of interpreting the slope and intercept of a linear model in the context of data S.ID.7  Additionally, students are given two linear models and are asked to interpret and use these models to make future predictions. |
| **Sample Activity 1.1** | [Let Down Your Hair](http://mathalicious.com/lessons/let-down-your-hair) , Mathalicious  **WHAT:** In the animated film *Tangled*, Rapunzel’s hair is long - really long! How fast must her hair grow in order for it to get that long? To address this question, students collect data from photographs of a real-life Rapunzel named Ida Pedersen, who did not cut her hair for ten years. They create a scatterplot of Ida’s hair length over time, fit a line to match the data 8.SP.2, and interpret the slope of this line as her hair’s growth rate 8.SP.3 MP.4. Then, they analyze the realism of the portrayal of Rapunzel’s super-long locks.  **WHY:** A lesson was purposefully chosen where there is no question about whether one variable depends on the other, in order to be able to focus on what students remember about linear functions 8.F.4 without the complicating factor of evaluating how well-correlated the variables are. This task is focused on 8th grade standards so can be used to review as students need. |
| **Sample Activity 1.2** | [Texting and Grades II](http://www.illustrativemathematics.org/illustrations/1028), Illustrative Mathematics  **WHAT**: Students are given a graph that shows Number of Texts Sent vs. GPA for 52 high school students. The regression equation is also given. Students are asked to interpret the quantities within the regression equation in the context of the data. Students may not have seen the regression equation in middle school so the notation of the regression equation may need to be explained in this problem.  **WHY**: A check for general understanding of the meaning of the slope and y-intercept quantities within a given regression line equation S.ID.7. This could be used if your students need more opportunities to practice interpreting a linear model. |
| **Sample Activity 1.3** | [Olympic Men’s 100-meter dash](http://www.illustrativemathematics.org/illustrations/1554), Illustrative Mathematics  **WHAT**: In this task, students are given a scatterplot, equation, and graph that show how the finishing time for the men’s 100-meter dash has changed over the past 100+ years. Students are asked to interpret what the given information means about the data in context S.ID.6. Then, students are asked to use the model to extrapolate the finish time from an earlier Olympics, and asked to determine a realistic domain. [Here’s an accompanying video](http://www.nytimes.com/interactive/2012/08/05/sports/olympics/the-100-meter-dash-one-race-every-medalist-ever.html).  **WHY**: Students are still interpreting a given model, but going further by asking students to use the model to make predictions MP.4, and determine a reasonable domain S.ID.7. This could be used if your students need more opportunities to practice interpreting a linear model. |
| **Focus Standards** | S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.  S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related. |
| **Mathematical Practices** | MP.4 |

|  |  |
| --- | --- |
| **Section 2:** 3 days | **Getting More Precise in Assessing Fit** |
| **Mathematical Goals** | Students will...   * Create and interpret a line of best fit S.ID.6 * Quantify the goodness of fit of a small data set by plotting and analyzing residuals S.ID.6b * Learn steps to find the “best” line using technology of your choice |
| **Narrative overview of section**  (and how the standards are achieved) | Until now, students have estimated the location of best-fit lines by eyeballing. We want to start quantifying how good or bad a fit is through plotting and analyzing residuals S.ID.6.b, so that we have some basis for knowing we have the “best” fit line. |
| **Sample Activity 2.1** | [Laptop Battery Charge 2](https://www.illustrativemathematics.org/illustrations/1559), Illustrative Mathematics  **WHAT:** Students are given a series of times and laptop battery levels, like 41% at 9:11AM. The data includes updated battery levels at different times, as the battery is charging. Students are asked to draw conclusions about when the battery will be charged, the rate at which it’s charging, and how long it would take to fully charge from 0%. While students are asked to interpret and understand the given data MP.1, students are not told how to represent the given data and thus must use appropriate tools strategically to complete this task MP.5. This is a great task to incorporate technology; how to plot points and create and interpret the line of best fit S.ID.6.  Note: Using the given times may pose a challenge. When setting up the axis for this graph students must understand that the initial time is set as time 0 and all future times build from there. The teacher may need to scaffold this part of the problem. Additionally, the teacher may use the [Laptop Battery Charge](https://www.illustrativemathematics.org/illustrations/1558) with 8th grade standards to scaffold as needed.  **WHY:**  The purpose of this task is for students to demonstrate an understanding of how to create and interpret given bivariate data. Students demonstrate that they can closely examine data and notice, both graphically and through average rate of change calculations F.IF.6, when something isn’t a perfect linear fit. Despite the fact that the data isn’t completely linear, students should be able to note that a linear model does give a fairly accurate understanding of the relationship between time and the laptop battery charge for this given scenario MP.4. |
| **Sample Activity 2.2** | [Used Subaru Foresters I](http://www.illustrativemathematics.org/illustrations/941), Illustrative Mathematics  **WHAT:** Students use technology to make two different scatterplots in this problem S.ID.6. They then generate regression lines and interpret them to compare the two sets of data, create a residual plot S.ID.6.b and argue which is a better predictor of used car price; age or mileage. Requiring a robust justification verbally or in writing will ensure students’ engagement with MP.3. Students use the regression equation found using technology to predict the price of a car given the age and mileage. By completing this task they are necessarily modeling with mathematics MP.4.  **WHY:** It’s hard (impossible?) to understand why the correlation coefficient is important without having two dependent variables (related to the same independent variable) to compare. This task gives students an opportunity to practice their recently-acquired tech skills, create and understand the meaning of a residual plot S.ID.6.b, while learning about the significance of the correlation coefficient S.ID.8. |
| **Focus Standards** | S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.  S.ID.6.b Informally assess the fit of a function by plotting and analyzing residuals.  S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.  F.IF.6 Calculate and interpret the average rate of change of a function (presented symbolically or as a table) over a specified interval. Estimate the rate of change from a graph.★ |
| **Mathematical Practices** | MP1, MP.3, MP.4, MP.5 |

|  |  |
| --- | --- |
| **Section 3:** 2 days | **Interpreting the Correlation Coefficient** |
| **Mathematical Goals** | Students will...   * Know that the correlation coefficient, r, is a measure of how linear the data are S.ID.8 * Know that r values closer to 1 imply a higher degree of correlation S.ID.8 * Learn how to compute correlation coefficient using the technology of your choice S.ID.8 * Practice steps to find the “best” line using technology of your choice S.ID.6a * Practice interpreting the meaning of the slope and intercept in the context S.ID.7 * Fit a linear function for a scatter plot that suggests a linear association S.ID.6c |
| **Narrative overview of section**  (and how the standards are achieved) | Students now know that a calculator has the power to find the “best” line, and that it has something to do with residuals. But how can we gauge just how good that best line is? Conversely, how close to linear is a linear-ish set of data? Students will learn to read the correlation coefficient from their technology of choice when computing the best fit line, and how to interpret its value S.ID.8. |
| **Sample Activity 3.1** | [Hitting the Slopes](http://mathalicious.com/lessons/hitting-the-slopes), Mathalicious  **WHAT**: It seems that certain countries are perennial powerhouses in the Winter Games. So, is there a way to use existing data to predict how many medals an individual nation will end up taking home? Two researchers think they may have found a solution. In this lesson, students create and use scatterplots S.ID.6 and linear regression S.ID.6c to examine several variables that may help predict Olympic performance MP.5. Students graphically analyze different factors such as GDP vs. number of medals won, interpreting the slope and intercept in the specific context S.ID.7. Are the Winter Games largely decided before the opening ceremonies even start?  **WHY**: This lesson is an opportunity to practice interpreting the meaning of the correlation coefficient S.ID.8, as it attempts to correlate the number of medals a country wins vs several different independent variables. Students demonstrate their ability to read and interpret scatterplots and linear regression by making supported predictions based on the given data MP.3. |
| **Sample Activity 3.2** | [Impact of a Superstar](http://illuminations.nctm.org/Lesson.aspx?id=2303), NCTM Illuminations  **WHAT**: Students will plot total points vs minutes played by each of the starters on the Los Angeles Lakers and Detroit Pistons MP.5. Then, one at a time, students will remove one player's data from the set and determine what effect, if any, the removal of that player's data has on the line of best fit and correlation coefficient S.ID.8. For the Lakers, students will notice that the correlation coefficient is 0.75 when the data for all players is considered. However, when the data for Kobe Bryant is removed, the *r*‑value increases to 0.95; when the data for any other player is removed, the correlation coefficient either stays the same or decreases. This indicates that the data for Kobe Bryant might be an outlier MP.7.  **WHY**: Students get a detailed view of how adding or removing one point can affect the correlation coefficient. This gives them a perspective on the meaning of the correlation coefficient S.ID.8, and also suggests one possible effect of an outlier S.ID.3 on the measures for a set of data. |
| **Focus Standards** | S.ID.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).  S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.  S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.  S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.  S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit. |
| **Mathematical Practices** | MP.3 MP.5 MP.7 |

|  |  |
| --- | --- |
| **Section 4:** 1 day | **Correlation vs Causation** |
| **Mathematical Goals** | Students will...   * Understand that two variables that are correlated are not necessarily causal S.ID.9 |
| **Narrative overview of section**  (and how the standards are achieved) | Students are now experts at figuring out how closely two sets of data are correlated, but they may be susceptible to the logical fallacy that correlation implies causation. Some well-chosen tasks can warn them against drawing such a false conclusion S-ID.9. |
| **Sample Activity 4.1** | [High Blood Pressure](http://www.illustrativemathematics.org/illustrations/1100), Illustrative Mathematics  **WHAT:** Students are presented with correlated data (time of television watched every day vs. likelihood of having high blood pressure) and asked to decide and justify whether that implies causation. Requiring students to make a thorough justification will ensure they engage with MP.3.  **WHY:** Students know how to determine how closely two variables are correlated, but have not yet been asked whether correlation implies causation S.ID.9. Through examples like this, students are introduced to the idea that just because two variables are correlated, doesn’t mean that one causes the other. |
| **Sample Activity 4.2** | [Pickles](https://docs.google.com/a/hightechhigh.org/file/d/0B784Ztw6k5CsbWhGY3c2NUVDTnM/edit?usp=drive_web) [& Test Scores](https://docs.google.com/file/d/0B784Ztw6k5CsbWhGY3c2NUVDTnM/edit), High Tech High, Jade White  **WHAT:** For two math classes, we are given a table with names, number of pickles eaten, and test scores. One of these shows a correlation and the other does not. Students are asked to make a prediction about correlation, then use technology to create scatterplots, regression lines, S.ID.6, S.ID.6a and correlation coefficients S.ID.8. The task does give specific direction in using technology, however, if students are afforded some choice , they can exercise MP.5. After interpreting these, they revisit their prediction.  **WHY:** This task is an opportunity to apply all the skills learned in this unit, but with a context that is clearly not causal. This task is good fodder for a discussion over whether correlation implies causation S.ID.9 MP.3. |
| **Focus Standards** | S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.  S.ID.6a 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.  S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.  S.ID.9 Distinguish between correlation and causation. |
| **Mathematical Practices** | MP.3, MP.5, |

|  |  |
| --- | --- |
| **Section 5:** 2 - 4 days | **Bringing it All Together** |
| **Mathematical Goals** | Students will...   * Interpret the meaning of the slope and intercept in the context S.ID.7 * Describe how two quantitative variables are related S.ID.6 * Use technology to create the line of best fit S.ID.6c * Fit a function to given data and use this to make further predictions S.ID.6a * Compute and understand that meaning of the correlation coefficient S.ID.8 * Understand that two variables that are correlated are not necessarily causal S.ID.9 |
| **Narrative overview of section**  (and how the standards are achieved) | In this section, students bring together all that they have learned in this unit. Students are either given or they collect data and, using technology, create the line of best fit and calculate the correlation coefficient. The final step is to analyze the situation and draw conclusions based on their findings. |
| **Sample Activity 5.1** | [Win at Any Cost](http://mathalicious.com/lessons/win-at-any-cost), Mathalicious  **WHAT**: Professional sports teams drop serious cash to try and secure the very best talent, and the dough can really pile up. But is all that money well spent? Conventional wisdom says that teams with higher payrolls ought to perform better than those with more modest means. Some leagues have even instituted limits on spending in order to make games more competitive. But do higher-spending teams really do better, and do salary caps actually level the playing field? In this lesson, students look at data for four major pro sports leagues and try to answer the question: Can you buy wins?  **WHY**: This lesson is placed here because it ties together all of the major components in this unit by applying them to model data MP.4 to answer a compelling question in context: do higher payrolls lead to more wins? Students use technology to create lines of best fit S-ID.6c, compare the relative strengths of different correlation coefficients S-ID.8, interpret the parameters of regression equations in a context S-ID.7, and discuss whether correlation implies causation, in this context S-ID.9. |
| **Sample Activity 5.2** | [Coffee and Crime](http://www.illustrativemathematics.org/illustrations/1307), Illustrative Mathematics  **WHAT:** In this task, data about the number of coffee shops and property crimes in the county are given. Students are initially given 8 data points and must describe the relationship S.ID.6, compute the correlation coefficient S.ID.8 and use this to describe and analyze the relationship between coffee shops and crime rates. In the later part of this task, students are presented with the least-squares equation and must use this for further predictions S.ID.6a S.ID.7. The last part of this task has students analyze and discuss MP.3 whether the given data supports causation for increased crime with an increased number of coffee shops S.ID.9.  **WHY:** The purpose of this task is to pull together all the big ideas of interpreting and analyzing bivariate data seen in this unit. Students demonstrate that they correctly know how to use technology to graph given data MP.4 and can accurately and with detail analyze and draw conclusions from data and the created graphs and calculations MP.6. |
| **Sample Activity 5.3** | Unit Project: [Lesson Plan](https://docs.google.com/a/hightechhigh.org/document/d/1hNyRF6VXOTa-m00inbPCGOsebJZnL82rZAoXBhu3bCo/edit) and [Analysis & Conclusion Requirements](https://docs.google.com/a/hightechhigh.org/document/d/1O4_hdNuLJHktH_e0JE34Q-xdKHKakUhDNaJQG5ESeXs/edit), High Tech High, Jade White  **WHAT:** In this project, students create and test an easily measurable hypothesis. Students collect data, record and graph their data and then complete an analysis of their data looking to determine if their hypothesis was correct. As part of the analysis, students must represent their data on a scatter plot S.ID.6, find the line of best fit S.ID.6c, interpret the slope and intercept in their given context S.ID.7 and calculate and interpret the correlation coefficient S.ID.8.  **WHY:**  This unit project allows students to demonstrate their ability to actively participate in the entire data collection, analysis and conclusion process. Students must make sense of the project and persevere through the process MP.1, attend to precision when measuring and modeling their data MP.4 MP.6, construct a viable argument in their conclusion based on their dataMP.3. |
| **Focus Standards** | S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.  S.ID.6a 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.  S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.  S.ID.7 Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.  S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.  S.ID.9 Distinguish between correlation and causation. |
| **Mathematical Practices** | MP.1, MP.3, MP.4, MP.6 |

|  |  |
| --- | --- |
| **Section 6:** 1 day | **Diagnostic Summative Assessment** |
| **Assessment Targets** | Assess students’ ability to   * create a scatterplot and a regression line using technology S.ID.6 * interpret the meaning of slope and intercept in context S.ID.7 * interpret the meaning of the correlation coefficient S.ID.8 * articulate the difference between correlation and causation S.ID.9 |
| **Sample Activities** | [Unit Assessment](https://docs.google.com/a/hightechhigh.org/file/d/0B784Ztw6k5CsVmgxckdkV1VLVnc/edit?usp=drive_web) ([document with answers but missing graphs](https://docs.google.com/a/hightechhigh.org/file/d/0B784Ztw6k5CsOVFMUG5OeElaaFE/edit?usp=drive_web)) |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1.1 | 1.2 | 1.3 | 2.1 | 2.2 | 3.1 | 3.2 | 4.1 | 4.2 | 5.1 | 5.2 | 5.3 |
| S.ID.3 |  |  |  |  |  |  | **X** |  |  |  |  |  |
| S-ID.6 |  |  | **X** | **X** | **X** | **X** |  |  | **X** |  | **X** | **X** |
| S.ID.6a |  |  |  |  |  |  |  |  | **X** |  | **X** |  |
| S.ID.6b |  |  |  |  | **X** |  |  |  |  |  |  |  |
| S.ID.6c |  |  |  |  |  | **X** |  |  |  | **X** |  | **X** |
| S-ID.7 |  | **X** | **X** |  |  | **X** |  |  |  | **X** | **X** | **X** |
| S-ID.8 |  |  |  |  | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** |
| S-ID.9 |  |  |  |  |  |  |  | **X** | **X** | **X** | **X** |  |
| F-IF.5 |  |  |  |  |  |  |  |  |  |  |  |  |
| F.IF.6 |  |  |  | **X** |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1.1 | 1.2 | 1.3 | 2.1 | 2.2 | 2.3 | 3.1 | 3.2 | 4.1 | 4.2 | 5.1 | 5.2 | 5.3 |
| MP.1 |  |  |  | **X** |  |  |  |  |  |  |  |  | **X** |
| MP.2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MP.3 |  |  |  |  | **X** |  | **X** |  | **X** | **X** |  | **X** | **X** |
| MP.4 | **X** |  | **X** | **X** | **X** |  |  |  |  |  | **X** | **X** | **X** |
| MP.5 |  |  |  | **X** |  |  | **X** | **X** |  | **X** |  |  |  |
| MP.6 |  |  |  |  |  |  |  |  |  |  |  | **X** | **X** |
| MP.7 |  |  |  |  |  |  |  | **X** |  |  |  |  |  |
| MP.8 |  |  |  |  |  |  |  |  |  |  |  |  |  |