Algebra 1

ELG HS.S.2: Summarize, represent, and interpret data on two categorical and quantitative variables

Vertical Progression:

7th Grade	**Draw informal comparative inferences about two populations.** ○ **7.SP.4** Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.
8th Grade	**8.SP.A Investigate patterns of association in bivariate data** ○ **8.SP.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. ○ **8.SP.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. ○ **8.SP.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.
Algebra 1	**ELG.MA.HS.S.2: Summarize, represent, and interpret data on two categorical and quantitative variables.** ○ **S-ID.5** Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data. ○ **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.6b** Informally assess the fit of a function by plotting and analyzing residuals. ○ **S-ID.6c** Fit a linear function for a scatter plot that suggests a linear association.
Algebra 2	**ELG.MA.HS.S.2 Summarize, represent, and interpret data on two categorical and quantitative variables** ○ **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.*

Students will demonstrate command of the ELG by:

- Calculating joint, marginal, and conditional relative frequency.
- Interpreting data, recognizing associations and trends within two-way frequency tables.
- Reading, interpreting, and plotting data using a scatter plot.
- Describing how data is related.
- Determining if a set of data can be represented by a linear or exponential function and finding either a linear or exponential function to fit the data.
- Assessing the fit of a function by analyzing residuals.

Vocabulary:

- conditional relative frequency
- exponential function
- joint frequency
- line of fit
- linear function
- marginal frequency
- residuals
- scatter plot
- two-way frequency table

Sample Instructional/Assessment Tasks:

1) **Standard(s): ELG.MA.HS.S.2 (S-ID.5)**

Longer School Days?

Source: Illustrated Mathematics
https://www.illustrativemathematics.org/content-standards/HSS/ID/B/5/tasks/2044

Item Prompt:

Each student in a random sample of students at a local high school was categorized according to gender (male or female) and whether they supported a proposal to increase the length of the school day by 30 minutes (oppose, support, no opinion). The following table summarizes the data for this sample.

Opinion on Proposal to Increase Length of School Day					
		Oppose	Support	No Opinion	Total
Gender	Male	50	40	20	110
	Female	40	40	10	90
	Total	90	80	30	200

a. What proportion of the students in this sample are male?
b. What proportion of the students in this sample support the proposal?
c. What proportion of the males in this sample support the proposal?
d. What proportion of the students in this sample who support this proposal are male?
e. Interpret the following joint relative frequency in the context of this problem: 10/200.
f. Interpret the following marginal relative frequency in the context of this problem: 30/200.
g. Interpret the following conditional frequency in the context of this problem: 50/110.
h. Interpret the following conditional frequency in the context of this problem: 20/110.
i. Interpret the following conditional frequency in the context of this problem: 20/30.

Correct Answer:

- **a.** $\frac{number\ of\ males}{total\ number\ of\ students} = \frac{110}{200} = 0.55$
- **b.** $\frac{number\ who\ support\ proposal}{total\ number\ of\ students} = \frac{80}{200} = 0.40$
- **c.** $\frac{number\ who\ support\ porposal}{total\ number\ of\ males} = \frac{40}{110} = 0.39$
- **d.** $\frac{number\ of\ males\ who\ support\ proposal}{number\ who\ support\ proposal} = \frac{40}{80} = 0.50$
- **e.** This is the proportion of students in the sample who are female and who have no opinion on the proposal.
- **f.** This is the proportion of students in the sample who have no opinion on the proposal.
- **g.** This is the proportion of males in the sample who oppose the proposal.
- **h.** This is the proportion of males in the sample who have no opinion on the proposal.
- **i.** This is the proportion of students in the sample who have no opinion on the proposal who are male.

2) **Standard(s): ELG.MA.HS.S.2 (S-ID.6)**

Used Subarus

Source: Illustrated Mathematics
https://www.illustrativemathematics.org/content-standards/HSS/ID/B/6/tasks/941

Item Prompt:

Jane wants to sell her Subaru Forester, but doesn't know what the listing price should be. She checks on craigslist.com and finds 22 Subarus listed. The table below shows age (in years), mileage (in miles), and listed price (in dollars) for these 22 Subarus. (Collected on June 6th, 2012 for the San Francisco Bay Area.)

Age	Mileage	Price	Age	Mileage	Price	Age	Mileage	Price	Age	Mileage	Price
8	109428	12995	10	67740	9888	10	161460	5995	3	30047	20850
5	84804	14588	11	97500	6950	4	68075	12999	8	107506	11988
3	55321	20994	6	36967	19700	3	30007	22900	11	89207	8995
3	57474	18991	12	148000	3995	8	66000	13995	13	141235	5977
1	11696	19981	2	29836	18990	10	93450	8488			
13	125260	6888	3	32349	21995	3	35518	22995			

a. Make appropriate plots with well-labeled axes that would allow you to see if there is a relationship between price and age and between price and mileage. Describe the direction, strength and form of the relationships that you observe. Does either mileage or age seem to be a good predictor of price?

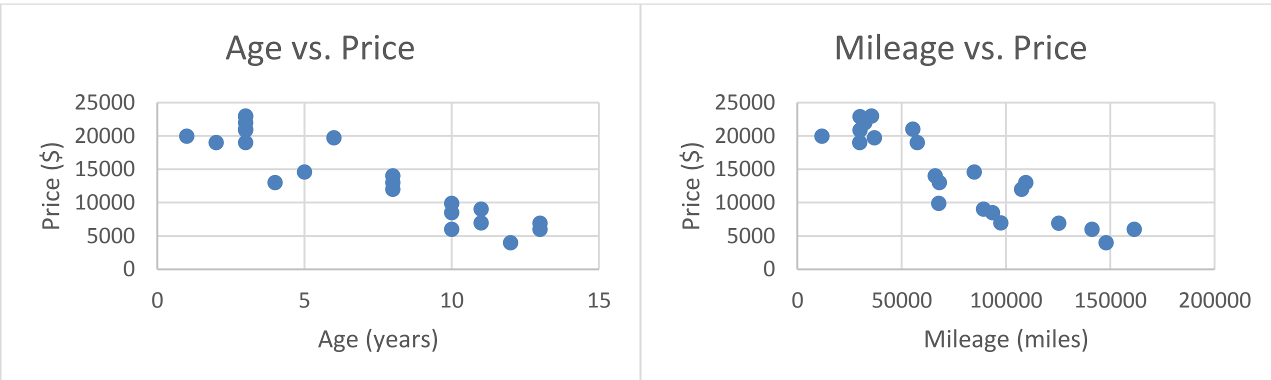
b. If appropriate, describe the strength of each relationship using the correlation coefficient. Do the values of the correlation coefficients agree with what you see in the plots?

c. Pick the stronger relationship and use the data to find an equation that describes this relationship. Make a residual plot and determine if the model you chose is a good one. Write a few sentences explaining why (or why not) the model you chose is appropriate.

d. If Jane's car is 9 years old with 95000 miles on it, what listing price would you suggest? Explain how you arrived at this price.

Correct Answer:

Since all variables are quantitative, we can make two scatterplots. It makes sense to choose price as the response variable in both scatterplots, with age and mileage as the explanatory variable.

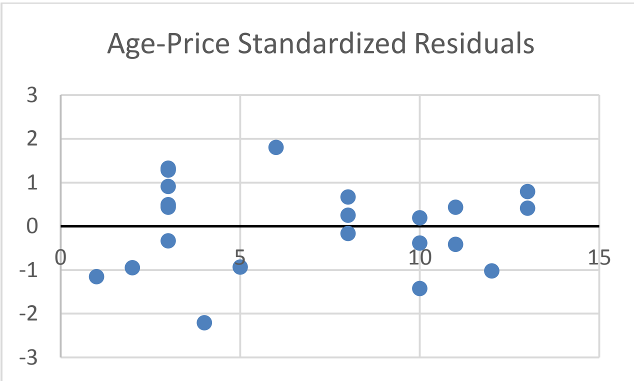


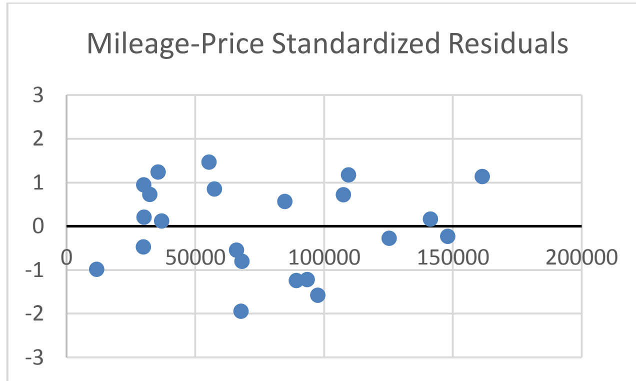
Both scatterplots show a strong negative linear trend. As age increases, price tends to decrease. As mileage increases, price tends to decrease. Both age and mileage seem to be good predictors of price. Since the scatterplots show linear relationships, it is appropriate to find the values of the correlation coefficients. Since we see a strong negative linear relationship we expect them to be negative and close to -1. Using technology, we find the value of the correlation coefficient for age and price to be -0.9251. The value of the correlation coefficient for mileage and price is -0.8995.

The relationship between price and age is slightly stronger than the relationship between price and mileage, so this solution gives the line of best fit equation for price vs. age. But since the values of the correlation coefficients are very close, the student could choose either age or mileage as a predictor. Since the scatterplot looks linear and the correlation is strong, we can find linear models: (let p = price, a = age, m = mileage)

$p = -1482.1a + 24248$ and $p = -0.1327m + 24266$

We can also use technology to find the residual plots.





Residual plots show us where the model has overestimated the values of cars (where residuals are negative) and where the model has underestimated the values of cars (positive residuals). If the pattern is truly linear, the residual plot should show no pattern and should be a random scattering of points. We see an example of that here. Since the original scatterplot of price vs. age looks straight, since the correlation is strong and since the residual plot shows no obvious pattern, our linear model is appropriate and can be used to make predictions.

Since we determined that age is a slightly better predictor than mileage, we use the linear model for predicting price based on age found in part (c). Plugging in Jane's information using age=9, we compute her predicted list price to be approximately $10,909.

3) **Standard: S**-ID.A.6

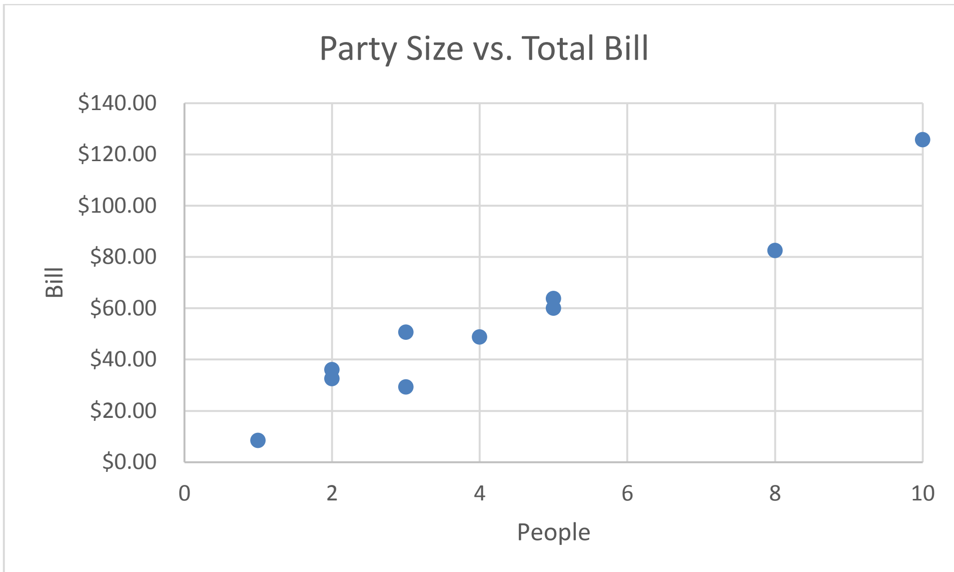
Restaurant Bill and Party Size

Source: Illustrative Mathematics
https://www.illustrativemathematics.org/content-standards/HSS/ID/B/6/tasks/2046)

Parts a-g are recommended. Parts h and i are included as possible extensions.

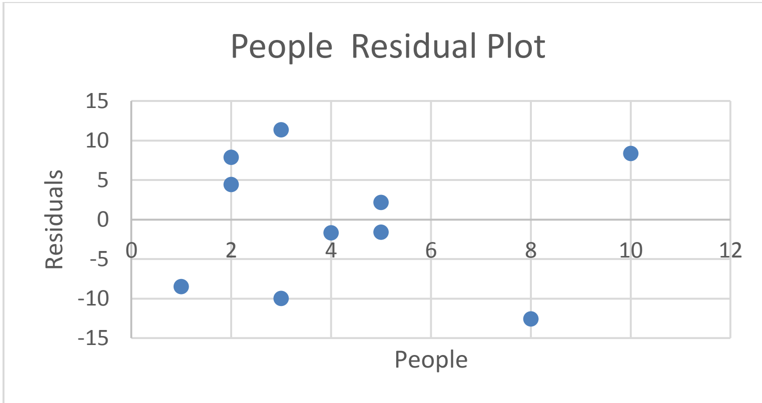
The owner of a local restaurant selected a random sample of dinner tables at his restaurant. For each table, the owner recorded the total amount of the dinner bill and the number of people at the table. The data are given in the table below. A scatterplot of the data is also shown.

People	Bill
1	$8.50
2	$36.00
2	$32.55
3	$29.30
3	$50.65
4	$48.75
5	$63.75
5	$60.00
8	$82.50
10	$125.75



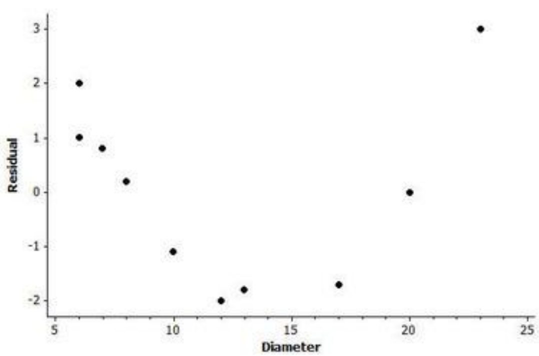
a) Does the relationship between number of people and the bill appear to be weak, moderate or strong?
b) Does the relationship between number of people and the bill appear to be linear?
c) The equation of the line of best fit is Bill = 5.80 + 11.15 · People. Sketch this line on the scatterplot.
d) Interpret the slope of the line of best fit in the context of this problem.
e) Note that the points in the scatterplot do not all lie on the line. What could explain this variability?
f) Use the equation of the line of best fit to estimate how much would you predict the bill to be for a party of 8?
g) There was one party of 8 whose bill was $82.50. Did they pay more than or less than the predicted amount? How much more or how much less did they pay?

The difference calculated in part (g) is called a residual. When each residual is calculated and plotted against the corresponding number of people, the resulting plot is called a residual plot. The residual plot for this data set is shown below.



h) Explain how one could use the residual plot in order to determine if a linear model is an appropriate way to describe the relationship between number of people and the amount of the bill.

A random sample of several pizza restaurants in the area was selected. Data was collected on the diameter and the price of the smallest cheese pizza sold at the restaurant. The line of best fit was fit to the data and the corresponding residual plot is given below.



i) Does there appear to be a linear relationship between the diameter and the price of a cheese pizza in this area? Use the residual plot to explain your answer.