Reliability_Class_Activity

George Paul (2021121006)

2024-02-20

```
rm(list = setdiff(ls(), lsf.str()))
```

```
# install.packages('readxl')
library(readxl)
```

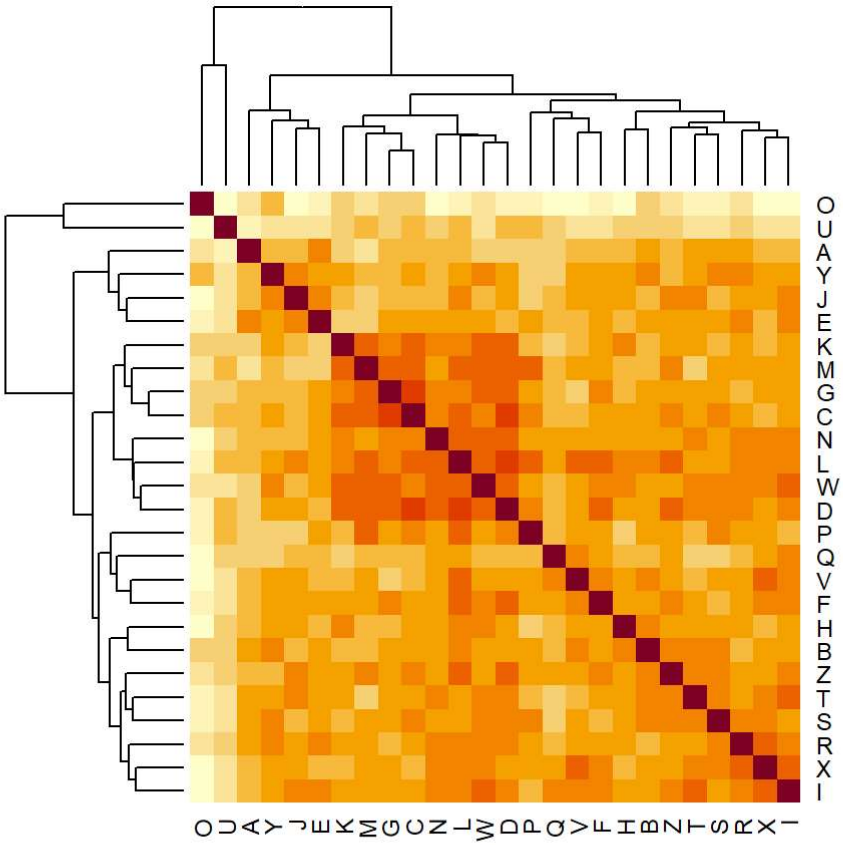
Question 1

```
excel_path <- 'D:\\FILES\\BRSM_Assignment_2_datasets.xlsx'
data <- read_excel(excel_path)
data
```

A <dbl>	B <dbl>	C <dbl>	D <dbl>	E <dbl>	F <dbl>	G <dbl>	H <dbl>	I <dbl>	J <dbl>
2	-1	-1	-1	-1	-1	4	-2	3	-2
-3	-2	-2	-3	-3	-4	-4	-2	-1	-3
2	3	1	0	-1	-1	2	-1	2	3
3	1	-1	-2	-1	-1	-4	-2	-3	0
-2	3	-1	1	-1	1	-2	1	3	2
-3	2	-2	0	-4	1	-2	-4	-1	-2
-1	-1	-2	-1	-3	-3	-3	-3	-1	-1
-3	-3	-2	-4	-4	-4	-4	2	-4	-3
2	-1	-2	-3	-2	-2	-4	-3	-2	-3
3	4	-1	-1	-2	0	-2	3	-2	-2

1-10 of 110 rows | 1-10 of 26 columns Previous **1** 2 3 4 5 6 … 11 Next

```
correlation_matrix <- cor(data)
heatmap(correlation_matrix,
        symm = TRUE,  # Display the upper and lower triangles symmetrically
)
```



Taking a look at the heat map, the most likely outliers among the participants are either *Participant O* or *Participant U*.

Question 2.1

```
# install.packages("psych")
library(psych)
```

```
data <- read_excel(excel_path, sheet=2)
data
```

JS1 <dbl>	JS2 <dbl>	JS3 <dbl>	JS4 <dbl>	JSAVG <dbl>	JP1 <dbl>	JP2 <dbl>	JP3 <dbl>	JP4 <dbl>	JPAVG <dbl>
9	9.5	9.75	7.875	9.03125	1	1.5	1.75	4.875	2.28125
5	6.5	6.25	7.125	6.21875	6	4.0	4.00	7.000	5.25000
4	5.0	3.50	6.750	4.81250	5	6.5	5.25	4.625	5.34375
6	7.0	5.50	4.750	5.81250	6	7.0	8.50	8.250	7.43750
7	7.5	6.75	7.375	7.15625	6	5.0	4.50	9.250	6.18750
1	2.5	3.25	3.625	2.59375	9	7.5	4.75	2.375	5.90625

JS1 <dbl>	JS2 <dbl>	JS3 <dbl>	JS4 <dbl>	JSAVG <dbl>	JP1 <dbl>	JP2 <dbl>	JP3 <dbl>	JP4 <dbl>	JPAVG <dbl>
1	5.5	7.75	7.875	5.53125	8	8.0	6.00	5.000	6.75000
5	7.5	4.75	5.375	5.65625	7	8.5	9.25	5.625	7.59375
6	4.0	3.00	2.500	3.87500	2	3.0	6.50	5.250	4.18750
1	2.5	3.25	6.625	3.34375	9	9.5	6.75	4.375	7.40625

1-10 of 30 rows Previous **1** 2 3 Next

```
calc_cb <- function(data) {
  avg_var <- mean(apply(data, 2, function(col) mean(var(col))))
  avg_var

  cov_mat <- cov(data)
  avg_cov <- sum(cov_mat[lower.tri(cov_mat)]) / sum(lower.tri(cov_mat))
  avg_cov
  item_count <- length(names(data))
  cb_alpha <- (item_count * avg_cov) / (avg_var + ((item_count-1) * avg_cov))
  cb_alpha
}
```

```
js.data <- data[, c("JS1", "JS2", "JS3", "JS4")]
js.data
```

JS1 <dbl>	JS2 <dbl>	JS3 <dbl>	JS4 <dbl>
9	9.5	9.75	7.875
5	6.5	6.25	7.125
4	5.0	3.50	6.750
6	7.0	5.50	4.750
7	7.5	6.75	7.375
1	2.5	3.25	3.625
1	5.5	7.75	7.875
5	7.5	4.75	5.375
6	4.0	3.00	2.500
1	2.5	3.25	6.625

1-10 of 30 rows Previous **1** 2 3 Next

```
calc_cb(js.data)
```

```
## [1] 0.8626397
```

Cronbach's Alpha for *Job Satisfaction* is ≈ 0.86.

Question 2.2

```
jp.data <- data[, c("JP1", "JP2", "JP3", "JP4")]
jp.data
```

JP1 <dbl>	JP2 <dbl>	JP3 <dbl>	JP4 <dbl>
1	1.5	1.75	4.875
6	4.0	4.00	7.000
5	6.5	5.25	4.625
6	7.0	8.50	8.250
6	5.0	4.50	9.250
9	7.5	4.75	2.375
8	8.0	6.00	5.000
7	8.5	9.25	5.625
2	3.0	6.50	5.250
9	9.5	6.75	4.375

1-10 of 30 rows Previous **1** 2 3 Next

```
calc_cb(jp.data)
```

```
## [1] 0.5576079
```

Cronbach's Alpha for *Job Performance* is ≈ 0.56.

Question 2.3

Considering a benchmark as $ $, the items for Job Satisfaction seem well correlated. All items test the same construct and they can be used together as a good gauge of the construct. Whereas, the same cannot be said for the Job Performance items. With $\mathrm{CB} \approx 0.56$, the items are not very internally consistent. We cannot surely say that all the *JP* items test the same construct. A point to note though is that we can easily increase the internal consistency of the *JP* items by removing *JP4* in particular:

```
jp.data.without4 <- data[, c("JP1", "JP2", "JP3")]
jp.data.without4
```

JP1 <dbl>	JP2 <dbl>	JP3 <dbl>
1	1.5	1.75
6	4.0	4.00
5	6.5	5.25
6	7.0	8.50
6	5.0	4.50
9	7.5	4.75
8	8.0	6.00
7	8.5	9.25
2	3.0	6.50
9	9.5	6.75

1-10 of 30 rows Previous **1** 2 3 Next

```
calc_cb(jp.data.without4)
```

```
## [1] 0.6843528
```

Note: Removing the other items was also considered, omitting JP4 remained the best option.

Question 3.1

Note: Only the first 5000 rows were considered due to limitations in the `shapiro.test()` function

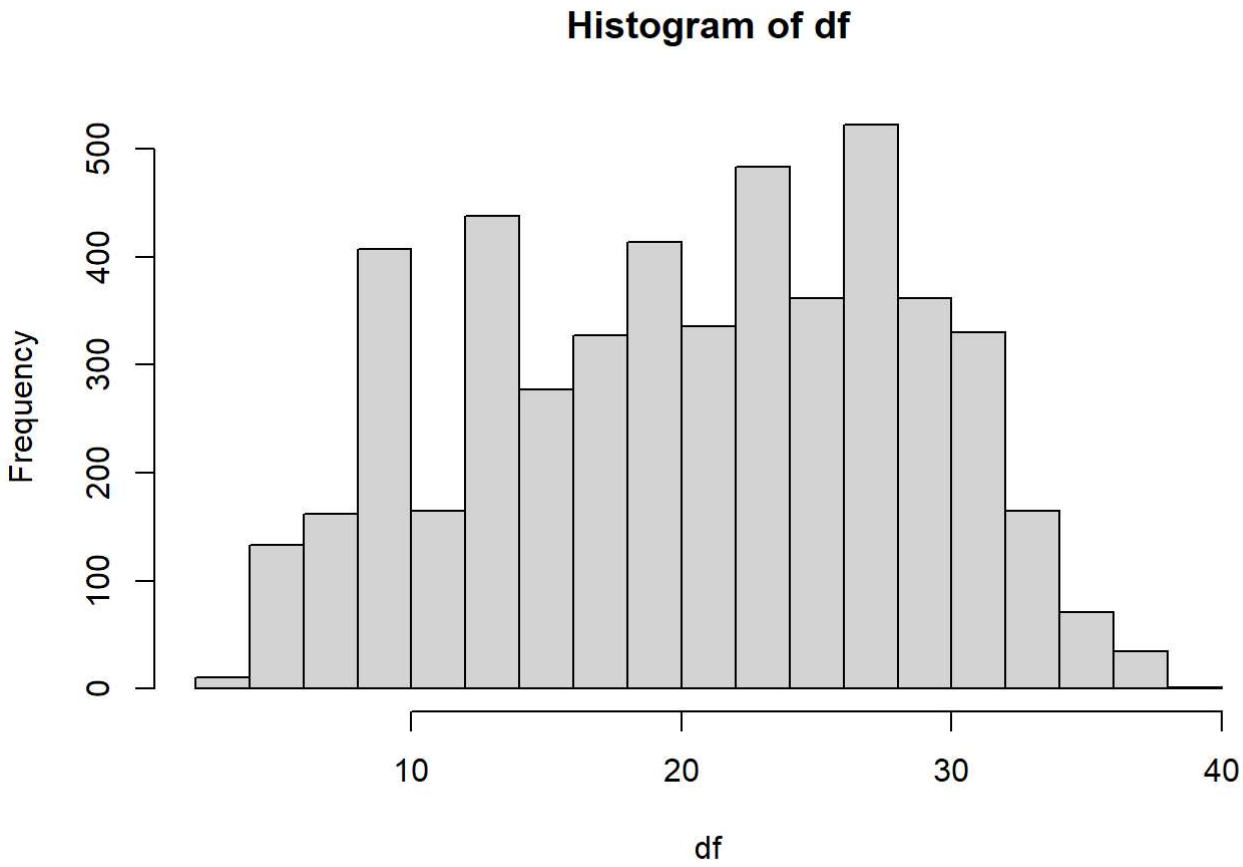
```
excel_path <- 'D:\\FILES\\BRSM_Assignment_2_datasets.xlsx'
data <- read_excel(excel_path, sheet=3)
data
```

datetime <dttm>	sea… <dbl>	holiday <dbl>	workingday <dbl>	weather <dbl>	temp <dbl>	atemp <dbl>	humidity <dbl>	windspe… <dbl>
2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0000
2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0000
2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0000

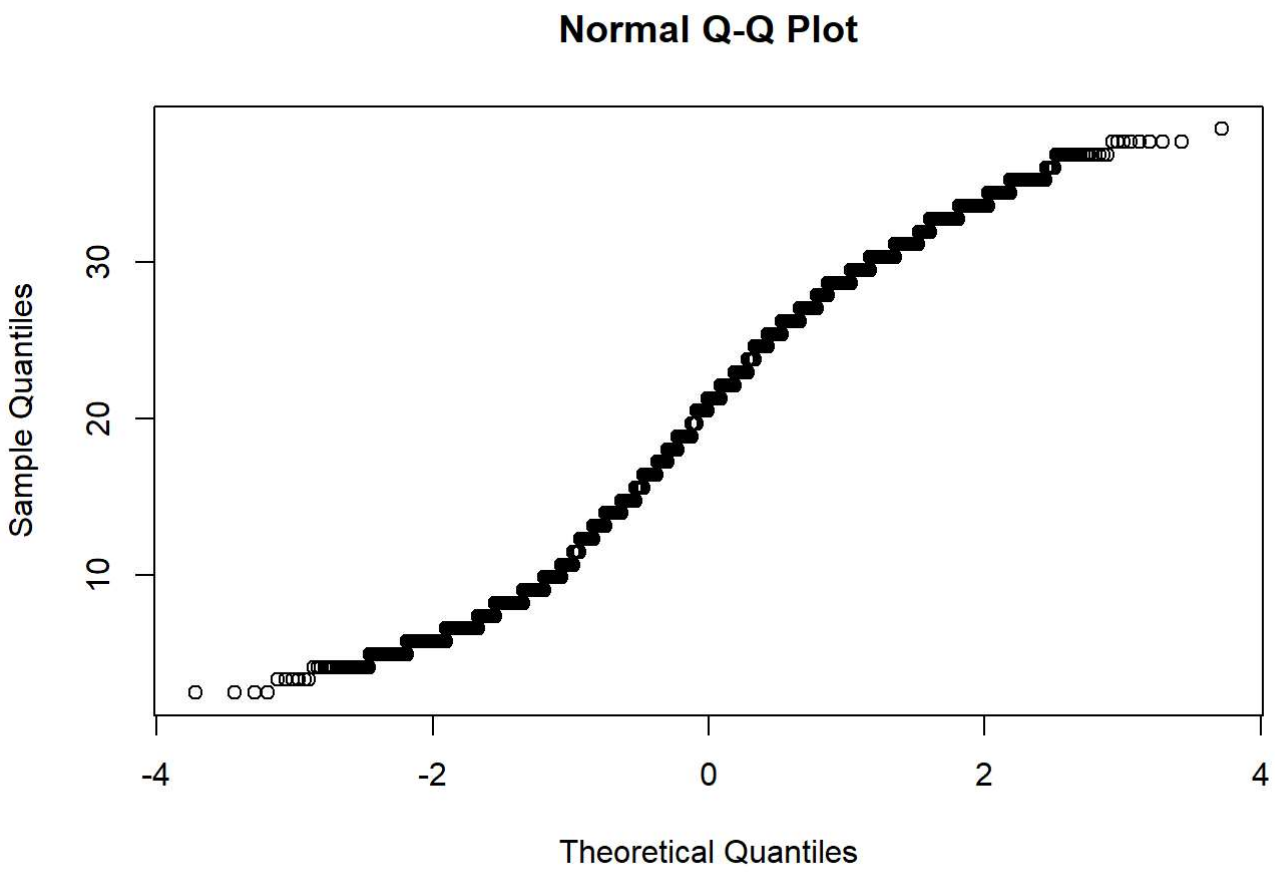
datetime <dttm>	sea… <dbl>	holiday <dbl>	workingday <dbl>	weather <dbl>	temp <dbl>	atemp <dbl>	humidity <dbl>	windspe… <dbl>
2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0000
2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0000
2011-01-01 05:00:00	1	0	0	2	9.84	12.880	75	6.0032
2011-01-01 06:00:00	1	0	0	1	9.02	13.635	80	0.0000
2011-01-01 07:00:00	1	0	0	1	8.20	12.880	86	0.0000
2011-01-01 08:00:00	1	0	0	1	9.84	14.395	75	0.0000
2011-01-01 09:00:00	1	0	0	1	13.12	17.425	76	0.0000

Temperature

```
df <- data$temp[1:5000]
hist(df)
```




```
qqnorm(df)
```

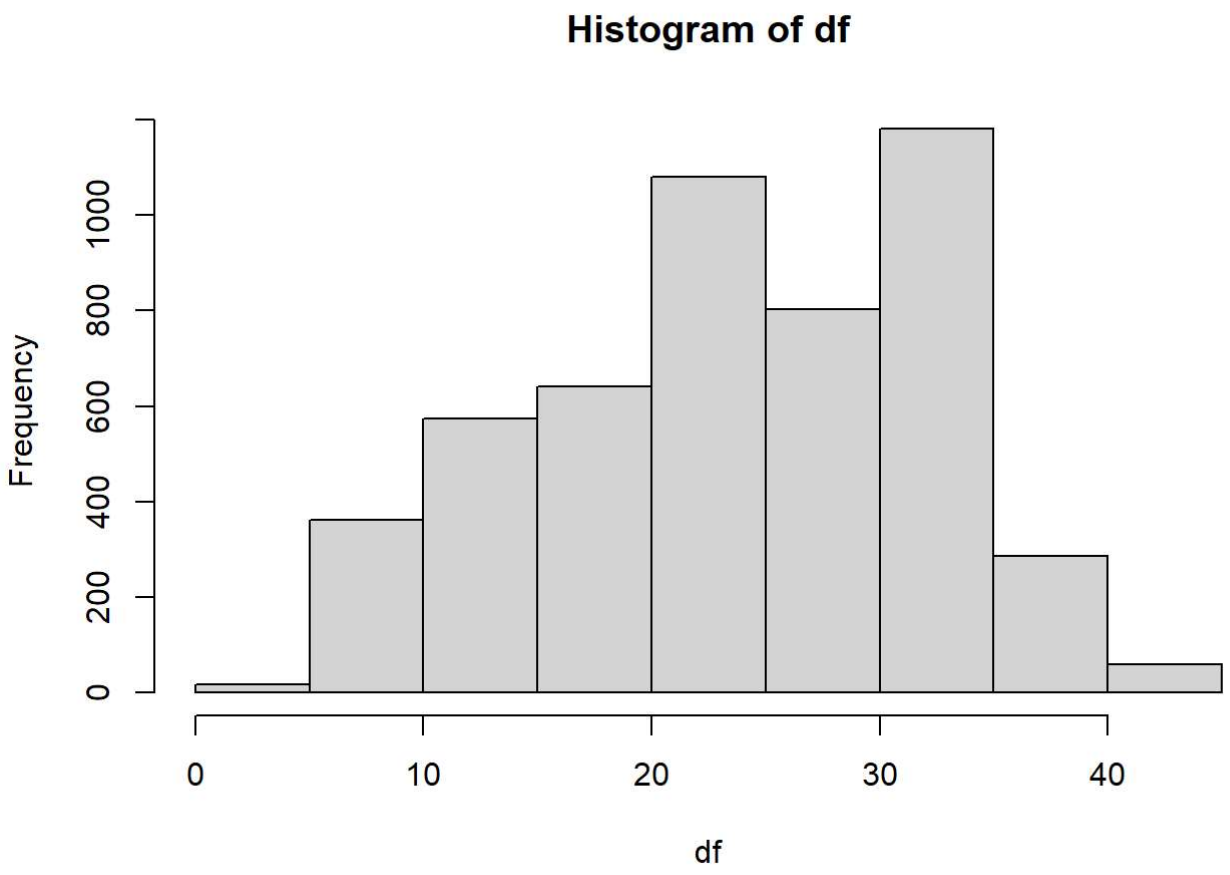



```
shapiro.test(df)$p.value
```

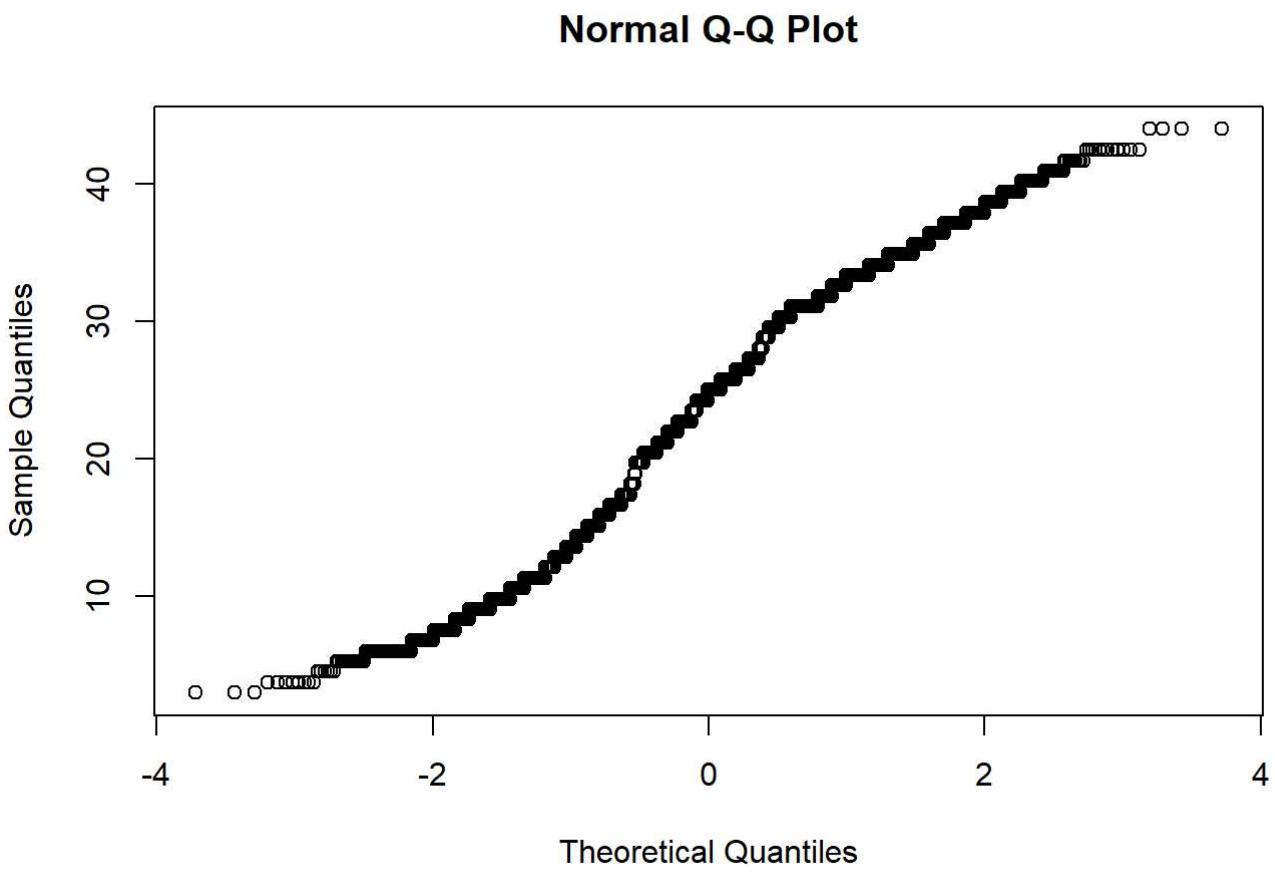
```
## [1] 1.522271e-28
```

Air Temp

```
df <- data$atemp[1:5000]
hist(df)
```


Histogram of df
Frequency
0
200
400
600
800
1000
0
10
20
30
40
df

```
qqnorm(df)
```

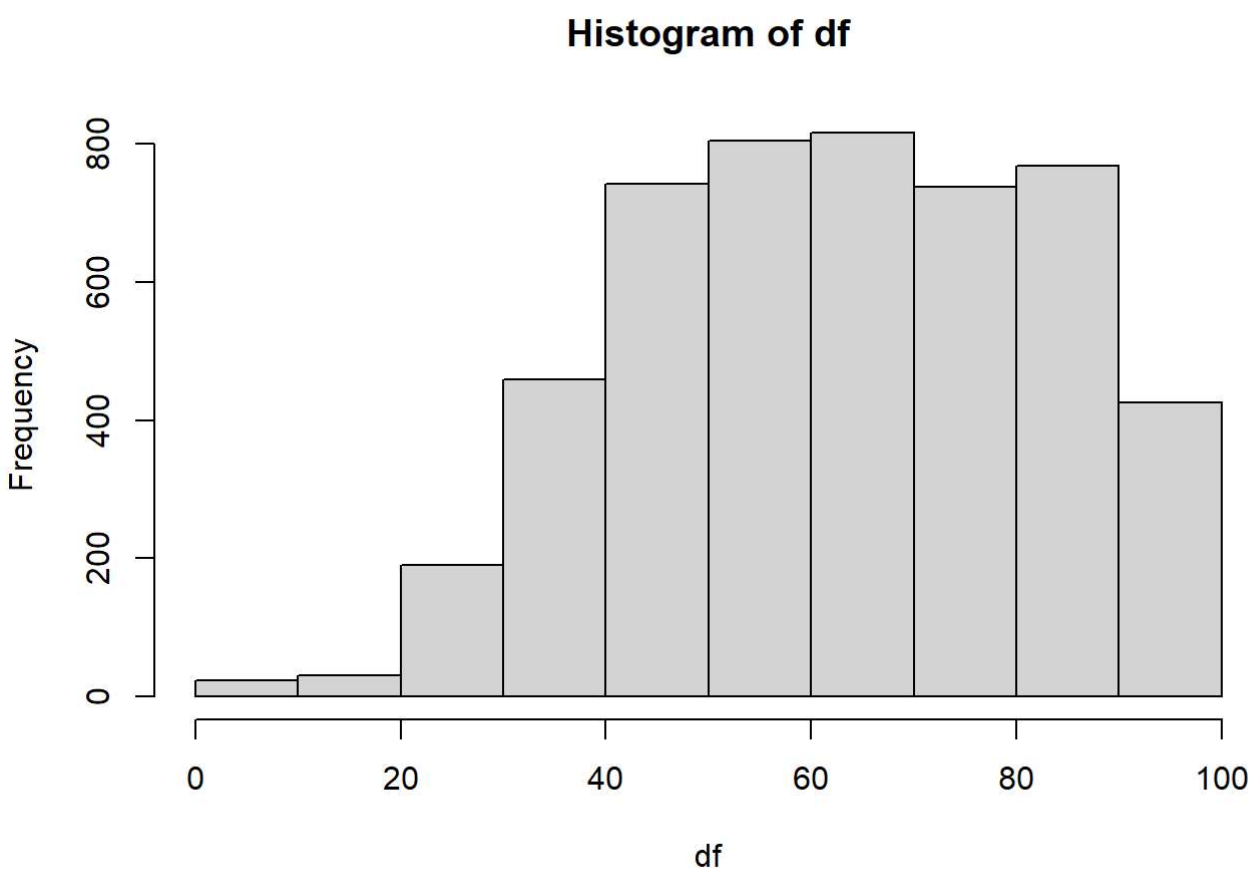



```
shapiro.test(df)$p.value
```

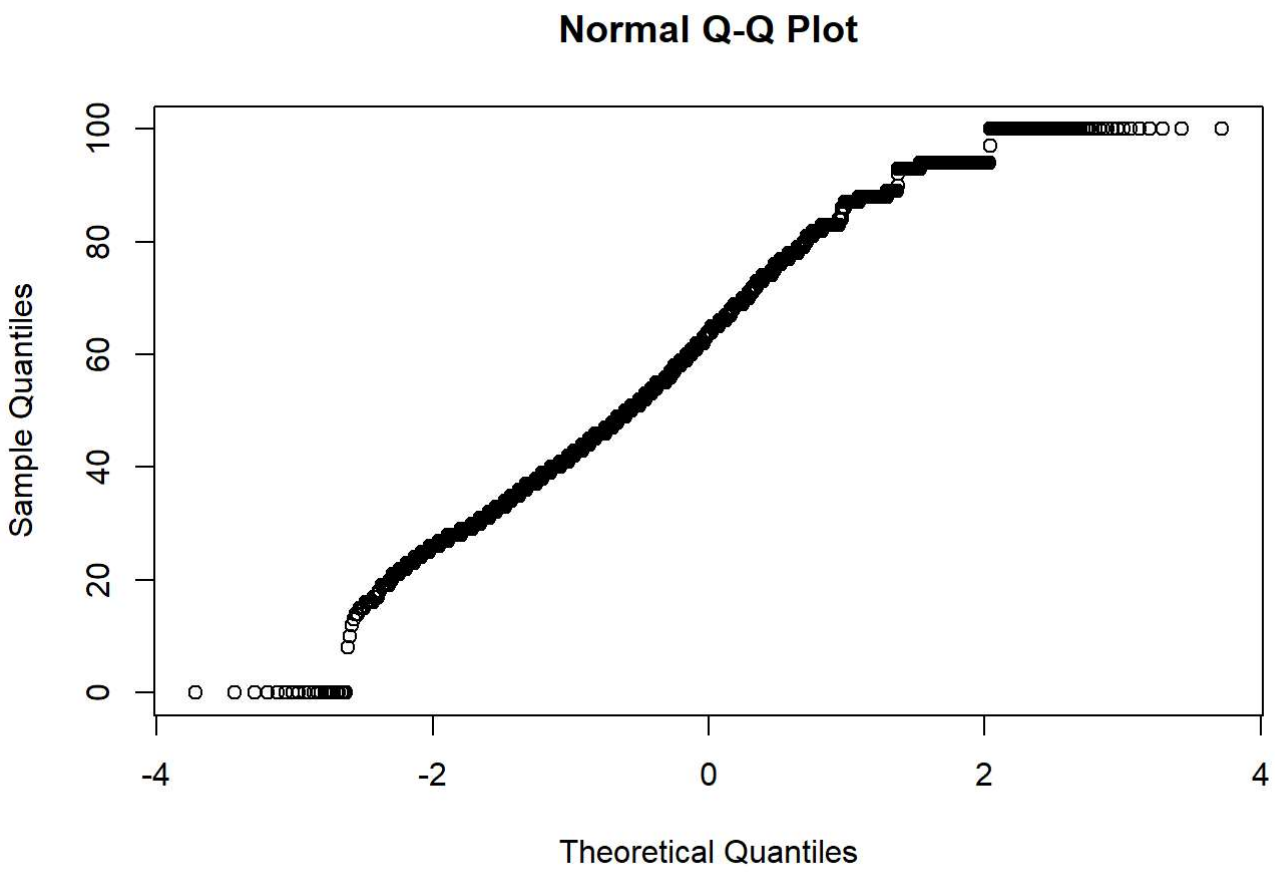
```
## [1] 7.380894e-29
```

Humidity

```
df <- data$humidity[1:5000]
hist(df)
```




```
qqnorm(df)
```

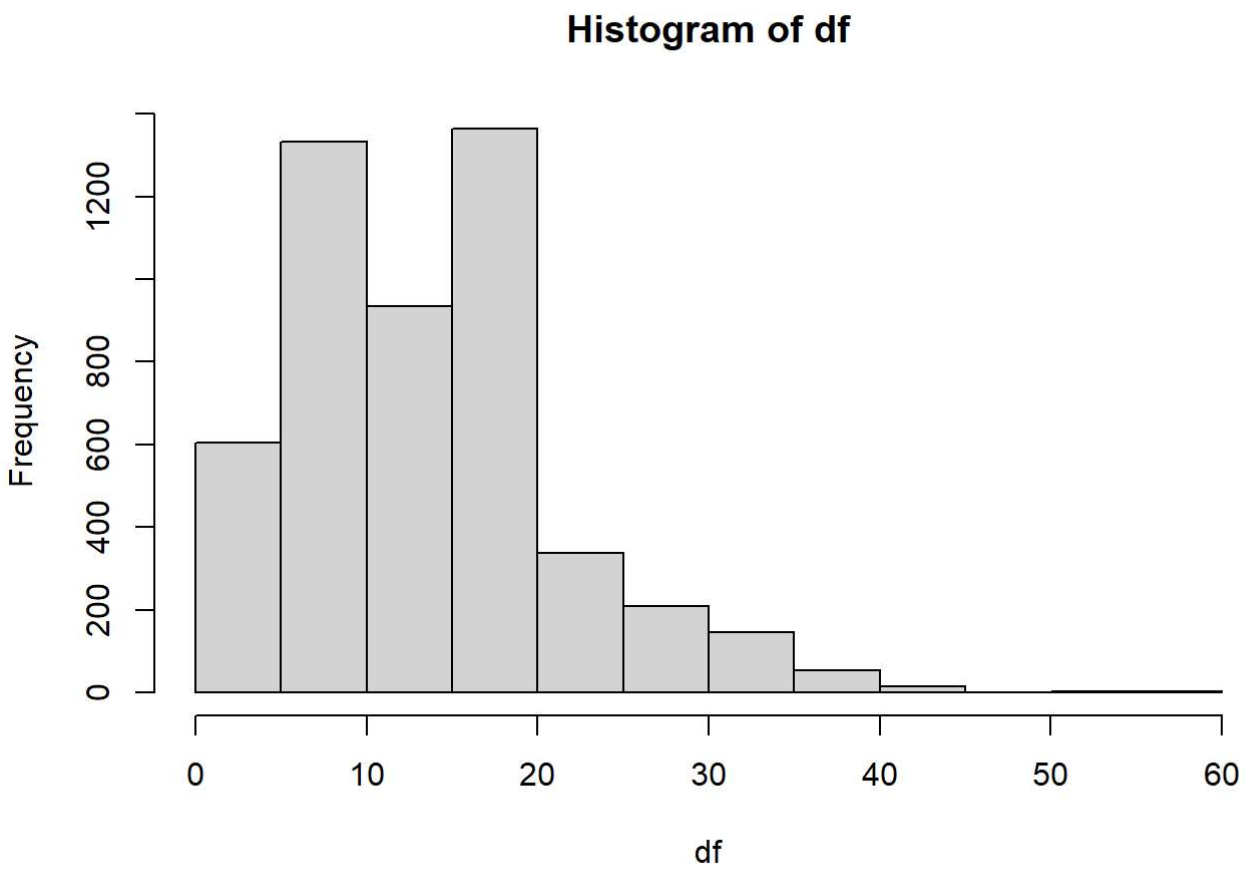



```
shapiro.test(df)$p.value
```

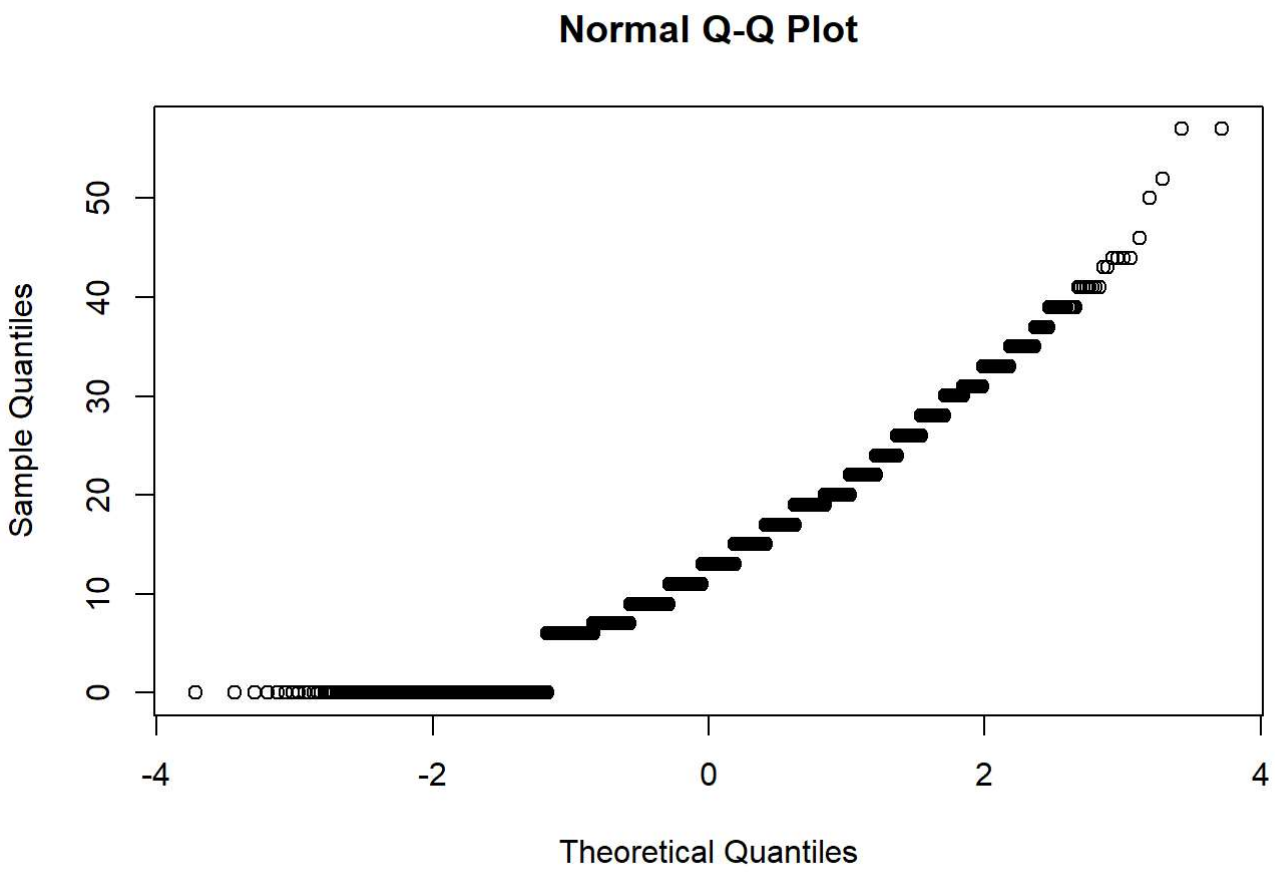
```
## [1] 1.318284e-25
```

Windspeed

```
df <- data$windspeed[1:5000]
hist(df)
```


Histogram of df
Frequency
0
200
400
600
800
1200
0
10
20
30
40
50
60
df

```
qqnorm(df)
```

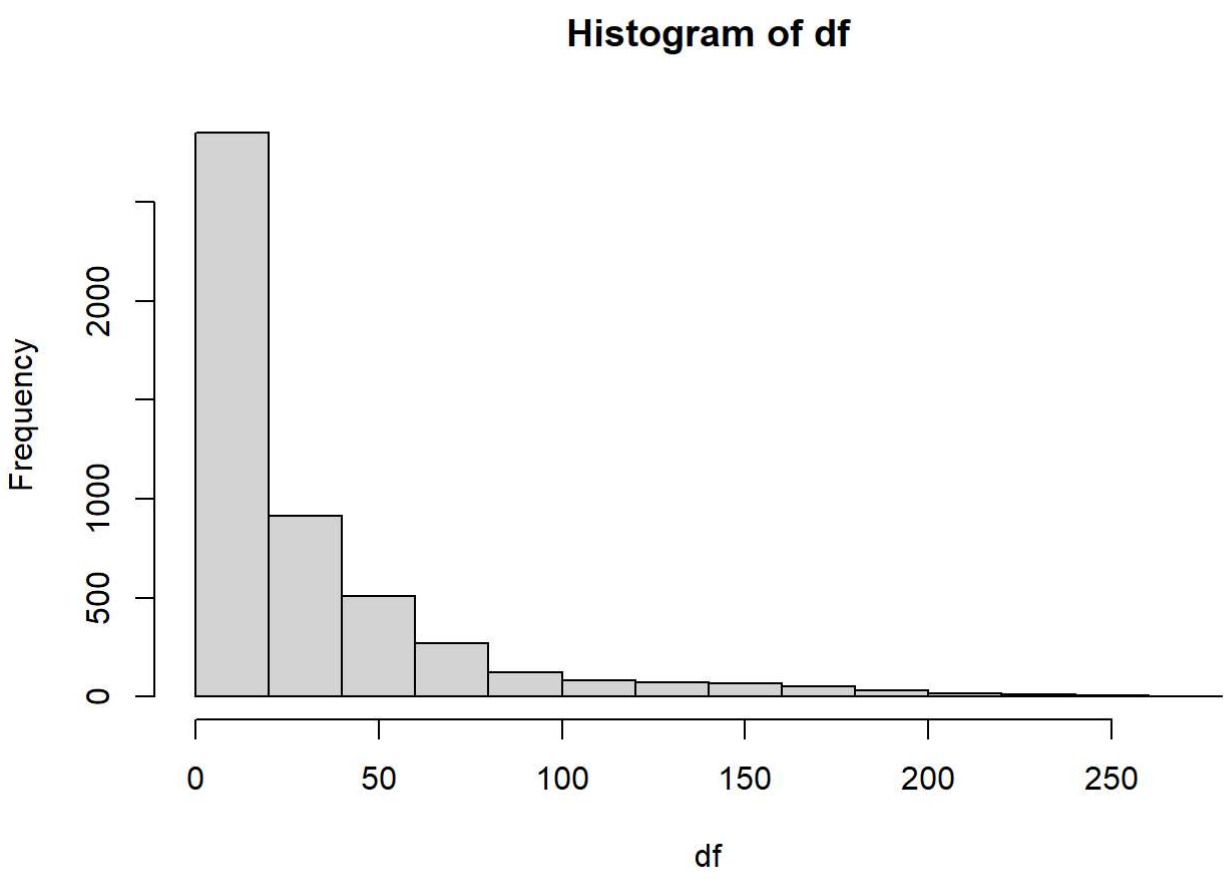



```
shapiro.test(df)$p.value
```

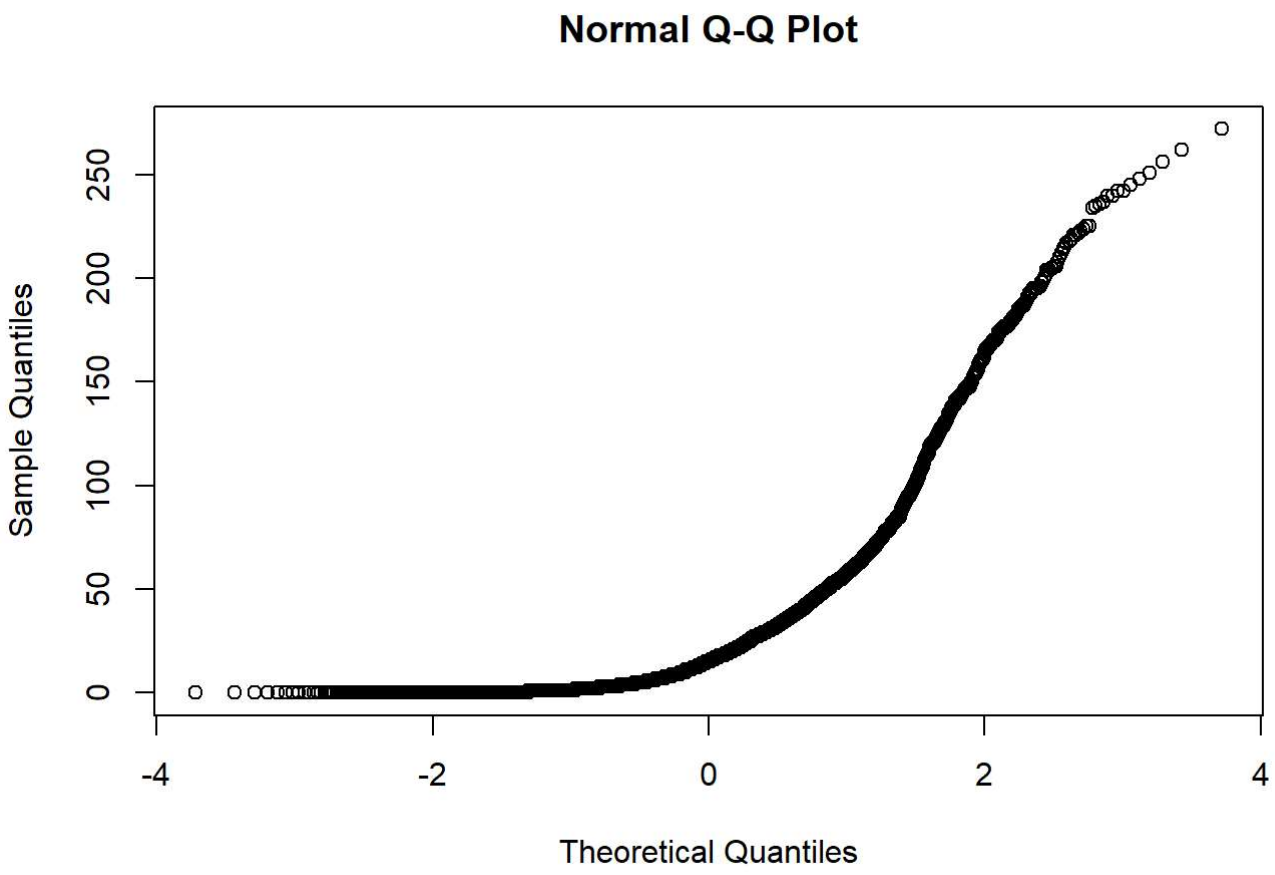
```
## [1] 9.358795e-35
```

Casual Users

```
df <- data$casual[1:5000]
hist(df)
```


Histogram of df
Frequency
0
500
1000
2000
0
50
100
150
200
250
df

```
qqnorm(df)
```

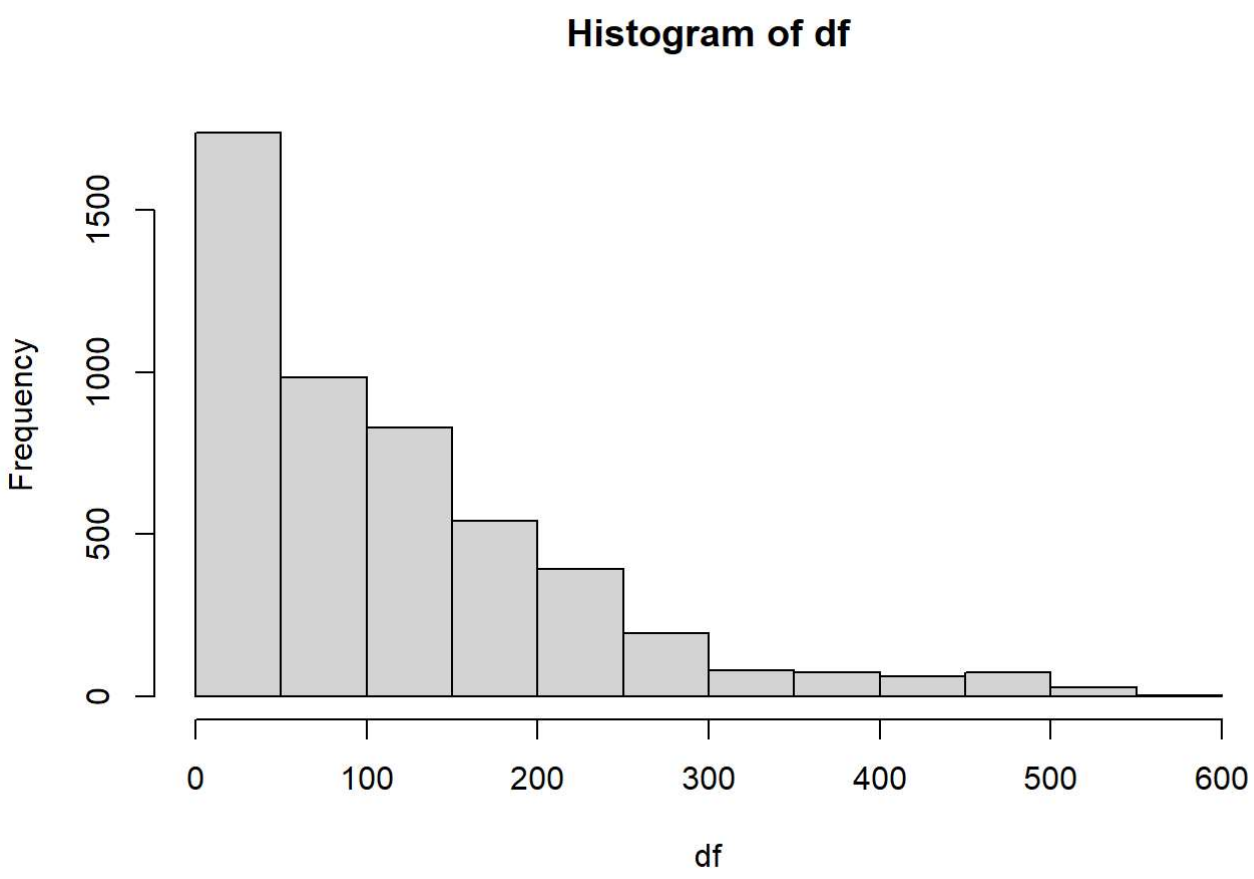



```
shapiro.test(df)$p.value
```

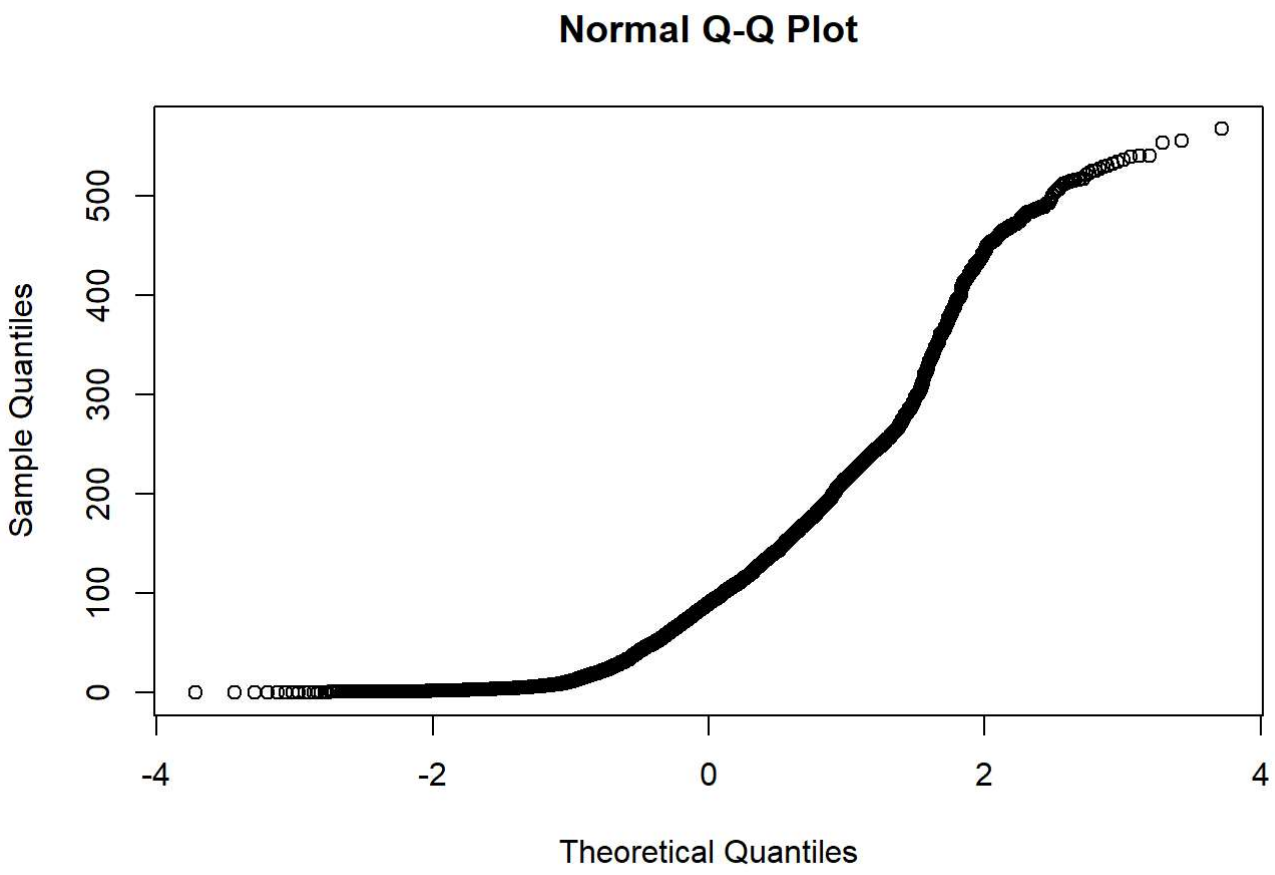
```
## [1] 4.802049e-68
```

Registered

```
df <- data$registered[1:5000]
hist(df)
```


Histogram of df
Frequency
0
500
1000
1500
0
100
200
300
400
500
600
df

```
qqnorm(df)
```




```
shapiro.test(df)$p.value
```

```
## [1] 2.198769e-54
```

Question 3.2

All data seems to be normally distributed according to the Shapiro-Wilk tests. Doubtfully, the data is left without transformations.