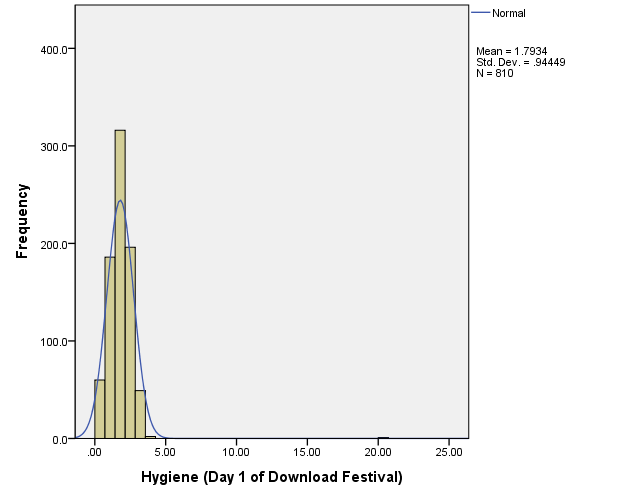
**Practical 4:**

**Part A:  Assessing Normality. (**DownloadFestival.sav)

**Question 1:** Check the normality for “day 1”.

① Graphical Displays

* 1. Normal Curve



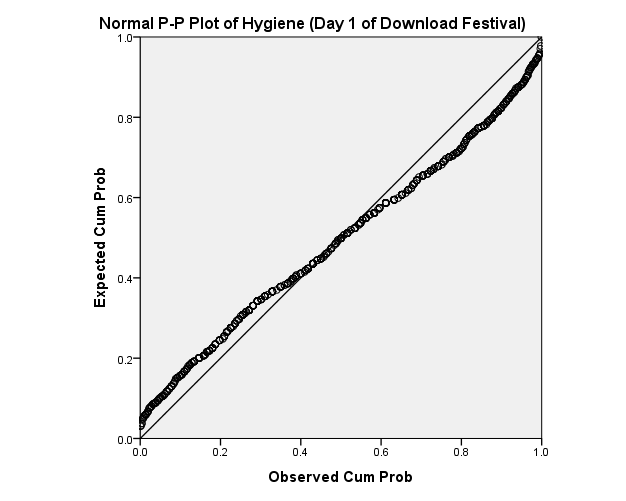
It’s not a normal distribution because there is a outlier

nomal curve -> day 1 에 대해서

graph > histogram > nomal curve

nomal이 아님> outlier 가 있으므로

* 1. P-P Plot

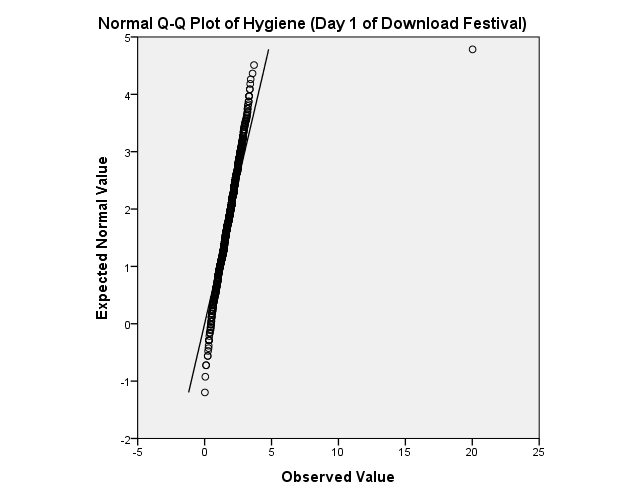


/nomal 이기 위해선 모든 점들이 line 내에 있어야 하지만 이 경우 벌어진 점들이 많이 있으므로 nomal이 아니다.

It’s not a normal distribution because there are many data points which are not near to the reference line.

Analyze > descriptive s~ > pp plot

* 1. Q-Q Plot



/It’s not nomal distribution , because there are some points that not in the R-line.

② Values of Kurtosis/Skewness

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| Hygiene (Day 1 of Download Festival) | | |
| N | Valid | 810 |
| Missing | 0 |
| Skewness | | 8.865 |
| Std. Error of Skewness | | .086 |
| Kurtosis | | 170.450 |
| Std. Error of Kurtosis | | .172 |

The values of skewness(8.865)and kurtosis(170.450)are far from 0.

Therefore, it’s not a normal distribution.

Analyze > d-s > frequency > statics 에서 Skewness , kurtosis체크할 것.

Skewness , kurtosis 가 0에 가까워야 nomal)(0.4 정도보다 작아야 nomal로 봄 크면 nomal 아님)

③ Kolmogorov-Smirnov and Shapiro-Wilk Test

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| Hygiene (Day 1 of Download Festival) | .083 | 810 | .000 | .654 | 810 | .000 |
| a. Lilliefors Significance Correction | | | | | | |

:It’s a normal distribution. > normality can be assume.

:It’s not a normal distribution.

K-S Test

D(810) = 0.083, sig = 0.000(<0.05)

This test is significant

Reject

Conclusion : It’s not a normal distribution.

만약 아래 설명과 같이 conflict 라 S-W썼다면 D(810) = 0.654

Analyze > ds > explore >depend list 에 넣고 plot > normality plot with tests.

앞칸 sig 는 >0.05고 sig < 0.05를 ,conflict 라 함 conflict 의 경우 S-W사용

아닌경우 K-S사용.

④ Identify the outlier in “day1”.

|  |  |
| --- | --- |
| ticknumb | day1 |
| 4158 | 20.02 |
|  |  |

별은 무조건 outlier 이므로 수정하거나 제거해야함

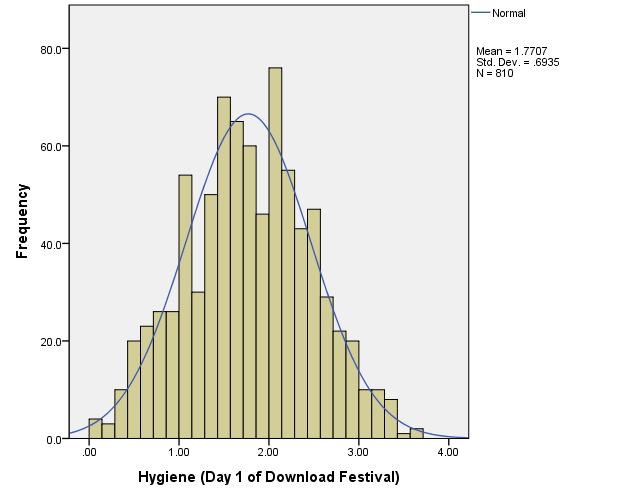
But 원의 경우는 별을 제거하거나 수정했을 때 정상범위로 인식 될 수 있음.

+ 이 경우 box plot을 사용해야 outlier 를 찾을 수 있음.

그래서 처음에 3.69를 넣었지만 아래로 내려가 확인 해 보니 outlier 가 아님을 알 수 있었음, 그래서 지웠음.

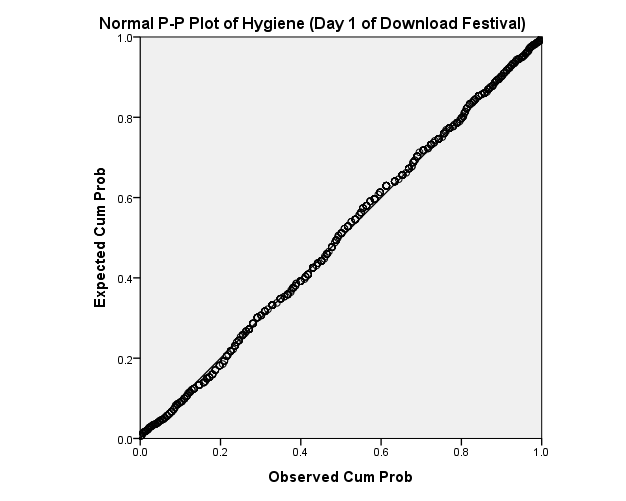
⑤ Modify the outlier value to “1.65” and repeat the following tests.

1. Normal Curve



It’s approximately normal because there is a symmetric line.

1. P-P Plot



모든 점들이 line 에 가깝게 거의 같이 있으므로 normal 이다.

It’s a normal because there are all point near to the line.

**Part B:  Assessing Homogeneity of Variance (Smoking\_Survey.sav)**

**Question 2:**

Check the homogeneity of variance for Age (grouped by Exercise Status)

Explore 에서 depend 는 scale , factor 는 nominal

Plot 들어가서 transformed 찍고감.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Test of Homogeneity of Variance** | | | | | |
|  | | Levene Statistic | df1 | df2 | Sig. |
| The age of the subject | Based on Mean | .985 | 2 | 56 | .380 |
| Based on Median | .499 | 2 | 56 | .610 |
| Based on Median and with adjusted df | .499 | 2 | 36.562 | .611 |
| Based on trimmed mean | .872 | 2 | 56 | .424 |

이게 levene 테스트 임.

:Homogeneity of variance can be assumed

:--- ----------------------------can not----------

Levene’s Test

F(2,56) = 0.985, sig = 0.380 ( >0.05)

This test is non-significant

Accept

Conclusion: The variances are about the same in different groups.

**Part C:  Correcting Data Problem.**

**Question 3: (EssayMarks.sav)**

① By using the COMPUTE function, transform essay marks into the following new

variables:

* + 1. Log\_Marks

Transform > Compute variable >

Log로 넣는 것을 의미함

* + 1. SquareRoot\_Marks

Root로 넣는 것을 의미함

* + 1. Reciprocal\_Marks

역수로 바꿔주는 것을 의미함 이 경우 arithmetic 에 함수가 없으므로 그냥 값을 분모로 넣는 ex ) 1/x로 사용

② Complete the following table for the above datasets :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Tests of Normality** | | | | | | | |  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | | | Statistic | df | Sig. | Statistic | df | Sig. | | Essay Mark (%) | .110 | 45 | .200\* | .977 | 45 | .518 | | Log\_Marks | .120 | 45 | .108 | .968 | 45 | .241 | | Sqrt\_Mars | .114 | 45 | .170 | .974 | 45 | .404 | | Reciprocal\_Marks | .133 | 45 | .043 | .947 | 45 | .038 | | \*. This is a lower bound of the true significance. | | | | | | | | a. Lilliefors Significance Correction | | | | | | |   작업 진행시 sig 가 원래 >0.05라서 normal 했지만  Transform 진행시 normal 이 아니게 되어버렸음 > 이는 아주 안좋은 경우임 |

③ Which is the BEST transformation? Why?

**The original data gave the best result for both tests(K-S, S-W)**

**Therefore, it’s not necessary to perform the transformation.**