



Microsoft: DAT210x Programming with Python for Data Science



Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▼ 5. Data Modeling

Lecture: Clustering

Quiz

**Lab: Clustering**

Lab

**Lecture: Splitting Data**

Quiz

**Lecture: K-Nearest
Neighbors**

Quiz

**Lab: K-Nearest Neighbors**

5. Data Modeling > Lab: Regression > Lab 8



Bookmark

Welcome to Module 5's Linear Regression Labs!

In order to complete the Linear Regression labs in this module, please make sure you download and unarchive this .zip file with all the datasets and files necessary.

Lab Assignment 8

Advances in medicine, an increase in healthcare facilities, and improved standards of care have all contributed to an increased overall life expectancy over the last few decades. Although this might seem like great achievement for humanity, it has also led to the abandonment of more elderly people into senior-care and assisted living communities. The morality, benefits, and disadvantages of leaving one's parents in such facilities are still debatable; however, the fact that this practice has increased the financial burden on both the private-sector and government is not.

In this lab assignment, you will be using the subset a life expectancy dataset, provided courtesy of the Center for Disease Control and Prevention's National Center for Health Statistics page. The page hosts many open datasets on topics ranging from injuries, poverty, women's health, education, health insurance, and of course infectious diseases, and much more. But the one you'll be using is their "Life expectancy at birth, at age 65, and at age 75, by sex, race, and origin" data set, which has statistics dating back from the 1900's to current, taken within the United States. We have included a copy of the

Lab



Lecture: Regression

Quiz



Lab: Regression

Lab



Dive Deeper

dataset as Module5/Datasets/**life_expectancy.csv**. The dataset only lists the life expectancy of whites and blacks, because throughout most of the collection period, those were the dominant two races that actively had their statistics recorded within the U.S.

Using linear regression, you will extrapolate how long people will live in the future. The private sector and governments mirror these calculations when computing social security payouts, taxes, infrastructure, and more. Fire up your Module5/**assignment8.py** starter code, and then complete the following:

1. Make sure the dataset has been properly loaded.
2. Create a linear model to use and re-use throughout the assignment. You can retrain the same model again, rather than re-creating a new instance of the class.
3. Slice out using indexing any records before 1986 into a brand new slice.
4. Have one slice for training and one for testing. First, map the life expectancy of white males as a function of age, or `WhiteMales = f(age)`.
5. Fit your model, draw a regression line and scatter plot with the convenience function, and then print out the actual, observed 2015 White Male life expectancy value from the dataset.
6. Repeat for BlackFemale.
7. Print out and display a correlation matrix for the dataset's features.
8. Answer the questions below.

Multiple Choice

(2 points possible)

Please enter a numeric value (e.g. 0, 1, 10.5, etc) which correctly answers the question(s) below:

Enter the 2014 life expectancy prediction for white males living within the U.S, using only data recorded prior to 1986:

?

Which relationship had the highest correlation?

?

You have used 0 of 2 submissions

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY
OPENedX



