Continuous Probability Distributions

Discrete Random Variable

1. There are GAPs between numbers
2. can only assume clearly separated values
3. usually from counting successes
4. Examples:
   - Number of bedrooms in house.
   - number of times late
   - number of times interrupted
   - number of Heads in 3 Flips
5. Discrete Probability Distribution

Continuous Random Variable

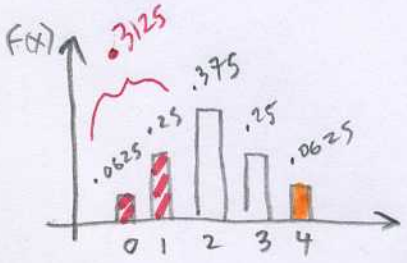
1. No GAPs between numbers
2. can assume any value in interval or collection of intervals
3. depends on measuring instrument
4. Examples:
   - weight of cereal box
   - score on statistics Excel Test
   - Salary Earned (Dollars)
   - Miles Auto Tires last
5. Continuous Probability Distribution

Both:

A listing of all the Random Variable outcomes of an Experiment & the probability associated with each outcome

6. Types:
   1. Binomial Distribution
   2. Poisson Distribution
   3. Hypergeometric
7. Questions we are allowed to ask:
   1. Probability of exactly X
   2. Probability between 2 x values

$P = .5$
$n = 4$



1. $P(x=4) = .0625$
2. $P(x \leq 1) = .25 + .0625 = .3125$

8. How we calculate:
   - Add individual Probabilities

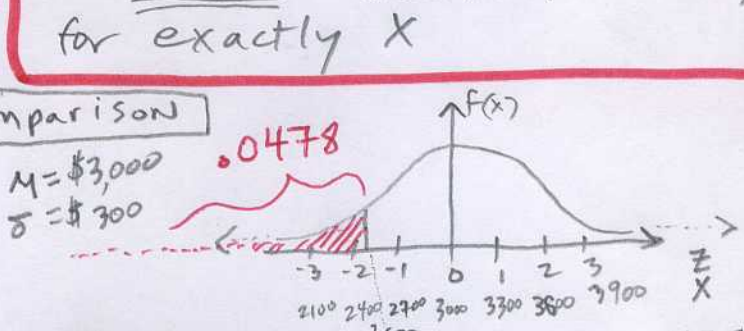
Because line has No area = actually $P(x=3) = 0$

6. Types we will study in this chapter:
   1. uniform Probability Distribution
   2. Normal (Bell) Prob. Distribution
   3. Exponential Prob. Distribution
7. Questions we are allowed to ask:
   1. Probability between 2 x values

we cannot calculate Probability for exactly X

Visual comparison

$M = \$3,000$
$\sigma = \$300$



1. $P(x \leq 2500) = P(x < 2500) = .0478$ *

8. How we calculate:
   - Find area under curve between 2 values
   - Area = Probability

* Because a line has No area, we cannot calculate $P(x=3)$ or $P(x=.12)$

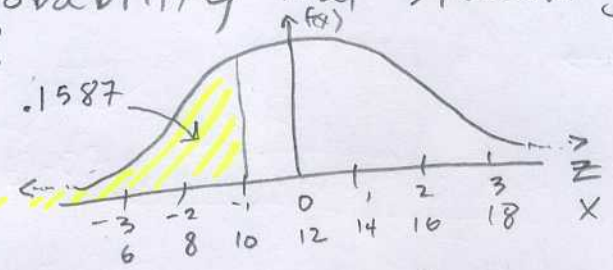
Continuous Probability Distributions:

We ask questions like:

Q: what is probability that student gets less than 10 on quiz?

$\mu = 12$
$\sigma = 2$

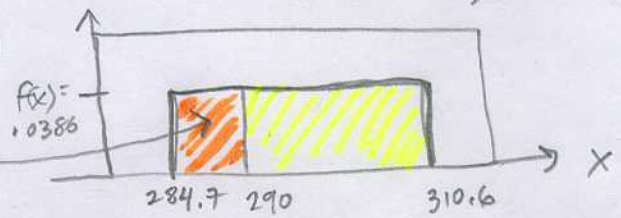
All area = 1
$P(X \leq 10) = .1587$



A: The probability will come from the area under the curve from $-\infty$ to 10 (x) or $-\infty$ to -1 (Z)

Q: what is probability that a top ten PGA golfer will drive between 284.7 & 290 yards?

All area = 1 (colored area)
$P(284.7 \leq X \leq 290) = .2046$



A: The probability will come from the area from 284.7 to 290

---

* Area = Probability
* Lines Don't have area, thus:
  1. Probability of exact value = 0 $P(X = 3) = 0$
  2. $P(X \leq 3) = P(X < 3)$

Discrete functions

Binomial $F(x) = \frac{n!}{x!(n-x)!} * p^n * (1-p)^{(n-x)}$

① can calculate probability

Continuous Functions

Normal $f(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-(x-\mu)^2/2\sigma^2}$

① cannot calculate Probability Directly

② continuous Probability functions, f(x) (called Density functions) can calculate the height of curve & can help to plot the curve & area, but we must use Integral Calculus to find Area. Luckily →

→ Instead of Integral calculus we can use Excel functions that will calculate the area under the curve.

Note about chapter 6

In chapter 6 we are dealing with population data that make up Distribution.

μ = mean of population

σ = Standard Deviation of population

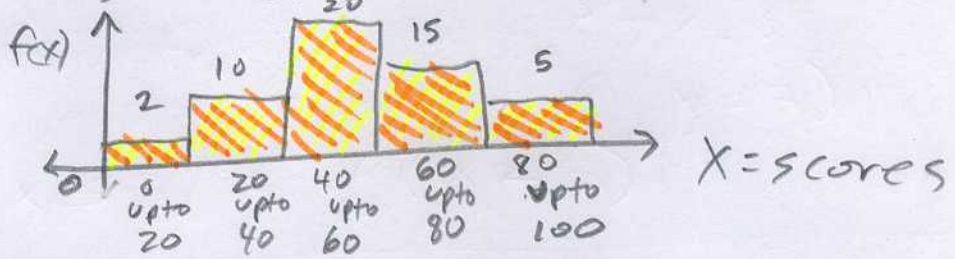
Later we will deal with sample Data (after we learn "Central Limit Theorem"):

$\bar{X}$ = mean of sample

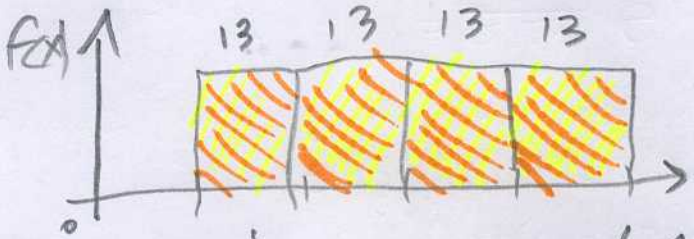
s = SD of sample

Visual Introduction to Uniform Probability Distr.

Imagine the distribution of statistics Quiz scores looked like this:



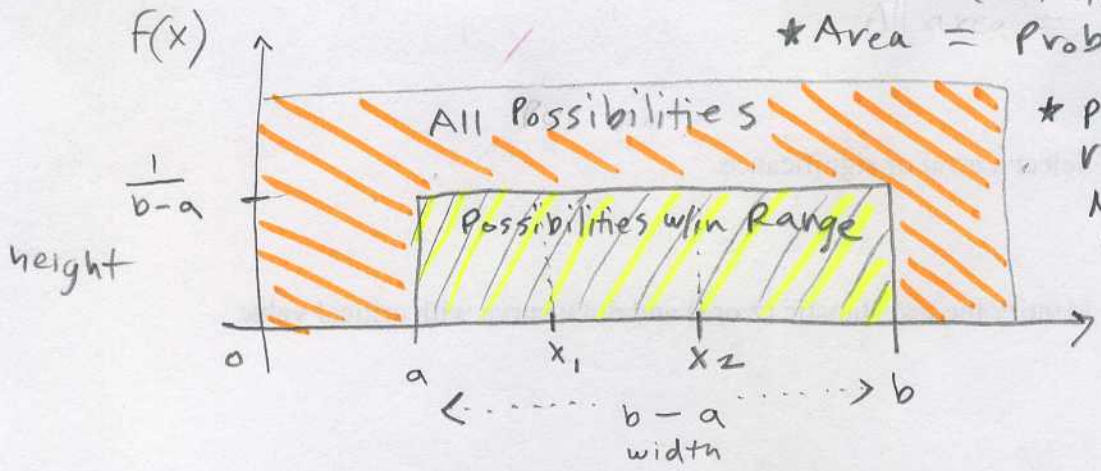
Now imagine if it looked like this:



Uniform Distribution!!

① Uniform Probability Distribution

* All area within Range = 1
* Area = Probability



* Probability over a range of values, Not for a particular value.

$$\text{Area} = 1 = \text{height} * \text{width} = \text{height} * (b-a)$$

$$1 = \text{height} * (b-a)$$

$$\frac{1}{(b-a)} = \text{height} \quad \text{(height is not a probability)}$$

② UNIFORM PROBABILITY DENSITY FUNCTION

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le X \le b \\ 0 & \text{elsewhere} \end{cases} \qquad \text{also } f(x) \ge 0 \text{ for all } X$$

a = Minimum value
b = Maximum value

X_1 = particular value #1
X_2 = particular value #2

$$\text{Area} = \text{height} * \text{width} = \text{Probability} = P(X_1 \le X \le X_2) = \frac{1}{b-a} * (X_2 - X_1)$$

Area = Probability = Area under graph of $f(x)$ over the interval X_1 to X_2.

$P(X = X_1)$ = Probability for particular value = 0. Lines have NO Area

(Pop Data)

μ = mean = $E(X) = \frac{a+b}{2}$

σ = Standard Deviation = $\sqrt{\frac{(b-a)^2}{12}}$

Note: Because Lines have No Area
$P(1 \le X \le 2) = P(1 < X < 2)$
endpoints included or Not!!

* height of probability Density function is not a probability. Lines have no area.

Example of uniform Probability Distribution:

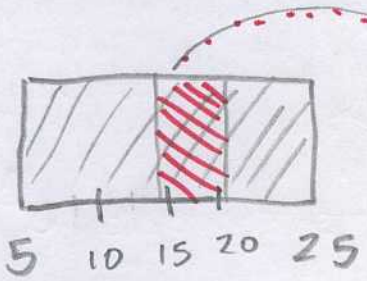
suppose the time that you wait on the telephone for a live representative of your phone company to discuss your problem with you is uniformly distributed between 5 & 25 minutes.

1. what is the mean wait time?
2. what is standard deviation?
3. what is probability that you will wait between 15 & 20 minutes

variables:

X_1 = 15 minutes.
X_2 = 20 minutes.
min = a = 5 minutes
max = b = 25 minutes

$\frac{1}{20} = \frac{1}{25-5}$



$$\mu = \frac{a+b}{2} = \frac{5+25}{2} = 15 \text{ minutes}$$

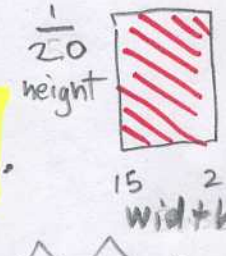
$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(25-5)^2}{12}} = \sqrt{\frac{400}{12}} = 5.77 \text{ minutes}$$

$$P(15 < X < 20) = \frac{1}{25-5} * (20-15) = .05 * 5 = .25$$

Finding Probability for uniform Distribution is a simple Geometry problem of

width * height = Area = Probability.

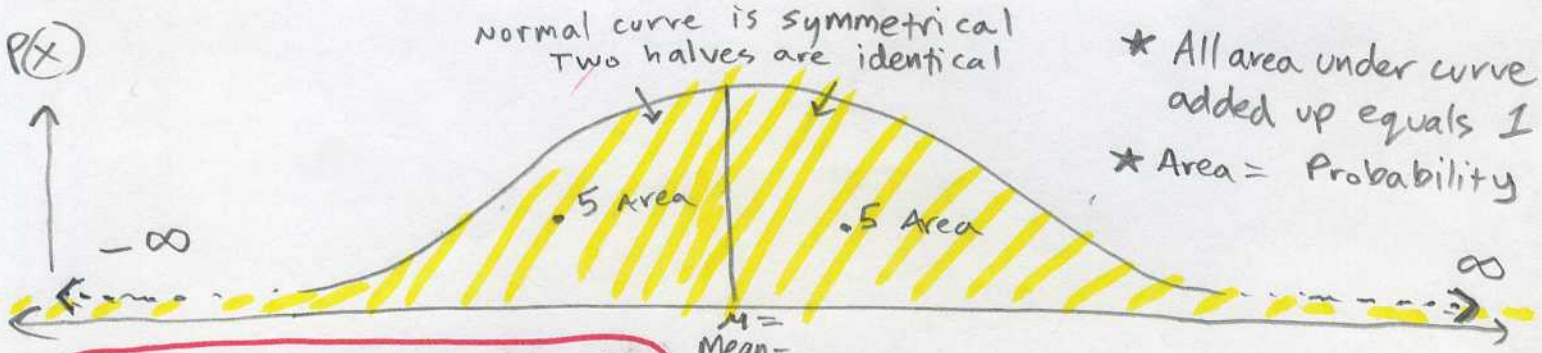
$(20-15) * \frac{1}{20} = \frac{5}{20} = .25$



A: The average wait time is 15 minutes & the probability of waiting between 15 & 20 minutes is .25

* Because Lines have No Area
$P(15 \leq X \leq 20) = P(15 < X < 20)$

Normal (Bell) Probability Distribution



★ All area under curve added up equals 1

★ Area = Probability

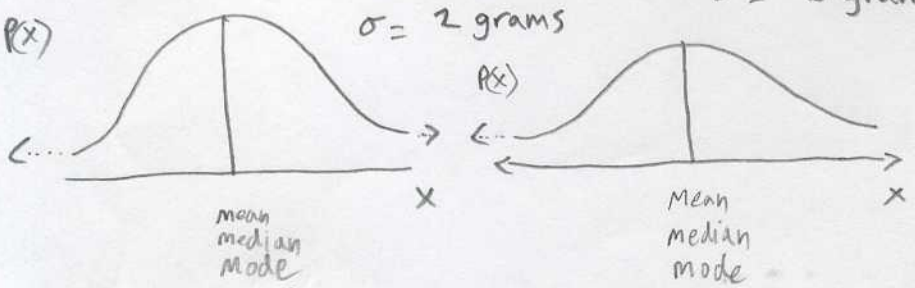
This is a Frequency Polygon (FP) that extends infinitely in both directions - the area under the FP line is our Probability.

$-\infty \leftarrow$ Asymptotic in Both Directions $\rightarrow \infty$

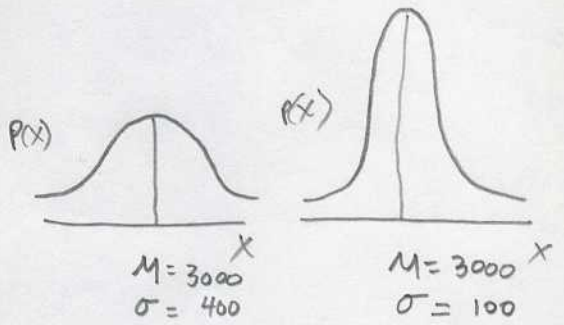
Family of Normal Probability Distributions

As long as the Frequency Polygon looks like a bell shape when you graph it, then you have a Normal Probability Distribution. There are many Distributions that tend to follow the Normal (Bell) Distribution

weight of cerial box



Salaries



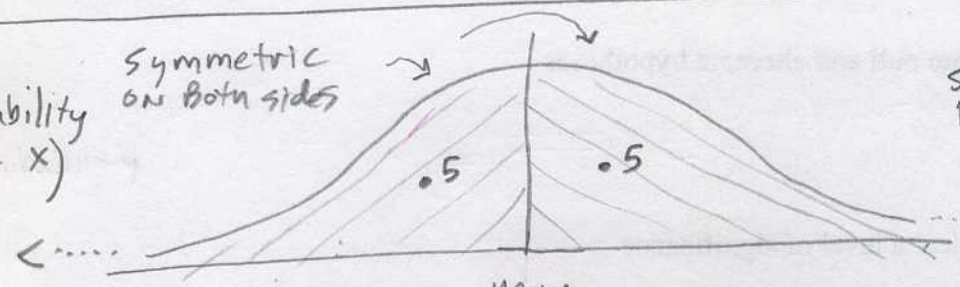
Excel:

For a normal probability distribution:

$$= \text{NORMDIST}(X, \mu, \sigma, 1)$$

← cumulative from $-\infty$ to X

(3) Normal (Bell) shaped Distribution



Area = Probability (Between 2 X values)

Standard Deviation = σ

All Area = 1

This is a Frequency Polygon that extends infinitely in both directions

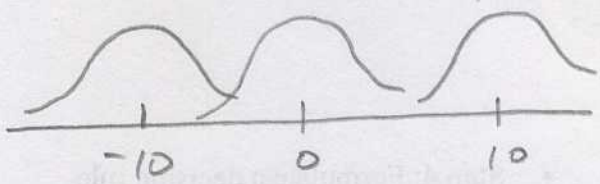
Mean
Median
Mode
M

(1) lots of different Normal Distributions. M determines location on X axis. σ determins height (shape) of distribution.

(2) Highest point is in in middle, where mean = median = Mode.

(3) mean can be any value: negative, zero, positive. Mean determines location

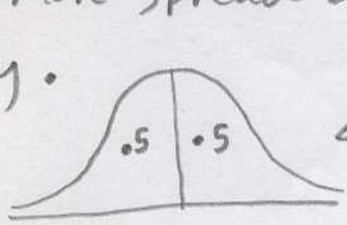
All are same shape (height), but the mean is at a different location on X-axis

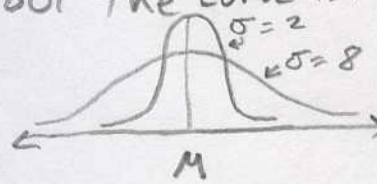


$\sigma = 2$ for all these

(4) Normal curves are symmetric on both sides of mean. Area on each is .5. The curve extends (without ever touching x-axis) in both directions $-\infty$ and ∞. Skew = 0

(5) standard deviation determines shape or height. The bigger σ is the flatter & more spread out the curve is.





(6) All are = 1 = All probability.
Each half = Area = .5

(7) Normal Random Variable:

(a) 68.3% of values are within +/- 1 standard Deviation of mean.
(b) 95.4% of values are within +/- 2 standard Deviations of mean.
(c) 99.7% of values are within +/- 3 Standard Deviations of mean.

Empirical (Normal) Rule ↗

(8) cannot calculate probability for a particular X, only between 2 X values

(4) NORMAL PROBABILITY DENSITY FUNCTION

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

μ = mean
σ = standard deviation
$\pi = \frac{C}{D} \approx 3.14159 =$ PI() (Excel)

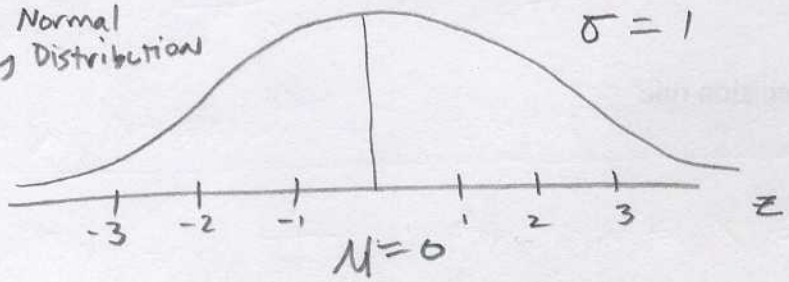
$e \approx 2.71828 =$ EXP(1) ← argumentless function in Excel

1 as only argument in EXP Excel function

$$*e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

(5) Standard Normal Probability Distribution

Same as Normal Probability Distribution except
$\mu = 0$
$\sigma = 1$



$$z = \frac{X-\mu}{\sigma} = \#\text{ SD away from mean}$$

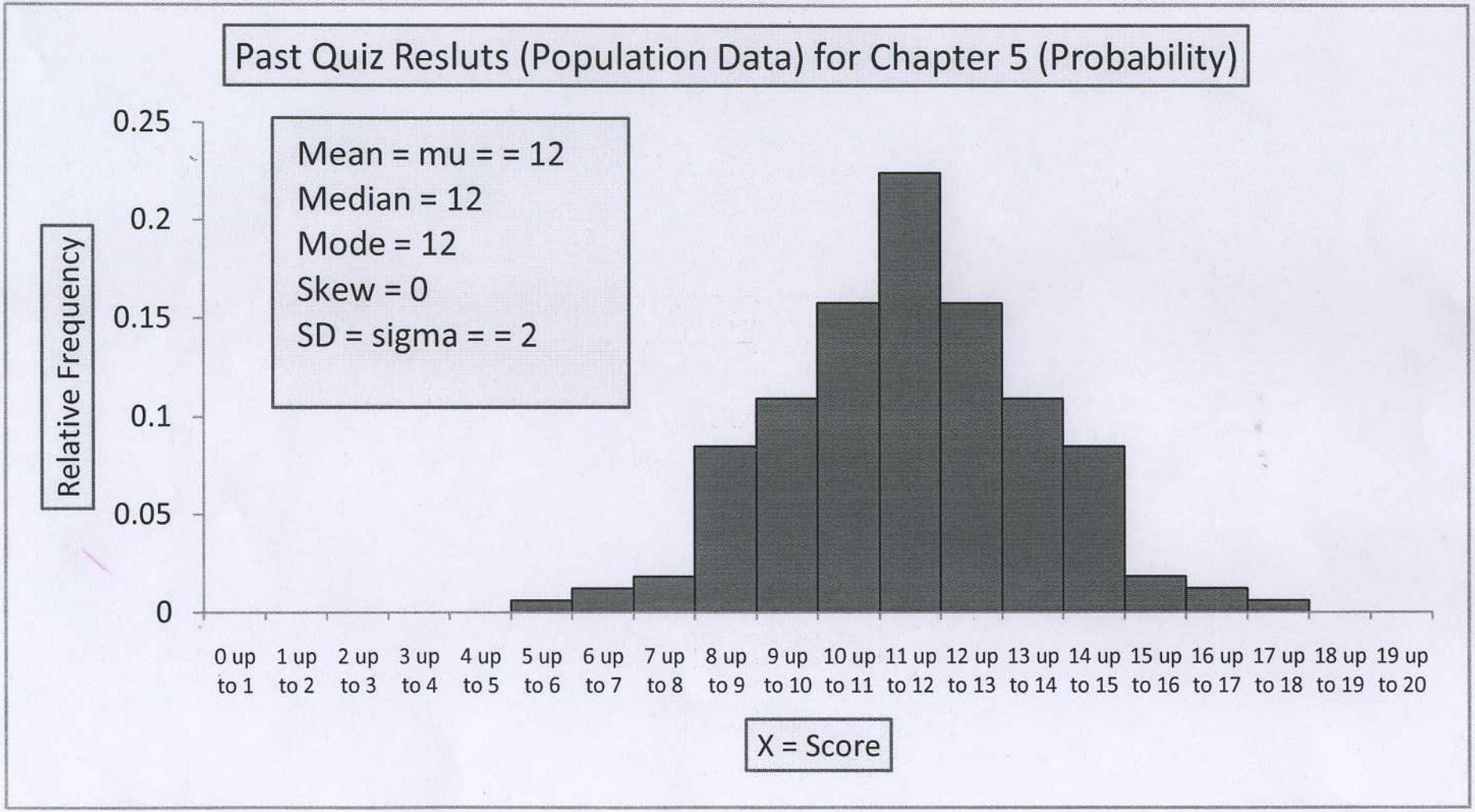
When you do division denominator is always 1σ

(6) Standard Normal Probability Density Function

$$f(z) = \frac{1}{\sqrt{2\pi}} * e^{-z^2/2}$$

$\pi \approx 3.14159 =$ PI()
$e \approx 2.71828 =$ EXP(1)
$z = \frac{x-\mu}{\sigma}$

(P. 9)


Past Quiz Resluts (Population Data) for Chapter 5 (Probability)
Mean = mu = = 12
Median = 12
Mode = 12
Skew = 0
SD = sigma = = 2
Relative Frequency
0.25
0.2
0.15
0.1
0.05
0
0 up to 1
1 up to 2
2 up to 3
3 up to 4
4 up to 5
5 up to 6
6 up to 7
7 up to 8
8 up to 9
9 up to 10
10 up to 11
11 up to 12
12 up to 13
13 up to 14
14 up to 15
15 up to 16
16 up to 17
17 up to 18
18 up to 19
19 up to 20
X = Score

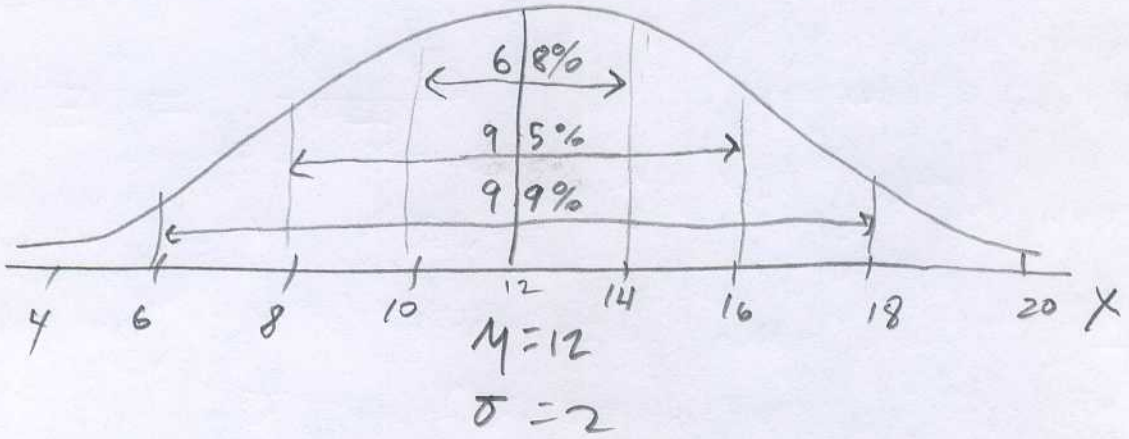
Example :

⇒ Max score for Statistics Quiz = 20

⇒ Population mean = μ = 12

⇒ Pop. Standard Deviation = σ = 2

⇒ The shape of the Distribution from Past Data is Normal (Bell)



Why is this true:

P(1.2<x<1.5) = P(1.2<=x<=1.5)

For Continuous Probability Distributions?

Continuous variables and the associated Probability Distributions are all about area. We will not talk about P(x = 1.2), for example, we will only talk about area (area = probability) between two x values. The reason why we cannot talk about P(x = 1.2) is because it is a line and lines have no area. Thus:

P(1.2<x<1.5) = P(1.2<=x<=1.5) because lines have no area.

f(x = 1.2) and other single x calculations for Continuous Probability Distributions do exist and can be calculated, but they are not probabilities, they are "heights" (how tall).

In the Uniform Probability section, they do not do such a good job of talking about this. They allude to it on **page 248** about half way down in the paragraph with the bold number two before it. They say: **"It also means that the probability of a continuous random variable assuming a value in any interval is the same whether or not the end points are included."** That means this:

P(1.2<x<1.5) = P(1.2<=x<=1.5)

In the section on Standard Normal Probability Distribution on **page 254** in the margin, they explicitly state that:

P(z<=1.00) = P(z<1.00)

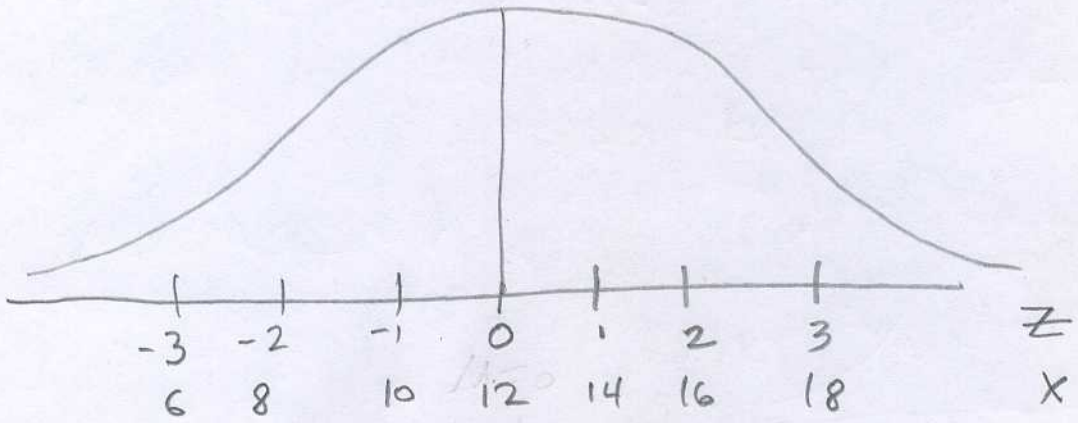
I hope that helps.

(7) Converting TO The Standard Normal Random Variable

$$\left\{\begin{array}{c}\text{Standard}\\ \text{Normal}\\ \text{Random}\\ \text{Variable}\end{array}\right\} = Z = \frac{X-\mu}{\sigma}$$

(similar to chapter 3 when we calculated $z\text{-score} = \frac{X-\bar{X}}{s}$ Sample data)

Z = Standard Normal Random Variable
X = particular X value
μ = population mean
σ = population standard Deviation



μ = median = mode = 0 Z

$\sigma = 1$

Example Statz Quiz:

$\mu = 12$
$\sigma = 2$
$X_1 = 10$
$X_2 = 12$
$X_3 = 14$
$X_4 = 17$

$$Z_1 = \frac{10-12}{2} = -1$$
$$Z_2 = \frac{12-12}{2} = 0$$
$$Z_3 = \frac{14-12}{2} = 1$$
$$Z_4 = \frac{17-12}{2} = \frac{5}{2} = 2.5$$

standard Deviations above or below mean

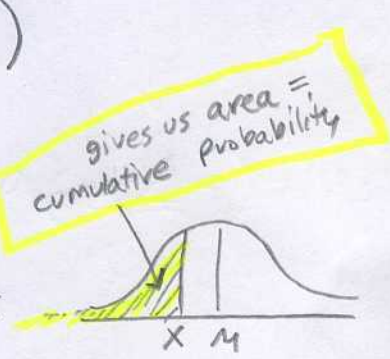
(8) NORMDIST Function

calculating Probabilities for a Normal (Bell) Distribution when you have:

x = particular x value
μ = population mean
σ = population standard deviation

= NORMDIST(x, M, σ, cumulative)

cumulative = 1 (or TRUE)
Returns area from $-\infty$ to x



or

cumulative = 0 (or False)
Returns the height of the curve at a particular x value

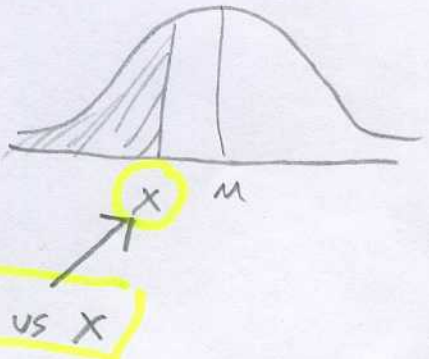
*Excel Help:

1 or TRUE returns the "cumulative distribution function".
0 or FALSE returns the "probability mass function".

(9) NORMINV Function

It takes the cumulative probability ($-\infty$ to x) & will give you the x value

= NORMINV(probability (cumulative), M, σ)

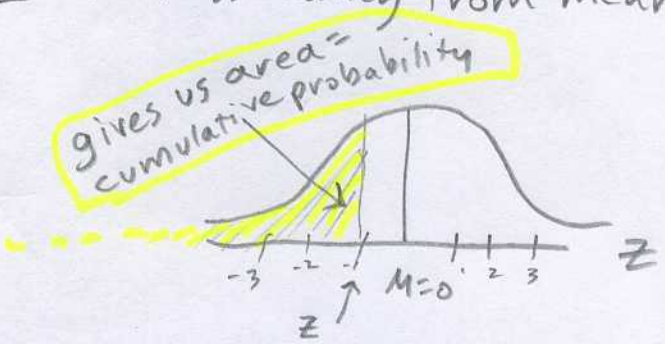


(10) NORMSDIST function

calculating probabilities for a Standard (z) Normal Distribution

$$z = \frac{x - \mu}{\sigma} = \#\text{ of SD 'away from mean.}$$

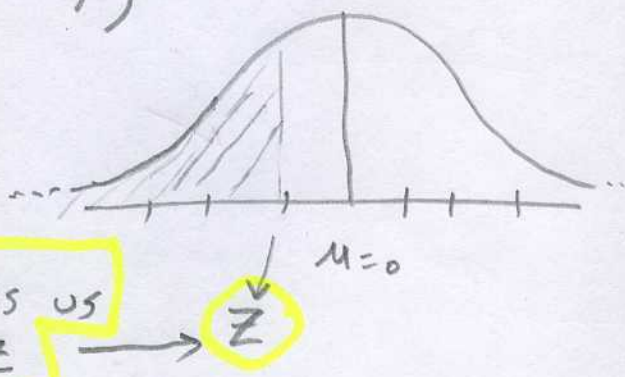
$$= \text{NORMSDIST}(z)$$



(11) NORMSINV function

It takes the cumulative probability ($-\infty$ to x) and gives you the associated z-score.

S = NORMSINV (cumulative probability)

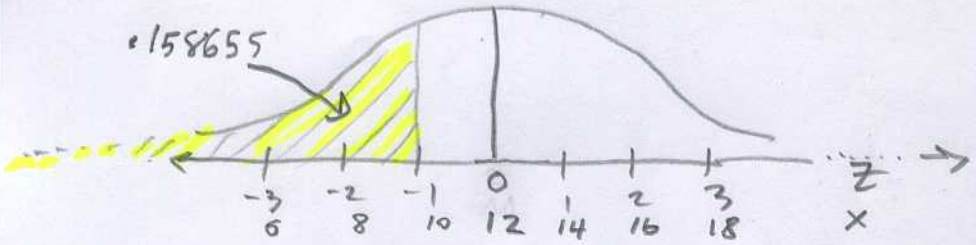


"S" in function name means:

"Standard" Normal probability Distribution.

(12) 3 types of Probabilities to calculate:

(1) Probability that X or Z will be less than or equal to a given value.



$X = 10$ = particular score
$\mu = 12$ = mean
$\sigma = 2$ = SD
$Z = \frac{10-12}{2} = \frac{-2}{2} = -1$

= NORMDIST(10,12,2,1) = .158655

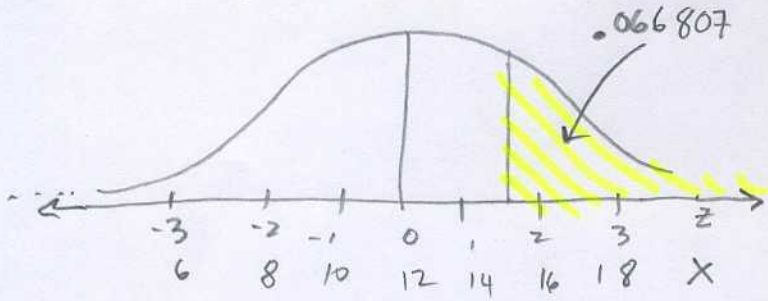
S = NORMSDIST(-1) = .158655

$P(X < 10) = P(X \leq 10) = P(Z < -1) = P(Z \leq -1) = .1586$

= NORMINV(.158655,12,2) = 10 } Inverse

S = NORMSINV(.158655) = -1

(2) Probability that X or Z will be greater than or equal to a given value



$X = 15$
$\mu = 12$
$\sigma = 2$
$Z = \frac{15-12}{2} = 1.5$

$P(X > 15) = P(X \geq 15) = P(Z > 15) = P(Z \geq 15) = .066807$

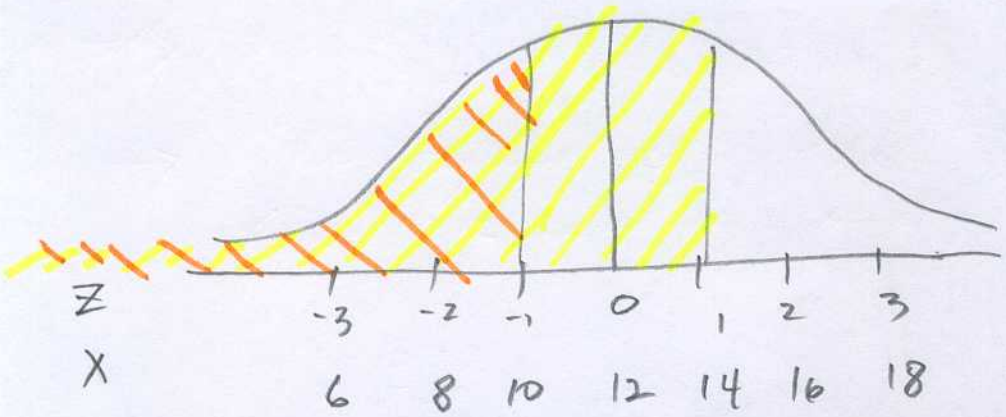
= 1 − NORMDIST(15,12,2,1) = .066807

S = 1 − NORMSDIST(1.5) = .066807

Inverse { = NORMINV(1−.066807,12,2) = 15

S = NORMSINV(1−.066807) = 1.5

(3) Probability that X or Z will be between X_1 and X_2



$X_1 = 10$

$X_2 = 14$

$M = 12$

$\sigma = 2$

$Z_1 = \frac{10-12}{2} = -1$

$Z_2 = \frac{14-12}{2} = 1$

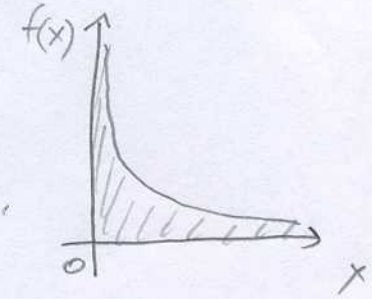
RULE:
Bigger Area − smaller Area = Find Area (Probability) between X_1 & X_2.

= NORMDIST(14, 12, 2, 1) − NORMDIST(10, 12, 2, 1) = .682

S = NORMSDIST(1) − NORMSDIST(−1) = .682689

(13) Exponential Probability Distribution

useful in computing probabilities for the time it takes to complete a task or distance between similar occurences.



Examples:
* Time between arrivals at car wash
* Time to take a test
* Distance between potholes on a road

(14) Exponential Density Function

$$f(x) = \frac{1}{\mu} * e^{-x/\mu} \quad \text{for } x \geq 0, \mu > 0$$

★ Property of Exponential Dist. → μ = mean = σ = standard Deviation

$e \approx 2.71828$ = EXP(1) in Excel

(15) Exponential cumulative Probabilities

$$P(X \leq X_0) = 1 - e^{-X_0/\mu}$$

X_0 = particular X

(16) EXPONDIST Function

called "lambda"

= EXPONDIST (X, $\frac{1}{\mu}$, cumulative)

1 or TRUE to get cumulative from 0 to X

0 or FALSE to get height

Ⓐ continuous Random Variable probability Distribution → Find Area Between 2 X values

(17) Relationship between Poisson & Exponential Distributions

(A) Poisson Distribution provides an appropriate description of the Number of occurrences per interval

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \quad \text{or} \quad =\text{POISSON}(x, \mu, \text{cum...})$$

↑ Excel Function

Exponential Distribution provides an appropriate description of the length of the interval between occurences.

Density: $f(x) = \frac{1}{\mu} e^{-x/\mu}$ for $x \geq 0, \mu > 0$

cumulative: $P(x \leq x_0) = 1 - e^{-x_0/\mu}$

(B) If arrivals follow a Poisson Distribution, the time between arrivals must follow an Exponential Distribution.

Note on Exponential:

The skewness measure for exponential Distributions is 2.

Exponential Example:

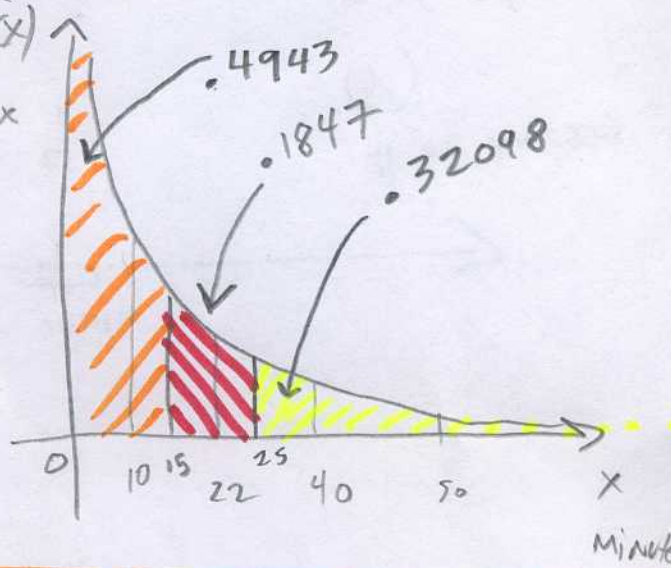
(Time in line)

The average time to get to a Disney ride during peak hours follows a Exponential Distribution.

μ = time to stand in line = 22 minutes

cumulative Exponential Formula: $= 1 - e^{-x/\mu}$

x = particular x
μ = mean
e = 2.7182



Probability stand in line for 15 or less

$P(X \leq 15) = 1 - e^{-\frac{15}{22}}$

$= .494303$

$P(X \leq 15) =$ =EXPONDIST(15, 1/22, 1) = .494303

Probability stand in line for 25 minutes or more

$P(x \geq 25) = e^{-\frac{25}{22}} = .320984117$

$= 1 - EXPONDIST(25, 1/22, 1) = .32098$

$P(15 \leq X \leq 25) = EXPONDIST(25, 1/22, 1) - EXPONDIST(15, 1/22, 1)$

$= .1847$

* Area between 2 ⇒ Big Area − small Area