

Exam 01: Introduction to Forecasting and its Basic Tools

⚠ This is a preview of the published version of the quiz

Started: Feb 18 at 3:38pm

Quiz Instructions

What/Why/Prep?

The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 06, including your thorough reading up to the end of Chapter 3.2 of the book. **Note that simple exponential smoothing is NOT covered for this exam.**

Required Rmd File

[isa444exam01.Rmd](#)

General Guidelines:

- Individual exam. You are **NOT** allowed to share any materials or discuss this exam with your colleagues.
- This is **NOT** a timed exam.
- You do **NOT** have to complete this assignment in one sitting.
- I will not answer any questions pertaining to the exam to ensure fairness. If something is not clear, please state your assumptions.
- Proctorio is **NOT** required for this assignment;
 - Each student has an individualized exam -- but any communication with other students will result in a F for the graded assignment.
- You will need to have R installed or accessible and you are highly encouraged to have read up to the end of Chapter 3.2 of the book. For those of you who had issues with R, I would recommend using the Virtual Desktop and ensuring that all files are stored on the M Drive (or Google Drive).
- See [isa444exam01.html](#) for more details.

Correct Answers

N/A given that this is an individualized exam. You will get feedback on your exam when I grade it.

As a part of this exam, you are expected to document your analysis in this Markdown document [isa444exam01.Rmd](#). A knitted version of the HTML generated from the Markdown document is available [isa444exam01.html](#). **Please carefully read the instructions in Section 1.1-1.3. You should divide the R Markdown into 2 major sections.**

Section 2 should address all the questions related to the COVID data, which is generated in Section 1.2.

Section 3 should address the crypto related questions, which are generated based on the code highlighted in this exam (i.e., it is provided within a later question).

Question 1

5 pts

Based on the [isa444exam01.Rmd](#), the generated data frame titled **exam01DF** contains variables.

Question 2

5 pts

Based on the [isa444exam01.Rmd](#), the generated data frame titled **exam01DF** contains observations.

Question 3**5 pts**

Based on the [isa444exam01.Rmd](#), one can categorize the generated data frame titled **exam01DF** as a panel dataset.

☐ True

☐ False

Question 4**5 pts**

Based on the [isa444exam01.Rmd](#), the generated data frame titled **exam01DF** is grouped by the "id" variable.

☐ True

☐ False

Question 5**5 pts**

Based on the [isa444exam01.Rmd](#), create an interactive panel plot of your five counties based on your **covidExamDF** object. Based on the plot and interpretation, assess whether the data exhibits any seasonal pattern in the variable titled **newCases** (*note that it denotes the newCases aggregated for that week*).

☐ The data does NOT exhibit any seasonal patterns

☐ The data DOES exhibit a seasonal pattern

Question 6**5 pts**

Based on the [isa444exam01.Rmd](#), create an interactive panel plot of your five counties based on your **covidExamDF** object. Based on the plot and interpretation, assess whether the variable titled **newCases** can be assumed to be (weakly) stationary for all of your five counties.

☐ True☐ False**Question 7****5 pts**

In class, we have talked about measures for summarizing data. Building on the code below, compute the correlation between the newCases for `id == '5fc72c'` (i.e., Iowa, Story County) and the new Cases for `id = '09b187b8'` (Georgia, Richmond): **Note you will need to have the packages loaded.**

```
fadelData = exam01DF %>% filter(id %in% c("09b187b8", "25a4877c",  
"5fc72c", "76aae175", "af58129c")) %>%  
  select(id, key_google_mobility, population, date, confirmed) %>% # selecting  
  only five variables  
  mutate(newCases = confirmed - lag(confirmed))  
  
fadelData %<>% tq_transmute(select = newCases, mutate_fun =  
  apply.weekly, FUN = sum)
```

Note that each of them will contain a NA value at the beginning. Use the following argument inside the cor function (use = 'complete.obs') to remove the NAs from the calculation. Report the correlation value to the third decimal place (i.e. 0.XYZ).

Question 8**5 pts**

If you were to run this code (you would need to have the packages in the beginning of your Markdown Loaded):

```
fadelData = exam01DF %>% filter(id %in% c("09b187b8", "25a4877c",  
"5fcdf72c", "76aae175", "af58129c")) %>%  
  select(id, key_google_mobility, population, date, confirmed) %>% # selecting  
  only five variables  
  mutate(newCases = confirmed - lag(confirmed))  
  
fadelData %<>% tq_transmute(select = newCases, mutate_fun =  
  apply.weekly, FUN = sum)
```

Explain what does it mean that the county with id=25a4877c (i.e., Texas, Panola County) had a -26 on 2020-06-21? Please provide your interpretation/explanation as to why this is possible given your knowledge of COVID-19 reporting and differencing.

p



0 words

**Question 9****5 pts**

Based on the *fadelData* and without performing any calculations, which of the following methods would likely produce more accurate results if we were to forecast the number of newCases for Georgia, Richmond (i.e. id = '09b187b8') on 2021-02-20?

☐ Naive Forecast

☐ MA(4)

Question 10

5 pts

If you were to filter the *fadelData* to just Georgia, Richmond (id = '09b187b8'), what is the **95% upper prediction limit** for your naive forecast on 2021-02-20? **Please report your answer rounded to the second decimal place.**

You can build on the following code:

```
df = fadelData %>% filter(id=='09b187b8') %>% ungroup()
```

Question 11

5 pts

When I filtered the *fadelData* to just Georgia, Richmond (id = '09b187b8'), I computed the MAPE for the naive Forecast and it was 36.31% **Please comment on why this awful performance for the naive forecast is expected based on what you learned in class. Your answer should be limited to < 15 words.**

p



0 words



Question 12

5 pts

In class, we have talked about using smoothing with a moving average. In the context of COVID-19, it is common to plot such data using a 7-day moving average. For example, see [Texas Coronavirus Cases & Deaths by County | View the Map | USAFacts](https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/texas) (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/texas>). Using what you have learned in R, please replicate the "NEW KNOWN CASES PER DAY" chart using R. Below, you should include the R code that you used to replicate this chart. Additionally, at the end of this exam, you will be asked to include the HTML generated from your R Markdown with answers to all the questions.

Hint: You should start with the exam01DF and filter the Data to just Texas.

For the purpose of this answer, you should: (a) provide your code, and (b) upload a static figure of the result in the text box below.

p



0 words



Question 13

5 pts

In class, we have talked about using smoothing with a moving average. In the context of COVID-19, it is common to plot such data using a 7-day moving average. For example, see [Texas Coronavirus Cases & Deaths by County | View the Map | USAFacts](https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/texas) (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/texas>). If you were to focus on the "NEW KNOWN CASES PER DAY" chart, you would clearly see that such a practice may not always be the best. For example, on Dec 21, there were 184,937 new cases reported in Texas. If you compare this data point to all other days, one can conclude that this is likely due to a change in the data collection protocol. However, the impact of this data point on the 7-day moving average lasted an entire week (as shown in the pink line). **Propose an alternative smoothing strategy that would have eliminated the impact of this outlier on the smoothed curve. The smoothing strategy can be simple and will likely be something that we did not directly discuss in class, but builds logically on the class material and/or the previous prereqs.**

p



0 words



The following questions are based on extracting cryptocurrencies. **Please add it to your Markdown document. The code requires the same packages as those listed in the packages code chunk.**

The code below should be your starting point for any/all of these questions:

```
cryptoCurrencies = c('BTC-USD', 'ETH-USD', 'USDT-USD', 'BNB-USD', 'ADA-USD', 'DOT-USD', 'XRP-USD', 'LTC-USD', 'LINK-USD', 'BCH-USD', 'XLM-USD', 'DOGE-USD', 'USDC-USD', 'UNI-USD')
```

```
userID = "megahefm" # <-- replace my userID with your unique ID
```

```
seedNum = char2seed(userID, set = FALSE)
```

```
set.seed(seedNum); coinSymbol = sample(cryptoCurrencies, size=1, replace = F) %>% sort()
```

```
dfCrypto = tq_get(x = coinSymbol, from = '2020-11-02', to = '2021-02-17')
```

Question 14

5 pts

Based on your unique username, investigate whether there is a week-day effect in your data through a seasonal plot. You should have:

- the day on the x-axis (see the `wday()` function from `lubridate`);
- the closing price of the crypto on the y-axis; and

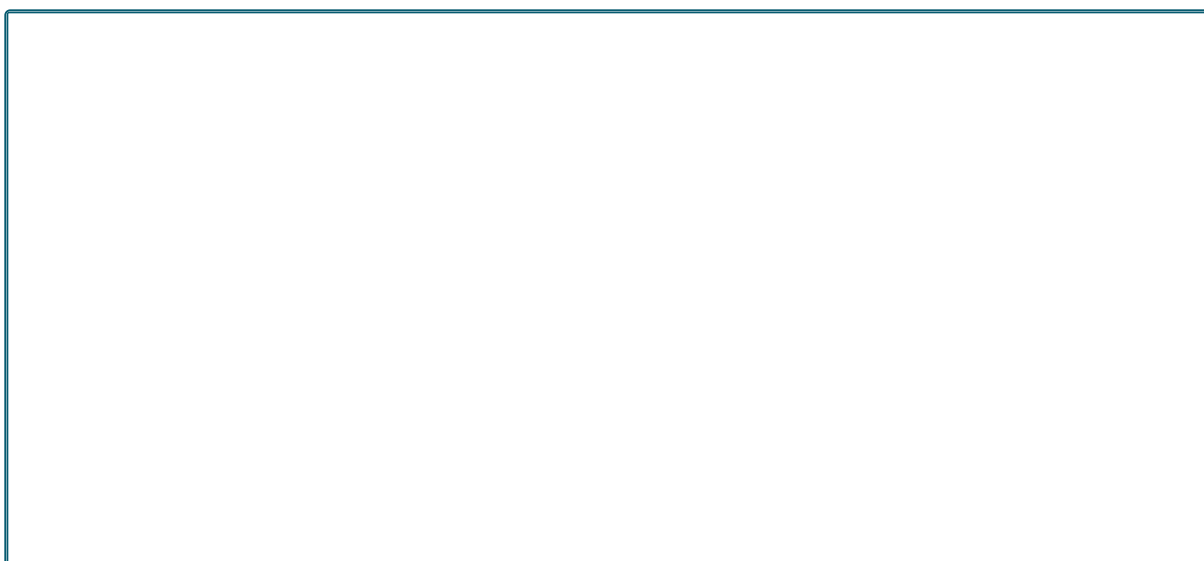
- the data grouped and colored by week (which should be a factor variable). You can capitalize on the week() from lubridate to extract the week number.

Your plot should be limited to only dates in 2021. **Based on your unique data and plot, assess whether there is a consistent day-of-the-week effect in your cryptocurrency. The text box below should contain both your plot and your text response.**




Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ |

 ▾  ▾  ▾  ▾ |  |  ▾  ▾  |   ▾  



p

  | 0 words |   

Question 15

5 pts

Based on your previous question, **propose a simple trading strategy that can be used for maximizing your profit. Without investigating all the charts, here are some possible strategies to chose from:**

- **Sell on Friday and buy on Sunday (or similar if there is evidence to suggest a dip of price over the weekend).**
- **Buy and hold (if the price seems to be going up in 2021 and there are no clear day-of-the-week effects).**

- **Sell (if the price seems to be going down in 2021)**
- **The price of the coin seems to be relatively steady and hence, there are no obvious choices for trading.**

Please pick one of the four strategies above. For those who want to learn more about their specific coin, please click on the name of the coin at: [Cryptocurrency Prices, Charts And Market Capitalizations | CoinMarketCap](https://coinmarketcap.com/) (<https://coinmarketcap.com/>) for more details

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ |

 ▾  ▾  ▾  ▾ |  |  ▾  ▾  |   ▾ \sqrt{x} 

p



0 words



Question 16

5 pts

Based on your data, **provide the code** that can be used to answer the following question:

What is the %change in the closing price of your stock over the 108 days of data?

The text box below should contain **the code** used for computation and the **answer to the question**.

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ T^2 ▾ | ▾  ▾  ▾  ▾ |  |  ▾  ▾  |   ▾ \sqrt{x} 

p



0 words

**Question 17****5 pts**

Using the same time frame (2020-11-02 up to and including 2021-02-17), what is the %change in the closing price of the S&P 500 (^GSPC)? Report your answer in percent and to the nearest two decimal places i.e., it should be something like YY.ZZ.

Question 18**5 pts**

What is the interpretation of the difference in the number of observations between the *dfCrypto* and the data frame that you used to acquire the S&P 500 data?

Edit View Insert Format Tools Table

12pt ∨ Paragraph ∨ | **B** *I* U A ∨  ∨ T^2 ∨ | ∨  ∨  ∨  ∨ |  |  ∨  ∨  |   ∨ \sqrt{x} 

p



0 words

**Question 19****5 pts**

Please upload your .Rmd file, with your code, commented and the file divided into three major sections (Section 1: provided by Fadel, Section 2 to cover the COVID questions, and Section 3 to cover the Crypto questions). The Rmd file should contain any calculations made by you. It should knit, without any errors, for you to receive credit for this question.

Upload

Question 20**5 pts**

Please upload the HTML generated from the Markdown. The file divided into three major sections (Section 1: provided by Fadel, Section 2 to cover the COVID

questions, and Section 3 to cover the Crypto questions). The Rmd file should contain any calculations made by you. It should knit, without any errors, for you to receive credit for this question.

Upload

Choose a File

Quiz saved at 3:38pm

Submit Quiz