

A Statistical Analysis of Annual Gas Prices from 1976-2005

Level/Course:

This lesson can be used and modified for teaching High School Math, Foundations of Algebra, Algebra I, Algebra II, Discrete Math, and Advanced Functions and Modeling.

Part I: Mean, Median, Mode, & Range

Objective

- Calculate, use, and interpret the mean, median, mode, range for a set of data.

Activity

Have the students find the mean yearly gas prices from 1976-2005 and write their answers in the column named "Annual Mean." Have the students do the attached worksheet titled "Mean, Median, Mode, & Range." You can also have the students create various lists by looking at certain months or a select group of years and analyze the mean or median of each individual month or the median of each year. Teach the students how to use the "SortA(" and "SortD(" functions on the calculator to rearrange the data in either ascending or descending order so that they can more easily find the range and modes.

Assessment

Students will be graded not only on mathematical computations but also on their analysis of interpretation questions written in short answer form.

Part II: Scatter Plots & Linear Regression

Objective

- Collect, organize, analyze, and display data (including scatterplots) to solve problems.
- Approximate a line of best fit for a given scatterplot; explain the meaning of the line as it relates to the problem and make predictions.

Activity

Have students look back at their answers in the "Annual" box beside each year. If you did not do the Part I activity then you can give them Table 2 which includes the needed information. Have the students create a table in either a graphing calculator or a computer program so that when a scatter plot is graphed, the years are on the x-axis and the annual mean is on the y-axis. Instruct each student to look at the scatter plot and determine if they can visual see any pattern or a positive or negative regression line. Then have the students actually plot the linear regression line that best fits the data and determine the projected cost of gas will be in the year 2030 and explain why they feel this may or may not be a good prediction.

Assessment

Students will be graded on their data plots and on their linear regression line. They will also be graded on their use of this data to predict and analysis of results. Finally the students will be graded on their group project which will be presented to the class.

Part III: Box Plots

Objective

- Collect, organize, analyze, and display data (including box plots) to solve problems.
- Calculate, use, and interpret the inter-quartile range for a set of data.
- Identify outliers and determine their effect on the a set of data.

Activity

Take the list of average annual gas prices (either from Table 2 or the chart they received and filled out in Part I) and create a box plot. The students can then trace the box plot to find the minimum, quartile 1, median, quartile 3, and maximum values. From here, have the students find the interquartile range, find outliers and then analyze whether there is bad data and if the scatter plot produces an even distribution or if the graph is skewed.

Assessment

Students will be graded on finding the key points of a box plot graph and their written description of the analysis of results with respect to good and bad data. Finally, the students will be graded on their group project that will be presented to the class.

Part IV: Scatter Plot & Median-Median Line

Objective

- Collect, organize, analyze, and display data (including scatterplots) to solve problems.
- Find the median-median line for a given scatterplot; explain the meaning of the line as it relates to the problem and make predictions.
- Compare two different types of regression lines and which is more accurate for the data.

Activity

The students create a scatter plot putting the years on the x-axis and the annual means on the y-axis and find a median-median line which could be a more consistent regression line. Using the median-median line, they can find the residuals and the root mean square error of the data. They can then use this information to extrapolate data. They should also create a line of best fit and explore the differences between the line of best fit and the median-median line and determine which is actually a more accurate line for the data.

Assessment

Students will be graded on the median-median line and the standard deviation. They will also be graded on their analysis of what the data means and their comparison of the two types of linear regression lines and their prediction of possible future gas prices.

Mean, Median, Mode, & Range Activity

Instructions: Using the attached chart that has the monthly gas prices from 1976-2005, answer the following questions.

1. Look at the chart for the monthly gas averages from 1976-2005. Fill in the column titled Annual with the mean of each year's data.
2. Further Analysis of Mean, Median, Mode and Range.
 - a. Find the median of gas prices for the year of 1982. _____
 - b. Find the median of gas prices for the year of 2001. _____
 - c. Find the range of the Annual gas prices from 1976-2005. _____
 - d. Find the range from January 1976 to August 2005. _____
 - e. Find the mode from January 1976 to December 1979. _____
 - f. Find the mode from January 2000 to August 2005. _____

Short Answer: On another sheet of paper answer the following questions.

3. Compare the means and medians of the years 1982 and 2001. What are the differences and why is there a difference. Which would be better when trying to determine what gas prices were for those two years?
4. Why do you think there is such a big range from January, 1976, to August, 2005? What types of events and reasons could there be for price increases over the 30 year period?

Scatter Plot and Linear Regression Line Activity

Instructions: Using the years as your x-axis and the annual mean as your y-axis, create a scatter plot and linear regression line to answer the following questions.

1. What is the slope of the linear regression line? _____

2. What is the Y-Intercept of the linear regression line? _____

3. What is the equation of the linear regression line in slope-intercept form?

4. Further analysis of your scatter plot and linear regression line.

a. Based on your linear regression line, what would be an estimated cost of gas in 2030?

b. Do you think this will really be the price of gas? Why or why not?

GROUP PROJECT: In your assigned group, research an alternative energy source that can be used to replace gasoline. Be prepared to present to the class the estimated cost of producing and maintaining that form of alternative energy as well as the positives/negatives of using that form of alternative energy.

Box Plot Activity

Instructions: Using the years as your x-axis and the annual mean as your y-axis, create a box plot to answer the following questions about the quality of the data.

1. What is the maximum value (the highest price of gas)? _____
2. What is the minimum value (the lowest price of gas)? _____
3. What is the median gas price? _____
4. What is the cost of gas at quartile one? _____
5. What is the cost of gas at quartile three? _____
6. Is the graph skewed to one side or balanced? _____
7. What is the interquartile range? _____
8. How many years are outliers and what are they? _____

Short Answer:

9. Is there bad data in this scatter plot? Why or why not?

10. Are we living in an outlier year? If so do you think time will prove that this is an outlier year or will there be a new trend in the price of gasoline? Explain your answer.

GROUP PROJECT: In your assigned group find all the outliers and research what events were happening in the United States and the world to determine what was going on that could have led to the price of gas being that much higher or lower than what it should have been. Be prepared to present your findings to the class in a creative manner.

Scatter Plot & Median-Median Line Activity

Instructions: Using the years as your x-axis and the annual mean as your y-axis, create a scatter plot and a median-median line to answer the following questions.

1. What is the equation of the median-median line? _____
2. What are the residuals of each year from 1976-2005?

1976	1977	1978	1979	1980	1981
1982	1983	1984	1985	1986	1987
1988	1989	1990	1991	1992	1993
1994	1995	1996	1997	1998	1999
2000	2001	2002	2003	2004	2005

3. What is the root mean square error? _____
4. What would be the cost of gas in 2030? _____
5. Based on your information, what would the maximum and minimum gas price be in 2030?

Short Answer: If you did the Scatter Plot & Linear Regression Line Activity then pull out your worksheet. If you did not do that activity, find the equation for the line of regression (a best fit line) and answer the following questions. If you need more space finish answers on the back.

5. Compare the possible gas prices from 2030 on the two different linear regression lines. Which price do you think will be more accurate and why?

6. Do you think whichever equation you picked as the best equation (line of best fit or median-median line) for this problem will always be a better regression line for all data analysis? Why?

TABLE 1: AVERAGE GAS PRICES BY MONTH FROM 1976-2005

Table comes from the U.S. Department of Labor: Bureau of Labor Statistics web site: <http://www.bls.gov>

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual Mean
1976	0.605	0.600	0.594	0.592	0.600	0.616	0.623	0.628	0.630	0.629	0.629	0.626	
1977	0.627	0.637	0.643	0.651	0.659	0.665	0.667	0.667	0.666	0.665	0.664	0.665	
1978	0.648	0.647	0.647	0.649	0.655	0.663	0.674	0.682	0.688	0.690	0.695	0.705	
1979	0.716	0.730	0.755	0.802	0.844	0.901	0.949	0.988	1.020	1.028	1.041	1.065	
1980	1.131	1.207	1.252	1.264	1.266	1.269	1.271	1.267	1.257	1.250	1.250	1.258	
1981	1.298	1.382	1.417	1.412	1.400	1.391	1.382	1.376	1.376	1.371	1.369	1.365	
1982	1.358	1.334	1.284	1.225	1.237	1.309	1.331	1.323	1.307	1.295	1.283	1.260	
1983	1.230	1.187	1.152	1.215	1.259	1.277	1.288	1.285	1.274	1.255	1.241	1.231	
1984	1.216	1.209	1.210	1.227	1.236	1.229	1.212	1.196	1.203	1.209	1.207	1.193	
1985	1.148	1.131	1.159	1.205	1.231	1.241	1.242	1.229	1.216	1.204	1.207	1.208	
1986	1.194	1.120	0.981	0.888	0.923	0.955	0.890	0.843	0.860	0.831	0.821	0.823	
1987	0.862	0.905	0.912	0.934	0.941	0.958	0.971	0.995	0.990	0.976	0.976	0.961	
1988	0.933	0.913	0.904	0.930	0.955	0.955	0.967	0.987	0.974	0.957	0.949	0.930	
1989	0.918	0.926	0.940	1.065	1.119	1.114	1.092	1.057	1.029	1.027	0.999	0.980	
1990	1.042	1.037	1.023	1.044	1.061	1.088	1.084	1.190	1.294	1.378	1.377	1.354	
1991	1.247	1.143	1.082	1.104	1.156	1.160	1.127	1.140	1.143	1.122	1.134	1.123	
1992	1.073	1.054	1.058	1.079	1.136	1.179	1.174	1.158	1.158	1.154	1.159	1.136	
1993	1.117	1.108	1.098	1.112	1.129	1.130	1.109	1.097	1.085	1.127	1.113	1.070	
1994	1.043	1.051	1.045	1.064	1.080	1.106	1.136	1.182	1.177	1.152	1.163	1.143	
1995	1.129	1.120	1.115	1.140	1.200	1.226	1.195	1.164	1.148	1.127	1.101	1.101	
1996	1.129	1.124	1.162	1.251	1.323	1.299	1.272	1.240	1.234	1.227	1.250	1.260	
1997	1.261	1.255	1.235	1.231	1.226	1.229	1.205	1.253	1.277	1.242	1.213	1.177	
1998	1.131	1.082	1.041	1.052	1.092	1.094	1.079	1.052	1.033	1.042	1.028	0.986	
1999	0.972	0.955	0.991	1.177	1.178	1.148	1.189	1.255	1.280	1.274	1.264	1.298	
2000	1.301	1.369	1.541	1.506	1.498	1.617	1.593	1.510	1.582	1.559	1.555	1.489	
2001	1.472	1.484	1.447	1.564	1.729	1.640	1.482	1.427	1.531	1.362	1.263	1.131	
2002	1.139	1.130	1.241	1.407	1.421	1.404	1.412	1.423	1.422	1.449	1.448	1.394	
2003	1.473	1.641	1.748	1.659	1.542	1.514	1.524	1.628	1.728	1.603	1.535	1.494	
2004	1.592	1.672	1.766	1.833	2.009	2.041	1.939	1.898	1.891	2.029	2.010	1.882	
2005	1.823	1.918	2.065	2.283	2.216	2.176	2.316	2.503					

TABLE 2: 1976-2005 ANNUAL AVERAGE CHART

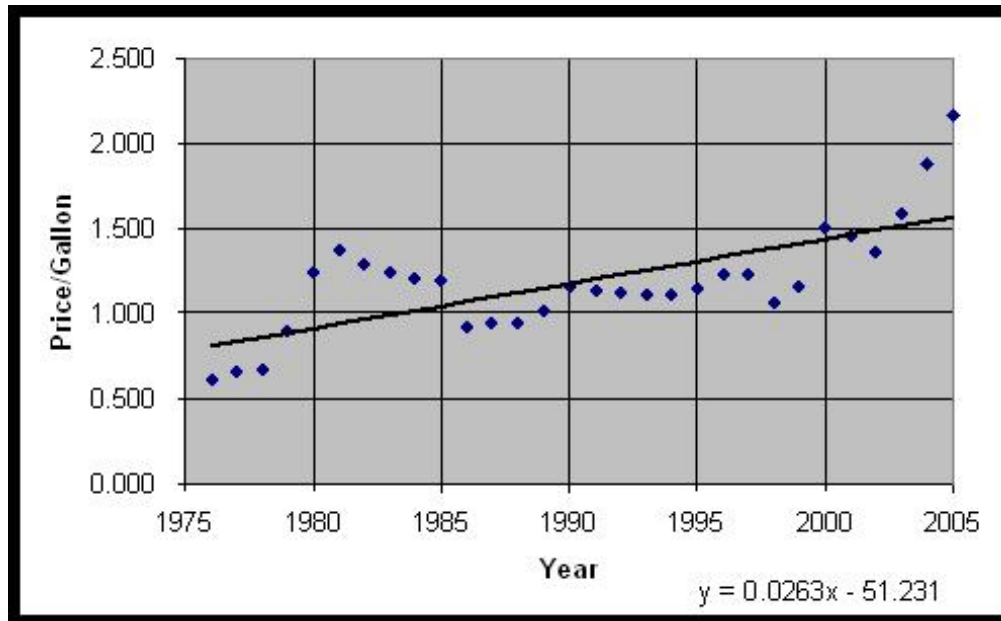
Year	Annual Mean
1976	0.614
1977	0.656
1978	0.670
1979	0.903
1980	1.245
1981	1.378
1982	1.296
1983	1.241
1984	1.212
1985	1.202
1986	0.927
1987	0.948
1988	0.946
1989	1.022
1990	1.164
1991	1.140
1992	1.127
1993	1.108
1994	1.112
1995	1.147
1996	1.231
1997	1.234
1998	1.059
1999	1.165
2000	1.510
2001	1.461
2002	1.358
2003	1.591
2004	1.880
2005	2.163

Teacher's Answer Key

Part I: Mean, Median, Mode, & Range

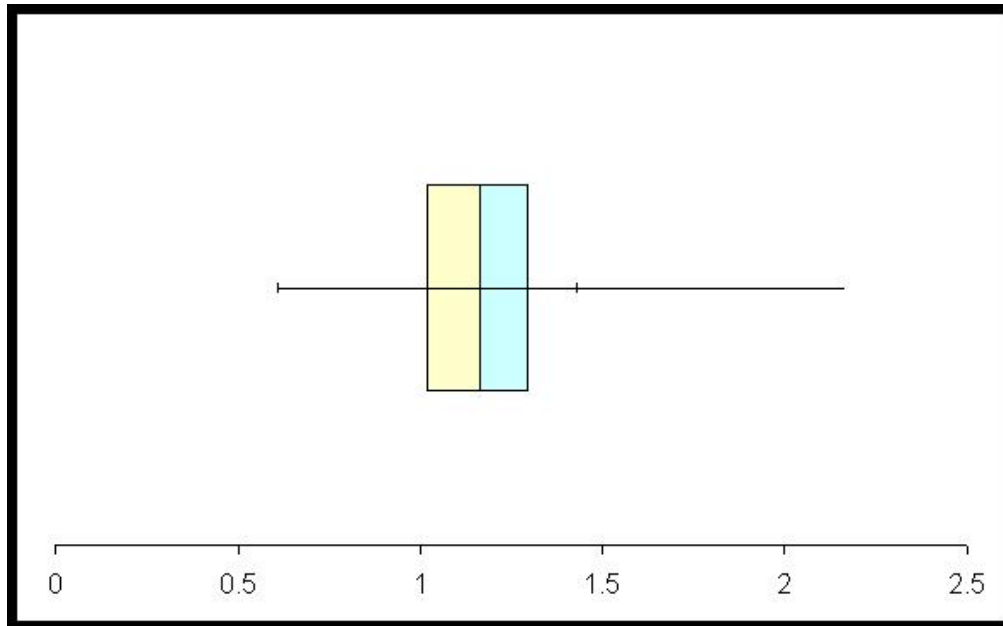
- 1) The answers for the chart can be found by looking at Table 2.
- 2)
 - a. 1.301
 - b. 1.477
 - c. 1.548
 - d. 1.911
 - e. 0.665
 - f. No Mode

Part II: Scatter Plots & Linear Regression



- 1) 0.0263
- 2) -51.231
- 3) $y = 0.0263x - 51.231$
- 4) a. 2.2306

Part III: Box Plots



- 1) 2.163
- 2) 0.614
- 3) 1.1645
- 4) 1.022
- 5) 1.296
- 6) Skewed Right
- 7) 0.411
- 8) 2004 and 2005

Part IV: Scatter Plot & Median-Median Line

- 1) $y = 0.010125x - 18.9116$
- 2)

1976	1977	1978	1979	1980	1981
-0.4814	-0.4485	-0.4457	-0.2228	0.1091	0.23198
1982	1983	1984	1985	1986	1987
0.13985	0.07473	0.0356	0.01548	-0.2697	-0.2588
1988	1989	1990	1991	1992	1993
-0.2709	-0.205	-0.0732	-0.1073	-0.1304	-0.1595
1994	1995	1996	1997	1998	1999
-0.1657	-0.1408	-0.0669	-0.074	-0.2592	-0.1633
2000	2001	2002	2003	2004	2005
0.1716	0.11248	-0.0007	0.22223	0.5011	0.77398

- 3) 0.279
- 4) about 1.6421
- 5) Maximum = 1.9211 and Minimum = 1.3631