**Continuous Probability Distribution**

## **1. Uniform (rectangle) Distribution:** fixed number of outcomes and probability of each outcome is the same. **P(x) = 1 / n equal outcomes.**

## Hence if n moves toward infinity (in dataset such as temperature, distance, income, mass, etc.). The probability moves toward zero.

* So think of a bar chart where the Hight of an infinite number of bars is zero! We call this ***continuous data.*** The probability of any specific outcome is zero.
* Instead of the probability of any specific outcome (which is zero), we can only find the probability over ***Specific interval***.

E(X) = Mean = (a+b)/2; Variance = δ2 = (b-a)2 /12; Std = δ = (b-a)/√12;

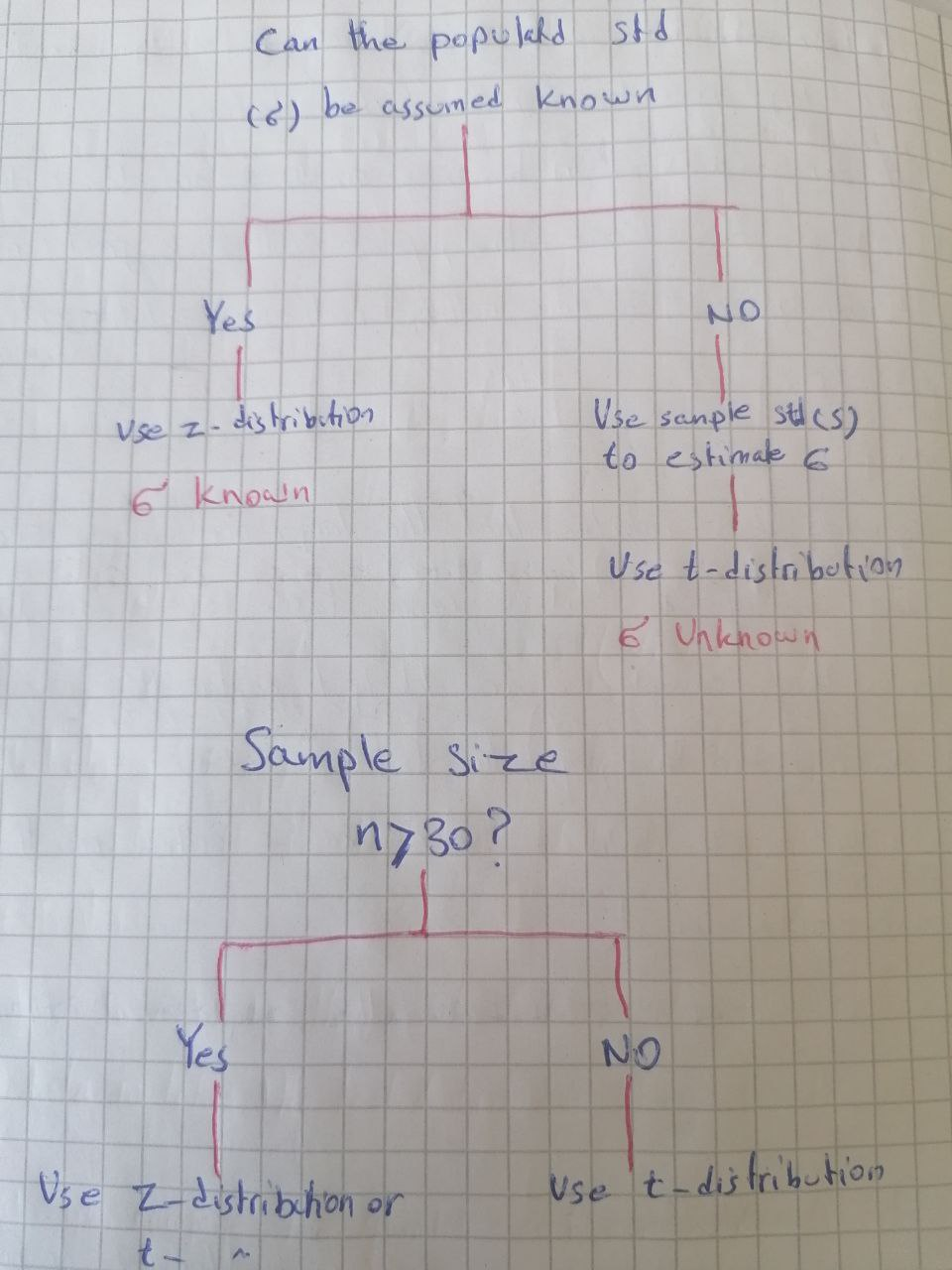
P(X) = (X2-X1)/(b-a)

**2. Normal (bell) Distribution:** the average (mean) tends to be very frequent while measures away from the mean are less frequent. µ (mean) can be any numerical value; slides distribution side-to-side. As much as δ is smaller we have narrower and taller tails and vice versa.

**2.1. The standard normal curve (Z-Distribution):** in this distribution µ = 0; δ= 1; and the area under the curve is 1. (when our sample size (n>30 or we know the δ of our population).

* In Excel **=NORM.DIST(x(z); µ(0); δ(1); TTUE/FALSE)**

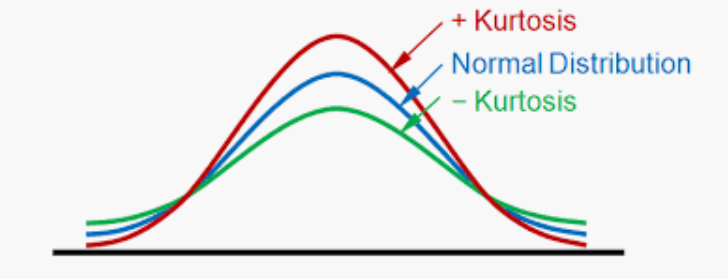
**2.2. The T-Distribution:** when our sample size is n=<30 and/or we don’t know our population’s variance/Std we use ***t-dist*** instead of ***z-dist.***

* The t-dist allows us to use small samples but in expenses of higher margin of error.
* It takes sample sizes into account using **df=n-1**; there is a different t-dist for any given **df**.
* The bell curve is squishier and fatter (tails) the smaller the n.
* However as n>30 and definitely n>100, the t-dist and z-dist become indistinguishable.
* 

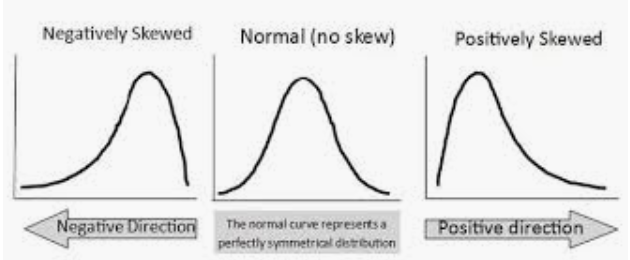
**V.V.Imp Note:** Before using normal distribution, first and foremost look at your data graphically first, before any analysis. Get to know the data. Look for patterns, potential problems, initial relationship, etc.

**2.3. Graphical Data Exploration:** by using a few simple visual tools, we can learn a tremendous amount of information about our data. Our data may have excess skew (lopsided), kurtosis (very fat tails), be bi-modal (two humps like camel), or follow a distribution other than normal distribution. Yet many statistical techniques **ASSUME** the data fits a normal distribution.

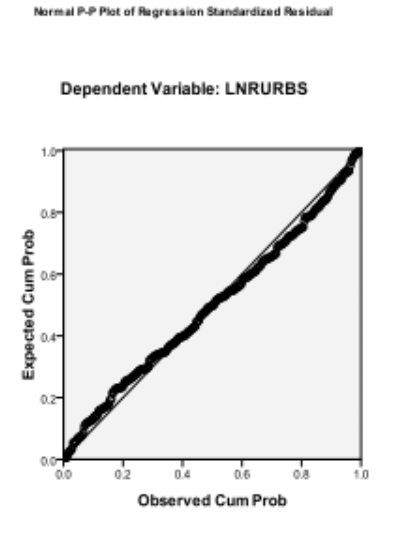
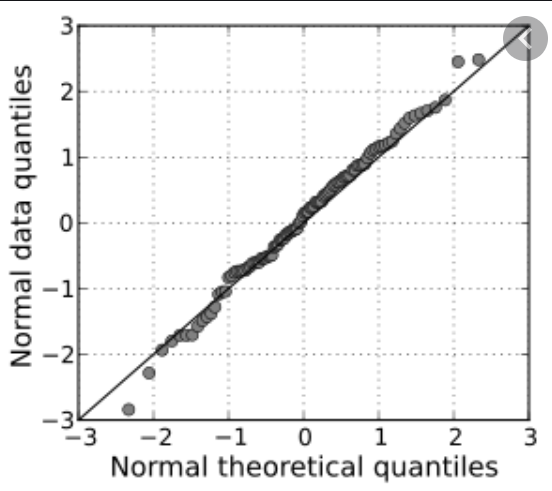
* **Graphical Tools for determining “normal” data:** 1) Histograms; 2) Steam and Leaf Plots; 3) Box Plots; 4) P-P Plots; 5) Q-Q Plots.
* **Excess Kurtosis:** More probability than expected in the tails of distribution due to the extreme values away from the mean. Probability (values) are pushed away from the mean and out towards the tails.



* **Excess Skewness:** More probability than expected is on one side of the distribution versus. Others; lopsided.



* **P-P Plots:** in a P-P plot, we compare the cumulative probability of our empirical data with an “**ideal test**” distribution (here: normal dist). Question to ask: **Do the points fall in straight line?** If our data matches the test distribution they should.
* **Q-Q Plots:** we compare the quantiles of our empirical data with the ideal**.** Question to ask: **Do the points fall in straight line?** If our data matches the test distribution they should.



Q-Q Plot