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

\* Homework 2

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

\* 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 2 on Canvas

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

\* Topic 1: Data Management in STATA

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

\* 1. Import the AssetReuturns file

clear

cd "C:\Users\haniu\OneDrive\Desktop\Deepa\Deepa\Fall second seven week\Business Appliation in ML\Homework 2"

\* C:\Users\haniu\OneDrive\Desktop\Deepa\Deepa\Fall second seven week\Business Appl

\* > iation in ML\Homework 2

import excel "AssetReturns.xlsx", sheet ("Sheet1") firstrow

\* (7 vars, 94 obs)

\* 2. Estimate the Annual Equity Risk Premium in two ways and create a new variable for each called ERP1 and ERP2

\* A: Annual S&P500 Return - Annual Treasury Bill

generate ERP1 = AnnualReturnSP500 - AnnualReturnTBills

\* B: Annual S&P500 Return - Annual Treasury Bonds

generate ERP2= AnnualReturnSP500- AnnualReturnTBonds

\* Which is the larger. Why?

sum ERP\*

\* Variable | Obs Mean Std. dev. Min Max

\*-------------+---------------------------------------------------------

\* ERP1 | 94 .0849362 .1982905 -.4615 .5162

\* ERP2 | 94 .0671074 .2108687 -.5665 .4927

\* 3. Rename your two new variables ERP\_Bills and ERP\_Bonds

ren ERP1 ERP\_Bills

ren ERP2 ERP\_Bonds

\* 4. Label the new variables Equity Risk Premium Bonds and Equity Risk Premium Bills

label variable ERP\_Bills "EquityRiskPremiumBills"

label variable ERP\_Bonds "EquityRiskPremiumBonds"

\* 5. Save your file under a new name: New\_Homework2

save "C:\Users\haniu\OneDrive\Desktop\Deepa\Deepa\Fall second seven week\Business Appliation in ML\Homework 2\New\_Homework2.dta"

\*file C:\Users\haniu\OneDrive\Desktop\Deepa\Deepa\Fall second seven week\Business Appliation in ML\Homework 2\New\_Homework2.dta saved

\* 6. Drop observations prior to 1940 and restimate the mean Equity Risk Premiums. Did they increase or decrease? Why?

drop if Year<1940

\* (12 observations deleted)

sum ERP\*

\* Variable | Obs Mean Std. dev. Min Max

\*-------------+---------------------------------------------------------

\* ERP\_Bills | 82 .0897963 .1726156 -.3792 .5162

\* ERP\_Bonds | 82 .0728915 .1905848 -.5665 .4927

\* The mean for both the variables have increased as most of the values dropped were negative thereby increasing the mean of the data. Also, the minumum value for ERP\_Bills has increased from -0.4615 to -0.3792 which shows that one of the values i.e.,-0.4615 was dropped thereby increasing the mean.

\* 7. Sort the data by inflation. Which year has the lowest inflation?

sort InflationRate

\* 8. Sort the data by inflation, with highest inflation first

gsort -InflationRate

\* 9. Merge the original dataset (Homework1) to a new dataset called RecessionDates

\*Is this a 1:1 merge, a m:1 merge, or a 1:m merge?

\*Why aren't some matched?

clear

import excel RecessionDates.xlsx, firstrow clear

ren year Year

merge 1:1 Year using New\_Homework2

\* Result Number of obs

\* -----------------------------------------

\* Not matched 74

\* from master 74 (\_merge==1)

\* from using 0 (\_merge==2)

\*

\* Matched 94 (\_merge==3)

\* -----------------------------------------

\* It is a 1:1 merge. the year column is common in both the files Recession Dates and Asset Return files and there is data missing for some years from the Asset Return file which is not present in the Recession Date file hence it will show blank values in the merge files which doesn't match for those years.

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

\* Topic 2: Statsistical Analysis in STATA

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

\* 10. Estimate the correlation between recessions and AnnualReturns on Corporate Bonds

correlate recession AnnualReturnCorporateBonds

(obs=94)

| recess~n A~eBonds

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

recession | 1.0000

Annua~eBonds | -0.0878 1.0000

\* 11. Run a linear regression with the y-Variable as Recession and the x-variables are: Annual Returns on S&P500, inflation, and Annual Returns on Real Estate

regress recession AnnualReturnSP500 InflationRate AnnualReturnRealEstate

\* Source | SS df MS Number of obs = 94

\*-------------+---------------------------------- F(3, 90) = 5.50

\* Model | 1.95353137 3 .651177124 Prob > F = 0.0016

\* Residual | 10.6528516 90 .118365018 R-squared = 0.1550

\*-------------+---------------------------------- Adj R-squared = 0.1268

\* Total | 12.606383 93 .135552505 Root MSE = .34404

\*-------------------------------------------------------------------------------

\* recession | Coefficient Std. err. t P>|t| [95% conf. interval]

\*--------------+----------------------------------------------------------------

\*AnnualRet~500 | -.5184188 .1860427 -2.79 0.006 -.8880251 -.1488126

\*InflationRate | -1.41147 1.080596 -1.31 0.195 -3.558263 .735322

\*AnnualRetur~e | -.8275749 .6921328 -1.20 0.235 -2.202618 .5474678

\* \_cons | .3003492 .0503616 5.96 0.000 .2002972 .4004013

\*-------------------------------------------------------------------------------

\* 12. Estimate the predicted values of the regression

predict Predictedrecession

\*(option xb assumed; fitted values)

\*(74 missing values generated)

\* 13. Estimate the residuals of the regression

predict Residualrecession, res

\* (74 missing values generated)

\* 14. Estimate the Mean-Squared Error

sum Residualrecession, detail

\* Residuals

\*-------------------------------------------------------------

\* Percentiles Smallest

\* 1% -.3522395 -.3522395

\* 5% -.3015367 -.3246011

\*10% -.2609816 -.3182504 Obs 94

\*25% -.1977083 -.3168103 Sum of wgt. 94

\*

\*50% -.1046123 Mean -5.55e-10

\* Largest Std. dev. .3384476

\*75% -.0237895 .8642056

\*90% .7152951 .8647521 Variance .1145468

\*95% .840659 1.014595 Skewness 1.801325

\*99% 1.102154 1.102154 Kurtosis 5.195504

\* 15. Run a Logistic Regression where the y-variable is Recession and the x-variable is Annual Return on Corporate Bonds and Inflation

logit recession AnnualReturnCorporateBonds InflationRate

\*Iteration 0: log likelihood = -41.262576

\*Iteration 1: log likelihood = -38.298118

\*Iteration 2: log likelihood = -38.141588

\*Iteration 3: log likelihood = -38.141098

\*Iteration 4: log likelihood = -38.141098

\*

\*Logistic regression Number of obs = 94

\* LR chi2(2) = 6.24

\* Prob > chi2 = 0.0441

\*Log likelihood = -38.141098 Pseudo R2 = 0.0756

\*

\*-------------------------------------------------------------------------------

\* recession | Coefficient Std. err. z P>|z| [95% conf. interval]

\*--------------+----------------------------------------------------------------

\*Annual~eBonds | -3.325744 3.999197 -0.83 0.406 -11.16403 4.512539

\*InflationRate | -17.8441 8.247193 -2.16 0.030 -34.0083 -1.679898

\* \_cons | -1.024286 .4196204 -2.44 0.015 -1.846727 -.2018453

\*-------------------------------------------------------------------------------