EIN 5226 Lecture 12

Probability Plots

Chapter 8

Note: Need Calculator and Probability Plot Handout

Karen E. Schmahl Ph.D., P.E.

What is the distribution?

Before analyzing a set of sample data, we must make an assumption as to the type of distribution it follows.

After collecting data:

Generate graphical summaries
- histogram
- may also find boxplot or dot plot useful
- run chart if data collected over time

Generate Descriptive Statistics
- measures of location: mean, median,
- measures of dispersion: standard deviation, range

Review graphs
Outliers, Bi/Mulit- modal – investigate causes
Shape – Typically looking first to see if from a normal distribution or not

Checks for normality

On the histogram
- Is the data symmetric and bell shaped?
- Is there no more than one outlier?

With the descriptive statistics
- Are the mean and median fairly close together?
- Is the range about equal to 6 times the standard deviation?

If no to these questions – likely not normal data.

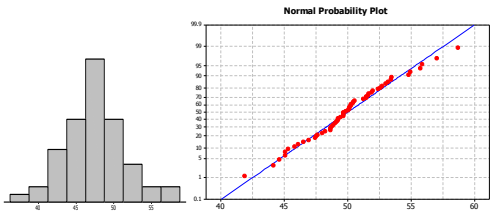
If yes to these questions - likely normal data

With limited amounts of data, or uncertainty in the answers - generate a normal probability plot

Probability Plot

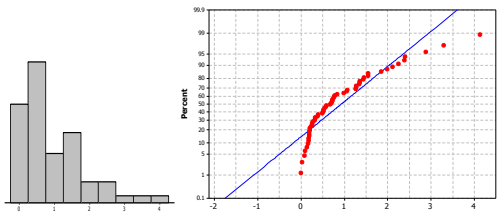
- Graphic used to determine if a sample of data might reasonably be assumed to come from a specific distribution
- Probability plots are available for various distributions
  - i.e., Normal, Lognormal, Exponential, Weibull)
- Subjective interpretation
  - A straight line indicates a fit to the distribution
  - The more curved the line, the less of a fit the data has
- Quantitative measures of goodness of fit available

Normal Probability Plot



A histogram for data that is close to normal will have a probability plot where the points fall reasonably close to a straight line.

Normal Probability Plot

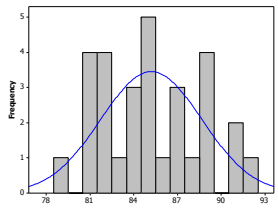


A histogram for data that is positively skewed will have a normal probability plot where the points fall into a curve.

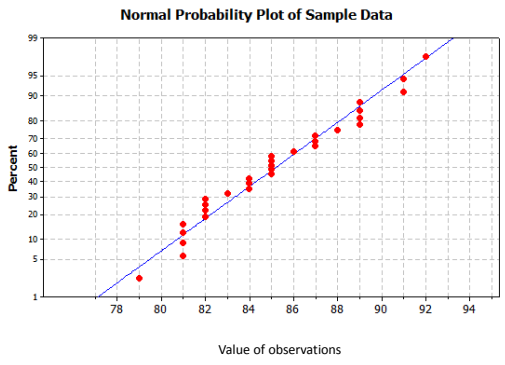
Sample Set of Data

The following set of sample data was collected.

82	87	85	84	85	87	79	92	81	86
81	89	91	85	82	82	89	81	85	89
88	89	85	81	91	87	84	84	82	83

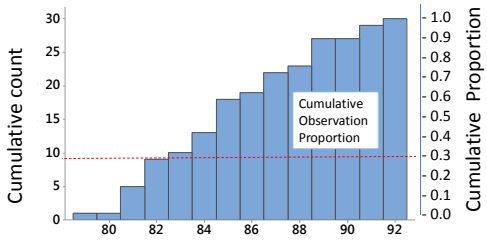


Mean = 85.2
Median = 85
Std Dev = 3.468
Range = 13

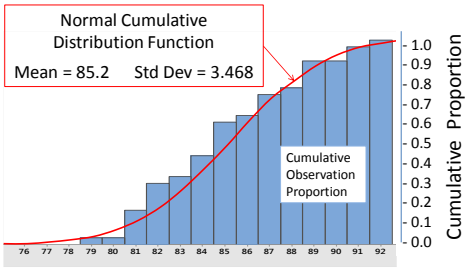


Frequency Table

		Frequency	Relative	Cumulative	
		Count	Frequency	Count	Rel. Freq
79	79	1	0.033	1	0.033
80	80	0	0.000	1	0.033
81	81	4	0.133	5	0.167
82	82	4	0.133	9	0.300
83	83	1	0.033	10	0.333
84	84	3	0.100	13	0.433
85	85	5	0.167	18	0.600
86	86	1	0.033	19	0.633
87	87	3	0.100	22	0.733
88	88	1	0.033	23	0.767
89	89	4	0.133	27	0.900
90	90	0	0.000	27	0.900
91	91	2	0.067	29	0.967
92	92	1	0.033	30	1.000
	Sum	30			



Example: Cumulative count at 82 (observations ≤ 82) = 9
Cumulative percentage = 0.30

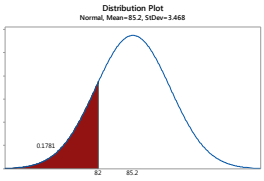


Sample Set of Data

The following set of sample data was collected.

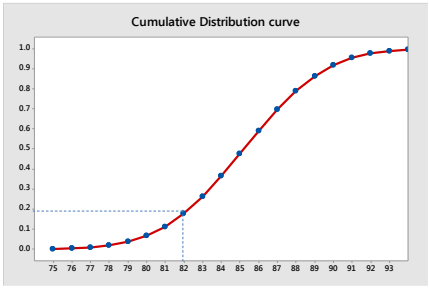
82	87	85	84	85	87	79	92	81	86
81	89	91	85	82	82	89	81	85	89
88	89	85	81	91	87	84	84	82	83

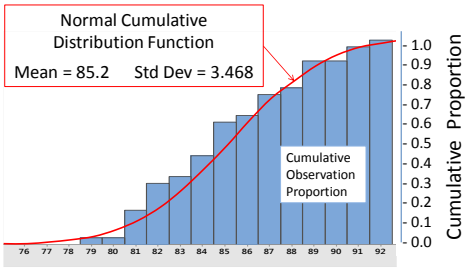
Mean = 85.2
Median = 85
Std Dev = 3.468
Range = 13

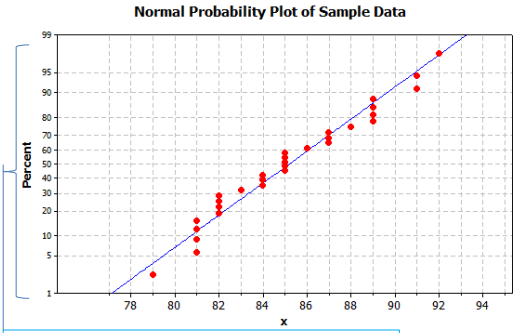


$$Z = \frac{82 - 85.2}{3.468} = -0.92$$

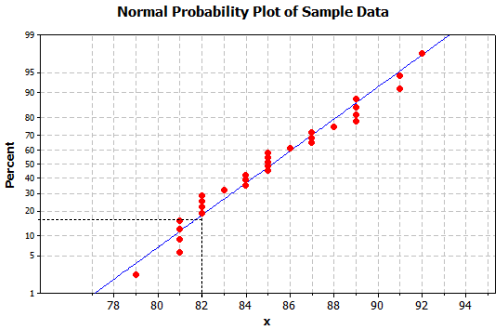
From Table P(Z<-0.92) = 0.1788







Y scale set up so that the Normal cumulative distribution function graphs as a straight line



You can read the expected proportion below a specified value directly off the graph.

Creating a Normal Probability Plot

(with normal probability plot paper)

1. Rank order the data from smallest to largest
2. Number the list from 1 to n. These are the i values.
3. For each i value, compute the value (i-0.5)/n
4. On normal probability graph paper, for each point plot

   (i-.5)/n as the percent on the y axis versus the associated value of the x
5. Evaluate the plot

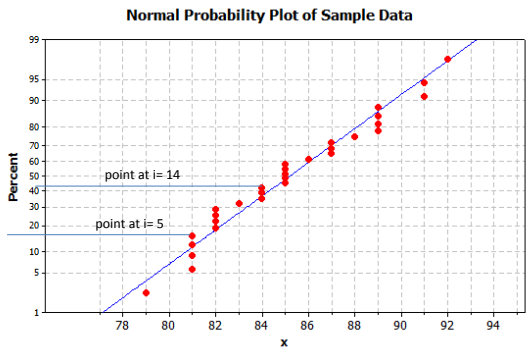
Example

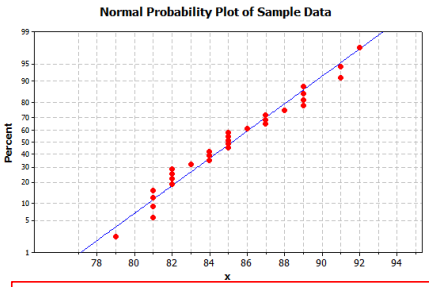
Sorted Observed Values	i	$\frac{i-0.5}{n}$
79	1	0.017
81	2	0.050
81	3	0.083
81	4	0.117
81	5	0.150
82	6	0.183
82	7	0.217
82	8	0.250
82	9	0.283
83	10	0.317
84	11	0.350
84	12	0.383
84	13	0.417
84	14	0.450
85	15	0.483

(Plus other 15 points)

$\longrightarrow \frac{i-.05}{n} = \frac{5-0.5}{30} = 0.150$

$\longrightarrow \frac{i-.05}{n} = \frac{14-0.5}{30} = 0.450$





If we assume the fitted line is appropriate for this distribution, for what value of x would 80% of the values be less than?

A) 87 B) 87.5 C) 88 D) 90

For the next several slides in this lecture you will need to use the Normal Plot worksheet found in Blackboard.

No voice

Probability Plot Worksheet

The following 10 data observations are collected from a process.

47.0 55.3 48.8 50.5 42.8 49.6 47.6 43.0 53.2 53.9

Analysis is to be performed to see if the assumption of normality is reasonable. Complete the data table on the worksheet, then answer the following questions:

What is the i=5 value from your initial data set (x)?

A) 42.8 B) 48.8 C) 49.6 D) 50.5

What is the (i-.5)/n value for data point 53.9?

A) 0.65 B) 0.75 C) 0.85 D) 0.95

Plot the points on your chart.

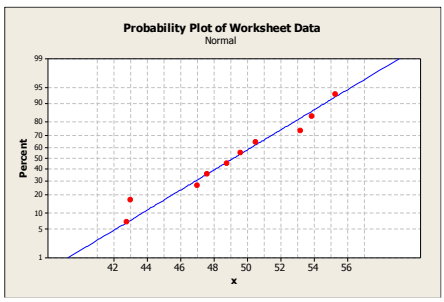
Is the assumption of normality a reasonable assumption?

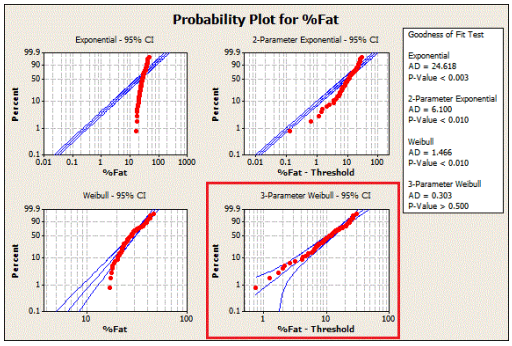
A) No, because the points are not equally spaced throughout the line.

B) No, because there is not enough data to determine if it is reasonable.

C) Yes, because the points fall in a reasonably straight line.

D) Yes, because the slope of a line drawn through the points would be negative.





http://blog.minitab.com/blog/adventures-in-statistics/how-to-identify-the-distribution-of-your-data-using-minitab

Related Assignments

Please see Blackboard for related assignments.