

Environmental Health Project Part 1 - Problem 3 (Extract linear relationship by doing linear regression)

March 28, 2023

0.0.1 Environmental Health Project Part 1 - Problem 3 (Extract linear relationship by doing linear regression)

Author: Kenneth Cochran

Date written: 3/24/23

```
[1]: import os
import numpy as np
import pandas as pd
import seaborn as sns
import statsmodels.api as sm
%matplotlib inline

result_dir = '~/Environmental Health project - part 1/result'
data_dir = '~/Environmental Health project - part 1/data'

in_file_name = 'Clean PFAS_J.xlsx'
out_file_name = "Filtered_R^2_values.csv"

in_file_full_name = os.path.join(result_dir, in_file_name)
out_file_full_name = os.path.join(result_dir, out_file_name)

data_in = pd.read_excel(in_file_full_name) #Reads the PFAS_J excel sheet into a
↳ data frame object
data = data_in.drop(['Unnamed: 0', 'SEQN'], axis = 1) #Drops the index column
↳ and the seqn label column.

column_points = data.columns #Gets the labels for the final data frame
length = len(column_points) #Gets the amount of labels of the dataframe for the
↳ column and index so the final dataframe can be shaped correctly
P_values = [] #Creates an empty list that the P-values of the F statistic will
↳ be stored in
R_squared_values = [] #Creates an empty list that the R^2 values will be stored
↳ in
```

```

for h in column_points: #Calculates the pvalue of the f-statistic and r^2 value
    ↪for each pair of variables
        for k in column_points:
            X2 = sm.add_constant(data[h])
            est = sm.OLS(data[k],X2)
            est2 = est.fit()
            P_values.append(est2.f_pvalue) #Appends the pvalue of the f-statistic
            ↪to the p_values list
            R_squared_values.append(est2.rsquared) #Appends the r^2 value to the
            ↪R_squared_values list

reshaped_p_values = np.array(P_values).reshape(length,length) #Reshapes the
            ↪P_value list into a numpy array of the correct shape
p_value_matrix = pd.DataFrame(reshaped_p_values, index = column_points, columns=
            ↪column_points) #Creates a dataframe of the P_values array with the
            ↪variable names as the columns and indexes
print("P-values for the F statistic")
display(p_value_matrix)

reshaped_R_squared_values = np.array(R_squared_values).reshape(length,length)
            ↪#Reshapes the R_squared_values list into a numpy array of the correct shape
R_squared_values_matrix = pd.DataFrame(reshaped_R_squared_values, index =
            ↪column_points, columns = column_points) #Creates a dataframe of the
            ↪R_squared_values array with the variable names as the columns and indexes
print("R^2 Values")
display(R_squared_values_matrix)

filtered_matrix = R_squared_values_matrix[(R_squared_values_matrix > 0.50) &
            ↪(R_squared_values_matrix < 0.999)] #Filters the R_squared_values matrix to
            ↪only show values greater than 0.50 and less than 0.999
print("Filtered R^2 Values")
display(filtered_matrix)

print("HeatMap of filtered R^2 values")
print(sns.heatmap(filtered_matrix))
filtered_matrix.to_csv(out_file_full_name) #Outputs the filtered matrix to a
            ↪csv file

```

P-values for the F statistic

	WTSB2YR	LBXPFDE	LBDPFDEL	LBXPFHS	LBDPFHSL	\
WTSB2YR	0.000000	1.216565e-02	8.742231e-02	6.738398e-01	2.447543e-01	
LBXPFDE	0.012166	0.000000e+00	4.022266e-19	3.549814e-09	1.127461e-01	
LBDPFDEL	0.087422	4.022266e-19	0.000000e+00	5.141260e-07	4.385772e-14	
LBXPFHS	0.673840	3.549814e-09	5.141260e-07	0.000000e+00	2.057209e-02	
LBDPFHSL	0.244754	1.127461e-01	4.385772e-14	2.057209e-02	0.000000e+00	
LBXMPAH	0.620278	3.627985e-06	6.925563e-03	1.781187e-03	2.580786e-01	
LBDMPAHL	0.060441	2.398585e-03	2.535216e-13	6.775958e-03	1.307300e-03	

LBXPFNA	0.704234	3.699566e-249	4.172597e-32	2.128936e-21	2.810656e-03
LBDPFNAL	0.318496	2.734041e-10	1.899994e-119	5.312075e-07	1.311831e-17
LBXPFUA	0.000034	0.000000e+00	9.528374e-12	1.752720e-17	2.437152e-01
LBDPFUAL	0.689471	7.748586e-43	3.634396e-98	2.524898e-04	7.024807e-03
LBXNFOA	0.419036	2.090434e-52	1.821827e-13	1.315075e-92	4.426671e-03
LBDNFOAL	0.225943	1.347088e-01	7.736845e-14	9.077259e-02	1.715138e-211
LBXBFOA	0.916885	9.954320e-01	1.372519e-01	6.695324e-01	6.078061e-01
LBDDBFOAL	0.585679	9.848480e-01	8.668749e-01	4.257443e-01	1.187321e-02
LBXNFOS	0.000062	4.428897e-272	1.979109e-16	1.951596e-41	1.484979e-02
LBDNFOSL	0.242918	1.662319e-01	4.805563e-12	1.164654e-01	4.999992e-260
LBXMFOS	0.341906	7.133078e-72	2.617623e-20	8.273241e-55	7.729233e-04
LBDMFOSL	0.158680	2.603343e-02	1.358466e-22	1.353268e-02	1.601705e-112

	LBXMPAH	LBDMPAHL	LBXPFNA	LBDPFNAL	\
WTSB2YR	6.202776e-01	6.044106e-02	7.042343e-01	3.184958e-01	
LBXPFDE	3.627985e-06	2.398585e-03	3.699566e-249	2.734041e-10	
LBDPFDEL	6.925563e-03	2.535216e-13	4.172597e-32	1.899994e-119	
LBXPFHS	1.781187e-03	6.775958e-03	2.128936e-21	5.312075e-07	
LBDPFHSL	2.580786e-01	1.307300e-03	2.810656e-03	1.311831e-17	
LBXMPAH	0.000000e+00	5.482862e-53	9.057528e-17	1.024298e-04	
LBDMPAHL	5.482862e-53	0.000000e+00	1.082440e-07	3.151677e-11	
LBXPFNA	9.057528e-17	1.082440e-07	0.000000e+00	3.255755e-32	
LBDPFNAL	1.024298e-04	3.151677e-11	3.255755e-32	0.000000e+00	
LBXPFUA	9.168905e-04	1.634362e-01	8.023072e-213	3.057959e-07	
LBDPFUAL	8.346154e-03	1.167254e-05	1.027077e-58	1.030531e-39	
LBXNFOA	7.671387e-06	5.369797e-03	1.404475e-100	3.301223e-13	
LBDNFOAL	2.988431e-01	1.582699e-02	1.395658e-02	5.568391e-21	
LBXBFOA	3.311484e-03	9.244924e-02	5.495355e-01	2.192428e-01	
LBDDBFOAL	5.074127e-01	1.364773e-01	3.131969e-01	3.224013e-01	
LBXNFOS	1.214987e-24	8.160602e-09	1.163359e-160	8.237619e-13	
LBDNFOSL	3.392548e-01	3.450291e-02	2.289580e-02	3.820558e-18	
LBXMFOS	1.290847e-45	1.064694e-17	2.252051e-119	7.984342e-21	
LBDMFOSL	2.441593e-01	2.317813e-02	2.572047e-04	1.997581e-35	

	LBXPFUA	LBDPFUAL	LBXNFOA	LBDNFOAL	LBXBFOA	\
WTSB2YR	3.438658e-05	6.894711e-01	4.190357e-01	2.259430e-01	0.916885	
LBXPFDE	0.000000e+00	7.748586e-43	2.090434e-52	1.347088e-01	0.995432	
LBDPFDEL	9.528374e-12	3.634396e-98	1.821827e-13	7.736845e-14	0.137252	
LBXPFHS	1.752720e-17	2.524898e-04	1.315075e-92	9.077259e-02	0.669532	
LBDPFHSL	2.437152e-01	7.024807e-03	4.426671e-03	1.715138e-211	0.607806	
LBXMPAH	9.168905e-04	8.346154e-03	7.671387e-06	2.988431e-01	0.003311	
LBDMPAHL	1.634362e-01	1.167254e-05	5.369797e-03	1.582699e-02	0.092449	
LBXPFNA	8.023072e-213	1.027077e-58	1.404475e-100	1.395658e-02	0.549536	
LBDPFNAL	3.057959e-07	1.030531e-39	3.301223e-13	5.568391e-21	0.219243	
LBXPFUA	0.000000e+00	2.472172e-46	3.581948e-57	2.585995e-01	0.882891	
LBDPFUAL	2.472172e-46	0.000000e+00	2.073154e-14	3.752043e-03	0.847820	
LBXNFOA	3.581948e-57	2.073154e-14	0.000000e+00	2.281975e-02	0.007857	
LBDNFOAL	2.585995e-01	3.752043e-03	2.281975e-02	0.000000e+00	0.896623	

LBXBFOA	8.828912e-01	8.478197e-01	7.856689e-03	8.966232e-01	0.000000
LBDBFOAL	8.644944e-01	2.102010e-01	6.896620e-01	3.562657e-03	0.000000
LBXNFOS	1.998296e-203	4.042544e-32	2.778021e-29	5.473049e-02	0.581420
LBDNFOSL	2.780824e-01	6.159676e-04	3.572382e-02	0.000000e+00	0.968267
LBXMFOS	2.238158e-43	8.684101e-24	7.060924e-58	1.127562e-02	0.668243
LBDMFOSL	1.097956e-01	5.886884e-05	1.647223e-03	5.303248e-163	0.523244

	LBDBFOAL	LBXNFOS	LBDNFOSL	LBXMFOS	LBDMFOSL
WTSB2YR	0.585679	6.178785e-05	2.429183e-01	3.419059e-01	1.586800e-01
LBXPFDE	0.984848	4.428897e-272	1.662319e-01	7.133078e-72	2.603343e-02
LBDPFDEL	0.866875	1.979109e-16	4.805563e-12	2.617623e-20	1.358466e-22
LBXPFHS	0.425744	1.951596e-41	1.164654e-01	8.273241e-55	1.353268e-02
LBDPFHSL	0.011873	1.484979e-02	4.999992e-260	7.729233e-04	1.601705e-112
LBXMPAH	0.507413	1.214987e-24	3.392548e-01	1.290847e-45	2.441593e-01
LBDMPAHL	0.136477	8.160602e-09	3.450291e-02	1.064694e-17	2.317813e-02
LBXPFNA	0.313197	1.163359e-160	2.289580e-02	2.252051e-119	2.572047e-04
LBDPFNAL	0.322401	8.237619e-13	3.820558e-18	7.984342e-21	1.997581e-35
LBXPFUA	0.864494	1.998296e-203	2.780824e-01	2.238158e-43	1.097956e-01
LBDPFUAL	0.210201	4.042544e-32	6.159676e-04	8.684101e-24	5.886884e-05
LBXNFOA	0.689662	2.778021e-29	3.572382e-02	7.060924e-58	1.647223e-03
LBDNFOAL	0.003563	5.473049e-02	0.000000e+00	1.127562e-02	5.303248e-163
LBXBFOA	0.000000	5.814197e-01	9.682669e-01	6.682426e-01	5.232444e-01
LBDBFOAL	0.000000	1.378851e-01	5.541689e-02	2.609837e-01	7.327539e-01
LBXNFOS	0.137885	0.000000e+00	7.399356e-02	3.051583e-262	4.681199e-03
LBDNFOSL	0.055417	7.399356e-02	0.000000e+00	1.906903e-02	3.768553e-127
LBXMFOS	0.260984	3.051583e-262	1.906903e-02	0.000000e+00	1.161643e-04
LBDMFOSL	0.732754	4.681199e-03	3.768553e-127	1.161643e-04	0.000000e+00

R^2 Values

	WTSB2YR	LBXPFDE	LBDPFDEL	LBXPFHS	LBDPFHSL	LBXMPAH \
WTSB2YR	1.000000	3.257887e-03	0.001515	0.000092	0.000702	0.000127
LBXPFDE	0.003258	1.000000e+00	0.040605	0.017930	0.001305	0.011073
LBDPFDEL	0.001515	4.060514e-02	1.000000	0.013001	0.029148	0.003778
LBXPFHS	0.000092	1.793030e-02	0.013001	1.000000	0.002780	0.005055
LBDPFHSL	0.000702	1.304780e-03	0.029148	0.002780	1.000000	0.000664
LBXMPAH	0.000127	1.107345e-02	0.003778	0.005055	0.000664	1.000000
LBDMPAHL	0.001828	4.772580e-03	0.027408	0.003798	0.005349	0.114722
LBXPFNA	0.000075	4.456532e-01	0.069662	0.045752	0.004623	0.035265
LBDPFNAL	0.000516	2.047571e-02	0.244436	0.012968	0.037172	0.007803
LBXPFUA	0.008867	6.913014e-01	0.023809	0.036886	0.000705	0.005688
LBDPFUAL	0.000083	9.308288e-02	0.205087	0.006928	0.003765	0.003605
LBXNFOA	0.000339	1.134967e-01	0.027736	0.194479	0.004196	0.010337
LBDNFOAL	0.000761	1.160698e-03	0.028585	0.001484	0.393478	0.000560
LBXBFOA	0.000006	1.701375e-08	0.001146	0.000095	0.000137	0.004468
LBDBFOAL	0.000154	1.872153e-07	0.000015	0.000329	0.003280	0.000228
LBXNFOS	0.008295	4.751856e-01	0.034493	0.090056	0.003076	0.053049
LBDNFOSL	0.000708	9.943664e-04	0.024489	0.001278	0.459858	0.000474
LBXMFOS	0.000469	1.536623e-01	0.043291	0.118552	0.005851	0.099056

LBDMFOSL	0.001031	2.568094e-03	0.048446	0.003160	0.231848	0.000704
----------	----------	--------------	----------	----------	----------	----------

	LBDMPAHL	LBXPFNA	LBDPFNAL	LBXPFUA	LBDPFUAL	LBXNFOA \
WTSB2YR	0.001828	0.000075	0.000516	0.008867	0.000083	0.000339
LBXPFDE	0.004773	0.445653	0.020476	0.691301	0.093083	0.113497
LBDPFDEL	0.027408	0.069662	0.244436	0.023809	0.205087	0.027736
LBXPFHS	0.003798	0.045752	0.012968	0.036886	0.006928	0.194479
LBDPFHSL	0.005349	0.004623	0.037172	0.000705	0.003765	0.004196
LBXMPAH	0.114722	0.035265	0.007803	0.005688	0.003605	0.010337
LBDMPAHL	1.000000	0.014542	0.022621	0.001008	0.009925	0.004015
LBXPFNA	0.014542	1.000000	0.069900	0.395401	0.126712	0.209649
LBDPFNAL	0.022621	0.069900	1.000000	0.013514	0.086322	0.027146
LBXPFUA	0.001008	0.395401	0.013514	1.000000	0.100593	0.123499
LBDPFUAL	0.009925	0.126712	0.086322	0.100593	1.000000	0.029890
LBXNFOA	0.004015	0.209649	0.027146	0.123499	0.029890	1.000000
LBDNFOAL	0.003017	0.003132	0.044810	0.000662	0.004351	0.002686
LBXBFOA	0.001469	0.000186	0.000783	0.000011	0.000019	0.003661
LBDBFOAL	0.001150	0.000528	0.000508	0.000015	0.000815	0.000083
LBXNFOS	0.017104	0.315216	0.026239	0.381682	0.069692	0.063408
LBDNFOSL	0.002317	0.002683	0.038388	0.000610	0.006069	0.002287
LBXMFOS	0.037378	0.244303	0.044456	0.094245	0.051132	0.124969
LBDMFOSL	0.002672	0.006910	0.076967	0.001326	0.008342	0.005129

	LBDNFOAL	LBXBFOA	LBDBFOAL	LBXNFOS	LBDNFOSL \
WTSB2YR	0.000761	5.653121e-06	1.542280e-04	0.008295	7.075504e-04
LBXPFDE	0.001161	1.701375e-08	1.872153e-07	0.475186	9.943664e-04
LBDPFDEL	0.028585	1.145723e-03	1.458571e-05	0.034493	2.448869e-02
LBXPFHS	0.001484	9.454499e-05	3.292499e-04	0.090056	1.278300e-03
LBDPFHSL	0.393478	1.367060e-04	3.280206e-03	0.003076	4.598584e-01
LBXMPAH	0.000560	4.468251e-03	2.280554e-04	0.053049	4.739314e-04
LBDMPAHL	0.003017	1.468529e-03	1.150251e-03	0.017104	2.317377e-03
LBXPFNA	0.003132	1.858898e-04	5.278368e-04	0.315216	2.683335e-03
LBDPFNAL	0.044810	7.831453e-04	5.081399e-04	0.026239	3.838829e-02
LBXPFUA	0.000662	1.126326e-05	1.511716e-05	0.381682	6.104743e-04
LBDPFUAL	0.004351	1.911532e-05	8.146389e-04	0.069692	6.068762e-03
LBXNFOA	0.002686	3.660997e-03	8.276403e-05	0.063408	2.286631e-03
LBDNFOAL	1.000000	8.762669e-06	4.399468e-03	0.001914	5.940622e-01
LBXBFOA	0.000009	1.000000e+00	5.416897e-01	0.000158	8.214942e-07
LBDBFOAL	0.004399	5.416897e-01	1.000000e+00	0.001142	1.902935e-03
LBXNFOS	0.001914	1.577581e-04	1.142044e-03	1.000000	1.655586e-03
LBDNFOSL	0.594062	8.214942e-07	1.902935e-03	0.001656	1.000000e+00
LBXMFOS	0.003328	9.533139e-05	6.556537e-04	0.462708	2.848349e-03
LBDMFOSL	0.319032	2.114978e-04	6.052142e-05	0.004143	2.581965e-01

	LBXMFOS	LBDMFOSL
WTSB2YR	0.000469	0.001031
LBXPFDE	0.153662	0.002568
LBDPFDEL	0.043291	0.048446

LBXPFHS	0.118552	0.003160
LBDPFHSL	0.005851	0.231848
LBXMPAH	0.099056	0.000704
LBDMPAHL	0.037378	0.002672
LBXPFNA	0.244303	0.006910
LBDPFNAL	0.044456	0.076967
LBXPFUA	0.094245	0.001326
LBDPFUAL	0.051132	0.008342
LBXNFOA	0.124969	0.005129
LBDNFOAL	0.003328	0.319032
LBXBFOA	0.000095	0.000211
LBDDBFOAL	0.000656	0.000061
LBXNFOS	0.462708	0.004143
LBDNFOSL	0.002848	0.258196
LBXMFOS	1.000000	0.007681
LBDMFOSL	0.007681	1.000000

Filtered R² Values

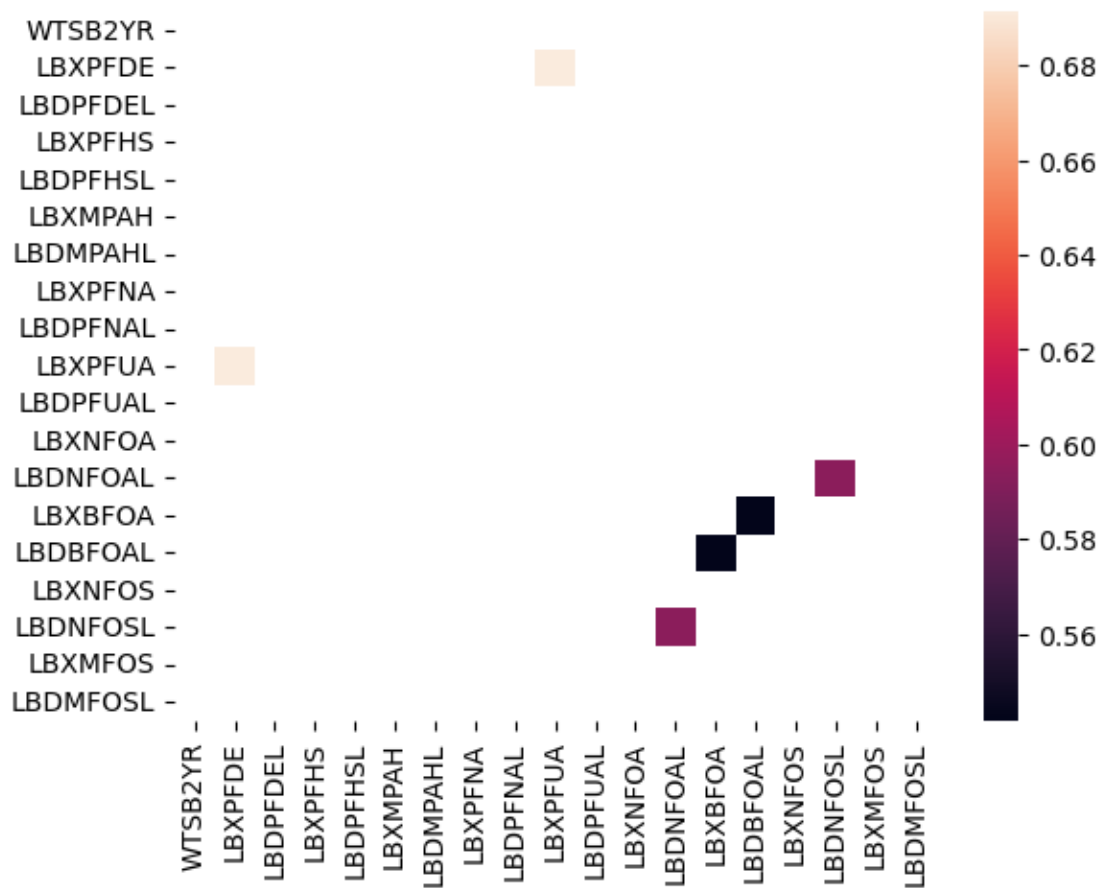
	WTSB2YR	LBXPFDE	LBDPFDEL	LBXPFHS	LBDPFHSL	LBXMPAH	LBDMPAHL	\
WTSB2YR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFDE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDPFDEL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFHS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDPFHSL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXMPAH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDMPAHL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFNA	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDPFNAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFUA	NaN	0.691301	NaN	NaN	NaN	NaN	NaN	
LBDPFUAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXNFOA	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDNFOAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXBFOA	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDDBFOAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXNFOS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDNFOSL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXMFOS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDMFOSL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	LBXPFNA	LBDPFNAL	LBXPFUA	LBDPFUAL	LBXNFOA	LBDNFOAL	LBXBFOA	\
WTSB2YR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFDE	NaN	NaN	0.691301	NaN	NaN	NaN	NaN	
LBDPFDEL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFHS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDPFHSL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXMPAH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBDMPAHL	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LBXPFNA	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

LBDPFNAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBXPFUA	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBDPFUAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBXNFOA	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBDNFOAL	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBXBFOA	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBDBFOAL	NaN	NaN	NaN	NaN	NaN	NaN	0.54169
LBXNFOS	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBDNFOSL	NaN	NaN	NaN	NaN	NaN	0.594062	NaN
LBXMFOS	NaN	NaN	NaN	NaN	NaN	NaN	NaN
LBDMFOSL	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	LBDBFOAL	LBXNFOS	LBDNFOSL	LBXMFOS	LBDMFOSL
WTSB2YR	NaN	NaN	NaN	NaN	NaN
LBXPFDE	NaN	NaN	NaN	NaN	NaN
LBDPFDEL	NaN	NaN	NaN	NaN	NaN
LBXPFHS	NaN	NaN	NaN	NaN	NaN
LBDPFHSL	NaN	NaN	NaN	NaN	NaN
LBXMPAH	NaN	NaN	NaN	NaN	NaN
LBDMPAHL	NaN	NaN	NaN	NaN	NaN
LBXPFNA	NaN	NaN	NaN	NaN	NaN
LBDPFNAL	NaN	NaN	NaN	NaN	NaN
LBXPFUA	NaN	NaN	NaN	NaN	NaN
LBDPFUAL	NaN	NaN	NaN	NaN	NaN
LBXNFOA	NaN	NaN	NaN	NaN	NaN
LBDNFOAL	NaN	NaN	0.594062	NaN	NaN
LBXBFOA	0.54169	NaN	NaN	NaN	NaN
LBDBFOAL	NaN	NaN	NaN	NaN	NaN
LBXNFOS	NaN	NaN	NaN	NaN	NaN
LBDNFOSL	NaN	NaN	NaN	NaN	NaN
LBXMFOS	NaN	NaN	NaN	NaN	NaN
LBDMFOSL	NaN	NaN	NaN	NaN	NaN

HeatMap of filtered R² values
AxesSubplot(0.125,0.11;0.62x0.77)



These R^2 values indicate that there is very little positive correlation between the pairs of variables. The highest R^2 value is between LBXPFDE and LBXPFUA.