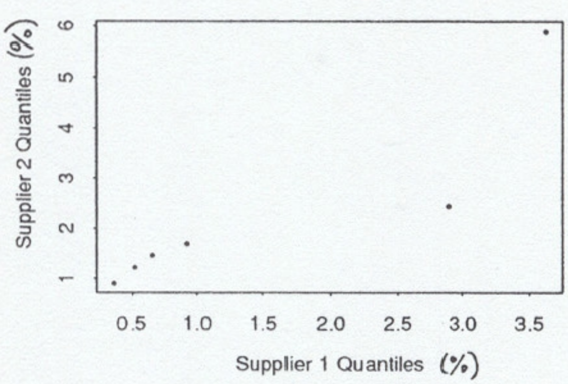
- **Problem 1:**
(a) Provide a QQ-plot for the two data sets below. **Note:** Be careful about the size of the two data sets. [10 pts]
Solution:

Use the $(i-.5)/n$ quantiles for the smaller data set.

i	$\frac{i-.5}{6}$	$Q_1(\frac{i-.5}{6})$	$Q_2(\frac{i-.5}{6})$
1	.08	.37	.91
2	.25	.52	1.22
3	.42	.65	1.47
4	.58	.92	1.70
5	.75	2.89	2.45
6	.92	3.62	5.89

The above quantiles for the Supplier 2 data were obtained by interpolation from the following table.

i	$\frac{i-.5}{8}$	$Q_2(\frac{i-.5}{8})$
1	.06	.89
2	.19	.99
3	.31	1.45
4	.44	1.47
5	.56	1.58
6	.69	2.27
7	.81	2.63
8	.94	6.54

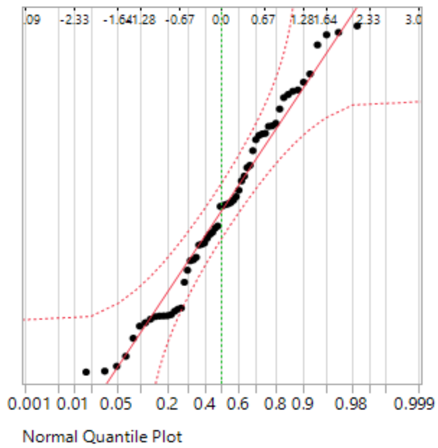


(b) What can you say about the shape of the two data sets? [5 pts]
Solution: It seems the data of the two supliers have **the same shape** (except one quantile at $(2.89, 2.45)$.

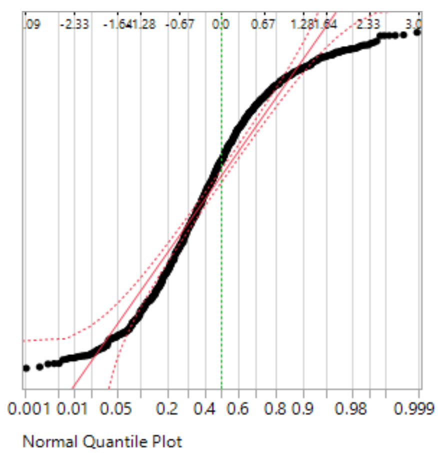
- **Problem 2:** (a) Explain the usefulness of theoretical QQ-plots [5 pts]
Sulotion: Theoretical QQ-plotting allows you to roughly check to see if **a data set has s shape which is similar to some theoretical distribution.** This can be useful in **identifying a theoretical (probability) model to represent how the process is generating**

data. Such a model can be then used to make inferences (conclusions) about the process.

(b) There is an excel file named "ballbearings.csv" as the relevant material of homework 3 on the course page (here). Use JMP to draw a Normal probability plot for **Group1** and **Group2** in the excel file separately.(You may use the tutorial, and just copy and paste the two QQ-plots)[5 pts for each plot]
Solution: For Group I, the Normal QQ-plot is



and for Group II:

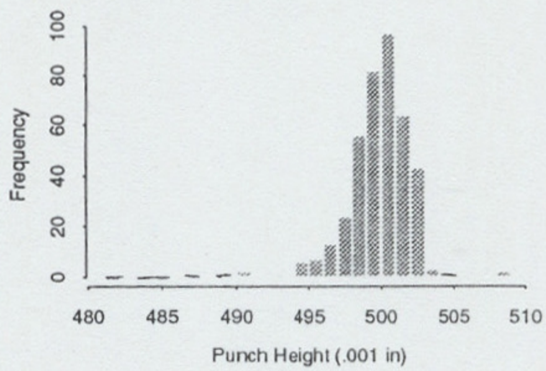


(c) Comment on the two QQ-plots you draw in part (b) of how similar the shapes of the data are to the theoretical quantile of Normal distribution and explain why.[10 pts]
The answer to the question is that the two data set **DO NOT HAVE SIMILAR SHAPE WITH THE NORMAL DISTRIBUTION.** The reason is that the QQ-plot of the two Groups of the data do not lie around the straight line in the QQ-plot of Normal distribution.(you get the full credit for answers close to this); however, you might come with the discussion that the two data sets are somehow similar to Normal distribution in shape except the extreme quantiles, especially for Group I where the data lie around the straight line better than Group II.

- **Problem 3:** End Chapter Exercise, Problem 9 (page 116) [part (a) 10 pts- part (b) 5 pts].
Note: Summarizing the data means you should find the sample mean, sample standard deviation, median, IQR and range of the data. Then draw appropriate plots to discuss the

distributional shape of the data
Solution:

(a) $\bar{x} = 500.24$, $s = 2.60$, $Median = 501$, $IQR = 3.0$, $R = 27.0$. The following histogram is easily constructed using the data as given.



The histogram shows that the distribution is truncated on the right. It appears that the supplier has inspected the punches and removed almost all which had punch heights greater than .505 inches or less than .495 inches. It seems that the supplier's equipment is not capable of meeting the specifications, because there would be a fair number of punches outside of specifications if no inspection were done.

(b) If a cut piece of material has hole diameters with a large amount of variability, it may be difficult to further process the piece. The punched holes may be filled with another part that can be made to have a uniform diameter. It may be easy to change the mean diameter of the filling part to accommodate the mean punch height, but if there is too much variability in punch heights, it will be impossible to make the filling part so that it fits consistently.

- **Problem 4:** There is a study with 5 factors, each with 3 levels, how many observations do we need to have a *full factorial study*? [5 pts]
  Solution: $3^5 = 243$

  Note: If the factors in a design all have the same number of levels, all possible combinations for a full factorial study has
  $$(\#of\ levels)^{(\#of\ factors)}$$

- **Problem 5:** Frequently, several measurements of a quantity are made under similar conditions using a single measuring device and are then averaged to produce a final value. Which of the aspects of measurement can be improved by this measurement and averaging process?. [5 pts]
  Solution: Accuracy (or unbiasness) is how close a measurement is to the true value "on average"

- **Problem 6:** Calculate the variance for the following samples (*note: if you are neat with your work, you may notice a pattern*):[3 pts each]

1. Sample 1: -1.05, -1.0, -0.5, 0.15, 0.6, 0.65, 0.7, 1.25

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{1}{8}(-1.05 + (-1.0) + (-0.5) + 0.15 + 0.6 + 0.65 + 0.7 + 1.25) \\
&= 0.1 \\
s^2 &= \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 \\
&= \frac{1}{n-1}((x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + (x_4-\bar{x})^2 + (x_5-\bar{x})^2 + (x_6-\bar{x})^2 + (x_7-\bar{x})^2 \\
&+ (x_8-\bar{x})^2) \\
&= \frac{1}{8}((-1.05-0.1)^2 + (-1.0-0.1)^2 + (-0.5-0.1)^2 + (0.15-0.1)^2 + (0.6-0.1)^2 \\
&+ (0.65-0.1)^2 + (0.7-0.1)^2 + (1.25-0.1)^2) \\
&= 0.73 \\
s &= \sqrt{s^2} = 0.85
\end{aligned}$$

2. Sample 2: -2.1, -2.0, -1.0, 0.3, 1.2, 1.3, 1.4, 2.5

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
&= \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{1}{8}(-2.1 + (-2.0) + (-1.0) + 0.3 + 1.2 + 1.3 + 1.4 + 2.5) \\
&= 0.2 \\
s^2 &= \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 \\
&= \frac{1}{n-1}((x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + (x_4-\bar{x})^2 + (x_5-\bar{x})^2 + (x_6-\bar{x})^2 + (x_7-\bar{x})^2 \\
&+ (x_8-\bar{x})^2) \\
&= \frac{1}{8}((-2.1-0.2)^2 + (-2.0-0.2)^2 + (-1.0-0.2)^2 + (0.3-0.2)^2 + (1.2-0.2)^2 \\
&+ (1.3-0.2)^2 + (1.4-0.2)^2 + (2.5-0.2)^2) \\
&= 2.93 \\
s &= \sqrt{s^2} = 1.71
\end{aligned}$$

3. Sample 3: -4.2, -4.0, -2.0, 0.6, 2.4, 2.6, 2.8, 5.0

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= 0.4 \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 \\ &= 11.72 \\ s &= \sqrt{s^2} = 3.42\end{aligned}$$

4. Sample 4: -8.4, -8.0, -4.0, 1.2, 4.8, 5.2, 5.6, 10.0

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= 0.8 \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 \\ &= 46.90 \\ s &= \sqrt{s^2} = 6.84\end{aligned}$$

5. Sample 5: -16.8, -16.0, -8.0, 2.4, 9.6, 10.4, 11.2, 20.0

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= 1.6 \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 \\ &= 187.64 \\ s &= \sqrt{s^2} = 6.8413.69\end{aligned}$$

- **Problem 7:** Mechanical engineers were interested in studying the effects of 2 chemical compounds (low Ca, high Ca) and 3 uni-axial pressure (P1, P2, P3) on metal bars lifetime. A total of 12 specimen were assigned to the possible combinations with two metal bars in each treatment. The lifetime of the bars were recorded after each run of the experiment.

  1. How many possible combinations of *compound* $\times$*pressure* are there available for a full factorial study? Draw a table for this design to get full credit.[5 pts]
     Solution: There are $2 \times 3 = 6$ possible combinations for a full factorial study. The following table shows a full factorial study with two specimen per treatment.

	P1	P2	P3
Low Ca	||	||	||
High Ca	||	||	||

2. What is the response variable in this study?[2 pts]
   Solution: The response variable is the metal bars lifetime
3. What are experimental variables in this study? [2 pts]
   Solution: The chemical compound and uni-axial pressure are the experimental variables.
4. What type of experimental variables are they in part 3 above?(Be careful with this question, we already know they are experimental variables and not response variable)[2 pts]
   Solution: The chemical compound and uni-axial pressure are both **qualitative (categorical)** variables.
5. For this full factorial study with two factors chemical compounds and uni-axial pressure, the six experimental runs are labeled as:
   No. 1: low- P1, No. 2: low-P2, No. 3: low-P3,
   No. 4: high- P1, No. 5: high- P2, and No. 6: high- P3.
   Based on the following random digits

   49502 18963 63920 39544 25804

   Which experiment should be done last?[4 pts]
   Solution: This exercise directly asks to use a random number table to randomly run the experiments. Using the digits, experimental run **NO. 3** will be the last to do.

Total: 95 pts