

Probability Overview

Dr. Farshid Alizadeh-Shabdiz
Spring 2021

What do we cover

- Quantitative vs quantitative data

What do we cover

- Pdf, pmf, cdf
- Mean - median – sd – variance
 - Population vs sample
- Outlier detection
 - Supervised
 - unsupervised
- Histogram
- Boxplot

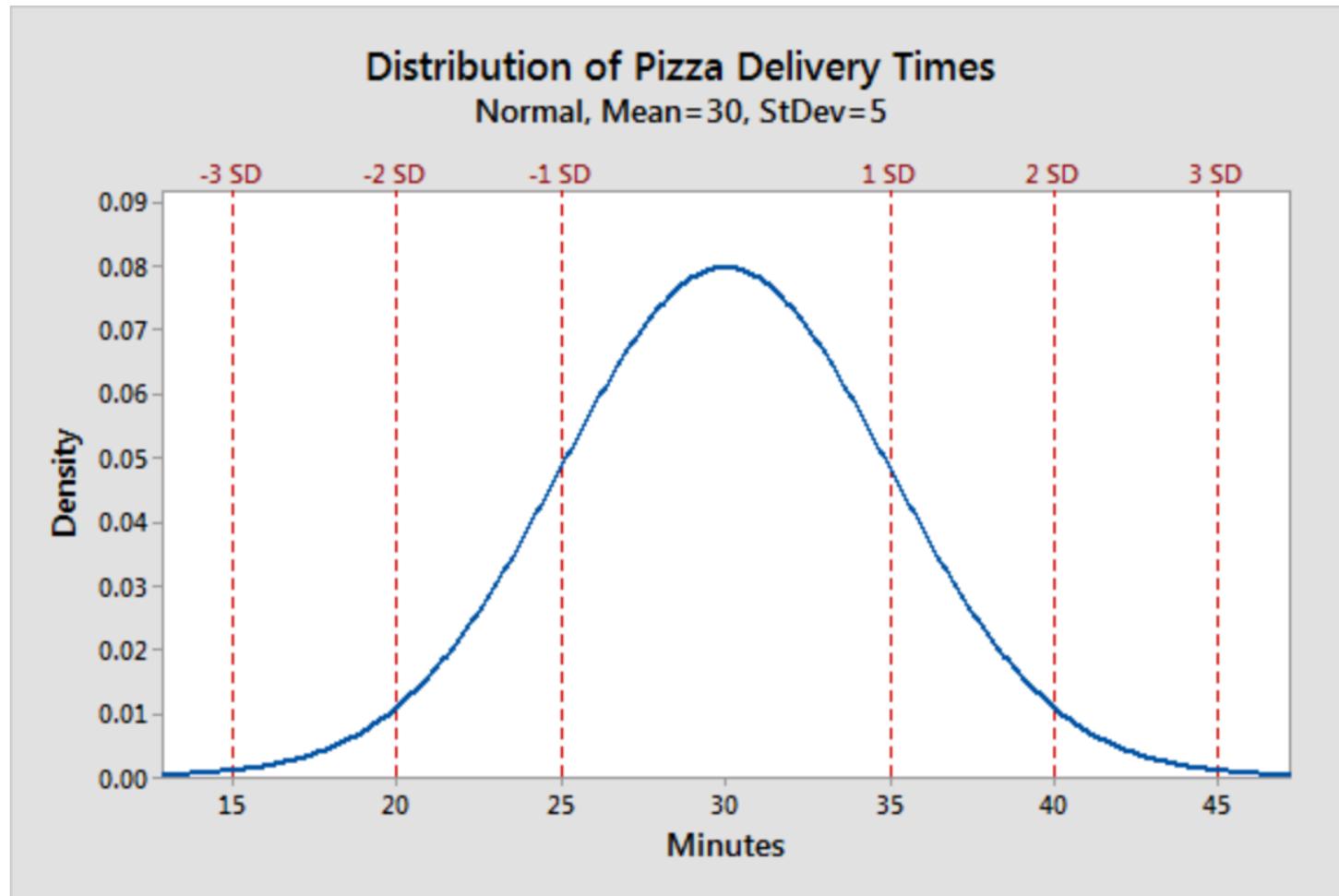
Important RV – Gaussian Distribution

- It is also called Normal Distribution
- Pdf:

$$P_x[x] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

- We write it as $N(\mu, \sigma^2)$. Normal distribution is defined with its mean and standard deviation.

Important RV – Standardization of Normal Distribution



Ref: <https://statisticsbyjim.com/basics/normal-distribution/>

$$P_x[x] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Normal Distribution

- It is symmetric
- Fully characterized by its mean and standard deviation

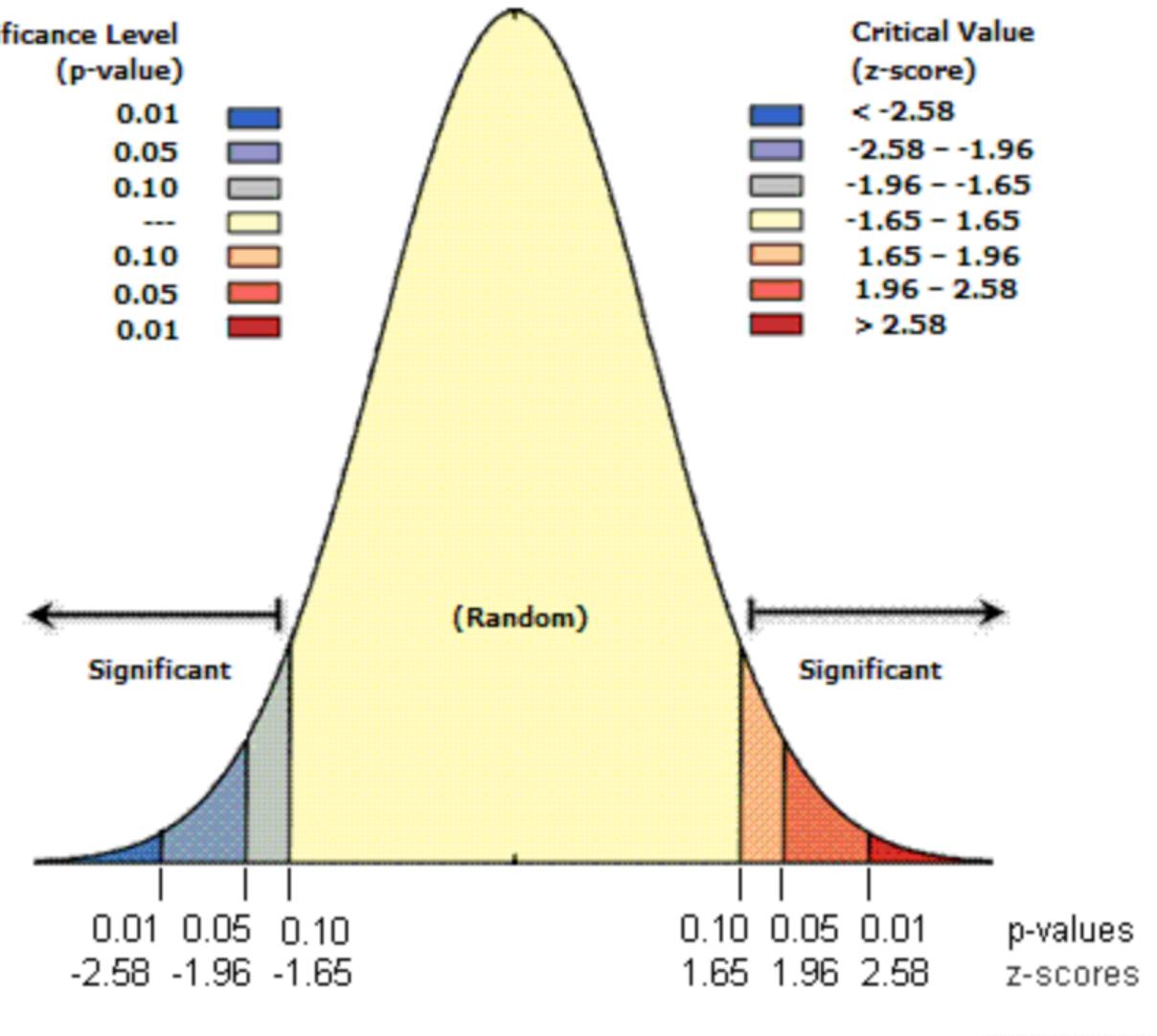
Important RV – Standardization of Normal Distribution

- If $N(\mu, \sigma^2)$ – z-score is calculated as follows

$$z = \frac{(x - \mu)}{\sigma}$$

- Pdf of normal distribution will become:

$$P_x[x] = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$



Reference: Picture from ArcGIS Pro

Log-Normal Distribution

- If U is a normal distribution, $X=e^U$ will have lognormal distribution

$$P_x[x] = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

Central Limit Theorem and Normal Distribution

- Mean of arbitrary random variables with arbitrary distribution approaches normal distribution as the number of random variables goes to infinity

