

INTERPRETATION OF TEST DATA USING STATISTICS

*Nowak, A.S., Collins K.R. Reliability of structures.
McGraw-Hill Higher Education 2000*

PROBABILITY PAPER

Probability paper can be used to graphically determine whether a set of experimental data can be described by the normal distribution.

It is possible to construct probability paper for other distributions.

The basic idea behind normal probability paper is to redefine the vertical scale so that the normal CDF will plot as a straight line.

If a set of data plotted on normal probability paper plots as a straight line, then it is reasonable to model the data using a normal CDF.

The slope and y intercept of the graph can be used to determine the mean and standard deviation of the distribution.

Today, with the availability of spreadsheet programs and computers, it is very easy to achieve the same effect of commercial normal probability paper by performing a simple mathematical transformation and plotting a standard linear (xy) graph.

For any realization x of the normal random variable X , the corresponding standardized value is

$$z = \frac{X - \mu_X}{\sigma_X} = \left(\frac{1}{\sigma_X} \right) z + \left(\frac{-\mu_X}{\sigma_X} \right)$$

The corresponding probability based on the normal CDF

$$F_X(x) = p = \Phi \left(\frac{X - \mu_X}{\sigma_X} \right)$$

The inverse of the above equation

$$\Phi^{-1}(p) = z = \left(\frac{1}{\sigma_x} \right) z + \left(\frac{-\mu_x}{\sigma_x} \right)$$

The equation represents a linear relationship between $z = \Phi^{-1}(p)$ and x .

If $\Phi^{-1}(p)$ versus x is plotted on standard (linear) graph paper, a straight-line plot will result.

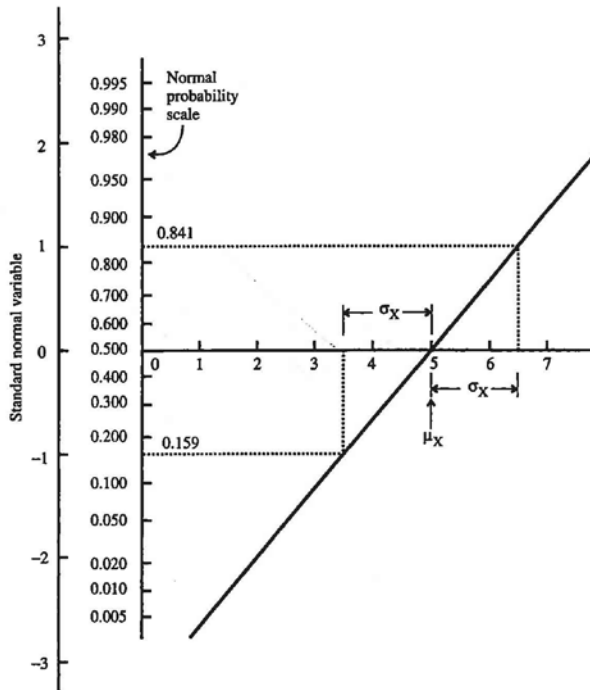


FIGURE Interpretation of a straight-line plot on normal probability paper in terms of the mean and standard deviation of the normal random variable.

The procedure is as follows:

1. Arrange the data values $\{x\}$ in *increasing* order.

The first (lowest) value of x will be denoted as x_1 the next value as x_2 , and so on, up to the last (largest) value x_N .

Do not discard repeated values.

2. Associate with each x_i a cumulative probability p_i equal to

$$p_i = \frac{1}{N + 1}$$

3. If commercial normal probability paper is being used, then plot the (x_i, z_i) and go to Step 6. Otherwise, go to Step 4.

4. For each p_i determine $z_i = \Phi^{-1}(p_i)$.

5. Plot the coordinates (x_i, z_i) on standard linear graph paper by hand or using a computer.

6. If the plot appears to follow a straight line, then it is reasonable to conclude that the data can be modeled using a normal distribution.

Sketch a "best-fit" line for the data.

The slope of the line will be equal to $1/\sigma_x$, and the value of x at which the probability is 0.5 (or $z = 0$) will be equal to μ_x .

Alternatively, you can plot a reference line using the sample mean \bar{x} and sample standard deviation s_x .

If the data do not appear to follow a straight line, then a normal distribution is probably not appropriate.

EXAMPLE.

Consider the following set of 9 data points:

$$\{x\} = \{6.5, 5.3, 5.5, 5.9, 6.5, 6.8, 7.2, 5.9, 6.4\}.$$

Plot the data on normal probability paper.

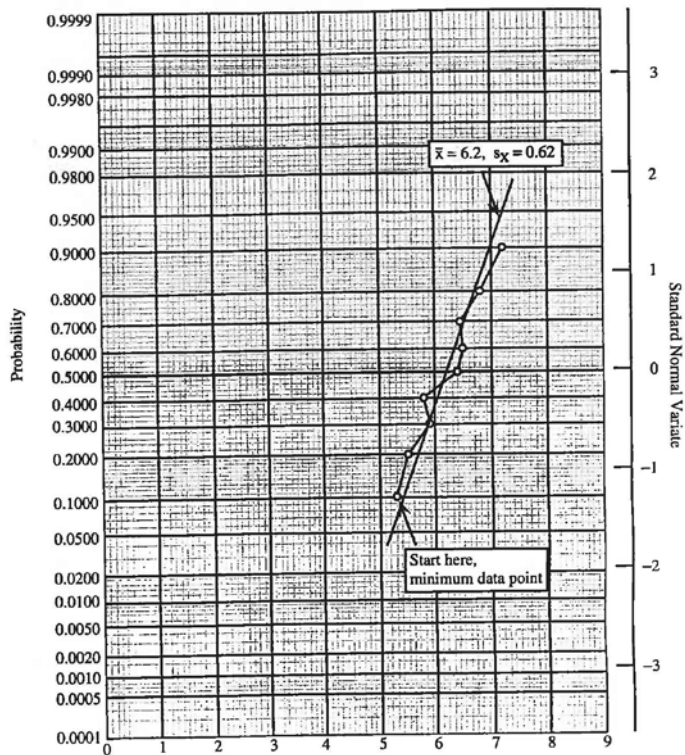
Solution.

It is convenient to carry out Steps 1 and 2 by setting up a table as seen in Table

TABLE 2.1 Data table for Example 2.7

Index value, i	x_i (in increasing order)	Probability, $p_i = i/(N + 1)$	$z_i = \Phi^{-1}(p_i)$
1	5.3	0.1	-1.282
2	5.5	0.2	-0.842
3	5.9	0.3	-0.524
4	5.9	0.4	-0.253
5	6.4	0.5	0
6	6.5	0.6	0.253
7	6.5	0.7	0.524
8	6.8	0.8	0.842
	--	0.9	1.282

The values of (x_i, p_i) are plotted on probability paper in Figure.



We would obtain the same graph if we plotted (x_i, z_i) and use the linear scale shown on the right side of Figure.

The data plotted in Figure appear to follow (at least approximately) a straight line and thus we might conclude that the data follow a normal distribution.

For comparison, a "reference" straight line is plotted based on the sample statistics $\bar{x} = 6.2$ and $s_x = 0.62$.

HISTOGRAM

The basic idea is to count the number of data points that fall into predefined intervals and then make a bar graph.

By looking at the bar graph, you can observe trends in the data and visually determine the "distribution" of the data.

EXAMPLE:

Suppose we test 100 concrete cylinders and experimentally determine the compressive strength for each specimen.

We then establish intervals of values and count the number of observed values that fall in each interval.

This is shown in Table.

Concrete strength, f'_c (150 psi interval)	Number of observations in interval	Frequency of occurrence	Cumulative frequency
Below 1500 psi	0	0.00	0.00
1500–1650	1	0.01	0.01
1650–1800	1	0.01	0.02
1800–1950	3	0.03	0.05
1950–2100	3	0.03	0.08
2100–2250	8	0.08	0.16
2250–2400	12	0.12	0.28
2400–2550	11	0.11	0.39
2550–2700	10	0.10	0.49
2700–2850	13	0.13	0.62
2850–3000	9	0.09	0.71
3000–3150	8	0.08	0.79
3150–3300	6	0.06	0.85
3300–3450	3	0.03	0.88
3450–3600	4	0.04	0.92
3600–3750	4	0.04	0.96
3750–3900	2	0.02	0.98
3900–4050	1	0.01	0.99
4050–4200	1	0.01	1.00
4200–4350	0	0.00	1.00
4350–4500	0	0.00	1.00
Above 4500 psi	0	0.00	1.00

Then, for each interval, we calculate the relative "frequency of occurrence," which is the total number of observations for the interval divided by the total number of all observations.

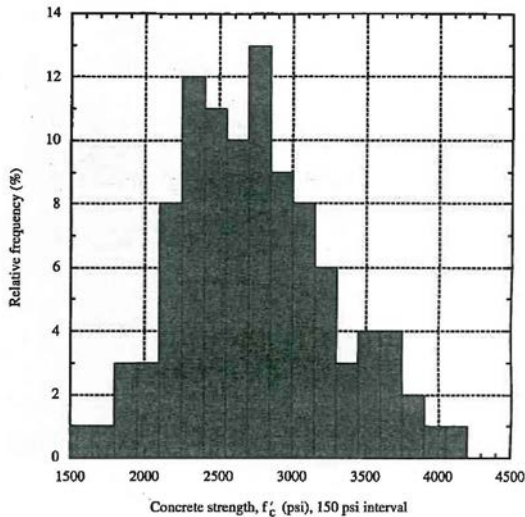
This corresponds to the percentage of all observations that fall in a particular interval.

This has been calculated in the third column of Table.

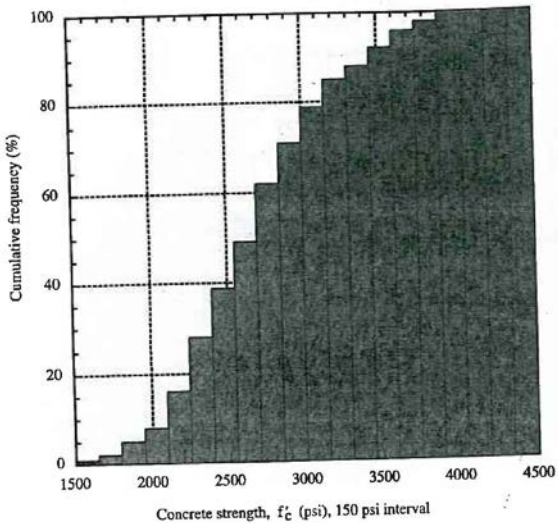
If, for each interval, we add up the frequency value for that interval and all intervals below it, we get a cumulative frequency value as shown in the last column of Table.

If we plot the values in column 3 of Table versus the interval values in column 1, we get a *relative frequency histogram* plot as seen in Figure .

If we plot the values in column 4 versus the interval values in column I, we get a *cumulative frequency histogram* as seen in Figure.



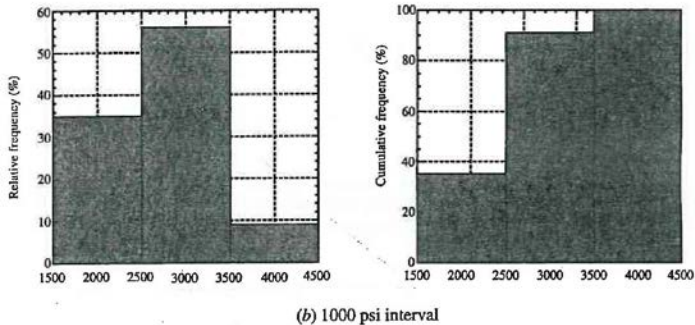
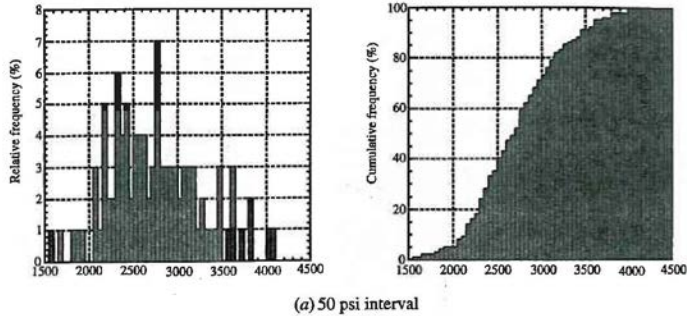
a)



b)

FIGURE Relative (a) and cumulative (b) frequency histogram for concrete strength.

Figure shows how the interval size can influence the overall appearance of frequency and cumulative frequency histograms.

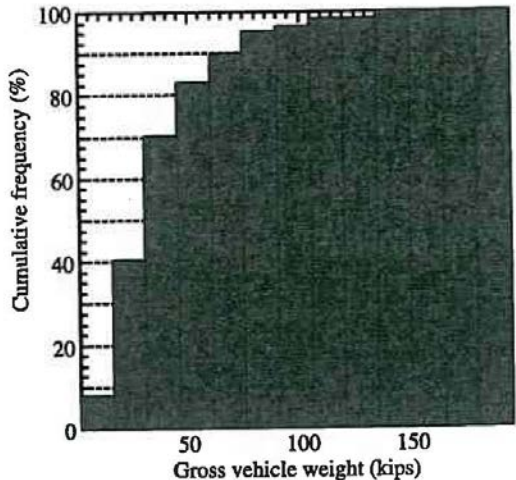
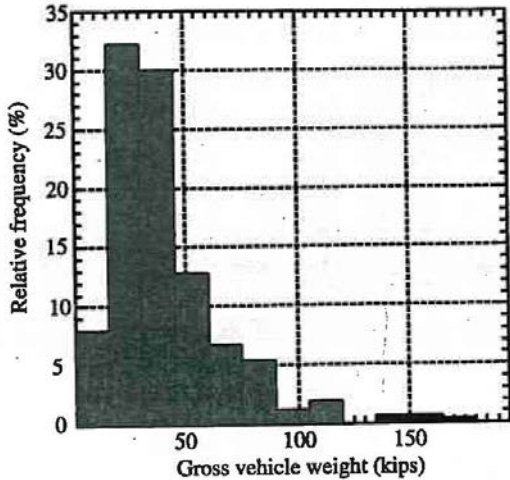


EXAMPLE.

Consider the set of values of vehicle weight recorded in Table.

Gross vehicle weight (15 kips interval)	Number of observations	Frequency of occurrence	Cumulative frequency
Below 15 kips	25	0.08	0.08
15.0–30.0	101	0.32	0.40
30.0–45.0	94	0.30	0.70
45.0–60.0	40	0.13	0.83
60.0–75.0	21	0.07	0.90
75.0–90.0	17	0.05	0.95
90.0–105.0	4	0.01	0.96
105.0–120.0	6	0.02	0.98
120.0–135.0	0	0.00	0.98
135.0–150.0	2	0.01	0.99
150.0–165.0	2	0.01	1.00
165.0–180.0	1	0.00	1.00
Above 180.0 kips	0	0.00	1.00

Figures can be plotted from calculations of relative and cumulative frequency values for the intervals



Relative and cumulative frequency histogram