# Quantitative Investment Analysis Midterm

*Refik Türkeli*

*25/11/2018*

## Contents

# 1 Question 1

Bayer Pharmaceuticals (BP) has branched out and is now promoting a vaccine which is supposed to prevent statisticitis, a nasty disease afflicting students at a University. They are trying to convince the university's health center that the vaccine is worth stocking. The campus health center has agreed to stock and administer the vaccine if it sells well enough, at least `mu` = 50 vaccines per week. VP arranges for the health center to conduct a 40 week trial with the goal of convincing the center to stock the vaccine. They find that the center has administered an average of X = 55 vaccines per week over that period with a std. dev. of s = 7.99.

```
sample_mean <- 55
sample_sd <- 7.99
sample_n <- 40
conf_int <- 0.95
null_value <- 50
```

## 1.1  a.

### 1.1.1  Q. Compute a 95% confidence interval for `mu`, the mean number of vaccines the health center will administer per week in the long run. Does it look like the health center will be willing to stock the vaccine? Explain.

### 1.1.2  Answer

```
standard_error <- sample_sd / sqrt(sample_n)
significance <- 1 - conf_int
upper_z <- qnorm(significance / 2, lower.tail = FALSE)
scaled_z <- upper_z * standard_error
min_mean <- sample_mean - scaled_z
max_mean <- sample_mean + scaled_z
```

$$\sigma_{\bar{X}} = 1.26$$
$$z_{0.025} = 1.96$$
$$\mu = 55 \pm 2.48$$
$$52.52 \leq \mu \leq 57.48$$

It looks like health center will stock the vaccine because the lower point of sales estimate is higher than the required minimum of 50.

## 1.2  b.

### 1.2.1  Q. Perform the hypothesis test using test statistics that BP will conduct to convince the health center the vaccine is worth stocking. In other words, give the null and alternative hypotheses, both mathematically and in words, and explain your reasoning.

### 1.2.2  Answer

$$H_0 : \mu < 50$$
$$H_a : \mu \geq 50$$

Null hypothesis is that the vaccine won't sell enough worth stocking. Alternative hypothesis is that it will sell more than the required minimum of 50

```
test_statistic <- (sample_mean - null_value) / standard_error
rejection_point <- qnorm(significance, lower.tail = FALSE)
rejected <- test_statistic > rejection_point
```

$$\text{Test Statistic} = 3.96$$
$$\text{Rejection Point} = 1.64$$

Test statistic is larger than the rejection point. Null hypothesis is rejected. This means the vaccine will sell enough and is worth stocking.

## 1.3   c.

### 1.3.1   Q. Conduct the hypothesis test using p-value and state your real-conclusions assuming you need to be 95% sure the vaccine is worthwhile before you stock it. Is this consistent with your result from (b)?

### 1.3.2   Answer

```
p_value <- pnorm(test_statistic, lower.tail = FALSE)
```

$$\alpha = 5\%$$
$$p = 0.0038\%$$

p value is less than the significance level meaning that the chances of getting the result from experiment by chance is lower than the accepted level of error. This is consistent with the result from b.

## 1.4 Extra

```
m <- null_value; se <- standard_error; min_m <- m - 5 * se; max_m <- m + 5 * se
reject_m <- m + rejection_point * se; test_m <- m + test_statistic * se
ggplot(tibble(x = c(min_m, max_m)), aes(x)) +
    ylab("P(mean)") + xlab("mean") +
    stat_function(fun = dnorm, args = list(mean = m, sd = se)) +
    stat_function(fun = dnorm, args = list(mean = m, sd = se), geom = "area", fill = "red",
                  alpha = 0.5, xlim = c(reject_m, max_m)) +
    geom_vline(xintercept = test_m, color = "blue") +
    scale_x_continuous(breaks = round(c(min_m, m, reject_m, test_m, max_m), 2))
```
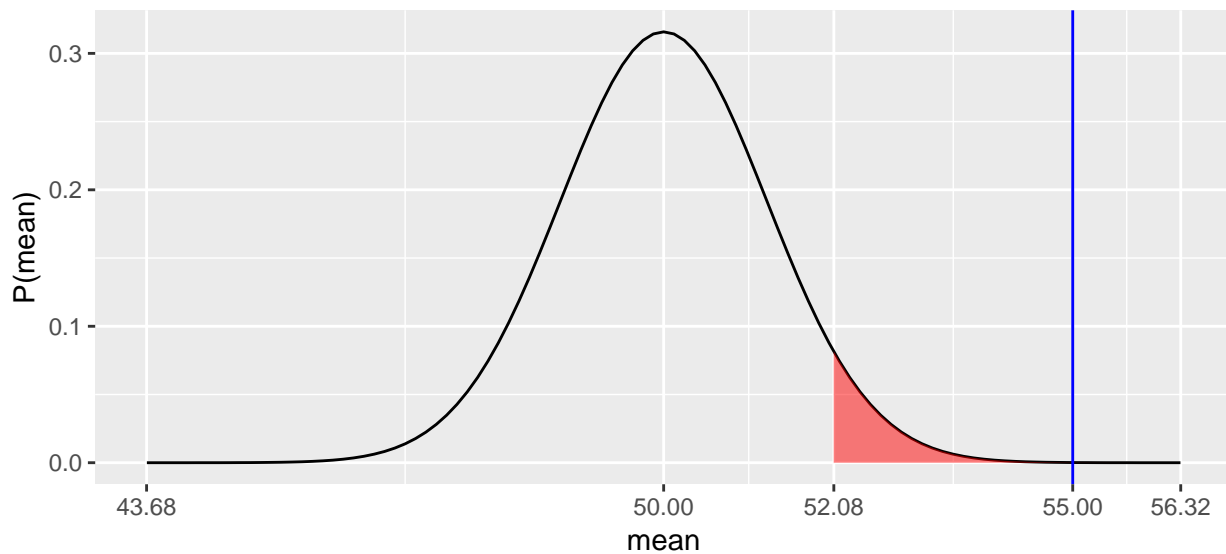


Red area is the rejection zone according to the confidence interval. In this case it is 5%. Our sample is on the blue line. Since it is inside the red area, we reject the null hypothesis. The area to the right side of the blue line is the p value. It is too small to be noticed on this plot.

## 2   Question 2

A car manufacturer receives its air conditioning units from 3 different suppliers, A; B; and C. 20% of its units come from supplier A, 30% from supplier B and 50% from supplier C. It is known that 10% of the units from supplier A are defective, 8% of units from supplier B are defective and 5% of units from supplier C are defective. If a unit is selected at random and is found to be defective, what is the probability that it came from:

```
supply <- matrix(c(0.2, 0.3, 0.5, 0.1, 0.08, 0.05), nrow = 3, ncol = 2,
                 dimnames = list(c("A", "B", "C"), c("Weight", "Defective")))
matrix_perc(supply)
```

|   | Weight | Defective |
|---|--------|-----------|
| A | 20%    | 10%       |
| B | 30%    | 8%        |
| C | 50%    | 5%        |

## 2.1  a.

### 2.1.1  Q. Supplier A? Write probability functions and explain your answer.

### 2.1.2  Answer

```
p_defective <- sum(supply[,"Weight"] * supply[,"Defective"])
p_a_and_defective <- supply["A", "Weight"] * supply["A", "Defective"]
p_a_given_defective <- p_a_and_defective / p_defective
```

$$P(Defective) = 6.9\%$$
$$P(A \cap Defective) = 2\%$$
$$P(A \mid Defective) = 28.99\%$$

Because:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

## 2.2  b.

### 2.2.1  Q. Supplier C? Write probability functions and explain your answer.

### 2.2.2  Answer

```
p_c_and_defective <- supply["C", "Weight"] * supply["C", "Defective"]
p_c_given_defective <- p_c_and_defective / p_defective
```

$$P(C \cap Defective) = 2.5\%$$
$$P(C \mid Defective) = 36.23\%$$

# 3  Question 3

Garanti believes the following probability distribution exists for its stock.

```
garanti <- matrix(c(0.25, 0.45, 0.30, 0.25, 0.10, -0.15), nrow = 3, ncol = 2,
                  dimnames = list(c("Boom", "Normal", "Recession"),
                                  c("Probability", "Return")))
matrix_perc(garanti)
```

|           | Probability | Return |
|-----------|-------------|--------|
| Boom      | 25%         | 25%    |
| Normal    | 45%         | 10%    |
| Recession | 30%         | -15%   |

## 3.1 a.

### 3.1.1 Q. What is the expected return of Garanti stock?

### 3.1.2 Answer

```
expected_return <- sum(garanti[,"Probability"] * garanti[,"Return"])
```

$$\mathrm{E}\left(R\right) = 6.25\%$$

## 3.2 b.

### 3.2.1 Q. What is the expected variance and standard deviation of Garanti stock?

### 3.2.2 Answer

```
expected_variance <-
    sum(((garanti[,"Return"] - expected_return) ^ 2) * garanti[,"Probability"])
expected_stdev <- sqrt(expected_variance)
```

$$\sigma^2\left(R\right) = 2.3\%$$
$$\sigma\left(R\right) = 15.16\%$$

## 3.3 c.

### 3.3.1 Q. What is the coefficient of variation on the company's stock?

### 3.3.2 Answer

```
cv <- expected_stdev / expected_return
```

$$\widehat{c_{\mathrm{v}}} = 2.42$$

## 3.4 d.

### 3.4.1 Q. Assuming that the probabilities are calculated empirically based on the last 20 years annual returns, Can we conclude that the expected return on Garanti is higher than the 2% benchmark index return, using a 5% level of significance?

```
sample_n <- 20
benchmark <- 0.02
significance <- 0.05
```

### 3.4.2 Answer

```
sample_error <- expected_stdev / sqrt(sample_n)
test_statistic <- (expected_return - benchmark) / sample_error
rejection_point <- qt(significance, sample_n - 1, lower.tail = FALSE)
```

$$H_0 : \mathrm{E}\left(R\right) \leq 2\%$$
$$H_a : \mathrm{E}\left(R\right) > 2\%$$
$$\sigma_{\bar{X}} = 3.39\%$$
$$\text{Test Statistic} = 1.25$$
$$\text{Rejection Point} = 1.73$$

We can't conclude that Garanti will bring returns better than the benchmark with the given level of significance. If Test Statistic was larger than the Rejection Point we could have rejected the null hypothesis of returns being lesser than the benchmark.

## 3.5 e.

### 3.5.1 Q. Test whether Garanti's expected variance is equal to 2% with a 90% confidence interval?

```
conf_int <- 0.90
s2 <- 0.02
```

### 3.5.2 Answer

```
significance <- 1 - conf_int
deg_f <- sample_n - 1
f_stat <- expected_variance / s2
f_rejection <- qf(significance / 2, deg_f, deg_f, lower.tail = FALSE)
```

$$F = 1.15$$
$$\text{Rejection Point} = 2.17$$

It is not possible to reject the null hypothesis of Garanti's variance being equal to 2%, because the F value is not bigger than the Rejection Point.

## 3.6 Extra

```
ggplot(tibble(x = c(0, 3)), aes(x)) +
    ylab("P(F)") + xlab("F") +
    stat_function(fun = df, args = list(deg_f, deg_f)) +
    stat_function(fun = df, args = list(deg_f, deg_f), geom = "area", fill = "red",
                  alpha = 0.5, xlim = c(f_rejection, 3)) +
    geom_vline(xintercept = f_stat, color = "blue") +
    scale_x_continuous(breaks = round(c(0, f_stat, f_rejection, 3), 2))
```
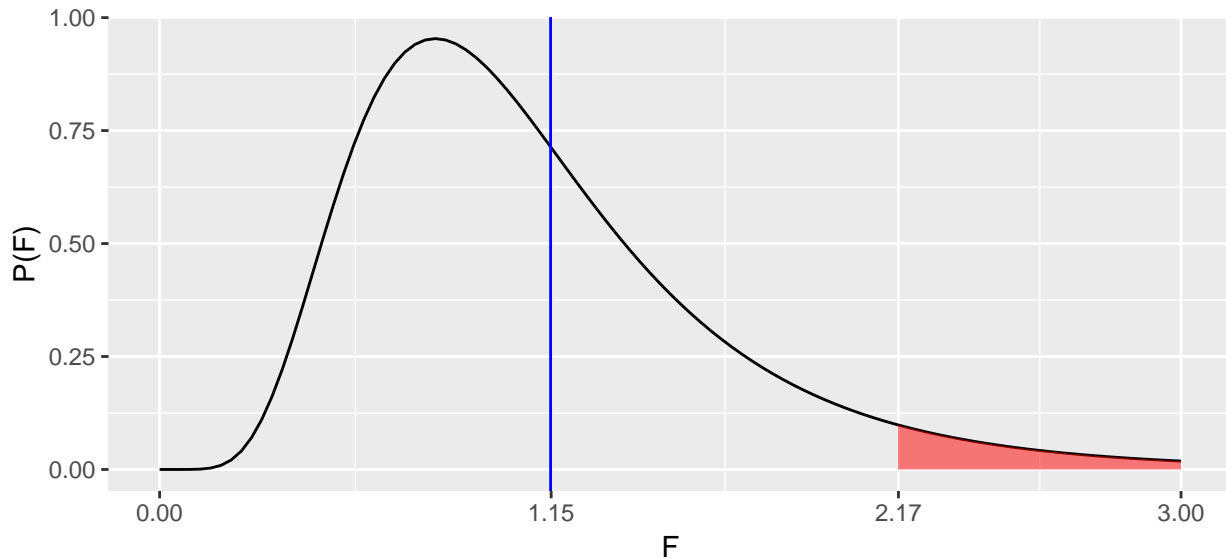
On F-tests, null hypothesis is that the variance is equal. If the ratio of sample variance and benchmark variance was in the red area, we could have rejected the null hypothesis. In this case we can't because samples F statistic which is the blue line is not in the red area. Red area is 5% in our test.

# 4 Question 4

The checking accounts of Garanti Bank are categorized by age of account and balance in account. We are going to select an account at random from this group of 8,000 accounts.

```
account_n <- matrix(
    c(1750, 1250, 2000, 1000, 1250, 750), nrow = 2, ncol = 3,
    dimnames = list(c("<5y", ">5y"), c("0-10k", "10k-100k", "100k+")))
kable(account_n, format.args = list(big.mark = ','))
```

|      | 0-10k | 10k-100k | 100k+ |
|------|-------|----------|-------|
| <5y  | 1,750 | 2,000    | 1,250 |
| >5y  | 1,250 | 1,000    | 750   |

```
balance_n <- colSums(account_n)
kable(t(balance_n), format.args = list(big.mark = ','))
```

| 0-10k | 10k-100k | 100k+ |
|-------|----------|-------|
| 3,000 | 3,000 | 2,000 |

```
age_n <- rowSums(account_n)
kable(t(age_n), format.args = list(big.mark = ','))
```

| <5y | >5y |
|-------|-------|
| 5,000 | 3,000 |

## 4.1  a.

### 4.1.1  Q. What is the probability of selecting an account that is 5 or more years old and has an account balance of TL15,000.

### 4.1.2  Answer

```
account_total <- sum(account_n)
p_gt5y_and_15k <- account_n[">5y", "10k-100k"] / account_total
```

$$P(year > 5 \ \cap \ 15,000) = 12.5\%$$

## 4.2  b.

### 4.2.1  Q. What is the probability of selecting an account that is less than 5 years old given that the account balance is TL128,000.

### 4.2.2  Answer

```
p_lt5y_given_128k <- account_n["<5y", "100k+"] / balance_n["100k+"]
```

$$P(year < 5 \ \cap \ 128,000) = 62.5\%$$

## 4.3  c.

### 4.3.1  Q. What is the probability of selecting an account with a TL5,000 balance.

### 4.3.2  Answer

```
p_5k <- balance_n["0-10k"] / account_total
```

$$P(5,000) = 37.5\%$$

## 4.4 d.

### 4.4.1 Q. What is the conditional probability that the account balance is less than 9,999TL, given that the account is more than 5 years old.

### 4.4.2 Answer

```
p_lt10k_given_gt5y <- account_n[">5y", "0-10k"] / age_n[">5y"]
```

$$P(balance < 10,000 \ \cap \ year > 5) = 41.67\%$$

## 4.5 e.

### 4.5.1 Q. What is the conditional probability that the account is less than 5 years old, given that its balance is TL75,000.

### 4.5.2 Answer

```
p_lt5y_given_75k <- account_n["<5y", "10k-100k"] / balance_n["10k-100k"]
```

$$P(year < 5 \ | \ 75,000) = 66.67\%$$

## 4.6 f.

### 4.6.1 Q. Are age of account and balance in account independent at Garanti Bank? Why or why not?

### 4.6.2 Answer

```
p_lt5y <- age_n["<5y"] / account_total
p_10k <- balance_n["0-10k"] / account_total
p_lt5y_and_10k <- account_n["<5y", "0-10k"] / account_total
p_lt5y_times_10k <- p_lt5y * p_10k
independent <- p_lt5y_and_10k == p_lt5y_times_10k
```

$$P(year < 5) = 62.5\%$$
$$P(balance < 10,000) = 37.5\%$$
$$P(year < 5 \ \cap \ balance < 10,000) = 21.88\%$$
$$P(year < 5) \times P(balance < 10,000) = 23.44\%$$

Not independent because this condidition is not satisfied:

$$P(A \cap B) = P(A) \times P(B)$$

## 4.7 g.

**4.7.1 Q. Suppose 25 accounts are drawn at random from Garanti Bank. Let "F" be the event "At least 5 accounts are less than 5 years old". State the complementing event "F".**

**4.7.2 Answer**

Less than 5 accounts are less than 5 years old.

# 5 Question 5

Below you can see the daily closing prices for four listed stocks.

```
closing <- as.matrix(read.csv(here("q5_closing.csv"), row.names = "Date"))
kable(head(closing, 5))
```

|            | Garanti | TurkTelekom | Aselsan | Ulker | BIM   |
|------------|---------|-------------|---------|-------|-------|
| 04/01/2016 | 6.66    | 5.13        | 8.51    | 17.52 | 48.69 |
| 05/01/2016 | 6.63    | 5.15        | 8.62    | 18.20 | 48.88 |
| 06/01/2016 | 6.77    | 5.09        | 8.68    | 17.96 | 48.26 |
| 07/01/2016 | 6.80    | 5.05        | 8.70    | 17.99 | 48.59 |
| 08/01/2016 | 6.68    | 4.96        | 8.84    | 17.58 | 47.62 |

First 5 rows of total 472 are shown.

## 5.1 a.

**5.1.1 Q. Build a portfolio from these five stocks that would minimize the variance of the portfolio.**

```
returns <- tail(closing, -1) / head(closing, -1) - 1
matrix_perc(head(returns, 5))
```

|            | Garanti | TurkTelekom | Aselsan | Ulker  | BIM    |
|------------|---------|-------------|---------|--------|--------|
| 05/01/2016 | -0.45%  | 0.39%       | 1.29%   | 3.88%  | 0.39%  |
| 06/01/2016 | 2.11%   | -1.17%      | 0.7%    | -1.32% | -1.27% |
| 07/01/2016 | 0.44%   | -0.79%      | 0.23%   | 0.17%  | 0.68%  |
| 08/01/2016 | -1.76%  | -1.78%      | 1.61%   | -2.28% | -2%    |
| 11/01/2016 | 1.35%   | 0.81%       | 2.04%   | -0.57% | -0.97% |

First 5 rows of total 471 are shown.

### 5.1.2 Answer

```
get_pf_return <- function(wt) returns %*% wt
get_pf_var <- function(wt) as.numeric(var(get_pf_return(wt)))
gt0_constaint <- function(wt) wt
eq1_contraint <- function(wt) sum(wt) - 1
optimized <- slsqp(rep(0.2, 5), get_pf_var,
                    hin = gt0_constaint, heq = eq1_contraint)
opt_wt <- set_names(optimized$par, colnames(returns))
opt_var <- optimized$value
```

Optimized weights:

```
matrix_perc(t(opt_wt))
```

| Garanti | TurkTelekom | Aselsan | Ulker | BIM |
|---|---|---|---|---|
| 6.58% | 24.38% | 21.2% | 16.08% | 31.76% |

$$\min(\sigma_p^2) = 0.01127\%$$

## 5.2  b.

### 5.2.1  Q. The daily returns on your portfolio. Show descriptive stats of the daily returns of your portfolio.

### 5.2.2  Answer

```
opt_pf_return <- get_pf_return(opt_wt)
colnames(opt_pf_return) <- "Portfolio Return"
matrix_perc(head(opt_pf_return, 5))
```

|  | Portfolio Return |
|---|---|
| 05/01/2016 | 1.09% |
| 06/01/2016 | -0.61% |
| 07/01/2016 | 0.13% |
| 08/01/2016 | -1.21% |
| 11/01/2016 | 0.32% |

```
descriptive <- describe(opt_pf_return, trim = 0)
rownames(descriptive) <- "Portfilio Return"
kable(t(select(descriptive, -vars)))
```

|          | Portfilio Return |
|----------|------------------|
| n        | 471.0000000      |
| mean     | 0.0013758        |
| sd       | 0.0106166        |
| median   | 0.0014284        |
| trimmed  | 0.0013758        |
| mad      | 0.0084356        |
| min      | -0.0584997       |
| max      | 0.0341337        |
| range    | 0.0926335        |
| skew     | -0.4244545       |
| kurtosis | 2.8656956        |
| se       | 0.0004892        |

## 5.3   c.

### 5.3.1   Q. Calculate the Sharpe ratio of the portfolio based on daily returns (assume 0% risk-free rate)

### 5.3.2   Answer

```
sharpe <- descriptive[,"mean"] / descriptive[,"sd"]
```

$$S = 0.13$$

## 5.4   d.

### 5.4.1   Q. Construct a 95% confidence interval for the daily returns on your portfolio, assuming that the portfolio returns are normally distributed.

### 5.4.2   Answer

```
conf_int <- 0.95
significance <- 1 - conf_int
upper_z <- qnorm(significance / 2, lower.tail = FALSE)
scaled_z <- upper_z * descriptive[,"se"]
min_mean <- descriptive[,"mean"] - scaled_z
max_mean <- descriptive[,"mean"] + scaled_z
```

$$\sigma_{\bar{X}} = 0.05\%$$
$$z_{0.025} = 1.96$$
$$\mu_R = 0.14\% \pm 0.1\%$$
$$0.04\% \leq \mu_R \leq 0.23\%$$

## 5.5   e.

### 5.5.1   Q. Calculate the probability of your portfolio will earn a daily return of 0.5% or more; assuming that the returns are normally distributed.

### 5.5.2   Answer

```
test_stat <- (descriptive[,"mean"] - 0.005) / descriptive[,"se"]
prob_of_more <- pnorm(test_stat)
```

$$\text{Test Statistic} = -7.41$$

$$P(R_p > 0.5\%) = 6.3839311 \times 10^{-14}$$

## 5.6   Extra

```
returns_df <- returns %>%
    as_tibble(rownames = "Date") %>%
    mutate_at("Date", dmy) %>%
    gather("Company", "Return", -Date)
returns_pf_df <- cbind(opt_pf_return, get_pf_return(rep(0.2, 5))) %>%
    `colnames<-`(c("1. Minimum Variance", "2. Equal Weight")) %>%
    as_tibble(rownames = "Date") %>%
    mutate_at("Date", dmy) %>%
    gather("Portfolio", "Return", -Date)
ggplot(returns_df, aes(x = Date, y = Return)) +
    geom_line(aes(group = Company), color = "red", alpha = 0.3) +
    geom_line(aes(x = Date, y = Return), color = "blue",
              data = filter(returns_pf_df, Portfolio == "1. Minimum Variance")) +
    coord_cartesian(ylim = c(-0.1, 0.1))
```
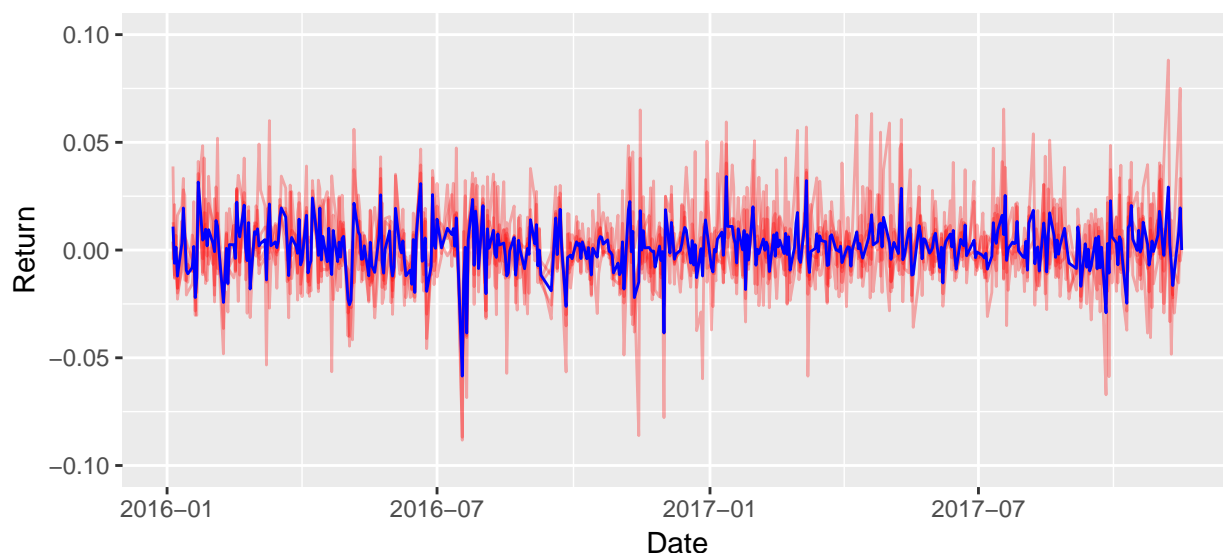


Red lines are returns for individual companies. The blue line is the minimum variance portfolio return. It can be seen that portfolio return has less deviation.

```
ggplot(returns_pf_df, aes(Return)) +
    geom_density(aes(fill = Portfolio), alpha = 0.5) +
    scale_colour_manual(values = c("blue", "red"))
```
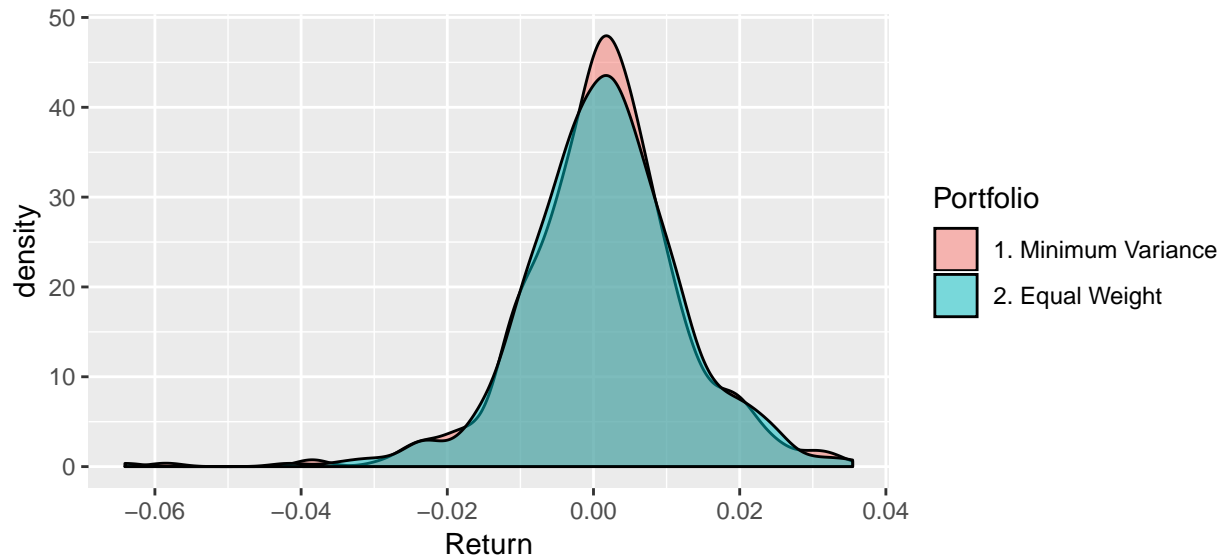


This is a density plot for the optimized minimum variance portfolio's returns. Overlayed on top of it is another portfolio, with the same companies, but with equal weights of 20% each. It can be observed that the optimization managed to make the distribution more peaked, closer to the center with less variance.

# 6  Question 6

It is known that 65% of MS in Finance students think their Statistics lecturer is cool. In a certain sample, 10 students are picked at random and it is required to calculate the probability that less than 3 of the 10 think that their lecturer is cool.

```
p <- 0.65
n <- 10
```

## 6.1  a.

### 6.1.1  Q. Name a probability distribution that can be used for modelling this situation stating one necessary assumption for this model to be valid.

### 6.1.2  Answer

Binomial distribution is appropriate for modeling. Students response shouldn't be effected from each other, meaning independence is assumed.

## 6.2   b.

**6.2.1   Q. Use the model to calculate the required probability less than 3 of the 10 think that their lecturer is cool.**

**6.2.2   Answer**

```
less_than_3 <- pbinom(3, n, p)
```

$$P(X \leq 3) = 2.6\%$$

## 6.3   c.

**6.3.1   Q. Calculate the mean and variance of this distribution.**

**6.3.2   Answer**

```
mean <- n * p
variance <- n * p * (1 - p)
```

$$\mu = 6.5$$
$$\sigma = 2.275$$