Assessing Normality

There are Statistical processes that **require data** to come from a **Normal Distribution**. As a result, we need to verify that a sample set of data comes from a random variable that is Normally Distributed. What is described below are methods that can be used to address that concern.

Methods

Visual Inspection of a **Histogram**

Identifying Outliers

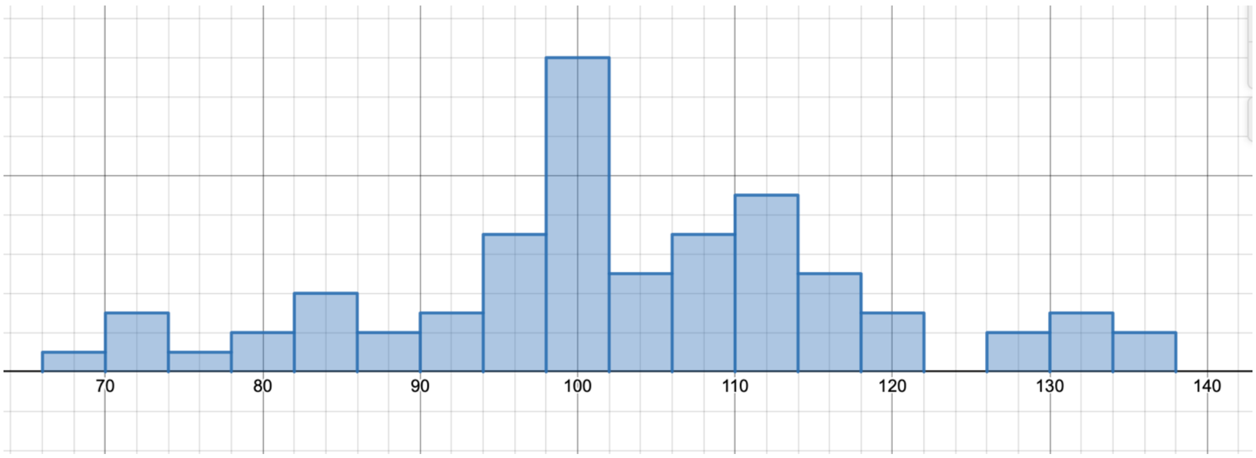
Construct a **Normal Quantile Plot or Quartile Quantile Plot**

Consider the following set of data that we know is **Normally Distributed**. We know **IQ Scores** are Normally Distributed with a Mean of 100 and a Standard Deviation of 15.

Data

	A	B	C	D	E	F	G
1							
2	**IQ Scores**						
3	**Mean**	100					
4	**SD**	15					
5							
6		115	124	92	105	100	
7		85	88	99	95	114	
8		83	103	106	100	97	
9		106	110	90	96	110	
10		83	125	128	90	89	
11		67	116	114	85	106	
12		112	89	100	108	88	
13		114	75	119	36	122	
14		115	111	70	91	124	
15		101	112	109	103	99	
16		121	113	102	114	86	
17		103	102	101	102	90	
18		89	93	92	109	96	
19		99	125	84	85	91	
20		120	117	125	96	124	
21							
22	**Mean**	**102.2**					
23	**SD**	**17.0**					
24							

Histogram

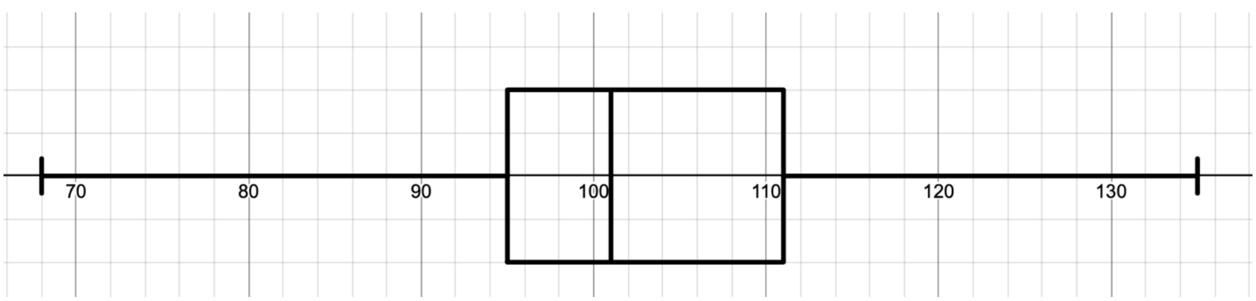


This could be Normally Distributed?

If a Histogram departs radically from a Bell-Shaped curve, you can conclude that data is not Normally Distributed. If the data approximates a Bell-Shaped curve, you can conclude the data is Normally distributed.

Outliers

If there is more than one outlier, then your data might not have a Normal Distribution. One outlier may just be an error and you can through it out as it may have showed up by chance. Just one outlier can have a dramatic influence, so be careful. I constructed a boxplot in Demos to identify potential outliers.



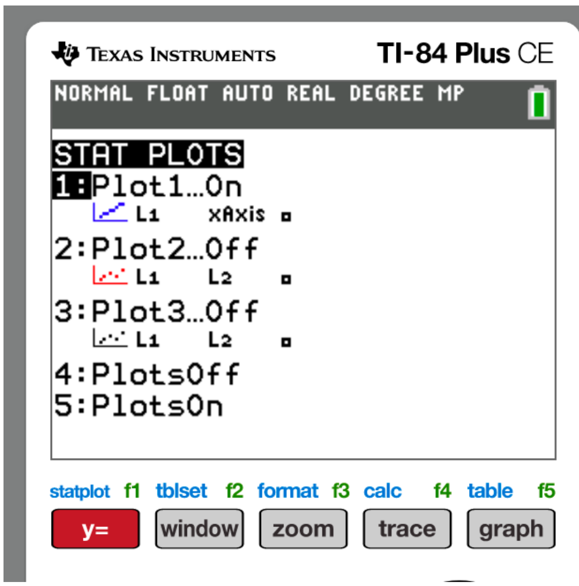
No Outliers!

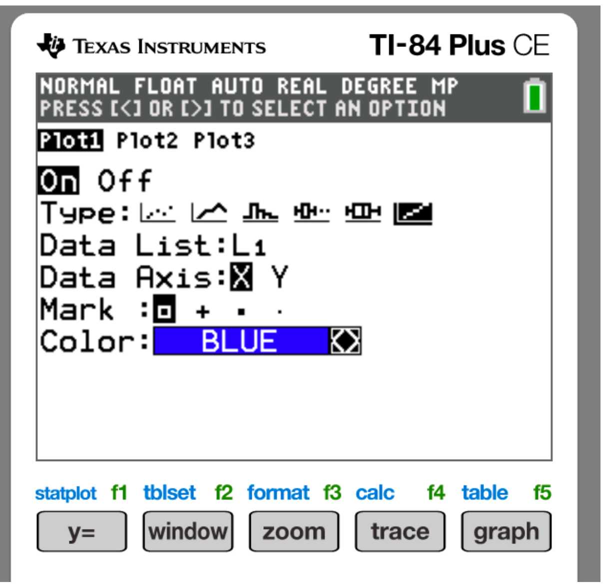
Normal Quantile Plot

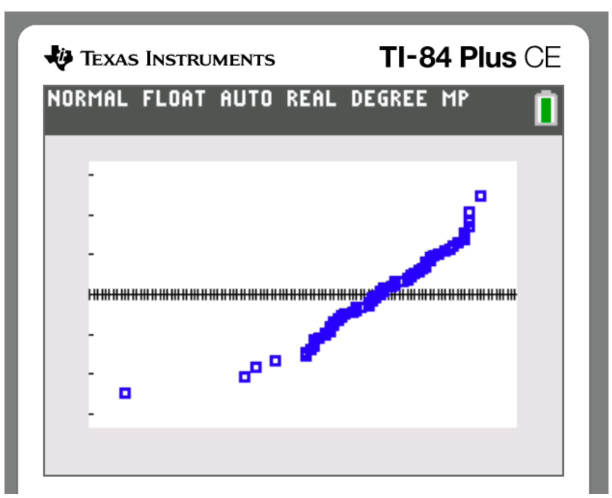
A **Normal Quantile Plot** is also known as a **Normal Probability Plot.** Is a graph of points (x, Z_y) where each x value is the original data value, and the y value is the corresponding z-value that is from the **Standard Normal Probability Distribution**.

If your **Histogram** is symmetric and you have **no more than one outlier**, use technology to generate a **Normal Quantile Plot**.

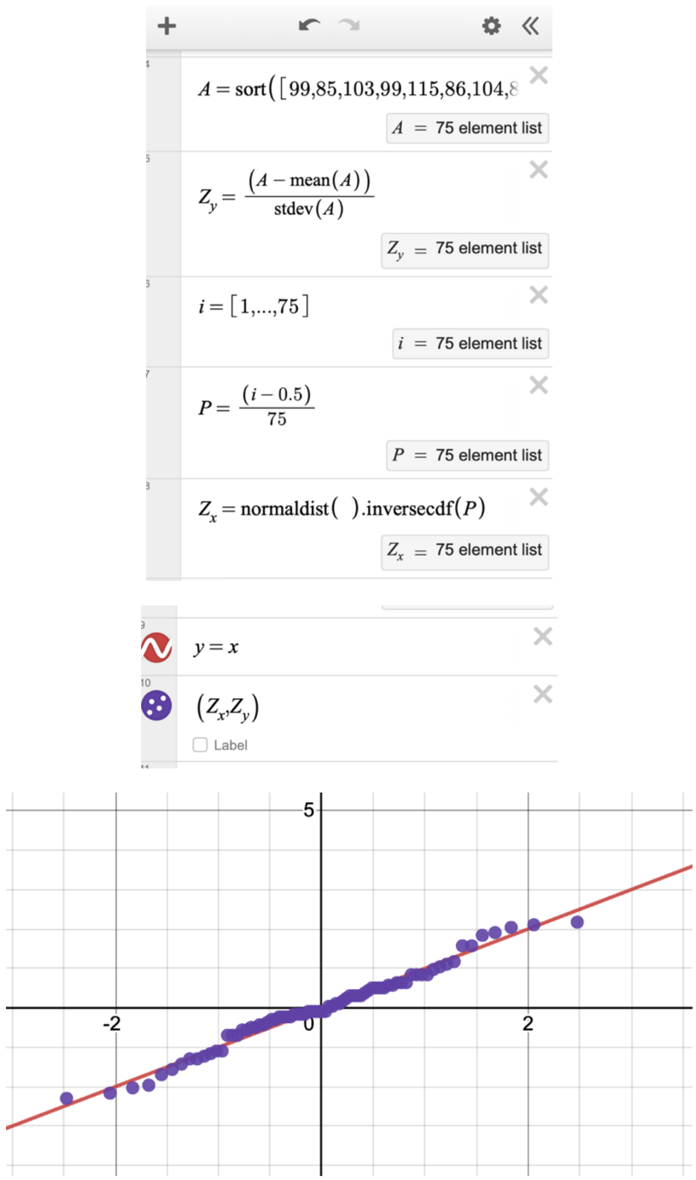
Normal Quantile Plot (x, Z_y)







Quartile Quantile Plot



If the data comes from a Normal Distribution, then you should see the data values (x,y) "reasonably" sketch a **straight line and does not show some sort of symmetric pattern**. In this case, we can conclude that the data is from a Normal Distribution. However, we knew that from the beginning as these were **IQ Scores.**

The Data is Normal!

If you **do not see your data line up in a straight-line pattern** or your data shows some sort of **symmetric properties**, you can conclude the data is not from a Normal Distribution. The following scenario illustrates this point.

Let’s now consider the **die roll simulation** of data as we know it comes from a **Uniform Distribution** and is clearly not Normally Distributed.

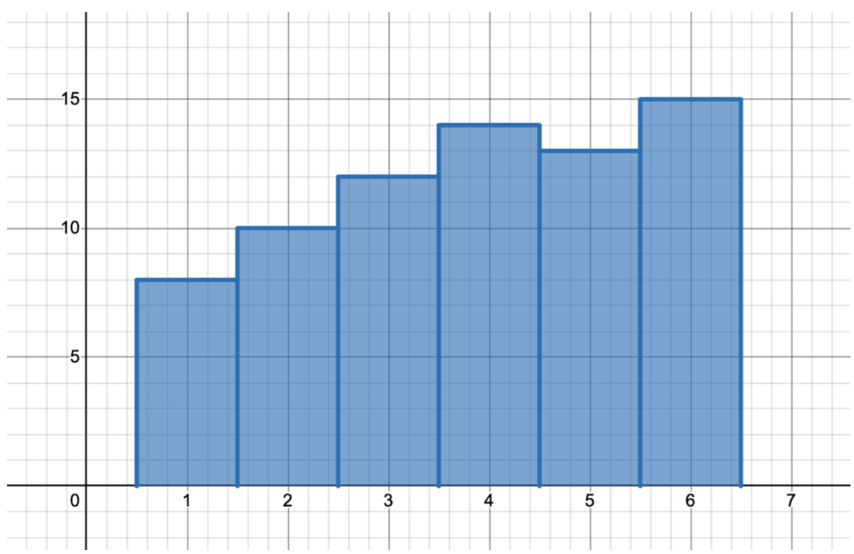
72 Die Rolls

	A	B	C
2	**Roll a Die**		
3	**S={1,2,3,4,5,6}**		
4		**x**	**z**
5		3	-0.15
6		2	-0.74
7		5	1.02
8		2	-0.74
9		2	-0.74
10		5	1.02
11		4	0.44
12		5	1.02
13		5	1.02
14		3	-0.15
15		3	-0.15
16		4	0.44
17		3	-0.15
18		4	0.44
19		3	-0.15
20		4	0.44
21		6	1.61
22		6	1.61
23		6	1.61

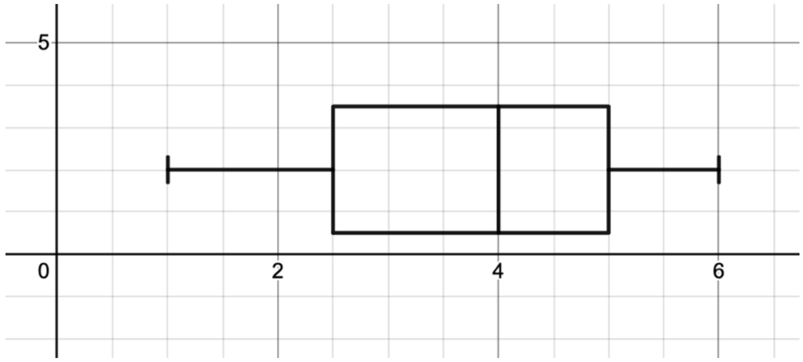
	A	B	C
0		4	0.44
1		1	-1.33
2		6	1.61
3		1	-1.33
4		1	-1.33
5		6	1.61
6		5	1.02
7		3	-0.15
8		4	0.44
9		6	1.61
0		6	1.61
1		3	-0.15
2		4	0.44
3		1	-1.33
4		4	0.44
5		2	-0.74
6		4	0.44
7			
8	**Mean**	3.82	
9	**SD**	1.66	

Histogram

Does Not look Normally Distributed. Looks Left Skew at this Point. However, we know this is a Uniform Distribution. We can stop here!



Outliers No Outliers!

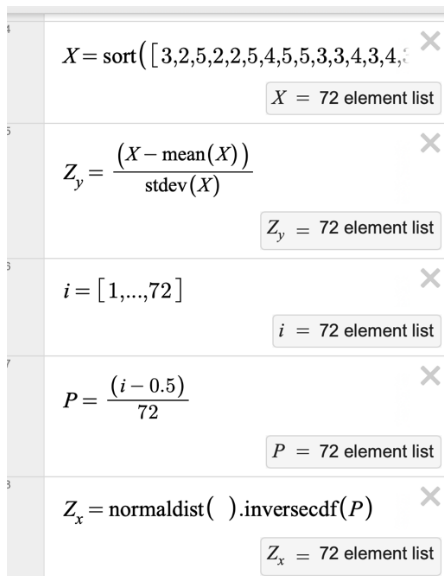


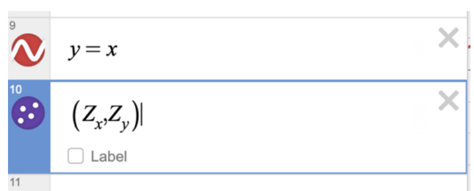
Distribution looks left Skew, but we know it is Uniform.

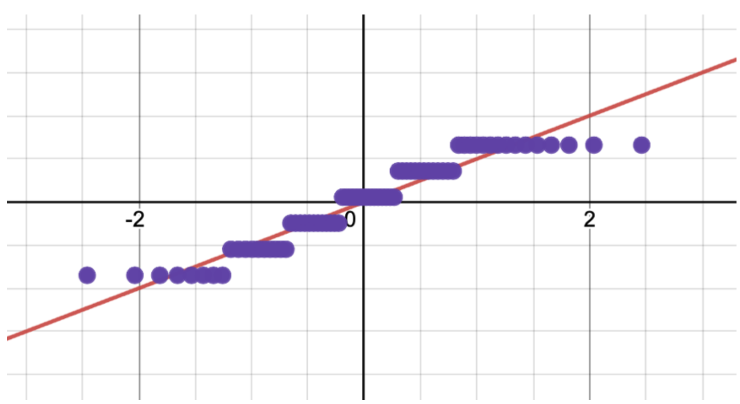
Quantile Quartile Plot Desmos

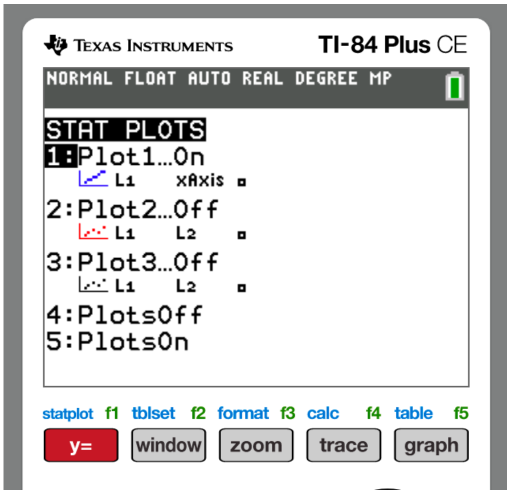
https://www.youtube.com/watch?v=Wa12Dvviuzs

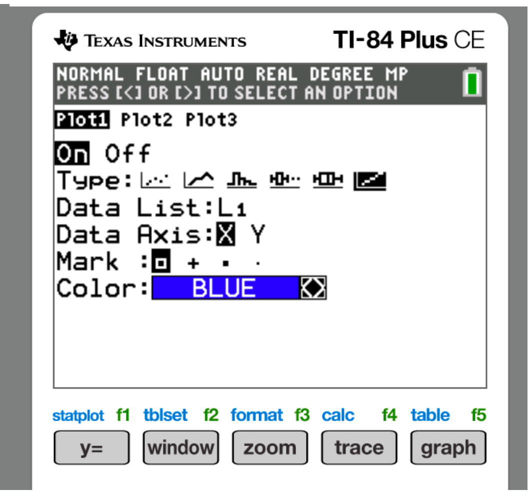
https://www.youtube.com/watch?v=okjYjClSjOg

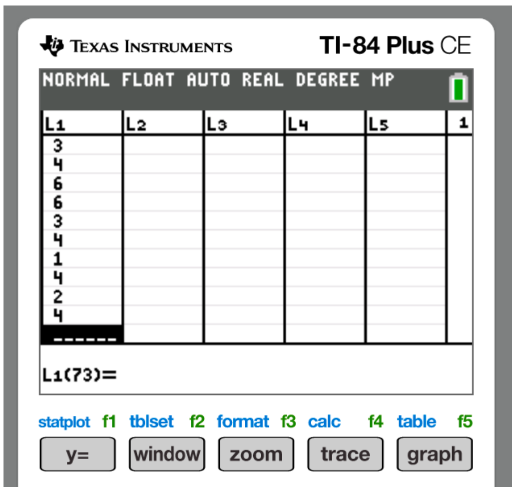


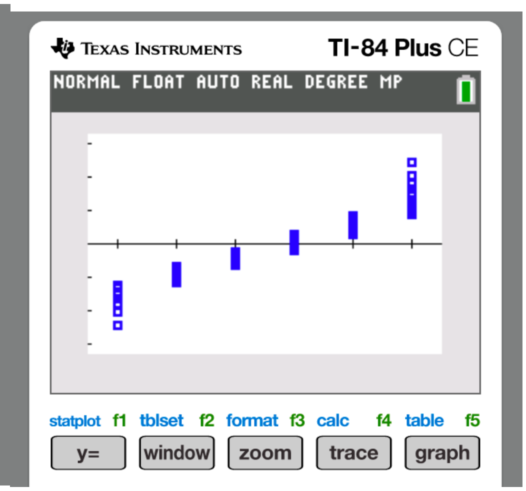












The values do not line up in a straight line and have some sort of symmetry properties.

Therefore, they are not Normally Distributed!