Lab Report #1: Statistical Properties of Stock Returns

Course: RSM338H1S

Term: Winter 2026

Instructor: Professor Kevin Mott

Instructions:

- This assignment is to be completed **individually**.
- Submit your lab report as a **PDF** to Crowdmark via Quercus. Export your Jupyter Notebook to PDF before uploading.
- There is no page limit, but be concise. A good report is thorough but not padded.

Marking:

- **75%** — Coding and results (correct implementation, complete answers to all parts, appropriate choice of methods, accurate numerical output, properly formatted tables and figures)
- **25%** — Overall quality (clear and professional writing, thoughtful interpretation of results, demonstrated understanding of the underlying concepts, logical flow and narrative structure)

Writing Expectations: Your report should read as a **coherent narrative**, not just code with scattered comments. Use section headers to notate which question you're working on. Before each code block, briefly explain what you are about to do and why. After results appear, interpret what you see. A reader should be able to understand your analysis even if they skipped the code cells. Use a professional tone, and make sure to use connective sentences that link the exercises and your analysis.

You may use AI coding assistants (ChatGPT, Copilot, Claude, etc.) to help write code, but you must be able to explain what every line does. The text you write around the code is what demonstrates your understanding. You are ultimately responsible for your own work. **If you use an AI tool, you must disclose in a note at the end of your report. Mention which tool you used, which tasks you asked it to complete, and discuss your (dis)satisfaction with its assistance.**

Assignment: You are a junior quantitative analyst at an asset management firm. The portfolio managers are reviewing their exposure to small-cap versus large-cap equities and have asked you to prepare a report on the statistical properties of stock returns across different size categories. They want to understand: How do returns vary by firm size? Are returns normally distributed, or do they exhibit fat tails and skewness? Is there evidence of seasonality (such as the "January effect")? Your analysis will inform their asset allocation decisions.

Data: The data for this assignment come from Kenneth R. French's Data Library:

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

This is a canonical source for empirical asset pricing research. You are encouraged to explore the website to understand how the portfolios are constructed, what the variables mean, and what other data are available. The documentation pages for each dataset are particularly useful.

You are provided with three files:

- `Portfolios_Formed_on_ME.txt` — monthly returns for portfolios sorted by market equity (firm size), including 10 size deciles
- `F-F_Research_Data_Factors.txt` — monthly Fama-French factor returns
- `F-F_Research_Data_Factors_daily.txt` — daily Fama-French factor returns

Problem 1: Monthly Size Portfolio Returns

For this problem, use the **monthly** data files.

Data Preparation:

(i) Read the documentation on the French Data Library website to understand how these portfolios and factors are constructed.

(ii) Load both files into pandas DataFrames. The raw files have a particular format—you will need to identify where the relevant data begins, parse the columns appropriately, and assign meaningful column names.

(iii) From the size portfolios file, extract the **value-weighted** returns for the **10 size decile portfolios** (Lo 10 through Hi 10).

(iv) From the factors file, compute the **market return** as $r_{\text{mkt}} = \texttt{Mkt-RF} + \texttt{RF}$.

(v) Merge these into a single DataFrame with 11 return series: the market and the 10 size deciles.

(vi) **Convert to log returns.** The raw data are simple (arithmetic) returns expressed in percentage points. Convert them to log returns using:

$$r_{\text{log}} = \ln\left(1 + \frac{r_{\text{simple}}}{100}\right)$$

Use log returns for all subsequent analysis.

Tasks:

(a) Compute the following summary statistics for the monthly returns of all 11 portfolios (market and 10 size deciles):

- Mean
- Standard deviation
- 25th, 50th (median), and 75th percentiles

Present your results in a well-formatted table. For the 10 size deciles (excluding the market), plot mean returns against decile number and include a line of best fit. Is there a systematic relationship between firm size and average returns?

(b) Compute the coefficients of skewness and excess kurtosis for each of the 11 portfolios. For each portfolio, perform an appropriate test of whether the returns are normally distributed and report whether you reject normality at the 5% significance level.

(c) Repeat parts (a) and (b), but separately for:

- Returns in the month of **January only**
- Returns in **all other months** (February–December)

Do you observe any differences? In which size deciles are the differences most pronounced? This phenomenon is known as the "January effect." Briefly discuss what you find and offer a hypothesis for why this pattern might exist.

(d) The portfolios in this dataset are **value-weighted**, meaning larger firms within each decile receive more weight. The French Data Library also provides equal-weighted versions. In 2–3 sentences, explain the difference between value-weighted and equal-weighted portfolio returns, and discuss which might be more relevant for a typical investor.

Problem 2: Constructing Weekly Returns from Daily Data

For this problem, use the **daily** data file.

Data Preparation:

(i) Load the daily factors file into a pandas DataFrame. As with the monthly files, the raw file has header text before the data begins—you will need to identify where the data starts and parse accordingly.

(ii) The date column is in `YYYYMMDD` format (e.g., 19260701 for July 1, 1926). Parse this into a proper datetime index.

(iii) Compute the **market return** as $r_{\text{mkt}} = \texttt{Mkt-RF} + \texttt{RF}$, just as in Problem 1.

(iv) **Convert to log returns** using the same formula as before:

$$r_{\log} = \ln\left(1 + \frac{r_{\text{simple}}}{100}\right)$$

(v) Check your data: you should have daily observations from 1926 through the present.

Tasks:

(a) Construct a weekly return series from the daily market returns. Since you are working with log returns, the weekly log return is simply the sum of the daily log returns within that week:

$$r^{\log}_{\text{weekly}} = r^{\log}_1 + r^{\log}_2 + \cdots + r^{\log}_k$$

This is one of the key advantages of log returns—they aggregate additively over time.

Define weeks however you find sensible (e.g., Monday–Friday, or Wednesday–Tuesday). Some weeks will have fewer than 5 trading days due to holidays. You must decide how to handle these weeks: include them as-is, exclude them, or something else. Briefly justify your choice in your report.

(b) Compute the following summary statistics for your weekly market returns:

- Mean
- Standard deviation
- 25th, 50th (median), and 75th percentiles
- Skewness and excess kurtosis

Test whether weekly returns are normally distributed.

(c) Compare the distributional properties of weekly market returns (from this problem) to monthly market returns (from Problem 1). Are weekly returns "more normal" or "less normal" than monthly returns? What does the Central Limit Theorem suggest about how normality should change as the return horizon increases?

