Algebra I/ Math I Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Linear Regression

A Residuals Enrichment Activity

**Iris Activity**

Today, we are going to look at data collected on three different species of Iris. All three species are similar in appearance, but differ in length and width of their petals and sepals. When an iris is observed growing in the wild, these differences in length and width can be used to help classify the iris.

Before we begin, it is important to always know the story behind the data. So, first a little background:

From the Bioinformatics Research Group at South Dakota State University:

<http://www.sdstate.edu/mathstat/bioinformatics/>

The Iris flower data set is a multivariate data set introduced by Sir Ronald Aylmer Fisher (1936) as an example of discriminant analysis. It is sometimes called Anderson’s Iris data set because Edgar Anderson collected the data to quantify the geographic variation of Iris flowers in the Gaspé Peninsula. The dataset consists of 50 samples from each of the three species of Iris flowers (Iris setosa; Iris virginica and Iris veriscolor). Four features were measured from each sample, they are the length and the width of sepal and petal. Based on the combination of the four features, Fisher developed a linear discriminant model to determine which species they are.

|  |  |
| --- | --- |
| The **Gaspé Peninsula**, or ***Gaspésie*** (official name), is a peninsula along the south shore of the Saint Lawrence River in Quebec, Canada. | [http://t1.gstatic.com/images?q=tbn:ANd9GcRYgR-hOJ7gNFYVZ4CauEvHdp3q4QUw4vdJhtbCR-l3VZrpMgcU](http://www.google.com/url?sa=i&source=images&cd=&cad=rja&docid=8G-YpiVie3M32M&tbnid=4latK21DKA0LMM:&ved=0CAgQjRwwAA&url=http://www.fs.fed.us/wildflowers/beauty/iris/flowers.shtml&ei=TLsCUvsRhMqqAejBgOgK&psig=AFQjCNGk_0KmeBDKs7mognsNMpqq39o46Q&ust=1375997132054015)  <http://www.fs.fed.us/wildflowers/beauty/iris/flowers.shtml> |

|  |  |  |
| --- | --- | --- |
| [http://upload.wikimedia.org/wikipedia/commons/thumb/5/56/Kosaciec_szczecinkowaty_Iris_setosa.jpg/220px-Kosaciec_szczecinkowaty_Iris_setosa.jpg](http://en.wikipedia.org/wiki/File:Kosaciec_szczecinkowaty_Iris_setosa.jpg)  Iris setosa | [http://upload.wikimedia.org/wikipedia/commons/thumb/4/41/Iris_versicolor_3.jpg/220px-Iris_versicolor_3.jpg](http://en.wikipedia.org/wiki/File:Iris_versicolor_3.jpg)  Iris versicolor | [http://upload.wikimedia.org/wikipedia/commons/thumb/9/9f/Iris_virginica.jpg/220px-Iris_virginica.jpg](http://en.wikipedia.org/wiki/File:Iris_virginica.jpg)  Iris virginica |

The Iris data that we will use in this activity is taken from <http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.names> also available <http://en.wikipedia.org/wiki/Iris_flower_data_set>.

In the problems that follow, you are to identify the type of iris species based on the information provided.

1. The scatterplot below is a graph of petal length vs. petal width for the 150 irises in Fisher’s data set. Five new iris samples were collected and their petal length and width recorded in the table at right. Use the graph to help identify the iris species for each of the 5 samples.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mark  SE – for Iris setosa  VE – for Iris versicolor  VI – for Iris virginica  Five New Iris Samples   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Petal Width | 2.2 | 0.3 | 1.1 | 1.6 | 1.4 | | Petal Length | 5.8 | 2.2 | 3.6 | 4.9 | 2.8 | | Iris Species |  |  |  |  |  |   For which ordered pair above do you feel least confident about your classification? Why?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  ­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

1. In the scatterplots below, the versicolor and virginica have been separated and the equation of the least squares regression lines has been found for each. (See equation under each graph)

|  |  |
| --- | --- |
|  |  |

Let’s take another look at the iris with petal width 1.6 cm and petal length 4.9 cm. Use the equation of the least squares regression line for the virginica data set and the versicolor data set to determine the predicted petal width for a petal with a length of 1.6 cm.

|  |  |
| --- | --- |
| Iris virginica  Predicted petal length for a petal width of 1.6 cm.  Work:  Length: \_\_\_\_\_\_\_\_\_\_  Residual: (Actual – Predicted) \_\_\_\_\_\_\_\_\_\_\_ | Iris vesicolor  Predicted petal length for a petal width of 1.6 cm.  Work:  Length: \_\_\_\_\_\_\_\_\_\_  Residual: (Actual – Predicted) \_\_\_\_\_\_\_\_\_\_\_ |

Which species has a predicted petal length closest to the actual length of 4.9 cm? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How would you classify the iris with petal width 1.6 cm and petal length 4.9 cm? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Does this agree with your answer in #1 above? If not, which answer do you feel is the best classification? Explain.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. The scatterplot below shows a graph of sepal length vs. sepal width for all three species of iris.

|  |  |
| --- | --- |
|  | 1. If the length and width for an unidentified iris were given, which type of iris would be the easiest to identify? Explain.   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   1. An iris has a sepal length of 5.0 cm and a sepal width of 3.4 cm. Which species is this iris most likely to be?   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| 1. Royce was asked to classify an iris with a sepal length of 6.0 cm and a sepal width of 3.0 cm. Royce concluded that this iris is a virginica. Do you agree? Explain.   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | |

1. Ruth wanted to spend more time investigating the sepal length and sepal width of irises. She calculated the following equations for the least squares regression line for setosa and versicolor:

Setosa *Sepal\_Width\_in\_cm = 0.807 Sepal\_Length\_in\_cm - 0.62*

Versicolor *Sepal\_Width\_in\_cm = 0.320 Sepal\_Length\_in\_cm + 0.87*

Use Ruth’s equations to classify an iris with the following sepal length and width:

(Be sure to show organized work to support your classification.)

1. Sepal Length 5.0 cm and Sepal Width 2.6 cm

b. Sepal Length 5.6 cm and Sepal Width 3.6 cm