***Exploring Relationships between Quantitative Variables***

***Part 1 – Ice Cream Data***

**1. The science of statistics involves four basic components, the first of which is formulating questions. Suppose we are interested in the question, “Does the number of calories in ice cream relate to its sugar content?” Do you think there may be a relationship between sugar content and calories? If so, describe what you think this relationship may be. If not, explain why not.**

**The following table gives the calories, grams of fat, and grams of sugar in a ½ cup serving of each of several flavors of Breyer’s ice cream. To open the Fathom file containing this data, go to the CAMP website and click on the link. After opening the file, expand the window so that you will have plenty of work space.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Flavor** | **Calories** | **Fat** | **Sugar** | **Flavor** | **Calories** | **Fat** | **Sugar** |
| Carmel fudge | 160 | 7 | 18 | Banana fudge chunk | 170 | 9 | 19 |
| Vanilla | 140 | 8 | 15 | Vanilla fudge brownie | 160 | 9 | 16 |
| French vanilla | 150 | 8 | 15 | Cherry chocolate chip | 150 | 8 | 17 |
| Van/choc/straw | 140 | 8 | 15 | Peanut butter & fudge | 170 | 10 | 15 |
| Butter pecan | 170 | 11 | 14 | Dulce de Leche | 150 | 7 | 19 |
| Mint chocolate chip | 160 | 9 | 17 | Mocha almond fudge | 170 | 9 | 15 |
| Strawberry | 120 | 6 | 15 | Butter almond | 160 | 10 | 14 |
| Rocky road | 150 | 8 | 19 | Calcium rich vanilla | 130 | 7 | 14 |
| Cookies & cream | 160 | 8 | 16 | Carmel praline crunch | 180 | 9 | 19 |
| Vanilla fudge twirl | 140 | 7 | 15 | Fresh banana | 140 | 5 | 16 |
| Peach | 130 | 6 | 16 | Homemade vanilla | 140 | 7 | 13 |
| Coffee | 140 | 8 | 15 | Extra creamy vanilla | 150 | 8 | 14 |
| Cherry vanilla | 140 | 8 | 16 | Extra creamy chocolate | 140 | 7 | 15 |
| Chocolate chip | 160 | 9 | 17 | Take two | 150 | 8 | 15 |
| Choc chip cookie dough | 170 | 9 | 17 | Take two (sherbet) | 130 | 4.5 | 17 |
| Van/choc fudge checks | 170 | 9 | 17 |

**2. Construct a scatterplot of calories vs. sugar content by clicking on the Graph icon and dragging a graph to the work space. Now, click on the “sugar” column label, and drag the label to the bottom (horizontal axis) of the graph. Then drag the “calories” label from the table to the left side (vertical axis) of the graph.**

**3. Does the scatterplot suggest that there is some relationship between the sugar content and calories in ice cream? If so, describe what the scatterplot tells you about this relationship.**

**4. Repeat #2 & 3 to examine a possible relationship between fat content and calories. Does there appear to be a relationship between these variables? If so, describe this relationship.**

**5. Does every flavor of ice cream with higher fat content have more calories than every flavor with lower fat content? Why do we still say that there is a positive association between fat content and calories?**

**6. Which pair of variables, sugar content and calories OR fat content and calories, appears to have a stronger association? Explain. What might this suggest about the effect of sugar content and fat content on the number of calories in ice cream?**

**Let’s look more closely at the relationship between the number of calories and the sugar content of ice cream and see if we can describe this relationship using a mathematical model and if we can find out more about the strength of this relationship.**

**7. Click on your graph of calories vs. sugar content. Go to the Graph menu at the top of the window and choose “Add Movable Line”. You can move this line up or down by clicking toward the middle of the line, or you can tilt the line by clicking toward either end of the line. Place the movable line in a position that you think best represents the trend of the data points. Record the equation of your linear model below. (Fathom will record the equation of the line in slope-intercept form using the labels for the variables rather than  and , so the equation given will look like calories = *m*  sugar + *b*, where *m* is the slope and *b*  is the -intercept.)**

 **Linear model using movable line:**

 ** = \_\_\_\_\_\_\_\_\_\_   + \_\_\_\_\_\_\_\_\_\_ where  = calories and  = grams of sugar**

**8. According to this model, about how many calories would we expect a serving of ice cream containing 18 grams of sugar to have?**

**9. There is only one flavor in our list, Caramel Fudge, for which one serving contains 18 grams of sugar. How many calories does a serving of Caramel Fudge actually have? How well did your linear model predict the number of calories in a serving of Caramel Fudge ice cream based on its sugar content?**

**10. When we use a mathematical model to describe real data, the model rarely fits the data exactly. We can represent the variation between the model and the actual data using residuals. The residual value for a data point is the difference between the actual  value and the  value predicted by the model (). Find the residual value for Caramel Fudge. Is its value positive or negative? Why?**

**11. In general, when will a residual value be positive? When will it be negative? When will it be zero?**

**12. If some residual values are positive, some are negative, and some may be zero, what might the sum of all the residuals for a data set be equal to or close to?**

**13. Instead of looking at the sum of the residuals, we typically will square each residual value and then add the squared values together. This is called the *sum of squared residuals* (SSR). If we want a model that provides the best fit to our data, will we want the SSR to be as large or as small as possible? Why?**

**14. Click on your graph of calories vs. sugar content, go to the Graph menu and choose “Show Squares”. From the information given in #10 and #13 above, what do the squares on the graph represent?**

**15. Adjust the movable line, looking at the squares and at the given SSR value, until it provides the best possible fit to the data. Record the equation of your linear model below.**

 **Linear model using movable line with squares shown:**

 ** = \_\_\_\_\_\_\_\_\_\_   + \_\_\_\_\_\_\_\_\_\_ where  = calories and  = grams of sugar**

**16. Now go back to the Graph menu, click again on “Show Squares” (to remove the squares) and choose “Least-Squares Line”. The least-squares line is the linear model that mathematically minimizes the SSR. Sometimes this line is called the “line of best fit”. (Note: When you do “linear regression” in your calculator, you are finding the least-squares line.) Record the equation of the least-squares line.**

 **Least-Squares Line:**

 ** = \_\_\_\_\_\_\_\_\_\_   + \_\_\_\_\_\_\_\_\_\_ where  = calories and  = grams of sugar**

**17. How close was your linear model in #15 to the least-squares line in #16?**

**18. The least-squares line is the best linear model for this data, but how well does this line seem to represent the data? Would you be comfortable using this model to predict the number of calories in a serving of ice cream based on its sugar content? Why or why not?**

**19. Following the equation of the least-squares line at the bottom of the graph, a value called the Coefficient of Determination () is also given. This value tells us what percent of the variation in values can be explained by the linear relationship between  and . Based on our data, what percent of the variation in numbers of calories among the different flavors of ice cream can be explained by the linear relationship between calories and sugar content?**

**20. Does the sugar content of ice cream seem to be a good predictor of the number of calories in the ice cream? Why or why not?**

**21. Go to your graph of calories vs. fat content and obtain the least-squares line along with the  value. Does the fat content of ice cream seem to be a good predictor of the number of calories in the ice cream? Why or why not?**

**22. What conclusions can you draw from your work with this data?**

***Part 2 – CAMP Survey Data***

**Go back to the CAMP website and click on the link to the CAMP Survey data. Use Fathom to analyze possible relationships between quantitative variables using our survey results. Have fun! ☺**