### BOX PLOT AND Q-Q PLOT FOR GEOMETRIC AND HARMONIC OBSERVATIONS

Mian Arif Shams Adnan Indiana University Bloomington Bloomington, Indiana 47401 maadnan@iu.edu

#### Abstract

Since there are various types of data (Arithmetic, Geometric and Harmonic) in literature, various types of Box Plots and Q-Q plots are required to demonstrate their Normality.

# 1. Introduction

In traditional statistics, instead of assuming the dependent ordered observations we assume that all the observations are independent among themselves [Hogg et al, 2013] whereas we are presenting all the observations on an ordered real line. We discuss here the Box plot and Q-Q plot of three types of symmetric distributions [Sharna, Adnan and Imon, 2015] including (i) Arithmetically Symmetric, (ii) Geometrically Symmetric and (iii) Harmonically Symmetric.

#### **2. Box Plot for Quartiles**

For Arithmetic Symmetric Distribution the Arithmetic Box Plot will be of the following type



to check, 
$$Q_3 - Q_2 = Q_2 - Q_1$$
.

In case of, Geometrically Symmetric Distribution the Geometric Box Plot will be of the following type

<i>Q</i> <sub>3</sub>		
$Q_1$		
		$Q_2$

to verify, 
$$\frac{Q_3}{Q_2} = \frac{Q_2}{Q_1}$$
.

For Harmonically Symmetric Distribution the Harmonic Box Plot will be of the following type



Iff, 
$$\frac{1}{Q_2} - \frac{1}{Q_3} = \frac{1}{Q_1} - \frac{1}{Q_2}$$
.

#### 3. Outlier Detection for Quartiles using Box-plot

In case of arithmetically symmetric or normal distribution we term one observation will be an outlier if it falls out of the following fences:

Lower Fence = 
$$Q_1 - \frac{3}{2}(Q_3 - Q_1)$$
  
Upper Fence =  $Q_3 + \frac{3}{2}(Q_3 - Q_1)$ .

For geometrically symmetric or normal distribution we term one observation will be treated as outlier if it falls out of the following fences:

> Lower Fence =  $\frac{1}{2} \left( \frac{Q_3}{Q_1} \right)$ Upper Fence =  $2 \left( \frac{Q_3}{Q_1} \right)$

Due to harmonically symmetric or normal distribution one observation will be an outlier if it cannot maintain the following limits:

Lower Fence  $=\frac{1}{Q_3} - \frac{3}{2} \left(\frac{1}{Q_1} - \frac{1}{Q_3}\right)$ Upper Fence  $=\frac{1}{Q_1} + \frac{3}{2} \left(\frac{1}{Q_1} - \frac{1}{Q_3}\right)$ .

# 3.1 Outlier Detection for Octiles using box-plot

In case of arithmetically symmetric or normal distribution we term one observation will be an outlier if it falls out of the following fences:

Lower Fence =  $O_1 - \frac{5}{6}(O_7 - O_1)$ Upper Fence =  $O_7 + \frac{5}{6}(O_7 - O_1)$ . For geometrically symmetric or normal distribution we term one observation will be treated as outlier if it falls out of the following fences: Lower Fence =  $\frac{1}{1.65}(\frac{O_7}{O_1})$ 

Upper Fence =  $1.65(\frac{O_7}{O_1})$ 

Due to harmonically symmetric or normal distribution one observation will be an outlier if it cannot maintain the following limits:

Lower Fence  $=\frac{1}{O_7} - \frac{5}{6}(\frac{1}{O_1} - \frac{1}{O_7})$ Upper Fence  $=\frac{1}{O_1} + \frac{5}{6}(\frac{1}{O_1} - \frac{1}{O_7})$ .

#### 4. Goodness of Fit Test for Quartiles

Now for the concerned sequences or progression-ed typed data the following statistics are true:

Arithmetic Fit (1):  $\frac{\left(Q_2 - \frac{Q_1 + Q_3}{2}\right)^2}{\frac{Q_1 + Q_3}{2}} \sim \chi_1^2,$ 

Geometric Fit (2): 
$$\frac{(Q_2 - \sqrt{Q_1 Q_3})^2}{\sqrt{Q_1 Q_3}} \sim \chi_1^2$$

Harmonic Fit (3): 
$$\frac{\left(Q_2 - \frac{2}{\frac{1}{Q_1} + \frac{1}{Q_3}}\right)^2}{\frac{2}{\frac{1}{Q_1} + \frac{1}{Q_3}}} \sim \chi_1^2.$$

(Qi) If  $\chi^2 < \chi^2_1$ , we will accept  $H_0$ . Otherwise we will fail to accept  $H_0$ .

(Qii) If any two of the aforesaid hypotheses are accepted, then

$$\frac{\chi^2}{\chi^2_1} \sim F_{1,1}$$

If  $F < F_{1,1}$ , we will accept the hypothesis concerned to the

denominator of the F test statistic, otherwise accept that of the numerator of the F statistic.

## 5. Q-Q Plot Three Types of Normal Data

Q-Q Plot for Arithmetic Normal Data



# Q-Q Plot for Geometric Normal Data

Ratio of Successive Sample Quantiles

**Theoretical Quantiles** 

#### Q-Q Plot for Harmonic Normal Data

1

Sample Quantiles





#### 6. Comparison among the arithmetically, geometrically and harmonically symmetric or normal distributions

For arithmetic symmetric or normal distribution:

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

$$Z = \frac{x-\mu}{\sigma}.$$

In case of geometrically symmetric or normal distribution:

$$f(x) = \frac{e^{-\frac{1}{2} \left(\frac{\log x - E(\log x)}{(E(\log x - E(\log x))^2)^{0.5}}\right)^2}}{(E(\log x - E(\log x))^2)^{0.5} \sqrt{2\pi}}$$

$$Z = \frac{\log x - E(\log x)}{(E(\log x - E(\log x))^2)^{0.5}}.$$

For harmonically symmetric or normal distribution:

$$f(x) = \frac{e^{-\frac{1}{2} \left(\frac{\frac{1}{x} - E(\frac{1}{x})}{\left(E(\frac{1}{x} - E(\frac{1}{x}))^2\right)^{0.5}}\right)^2}}{\left(E(\frac{1}{x} - E(\frac{1}{x}))^2\right)^{0.5} \sqrt{2\pi}}$$

$$Z = \frac{\frac{1}{x} - E(\frac{1}{x})}{\left(E(\frac{1}{x} - E(\frac{1}{x}))^2\right)^{0.5}}.$$

For arithmetically symmetric or normal distribution, the first four moments are

$$\mu = E(x) = \sum xp(x)$$
  

$$\sigma^{2} = \sum (x - \mu)^{2} p(x)$$
  

$$\mu_{3} = \sum (x - \mu)^{3} p(x)$$
  

$$\mu_{4} = \sum (x - \mu)^{4} p(x)$$

For geometrically symmetric or normal distribution, the first two moments are

$$\mu = G(x) = (x_1 x_2 \dots x_N)^{\frac{1}{N}}$$

$$\sigma = \left(\frac{x_2}{x_1} \frac{x_3}{x_2} \dots \frac{x_N}{x_{N-1}}\right)^{\frac{1}{N-1}}$$
$$= \left(\frac{x_N}{x_1}\right)^{\frac{1}{N-1}}$$

$$\mu = G(x) = x_1^{P(x_1)} x_2^{P(x_2)} \dots x_N^{P(x_N)}$$

σ

$$= \left(\frac{x_2^{P(x_2)}}{x_1^{P(x_1)}} \frac{x_3^{P(x_3)}}{x_2^{P(x_2)}} \dots \frac{x_N^{P(x_N)}}{x_{N-1}^{P(x_{N-1})}}\right)^{\frac{1}{N-1}}$$
$$= \left(\frac{x_N^{P(x_N)}}{x_1^{P(x_1)}}\right)^{\frac{1}{N-1}}$$

For Harmonically symmetric or normal distribution, the first two moments are

 $\mu = H(x)$   $= \frac{N}{x_1^{-1} + x_2^{-1} + \dots + x_N^{-1}}$   $\sigma^2 = \frac{N}{\left(x_1^{-1} - \frac{N}{x_1^{-1} + x_2^{-1} + \dots + x_N^{-1}}\right)^2}$   $+ \left(x_2^{-1} - \frac{N}{x_1^{-1} + x_2^{-1} + \dots + x_N^{-1}}\right)^2$   $+ \dots + \left(x_N^{-1} - \frac{N}{x_1^{-1} + x_2^{-1} + \dots + x_N^{-1}}\right)^2$ 

## Conclusion

Since a statistical data analyst does not know entirely the population before carrying out a study, assuming only arithmetic normality assumptions is frequently misleading the statistician to obtain the true results. Statisticians should not only base on the three types of means for three types of progressions, but also should know what is core increment and exponent of sequence of the entire data along with its general term which is the driving force of deciding various features of the said distribution.

#### References

Hogg, Mckean and Craig.Introduction to MathematicalStatistics. 2013. Seventh Edition.Pearson publisher.

Sharna, Adnan, Imon (2015). How Normal is Normal: How Symmetric is Symmetric: How Local is the Location for a Symmetric Distribution. Proc. JSM 2015. American Statistical Association.