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

\* 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

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

clear

\*1. Change directory

\*DONE\*

\*cd "/Users/henryvelasquez/Documents/MBA/MBA S3/2nd 7/Machine Learning in Finance K579/HW7/"

\*2. Import the Excel file CountryRiskData

\*DONE\*

import excel "/Users/henryvelasquez/Documents/MBA/MBA S3/2nd 7/Machine Learning in Finance K579/HW7/CountryRiskData.xlsx", sheet("Sheet1") firstrow

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

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

foreach var in corruption peace legal gdpgrowth {

egen mean = mean(`var')

egen sd = sd(`var')

gen zscore\_`var' = (`var' - mean)/sd

capture drop mean sd

}

\*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 legal corruption) (lfit legal corruption)

\*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 corruption peace gdpgrowth

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

predict predDefaultSpread

predict resDefaultSpread, residuals

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

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

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

predict predDefaultSpread\_K

gen resDefaultSpread\_K = defaultspread - predDefaultSpread\_K

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

summarize resDefaultSpread\_K, detail

\*variance .0008

\*rerun for resDefaultSpread

summarize resDefaultSpread\_K, detail

\*variance .0005

\*linear is better. you decide what's important

\*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 corruption peace gdpgrowth

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

predict resERP, residuals

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

discrim knn zscore\_corruption zscore\_peace zscore\_gdpgrowth, group(equityriskpremium) k(3) ties(random)

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

predict predERP\_K

gen resERP\_K = equityriskpremium - predERP

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

summarize resERP, detail

\*var .0247

summarize resERP, detail

\*Var .0247

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

\* Topic 2: K-Means

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

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

cluster kmeans peace legal gdpgrowth, k(3) keepcenters

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

bysort \_clus\_1: sum defaultspread equityriskpremium

/\*

-> \_clus\_1 = 1

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

defaultspr~d | 37 .0201081 .0353765 0 .175

equityrisk~m | 37 .0657784 .0411159 .0424 .2458

-----------------------------------------------------------------------------------------------------

-> \_clus\_1 = 2

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

defaultspr~d | 50 .030346 .0230306 .0042 .1021

equityrisk~m | 50 .07767 .0267595 .0473 .1611

-----------------------------------------------------------------------------------------------------

-> \_clus\_1 = 3

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

defaultspr~d | 22 .0360864 .018835 .006 .0766

equityrisk~m | 22 .0843455 .0218784 .0494 .1314

\*/

\*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\_gdpgrowth

/\*

-> \_clus\_1 = 1

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

zscore\_peace | 40 -.3514558 1.036485 -2.013102 2.197218

zscore\_legal | 40 .7646456 1.075206 -1.466251 2.154016

zscore\_gdp~h | 40 -.9196536 .941907 -4.726112 -.0591046

---------------------------------------------------------------------------------------------------------------

-> \_clus\_1 = 2

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

zscore\_peace | 54 .1280128 .9970312 -1.399865 2.964307

zscore\_legal | 54 -.3104442 .7627327 -2.242855 1.353394

zscore\_gdp~h | 54 .0615885 .3224037 -.8731493 .6437744

---------------------------------------------------------------------------------------------------------------

-> \_clus\_1 = 3

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

zscore\_peace | 27 .2646489 .8250797 -1.022821 2.637102

zscore\_legal | 27 -.5119206 .5688297 -1.775583 .633562

zscore\_gdp~h | 27 1.239273 .4108496 .6371436 2.005847

---------------------------------------------------------------------------------------------------------------

-> \_clus\_1 = .

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

zscore\_peace | 0

zscore\_legal | 0

zscore\_gdp~h | 0

\*/

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

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

egen min\_peace = min(zscore\_peace)

egen max\_peace = max(zscore\_peace)

capture drop RiskLevel

gen RiskLevel = "High" if mean\_peace == max\_peace

replace RiskLevel = "Low" if mean\_peace == min\_peace

replace RiskLevel = "Moderate" if RiskLevel=="" & mean\_peace !=.

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

\*What do you find?

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

\* Topic 3: Principal Components Analysis

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

\*20. Estimate the principal components

pca zscore\*

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

\* Hint: This is available from the pca output

corr equityriskpremium defaultspread f1 f2 f3 f4

\*21. Create new variables for each component using

predict f1 f2 f3 f4, score

reg defaultspread f1 f2 f3 f4

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

\*Factor 1