

### Ryan-Joiner Test for Normality

There are a few different tests to see if data is normally distributed, and some work better than others in different situations. We use the Ryan-Joiner test which gives fairly good results and can be run using the TI 83/84 calculator.

For this test, the null and alternative hypotheses are as follows.

$H_0$ : The data is normally distributed.

$H_1$ : The data is not normally distributed.

The test statistic for this test is the correlation coefficient between two lists of numbers.

The Ryan-Joiner test uses the critical value approach. Since there are different critical values for each confidence level, we will only use 95% confidence. Since the confidence is fixed at 95%, all that is needed to determine a critical value is the sample size. The critical values for this test can be found in the table on page 4 of these notes and on the website in the *Supplementary Statistics Topics* page under *Ryan-Joiner Lists and Critical Values*.

The Ryan-Joiner test decision criteria and conclusions are as follows.

If the correlation coefficient is less than the critical value,  $H_0$  is rejected with the conclusion:

With 95% confidence, the evidence is strong enough  
to say the data is not normally distributed.

If the correlation coefficient is not less than the critical value,  $H_0$  is kept with the conclusion:

With 95% confidence, the evidence is not strong enough  
to say the data is not normally distributed.

**Example Problem** Test whether the test scores given below (p.125, Sullivan, *Fundamentals of Statistics*, 5<sup>th</sup> ed., Pearson Education, Inc. 2018) are normally distributed at 95% confidence.

Test Scores	82	77	90	71	62	68	74	84	94	88
-------------	----	----	----	----	----	----	----	----	----	----

Since there are 10 data values, you first make a list of numbers from 1 to 10. These are called index values or indices.

Index( $i$ )	1	2	3	4	5	6	7	8	9	10
--------------	---	---	---	---	---	---	---	---	---	----

Next you apply the following function, called the Blom function after someone named Blom, to each of the index values,  $i$ , where  $n$  is the sample size.

$$f(i) = \frac{i - 0.375}{n + 0.25},$$

In this case,  $n = 10$ . This function sends the integers from one to the sample size to the interval between 0 and 1 in a linear fashion.

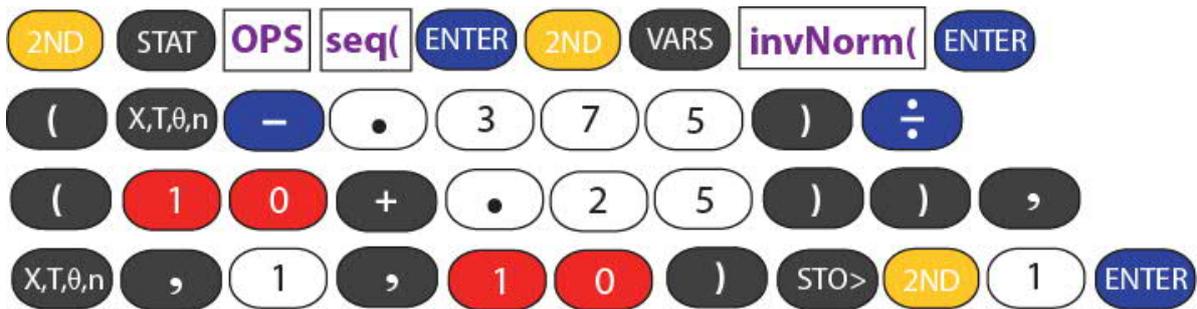
$i$	1	2	3	4	5	6	7	8	9	10
$f(i)$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939

Next, you must find the standard scores corresponding to each  $f(i)$ . To accomplish this, you must do the following for each value of  $f(i)$ : press “2ND” and “VARs”, choose “invNorm(“, press “ENTER”, input the value of  $f(i)$ , press “)” and press “ENTER” again.

$f(i)$	0.061	0.159	0.256	0.354	0.451	0.549	0.646	0.744	0.841	0.939
$\text{invNorm}(f(i))$	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.38	0.66	1.00	1.55

This list of inverse normal values of the  $f(i)$  will be compared to the ordered listing of the original data by using the correlation coefficient.

The diagram below provides a simpler way to generate this last list using the TI 83/84 calculator. Rectangles represent options you choose. All other symbols represent buttons you press.



Basically, the first row tells the calculator to create a sequence using the inverse normal function. The second row enters the numerator of the Blom function and the third row enters the denominator of the Blom function. The last row tells the calculator what variable is being used to create the sequence values, where this variable starts, where this variable stops, and finally the last row tells the calculator to store the results of the calculations in the list L1.

If you have a different sample size, you would put it where the red buttons are (“1” and “0” in this case) in the diagram.

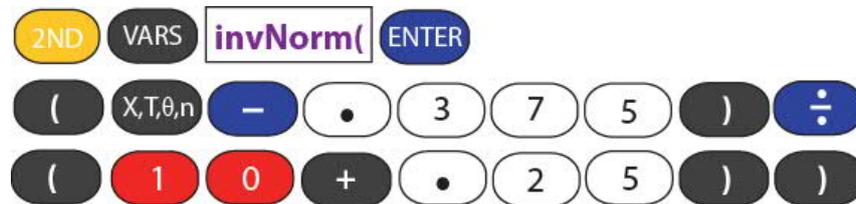
Before you press “ENTER” for the last time, your calculator screen should look like this:

```
seq(invNorm((X-.
375)/(10+.25)),X
,1,10)->L1
```

If the screen below appears after pressing the first “ENTER”, you will have to use the procedure that starts below this screen.



First put the cursor next to *Expr:* and follow the diagram below.



Then put the cursor next to *Variable:* and press the “X,T,θ,n” button. Next to *start:* enter 1. Next to *end:* enter 10. Next to *step:* enter 1. Your calculator screen should look like the one below.

If you have a different sample size, you would put it next to *Expr:* where the red buttons are (“1” and “0” in this case) and also next to *end:*



Now highlight *Paste* and press “ENTER”. You have now defined the sequence of inverse normal values and you just need to tell the calculator where to put the sequence. To do this, just press “STO>”, “2ND”, “1” and “ENTER”.

If you have either type of calculator and you press “STAT” and “ENTER” again, you will see the following numbers under L1.

invNorm( $f(i)$ )	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.38	0.66	1.00	1.55
-------------------	-------	-------	-------	-------	-------	------	------	------	------	------

The next step is to enter the original data set under L2 in the calculator.

Original Data	82	77	90	71	62	68	74	84	94	88
---------------	----	----	----	----	----	----	----	----	----	----

Now you must arrange this data in order from smallest to largest value. Press “STAT”, choose “SortA(“ and press “ENTER”. Press “2ND”, 2, “)” (right parenthesis), and “ENTER”. If you press “STAT” and “ENTER”, you will see the data under L2 is now in order.

Original Data	82	77	90	71	62	68	74	84	94	88
Ordered Data	62	68	71	74	77	82	84	88	90	94

To find the correlation coefficient between the inverse normal values and the ordered data, first make sure that the inverse normal values shown below are under L1 and ordered data values shown below are under L2.

invNorm( $f(i)$ )	-1.55	-1.00	-0.66	-0.38	-0.12	0.12	0.38	0.66	1.00	1.55
Ordered Data	69	73	73	78	79	81	90	92	95	97

Then press “STAT”, choose “TESTS” and “LinRegTTest”, and then press “ENTER”. Make sure L1 is next to “Xlist:” and L2 is next to “Ylist:”. Next to “ $\beta$  &  $\rho$ ” choose “ $\neq 0$ ”. Finally, highlight “Calculate” and press “ENTER”.

The correlation coefficient is the value of  $r$ . In this case, the coefficient is about 0.993.

According to the table below, for a sample of size 10, the critical value is 0.918. The correlation coefficient we computed, 0.993, is more than 0.918. As a result, we keep  $H_0$ .

CONCLUSION: With 95% confidence, the evidence is not strong enough to say that the data is not normally distributed.

Example Problem Test to see if the data from #19 p. 135 (Sullivan, *Fundamentals of Statistics*, 5<sup>th</sup> ed., Pearson Education, Inc. 2018) is normally distributed using 95% confidence.

### Some Ryan-Joiner Critical Values for 95% Confidence

Sample Size	8	9	10	11	12	13	14	15
Critical Value	0.906	0.912	0.918	0.923	0.928	0.932	0.935	0.939
Sample Size	16	17	18	19	20	21	22	23
Critical Value	0.941	0.944	0.946	0.949	0.951	0.952	0.954	0.956
Sample Size	24	25	26	27	28	29	30	31
Critical Value	0.957	0.959	0.960	0.961	0.962	0.963	0.964	0.965
Sample Size	32	33	34	35	36	37	38	39
Critical Value	0.966	0.967	0.968	0.969	0.969	0.970	0.971	0.971
Sample Size	40	41	42	43	44	45	46	47
Critical Value	0.972	0.973	0.973	0.974	0.974	0.974	0.975	0.976
Sample Size	48	49	50	51	52	53	54	55
Critical Value	0.976	0.976	0.977	0.977*	0.977*	0.978*	0.978*	0.979

Source: Stephen W. Looney & Thomas R. Gullledge, Jr. “Use of the Correlation Coefficient with Normal Probability Plots”, *The American Statistician*, Vol. 39, No. 1 (Feb. 1985), pp. 75-79.

\*These values were obtained using linear interpolation between the values given at 50 and 55.