

# Introduction to Statistics, Finance : A Case Report

Lectures in Descriptive Statistics and Statistical Analysis , Technical University of Denmark

Oriade Latifah Simpson

Challenge Accepted May 1, 2023

Write a short description of the data. Which variables are included in the dataset? Are the variables quantitative or categorised? How many observations are there? Which time period is covered by the observations? What is the date of the first and last observations? Are there any missing values?

An Exchange Trade fund is a pooled investment security, that works like a mutual fund.

The subset data consist of weekly returns of four ETFs and their return date. This dataset consists of 454 observations (rows) and 5 variables (columns). The column names are: t, AGG, VAW, IWN and SPY. The column entitled t contains dates which are categorical variables and the other columns contain quantitative random variables. The time span runs from the 5<sup>th</sup> of May 2006 until the 8<sup>th</sup> of May 2015. There are no missing values in the data frame.

Make a density histogram of the weekly returns from the ETF called AGG.

- Use this histogram to describe the empirical distribution of the returns.
- Is the empirical density symmetrical or skewed? Can the returns be both positive and negative? Is there much variation to be seen in the observations?

The histogram of the empirical density is somewhat symmetrical around zero. It is true that the returns can be either positive or negative. The coefficient of variation is high at around 22.48 which shows there is a lot of variation in the observations, (See Figure 1).

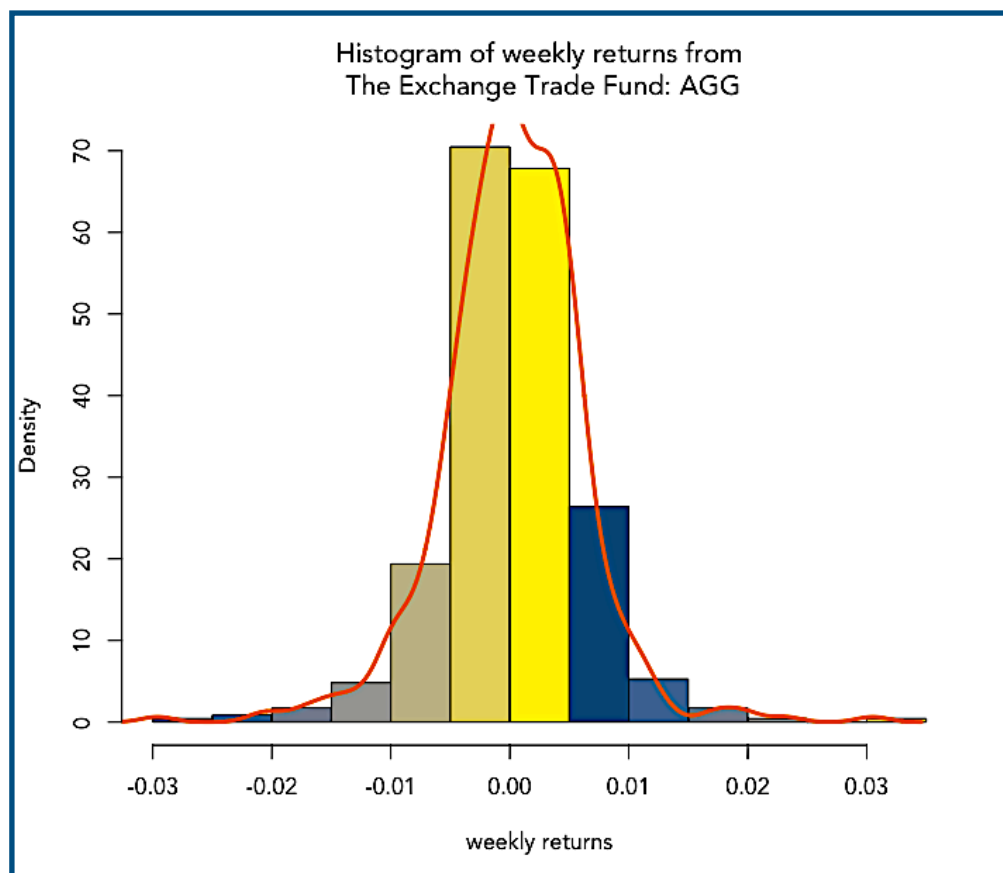


Figure 1

Make plots illustrating the weekly return over time for each of the four ETFs.

- Use these plots to describe the development of weekly returns in words.
- Does the level of return seem to change over time? Are there specific periods of time during which the weekly returns are notably different? Are there overall differences in the returns from the different ETFs?

The plots show that the weekly return is centered about zero for all four funds. The level of return fluctuates at the time points around 2009 and 2012 for the VAW and IWN funds , (See Figure 2).

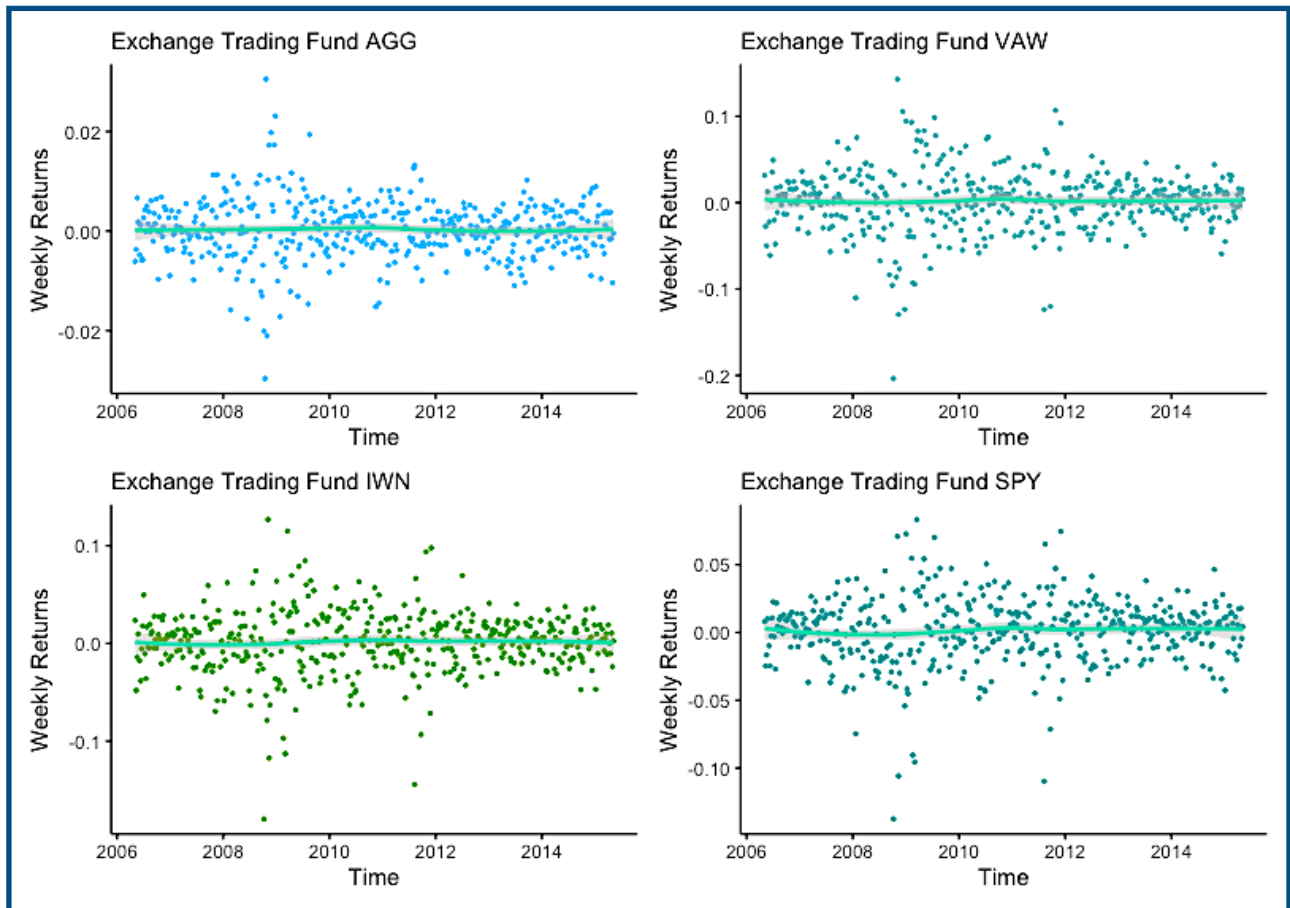


Figure 2

Make a boxplot of the weekly returns by ETF.

Use this plot to describe the empirical distribution of the weekly returns from each of the four ETFs. Are the distributions symmetrical or skewed? Are there differences between the distributions? If so, what are the differences? Are there extreme observations?

There are extreme observations for all four funds. The VAW and IWN funds show similarity in the spread of their distribution. The distribution of the SPY weekly return and VAW weekly return is slightly negatively skewed as the median is closer to the top of the box. The other distributions appear to be symmetrical, (see Figure 3).

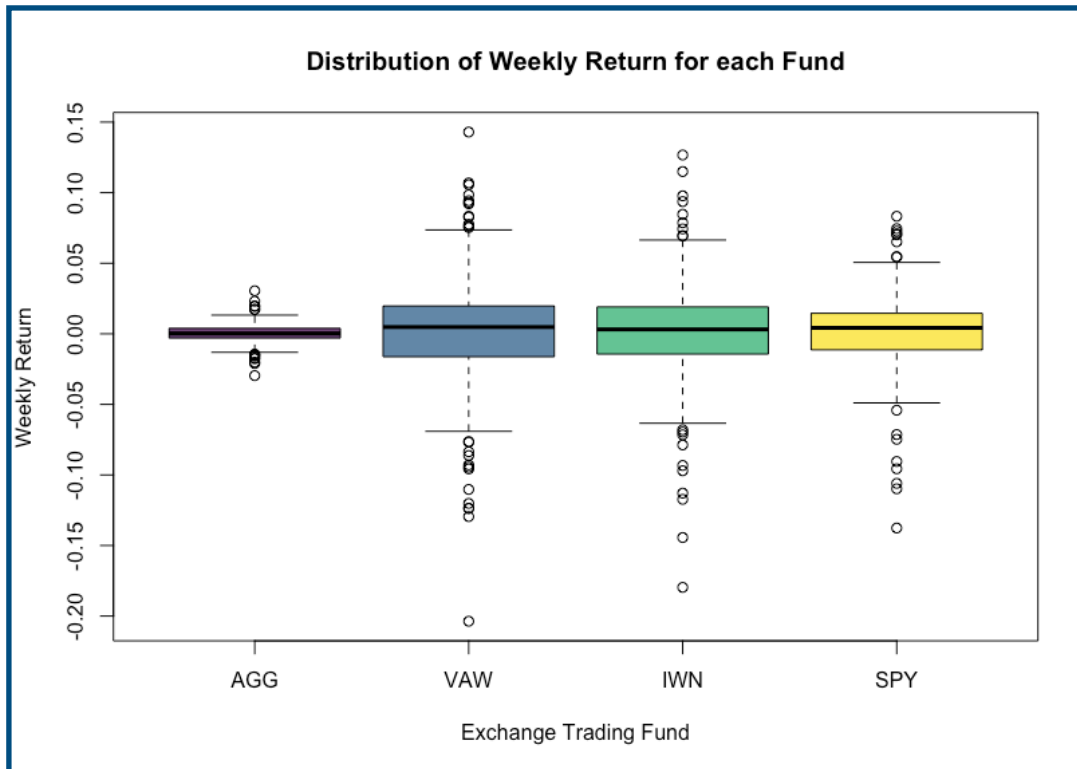


Figure 3

- Fill in the empty cells in the table by computing the relevant summary statistics for the four ETFS.
- Which additional information may be gained from the table compared to the boxplot?

The additional information to be gained from the table are the sample standard deviation values and sample variance values.

ETF	Number of Observations	Sample Mean	Sample Variance	Standard Deviation	Lower Quartile	Median	Upper Quartile
AGG	454	0.00027	0.0000	0,006	-0.0029	0.00237	0.00389
VAW	454	0.00179	0.0013	0.0036	-0.0162	0.00479	0.0197
IWN	454	0.001188	0,001	0,032	-0.01434	0.003119	0.01906
SPY	454	0.00136	6e -04	0,025	-0.011357	0.00421	0.01452

Specify separate statistical models describing weekly return for each of the four ETFs.  
Estimate the parameters of the models.

The model describing the weekly return for each of the four ETFs is  $X \sim N(\mu, \sigma^2)$

ETF	Parameter Mean	Parameter Standard deviation
<b>AGG</b>	0.0002551747	0.0002794381
<b>VAW</b>	0.001791923	0.0001650772
<b>IWN</b>	0.01188579	0.001553648
<b>SPY</b>	0.001363392	0.001148547

State the formula for a 95% confidence interval of the mean weekly return from AGG.  
Insert values and calculate the interval. Compute the corresponding intervals for the remaining ETFs and fill in the table below.

$$\hat{\mu} = \bar{x} \pm t \frac{1 - \alpha}{2} \cdot \frac{s}{\sqrt{n}}$$

ETF	Lower Bound of CI	Upper bound of CI
<b>AGG</b>	-0.0002854073	0.0008169213
<b>VAW</b>	-0.001534208	0.005121788
<b>IWN</b>	-0.001765174	0.004140533
<b>SPY</b>	-0.00092596	0.003646171

Perform a hypothesis test to investigate the mean weekly return from AGG.

There is no evidence against the null hypothesis that the mean weakly return from AGG is equal to zero. The p-value is listed as 0.3439. This is more than the 0.05 significance level.

Perform a hypothesis test to investigate the mean weekly return difference between two ETFs. Use VAW and AGG.

The null hypothesis is the mean weekly return of VAW is equal to the mean weekly return of AGG. The t-statistic is 0.94757. There is no evidence against the null hypothesis. The p-value is 0.3738 which is above the significance level  $\alpha$ .

Comment on whether it was necessary to carry out the statistical test in the previous question, or if the same conclusion could have been drawn using confidence intervals from before.

If the confidence interval for a difference between Exchange Trade Fund includes zero, it means that there is a high chance of finding no difference between mean weekly returns of each Exchange Trade Fund.

k) State the formula for computing the correlation between two Exchange Trade Funds; VAW and IWN. Determine the values and create a scatterplot illustrating the returns for the two Exchange Trade Funds.

$$r = \frac{1}{n - 1} \cdot \sum_{i=1}^n \frac{(x_i - \bar{x})}{sx} \cdot \frac{(y_i - \bar{y})}{sy}$$

The correlation between Exchange Trade Funds VAW and IWN is 0.8516407. There are many points clustered around zero, however there are extreme points in the bottom left corner of the scatterplot, (see figure 4).

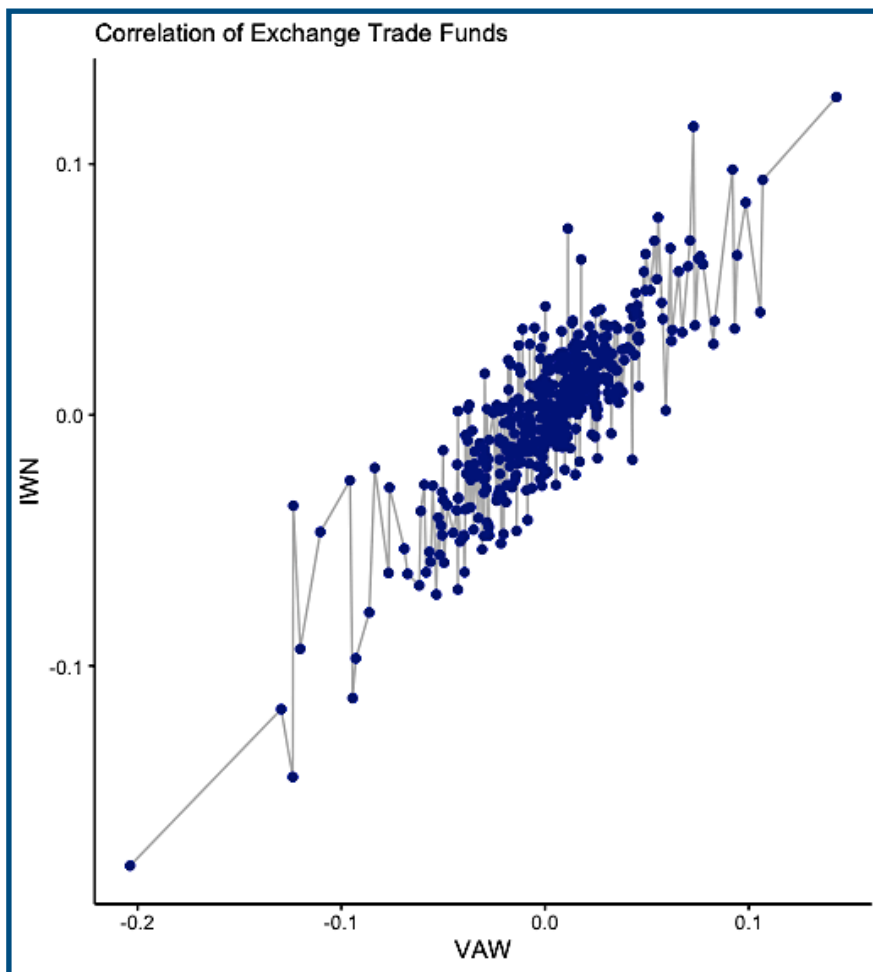


Figure 4