

8th CCSS D5 Statistics-Line of Fit Conceptual Foundation (2-3 weeks)

Domain 5: Statistics and Probability 8.SP

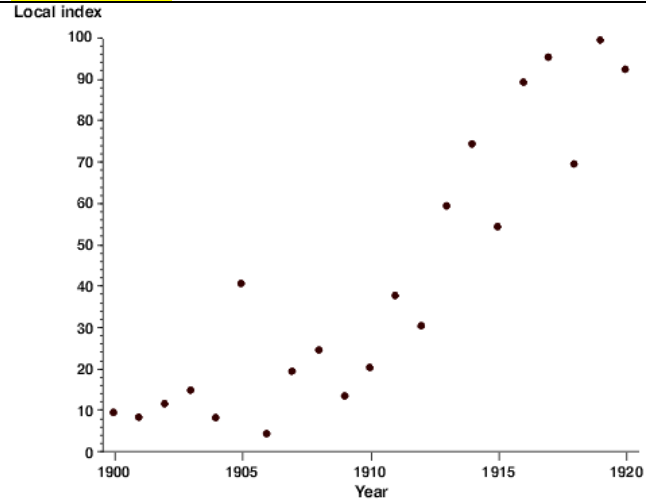
D5 Cluster 1: Investigate patterns of association in bivariate data.

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

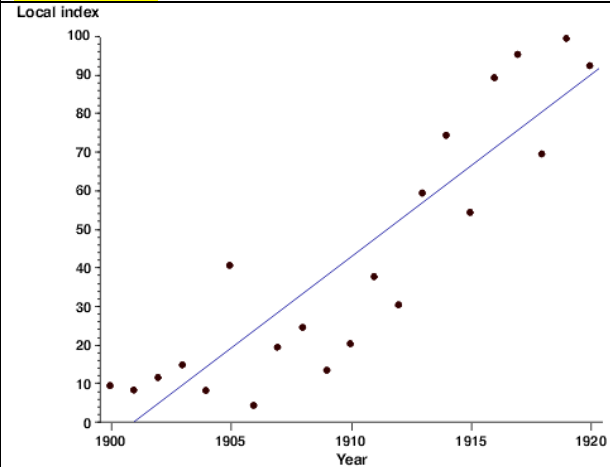
- 5.1.1 I can construct and interpret scatter plots. I can describe the relationships shown in a scatter-plot (identifying patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.).
- 5.1.2 I can sketch a line of best fit on a scatter plot, justify the location of the line; and explain why or why not a given line is a good fit.
- 5.1.3 I can write the equation of a line of best fit and use it to make predictions.
- 5.1.4 I can explain what the slope and y-intercept mean in terms of the situation.
- 5.1.5 I can construct two-way frequency and relative frequency tables to summarize bivariate categorical data (two variables collected from the same subjects).
- 5.1.6 From two-way frequency and relative frequency tables, I can describe, interpret, and justify inferences in patterns of association between two variables.

Scatter Plots, Line of Fit, Patterns of Association

Scatter Plot

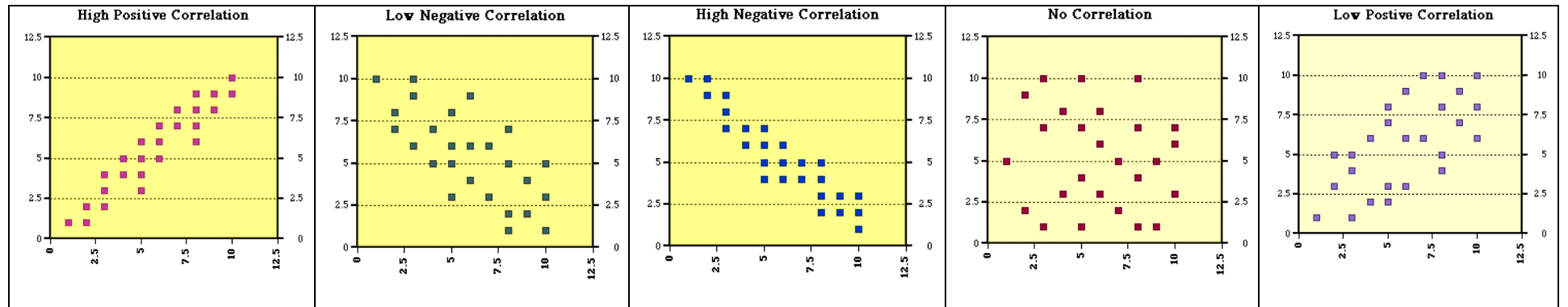


Line of Fit



Patterns of Association (clustering, outliers, positive/negative association, linear/nonlinear association)

- The line of best that rises quickly from left to right is called a **positive correlation**.
- The line of best that falls down quickly from left to the right is called a **negative correlation**.
- **Strong** positive and negative correlations have data points very close to the line of best fit.
- **Weak** positive and negative correlations have data points that are not clustered near or on the line of best fit.
- Data points that are not close to the line of best fit are called **outliers**.
- When data points cluster together the group of data points which are close together are called **clusters or clustering**.
- **Linear/nonlinear association** relates to whether or not the data points fall in a proportional linear pattern or otherwise.



TWO-WAY FREQUENCY TABLES to display BIVARIATE CATEGORICAL DATA

A two-way frequency table is a useful tool for examining relationships between categorical variables. The entries in the cells of a two-way table can be *frequency counts* or *relative frequencies*. The table summarizes and shows how often a value occurs.

The two-way frequency table below shows the favorite leisure activities for 50 adults – 20 men and 30 women using frequency counts.

	Dance	Sports	Movies	TOTAL
Women	16	6	8	30
Men	2	10	8	20
TOTAL	18	16	16	50

The two-way frequency table below shows the favorite leisure activities for 50 adults – 20 men and 30 women using relative frequencies. Generally *relative frequencies* are written as a decimal or percentage. It is the ratio of the actual number of favorable events to the total possible number of events; often taken as an estimate of probability. You can use the relative frequency to determine how often a value may occur in the future.

	Dance	Sports	Movies	TOTAL
Women	0.32	0.12	0.16	0.60
Men	0.04	0.20	0.16	0.40
TOTAL	0.36	0.32	0.32	1.00

Relative Frequency of Table

Two-way tables can show relative frequencies for the whole table, for rows or for columns. The table above shows relative frequencies for the whole table. The tables below show relative frequencies for rows and the relative frequencies for columns.

	Dance	Sports	Movies	TOTAL
Women	0.53	0.20	0.27	1.00
Men	0.10	0.50	0.40	1.00
TOTAL	0.36	0.32	0.32	1.00

Relative Frequency of Row

	Dance	Sports	Movies	TOTAL
Women	0.89	0.38	0.50	0.60
Men	0.11	0.62	0.50	0.40
TOTAL	1.00	1.00	1.00	1.00

Relative Frequency of Column

Each type of relative frequency table makes a different contribution to understanding the relationship between gender and preferences for leisure activities. For example, “Relative Frequency for Rows” table most clearly shows the probability that each gender will prefer a particular leisure activity. For instance, it is easy to see that the probability that a man will prefer movies is 40%; the probability that a woman will prefer movies is 27%; and so on.

Describe, interpret, and justify inferences in PATTERNS OF ASSOCIATION between two variables

NOTE: It is highly recommended that students collect and interpret real data.

1. Collect real data. CONTEXT is important—interpreting data and graphs relies on real-life/-world situations.
2. Start with a question. (How two variables are related—scatter plots. The frequency of an event—frequency table.)
3. Plan the data collection. Collect the data.
4. Organize the data into **scatter plots** (line of fit as appropriate) or **frequency tables**.
5. Interpret the data and the plots using the proper tools and vocabulary (patterns of association).
6. Draw conclusions about the data and the population as related to the sample.

7. Observe and question their peers' data, plots, and conclusions.

Scatter Plot/Line of Best Fit data collection: Several good lessons are included to address data collection and organization. <http://middlemathccss.wordpress.com/8th-grade-math/8th-d5-statistics/constructinterpret-scatter-plots-describe-relationships-clustering-outliers-increasingdecreasinglinearnonlinear-associations/>

Two-way Frequency/Relative-frequency Table data collection examples:

1. Collect data from students in your class on whether or not they have a curfew on school nights. Collect data from students on whether or not they have assigned chores at home. Is there evidence that those who have a curfew also tend to have chores?
2. How many miles to get to school? How many tardies? Is there evidence of correlation?
3. How many siblings? How many siblings do parents have? Do parents with lots of siblings have more children?

Related Interactive sites:

<http://www.shodor.org/interactivate/activities/ScatterPlot/> (interactive)

<http://staff.argyll.epsb.ca/jreed/math9/strand4/scatterPlot.htm> (interactive)

Statistical Questions Students Should Be Able to Answer

6th Grade: (6th—Histograms, Dot Plots, Box Plots, Frequency Tables (one-variable numerical))

- What was the statistical question asked?
- How many observations were made or people were surveyed to gather the data?
- What was measured or counted by the statistical question?
- How was the attribute measured or counted?
- What unit of measurement is being used to describe the data?
- What is the shape of the data?
- Where are there clusters of data or gaps in the data?
- Are there any outliers in the data?
- Is mean or median a better descriptor for the measure of center for a particular set of data?
- What is the mean absolute deviation of the data points in the set?
- What is the range of the data?
- What is the lower quartile (Q1) or median of the lower half of the data?
- What is the upper quartile (Q3) or median of the upper half of the data?
- What can be observed about the statistical question asked?

7th Grade—same questions as 6th grade, but comparing two populations (7th—Dot/Box Plots)

- What are the measures of variation found in the data? (range, quartiles, interquartile range)
- Describe the visual overlap of the two populations? What can we learn about the populations from the visual overlap? How are they related or not related?
- What can be observed about the statistical question asked?
- What are the patterns of association
- What are our conclusions?

8th Grade—most of the questions from 6th-7th are appropriate for 8th concepts also (8th—Scatter Plots/Line of Best Fit, Frequency Tables)

Scatter Plots/Line of Best Fit

- What is the relationship between the two variables?
- What are the patterns of association shown on the graph of the data? (clustering, outliers, positive/negative association, linear/nonlinear association)
- What is the rate of change in the line of fit?
- What is the y intercept and what does it mean?
- What predictions might be made from the data and graph?

Frequency Tables (the frequency tables in 6th were one variable—numerical—for dot, histograms, box plots)

- What are the two variables
- What are the patterns of association between the two variables?
- What inferences might be made from observing the patterns of association between the two variables?
- How can you and justify the inferences made from the data and graphs?