Statistical Methods and Data Analysis I
Lecture 8: Portfolio Optimization. Normal Distribution

Oleg Goldshmidt
oleg.goldshmidt@post.idc.ac.il

Arison School of Business
Interdisciplinary Center (IDC)
Herzliya, Israel

March 26, 2019

Linear Combination of Random Variables: Stock Portfolio

Scenario: constructing an investment portfolio

I want to invest a fixed sum in a portfolio of stocks for 1 year. I can invest in US treasury bonds practically without any risk, but that will only give me a 2.5% rate of return.
Instead, I am considering investing in 2 very large companies:

1. Cisco Systems (CSCO): computer networking equipment.
2. Unilever (ULVR): food, beverages, cleaning products, personal care *(familiar brands include Ben & Jerry's, Lipton, Hellmann's, Knorr, Dove, Magnum, Vaseline)*.

The two companies' rates of return are *independent* random variables: different markets, different geographies, etc. *How should I divide my money between these two stocks?*

Linear Combination of Random Variables: Stock Portfolio

Scenario: constructing an investment portfolio

I want to invest a fixed sum in a portfolio of stocks for 1 year. I can invest in US treasury bonds practically without any risk, but that will only give me a 2.5% rate of return.
Instead, I am considering investing in 2 very large companies:

1. Cisco Systems (CSCO): computer networking equipment.
2. Unilever (ULVR): food, beverages, cleaning products, personal care *(familiar brands include Ben & Jerry's, Lipton, Hellmann's, Knorr, Dove, Magnum, Vaseline).*

The two companies' rates of return are *independent* random variables: different markets, different geographies, etc. *How should I divide my money between these two stocks?*

Basic data about the distribution of the rates of return for these two candidates:

CISCO $\mu_{CSCO} = 0.13, \quad \sigma_{CSCO} = 0.15$

Unilever $\mu_{ULVR} = 0.29, \quad \sigma_{ULVR} = 0.20$

IDC HERZLIYA

Stock Portfolio: Analysis

Stock Portfolio: Analysis

- Either Cisco or Unilever will give me a better *expected return* (13% and 29%, respectively) than the "risk-free" $r = 0.025$ of US treasury bonds.

Stock Portfolio: Analysis

- Either Cisco or Unilever will give me a better *expected return* (13% and 29%, respectively) than the "risk-free" $r = 0.025$ of US treasury bonds.
- Unilever has a higher *expected return* than Cisco — *should I invest all my money in Unilever?*

Stock Portfolio: Analysis

- Either Cisco or Unilever will give me a better *expected return* (13% and 29%, respectively) than the "risk-free" $r = 0.025$ of US treasury bonds.
- Unilever has a higher *expected return* than Cisco — *should I invest all my money in Unilever?*
- But uncertainty, expressed as the *standard deviation of returns*, is also higher for Unilever! I would like to maximize the return while minimizing the risk...

Stock Portfolio: Analysis

- Either Cisco or Unilever will give me a better *expected return* (13% and 29%, respectively) than the "risk-free" $r = 0.025$ of US treasury bonds.
- Unilever has a higher *expected return* than Cisco — *should I invest all my money in Unilever?*
- But uncertainty, expressed as the *standard deviation of returns*, is also higher for Unilever! I would like to maximize the return while minimizing the risk...
- Actually, the important part is the *excess rate of return over the risk-free rate r*. *After all, if my portfolio yields 2.5% then I'd rather invest in the (virtually risk-free) bonds.*

Stock Portfolio: Analysis

- Either Cisco or Unilever will give me a better *expected return* (13% and 29%, respectively) than the "risk-free" $r = 0.025$ of US treasury bonds.
- Unilever has a higher *expected return* than Cisco — *should I invest all my money in Unilever?*
- But uncertainty, expressed as the *standard deviation of returns*, is also higher for Unilever! I would like to maximize the return while minimizing the risk...
- Actually, the important part is the *excess rate of return over the risk-free rate r*. *After all, if my portfolio yields 2.5% then I'd rather invest in the (virtually risk-free) bonds.*
- A useful measure of portfolio performance is the ration of its expected rate of return, adjusted by the risk-free rate, to the standard deviation of its rate of return:

$$\textit{Sharpe ratio} = \frac{\mu - r}{\sigma}$$

This is what I want to maximize to balance the expected return and the risk.

Stock Portfolio: Analysis (Cont.)

Stock Portfolio: Analysis (Cont.)

- If I invest fraction $a \in [0, 1]$ of my money in Cisco and $(1 - a)$ in Unilever, the statistics of my portfolio's rate of return will be described as

$$\mu = a\mu_{CSCO} + (1 - a)\mu_{ULVR} = 0.13a + 0.29(1 - a)$$

$$\begin{aligned} \sigma &= \sqrt{a^2\sigma^2_{CSCO} + (1 - a)^2\sigma^2_{ULVR}} \\ &= \sqrt{(0.13a)^2 + (0.29(1 - a))^2} \end{aligned}$$

Stock Portfolio: Analysis (Cont.)

- If I invest fraction $a \in [0, 1]$ of my money in Cisco and $(1 - a)$ in Unilever, the statistics of my portfolio's rate of return will be described as

$$\mu = a\mu_{CSCO} + (1 - a)\mu_{ULVR} = 0.13a + 0.29(1 - a)$$

$$\begin{aligned} \sigma &= \sqrt{a^2\sigma^2_{CSCO} + (1 - a)^2\sigma^2_{ULVR}} \\ &= \sqrt{(0.13a)^2 + (0.29(1 - a))^2} \end{aligned}$$

- Let's construct the Sharpe ratio as a function of a (the relative fraction of CSCO in the portfolio):

$$Sharpe\ ratio = \frac{0.13a + 0.29(1 - a) - 0.025}{\sqrt{(0.13a)^2 + (0.29(1 - a))^2}}$$

How do we maximize it?

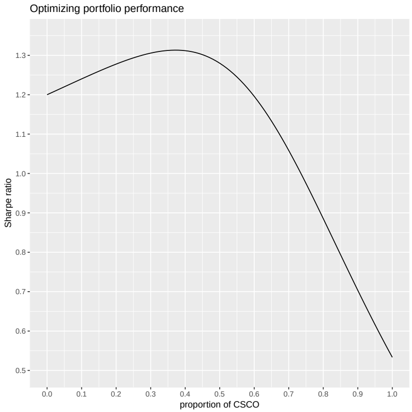
- If I invest fraction $a \in [0,1]$ of my money in Cisco and $(1-a)$ in Unilever, the statistics of my portfolio's rate of return will be described as

$$\mu = a\mu_{CSCO} + (1-a)\mu_{ULVR} = 0.13a + 0.29(1-a)$$

$$\sigma = \sqrt{a^2\sigma^2_{CSCO} + (1-a)^2\sigma^2_{ULVR}} = \sqrt{(0.13a)^2 + (0.29(1-a))^2}$$

- Let's construct the Sharpe ratio as a function of a (the relative fraction of CSCO in the portfolio):

$$Sharpe\ ratio = \frac{0.13a + 0.29(1-a) - 0.025}{\sqrt{(0.13a)^2 + (0.29(1-a))^2}}$$

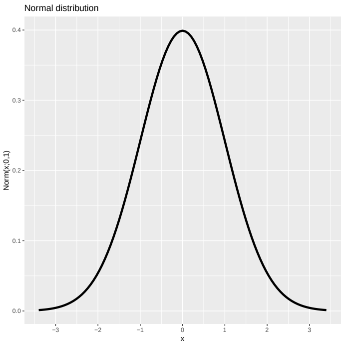


How do we maximize it?

Normal Distribution (The Bell Curve)

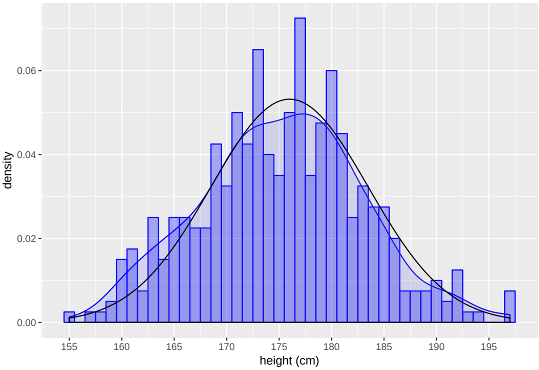
- Unimodal, symmetric, bell-shaped curve
- Many random variables are nearly normally distributed
  - But hardly any is exactly normal
- Normal distribution with mean μ and standard deviation σ is often denoted as $N(x;\mu,\sigma)$

$$N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

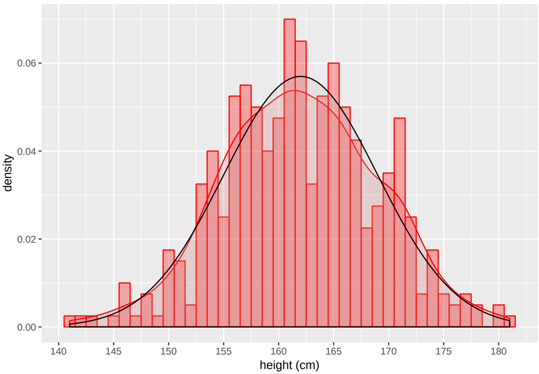


Is Height Distribution Normal?

Distribution of men's height

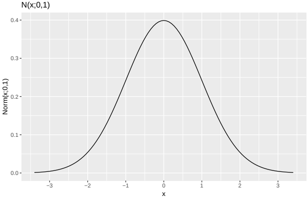


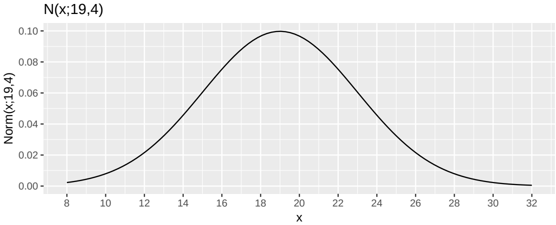
Distribution of women's height

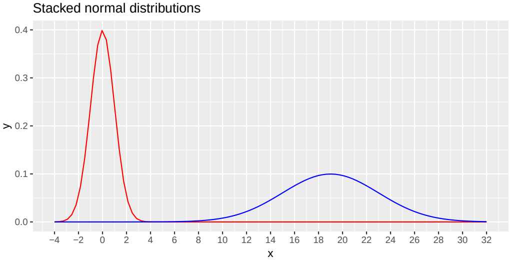


Normal Distributions With Different Parameters

$N(x; 0, 1)$ $N(x; 19, 4)$







Standardizing Normal Distributions

Scenario: college admission

Two American students, Rebecca and Joseph, apply to IDC. Their records show that Rebecca took the Scholastic Aptitude Test (SAT) and scored 1800, while Joseph took the American College Test (ACT) and scored 24.

How will the IDC decide which of the candidates will get the only remaining place in the Business and Data Analytics program?

Standardizing Normal Distributions

Scenario: college admission

Two American students, Rebecca and Joseph, apply to IDC. Their records show that Rebecca took the Scholastic Aptitude Test (SAT) and scored 1800, while Joseph took the American College Test (ACT) and scored 24.

How will the IDC decide which of the candidates will get the only remaining place in the Business and Data Analytics program?

- It is known that SAT scores are normally distributed with $\mu = 1500$ and $\sigma = 300$, while ACT scores are normally distributed with $\mu = 21$ and $\sigma = 5$.

IDC HERZLIYA

Standardizing Normal Distributions

Scenario: college admission

Two American students, Rebecca and Joseph, apply to IDC. Their records show that Rebecca took the Scholastic Aptitude Test (SAT) and scored 1800, while Joseph took the American College Test (ACT) and scored 24.

How will the IDC decide which of the candidates will get the only remaining place in the Business and Data Analytics program?

- It is known that SAT scores are normally distributed with $\mu = 1500$ and $\sigma = 300$, while ACT scores are normally distributed with $\mu = 21$ and $\sigma = 5$.
- Note that the normal distribution depends only on

$$z = \frac{x - \mu}{\sigma}, \qquad N(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}}$$

Z-Scores

Z-Scores

- We cannot directly compare Rebecca's SAT score with Joseph's ACT score.

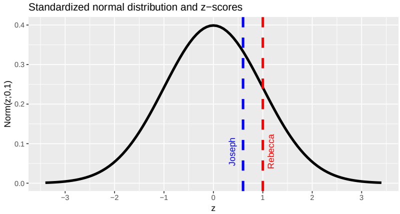
Z-Scores

- We cannot directly compare Rebecca's SAT score with Joseph's ACT score.
- *But we can compare the normalized distances from the mean, called* ***z-score****.*

Z-Scores

- We cannot directly compare Rebecca's SAT score with Joseph's ACT score.
- *But we can compare the normalized distances from the mean, called* **z-score**.

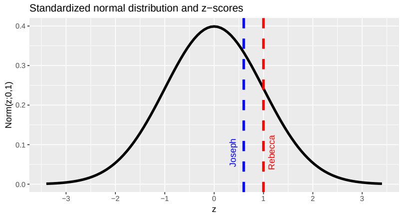
  - Rebecca's z-score is $\dfrac{1800 - 1500}{300} = 1$
  - Joseph's z-score is $\dfrac{24 - 21}{5} = 0.6$



Z-Scores

- We cannot directly compare Rebecca's SAT score with Joseph's ACT score.
- *But we can compare the normalized distances from the mean, called* **z-score**.

  - Rebecca's z-score is $\dfrac{1800-1500}{300}=1$
  - Joseph's z-score is $\dfrac{24-21}{5}=0.6$

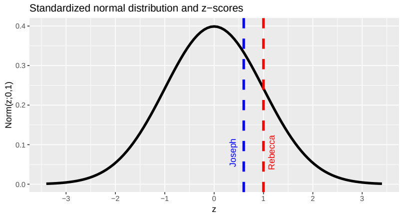


- In other words, Rebecca's score is 1 standard deviation above the mean, and Joseph's score is only 0.6 standard deviations above the mean.

Z-Scores

- We cannot directly compare Rebecca's SAT score with Joseph's ACT score.
- *But we can compare the normalized distances from the mean, called* ***z-score****.*

  - Rebecca's z-score is $\dfrac{1800-1500}{300} = 1$
  - Joseph's z-score is $\dfrac{24-21}{5} = 0.6$



- In other words, Rebecca's score is 1 standard deviation above the mean, and Joseph's score is only 0.6 standard deviations above the mean.
- *The only vacant place in the Business and Data Analytics program should be awarded to Rebecca.*

Z-Scores

Definition

A *standardized score*, or *z-score*, of an observation is the number of standard deviations that if falls above of below the mean:

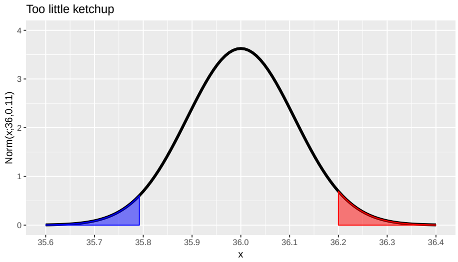
$$z = \frac{observation \; - \; mean}{standard \; deviation} = \frac{x - \mu}{\sigma}$$

- z-scores are defined for arbitrary distributions
- only for *normal* distribution can we use z-scores to calculate percentiles
- the often-used rule of thumb is that observations with $|z| > 2$ (more than 2σ from the mean in either direction) are unusual
  - assuming that is dangerous: at least check that your distribution does not have fatter tails than the normal one (that falls off very fast indeed)

Quality Control

Heinz Quality Control

At Heinz ketchup factories the amount of ketchup that goes into each bottle is supposed to be normally distributed with a mean of 36 oz and a standard deviation of 0.11 oz.
Once every 30 minutes a bottle is selected at random from the production line and the amount of ketchup in it is measured precisely. If it is below 35.8 oz or above 36.2 oz then the bottle fails the inspection.
What percent of bottles have less than 35.8 oz of ketchup?



$$z = \frac{35.8 - 36}{0.11} = -1.82$$

In R:

$\texttt{pnorm}(-1.82) \to 3.44\%$

Computing Percentiles In R

Computing Percentiles In R

- *What percentage of a normal distribution lies within 1σ of the mean?*

$$\texttt{pnorm(1)} - \texttt{pnorm(}-\texttt{1)} \approx \texttt{0.68}$$

Computing Percentiles In R

- *What percentage of a normal distribution lies within 1σ of the mean?*

$$\texttt{pnorm}(1) - \texttt{pnorm}(-1) \approx 0.68$$

- *What percentage of a normal distribution lies within 2σ of the mean?*

$$\texttt{pnorm}(2) - \texttt{pnorm}(-2) \approx 0.95$$

Computing Percentiles In R

- *What percentage of a normal distribution lies within 1σ of the mean?*

$$\texttt{pnorm}(1) - \texttt{pnorm}(-1) \approx 0.68$$

- *What percentage of a normal distribution lies within 2σ of the mean?*

$$\texttt{pnorm}(2) - \texttt{pnorm}(-2) \approx 0.95$$

- *What percentage of a normal distribution lies within 3σ of the mean?*

$$\texttt{pnorm}(3) - \texttt{pnorm}(-3) \approx 0.997$$

IDC HERZLIYA

Computing Percentiles In R

- *What percentage of a normal distribution lies within 1σ of the mean?*

$$\texttt{pnorm(1)} - \texttt{pnorm(}-\texttt{1)} \approx \texttt{0.68}$$

- *What percentage of a normal distribution lies within 2σ of the mean?*

$$\texttt{pnorm(2)} - \texttt{pnorm(}-\texttt{2)} \approx \texttt{0.95}$$

- *What percentage of a normal distribution lies within 3σ of the mean?*

$$\texttt{pnorm(3)} - \texttt{pnorm(}-\texttt{3)} \approx \texttt{0.997}$$

- *What is the z-score that corresponds to 95% of the distribution lying between $-z$ and $+z$?*

$$\texttt{pnorm(1.96)} - \texttt{pnorm(}-\texttt{1.96)} = \texttt{0.9500042} \approx \texttt{0.95}$$