

5 - Standard Deviation as a Measure of Dispersion

Objective

Showing how to use the frequency graphs that we have learned, to plot the daily returns from the IVV etf.

Loading the libraries and the Data

```
library(tidyverse)
library(readxl)
library(ggplot2)
library(readxl)

df <- read_excel('IVV.xlsx')
df

## # A tibble: 5,449 x 2
##   Date                Close
##   <dtm>              <dbl>
## 1 2000-05-22 00:00:00 -0.00622
## 2 2000-05-23 00:00:00 -0.0152
## 3 2000-05-24 00:00:00  0.0150
## 4 2000-05-25 00:00:00 -0.00917
## 5 2000-05-26 00:00:00 -0.00451
## 6 2000-05-30 00:00:00  0.0329
## 7 2000-05-31 00:00:00  0.00285
## 8 2000-06-01 00:00:00  0.0160
## 9 2000-06-02 00:00:00  0.0198
## 10 2000-06-05 00:00:00 -0.00655
## # ... with 5,439 more rows
```

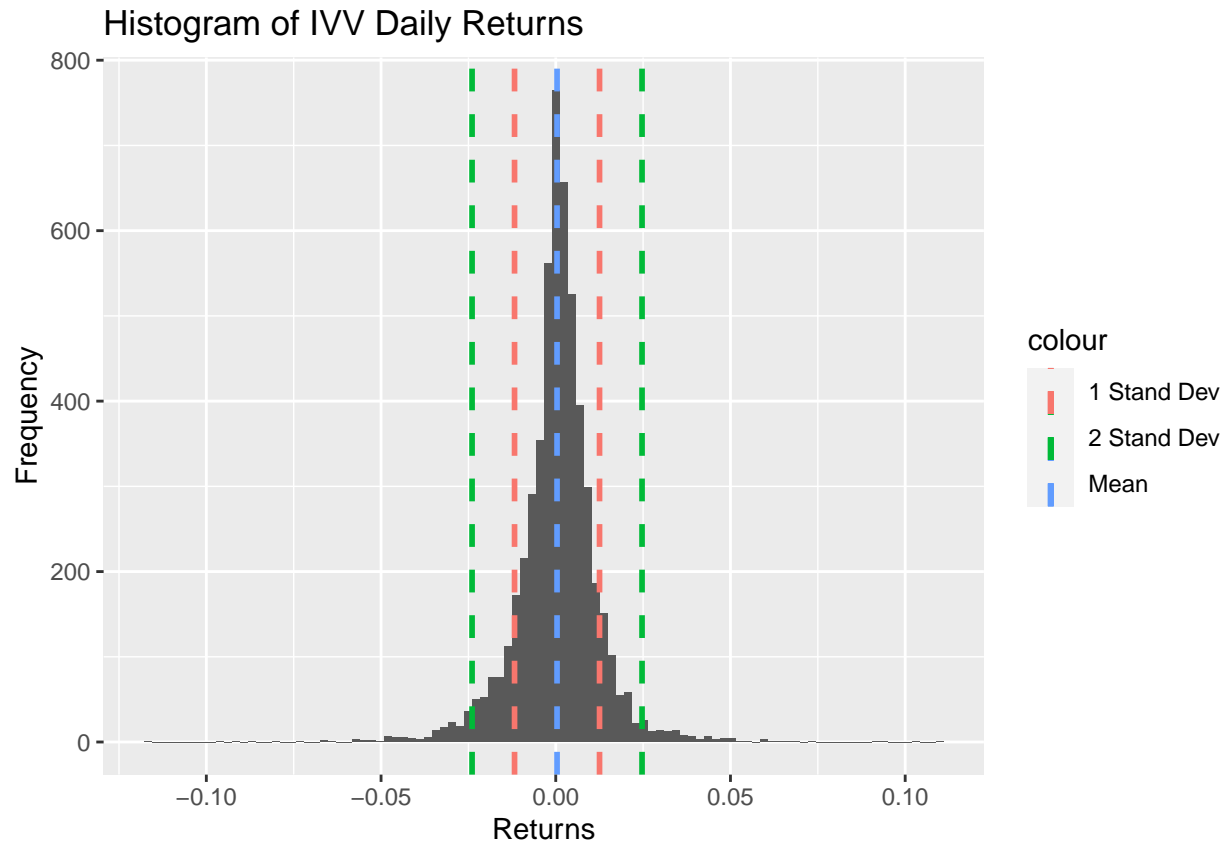
Histogram

```
p <- ggplot(df, aes(Close))

mean_ = mean(df$Close)
sd_ = sd(df$Close)

p + geom_histogram(bins = 100,) +
  ggtitle("Histogram of IVV Daily Returns") +
  ylab("Frequency") +
  xlab("Returns") +
  geom_vline(aes(xintercept = mean_, color = "Mean" ),linetype="dashed", size=1) +
```

```
geom_vline(aes(xintercept = mean_+sd_, color = "1 Stand Dev"),linetype="dashed", size=1) +
geom_vline(aes(xintercept = mean_-sd_, color = "1 Stand Dev"),linetype="dashed", size=1) +
geom_vline(aes(xintercept = mean_-2*sd_, color = "2 Stand Dev"),linetype="dashed", size=1) +
geom_vline(aes(xintercept = mean_+2*sd_, color = "2 Stand Dev"),linetype="dashed", size=1)
```

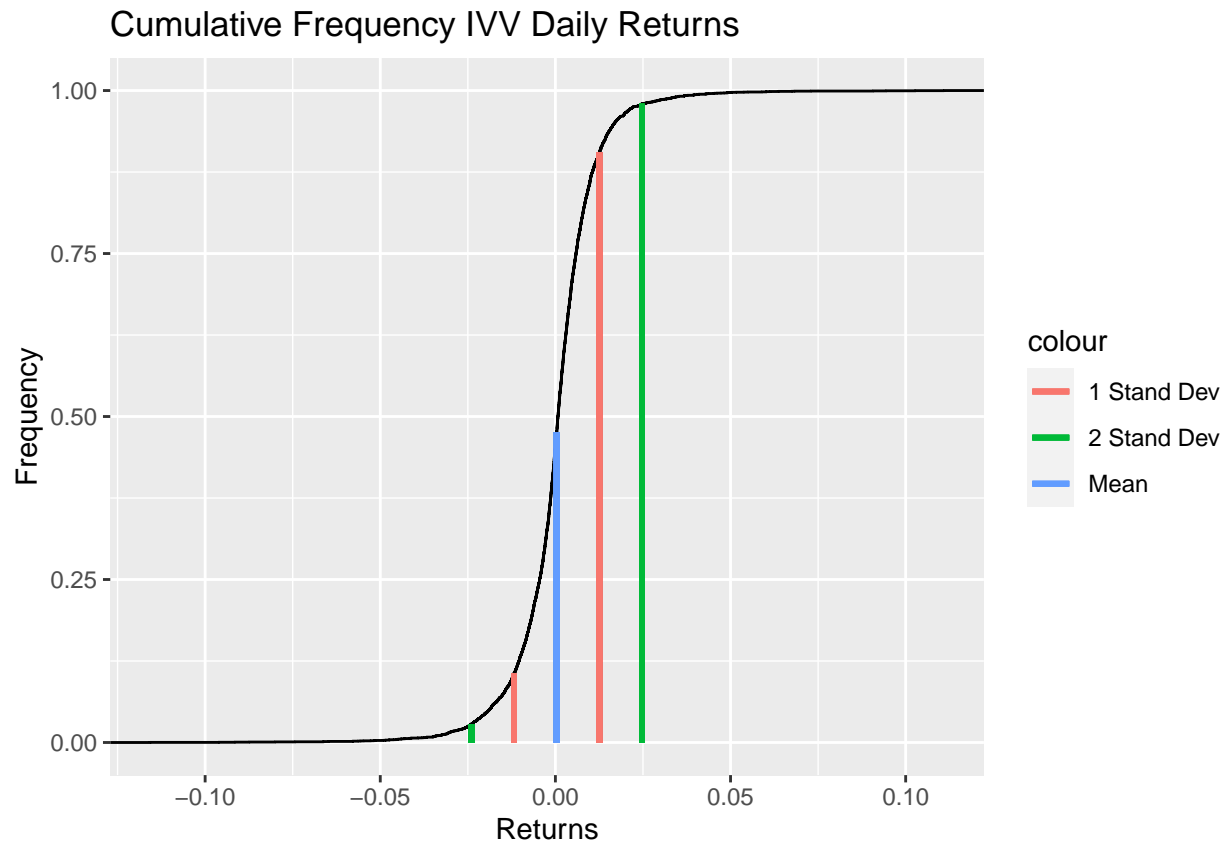


Cumulative Distribution

```
cdf = ecdf(df$Close)

p + stat_ecdf(geom = "step") +
  ggtitle("Cumulative Frequency IVV Daily Returns") +
  ylab("Frequency") +
  xlab("Returns") +
  geom_segment(aes(x = mean_, y = 0, xend = mean_, yend = cdf(mean_),
    colour = "Mean"), data = df, size=1)+
  geom_segment(aes(x = mean_+sd_, y = 0, xend = mean_+sd_,
    yend = cdf(mean_+sd_), colour = "1 Stand Dev"),
    data = df, size=1)+
  geom_segment(aes(x = mean_-sd_, y = 0, xend = mean_-sd_,
    yend = cdf(mean_-sd_), colour = "1 Stand Dev"),
    data = df, size=1) +
  geom_segment(aes(x = mean_+2*sd_, y = 0, xend = mean_+2*sd_,
    yend = cdf(mean_+2*sd_), colour = "2 Stand Dev"),
```

```
data = df,size=1)+
geom_segment(aes(x = mean_-2*sd_, y = 0, xend = mean_-2*sd_,
                yend = cdf(mean_-2*sd_), colour = "2 Stand Dev"),
data = df, size=1)
```



Numerical Evaluation

```
#How much of the data is between 1 Stand Deviation?
print(cdf(mean_+sd_) - cdf(mean_-sd_))
```

```
## [1] 0.7992292
```

```
#How much of the data is between 2 Stand Deviation?
print(cdf(mean_+2*sd_) - cdf(mean_-2*sd_))
```

```
## [1] 0.9511837
```

Boxplot

```
boxplot(df$Close)  
title("Boxplot of IVV Historical Returns")
```

