STAT 8120 – Applied Experimental Design

Lab 1 Report – Due 1/12/2020

Connor Armstrong

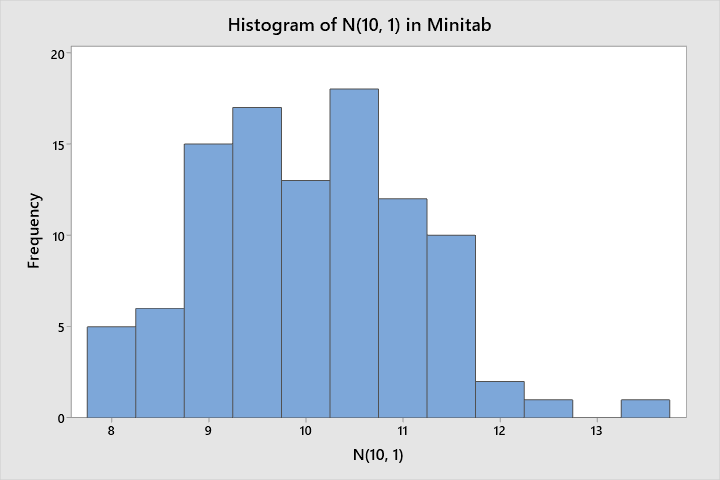
The purpose of this report is to fulfill the requirement for Module 1 Lab, according to the supplied lab documentation, *S8120Lab1d020517.pdf*. SAS and Minitab are utilized in order to output some random and normally distributed data which will we analyzed in this report. These programs are also used as analytical tools to address the questions in the lab document.

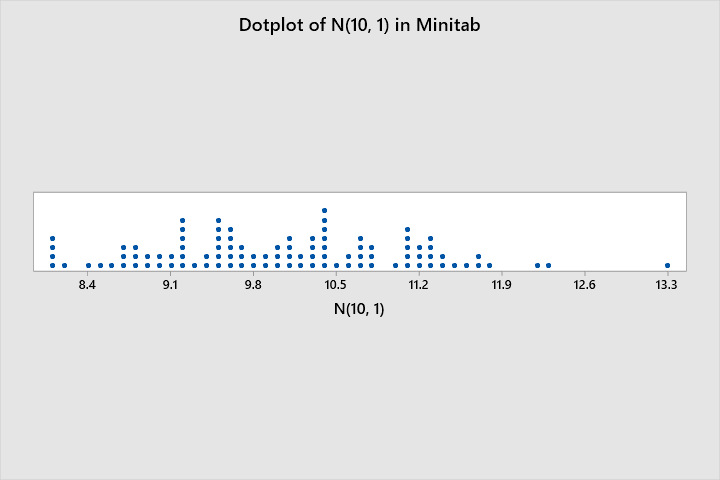
1. *Simulate n=100 values from a N(10, 1) distribution (Mean=10, σ2=1). [Commands: Calc > Random Data > Normal with σ as input NOT σ2.] Plot a histogram, Dot Plot and Probability Plot of the data. Calculate the descriptive statistics, compare to population values, and test for Normality.*

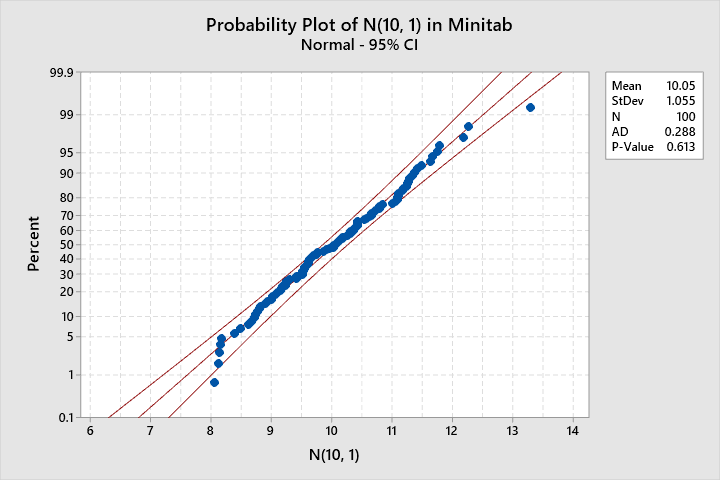
Table 1: 100 Normally Distributed Values with N(10, 1) Generated using Minitab

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11.43 | 8.82 | 9.50 | 10.37 | 8.61 | 10.54 | 13.31 | 9.53 | 11.09 | 9.95 |
| 9.55 | 11.78 | 9.10 | 11.27 | 8.06 | 8.38 | 10.02 | 9.75 | 8.89 | 10.66 |
| 11.66 | 10.09 | 10.31 | 8.73 | 8.12 | 11.28 | 10.79 | 11.24 | 10.72 | 8.17 |
| 10.71 | 11.10 | 10.13 | 11.63 | 10.62 | 9.54 | 9.23 | 10.79 | 9.61 | 12.18 |
| 9.42 | 11.06 | 10.84 | 9.19 | 8.13 | 11.34 | 8.79 | 10.17 | 9.65 | 9.86 |
| 10.38 | 9.16 | 8.15 | 12.26 | 9.70 | 9.00 | 10.66 | 10.42 | 10.26 | 8.72 |
| 9.52 | 10.11 | 10.43 | 10.27 | 8.48 | 10.43 | 9.92 | 9.29 | 11.00 | 9.51 |
| 9.68 | 9.42 | 9.01 | 10.18 | 9.15 | 8.77 | 11.09 | 11.10 | 10.06 | 11.19 |
| 10.37 | 9.58 | 9.61 | 11.38 | 9.06 | 10.58 | 11.25 | 10.31 | 9.76 | 9.24 |
| 11.18 | 10.43 | 9.63 | 8.93 | 11.49 | 9.60 | 8.67 | 10.03 | 9.23 | 11.75 |

The 100 values in Table 1 were generated with Minitab with a mean of 10 and a standard deviation of 1. These values should approximate a normal distribution with those statistics.







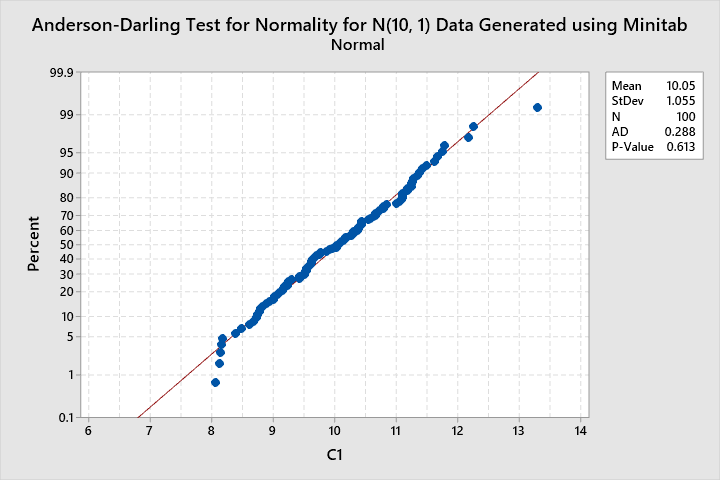
The data approximately follows the contours of a normal distribution in the probability plot. There appear to be at least 2 outliers, but this is not unexpected.

Table 2: Descriptive Statistics for N(10, 1) Generated using Minitab

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N\*** | **Mean** | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** |
| C1 | 100 | 0 | 10.051 | 0.106 | 1.055 | 8.055 | 9.233 | 10.044 | 10.791 | 13.306 |

The calculated statistics for the N(10, 1) data seems appropriate. The mean and standard deviation are close to the requested values.

The following is a Anderson-Darling test for normality completed using Minitab:



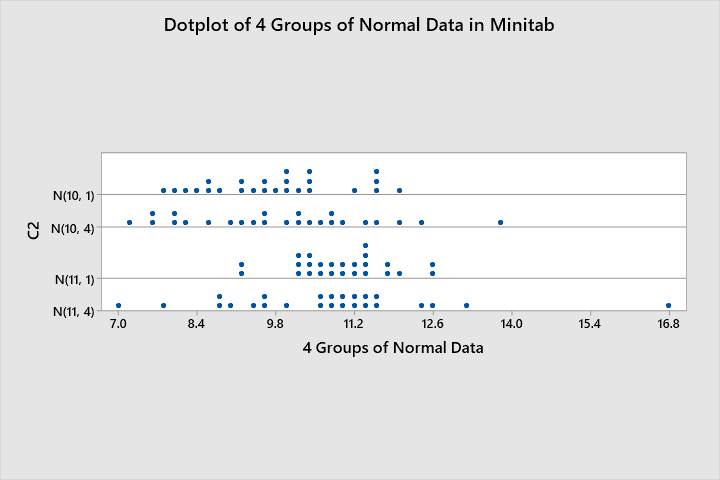
According to the Anderson-Darling Test for Normality, the data is normal. This is known because the P-Value in the above chart is greater than 0.05 (1).

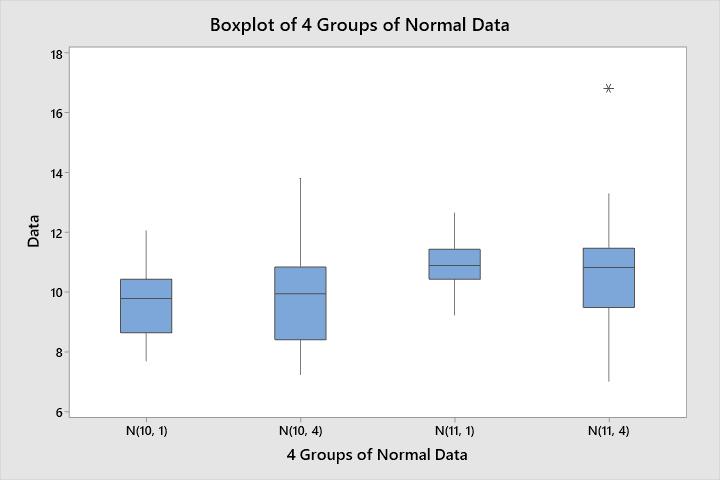
*2) Simulate 4 groups of Normal data labeled: N(10,1), N(10, 4), N(11, 1), and N(11,4) each having n=25 values. You must use statistical tools to uncover difference between the 4 groups. This is like an experimental design problem where initially one is unaware of similarities or differences between studied groups. Use descriptive statistics, Dot Plot, Box Plot, etc. to compare the 4 groups. Compare the groups in a brief summary of what you have learned from your analysis. Compare your conclusions to what you know to be true from the simulation population parameters.*

Table 3: 4 Groups of 25 Normally Distributed Values Generated using Minitab

|  |  |  |  |
| --- | --- | --- | --- |
| **N(10, 1)** | **N(10, 4)** | **N(11, 1)** | **N(11, 4)** |
| 11.62 | 10.20 | 10.42 | 10.55 |
| 8.76 | 7.58 | 10.12 | 12.54 |
| 9.55 | 7.26 | 10.55 | 7.03 |
| 9.20 | 8.19 | 9.26 | 9.50 |
| 9.64 | 9.58 | 9.25 | 10.94 |
| 11.17 | 12.36 | 10.46 | 11.19 |
| 11.56 | 10.73 | 10.90 | 9.50 |
| 10.37 | 8.08 | 10.28 | 8.89 |
| 11.53 | 10.23 | 10.61 | 11.11 |
| 9.92 | 11.43 | 11.73 | 10.83 |
| 10.03 | 13.81 | 11.43 | 11.52 |
| 10.11 | 11.60 | 10.46 | 9.91 |
| 8.35 | 9.61 | 12.64 | 13.29 |
| 10.03 | 11.94 | 10.18 | 10.94 |
| 10.41 | 7.69 | 12.50 | 11.69 |
| 9.32 | 8.65 | 11.74 | 9.48 |
| 8.55 | 9.08 | 11.45 | 12.39 |
| 7.71 | 10.53 | 11.42 | 7.82 |
| 8.55 | 9.32 | 11.36 | 10.74 |
| 9.19 | 10.72 | 10.84 | 8.82 |
| 9.80 | 10.44 | 10.74 | 16.83 |
| 12.06 | 8.04 | 11.08 | 11.42 |
| 8.17 | 9.96 | 11.11 | 10.57 |
| 10.47 | 10.96 | 11.28 | 11.37 |
| 7.96 | 9.14 | 12.08 | 8.95 |

The above table was generated using Minitab and the data within each column was generated according to the normal distribution atop their respective column.





It is clear upon inspecting the dotplots and boxplots for each group that each group has very different characteristics. For example, the first and third groups seem to be more tightly distributed than the second and fourth groups. We can see that the fourth group has an outlier around 17. If the underlying function for each group was not known, a keen eye might be able to tell that the first two groups seem to be distributed around lower means than the second two groups. Exact figures of descriptive statistics for each group will be explored in the following page.

Table 4: Descriptive Statistics for 4 Normally Distributed Groups Generated using Minitab

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Distribution** | **N** | **N\*** | **Mean** | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** |
| N(10, 1) | 25 | 0 | 9.760 | 0.243 | 1.216 | 7.710 | 8.653 | 9.797 | 10.442 | 12.057 |
| N(10, 4) | 25 | 0 | 9.885 | 0.328 | 1.638 | 7.257 | 8.419 | 9.955 | 10.848 | 13.808 |
| N(11, 1) | 25 | 0 | 10.955 | 0.170 | 0.850 | 9.250 | 10.439 | 10.902 | 11.438 | 12.640 |
| N(11, 4) | 25 | 0 | 10.713 | 0.388 | 1.941 | 7.027 | 9.492 | 10.833 | 11.472 | 16.830 |

Basic statistics were run for each group which is displayed in the above table. The means and standard deviation for each group are reasonably close to the input values, considering the small sample size for each group. One would expect for the calculated mean and standard deviation to approach the input values for larger sample sizes. The question was interpreted as for N(10, 25) that 25 is the variance. Therefore, for N(10, 4), a standard deviation of 2 was input into Minitab.

*3) Compare the variances of the five groups in (2) using Stat > ANOVA > Test for Equal Variance.*

Firstly, there were 4 groups requested in question 2. The following is the output of the Test for Equal Variance output in Minitab for the 4 groups in question 2:

**Test for Equal Variances: C1 versus C2**

Method

|  |  |
| --- | --- |
| Null hypothesis | All variances are equal |
| Alternative hypothesis | At least one variance is different |
| Significance level | α = 0.05 |

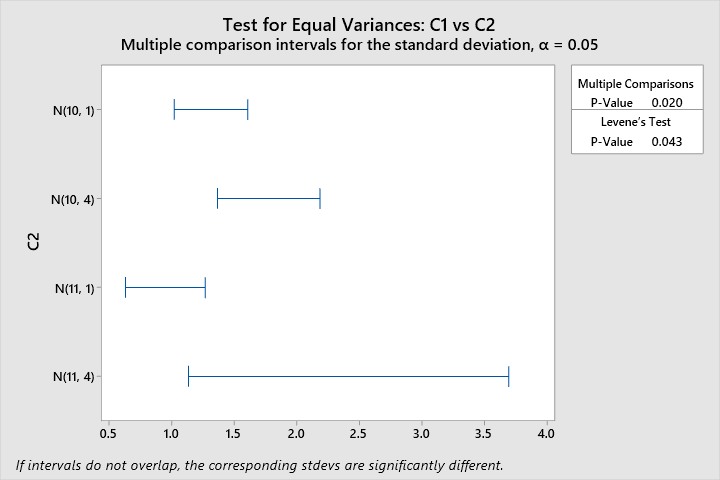
95% Bonferroni Confidence Intervals for Standard Deviations

|  |  |  |  |
| --- | --- | --- | --- |
| **C2** | **N** | **StDev** | **CI** |
| N(10, 1) | 25 | 1.21589 | (0.92841, 1.76915) |
| N(10, 4) | 25 | 1.63787 | (1.17379, 2.53909) |
| N(11, 1) | 25 | 0.85023 | (0.60239, 1.33325) |
| N(11, 4) | 25 | 1.94063 | (1.10447, 3.78829) |

Individual confidence level = 98.75%

Tests

|  |  |  |
| --- | --- | --- |
| **Method** | **Test Statistic** | **P-Value** |
| Multiple comparisons | — | 0.020 |
| Levene | 2.82 | 0.043 |



Because the p-values for both tests are lower than the significance level of 0.05, it can be stated with a 95% confidence rate that the alternate hypothesis is more accurate than the null hypothesis in this case. Specifically, that at least one variance of the 4 groups is different than the others. Of course, this was contrived to be so, but this is an interesting exercise to establish a certain level of familiarity with the software utilized in this course.

*4) To use SAS, it is necessary to stack the data in (2) into one column AllData using the Stack command [Data > Stack > Columns > select columns]. Create a GROUP identifier from the Stack command index. Copy the Minitab GROUP and ALLData columns into an Excel spreadsheet to use the SAS “Import Data” routine. Prepare the computer output in (2) above using SAS. Consider any SAS Procedure, Proc UNIVARIATE and BOXPLOT are useful.*

The SAS output will be found at the end of this document. The justification for providing the full output is that it was explicitly requested to perform these operations in the question. The SAS code is below:

libname lab1 "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 1\Lab 1";

**run**;

**proc** **import** datafile = "C:\Users\conno\OneDrive\Desktop\STAT 8120 - Applied Experimental Design\Module 1\Lab 1\Lab 1\_4.xlsx"

out = lab1.data

DBMS = xlsx Replace;

getnames=no;

**run**;

**data** LAB1.DATA;

set LAB1.DATA;

Rename A = ALLData B=GROUP;

**run**;

**data** LAB1.DATA;

set LAB1.DATA;

label ALLData='ALLData';

label GROUP='GROUP';

**run**;

**proc** **contents** data = lab1.data;

**run**;

/\*DOTPLOT\*/

ods rtf file='temp.rtf';

**PROC** **UNIVARIATE** DATA = LAB1.DATA;

BY GROUP;

**RUN**;

ods rtf close;

/\*BOXPLOT\*/

ods rtf file='temp.rtf';

**proc** **sgplot** data=lab1.data;

vbox alldata/ category=group;

**run**;

ods rtf close;

References

1. Normality Test in Minitab: Minitab with Statistics, Simplilearn, https://www.simplilearn.com/normality-test-in-minitab-article

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 25 | **Sum Weights** | 25 |
| **Mean** | 9.75991425 | **Sum Observations** | 243.997856 |
| **Std Deviation** | 1.21589395 | **Variance** | 1.47839809 |
| **Skewness** | 0.18324773 | **Kurtosis** | -0.7511684 |
| **Uncorrected SS** | 2416.87971 | **Corrected SS** | 35.4815541 |
| **Coeff Variation** | 12.4580392 | **Std Error Mean** | 0.24317879 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 9.759914 | **Std Deviation** | 1.21589 |
| **Median** | 9.796581 | **Variance** | 1.47840 |
| **Mode** | . | **Range** | 4.34684 |
|  |  | **Interquartile Range** | 1.65865 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 40.13473 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 12.5 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 162.5 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 12.05726 |
| **99%** | 12.05726 |
| **95%** | 11.62133 |
| **90%** | 11.55532 |
| **75% Q3** | 10.41442 |
| **50% Median** | 9.79658 |
| **25% Q1** | 8.75577 |
| **10%** | 8.16815 |
| **5%** | 7.95511 |
| **1%** | 7.71042 |
| **0% Min** | 7.71042 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 7.71042 | 18 | 11.1661 | 6 |
| 7.95511 | 25 | 11.5295 | 9 |
| 8.16815 | 23 | 11.5553 | 7 |
| 8.34752 | 13 | 11.6213 | 1 |
| 8.54813 | 17 | 12.0573 | 22 |

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 25 | **Sum Weights** | 25 |
| **Mean** | 9.88519727 | **Sum Observations** | 247.129932 |
| **Std Deviation** | 1.63786759 | **Variance** | 2.68261023 |
| **Skewness** | 0.36036759 | **Kurtosis** | -0.1284918 |
| **Uncorrected SS** | 2507.31077 | **Corrected SS** | 64.3826454 |
| **Coeff Variation** | 16.5688912 | **Std Error Mean** | 0.32757352 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 9.885197 | **Std Deviation** | 1.63787 |
| **Median** | 9.955167 | **Variance** | 2.68261 |
| **Mode** | . | **Range** | 6.55063 |
|  |  | **Interquartile Range** | 2.08616 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 30.17703 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 12.5 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 162.5 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 13.80784 |
| **99%** | 13.80784 |
| **95%** | 12.35883 |
| **90%** | 11.94460 |
| **75% Q3** | 10.73453 |
| **50% Median** | 9.95517 |
| **25% Q1** | 8.64836 |
| **10%** | 7.69322 |
| **5%** | 7.58063 |
| **1%** | 7.25721 |
| **0% Min** | 7.25721 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 7.25721 | 28 | 11.4329 | 35 |
| 7.58063 | 27 | 11.6042 | 37 |
| 7.69322 | 40 | 11.9446 | 39 |
| 8.04154 | 47 | 12.3588 | 31 |
| 8.08041 | 33 | 13.8078 | 36 |

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 25 | **Sum Weights** | 25 |
| **Mean** | 10.9553267 | **Sum Observations** | 273.883166 |
| **Std Deviation** | 0.85023302 | **Variance** | 0.72289619 |
| **Skewness** | -0.0266226 | **Kurtosis** | 0.09648572 |
| **Uncorrected SS** | 3017.82906 | **Corrected SS** | 17.3495085 |
| **Coeff Variation** | 7.7609098 | **Std Error Mean** | 0.1700466 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 10.95533 | **Std Deviation** | 0.85023 |
| **Median** | 10.90174 | **Variance** | 0.72290 |
| **Mode** | . | **Range** | 3.39045 |
|  |  | **Interquartile Range** | 0.97119 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 64.42544 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 12.5 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 162.5 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 12.64044 |
| **99%** | 12.64044 |
| **95%** | 12.50442 |
| **90%** | 12.07944 |
| **75% Q3** | 11.42621 |
| **50% Median** | 10.90174 |
| **25% Q1** | 10.45502 |
| **10%** | 10.12473 |
| **5%** | 9.26399 |
| **1%** | 9.24998 |
| **0% Min** | 9.24998 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 9.24998 | 55 | 11.7348 | 60 |
| 9.26399 | 54 | 11.7373 | 66 |
| 10.12473 | 52 | 12.0794 | 75 |
| 10.18419 | 64 | 12.5044 | 65 |
| 10.27718 | 58 | 12.6404 | 63 |

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 25 | **Sum Weights** | 25 |
| **Mean** | 10.7126368 | **Sum Observations** | 267.815921 |
| **Std Deviation** | 1.94063052 | **Variance** | 3.76604682 |
| **Skewness** | 0.99466398 | **Kurtosis** | 3.2309228 |
| **Uncorrected SS** | 2959.39983 | **Corrected SS** | 90.3851237 |
| **Coeff Variation** | 18.1153394 | **Std Error Mean** | 0.3881261 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 10.71264 | **Std Deviation** | 1.94063 |
| **Median** | 10.83304 | **Variance** | 3.76605 |
| **Mode** | . | **Range** | 9.80280 |
|  |  | **Interquartile Range** | 1.91640 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 27.60092 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 12.5 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 162.5 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 16.83020 |
| **99%** | 16.83020 |
| **95%** | 13.28975 |
| **90%** | 12.53743 |
| **75% Q3** | 11.42067 |
| **50% Median** | 10.83304 |
| **25% Q1** | 9.50427 |
| **10%** | 8.81816 |
| **5%** | 7.81534 |
| **1%** | 7.02740 |
| **0% Min** | 7.02740 |

| **Extreme Observations** | | | |
| --- | --- | --- | --- |
| **Lowest** | | **Highest** | |
| **Value** | **Obs** | **Value** | **Obs** |
| 7.02740 | 78 | 11.6912 | 90 |
| 7.81534 | 93 | 12.3891 | 92 |
| 8.81816 | 95 | 12.5374 | 77 |
| 8.88650 | 83 | 13.2898 | 88 |
| 8.95172 | 100 | 16.8302 | 96 |

SAS Output – Boxplots of 4 Random Normal Groups

