CARNEGIE MELLON UNIVERSITY DATA, INFERENCE & APPLIED MACHINE LEARNING (COURSE 18-785) ASSIGNMENT 3

INSTRUCTIONS

- Submissions should be made via canvas.
- **Single** Python/MATLAB code file(.ipynb or .m) [**Do not Submit checkpoints for .ipynb**]. In addition, each line of code should be documented by text. This demonstrates that the code is unique and owned by the student
- Assignment report(.pdf) with full evidence that the assignment was completed by the student and demonstrate a full understanding of each step in the process including textual descriptions of each result (statistics, table, graph etc) represents and insights that can be gained
- Indicate the libraries you have used in your code at the beginning of the report (After the title page)
- Data files (as given)

Submission process:

- 1. Put source code file and data files in a single folder
- 2. Name of the folder should be the same as your andrew ID
- 3. Zip this folder and attach the zipped file on assignment submission page (CANVAS)
- 4. After attaching zipped file, click on "Add Another File" from assignment submission page and attach your report
- 5. Submit your assignment

N.B. This process will allow us to compile your reports in **Turnitin** to check for plagiarism.

Specific reasons for a submission being classified as incomplete include:

- Failure to correctly name your folder with your Andrew ID, report, and code file with andrewID_DIAML_AssignmentNo. For example, mcsharry_DIAML_Assignment1, mcsharry_DIAML_Assignment2 and mcsharry_DIAML_Assignment3.
- A missing report describing the steps, results, and insights
- A missing dataset required for running the code
- A missing code file such as .ipynb or .m file
- An error in the file path needed to run the code

The student is responsible for checking that their submission is complete. Students will lose 10% as for late submission even if the submission is repaired during the 24 hours after the deadline has passed, and receive 0 for the assignment if it is not repaired.

The submission deadline is on Monday 03, October, 2022 17:59 Eastern Time (ET) /

Monday 03, October, 2022 23:59 Rwandan Time (CAT).

Question				Format	Value
Daily energy intake in kJ was measured for 11 women (Altman, 1991):				Six numbers	20%
5260, 5470, 5	We wish to investigate whether the women's energy intake deviates				
systematically from a recommended value of 7725 kJ. Assuming this data comes from a normal distribution; use a t-test to test whether the				answers.	
	na p-value. Fin	any exp	iain if the null hypothesis is rejected		
				Ovalitativa	20%
				l -	20%
				_	
				Code.	
	_				
Lisewhere	01	31	7.1		
Is this differe	nce of 74 versi	ıc 57 cia	nificant or is it simply due to		
one-sample, two-sample or paired test is appropriate. Show the steps of					
3 Use data from the World Bank Indicators for 2013 to study the				Graph,	20%
relationship between Fertility rate, total (births per woman) versus GDP				Correlation	
per capita PPP (current international \$). Make a carefully labelled graph with one dot per country. Estimate the correlation coefficient and				coefficient.	
				Interpretation.	
give your inte	erpretation.				
					20%
monthly indices (Post '91)). Graph the time series and label it carefully. Construct the autocorrelation function (ACF) of the monthly returns defined as $r(t) = [p(t)/p(t-1)]-1$ and show the values for lags of one up to 20 using a bar-graph. Indicate the values of the ACF using horizontal lines that would correspond to a statistically significant result at p<0.05. From the ACF of monthly data is there evidence of seasonality? Is there					
				1 ^	
				answers.	
		11at 15 tl	ic annuanzed return over uns period		
		from he	ere (ticker = ^FTSE) over the same	Graph.	20%
period (01-Jan-1991 to 31-Dec-2016). Plot the cumulative returns from				1 ~1 mp	
			6). Plot the cumulative returns from	Average	
period (01-Ja	ın-1991 to 31-I	Dec-2016		Average Annualized	
period (01-Ja the House ma	n-1991 to 31-I arket (using the	Dec-2016 price da	ata from question 4) and the	Annualized	
period (01-Ja the House ma FTSE100 ind	n-1991 to 31-I arket (using the lex on the same	Dec-2016 price da graph v	ata from question 4) and the with the time series normalized such	Annualized return.	
period (01-Ja the House ma FTSE100 ind that each star	n-1991 to 31-I arket (using the lex on the same ts at 100 in Jar	Dec-2016 price da graph v 1-1991.	ata from question 4) and the	Annualized	
	Daily energy 5260, 5470, 3 We wish to in systematicall data comes for distribution or right-tail or the standard deviation of freedom and or not. A Guinness of Guinness served elsewing GOES tasted elsewing the Location Ireland Elsewhere Is this different natural, rando one-sample, calculating the two-tailed test Use data from relationship to per capita PP graph with one give your into the defined as reconstruct the defin	Daily energy intake in kJ was 5260, 5470, 5640, 6180, 639. We wish to investigate whet systematically from a recomdata comes from a normal distribution might have a maright-tail or two-tailed test is standard deviation, standard of freedom and p-value. Fin or not. A Guinness Overall Enjoym Guinness served in an Irish served elsewhere around the mean GOES score of 74, whasted elsewhere was 57. The Location Sample Size Ireland 42 Elsewhere 61 Is this difference of 74 versus natural, random variation? Use data from the World Barelationship between Fertility per capita PPP (current integraph with one dot per coungive your interpretation. Load in monthly average how and the monthly indices (Post '91). Construct the autocorrelation defined as r(t) = [p(t)/p(t-1)] 20 using a bar-graph. Indicatines that would correspond From the ACF of monthly dated a trend in the time series? We as a percentage?	Daily energy intake in kJ was measure 5260, 5470, 5640, 6180, 6390, 6515. We wish to investigate whether the systematically from a recommended data comes from a normal distribution distribution might have a mean of 7 right-tail or two-tailed test is appropriated deviation, standard error of of freedom and p-value. Finally export not. A Guinness Overall Enjoyment Scot Guinness served in an Irish pub tast served elsewhere around the globe. In the final distribution in the globe of the globe. In the globe of	Daily energy intake in kJ was measured for 11 women (Altman, 1991): 5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770. We wish to investigate whether the women's energy intake deviates systematically from a recommended value of 7725 kJ. Assuming this data comes from a normal distribution; use a t-test to test whether the distribution might have a mean of 7725 kJ. Explain whether a left-tail, right-tail or two-tailed test is appropriate. Give the sample mean, sample standard deviation, standard error of the mean (SEM), t statistic, degrees of freedom and p-value. Finally explain if the null hypothesis is rejected or not. A Guinness Overall Enjoyment Score (GOES) was used to test if Guinness served in an Irish pub tastes significantly better than pints served elsewhere around the globe. Pints consumed in Ireland received a mean GOES score of 74, while the average GOES score for Guinness tasted elsewhere was 57. The full results were as follows: Location Sample Size Mean Standard Deviation Ireland 42 74 7.4 Elsewhere 61 57 7.1 Is this difference of 74 versus 57 significant, or is it simply due to natural, random variation? Use a t-test and explain whether a one-sample, two-sample or paired test is appropriate. Show the steps of calculating the t statistic and explain whether a left-tailed, right-tailed or two-tailed test is required. Give the resulting p-value. Use data from the World Bank Indicators for 2013 to study the relationship between Fertility rate, total (births per woman) versus GDP per capita PPP (current international \$). Make a carefully labelled graph with one dot per country. Estimate the correlation coefficient and give your interpretation. Load in monthly average house price data in pounds sterling (£) from Jan 1991 to Dec 2016. Download the data from here (choose the file UK monthly indices (Post '91)). Graph the time series and label it carefully. Construct the autocorrelation function (ACF) of the monthly returns defined as r(t) = [p(t)/p(t-1)]-1 and show the values for l	Daily energy intake in kJ was measured for 11 women (Altman, 1991): 5260, 5470, 5640, 6180, 6390, 6515, 6805, 7515, 7515, 8230, 8770. We wish to investigate whether the women's energy intake deviates systematically from a recommended value of 7725 kJ. Assuming this data comes from a normal distribution; use a t-test to test whether the distribution might have a mean of 7725 kJ. Explain whether a left-tail, right-tail or two-tailed test is appropriate. Give the sample mean, sample standard deviation, standard error of the mean (SEM), t statistic, degrees of freedom and p-value. Finally explain if the null hypothesis is rejected or not. A Guinness Overall Enjoyment Score (GOES) was used to test if Guinness served in an Irish pub tastes significantly better than pints served elsewhere around the globe. Pints consumed in Ireland received a mean GOES score of 74, while the average GOES score for Guinness tasted elsewhere was 57. The full results were as follows: Location Sample Size Mean Standard Deviation Ireland 42 74 7.4 Elsewhere 61 57 7.1 Is this difference of 74 versus 57 significant, or is it simply due to natural, random variation? Use a t-test and explain whether a one-sample, two-sample or paired test is appropriate. Show the steps of calculating the t statistic and explain whether a left-tailed, right-tailed or two-tailed test is required. Give the resulting p-value. Use data from the World Bank Indicators for 2013 to study the relationship between Fertility rate, total (births per woman) versus GDP per capita PPP (current international \$\$). Make a carefully labelled graph with one dot per country. Estimate the correlation coefficient and give your interpretation. Load in monthly average house price data in pounds sterling (£) from Jan 1991 to Dec 2016. Download the data from here (choose the file UK monthly indices (Post *91)). Graph the time series and label it carefully. Construct the autocorrelation function (ACF) of the monthly returns defined as r(t) = [p(t)/p(t-1)]-1 and show the values for l