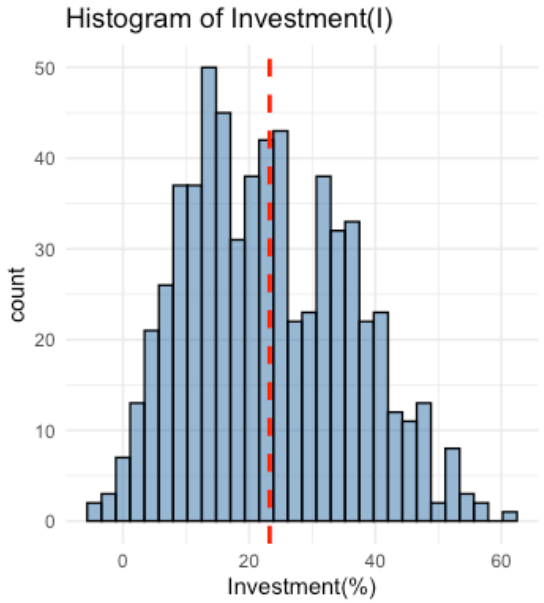
I. Exploring the data

1. Load the data and store it in an object called
- The dataset covers the period from year 2014 to 2021.
- There are 80 firms in the dataset.
- There are 640 observations in the dataset.
- The dataset is balanced.

2. Plot a histogram of Investment, and describe its distribution

Statistic	X1
N	640
Mean	23.24
Standard Deviation	12.93
Median	21.95
Trimmed Mean	22.75
MAD	14.47
Minimum	-4.69
Maximum	61.17
Range	65.86
Skewness	0.31
Kurtosis	-0.59
Standard Error	0.51

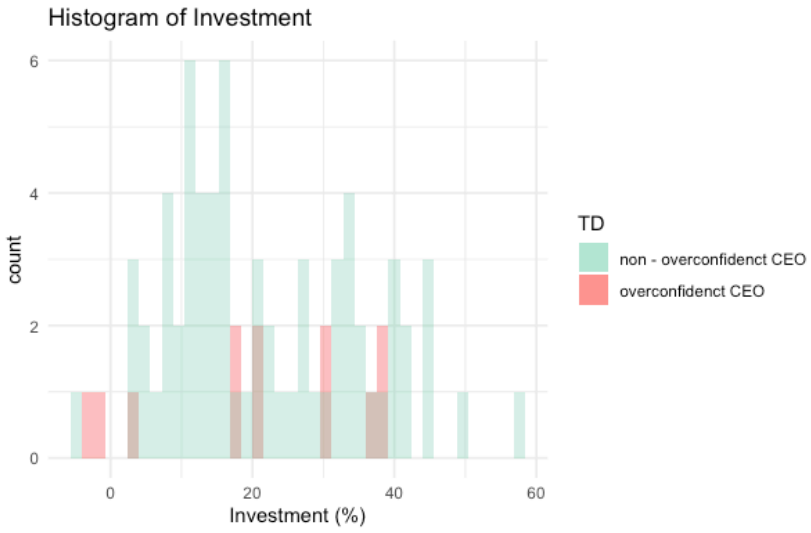


The data seems to be concentrated from 16% to 26%, which the mean value of the Investment(%) is 23.24258% and the median is 21.95447%. The distribution above is not symmetrical. In fact, the histogram has positive skew and negative kurtosis, which indicates that the distribution has a long tail to the right and thinner tails than normal distribution. The data range is from about -4.7% to about 61.2%, approximate range equals 65.86%. There seem to be no outlier in the histogram.

3. Plot the distribution of Investment, for CEOs with media portrayal classified as overconfident, i.e., TDi=1, and those who are not.

Overconfidence	X1
N	12
Mean	21.12
Standard Deviation	15.08
Median	20.83
Trimmed Mean	21.8
MAD	19.71
Minimum	-3.45
Maximum	38.93
Range	42.38
Skewness	-0.33
Kurtosis	-1.42
Standard Error	4.35

Non-Overconfidence	X1
N	68
Mean	21.97
Standard Deviation	13.67
Median	17.72
Trimmed Mean	21.38
MAD	14.7
Minimum	-4.69
Maximum	57.67
Range	62.36
Skewness	0.42
Kurtosis	-0.75
Standard Error	1.66



It is clear to see that the number of observations for overconfident CEOs is nearly 6 times smaller than the other type of CEOs in 2016. For CEOs portrayed as overconfidence, the mean and median are 21.12109% and 20.83045% respectively. In the non - overconfidence data, the mean (21.96709%) is very close to the mean value of Investment for overconfident CEOs , but the median is quite lower, which is 17.72203%. The values of Investment for non - overconfident CEOs tend to focus on the left side of the chart (more than 50% values of Investment < mean

value of the Investment). In contrast, most of the values belongs to the right of the overconfident CEOs distribution.

Both distributions have the same sign for kurtosis, but different for skew. The ranges of Investment(%) for overconfidence and non - overconfidence are about 42.38 and 62.36 respectively. The max value of Investment of overconfident CEOs is far smaller than the max value of the Investment of non - overconfident CEOs. There seem to be on probable outlier to the right and one to the left of the histogram of non - overconfident CEOs. Besides, It is hard to comment on the outliers about the distribution of Investment of overconfident CEOs.

II. Consider the following regression equation:

$$I_i=\beta_0+\beta_1 CF_i+\beta_2 size_i+\beta_3 MV_i+\beta_4 CG_i+\beta_5 TD_i+\beta_6(CF_i\times TD_i)+u_i$$

- Interpret the results of my estimation

	Dependent variable:	
	I	
	default SE	robust SE
	(1)	(2)
CF	1.043***	1.043***
	(0.069)	(0.073)
size	-0.758	-0.758
	(1.712)	(1.405)
MV	-0.166	-0.166
	(0.149)	(0.180)
CG	-0.642**	-0.642*
	(0.285)	(0.338)
TD	-5.126	-5.126
	(4.279)	(3.871)
CF:TD	0.101	0.101
	(0.192)	(0.149)
Constant	15.940	15.940
	(25.619)	(21.109)
Observations	80	80
R2	0.797	0.797
Adjusted R2	0.780	0.780
Residual Std. Error (df = 73)	6.472	6.472
F Statistic (df = 6; 73)	47.650***	47.650***
Note:	*p<0.1; **p<0.05; ***p<0.01	

- Evaluate whether the regression equation for CEOs portrayed as overconfident (TDi=1) is statistically significantly different to the regression equation for non-overconfident CEOs.

The F - test is appropriate method. The joint hypothesis can be written mathematically as:

$$H0{:}\beta5=\beta6=0 \qquad H1{:}\beta5 \cup \beta6 \neq 0$$

Using the function linearHypothesis, the result of p - value of the F - statistic = 0.2717, which is larger than 0.05, so that we can not reject $H0$H0. In conclusion, the regression equation for CEOs portrayed as overconfident can not be assumed statistically significantly different to the regression equation for non-overconfident CEOs at 5% significance level.

- The predicted investment percentage for CEOs portrayed as overconfident and other CEOs at the mean levels of cashflow, market value and size

With TD = 1:
Use the regression equation: Ii=10.814+1.144CFi−0.758$sizei$−0.166MVi−0.642CGi and the mean of independent variables for overconfident CEOs, we have the predicted value of I (21.12%).

With TD = 0:
Use the regression equation: Ii=15.940+1.043CFi−0.758$sizei$−0.166MVi−0.642CGi and the mean of independent variables for non - overconfident CEOs, we have the predicted value of I (21.97%).

III. Continue to use mydata, your subset from CEOdata, estimate a new specification of the model

$$I_i=\beta_0+\beta_1 CF_i+\beta_2 CG_i+\beta_3 MV_i+\beta_4 TD_i+\beta_5 CF_i \times TD_i+\beta_6 CF_{2i}+u_i$$

- Estimate the mode

```
Call:
lm(formula = I ~ CF + CG + MV + TD + CF:TD + I(CF^2), data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max
-15.0948  -3.8556  -0.5037   3.9468  17.2180

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.459497   2.274569   2.840  0.00584 **
CF           0.743460   0.263575   2.821  0.00617 **
CG          -0.611724   0.283586  -2.157  0.03429 *
MV          -0.177976   0.147102  -1.210  0.23023
TD          -5.104454   4.238158  -1.204  0.23232
I(CF^2)      0.007764   0.006539   1.187  0.23894
CF:TD        0.101824   0.189658   0.537  0.59298
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.419 on 73 degrees of freedom
Multiple R-squared:  0.7999,    Adjusted R-squared:  0.7835
F-statistic: 48.64 on 6 and 73 DF,  p-value: < 2.2e-16


t test of coefficients:

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.4594973  1.9304681  3.3461 0.001297 **
CF           0.7434604  0.2244145  3.3129 0.001439 **
CG          -0.6117241  0.3300797 -1.8533 0.067886 .
MV          -0.1779764  0.1738568 -1.0237 0.309360
TD          -5.1044540  4.1367718 -1.2339 0.221190
I(CF^2)      0.0077645  0.0052441  1.4806 0.143014
CF:TD        0.1018237  0.1585418  0.6423 0.522722
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Calculate and discuss the marginal effect of CF at CF = 10, when TD = 0 and when TD = 1

- marginal effect of CF at CF = 10 and TD = 1: 1.000574
- marginal effect of CF at CF = 10 and TD = 0: 0.8987499
- marginal effect of CF at CF = 10 and TD = 1 (alternative way): 1.000574
- marginal effect of CF at CF = 10 and TD = 0 (alternative way): 0.8987499

The marginal effect of CF when the CEOs are classified as overconfidence is larger than the marginal effect of CF of non - overconfident CEOs for any given CF, which implies that the instantaneous change in CF of overconfident CEOs on the Investment has bigger impact, while all other variables are held constant.

- Test the joint significance of the interaction between CF and TD and the quadratic term $CF2$

In order to test the joint significance of the interaction between CF and TD and the quadratic term $CF2$CF2, we use the F - test method. The hypothesis is:

$$H0{:}\beta5{=}\beta6{=}0 \quad H1{:}\beta5{\cup}\beta6{\neq}0$$

Using the linearHypothesis function, the result of p - value of the F - statistic = 0.3275, which is larger than 0.05, so that we can not reject $H0$H0.

IV. Estimate a model with fixed effects

- Estimate a linear-linear model with time and entity fixed effects. The dependent variable should be Investment

```
t test of coefficients:

       Estimate Std. Error t value  Pr(>|t|)
CF     1.014380   0.027716 36.5986 < 2.2e-16 ***
size  -0.128791   0.446697 -0.2883  0.773211
MV    -0.883537   0.470647 -1.8773  0.061012 .
CG    -0.714277   0.356726 -2.0023  0.045744 *
TD    -3.628456   1.299690 -2.7918  0.005425 **
CF:TD  0.020518   0.043329  0.4736  0.636007
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If CEOs are classified as overconfidence, then TD = 1, the estimated parameter of TD is become the intercept of the fixed effect model. Including a constant in addition to the entity and time fixed effects would result in perfect multicollinearity. The slope coeffient of CFi is expected to change from 1.014380 to 1.034898 (as (β1+β6)CFi).

The slope coefficient β5 of TD is statistically significant at 1% significane level (p - values < 0.01). That means we can reject the null hypothesis of β5 = 0, overconfident CEOs are statistically significant difference non - overconfident CEOs.