

San José State University  
Department of Applied Data Science

# DATA 220

## Mathematical Methods for Data Analysis

Spring 2021  
Instructor: Ron Mak

### Assignment #7

**Assigned:** Thursday, March 25, 2021  
**Due:** Thursday, April 8 at 5:30 pm  
**Team assignment, 100 points max**

#### Linear regression

In this assignment, you will perform linear regression analysis.

#### Two datasets

Find two datasets each of which has two variables which we'll call  $x$  and  $y$ , but they can have any names. You should have a suspicion or intuition that in each dataset, variable  $y$  is linearly dependent on variable  $x$ .

In two Jupyter notebooks, one for each dataset:

- Create a scatter plot of the data.
- Draw the least-squares regression line through the scatter plot.
- Calculate and print the coefficient of determination and the correlation coefficient.

#### Report

Include a report incorporated into your notebook for each dataset:

- Describe how variable  $y$  is dependent on variable  $x$ . What characteristics of the dataset produce this linear relationship?
- Does the correlation coefficient accurately reflect the strength of the relationship?
- Are the  $y$  values caused by the  $x$  values?

#### What to submit to Canvas

Submit your Jupyter notebooks: **Assignment #7**.

## Rubric

Your notebooks will be graded according to these criteria:

Criteria	Max points
<b>Suitable datasets</b> <ul style="list-style-type: none"><li>Dataset #1 with a linear relationship</li><li>Dataset #2 with a linear relationship</li></ul>	<b>20</b> <ul style="list-style-type: none"><li>10</li><li>10</li></ul>
<b>Charts</b> <ul style="list-style-type: none"><li>Dataset #1 scatter plot</li><li>Dataset #1 least-squares regression line</li><li>Dataset #1 coefficient of determination and correlation coefficient</li><li>Dataset #2 scatter plot</li><li>Dataset #2 least-squares regression line</li><li>Dataset #2 coefficient of determination and correlation coefficient</li></ul>	<b>60</b> <ul style="list-style-type: none"><li>10</li><li>10</li><li>10</li><li>10</li><li>10</li><li>10</li></ul>
<b>Report</b> (incorporated in the notebooks) <ul style="list-style-type: none"><li>Report #1</li><li>Report #2</li></ul>	<b>20</b> <ul style="list-style-type: none"><li>10</li><li>10</li></ul>