Ch. 6 Continuous Probability Distributions (P.1)

Discrete Random variable Ch. 5	Continuous Random variable Ch. 6
① GAPs between numbers 1, 2, 3...	① No Gaps between numbers 1 ... 2
② can only assume clearly separated values	② can assume any value in interval or collection of intervals
③ usually from counting successes	③ Depends on measuring instrument
④ Examples: - Number of Rooms used - number of customers - number of times late - Cards pulled from Deck	④ Examples - weight of ketchup bottle - salary Earned (Dollars) - Gas prices - score on test

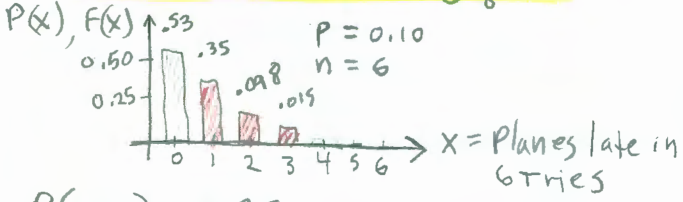
Both share definition Probability Distribution:

A description / Presentation of how the probabilities are distributed over the values of the Random Variable

⑤ Discrete Probability Distribution

⑥ Types: ① Binomial Distribution
② Poisson Distribution
③ Hypergeometric Distr.

⑦ Binomial Looks like:



$P(x=1) = 0.35$

$P(1 <= X <= 3) = 0.35 + 0.098 + 0.0015 = 0.47$

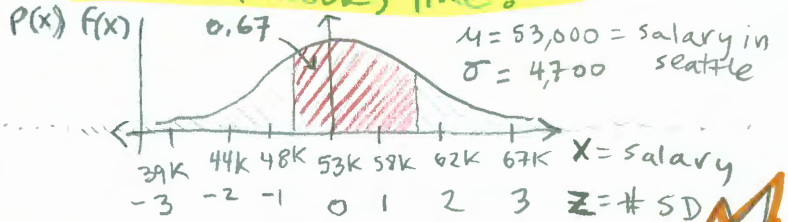
⑧ Questions you can ask:
① Probability of exactly X
② Probability between 2 X values

⑨ Probability Function calculates Probability & height of column

⑤ Continuous Probability Distribution

⑥ Types: ① Normal (Bell) Distribution
② Uniform Distribution
③ Exponential Distribution

⑦ Normal Looks like:



$P(X = 50{,}000) =$ CAN NOT DO!!!

$P(50{,}000 <= X <= 60{,}000) = 0.67$

CAN NOT CALCULATE Exactly X

⑧ Questions you can ask:
① Probability between 2 X values

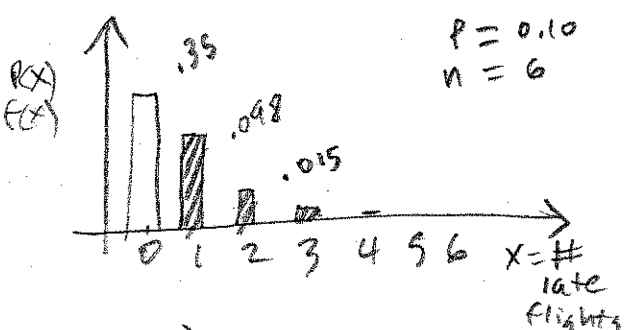
⑨ Probability Density function provides height of curve for plotting chart

Discrete

(10) Binomial Probability Function

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

(1) Tells you height of column
(2) calculates probability of an exact X value
(3) How we calculate probability between 2 X values:
Add each individual Prob.



$P(1 <= x <= 3) = 0.35 + 0.098 + 0.015$
$= 0.47$

or for 1 X value:

$P(1) = 0.35$

If you pick an X value like 50k the area under that point is zero. This means we cant calculate, or we can: it is zero.

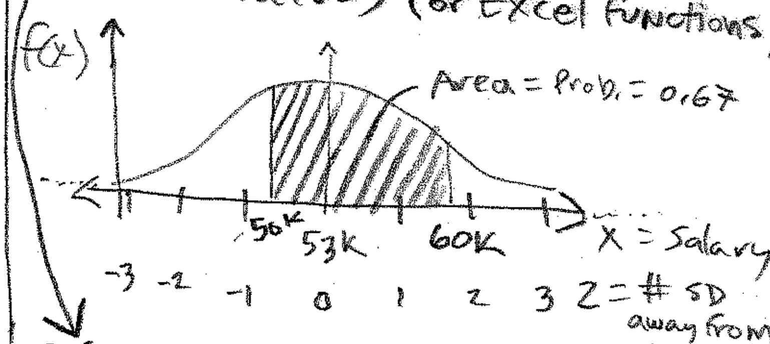
$P(50k) = 0$

Continuous

(10) Normal Probability Density Function

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

(1) calculates height of plotted curve for charting
(2) Does NOT calculate probability of an exact X.
(3) How we calculate Probability between 2 X values:
Integral calculus to calculate Area under curve between 2 X values (or Excel functions)



$P(50K <= x <= 60K) =$

= NORM.DIST(60K, 53K, 4.7K, 1) − NORM.DIST(50K, 53K, 4.7K, 1) =
0.67

(11) Continuous Probability Distributions

Area = Probability
All Area = 1
Because lines have No area:
(1) $P(X = 50K) = 0$
(2) $P(X \geq 50K) = P(X > 50K)$

(12) For continuous Probability Distributions

We Ask Questions like:

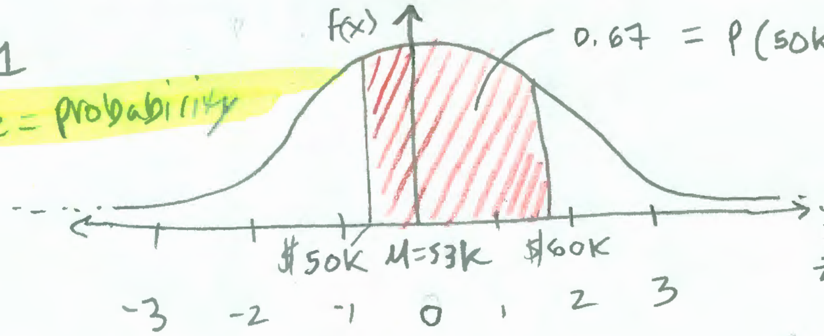
Q1: what is probability that we could get an accounting job in seattle area between #50k & #60k?

ALL Area = 1

(*) Area under curve = probability

$\mu = \$53K$

$\sigma = \$4.7K$



(Normal Example)

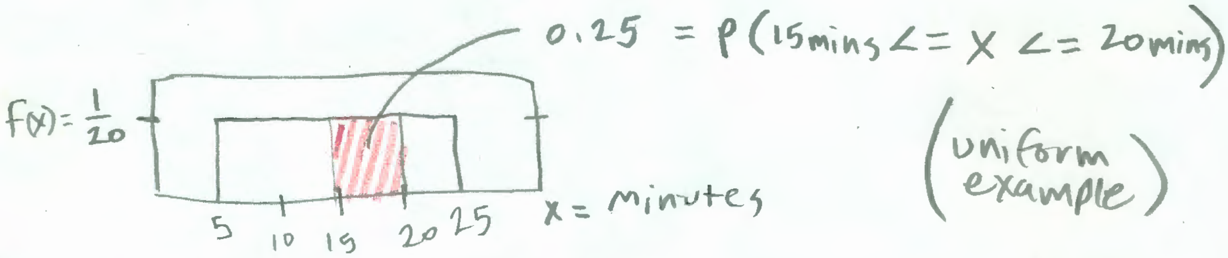
x = sarary

z = # SD away From Mean

A1: The probability comes from the area under the curve from X = 50K to X = 60K

Q2: what is the probability that we will wait on hold (on the phone) for customer service help between 15 and 20 minutes?

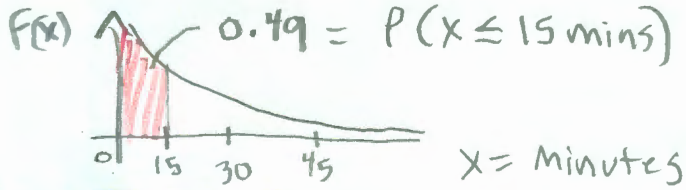
(*) Area in rectangle = probability



(uniform example)

A2: The probability comes from the area in the rectangle from 15 mins to 20 mins

Q3: what is probability that you will have to wait less than 15 mins in Disneyland line?



X = Minutes

A3: Prob. comes from the area under the curve from 0 mins to 15 mins.

(*) Area under curve = Probability

(Exponential Example)

→ 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.

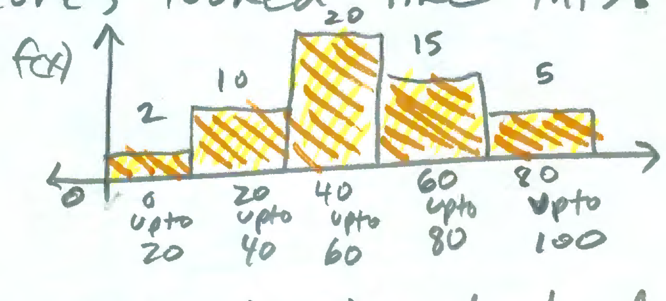
μ = mean of population
σ = Standard Deviation of population

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

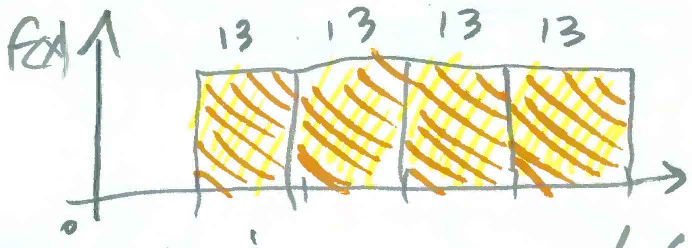
$\bar{X}$ = mean of sample
S = SD of sample

Visual Introduction to Uniform Probability Distr.

Imagine the distribution of statistics Quiz scores looked like this:



Now imagine if it looked like this:

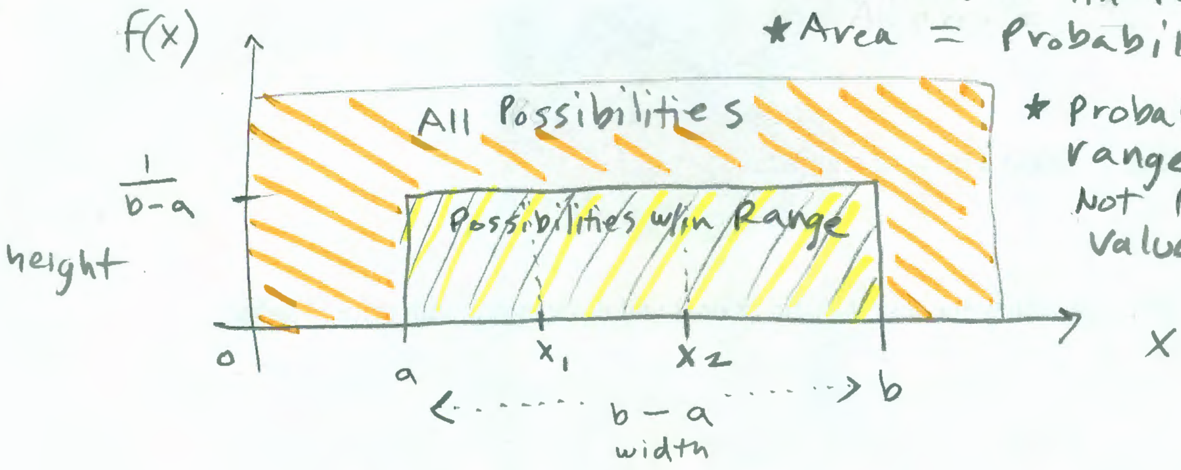


Uniform Distribution!!

① Uniform Probability Distribution

★ All area within Range = 1

★ Area = Probability



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

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

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

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

② UNIFORM PROBABILITY DENSITY FUNCTION

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq X \leq b \\ 0 & \text{elsewhere} \end{cases}$$

also $f(x) \geq 0$ for all X

a = Minimum value

b = Maximum value

x_1 = particular value #1

x_2 = particular value #2

Area = height * width = Probability = $P(x_1 \leq X \leq x_2) = \frac{1}{b-a} * (x_2 - x_1)$

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

$P(1 \leq X \leq 2) =$
$P(1 < X < 2)$

endpoints included or Not!!

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

p. 5

Example of uniform Probability Distribution:

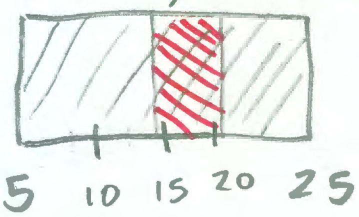
Suppose the time that you wait on the telephone for a live representative of your phone company to discuss your problem with you is uniformly distributed between 5 & 25 minutes.

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

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



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

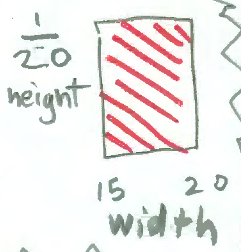
$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(25-5)^2}{12}} = \sqrt{\frac{400}{12}} = 5.77 \text{ minutes}$$

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

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 $P(15 \le X \le 20) = P(15 < X < 20)$

3

Normal (Bell) Probability Distribution

1)

Normal curve is symmetrical
Two halves are identical

★ 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

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

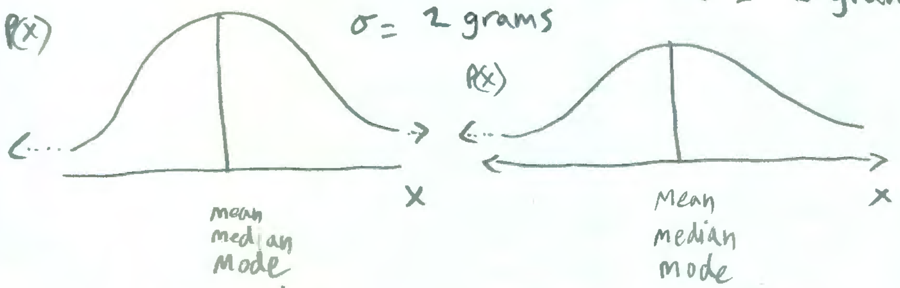
Family of Normal Probability Distributions

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:

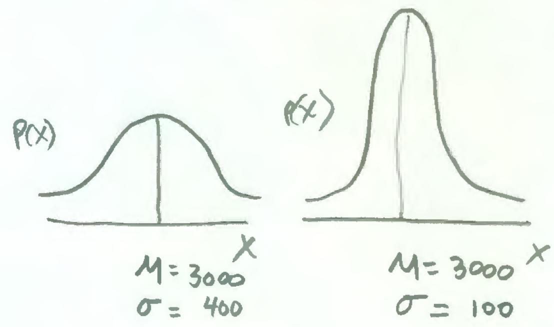
weight of cerial box

μ = 250 grams
σ = 2 grams

μ = 300 grams
σ = 3 grams



Salaries



Excel:

For a normal probability distribution:

= NORM.DIST (X, μ, σ, 1) ← cumulative from $-\infty$ to X

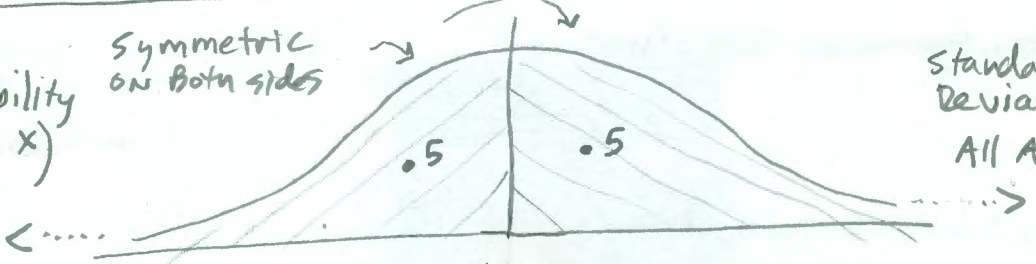
3 Normal (Bell) shaped Distribution

Symmetric on Both sides

a = Probability (Between 2 X values)

Standard Deviation = σ

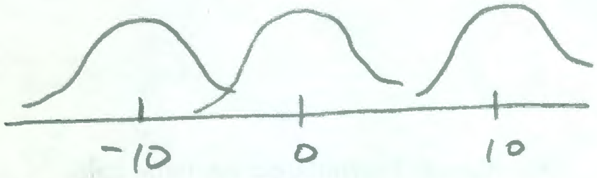
All Area = 1



This is a Frequency Polygon that extends infinitely in both directions

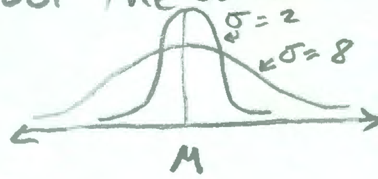
1. lots of different Normal Distributions. μ determines location on X axis. σ determins height (shape) of distribution.
2. Highest point is in in middle, where mean = median = Mode.
3. mean can be any value: negative, zero, positive. mean determines location

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

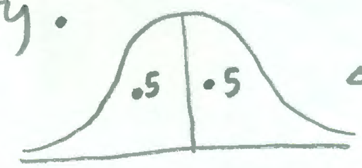


σ = 2 for all these

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 -∞ and ∞. skew = 0
5. standard deviation determines shape or height. The bigger σ is the flatter & more spread out the curve is.



6. All area = 1 = All probability. Each half = Area = .5



7. <u>Normal Random Variable:</u>

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

<u>Empirical (Normal) Rule</u>

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)

argumentless function in Excel

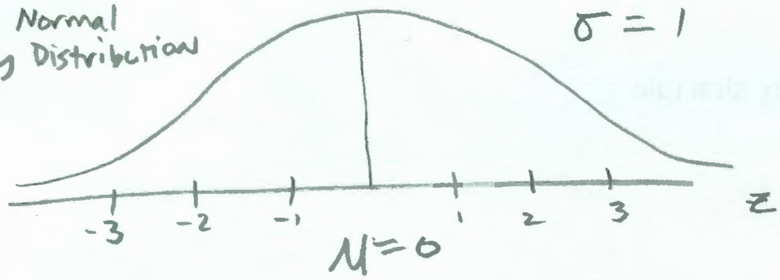
$e \approx 2.71828 = EXP(1)$

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

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

$\pi \approx 3.14159 = PI()$

$e \approx 2.71828 = EXP(1)$

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

7 Convert to 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 Ch. 3 when we calculate Z-score $\frac{x - \bar{x}}{s}$ Sample Data)

Z = Standard Normal Random Variable = # SD away from mean

X = particular X-value

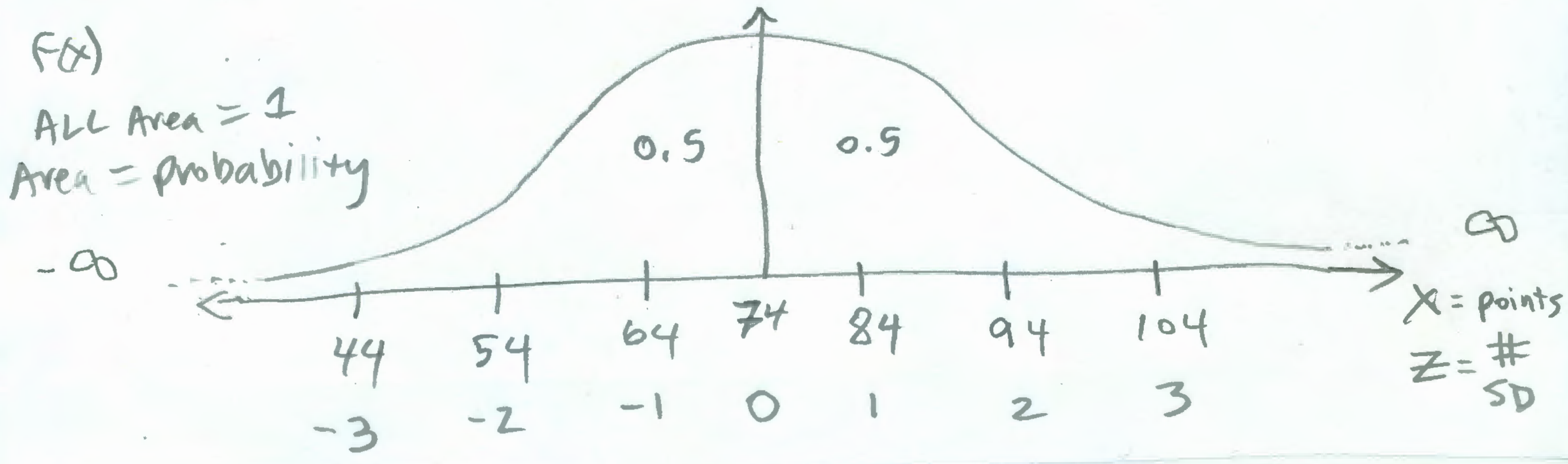
μ = population mean

σ = population Standard Deviation

Test Score Example:

$\mu = 74$ points on Test pop. data

$\sigma = 10$ points on Test pop. data



$$Z_0 = \frac{74 - 74}{10} = 0$$

$$Z_1 = \frac{84 - 74}{10} = 1$$

$$Z_{-1} = \frac{64 - 74}{10} = -1$$

Standard Normal curve shows # of Standard Deviations above / below mean.

8 Excel functions for Normal Prob. Distributions

a) when you have X – values:

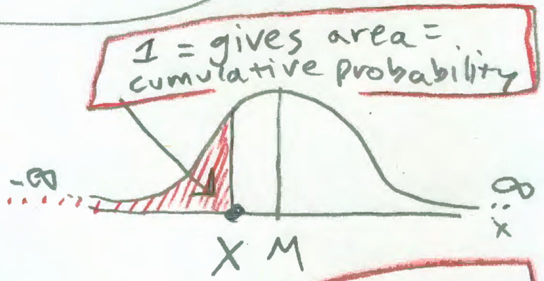
NORM.DIST (X, mean, standard-dev, cumulative)

X = particular x-value = x

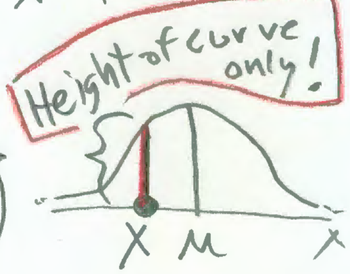
μ = population mean = mean

σ = pop. standard deviation = standard_dev

1 = TRUE = {Cumulative Distribution function} = {Returns Area (Probability) from $-\infty$ to X}



0 = FALSE = {Probability Mass function} = {Height of curve at particular X} = {Not Probability}



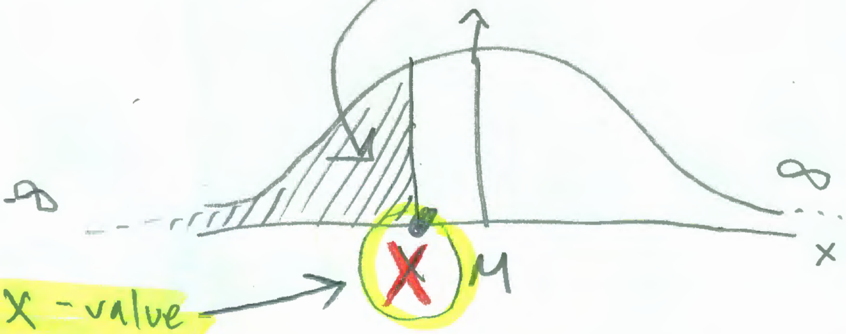
use when you want to plot chart

NORM.INV (cumulative Probability, mean, standard-dev)

cumulative Probability = Probability from $-\infty$ to X

μ = population mean = mean

σ = pop. standard deviation = standard_dev



NORM.INV gives us X-value

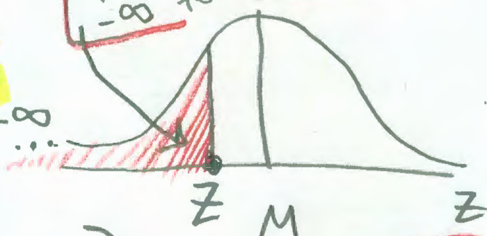
8 b when you have Z - values :

NORM.S.DIST (Z , cumulative)

Z = { Standard Normal Random variable } = { Number of Standard Deviations away from mean }

1 = TRUE = { cumulative Distribution function } = { Returns Area (Probability) from −∞ to Z }

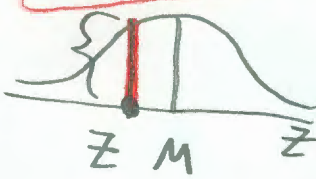
1 gives us area = probability from −∞ to Z



0 = FALSE = { Probability Mass Function } = { Height of curve at particular Z } = { Not Probability }

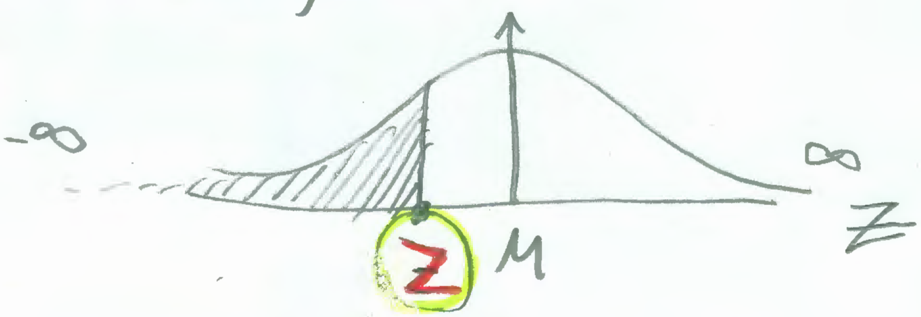
use when you want to plot chart

Height of curve only



NORM.S.INV (cumulative probability)

cumulative probability = probability from −∞ to Z



NORM.S.INV gives you Z

Get Probability (or curve Height):

x → NORM.DIST
z → NORM.S.DIST

.DIST gets you Probability or height

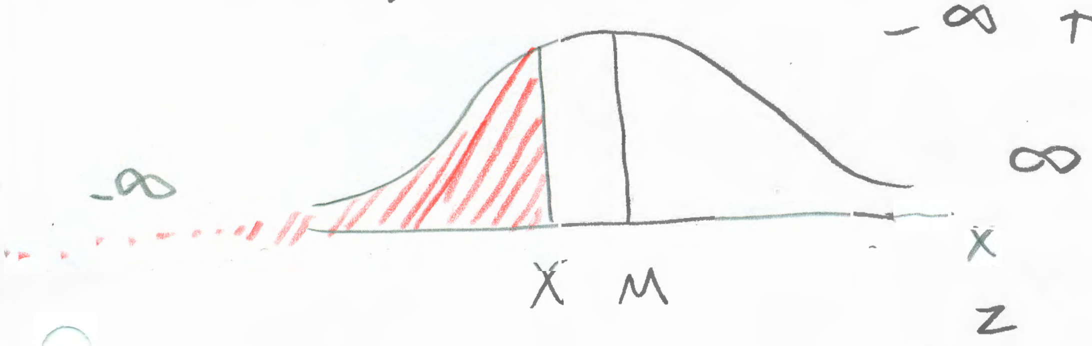
Get Particular value:

X → NORM.INV
z → NORM.S.INV

.INV gets you particular value

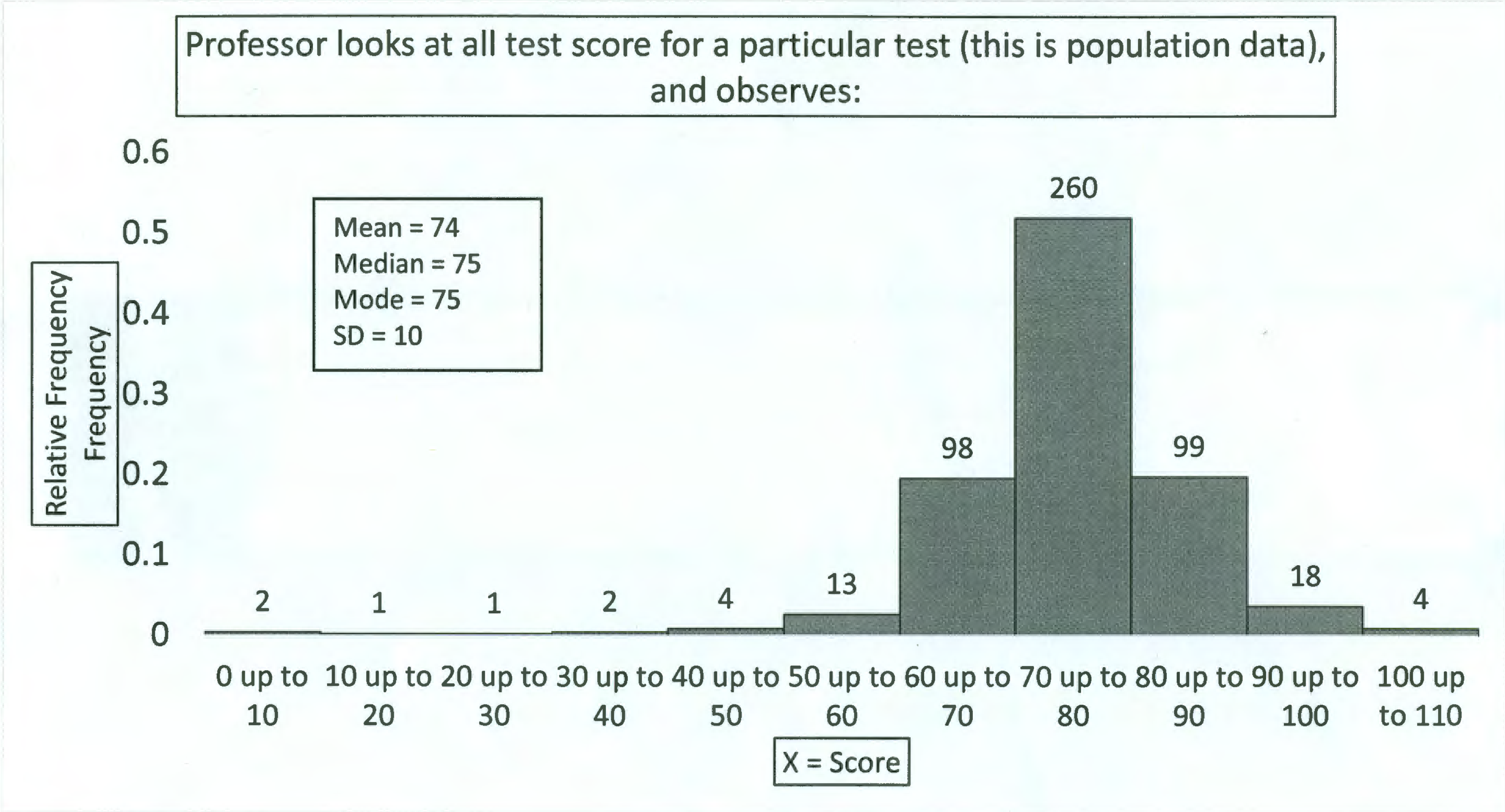
Key Function concept:

→ Functions calculate probability from $-\infty$ to X !!!!!



(P.13)

9 Example of when you can use Normal/Bell Prob. Distribution Model:



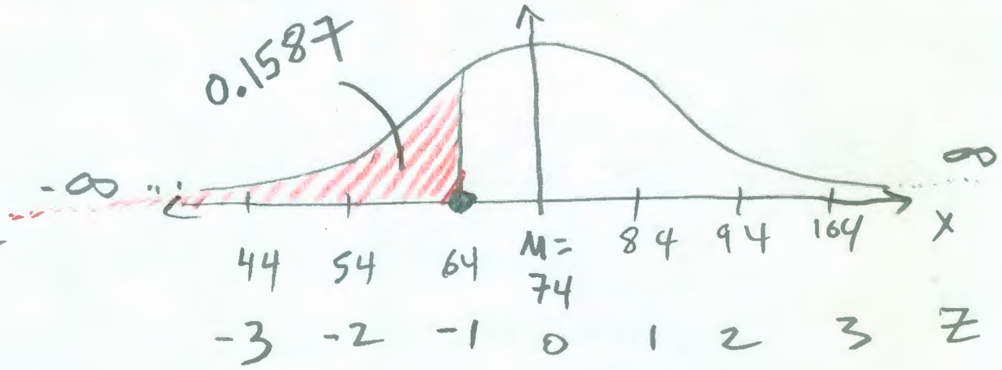
(★) Beause population data has a Normal / Bell shape distribution, we can use Normal Probability Distribution Model!!

3 Types of Probability to calculate:

① Probability that X or Z will be less than or equal to a particular value.

"What are chances that I will score 64 or less on test?"

$X = 64$
$\mu = 74$
$\sigma = 10$
$Z = \dfrac{64-74}{10} = -1$



=NORM.DIST(64,74,10,1) = 0.1587

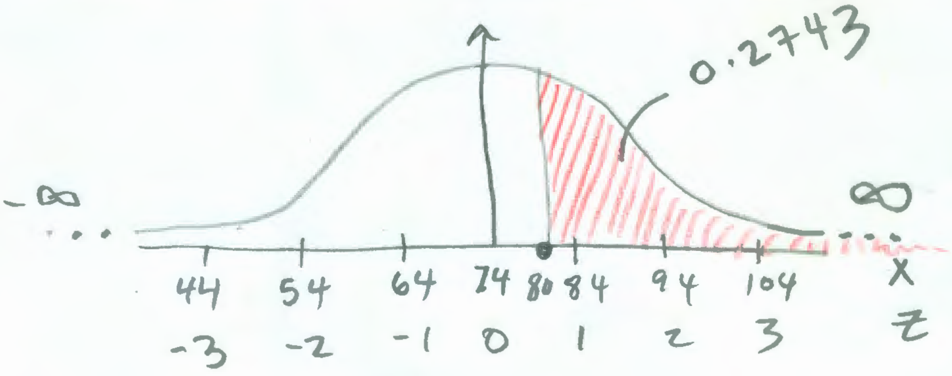
=NORM.S.DIST(-1,1) = 0.1587

$P(X \le 64) = P(X < 64) = P(Z \le -1) = P(Z < 1) = 0.1587$

② Probability that X or Z will be greater than or equal to a particular value

"What are the chances that I will score 80 or above?"

$X = 80$
$\mu = 74$
$\sigma = 10$
$Z = \dfrac{80-74}{10} = 0.6$



=1 - NORM.DIST(80,74,10,1) = 0.2743

=1 - NORM.S.DIST(0.6,1) = 0.2743

$P(X \ge 80) = P(X > 80) = P(Z \ge 0.6) = P(Z > 0.6) = 0.2743$

3 Probability that X or Z will be between two X values or two Z values.

"what are the chances that I will score between 75 and 90 points?"

$X_{lower} = 75$

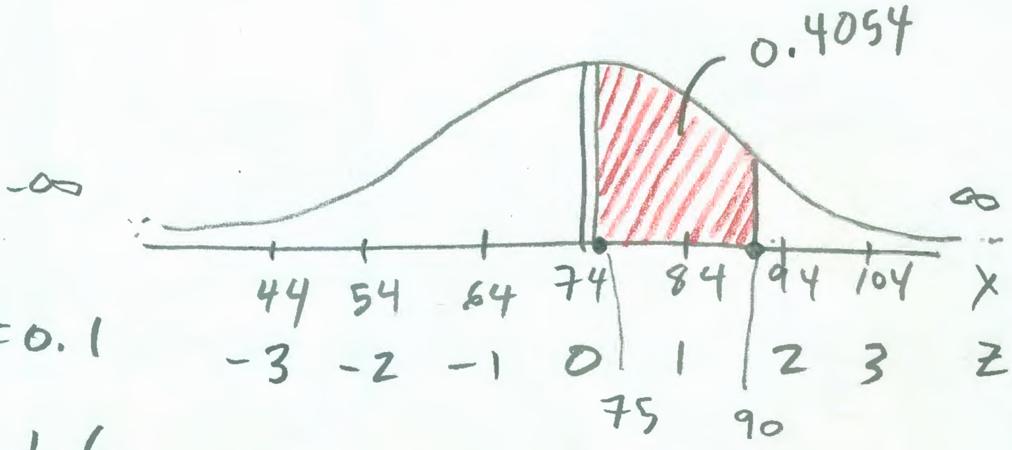
$X_{upper} = 90$

$\mu = 74$

$\sigma = 10$

$Z_{lower} = \frac{75-74}{10} = 0.1$

$Z_{upper} = \frac{90-74}{10} = 1.6$



Rule: Bigger Area − Smaller Area

= NORM.DIST(90,74,10,1) − NORM.DIST(75,74,10,1) = 0.4054

= NORM.S.DIST(1.6,1) − NORM.S.DIST(0.1,1) = 0.4054

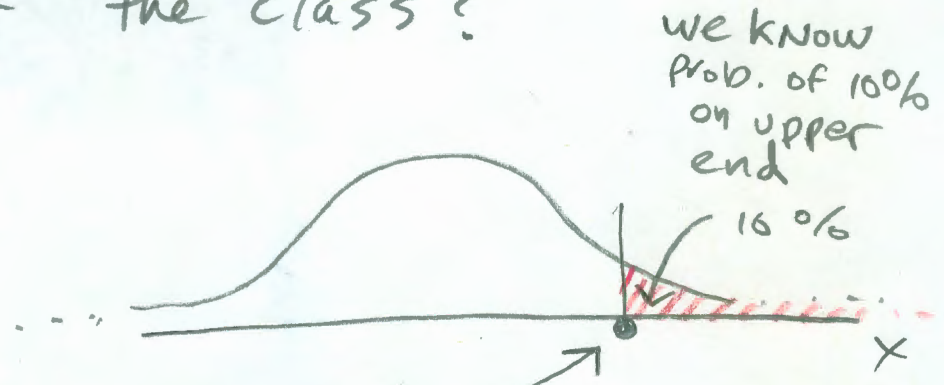
$P(75 \le X \le 90) = P(75 < X < 90) = P(0.1 \le Z \le 1.6) = P(0.1 < Z < 1.6) = 0.4054$

11 # Calculate X or Z given Probability

Example 1:

What is the score needed to be in the top 10% of the class?

$\mu = 74$
$\sigma = 10$



We need to solve for X or Z

Remember: The distribution functions always work from $-\infty$ to the X or Z!!

= NORM.INV(1-10%, 74, 10) = 86.8 points

= NORM.S.INV(1-10%) = 1.28 Standard Deviations above mean

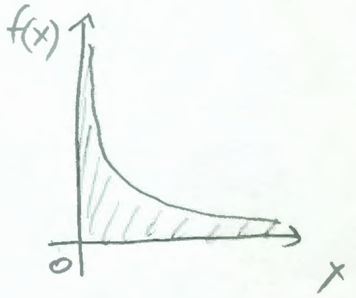
Example 2:

12 If Seattle Accountant Jobs pay on Average \$53,000 w/ a standard Deviation of \$4,700 & we assume these numbers represent a Normally distributed population data set, below what value do 40% of accountants earn?

= NORM.INV(40%, 53000, 4700) = \$51,809.00

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 EXPON.DIST Function

=EXPON.DIST(x, $\frac{1}{\mu}$, cumulative)

called "lambda"

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 <u>Number of occurrences per interval</u>

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

↑ Excel Function

Exponential Distribution provides an appropriate description of the <u>length of the interval between occurences.</u>

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

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

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

<u>Note on Exponential:</u>

The skewness measure for exponential Distributions is 2.

Exponential Example:

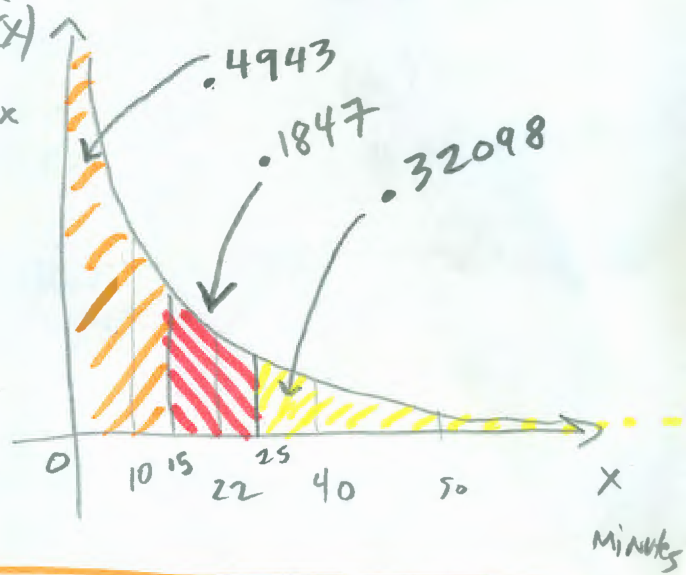
(Time in line)

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 = particlar x
μ = mean
e = 2.7182



probability stand in line for 15 or less

$$P(X \le 15) = 1 - e^{-\frac{15}{22}}$$
$$= .494303$$

$P(x \le 15)$ = =EXPON.DIST(15, 1/22, 1) = .494303

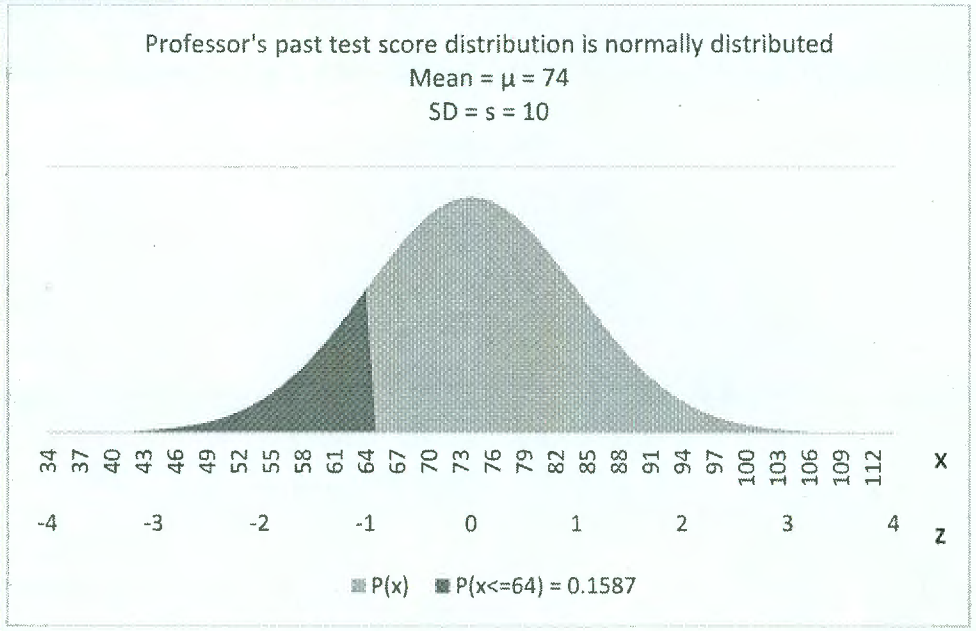
Probability stand in line for 25 minutes or more

$$P(x \ge 25) = e^{-\frac{25}{22}} = .32098417$$
= 1 − EXPON.DIST(25, 1/22, 1) = .32098

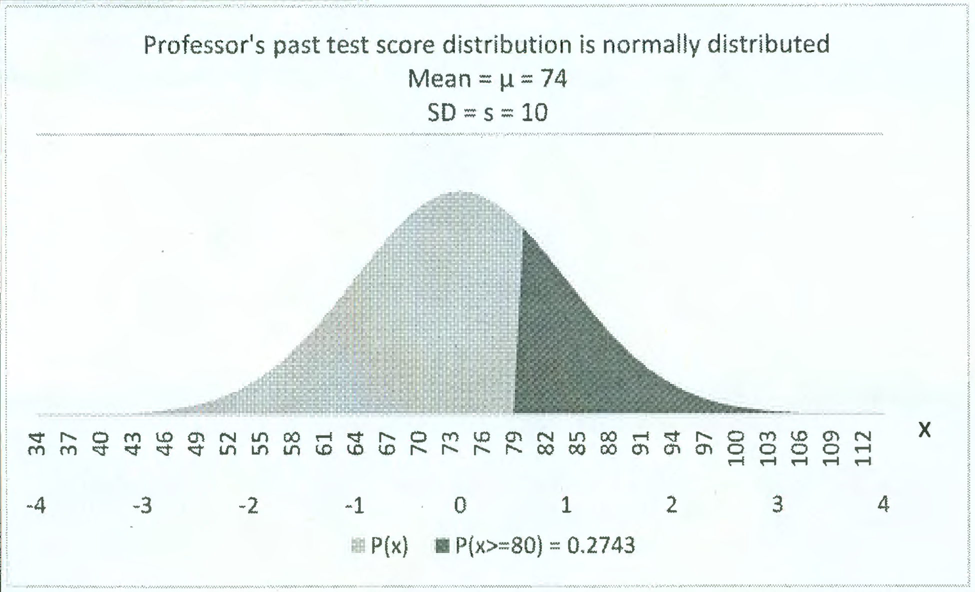
$P(15 \le X \le 25)$ = EXPON.DIST(25, 1/22, 1) − EXPON.DIST(15, 1/22, 1)
= .1847

* Area between 2 ⇒ Big Area − small Area

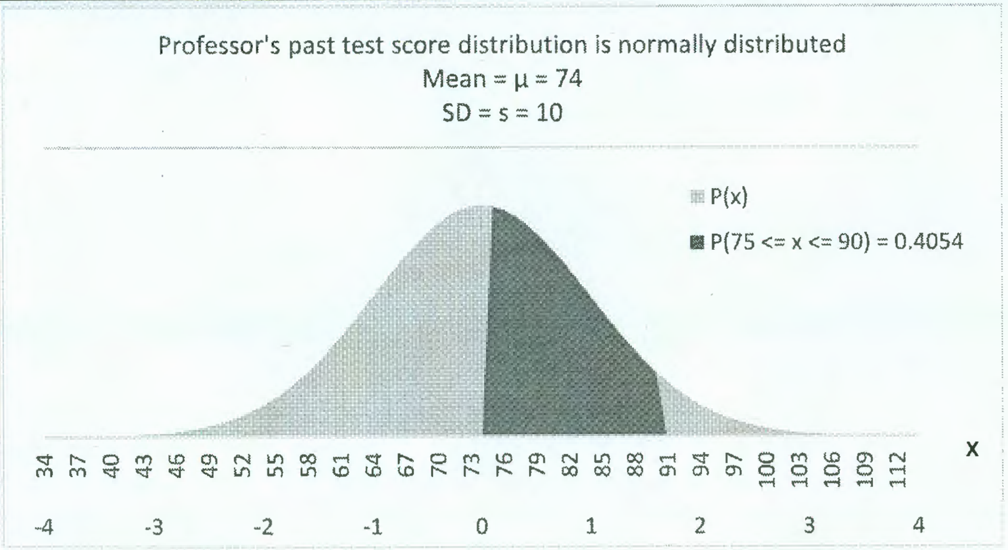
	A	B	C	D	E	F	G
1	**Professor's past test score distribution is normally distributed**						
2	Mean = μ	74					
3	SD = σ	10					
4	x	64					
5	Operator	<=					
6	P(x<=64)	0.158655254	=NORM.DIST(B4,B2,B3,1)				
7	z	-1	=(B4-B2)/B3				
8	P(z<=-1)	0.158655254	=NORM.S.DIST(B7,1)				
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							



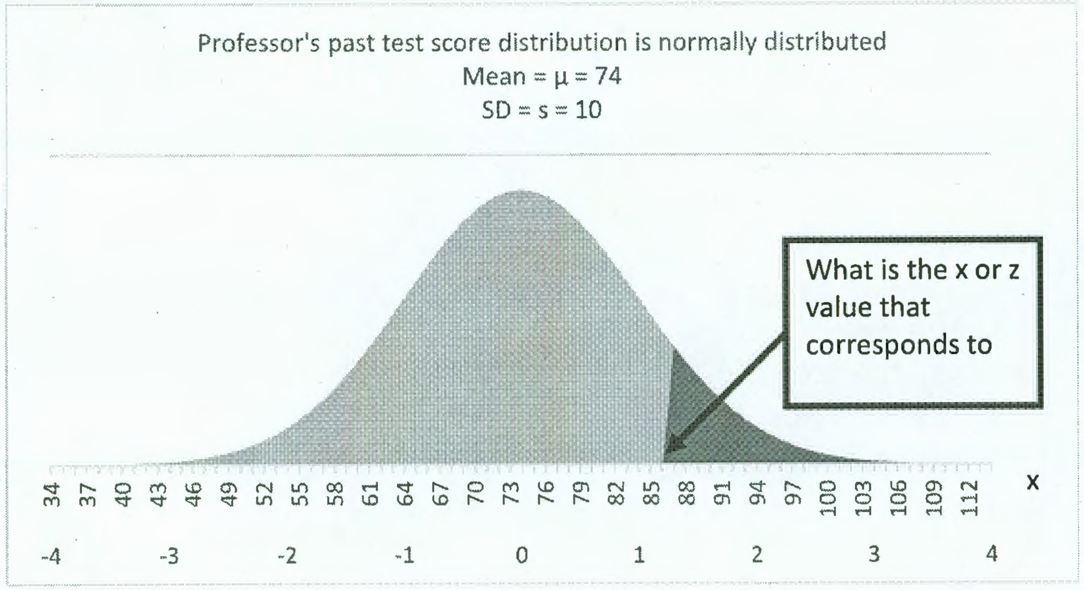
	A	B	C	D	E	F	G
1	**Professor's past test score distribution is normally distributed**						
2	Mean = μ	74					
3	SD = σ	10					
4	x	80					
5	Operator	>=					
6	P(x>=80)	0.274253118	=1-NORM.DIST(B4,B2,B3,1)				
7	z	0.6	=(B4-B2)/B3				
8	P(z>=0.6)	0.274253118	=1-NORM.S.DIST(B7,1)				



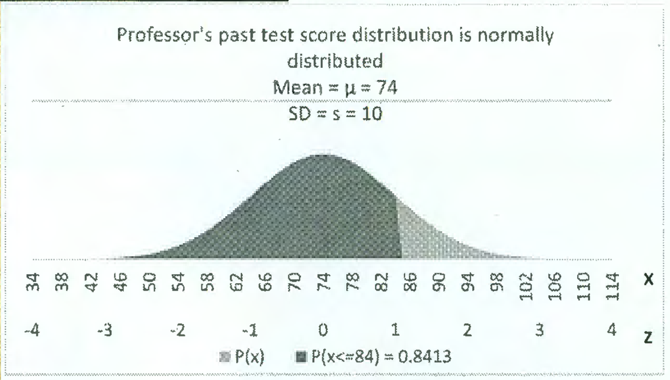
	A	B	C	D	E	F
1	**Professor's past test score distribution is normally distributed**					
2	Mean = μ	74				
3	SD = σ	10				
4	x lower	75				
5	x upper	90				
6	Operator	between				
7	P(75 <= x <= 90)	0.405372871	=NORM.DIST(B5,B2,B3,1)-NORM.DIST(B4,B2,B3,1)			
8	z lower	0.1	=(B4-$B$2)/$B$3			
9	z upper	1.6	=(B5-$B$2)/$B$3			
10	P(0.1 <= z <= 1.6)	0.405372871	=NORM.S.DIST(B9,1)-NORM.S.DIST(B8,1)			
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						



	A	B	C	D	E	F	G
1	**Professor's past test score distribution is normally distributed**						
2	Mean = μ	74					
3	SD = σ	10					
4	Probability	10%					
5	X value	86.81551566	=NORM.INV(1-B4,B2,B3)				
6	z value	1.281551566	=NORM.S.INV(1-B4)				
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							



	A	B	C	D
1	**Professor's past test score distribution is normally distributed**			
2	Mean = μ	74		
3	SD = σ	10		
4	x	84		
5	Operator	<=		
6	P(x<=84)	0.841344746	=NORM.DIST(B4,B2,B3,1)	
7	z	1	=(B4-B2)/B3	
8	P(z<=1)	0.841344746	=NORM.S.DIST(B7,1)	



	A	B	C	D
22	x	P(x)	z	P(x<=84) = 0.8413
23	34	1.3383E-05	-4	1.3383E-05
24	35	1.98655E-05	-3.9	1.98655E-05
25	36	2.91947E-05	-3.8	2.91947E-05
26	37	4.2478E-05	-3.7	4.2478E-05

Chart Instructions:

1) Highlight P(x) column of chart heights and create area chart
2) Design Ribbon Tab, Select Data button, Add second column of P(x) with chart heights
3) Select second column of P(x) in chart, Ctrl + 1, Add to Secondary Axis
4) Design Ribbon Tab, Edit Series: For P(x<=x) add z values to Horizontal Axis, For P(x) add x values to Horizontal Axis.
5) Click Plus Button, Axis, check Secondary Horizontal Axis
6) Select Top Axis, ctrl 1 to open Task Pane, then go to Series, then in Labels, Specify interval as 10
7) Still in Labels, select Label position Low
8) Above Labels in Tick Marks, Major and Minor: None
9) Adjust size of chart area to allow x-z text box
10) Insert Ribbon Tab, Text Box

D22 = ="P("&A4&B5&B4&")"&" = "&ROUND(B6,4)
C23 = =NORM.DIST(A23,$B$2,$B$3,0)
D23 = =(A23-$B$2)/$B$3
E23 = =IF(A23<=$B$4,B23,"")