Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
* Levels: B = Beginning, I = Intermediate, A = Advanced					
† Component: FQ = Formulate Questions, CD = Collect Data, AD = Analyze Data, IR = Interpret Results					
B.FQ	Students distinguish between a statistical solution and fixed answer. Statistical questions are posed within a context that anticipates variability.	Students recognize that statistical questions 1) are set in a context where one wants to know something; 2) are based in variability or uncertainty; 3) are always data based; and 4) have solutions that are approximations or estimates that emerge from data analysis, while a deterministic question is based upon exact calculations or theoretical deductions elicited from prior certain knowledge.	Possible work products might include the recognition that an appropriate question can address a given context. Work products include the identification of statistical questions based on existing data sets, or statistical questions that would require the production of new data. Students provide artifacts describing the difference between a statistical question and a deterministic question, and can provide examples of each.	Upper category responses will include the justification and critique of why certain questions are considered statistical in nature. A clear distinction is made between statistical questions that involve variability and deterministic questions that have fixed answers.	Lower category responses will include the identification of appropriate questions for a given context. Responses may blur the distinction between deterministic and statistical questions by neglecting the role of variability, or missing the importance of the data based context for a statistical question.
I.FQ.1	Students begin to pose their own questions.	Students formulate questions that would produce data appropriate for the context. These questions should demonstrate an understanding that a statistical question is one for which appropriate data can be collected and analyzed in order to determine a plausible answer that takes into account the variation in the data.	Possible work products might include the posing of questions by students for which it is possible to collect appropriate data. Students explain or identify the types of data that must be collected, and from whom or what it must be collected, in order to find plausible answers to the questions posed. Students identify characteristics of a question that make it appropriate for a given context.	Upper category responses will produce appropriate statistical questions that are possible to address with data and describe the type of data that would be appropriate to address the question in context.	Lower category responses will have students producing questions that are relevant to the context, but are not statistical and/or the data cannot be obtained to answer the question.
I.FQ.2	Students pose questions, address issues involving a group, and begin to recognize the distinction among a population, a census, and a sample.	Students demonstrate an understanding that a sample is a subset of a population and that a sample is taken when examining the entire population is not possible or feasable. Students can distinguish between situations where a small group (e.g., a classroom) is the entire population (census) and when it is a sample from a larger population (e.g., the classroom is used to answer a question about an entire grade level in a school).	Possible work products might include the posing of questions by students that demonstrates an understanding that data may be collected to answer questions about a larger group of interest. Students will determine if data required to answer a question represent a sample or census. Students will interpret reported statistics and be able to determine if data to address a question came from a sample or census.	Upper category responses will be able to identify whether collected data should be/are a sample or census when posing questions or interpreting reports of statistics. Students recognize the distinction between population, census, and sample.	Lower category responses will pose statistical questions but will not distinguish populations and samples, or do not recognize that a group can represent a sample or a population depending on the context.
A.FQ.1	Students formulate questions and determine how data can be collected and analyzed to provide an answer.	Students should be able to formulate questions and determine when a survey, observational study, or experimental study should be used to address them. Understanding when to collect data from a random sample and when to assign treatments randomly to available subjects is an important distinction for students to make.	Possible work products might include the development of questions by students for a given design that is provided. Other possible work products include critiques of questions and associated designs that are provided. Students should be able to distinguish among a question that is most appropriately addressed through a sample survey, observational study, or experiment.	Upper category responses will clearly formulate appropriate questions and designs to address the issue of interest. Limitations of the study will be clearly described or identified in context.	Lower category responses will contain vague questions or haphazard designs that do not use randomization to address the issue of interest. Explanations will often mimick textbook definitions without context for the current question or setting.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
B.CD.1	Students understand how to conduct and represent a census	Students devise and implement a method to collect and record data on selected variables, including numerical and categorical, from all students in the classroom or other well-defined small population. Students can represent a census of a classroom in a manner that preserves all information in the data.	Possible work samples might include the production of a detailed plan for conducting the census, creation of a list of reasons why the census may not reach all of the population, an explanation of why a "case study" presented is inappropriate (i.e., identify flaws). May involve the creation of a bargraph.	Upper category responses will detect errors in examples of suggested methods for conducting a census, be able to suggest a valid method of collection. Upper category responses will result in a complete data representation for all members in the population. It is clear that the representation contains all the information needed to answer the original question in context.	Lower category responses will suggest inappropriate methods of collection of data for a census. Lower category responses may also result in an incomplete data representation, a representation that does not tie directly back to the data, or incorrectly coded responses for a subset of members in the population.
B.CD.2	Students understand individual to individual natural variability	Students recognize and can explain why measuring the same variable on different people or objects or conducting repeated measures on the same person or object will generally result in differing outcomes. Students understand what is a reasonable amount of variability to anticipate for a set of data. Students may be able to identify informal reasons for variability within the data.	Possible work products might include the fact that students will anticipate and recognize potential variability as well as be able to identify possible sources of variability. Students might collect measurement data on, say, heights of students in the class and observe that there is variation among the numbers that reflect the actual variation in the heights when the students stand in a row. For this particular set of data they may also be able to identify what is reasonable to expect (e.g., you would not expect a 1st grader to be 6 feet tall).	Upper category responses will show that students understand to expect a reasonable difference between individuals. The student can explain that the variation in the data reflects the natural variation in the people or objects being observed or measured. A student can identify multiple reasons for variability in the data.	Lower category responses will recognize that there will be variability but will not have reasonable expectations for the amount of variability between individuals. Students are unlikely to identify potential reasons for variability in the data.
B.CD.3	Students understand how to conduct simple experiments with assignment of treatments.	Students demonstrate the understanding that an experiment can be designed to measure the effect of treatments by applying them to experimental units and recording values of a response variable.	Possible work products include students specifing or critiquing the design for a simple experiment to answer a research question of interest, which includes the random assignment of treatments to the experimental units in order to form different groups to be compared on a particular response variable. Students should be able to identify the experimental units, treatment groups, response variable and other variables.	Upper category responses will show understanding of all aspects of the design of a simple experiment, including how the treatments were applied to the different experimental units and what response variable will be obtained for comparative purposes.	Lower category responses will show understanding that the treatment groups are being compared, but they may not contain a clear description of the application of treatments to the experimental units or the response variable of interest. These responses may indicate a lack of understanding of why two groups were formed for comparison.
B.CD.4	Students understand induced variability attributable to an experimental condition.	Students recognize that there are multiple sources of variability (e.g. induced and natural) and explain how experimental treatments can create variation between groups. Students recognize how induced variability is different than individual to individual variability.	Possible work products might include the students recognizing and identifing possible sources of variability. Students are able to distinguish between variability induced by the treatments and the natural variability of the experimental units	Upper category responses will provide clear distinctions between different sources of variability with particular focus on how experimental treatments can induce variability.	Lower category responses will identify that there is variability in data but will not recognize that there is a difference between induced and natural variability.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
B.CD.5	Students understand how to represent categorical and numerical data in appropriate ways.	Students demonstrate the understanding necessary to create appropriate data displays. Students recognize that some displays are appropriate for categorical data whereas others are appropriate for numerical data. Students are able to produce and identify issues in producing dotplots, frequency tables, stemplots, picture graphs, bargraphs, scatterplots, and time series plots.	Possible work products might include students producing various data displays. They may also have to identify appropriate aspects of the display (e.g., scale). Data displays may be presented and students would have to identify particular issues with the display.	Upper category responses will illustrate a clear understanding of how to produce data displays as well as an understanding of potential issues that may be involved with specific data displays. Responses will articulate that the appropriateness of a data display is determined by the variable and features of the distribution.	Lower category responses will illustrate an understanding of how to construct displays; however they will lack in understanding potential issues associated with particular dispalys. Lower category responses will not indicate that certain displays are more appropriate than others for particular data sets.
I.CD.1	Students understand how to design and conduct nonrandom sample surveys and begin to use random selection.	Students understand how to design and conduct surveys from both non-random and randomly selected participants. Students explain why random samples can provide more unbiased information about a population than other types of samples, such as convenience samples or self selected samples.	Possible work products might include students being able to explain how to design and conduct a survey or opinion poll administered to a randomly selected sample of people. Random devices (e.g., rolling die, random number generator, etc.) may be used to model random sampling from a population. Students should understand that it is possible to make inferences from sample results to the population if randomness is involved, and acknowledge a range of variability between sample results and population characteristics.	Upper category responses will recognize a clear distinction between random and non-random samples for conducting surveys and that only random samples may lead to generalizations about the population. If surveys are described, there is enough detail so someone could read the description and carry out the randomization process.	Lower category responses will not recognize the distinction between random and non-random samples or do not acknowledge the limitations on generalization in non-random samples. Students may not be able to describe a randomization process.
I.CD.2	Students understand how to design and conduct comparative experiments and begin to use random assignment.	Students understand how to design and conduct comparitive experiments using random assignment and demonstrate correct methods for planning data collection for comparison of treatments. Students randomly assign treatments to experimental units.	Possible work products might include students explaining how to design and conduct a comparative experiment that includes random assignment. Students should understand that it may be possible to make causal claims if experiments are conducted with random assignment.	Upper category responses recognize a clear distinction between a comparative experiment with and without random assignment and the resulting claims that can be made based on each design. If designs are described, there is enough detail so someone could read the description and carry out the randomization process.	Lower category responses will not recognize the distinction between random and non-random assignment or do not acknowledge the limitations on causal claims if the assignment is not random. Students may not be able to describe a randomization process.
I.CD.3	Students understand how to represent categorical and numerical data in appropriate ways.	Students demonstrate the understanding necessary to create appropriate data displays. Students recognize that some displays are appropriate for categorical data whereas others are appropriate for numerical data. Students are able to produce and identify issues in producing histograms, boxplots, and relative frequency tables.	Possible work products might include students producing various data displays. They may also have to identify appropriate aspects of the display (e.g., scale). Data displays may be presented and students would have to identify particular issues with the display.	Upper category responses will illustrate a clear understanding of how to produce data displays as well as an understanding of potential issues that may be involved with specific data displays. Responses will articulate that the appropriateness of a data display is determined by the variable and features of the distribution.	Lower category responses will illustrate an understanding of how to construct displays; however they will lack in understanding potential issues associated with particular dispalys. Lower category responses will not indicate that certain displays are more appropriate than others for particular data sets.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
A.CD.1	Students understand what constitutes good practice in conducting a sample survey.	Students demonstrate understanding of sampling techniques; are able to critique a poorly constructed survey; and make suggestions for good questions. Features considered include whether the population is well-defined; whether the sampling procedure is random or non-random; and whether the objectivity or bias of questions will result in valid/invalid answers.	Possible work products might include a critique of a presented survey, a suggestion of how to carry out an appropriate survey in a given context, a statement of characteristics of a good survey, or production of unbiased questions in a given context.	Upper category responses will be able to distinguish between survey strategies with justification, be familiar with the types of biases potentially affecting survey results, will take account of both sampling technique and survey question construction in judging sample surveys, recall the four main features in survey construction (population, sample, randomization process, and accuracy of measurements).	Lower category responses will likely focus on either sampling strategy or survey construction but not both when judging a sample survey, only address 1 or 2 characteristics of a well constructed survey.
A.CD.2	Students understand what constitutes good practice in conducting an experiment.	Students are able to identify, discuss, and explain the aspects of best statistical practice for designing an experimental study, including: the clear identification of the statistical question to be investigated; the variables under investigation; and the random assignment of treatments to the experimental units.	Possible work products might include students designing or critiquing an experimental study to investigate a given statistical question. Aspects of statistical design are included in the study, such as attempts to control for bias and attention to groups receiving different treatments to allow for the possibility of making causal conclusions about differences between groups.	Upper category responses will include attention to clean data production, random selection, and random assignment in a statistical design. Students are able to explain why random assignment makes it possible to draw causal conclusions.	Lower category responses will include design flaws, such as lack of attention to the importance of including a control group, or lack of random assignment to conditions in the experiment. Confusion may exist between the importance and purpose of random selection and random assignment, or what is gained by including them in the experimental design.
A.CD.3	Students understand what constitutes good practice in conducting an observational study.	Students demonstrate a clear understanding that observational studies are used in situations in which randomized assignment of treatments is impractical or impossible; that in such studies, data are subject to bias; that they require all known variables related to the response to be measured; and that they produce results that cannot show cause and effect, but can give evidence of possible patterns or associations.	Possible work products might include students designing or critiquing an observational study to investigate a given statistical question. Aspects of statistical design are included in the study, such as attempts to control for bias and the possibility of being able to generalize results to an entire population (random selection). Students' work discusses why the study did not include random assignment (either it was not necessary or it is impractical), what was the response of interest, what were the other related variables that were measured, and what are the possible biases in the data.	Upper category responses will include attention to clean data production, identification of variables, and random selection in observational studies. Students recognize that observational studies cannot show cause and effect but can give evidence of possible patterns or associations.	Lower category responses will not attend to certain aspects of data production, variable identification, or random selection. Student does not recognize the limits of the interpretations that can be made from the results of an observational study (e.g., not cause and effect relationships).
A.CD.4	Students understand how to design and implement a data collection plan for statistical studies, including observational studies, sample surveys, and simple comparative experiments.	Students recognize the appropriate data collection method for statistical questions, including observational studies, sample surveys, and simple comparative experiments. Students should understand the advantages and disadvantages of each data collection method for specific statistical questions.	Possible work products might include the students selecting the most appropriate data collection method for a given statistical question. Students will be able to describe the reasons for selecting a particular approach and address the advantages/disadvantages of each method. Given a data collection method, students will select the most appropriate question.	Upper category responses will identify the correct data collection method for a statistical question or identify the most appropriate question given a data collection method. Students understand the advantages and disadvantages of each data collection method.	Lower category responses will not reflect understanding of all the advantages and disadvantages of each data collection method and may not recognize the difference between the three types of statistical studies.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
B.AD.1	Students compare individual to individual and individual to a group.	Students demonstrate ability to find an individual in a group/graph; to distinguish an individual from others; to understand/identify the attribute that is being compared; to have criteria for distinguishing "same" from "different", and "close" from "far apart" in relation to difference; and to identify whether a value is typical or an extreme value on a graph.	Possible work products might include describing how many data values are above or below a given individual value on a graph, placing an extreme value on a graph, be able to add a value to the appropriate group in pictograph or to increase the frequency of the appropriate bar in a bar graph, to describe "how far" or "how close" a value is to the modal value (or median/mean) using the scale of the graph.	Upper category responses will identify extreme values, will distinguish large and small differences between an individual value and a given mean, median or modal value, will identify "typical" values in clusters of data points on a graph.	Lower category responses will have difficulty identifying individual values in a cluster on a graph, struggle to place extreme values on a plot, may identify a column in a bar graph as a value of an individual rather than a frequency of a group, but will be able to place a categorical data value within the appropriate grouping.
B.AD.2	Students become aware of group to group comparison.	Students are able to recognize characteristics of groups and describe variation that occurs across two or more different groups. Students are able to identify attributes that would be important or interesting to compare across multiple groups drawn from multiple populations.	Possible work products might include verbal and written comparisons of group characteristics, including frequencies, centers, spreads, and shapes of the data from two or more groups, the creation and comparison of tabular and graphical displays of data from multiple groups, and the identification of variables to compare across groups in a data set.	Upper category responses will articulate meaningful comparisons of frequencies, centers, spreads, and/or shapes of graphical or tabular displays of distributions of data of multiple groups. Possible reasons for any differences between groups are discussed.	Lower category responses will include comparisons of individual data points or cases, or idiosyncratic descriptions that may be grounded in context but not in the data. Students can describe characteristics of a single group but may have trouble comparing across groups.
B.AD.3	Students understand and describe a distribution using appropriate tools such as a bar graph, dotplot, stem and leaf plot, tables, mean/median/mode/range, and modal category.	Students clearly see univariate data as a collection of measurements describing an aggregate in context rather than reflecting on single data values and who or what each value might represent. Students are able to describe a distribution's pattern in terms of shape, center, spread and possibly unusual data values that deviate from the pattern. Students demonstrate correct recognition of and ability to construct bar graphs, dotplots, stem and leaf plots, constructs and interprets tables; and can calculate and locate on graphs the mean/median/mode/range and modal category. These categories are used to characterize typicality or representativeness. The range is used to characterize spread.	Possible work products might include a display of measurement data with an explanation of shape, center, clusters and gaps, spread and seemingly unusual data values, with explanations that relate to the context in which the data was collected. Possible work samples might include production of tables and bar graphs with modal category labeled and of dotplots with mean/median/mode/range labeled. Median is described as the middle of an ordered set of data and the mean is described as a "fair share" with both locating a "middle" of a distribution, whereas the range is identified with the idea of "spread." Discussion might include how the skewness of the distribution affects the mean and the median.	Upper category responses will show the students plotting measurement data and explaining the key features of the distribution in terms of shape (humps, valleys, tails), center (typical value), clusters and gaps, and spread (range), relating these to the context. Responses will reflect a clear understanding of each of the tools, accurate description of the shape of distributions and how this relates to the measurement tools, comparison of mean, median, and mode and characteristics of distributions when they are the same or different.	Lower category responses will see data as a whole but concentrate on single points or particular areas of distributions to make comparisons. Student does not clearly relate the pattern of the distribution to the context. Student summarizes only one or two characteristics (shape, center or variability) of a distribution. Responses will identify but not be able to construct graphs of the distributions, identify but not interpret the mean/median/mode/range for the distributions.
B.AD.4	Students observe and explore association between two variables making appropriate use of scatterplots.	Students recognize contexts where a change in one attribute may be related to a change in another attribute. Students demonstrate the ability to construct scatterplots and describe positive and negative relationships, or lack of relationship. Strength of association is demonstrated by degree of spread about the "trend line" in a scatterplot.	Possible work products might include the construction and interpretation of scatterplots with written description of the trend and pattern of the association shown. Alternatively, these scatterplots could be provided.	Upper category responses will be able to state clearly how a scatterplot represents 2 variables for each data point. Responses will describe a variety of relationships (positive or negative, strong or weak) between attributes. Student can interpret the association in context.	Lower category responses will reflect struggle to explain how each point on a scatterplot represents two attributes for each case plotted. Responses may contain unlabeled graphical displays, graphs with inappropriate scaling, and weak interpretations. Students may believe that that all associations are "positive" and fail to recognize negative associations.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
I.AD.1	Students quantify descriptions of a data distribution, including variability. Students use appropriate tools for summarizing distributions including dotplots, histograms, five number summaries, boxplots, interquartile range, and mean absolute deviation.	Students demonstrate correct uses of graphical displays, numerical summaries, and relevant analyses, with appropriate interpretations and connections to context. Students demonstrate understanding of the 5-number summary and the mean absolute difference and how they describe variability, and an ability to interpret graphical representations of variability (clustering, spread, gaps, and outliers). Students analyze a collection of measurement data by constructing an appropriate plot (e.g., dotplots, histograms, and boxplots) to reveal the shape and to summarize the center, spread and any unusual features by means of appropriate statistics.	Possible work products might include the students presenting an analysis of measurement data using one or more graphical displays and summary statistics aligned with those displays, along with an explanation of why the plot and summary statistics were chosen. Possible work samples might include calculating the range of a data set from a table or graph, calculating the average of absolute differences from the mean for a small data set, writing an interpretation of a 5-number summary, calculating and interpreting the interquartile range; comparing the middle 50% of the data with the entire data set from a graphical representation, creating a box plot from a five-number summary. The summary statistics and graphical displays are then used to describe key features of the data distribution in context.	Upper category responses will reflect a clear understanding of the global features of the distribution and will include appropriate numerical summaries. Student can explain why the chosen plot and data summaries are appropriate for these data, as compared to others that might have been chosen, and describe what they mean in context. Upper category responses will use the five number summary, interquartile range, and MAD appropriately to describe variability in a data set, and will create and explain a boxplot (including density). Student demonstrates an understanding of the differing characteristics displayed by common plots, for example, a histogram versus a box plot. The analysis and explanation should reveal an understanding that the mean and mean absolute deviation are more appropriate for nearly symmetric data while the median and interquartile range might	Lower category responses will have limited understanding of the distribution and will have an incomplete graphical display or numerical summary. Lower category responses may be restricted to finding the range of the data and identifying key features of a boxplot including outliers, may appreciate the need to find differences from the mean but not the importance of the absolute value, may think the larger the section of a box plot the more data values it represents rather than it representing a greater spread of the same number of values as the other three parts of the plot. Student does not demonstrate clear reasoning as to why plots or summary statistics were selected, or misinterprets summary statistics in terms of the context of the data, or incorrectly calculates summary statistics.
I.AD.2	Students compare two or more distributions using graphical displays and numerical summaries.	Students provide explicit comparison between centers, spreads, and shapes of two or more distributions and are able to relate similarities and differences in the comparisons back to the context from which the original data were generated.	Possible work products might include students comparing two or more distributions using sample statistics such as measures of center (median, mean), spread (range, interquartile range, MAD), and shape (mounds, skewness, symmetry, tail behavior, gaps). Conclusions about differences (or similarities) between the distributions relate back to the context.	Upper category responses will compare the distributions as objects themselves, rather than individual cases within the distribution. Summary statistics are included in the comparison analysis. Clear connections back to the context are made when making comparisons between or among distributions.	Lower category responses will focus primarily on differences between particular data points or extreme cases within the distributions of data and neglect making comparisons across the entire distributions themselves, or might compare aspects of distributions such as center, shape, and spread without using appropriate summary statistics. The context might be neglected in the comparison process.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
I.AD.3	Students acknowledge sampling variability.	Students demonstrate the understanding that different random samples from the same population are not likely to be identical, are likely to have different centers, and can have different shapes and amounts of variability. Measures (statistics) obtained from a sample may be similar to but not the same as the corresponding measures (parameters) from the population.	Possible work products might include students recognizing that, given a description of a population and a sampling context, all samples will not have the same shape, mean, and variability and that, in repeated random sampling, sample means will vary less than the variability in the population.	Upper category responses will show the students stating that random samples from the same population will not likely be identical, and that the measure of center for each sample will not likely be the same, and they identify an appropriate range for the sample mean.	Lower category responses will expect that every sample will produce the same measure of center (and that a sample will be identical to the population in center and variability).
I.AD.4	Students recognize and describe association between two categorical variables by comparing relative proportions within two-way tables.	Students demonstrates correct use of appropriate numerical summaries including conditional relative frequencies, and clearly interpret these proportions in context.	Possible work products might include the construction and interpretation of two-way tables. Alternatively, these two-way tables could be provided, and students must compute, compare, and interpret conditional relative frequencies in context.	Upper category responses will include constructing two way tables from data descriptions and will correctly compute and interpret appropriate statistics for assessing association. Interpretations recognize that a possible association is present when conditional relative frequencies are different across rows or columns.	Lower category responses will be able to make two-way tables from data but may attempt to describe association in terms of frequencies rather than relative frequencies. Student may also choose inappropriate relative frequencies to compare.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
I.AD.5	Students describe and quantify the strength of association between two numerical variables and use tools for exploring association. They use the quadrant count ratio (QCR) as a measure of strength of association, and simple lines for modeling association between two numerical variables. Students also look at trends in time series plots.	Students demonstrate correct use of the approximate fit lines for describing association and to make predictions. Students use QCR to measure the strength of association. Students clearly interpret these in context.	Possible work products might include students sketching approximate fit lines on scatterplots and computing the QCR. Students describe and interpret the direction and strength of association between variables using approximate fit lines and the QCR, and use these to make predictions.	Upper category responses will include reasonable sketches of the scatterplot with the fit lines and correct computations of QCR and interpretations of both. Students can relate the value of the QCR to characteristics of the plot. Students make appropriate predictions using the approximate fit line.	Lower category responses will be able to sketch a reasonable line but not make appropriate predictions from it. They may make errors in calculating and interpreting the QCR.
A.AD.1	Students identify appropriate ways to summarize numerical data from statistical studies using graphical displays and numerical summary statistics.	Students demonstrate correct uses of graphical displays and summaries for numerical data, with appropriate interpretations and connections to context. Special emphasis is made on interpreting strandard deviation.	Possible work products might include the production of graphical displays and numerical summaries of the data, with interpretations that relate to the shape, center and spread of the data. Explanations might include why the median and IQR were chosen as measures of centrer and spread rather than the mean and standard deviation.	Upper category responses will reflect a clear understanding of the global features of the graphical displays and numerical summaries, and will include appropriate interpretations for numerical data along with reasons for their selection, all related to context.	Lower category responses will contain correct displays but may have incomplete or missing calculation and/or interpretation of the numerical summaries.
A.AD.2	Students understand how sampling distributions (developed through simulation) are used to describe the sample-to-sample variability of sample statistics.	Given a sampling distribution of some sample statistic, students identify and describe aspects of the sampling distribution such as the shape, range, mean, and standard deviation of the sample statistic, as well as the 'likely range' of that statistic (e.g. the range where 90 – 95% of the sample statistics are likely to fall). Students recognize that the variability of a statistic depends on the sample size.	Possible work products might include the students interpreting the outcomes of simulation models that generate empirical distributions of a sample statistic (e.g., frequency of an outcome in a sample, sample proportion, sample mean, sample range, etc.), analyzing the resulting empirical sampling distribution, and identifying and discussing the implications of the 'likely range' of that sample statistic, accounting for the sample size.	Upper category responses will involve accurate explanations of how simulations will generate sample statistics that can then be summarized in tabular or graphical sampling distributions. Student describes the shape, center, and spread of the sampling distribution, and identify a 'likely range' for the sample statistic. A clear distinction is made between the distribution of a sample of data, and the distribution of the sample statistic obtained from calculating that statistic for repeated samples of data, noting the effect of sample size.	Lower category responses will demonstrate some aspects of upper category responses but not all of them; e.g., Students may see that the center of a sampling distribution is the same as the center of the population but mistakenly expect the shape and/or variability of the sampling distribution to remain the same as those of the population..

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
A.AD.3	Students explain and measure association between two categorical variables with data summarized in a two-by-two table by use of differences in row or column proportions.	Students demonstrate a clear understanding that association between categorical variables with data summarized in a two-by-two table can be measured by differences in row or column proportions, and can correctly relate these differences to the context and goals of the study.	Possible work products might include the student collecting data on two categorical variables with two categories each, summarizing the data in two-by-two tables and clearly explaining how possible associations can be measured by differences in row or column proportions; relating the possible associations to the context and goals of the survey, experiment or observational study.	Upper category responses will reflect a clear understanding that possible association in bivariate categorical data from a statistical study is shown by row (column) proportions differing across columns (rows) in a two-way table, and that in a two-by-two table this association can be measured by differences in proportions. It then correctly ties these differences to the context and goals of the study, perhaps	Lower category responses will show proper displays of categorical data and may reflect a clear understanding of what patterns would suggest possible association, but does not clearly or correctly relate the possible association to a difference in proportions.
A.AD.4	Students recognize when the relationship between two numerical variables is reasonably linear, know how Pearson's correlation coefficient measures the strength of the linear relationship between two numerical variables, and understand the least squares criterion in line fitting.	Students can provide a reasonable estimate of the Pearson's correlation coefficient for a scatterplot; identify linear and non-linear relationships in scatterplots; correctly interpret the strength of a linear relationship based on Pearson's correlation coefficient; understand the magnitude of a correlation coefficient represents the strength of association (e.g., r = -.65 and r = .65 indicate the same strength); understand and able to calculate a residual; understand that any straight line other than the best fit line (by least squares) will have a larger sum of squared residuals than the best fit line.	Possible work products might include students investigating the relationship between two quantitative variables and providing descriptions of the nature and strength of the relationship; identifying and explaining the role of Pearson's correlation coefficient in measuring the strength of a linear relationship; identifying and explaining the role of the least squares criterion for determining the line of best fit for a scatterplot. For example, explaining how the magnitude of a correlation coefficient represents the strength of association; calculating and interpreting a residual; understanding that any straight line other than the best fit line (by least squares) will have a larger sum of squared residuals than the best fit line.	Upper category responses will demonstrate, for example, an understanding of how the magnitude of a correlation coefficient represents the strength of association; the calculation and interpretation a residual; and understanding that any straight line other than the best fit line (by least squares) will have a larger sum of squared residuals than the best fit line.	Lower category responses will confuse understanding of positive and negative correlation coefficient; can calculate the value of a residual but misinterpret its meaning; may see linear patterns when they do not exist. Students may not make the connection between the interpretation and the representation.
B.IR	Students interpret data for a group and acknowledge that results may be different in another group.	Students understand that the data being studied represents one group of people or objects, and, in general, cannot be generalized to other groups of similar people or objects. Students can give clear reasons why such generalization may be unreasonable and why groups might differ.	Possible work products might include interpretations of displays of data on categories or measurements from one group of people or objects, with suitable descriptions, summaries, and conclusions based on the analysis of the data. Students recognize that another group may have differing values for certain identified reasons.	Upper category responses will interpret information presented in data displays and summaries to inform conclusions about one group and recognizes that the conclusions may not apply to another group. Students can describe the reasons why data from one group does not apply to another.	Lower category responses will provide weak or vague interpretations of data. Student may understand that conclusions about one group may not apply to another, but is unable to explain why.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
I.IR.1	Students describe differences between two or more groups with respect to center, shape, and spread.	Students provide or select appropriate interpretations of graphical displays and numerical summaries to compare two or more groups in the context of a study.	Possible work products might include the interpretation of displays (e.g., side-by-side boxplots) and summaries (e.g., five number summaries), to produce or critique statements about differences between groups. The student should focus on center, spread, and shape for comparing distributions.	Upper category responses will illustrate a clear understanding of how to compare two or more distributions including center, spread, and shape.	Lower category responses will ignore at least one of the main features (center, spread, or shape) of a distribuiton. Inappropriate statistics may be used to make the comparison (e.g., means and standard deviations may be used to compare distributions that are strongly skewed and contain outliers). Finally, the interpretations in these responses may be weak or poorly connected to the context.
I.IR.2	Students recognize that random sampling is unbiased (fair) and serves as the basis for inference from a sample to a population.	Students recognize that samples must be randomly selected from the appropriate population in order to allow for generalizations that extend beyond the sample from which the data were collected. Students recognize that sampling procedures that are not random are not unbiased and do not allow for generalizations to the sampled population.	Possible work products might include whether a sample is unbiased (fair) and explanations of why some samples are not unbiased. Suggestions may be provided for how an unbiased (fair) sample could have been selected in a given context. Other work products include the identification of appropriate and/or inappropriate interpretations of the data given the way the sample was selected.	Upper category responses will be able to identify appropriate interpretations depending on how a sample was selected. Responses may include reasons for why sample selection justifies the interpretations. If samples were not unbiased, students will be able to provide suggestions to improve sample selection.	Lower category responses may not be able to recognize a non-random sample or explain why it is inappropriate to draw generalizations. Students may not be able to provide suggestions to improve sample selection. Students may confuse haphazard with random samples.
I.IR.3	Students begin to distinguish between an observational study and a designed experiment.	Students understand the difference between the interpretations that can be made from observational studies and experiments. Experimental design includes random assignment which allows for causal statements. Observational studies do not include random assignment; therefore produce results that cannot show cause and effect but interpretations are possible based on observed patterns or associations.	Possible work products might include the students interpreting experimental and observational studies and identifying appropriate conclusions that can be drawn from each of them. Given an interpretation, students will select whether such a conclusion is appropriate from an observational study or experiment. Given a data collection method, students will select the most appropriate interpretation.	Upper category responses will indicate that cause-and-effect conclusions cannot be drawn from observational studies, and recognize appropriate conclusions that can be drawn from experiments and observational studies. Students identify the correct data collection method for an interpretation or identifies the most appropriate interpretation given a data collection method.	Lower category responses may realize that random assignment is associated with experiments; however they are unable to make appropriate interpretations in context.
I.IR.4	Students understand basic interpretations of measures of association and recognize that "association" does not imply "cause and effect."	Students interpret measures of association in order to determine if there is a relationship between variables. Students recognize that "association" does not imply "cause and effect."	Possible work products might include students interpreting measures of association between two categorical variables or two numerical variables. Students may recognize different types of association (e.g., linear and non-linear). Students discuss possible association revealed by the pattern and recognize that association does not imply causation.	Upper category responses will interpret contingency tables and considers relative frequencies (by row or by column) in order determine if a relationship exists. The students interpret a scatterplot and measures of association (e.g., QCR and correlation coefficient) and provides an explanation of the strength, direction, and type (e.g., linear vs. non-linear). Students recognize that association does not imply causation.	Lower category responses will focus on cell frequencies rather than relative frequencies when determining if a relationship exists. Students interpret measures of association and can recognize a relationship exists; however they are not able to explain the strength, direction, and/or type of the relationship. Students may not recognize that association does not imply causation.

Code*†	Description	Evidence Statement	Possible Work Products	Upper Category Observable Features	Lower Category Observable Features
I.IR.5	Students recognize sampling variability in summary statistics, such as the sample mean and the sample proportion.	Students recognize that sample statistics vary with repeated sampling. Students interpret the sampling variability in a summary statistic. The students should be able to interpret the sampling variability from simulation studies of statistics (e.g., sample mean, sample proportion, grand total/sum, median, range, IQR, and MAD). Students recognize how sampling variability is influenced by sample size.	Possible work products might include the recognition that variation exists and it is related to the variation in the population and sample size. Other work products include the impact of sample size on the sampling variability of summary statistics. Finally, work products may include the interpretation of results of simulation studies.	Upper Category reponses will demonstrate clear interpretations of summary statistics and the sampling variability in the distributions of those statistics. Students understand that the sampling variability is influenced by the variability in the population and sample size.	Lower category results will contain errors in interpreting the variability in summary statistics and incorrect or imprecise explanations of the sampling variability.
A.IR.1	Students understand the meaning of statistical significance and the difference between statistical significance and practical significance.	Students decide whether an observed difference is something that would be likely to be observed by chance and whether or not this difference has any practical meaning. Students recognize that significance is demonstrated by a result that is unlikely to occur by chance. Students recognize that statistical, but not practical, significance is influenced by sample size.	Possible work products might include the students being able to interpret a basic resampling simulation and use sampling variability in order to decide whether an observed difference is due to chance or not; students being able to interpret practical significance within the context; students recognizing that practical significance is based on the size of the observed difference and not statistical significance.	Upper category responses will be able to interpret statistical significance, practical significance and basic resampling distributions. Students should distinguish between statistical and practical significance. Student recognizes that some differences are unlikely to be observed by chance.	Lower category responses will have a textbook definition of the difference between practical and statistical significance. They will not understand sampling distributions and simulations.
A.IR.2	Students understand the role of p-values in determining statistical significance.	Students demonstrate an ability to interpret a p-value to make an inference in the context of a study. Students should also be able approximate p-values using simulation or simulation results, especially for the difference in two means or two proportions.	Possible work products might include a detailed description of a real study with reported p-values and then the students are asked to interpret the p-values within the context of the study. A simulated distribution (e.g., for a difference in two means or two proportions) could be provided so that students estimate and interpret the p-value.	Upper category responses will correctly interpret the p-value in context as well as be able to provide a complete explanation of what a p-value represents for this study. For example, if we repeated this study many times the p-value represents the chance of observing a difference as extreme, or more extreme, than the one we actually observed.	Lower category responses will interpret the p-value in relation to some stated or assumed significance level, but will not be able to interpret the p-value within context. A common mistake is to interpret the p-value as a probability for the current set of data.
A.IR.3	Students interpret the margins of error and confidence intervals for an estimate of a population characteristic or the difference between two or more populations on that characteristic.	Students understand that there is sampling variability associated with summary statistics and uses the margin of error to form an interval (confidence interval) to estimate the characteristic. Students interpret the confidence interval(s) in context.	Possible work products might include the interpretation of margin of error and confidence intervals in context. Given an estimate and margin of error, students will form and interpret the confidence interval as capturing the true population characteristic a certain percentage of the time in repeated samples.	Upper category responses will reflect a correct interpretation of margin of error and confidence intervals in context. The correct interpretation relies on the distribution or repeated sampling.	Lower category responses will contain weak interpretations of the margin of error and confidence interval without making connections to context. For example an incorrect interpretation could be "the probability that the interval will capture the population characteristic is 0.95."