

# Course: AP Statistics

## Topic: Correlation

Subject Area: Math

Grade Levels: 10-12

Dates: 9/10, 9/13, 9/17, 9/23, 9/24, 10/8

### Lesson Overview

Students will research through MLS online listings for used homes for sale in their neighborhood areas or visit new home developments for listings. Using this data, students will find and interpret the correlation between square footage and listed sale price of homes in a designated neighborhood. Students will present their findings in poster form to the rest of their classmates.

#### Materials Included in this Lesson

- home comparison price list worksheet
- notes on scatterplots and correlation
- making scatterplots using Excel

#### Other Materials for this Lesson

- graphing calculator
- internet access to MLS listings
- Microsoft Excel

#### Skills the Student will Learn

- make a scatterplot by hand and using Excel
- interpret scatterplot
- compute and interpret the correlation coefficient

#### Student Deliverables

- MLS listing for at least 10 homes in chosen neighborhood
- completed data collection worksheet
- poster report on findings for chosen neighborhood

### Length of Lesson: 6 Days

#### Activity Day One (9/10/10)

Lecture and student notes on drawing and interpreting scatterplots for direction, form and strength.

#### Activity Day Two (9/13/10)

Lecture and student notes on correlation for linear data.

#### Activity Day Three (9/17/10)

Quiz on scatterplots and correlation. Students will be given the project information and begin collecting data for the project.

#### Activity Day Four (9/23/10)

Students will learn how to make scatterplots on Excel or to download their plots from their graphing calculators into a MS Word document. Students will be given time to work on their project in groups in the computer lab.

## Activity Day Five (9/24/10)

Students will be given time to work on their project in groups in the computer lab.

## Activity Day Six (10/8/10)

Students will turn in their project and present their findings to the class.

## Enrichment Suggestions

Students can find the correlation between development fees for their corresponding neighborhoods and the desirability of the neighborhood.

## Student Resources

[www.mlslistings.com](http://www.mlslistings.com) or new housing developments in EG area

## CA State Standards Met

<http://www.cde.ca.gov/ci/ma/cf/documents/math-ch2-8-12.pdf>

AP Probability & Statistics – Math content standard

- 
- |             |  |
|-------------|--|
| <b>13.0</b> | Students know what the <i>correlation coefficient of two variables</i> means and are familiar with the coefficient's properties.   |
| <hr/>       |  |
| <b>14.0</b> | Students organize and describe distributions of data by using a number of different methods, including frequency tables, histograms, standard line graphs and bar graphs, stem-and-leaf displays, scatterplots, and box-and-whisker plots. |
- 

## CTE Standards Met

<http://www.cde.ca.gov/ci/ct/sf/documents/ctestandards.pdf>

### Communication

Students understand the principles of effective oral, written, and multimedia communication in a variety of formats and contexts.

#### *2.2 Writing*

- c. Make distinctions between the relative value and significance of specific data, facts, and ideas.
- d. Include visual aids by employing appropriate technology to organize and record information on charts, maps, and graphs.

### Pathways standards: E-commerce pathway

A3.1 Students understand the benefits of online communication channels, such as chat rooms, news groups, list serves, and message boards, as they pertain to online advertising.

***Is there a relationship between square footage versus listed price for \_\_\_\_\_  
homes in \_\_\_\_\_? \_\_\_\_\_***  
name of neighborhood new or used

Options:

- Decide on whether you want to use new or used homes (for new homes: go to the site on a weekend and pick up the brochures/flyers, for used homes: use [www.mlslistings.com](http://www.mlslistings.com))
- Decide on the particular neighborhood (start with a zip code and narrow it down using the map)

You and your partner will explore whether or not there is a relationship between two quantitative variables. You will then make a poster about your process and findings. Make it look neat and professional by typing your information and using computer or calculator generated graphs. In grading your project, attention will be given to the following considerations:

- Did you follow instructions? Did you use the correct procedure?
- Does your poster have all the required components?
- Are your conclusions reasonable and justified?
- Is the information clear and easy to read or follow?
- Is your writing grammatically correct?

Your final poster is due on **Friday, Oct. 8** at the beginning of class. Late projects will **not** be accepted for any reason since you can submit them electronically through email by 8 a.m. on Friday if needed. While working in the computer lab, be sure to “save” often. Use the following list of tasks to guide you through the project. Before submitting your report, use the grading sheet as a final “check list”.

### Tasks (Things to do):

1. Collect a sample of at least 10 pieces of bivariate data for the two variables listed in the title. {For example: the ordered pair (2.7, 296) would represent house that is 2700 square foot in size and is listed for \$296,000 where the units are in thousands.} When using [www.mlslistings.com](http://www.mlslistings.com), decide on the zip code. Once the listing comes up, click on the picture of the house. The size of the house in sf is listed after the “Interior” description.)
2. Briefly explain the question/purpose of your project.
3. Define your explanatory and response variable (with units).
4. Create & type a table with your data in it. Be sure to include column headings.
5. Calculate the descriptive statistics for *each* of the two sets of data, (the 5 number summary, mean and standard deviation) one for the  $x$  and one for the  $y$ . Comment on anything interesting that you notice.
6. Make a scatter plot of your data using Excel. Be sure to properly label the axes and title your graph.
7. Find and interpret the correlation coefficient  $r$ .
8. Find and interpret  $r^2$  in context.
9. Find the LSRL. Draw this line on your scatter plot (in Excel this is called a trend line).
10. Verify the equation of the LSRL by finding  $a$  and  $b$  using the formulas.
11. Plot and label the coordinate  $(\bar{x}, \bar{y})$  on your scatter plot from task #6. Remember that this point should be on your LSRL. (You may add this point to your scatter plot by hand using a pen.)
12. Use your LSRL to determine a reasonable prediction by selecting an  $x$  value of your choice.
13. How does this predicted value,  $\$$ , compare to the  $\bar{y}$ , the average response variable in your data, that you obtained in your descriptive statistics from task #5?

14. Explain what the value of the slope in your equation represents (in context).
15. Explain what the value of the  $y$ -intercept in your equation represents. Is it reasonable?
16. Calculate the residuals and present this information in a table.

| $x$ | $y$ | $\hat{y}$ | Residual = $y - \hat{y}$ |
|-----|-----|-----------|--------------------------|
|     |     |           |                          |

17. On your **scatter plot** of the data from task #6, illustrate and identify two of these residuals, one positive and one negative, using vertical line segments. (You may add these segments to your scatter plot by hand using a pen and straight edge.)
18. Make a residual plot using Excel. Be sure to properly label the axes.
19. Interpret your residual plot.
20. If a linear model is not appropriate, perform the necessary transformation and graphs. Discuss the result of the final transformation that best fits your data. (Note that each of the above tasks would now apply to your new transformed LSRL.)
21. You may include in your report any final comments about the effectiveness of the procedure and/or any flaws that you see (i.e. any changes you would make if you were to do this project again). Also, go to this link and fill out the quick survey.

<http://spreadsheets.google.com/viewform?hl=en&formkey=dFFUT2J3T1VOUmE0NHBkbzdrWTIFQVE6MA#gid=0>

### Bivariate Data Project – Grading Sheet (Use the following as your check list for your report)

Code:      4 = Excellent                      3 = Good                      2 = Satisfactory                      1 = Unsatisfactory

#### General:

- |   |   |   |   |   |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | Followed instructions. Used correct procedures.                                 |
| 1 | 2 | 3 | 4 | Report is grammatically correct, free of spelling errors, concise & unambiguous |
| 1 | 2 | 3 | 4 | Report is pleasing neat, easy to follow and pleasing to look at.                |

#### Specific Task:

- |          |   |   |   |   |   |
|----------|---|---|---|---|---|
| #4       | 1 | 2 | 3 | 4 | Data collected. Table made. Labeled column headings on table (in context).          |
| #5       | 1 | 2 | 3 | 4 | Five number summary, mean and standard deviation for <b>each</b> variable.          |
| #6       | 1 | 2 | 3 | 4 | Scatter plot (with proper labels & title)   |
| #7       | 1 | 2 | 3 | 4 | Correlation and interpretation (form, type, strength)                               |
| #8       | 1 | 2 | 3 | 4 | R-squared and interpretation  |
| #9       | 1 | 2 | 3 | 4 | Equation of LSRL and drawing  |
| #10      | 1 | 2 | 3 | 4 | Verify LSRL by using the formulas for $a$ and $b$ .                                 |
| #11      | 1 | 2 | 3 | 4 | $(\bar{x}, \bar{y})$ plotted and labeled on LSRL                                    |
| #12 & 13 | 1 | 2 | 3 | 4 | Prediction for selected $x$ variable and comparison of predicted value to $\bar{y}$ |
| #14      | 1 | 2 | 3 | 4 | Interpret slope in context  |
| #15      | 1 | 2 | 3 | 4 | Interpret $y$ -intercept and comment on reasonableness                              |
| #16      | 1 | 2 | 3 | 4 | Table with $x$ , $y$ , $\hat{y}$ , and residuals                                    |
| #17      | 1 | 2 | 3 | 4 | Residual line segments on scatter plot (draw & identify a + and a – residual)       |
| #18      | 1 | 2 | 3 | 4 | Residual plot (with proper labels)  |
| #19      | 1 | 2 | 3 | 4 | Interpret residual plot   |
| #20      | 1 | 2 | 3 | 4 | Transformation (if necessary)   |
| #21      | 1 | 2 | 3 | 4 | Final comments & completed online survey  |

**Bivariate Data Project – Grading Sheet**

Code:      4 = Excellent                      3 = Good                      2 = Satisfactory                      1 = Unsatisfactory

**General:**

- 1 2 3 4    Followed instructions. Used correct procedures.  
1 2 3 4    Report is grammatically correct, free of spelling errors, concise & unambiguous  
1 2 3 4    Report is pleasing neat, easy to follow and pleasing to look at.

**Specific Task:**

- #4            1 2 3 4    Data collected. Table made. Labeled column headings on table (in context).  
#5            1 2 3 4    Five number summary, mean and standard deviation for **each** variable.  
#6            1 2 3 4    Scatter plot (with proper labels & title)  
#7            1 2 3 4    Correlation and interpretation (form, type, strength)  
#8            1 2 3 4    R-squared and interpretation  
#9            1 2 3 4    Equation of LSRL and drawing  
#10           1 2 3 4    Verify LSRL by using the formulas for  $a$  and  $b$ .  
#11           1 2 3 4     $(\bar{x}, \bar{y})$  plotted and labeled on LSRL  
#12 & 13    1 2 3 4    Prediction for selected  $x$  variable and comparison of predicted value to  $\bar{y}$   
#14           1 2 3 4    Interpret slope in context  
#15           1 2 3 4    Interpret  $y$ -intercept and comment on reasonableness  
#16           1 2 3 4    Table with  $x$ ,  $y$ ,  $\hat{y}$ , and residuals  
#17           1 2 3 4    Residual line segments on scatter plot (draw & identify a + and a – residual)  
#18           1 2 3 4    Residual plot (with proper labels)  
#19           1 2 3 4    Interpret residual plot  
#20           1 2 3 4    Transformation (if necessary)  
#21           1 2 3 4    Final comments & completed online survey.

**Final Score:**             
**80**

*Comments:*

