Excel Assignment# 2

Due on Thursday 12/19/2024 at 12 noon (Pacific Time)

INSTRUCTIONS:

- This assignment accounts for 10% of the final grade.
- You will need to **submit the Excel file** containing all of your analyses and explanations. Type all your answers in the Excel file. Clearly label your answers with the question numbers, e.g., 1.1)_____
- When done, email the Excel file to me (sada@sfsu.edu). Please name your file "DS604 Excel #2-LastnameFirstname."
- You can not collaborate, consult, or receive help from anyone. If you have any questions, please ask me, not a classmate. A late or missed submission will be counted as a zero.

Criteria for grading the assignment:

⇒ Your grade will be based on the completeness and correctness of your answers.

⇒ When asked to explain, you need to provide a thorough explanation. Do not provide statistical values without interpreting their meaning. **Submitting only Excel's output or plots without narratives will not earn you any points**.

⇒ When asked to submit a plot, ensure that the axes are labeled properly.

⇒ Each question will have points assigned to it. Go through the questions and make sure to answer all of them thoroughly.

Data for this assignment are in the Excel file "Data for **DS604 Excel Assignment 2.xlsx**"

Problem 1: Mutual Savings Bank

In year 2002 there was considerable concern within the mutual savings banks because monthly changes in deposits were getting smaller and monthly changes in withdrawals were getting bigger. Thus it was of interest to develop a short-term forecasting model to forecast the changes in end-of-month (DEOM) balance over the next few months. Table 1 shows 53 monthly observations (February 2002 to June 2006) of the changes in end-of-month balance.

Also presented in the table are the composite AAA bond rates (AAA) and the rates on U.S. Government 3-4 year bonds (3-4), and the changes in the 3-4 year bonds (D3-4). It was hypothesized that these three variables had an influence on the DEOM balance figures in the bank.

Using the data in Table 1, perform the following tasks. Prepare and print out the Excel outputs and the results of the analyses and bring them to the in-class exam. During the in-class exam, you will be asked questions related to these analyses.

1.1. Develop a correlation matrix among all four variables (DEOM, AAA, 3-4, and D3-4). Explain what you observe from the correlation matrix.

The correlation matrix exhibits the following:

- With Dependent "DEOM" Variable
  - (AAA) Bond rates are moderately positively correlated – 0.26
  - (D3-4) Change in Gov't Bond Rates are very weakly negatively correlated – (0.19)
  - (3-4) Gov't bond rates are moderately negatively correlated – (0.39)
- With Independent Variables
  - (D3-4) is weakly negatively correlated to (AAA) – (0.2)
  - (3-4) is moderately positively correlated to (AAA) – 0.59
  - (D3-4) is weakly negatively correlated to (3-4) – (0.2)

1.2. Develop a regression model to forecast the DEOM figures based on the three independent variables (AAA, 3-4, and D3-4). Evaluate the model the strength of the model using the appropriate test and comment on the strength of the model (using α = 0.01).

n – k – 1 = 48
alpha = 0.01

CV(t) = 5.841
CV(F) = 4.218

	t Stat	ABS t		CV
Intercept	-1.3313957	1.331	<	5.841
AAA	6.06494291	6.065	>	5.841
D3-4	-2.2773305	2.277	<	5.841
3-4	-7.2693784	7.269	>	5.841

The t-stats of independent variables (AAA) Bond Yield and (3-4) Government Bond Yield successfully reject the null hypothesis by exceeding the critical value. This proves these variables are significant predictors in predicting variation of DEOM figures. On the other hand, the change in bond yield variable does not successfully reject the null hypothesis, proving no significant predictability.

		ABS t		CV
Overall model	20.8536464	20.854	>	4.218

Performing the F-test starts with identifying the observed F in the output report which is 20.854. This exceeds the critical value and successfully rejects the null hypothesis, supporting at least one of the variables is statistically significant and proves predictability of variation by the model for DEOM

Regression Statistics					
Multiple R	0.74885123				
R Square	0.56077817				
Adjusted R Sq	0.53388704				
Standard Erro	2.11700988				
Observations	53				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	280.38129	93.46043	20.85364644	7.5906E-09
Residual	49	219.60481	4.48173082		
Total	52	499.9861			

Looking at the adjusted r-squared, the output report displays a moderately positive correlation sitting at 0.56. In terms of predictability, this isn't very strong and indicates only 56% of the variation can be predicted by the model.

	Se1	Se2	Se3
Ranges	2.11700988	4.23401975	6.35102963
	-2.1170099	-4.2340198	-6.3510296
Frequency	37	14	2
	69.81%	26.42%	3.77%
	69.81%	96.23%	100.00%
	68%	95%	99.70%
Diff	1.81%	1.23%	0.30%

In the analysis of the standard error, the table displays a normal distribution of residual data based on the residual analysis. This proves that the model is well fitted to the actual data.

1.3. Perform the Durbin-Watson test using α = 0.01 to check if autocorrelation exists. Show the details of your test and explain the results

DW-Test	
D	12.69333
DU	1.505
DL	1.284

The D Critical Value exceeds the upper value. Therefore, it fails to reject the null hypothesis, and proves no autocorrelation exists. This validates the results of the F-test and t-test, supporting that the model possesses significant predictability

1.4. Create dummy (indicator) variables to allow monthly seasonal components to be included in the regression model. Develop a new regression model, which includes these dummy variables and the three independent variables. Evaluate the model and comment on the strength of the model compared to the one in part 2.

n – k – 1 = 37
alpha = 0.01

CV (t) = 2.977
CV (F) = 2.606

Regression Statistics	
Multiple R	0.94171961
R Square	0.88683582
Adjusted R Sq	0.84514375
Standard Erro	1.22022989
Observations	53

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	14	443.405582	31.6718273	21.2710927	9.38775E-14
Residual	38	56.5805177	1.48896099		
Total	52	499.9861			

Adding in the dummy variables generates the following output. This regression analysis has a strong positive correlation, sitting at 0.87. This high coefficient of determination is interpreted as 87% of the variation can be accounted for by the model.

	t Stat	ABS		CV
Intercept	-0.3745667	0.37456668	<	2.977
AAA	10.1579564	10.1579564	>	2.977
D3-4	-3.1270767	3.12707671	>	2.977
3-4	-12.024991	12.0249913	>	2.977
January	-1.6584623	1.65846231	<	2.977
February	-2.2693893	2.26938928	<	2.977
March	-7.2399424	7.23994242	>	2.977
April	-3.3137835	3.31378349	>	2.977
May	-5.5867731	5.58677312	>	2.977
June	-3.4495974	3.44959742	>	2.977
July	-5.1801834	5.18018341	>	2.977
August	-5.9326804	5.93268036	>	2.977
September	-7.0730386	7.07303858	>	2.977
October	-3.7891977	3.78919768	>	2.977
November	-4.5188513	4.51885132	>	2.977

Looking at the t-test chart, most independent variables, 86%, successfully reject the null hypothesis. This indicates significant predictability for the three original variables and for the months March through November. January and February however fail the t-test and fail to reject the null hypothesis, indicating these variables add no explanation to the variation.

		ABS F		CV
Overall model	0	21.271	>	2.606

In evaluating the overall significance of the model, the output F-value exceeds the critical value. This indicates significant predictability of DEOM by the model.

	Se1	Se2	Se3
Ranges	1.22022989	2.44045979	3.66068968
	-1.2202299	-2.4404598	-3.6606897
Frequency	41	11	1
	77.36%	20.75%	1.89%
	77.36%	98.11%	100.00%
	68%	95%	99.70%
Diff	9.36%	3.11%	0.30%

When observing the standard error distribution of the model, it can be noted that a greater amount of the data compared to the first model is skewed towards the middle, accounting nearly 80% of the standard errors within the first deviation. Since errors could be potentially being underestimated, a Durbin Watson test should be constructed to see if autocorrelation exists.

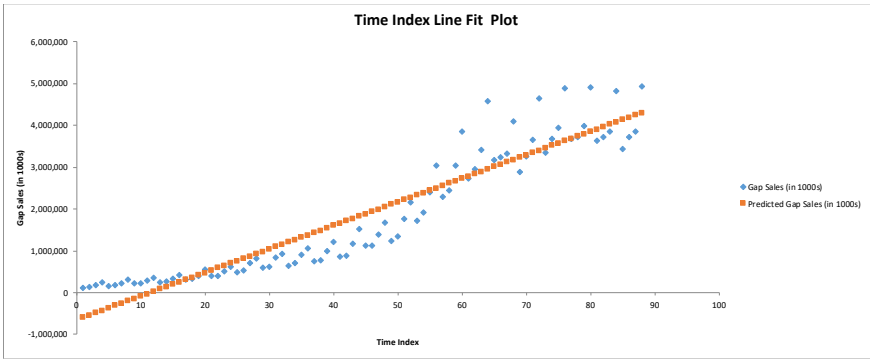
DW-Test	
D	2.43363946
DU	2.059
DL	0.863

Finally, when looking at the Durbin-Watson test, the D critical value exceeds the upper value. The test fails to reject the null hypothesis, proving no autocorrelation exists. This validates the t-test and F-test results, proving the model possess significant predictability in accounting the variation in DEOM.

Problem 2: Gap Sales

Table 2 presents the Gap's quarterly sales data for 1985 Q1 through 2006 Q4. Analyze the data and answer the following questions:

2.1. Create a time-series plot of the Gap Sales. Describe the pattern(s) of the time series you see in the graph.



The pattern of the actual data depicts a strong exponential trend with very little variability until t = 60 where the variation begins to take over the slope, becoming more significant as time progresses. It even seems to be leveling off, depicting an S-curve. This could indicate that Gap is suffering from a stagnation in sales growth and increased volatility.

2.2 Calculate the seasonal factors (SI) and then the seasonal indices (S) based on the 1985 Q1-2006 Q4 data. Are they consistent with your expectations? Explain in detail.

Qtr	SI
1	0.863516
2	0.875717
3	1.021998
4	1.237962

After calculating the seasonal factors, a seasonal index analysis can be performed and generates the following output. As seen in the data, quarters 1 (Jan, Feb, Mar) and 2 (April, May June) are below 1, indicating the sales values in these quarters are significantly lower on a seasonal basis. Quarters 3 (July, Aug, Sep) and 4 (Oct, Nov, Dec) is above 1.0, indicating sales values are significantly higher on a seasonal basis. These seasonal indices are expected because Gap operates in the consumer retail sector, which is cyclical. Much more shopping activity occurs in the latter half of the year as opposed to the former. From personal experience working for a company such as Vans, holiday promotions and back to school shopping drives the KPIs for these businesses in the latter half. The beginning of the year is always slow, especially with holiday returns, which can hurt the sales figures for beginning quarters due to loss in sales.