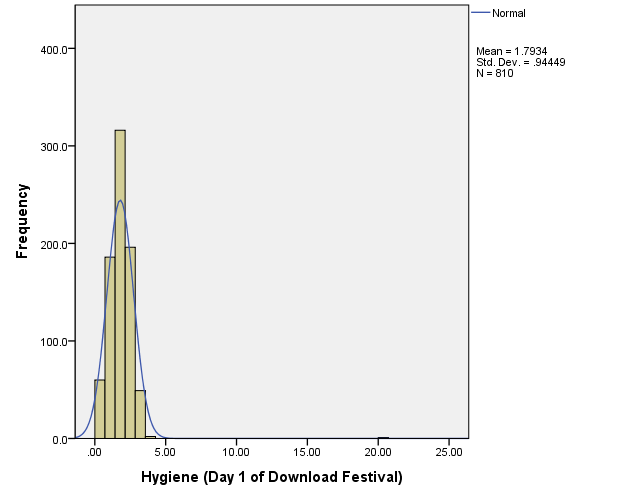
**Practical 4:**

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

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

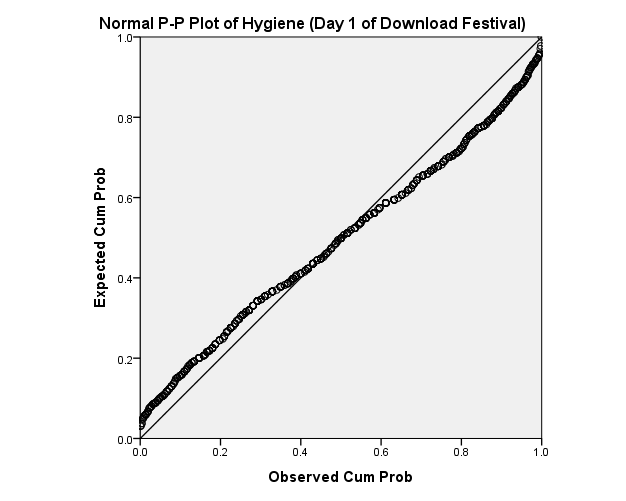
① Graphical Displays

* 1. Normal Curve



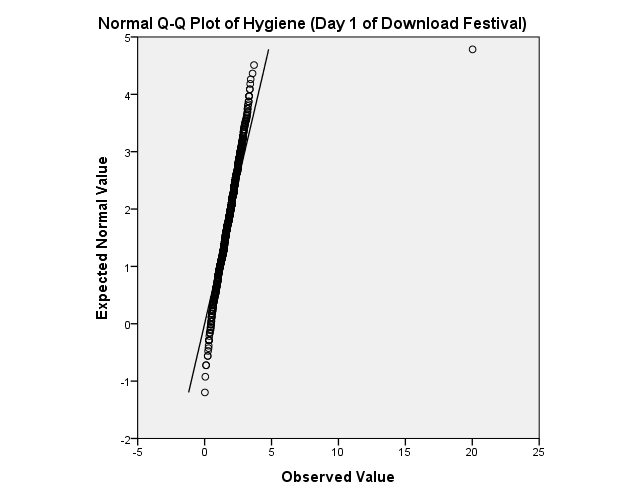
It’s not a normal distribution because there is an outlier

* 1. P-P Plot



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

* 1. Q-Q Plot



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

② Values of Kurtosis/Skewness // (Analyze -> de -> freq)

|  |  |  |
| --- | --- | --- |
| **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 for skewness (8.865) and kurtosis (170.450) are far from 0. Therefore, is is not a normal distribution

//0.4 이상차이나면 not normal

③ Kolmogorov-Smirnov and Shapiro-Wilk Test// (Analyze -> de -> explore)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **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 | | | | | | |

Normality can be assumed.

Normality cannot be assumed.

K-S Test:

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

This test is significant.

Reject

Conclusion : It’s not a normal distribution.

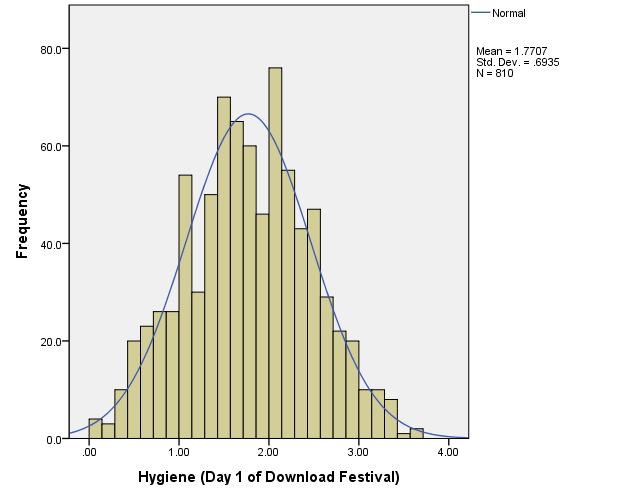
④ Identify the outlier in “day1”.

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

//(별로 찍혀있는 확실한 outlier)

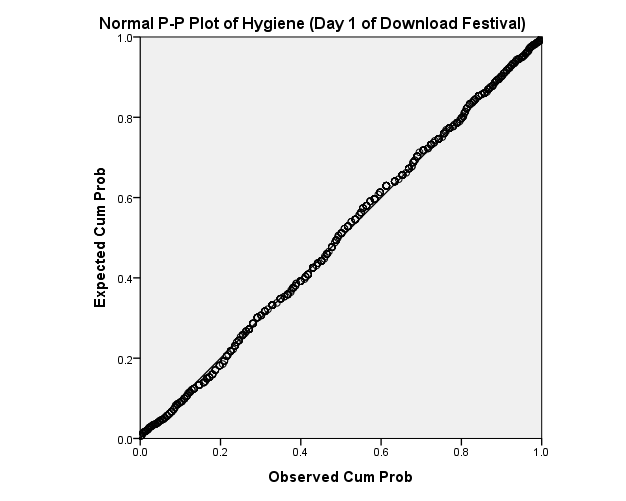
⑤ Modify the outlier value to “1.65” and repeat the following tests.

1. Normal Curve



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

1. P-P Plot



It’s a normal distribution because data points are near to the reference line.

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

**Question 2:**

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

// Analize -> de -> explore

// dependent list = age

//Factor list = exercise

// plots, check transformed power

**//조사의 사이즈는 상관없이 homo-variance는 인접**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |

Homogeneity of variance can be assumed.

Homogeneity of variance cannot be assumed.

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.

// F(df1,df2) = Leven

**Part C:  Correcting Data Problem.**

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

**// 결과값은 무조건 scale data**

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

variables:

* + 1. Log\_Marks // Transform -> Compute, Arithmetic -> Lg10
    2. SquareRoot\_Marks // T -> C , Sqrt
    3. Reciprocal\_Marks // 1/

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 |
| SquareRoot\_Marks | .114 | 45 | .170 | .974 | 45 | .404 |
| Reciprocal\_Mark | .133 | 45 | .043 | .947 | 45 | .038 |
| \*. This is a lower bound of the true significance. | | | | | | |
| a. Lilliefors Significance Correction | | | | | | |

③ Which is the BEST transformation? Why?

// data가 normal -> not need transformation

The original data gave the best result for both test(K-S and S-W). Therefore,it’s not necessary to perform the transformation.