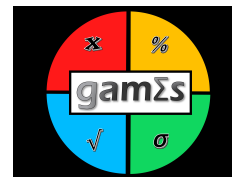


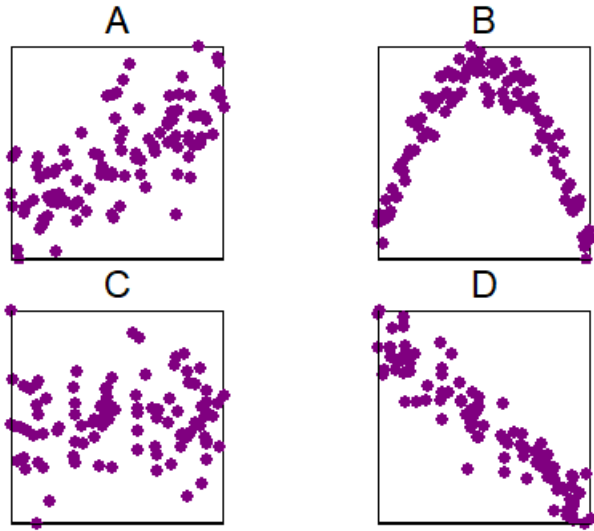
GAMES Practice Problems –Correlation and Simple Linear Regression



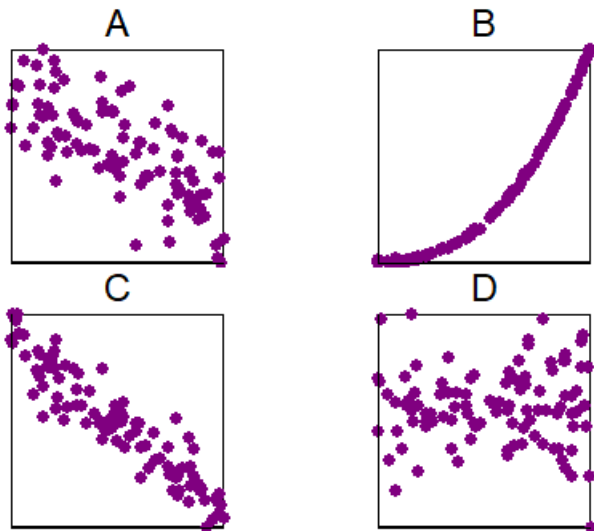
- For the following examples, describe the correlation between the variables in as much detail as possible. Which variable would you use as the independent variable which variable would be the dependent variable?
 - Oranges: weight in kilograms, weight in pounds
 - Graduate students: shoe size, course grades
 - Gasoline: the number of kilometers driven since the gas tank was filled and liters remaining in the gas tank.
- Consider the following daily data on number of employees and sales.

Employees	Sales
5	\$110,000
5	\$120,000
5	\$140,000
6	\$140,000
8	\$150,000
14	\$200,000
15	\$240,000
16	\$230,000
17	\$240,000
20	\$250,000

- Prepare a scatterplot of employees against sales.
 - Describe the direction of the association.
 - Describe the form of the underlying relationship.
 - Describe the strength of the underlying relationship?
 - Describe if outliers are present.
- If you had data for each pair of variables below and made a scatterplot, which variable would be the independent variable and which would be the response variable? Explain your decision. Describe what you would expect to see in a scatterplot. Discuss the likely direction, underlying relationship, and strength.
 - Climbing amount: altitude, temperature
 - Daily: ice cream sales and air conditioner sales
 - Humans: years of age, benchpress strength
 - Drivers: blood alcohol level, ability to react
 - Consider the following scatterplot diagrams. Identify which diagram has a correlation of $\{-.94, 0.06, 0.70, -0.09\}$



5. Consider the following scatterplot diagrams. Identify which diagram has a correlation of $\{0.95, -0.94, -0.71, -0.01\}$



6. A candidate for political office in Canada states that there is a correlation between committing a violent crime and playing violent video games. Criticize this statement using statistics.
7. A schoolteacher publishes a study showing that children's height in grade two is positively correlated with higher reading scores. Criticize this statement using statistics.
8. A class of economics majors explore the relationship between GDP and other factors such as life expectancy and happiness. The students reported the following. Identify their errors.
- A correlation of -0.75 shows that there is no association between GDP and clean drinking water.
 - There was a correlation of 0.51 between GDP and hemisphere.
 - There is a strong correlation between GDP and happiness of 1.14 .
 - The correlation between literacy in GDP was 0.87 , so an investment in education would increase a country's GDP.

9. Your assistant provides you with the following information: $\sigma_v = 30, \sigma_w = 15, \sigma_{v,w} = 225$. You're interested in using the variable w to predict v .

- Calculate the following:
 - the correlation coefficient.
 - the correlation of determination.
 - The beta coefficient: b_1
- What information would you need to calculate the The beta coefficient: b_0 ?
- How much variability in v is explained by the variability in w
- If w increased by 2, how would v change?

10. You find the following fictional information on the share price of Shopping Inc., s , and the equally fictitious Montreal Composite Index, m .

σ_s^2	625	σ_m^2	3025	$r_{s,m}$	0.7
$\bar{s}$	57.5	$\bar{n}$	105		

- Find the covariance between Shopping Inc. share price on the Montreal Composite Index.
 - Find the beta coefficients from regressing Shopping shares on the Montreal Composite Index.
 - If the Montreal Composite Index is 120, what share price would you predict for Shopping Inc. ?
 - Describe how the two series would appear on scatterplot.
 - How much of the variability in Shopping Inc. share prices is explained by the variability in the Montreal Composite Index?
11. Excel: Open the file GAMES_DescriptiveStats_Excel_PracticeProblems_Data.xlsx and sheet "Carbon emissions"
- Create a scatterplot of Western Europe CO2 on the y-axis and North America CO2 emissions on the x-axis. Format your graph appropriately.
 - Regress Western Europe CO2 on and North America CO2 emissions. Display the results on your graph. Are the conditions for linear regression satisfied?
 - This time, create a scatterplot of Western Europe CO2 on the **x-axis** and North America CO2 emissions on the **y-axis**. Now regress on North America CO2 on European CO2. Compare this result with regressing Europe on North America and answer the following questions:
 - Did the beta coefficients change? Explain why.
 - Did the r^2 change? Explain why.
12. Excel: Open the file GAMES_DescriptiveStats_Excel_PracticeProblems_Data.xlsx and sheet "Cryptocurrencies"
- Calculate the mean, median, sample standard deviation, range, and IQR for both Bitcoin and Ethereum. Create your answer in a table using Excel.
 - Add a regression line to your scatterplot. Do you think the conditions of inaccurate linear regression have been satisfied? Explain.

- (c) Add a regression line to your scatterplot. Describe what you see. Are the conditions for an accurate linear regression satisfied? Explain.
- (d) Calculate the beta coefficients and use these to compute the predicted Ethereum Price given the Bitcoin prices. Compute the Residuals and create a scatterplot with residuals on the y-axis and predicted Ethereum prices on the x-axis. Describe what you see. Based on your graph, are the conditions for an accurate simple linear regression satisfied?
- (e) Convert both the bitcoin in a theory and prices into daily percentage changes. For example in Excel, “ $= (B3 - B2) / B2$ ” where B3 is today’s price and B2 is yesterday’s price. Now perform your analysis in part d on these daily percentage changes. Describe what you see on your scatterplot of residuals and the actual ethereum prices. Based on your graph, are the conditions for an accurate simple linear regression satisfied?



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) <https://creativecommons.org/licenses/by-nc-nd/4.0/> Unless otherwise noted, all content in this video was created by Catherine Pfaff, Sumon Majumdar, and Robert J. McKeown. You are free to copy and share this material in any format, but you must give appropriate credit to the authors. This project is made possible with funding by the Government of Ontario and through eCampusOntario's support of the Virtual Learning Strategy. To learn more about the Virtual Learning Strategy visit: <https://vls.ecampusontario.ca/>.

