PROJECT-SET
Statistics Education for Teachers

Gas Chromatography

Grading Rubric with Answers

Task adopted from: http://lib.stat.cmu.edu/DASL/Datafiles/Chromatography.html] aligned with standard S-ID.6

Written task aligned with Reg_Loop 2

Grade each question below using the following three grades: Essentially Correct (E), Partially Correct (P), and Incorrect (I).

Example answers for each category are given below each question.

Once each of the questions is graded in this manner, the results will then be tallied to give an overall score for the task. For each E, a person will receive 1 point. For each P, a person will receive ½ point. For each I, a person will receive 0 points. The overall grade will be one of the following:

4 Complete Response (EEE)

3/3.5 Substantial Response (EEP, EPP, EEI,)

2/2.5 Developing Response (EPI, PPP, EII or PPI)

1/1.5 Minimal Response (PII or III)

When grading, please write the amount of points earned and the word to describe the points earned (complete, substantial, developing, or minimal) on the front page of each person's paper.

Gas Chromatography

Gas chromatography is a technique used to detect very small amounts of a substance. Lab technicians are interested in investigating whether a specific chromatograph was measuring accurately. To study the calibration of the gas chromatograph, five measurements were taken for each of four specimens containing different, but known amounts of the substance being studied. The purpose of the study is to calibrate the chromatograph by relating the actual amount of the substance to the chromatograph reading. Is the amount of the substance related to the chromatograph reading? If so, what is an appropriate model for the relationship?

The following table presents the measurement data:

Amount	Response
0.25	6.55
0.25	7.98
0.25	6.54
0.25	6.37
0.25	7.96
1	29.7
1	30
1	30.1
1	29.5
1	29.1
5	211
5	204
5	212
5	213
5	205
20	929
20	905
20	922
20	928
20	919

1. Find the sample linear regression equation using technology that models the relationship between gas amount and reading.

Response = 46.623(Amount) – 14.411

Grading Category	Solutions, Explanations, and Sample Answers
Essentially Correct (E)	An essentially correct answer will give the correct equation of the linear regression line.
Partially Correct (P)	A partially correct answer will give some indication of having fit a line to the data, but give the wrong equation.
Incorrect (I)	An incorrect answer fails to meet the criteria of E and P.

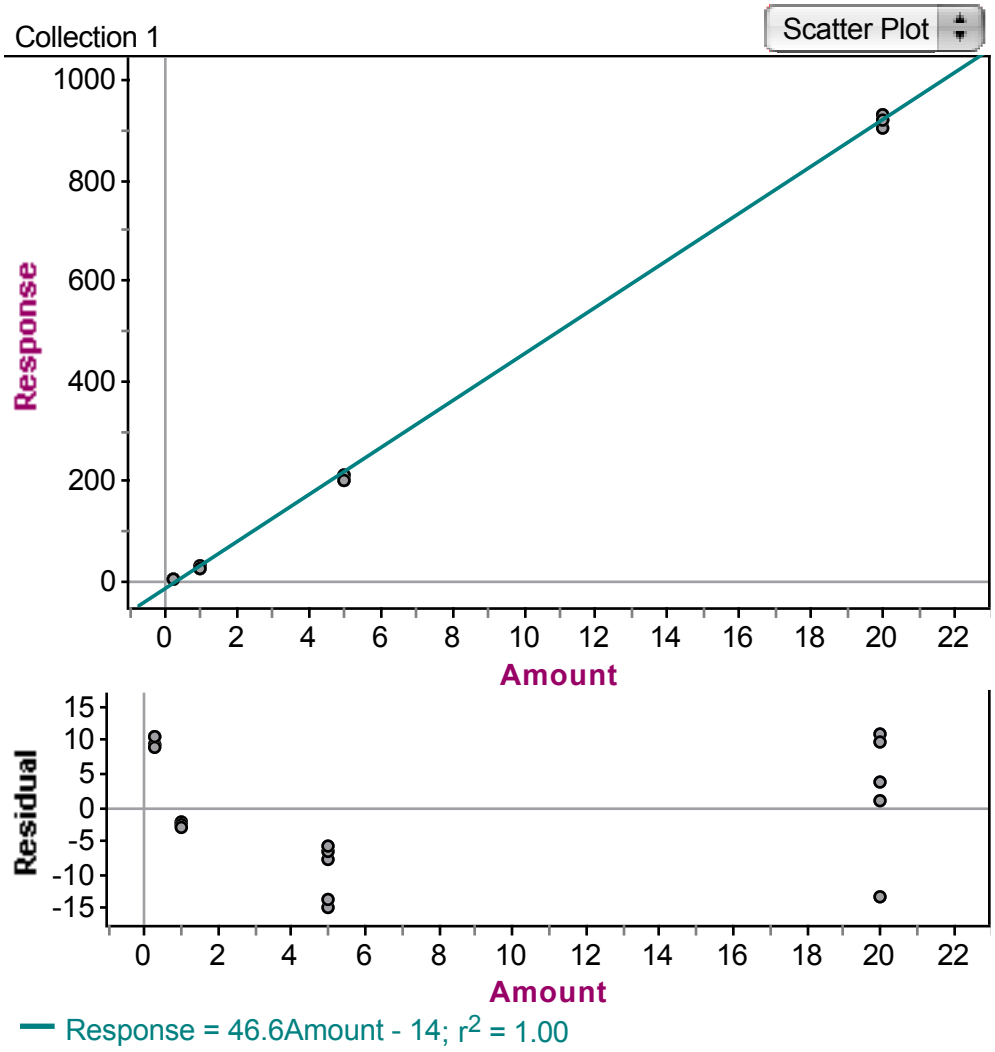
2. a) Using technology, show the residual plot.
   b) Upon studying the residual plot, do you have any concerns about the simple linear regression line as a model for the performance of the chromatograph?
   c) What is the sum of the residuals? Is this always the case? Why or why not?

Grading Category	Solutions, Explanations, and Sample Answers
Essentially Correct (E)	An essentially correct answer will answer all three questions correctly. For (a), a correct

	answer must show the residual plot. For (b), the correct answer must mention that the residuals are not evenly spread out above and below the 0 line. It must also illustrate an understanding on what the residuals tell us about overestimation and underestimation. For (c), the correct answer must state that the residuals always sum to 0. This happens by construction.
Partially Correct (P)	A partially correct answer will answer two of the three questions correctly.
Incorrect (I)	An incorrect answer fails to meet the criteria of E and P.

(b) It appears that for larger amounts of gas, the estimated model tends to underestimate the reading (only one observation of the 20s is below the regression line). At 5, the opposite is true – the regression tends to overestimate the amount. Ideally, we would like to see points scattered both above and below the 0 line on the residual plot in an even manner. From this plot, it appears that the model is not necessarily a good fit for this range of gas data. Although, the plot does illustrate that the residual values are relatively small (between 15 and-15). This seems small according to the overall readings in the data.

(c) The sum of the residuals is zero. This is always the case because the Ordinary Least Squares method of estimating the line of best fit minimizes the sum of the squared residuals in order to place the line. To do so, we can take the partial derivatives with respect to both alpha and beta and set them equal to zero. This minimization problem consequently makes it so the residuals sum to zero.



3. (a) Compute the average amount of gas ($\bar{x}$) and compute the average amount of reading ($\bar{y}$). Plot this point on the scatterplot. Where does this value lie in relation to the regression line?
(b) Describe how the technology estimates the regression line for these data. What does least-squares regression do in order to estimate the model?

(a) The average amount of gas is 6.5625 and the average reading for these data is 291.59. This point is on the regression line.
(b) The least-squares regression line minimizes the sum of squared residuals.

Grading Category	**Solutions, Explanations, and Sample Answers**
Essentially Correct (E)	An essentially correct answer will answer both (a) and (b) questions correctly. For (a), a correct answer must state and show that the average point is always on the regression line. For (b), the correct answer

	must mention that the regression line is found by minimizing the sum of the squared residuals.
Partially Correct (P)	A partially correct answer will answer one of the two questions correctly.
Incorrect (I)	An incorrect answer fails to meet the criteria of E and P.

4. Based on just the scatterplot, the estimated regression line, and the answers you gave above, does the linear model appear to provide a good description of the performance of the chromatograph?

   From looking at the scatterplot, it appears that the model provides a good description of the performance of the chromatograph. The points at every value of gas amounts appear to be tightly clustered around the line. Looking at the residual plot reveals something a bit different. For the specific values, the line either overestimates or underestimates the actual values. In general though, the residuals appear to be small so the underestimation and overestimation are not as much of a concern. Overall, the model does appear to provide a good description of the performance.

Grading Category	**Solutions, Explanations, and Sample Answers**
Essentially Correct (E)	An essentially correct answer will provide a well-supported case for whether the person believes it is a good model. Arguments for not being a good model include the answer about the residual given in #3. However, one may note that overall the residuals are very small.
Partially Correct (P)	A partially correct answer will provide some sound reason for why they state the model is either good or bad. The reason, however, may not be complete and only discuss one aspect of the data.
Incorrect (I)	An incorrect answer fails to meet the criteria of E and P.