\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Homework 7

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Instructions:

\* To create this document, first copy and paste the full text here into a .Do document (a STATA Do-File).

\* Below each question, write the code you used to answer the question

\* Next, write your actual answer to the question by commenting out your writing (by starting the line with a \*)

\* Next, copy and paste the entire document (my writing and yours) into a Word document. This will allow me to see your code on Canvas without downloading every homework.

\* The goal is that I should be able to copy and paste your entire text into a .Do File and run the code without any errors.

\* Finally, submit file as Homework 7 on Canvas

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Topic 1: K-Nearest Neighbors

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*1. Change directory

clear

cd "C:\Users\haniu\OneDrive\Desktop\Deepa\Deepa\Finance Core B\Business\Homework 7"

\*2. Import the Excel file CountryRiskData

import excel "CountryRiskData.xlsx", firstrow

\*3 Z-score scale Corruption, Peace, Legal, and GDP Growth variables

\* Use the Z-scored variables for the rest of the homework

egen Mean\_Corruption = mean(corruption)

egen Mean\_Peace = mean(peace)

egen Mean\_Legal = mean(legal)

egen Mean\_GDP = mean(gdpgrowth)

egen SD\_Corruption = sd(corruption)

egen SD\_Peace = sd(peace)

egen SD\_Legal = sd(legal)

egen SD\_GDP = sd(gdpgrowth)

gen zscore\_corruption = (corruption-Mean\_Corruption)/SD\_Corruption

gen zscore\_peace = (peace-Mean\_Peace)/SD\_Peace

gen zscore\_legal = (legal-Mean\_Legal)/SD\_Legal

gen zscore\_GDP = (gdpgrowth-Mean\_GDP)/SD\_GDP

\*4. Create a scatter plot with legal as the y-variable and corruption as the x-variable. Plot a line on top of the scatter plot

\* What do you notice about the line?

twoway (scatter zscore\_legal zscore\_corruption) (lfit zscore\_legal zscore\_corruption)

\* the variables are highly correlated

\*5. Run a linear regression with the Default Spread as the y-variable and Corruption, Peace, and GDP Growth as the x-variables

reg defaultspread zscore\_corruption zscore\_peace zscore\_GDP zscore\_legal

\*6. Estimate the residuals using the linear regression above

reg defaultspread zscore\_peace zscore\_GDP zscore\_legal

predict reg\_defaultspread, res

\*7. Predict Default Spread using K-Nearest Neighbors . Use 3 Neighbors

\* discrim knn normal\_peace normal\_legal normal\_gdpgrowth if testdata==0, group(defaultspread) k(3) ties(random)

discrim knn zscore\_peace zscore\_legal zscore\_GDP, group(defaultspread) k(3) ties(random)

\*8. Estimate the residuals from the K-Nearest Neighbors estimate

predict predict\_defaultspread

gen knn\_defaultspread = defaultspread - predict\_defaultspread

\*9 Which method (linear regression or KNN) performed better? Why?

\*Linear Regression

\*Linear regression works majorly with regards to variables and KNN works on which is better in the neighbour but its just to compare something close and not on the basis of variables

\*10. Run a linear regression with the Equity Risk Premium as the y-variable and Corruption, Peace, and GDP Growth as the x-variables

reg equityriskpremium zscore\_corruption zscore\_peace zscore\_GDP zscore\_legal

\*11. Estimate the residuals using the linear regression above

reg equityriskpremium zscore\_peace zscore\_GDP zscore\_legal

predict reg\_equityriskpremium, res

\*12. Predict Equity Risk Premium using K-Nearest Neighbors . Use 3 Neighbors

discrim knn zscore\_peace zscore\_legal zscore\_GDP, group(equityriskpremium) k(3) ties(random)

\*13. Estimate the residuals from the K-Nearest Neighbors estimate

predict predict\_equityriskpremium

gen knn\_equityriskpremium = equityriskpremium - predict\_equityriskpremium

\*14 Which method (linear regression or KNN) performed better? Why?

\*Linear Regression

\*Linear regression works majorly with regards to variables and KNN works on which is better in the neighbour but its just to compare something close and not on the basis of variables

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Topic 2: K-Means

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*15. Cluster countries into three groups based on peace, legal, and GDP growth.

cluster kmeans zscore\_peace zscore\_legal zscore\_GDP, k(3) keepcenters

\* we need to use zscores here as they would be calculating the distance and it needs to be in the same range so as to compare.

\* the range starts 1-3, which will be high risk to low risk

\*16. Estimate mean default spread and equity risk premium for each cluser

bysort \_clus\_1: sum defaultspread equityriskpremium

\*17. Based on these clusters, determine which cluster is high-risk, moderate-risk, and low-risk

\* Note that high levels in the peace index denote wars

bysort \_clus\_1: sum zscore\_peace zscore\_legal zscore\_GDP

\* 18. Rename the clusters: HighRisk, ModerateRisk, LowRisk

bysort \_clus\_1: egen mean\_legal = mean(zscore\_legal)

egen min\_legal = min(mean\_legal)

egen max\_legal = max(mean\_legal)

capture drop Risklevel

gen Risklevel = "High" if mean\_legal == max\_legal

replace Risklevel = "Low" if mean\_legal == min\_legal

replace Risklevel = "Moderate" if Risklevel=="" & mean\_legal!=.

\* last code = mean\_legal!=. means if i dont have data for it just drop those data and !=. means not equal to

\*19. Compare the Default Spread and Equity Risk Premium for each risk clusters

\*What do you find?

bysort Risklevel: sum zscore\_peace zscore\_legal zscore\_GDP defaultspread equityriskpremium

\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* Topic 3: Principal Components Analysis

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*20. Estimate the principal components

pca zscore\*

\*Eigenvalue is explaining how much that variable is explaining

\*

\*22. Estimate the variance accounted for from each component

\* Hint: This is available from the pca output

\*21. Create new variables for each component using

\* predict f1 f2 f3 f4, score

predict f1 f2 f3 f4, score

\*22. Determine which factor seems to drive the country's risk premium

corr defaultspread equityriskpremium f1 f2 f3 f4

\* here f1 is affecting equityriskpremium & defaultspread

reg equityriskpremium f1 f2 f3 f4

reg defaultspread f1 f2 f3 f4