GE2262 Business Statistics
Topic 3 Discrete & Continuous Probability Distributions

Lecturer: Dr. Iris Yeung
Room : LAU-7239
Tel No.: 34428566
E-mail: msiris@cityu.edu.hk

Outline

- Discrete Probability Distribution
- Binomial Distribution
- Continuous Probability Distribution
- Normal Distribution

Reference

Levine, D.M., Krehbiel, T.C. and Berenson, M.L., *Business Statistics: A First Course*, Pearson Education Ltd, Chapter 5-6

Part 1

- Discrete Probability Distribution
- Binomial Distribution
- Continuous Probability Distribution
- Normal Distribution

Random Variable

- Recall random experiment is a process which results in ONE of a number of possible outcomes, and sample Space (*S*) is the set of all possible outcomes of an experiment
- A random variable is a rule that assigns a **number** to each outcome in sample space S.
  - Random variables are usually denoted by capital letters like ***X*** and ***Y***
    - Events are usually denoted by capital letters like A, B, C
  - The number values that are assigned to the random variables *X* and *Y* are usually denoted by small letters ***x*** and ***y***
- Random variables can be classified into:
  - Discrete random variable
  - Continuous random variable

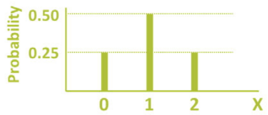
Discrete Random Variable and Discrete Probability Distribution

- Discrete random variable can take a countable number of values
- The probability distribution or probability mass function of a discrete random variable is the listing of the probability for each value of the random variable.
- The listing can be presented in the form of a table, chart and formula

EXPERIMENT	OUTCOME	RANDOM VARIABLE	VALUE OF RANDOM VARIABLES
Toss a fair coin two times	TT	*X = Number of heads*	X = 0
	TH		X = 1
	HT		X = 1
	HH		X = 2

Probability Distribution of X

X	Probability $P(X=x)$
0	1/4 = 0.25
1	2/4 = 0.50
2	1/4 = 0.25
Total	1



$$P(X=x)=\begin{cases}0.25 & \text{if } x=0,2\\ 0.50 & \text{if } x=1\end{cases}$$

Properties of Discrete Probability Distribution

1) $0 \le P(X=x) \le 1$
2) $\Sigma\, P(X=x) = 1$ where summation is over all possible values of *X* with non-zero probability

EXPERIMENT	OUTCOME	RANDOM VARIABLE	VALUE OF RANDOM VARIABLES
Toss a fair coin two times	TT	*X = Number of heads*	X = 0
	TH		X = 1
	HT		X = 1
	HH		X = 2

Probability Distribution of X

X	Probability $P(X=x)$
0	1/4 = 0.25
1	2/4 = 0.50
2	1/4 = 0.25
Total	1

Cumulative Distribution Function of Discrete Random Variable

- The cumulative distribution function of a discrete random variable *X* with probability distribution function P(X=*x*) is denoted by **F(*x*)** and defined for every number *x* by **F(*x*) = P(X≤*x*).** It gives the probability that the observed value of X will be less than or equal to *x*
- Example: Toss a fair coin two times, let X = number of heads

Probability Distribution of X

X	Probability $P(X=x)$	$F(x)$
0	1/4 = 0.25	0.25
1	2/4 = 0.50	0.75
2	1/4 = 0.25	1.00
Total	1	

- What is the probability of getting one or two heads?

P(1≤*X*≤2) = P(*X*=1) + P(*X*=2) = 0.5 + 0.25 = 0.75

Or P(1≤*X*≤2) =1-P(*X=0*) =1-0.25 = 0.75 (complement rule)

Or P(1≤X≤2) = P(X≤2) – P(X<1) = F(2) – F(0) = 1.0 – 0.25 = 0.75 (cumulative distribution function)

Mean and Variance of Discrete Random Variable

- The mean value (or expected value) of a discrete random variable *X* is a weighted average of all possible values of *X*
  - The weights are the probability associated with the value of the random variable X
  - $\mu = E(X) = \sum xP(X=x) = \sum xP(x)$
- The variance of a discrete random variable *X* is a weighted average of the squared deviation of *X* about the mean value

$$\sigma^2 = Var[X]$$
$$= E[(X-\mu)^2] = \sum (x-\mu)^2 P(X=x)$$
$$= E(X^2) - \mu^2 = \sum x^2 P(X=x) - \mu^2$$

- The standard deviation of a discrete random variable X is

$$\sigma = \sqrt{Var[X]}$$

Example : Toss Coin

- Let X = number of heads in tossing a fair coin two times
- The mean value of the number of heads:
- $\mu = E(X) = \sum xP(X = x)$
  $= (0)(0.25) + (1)(0.5) + (2)(0.25) = 1$
- The variance and the standard deviation of the number of heads:

$Var[X] = \sigma^2 =$
$= E[(X-\mu)^2] = \sum (x-\mu)^2 P(X=x)$
$= (0-1)^2(0.25) + (1-1)^2(0.5) + (2-1)^2(0.25) = 0.5$
or
$= E(X^2) - \mu^2 = \sum x^2 P(X=x) - \mu^2$
$= 0^2(0.25) + 1^2(0.5) + 2^2(0.25) - 1^2 = 1.5 - 1 = 0.5$
$\sigma = \sqrt{0.5} = 0.707$

Probability Distribution of X

X	Probability $P(X = x)$
0	1/4 = 0.25
1	2/4 = 0.50
2	1/4 = 0.25
Total	1

X	P(X=x)	X*P(x)	$(x-\mu)^2 P(x)$
0	0.25	0	0.25
1	0.50	0.5	0
2	0.25	0.5	0.25
Total	1.0	1.0	0.5
Mean = 1.0			
Variance = 0.5, standard deviation = 0.707			

Example : Investment Return

- The following table shows the return per $1,000 for Investment X and Y under different economic conditions. Which investment would you choose?

Economic Condition	P(X=x), P(Y=y)	Return for X	Return for Y
Recession	0.2	-$100	-$200
Stable Economy	0.5	+ 100	+ 50
Expanding Economy	0.3	+ 250	+ 350

$$E(X) = \mu_X = (-100)(.2) + (100)(.5) + (250)(.3) = \$105$$

$$\sigma_X^2 = (.2)(-100-105)^2 + (.5)(100-105)^2 + (.3)(250-105)^2$$
$$= 14,725 \qquad \sigma_X = 121.35$$

$$E(Y) = \mu_Y = (-200)(.2) + (50)(.5) + (350)(.3) = \$90$$

$$\sigma_Y^2 = (.2)(-200-90)^2 + (.5)(50-90)^2 + (.3)(350-90)^2$$
$$= 37,900 \qquad \sigma_Y = 194.68$$

Which investment has higher return? Which investment has higher variation?

Mean and Variance of Function of Random Variable

- For any two constants *a* and *b*
  - *E[aX+b]=aE[X]+b*
  - $Var[aX+b] = a^2 Var[X] = a^2\sigma^2$

Example: Magazine

- A small bookstore orders copies of a news magazine for its magazine rack each week. Let X be the weekly demand for the magazine, with the following probability distribution.

x	*1*	*2*	*3*	*4*	*5*	*6*
p(x)	*0.05*	*0.15*	*0.20*	*0.25*	*0.20*	*0.15*

- Suppose the store owner actually pays $25 for each copy of the news magazine and the selling price to customers is $40. The delivery charge of the magazines is $10 disregarding the number of copies ordered. Unsold copies will be returned to the publisher without any charge.
- Find the expected weekly revenue and the standard deviation.

Example: Magazine

X	P(x)	XP(x)	$(X-\mu)^2 P(x)$	X^2	X^2 P(X)
1	0.05	0.05	0.406125	1	0.05
2	0.15	0.3	0.513375	4	0.6
3	0.2	0.6	0.1445	9	1.8
4	0.25	1	0.005625	16	4
5	0.2	1	0.2645	25	5
6	0.15	0.9	0.693375	36	5.4
Total	1	3.85	2.0275	91	16.85

Solution:

$E[X] = \Sigma x P(x) = 3.85$

$E[X^2] = \Sigma x^2 P(x) = 16.85$

$Var[X] = E[X^2] - (E[X])^2 = 2.0275$

Let g(X) = revenue earned

$g(X) = (40-25)X - 10 = 15X-10$

$E[g(X)] = E[15X - 10] = 15E(X)-10$

$= 15*3.85 - 10 = \$47.75$

$Var[g(X)] = Var[15X - 10] = 15^2 Var[X]$

$= 225*2.0275=456.1875$

$SD[g(X)] = \$21.359$

For any two constants a and b

$E[aX+b]=aE[X]+b$

$Var[aX+b] = a^2 Var[X]= a^2\sigma^2$

Part 2

- Discrete Probability Distribution
- Binomial Distribution
- Continuous Probability Distribution
- Normal Distribution

Bernoulli / Binomial Experiment

- The experiment is repeated *n* times (*n* trials).
- Each trial has only two possible outcomes (denoted as success S and failure F).
- The probability of success, denoted by *p*, is the same for each trial.
  - The probability of failure for each trial is equal to *q=1-p*.
- The trials are independent (the outcome of a trial does not depend on the outcomes of previous trials).
- We are interested in the random variable *X* where X is the number of successes observed in *n* trials. Note the possible values of X are *0,1,2,…,n*.
- Example of a Bernoulli experiment : Toss a fair coin 4 times
  - number of trials *n* = 4
  - Each trial has two possible outcomes, may denote "head" as success, "tail" as failure
  - P(success) =P(head)=0.5, P(failure) = P(tail)= 0.5 for all 4 trials
  - Each toss is independent of the other
  - We are interested in X, the number of heads in 4 tosses

Binomial Distribution

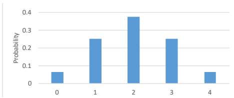
- Let *X* be the number of successes in *n* Bernoulli trials. The probability distribution of *X* is given by:

$$p(X=x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad x = 0,1,\cdots,n$$

$$Note: n! = n(n-1)(n-2)\cdots 1; \quad 0! = 1$$

- *X* is said to be a binomial random variable and has a binomial distribution with parameters *n* and *p*. We write *X~BIN(n,p)*.
- *n trials, X successes, (n-X) failures* SFSFSSSFSSF
- *Probability of* getting exactly *x* successes and (*n* – *x*) failures for a particular sequence= $P(\underbrace{SS\ldots S}_{x}\ \underbrace{FF\ldots F}_{n-x}) = p^x q^{n-x}$
- There are ${}_nC_x$ ways of getting *x* successes out of *n* trials.
- The probability of getting exactly *x* successes out of *n* trials is
  - $P(x) = {}_nC_x p^x q^{n-x}$ where $x = 0, 1, 2, \ldots, n$.
- The probability distribution of *X* is called the Binomial Distribution

Example : Toss Coin



- Let X = number of heads in 4 tosses
- n=4, P(success) =P(H)=0.5, P(failure) = P(T)= 0.5 in each trial

$$P(X=0) = {}_4C_0\, p^0 q^{4-0} = \frac{4!}{0!4!}(0.5)^0(0.5)^4 = 0.5^4 = 0.0625$$

$$P(X=1) = {}_4C_1\, p^1 q^{4-1} = \frac{4!}{1!3!}(0.5)(0.5)^3 = 4(0.5)^4 = 0.25$$

$$P(X=2) = {}_4C_2\, p^2 q^{4-2} = \frac{4!}{2!2!}(0.5)^2(0.5)^2 = 6(0.5)^4 = 0.375$$

$$P(X=3) = {}_4C_3\, p^3 q^{4-3} = \frac{4!}{3!1!}(0.5)^3(0.5) = 4(0.5)^4 = 0.25$$

$$P(X=4) = {}_4C_4\, p^4 q^{4-4} = \frac{4!}{4!0!}(0.5)^4(0.5)^0 = 0.5^4 = 0.0625$$

$$P(X \ge 3) = P(X=3) + P(X=4) = 0.25 + 0.0625 = 0.3125$$

$$P(X < 3) = P(X=0) + P(X=1) + P(X=2)$$

$$= 0.0625 + 0.25 + 0.375 = 0.6875$$

or

$$P(X < 3) = 1 - P(X \ge 3) = 1 - 0.3125 = 0.6875$$

$$P(2 \le X \le 3) = P(X=2) + P(X=3) = 0.375 + 0.25 = 0.625$$

Mean and Variance of a Binomial Variable

If $X \sim BIN(n,p)$, then the mean and variance of X are given by

Mean $\mu = E[X] = np$

Variance $\sigma^2 = Var[X] = np(1\text{-}p)$

- Example: Toss a fair coin 4 times. What is the mean and variance of the number of heads in 4 tosses?

Mean $\mu = E[X] = np = 4*0.5 = 2$

Variance $\sigma^2 = Var[X] = np(1\text{-}p) = 4*0.5*0.5 = 1$

Standard deviation $\sigma = \sqrt{Var[X]} = 1$

Example: Invoice Payment

Suppose the probability of an invoice payment being late is 0.10. What is the probability of having 3 late invoice payments in a group of 5 invoices?

X = no. of late invoice payment out of 5 invoices
X follows Binomial distribution ($n = 5, p = 0.1$)

The probability of having 3 late invoice payments in a group of 5 invoices:

$$P(X=3) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{(n-x)}$$

$$= \frac{5!}{3!(5-3)!} 0.1^3 (1-0.1)^{(5-3)}$$

$$= 0.0081$$

Example: Invoice Payment

- What is the probability that there are 3 or more late invoice payments?

$$P(X \ge 3) = P(X=3) + P(X=4) + P(X=5)$$
$$= 0.0081 + 0.00045 + 0.00001$$
$$= 0.00856$$

- What is the probability that there are less than 3 late invoice payments?

$$P(X < 3) = P(X=0) + P(X=1) + P(X=2)$$
$$= 1 - P(X \ge 3)$$
$$= 1 - 0.00856 = 0.99144$$

Example: Invoice Payment

What is the mean and variance of the number of late invoice payments?

Binomial Probability Distribution:

x_i	$P(X=x_i)$
0	0.59049
1	0.32805
2	0.0729
3	0.0081
4	0.00045
5	0.00001

$$\mu = \sum xP(X=x)$$
$$= (0)(0.59049) + (1)(0.32805) + (2)(0.0729)$$
$$+ (3)(0.0081) + (4)(0.00045) + (5)(0.00001)$$
$$= 0.5$$

$$\sigma^2 = \sum (x-\mu)^2 P(X=x)$$
$$= (0-0.5)^2 (0.59049) + (1-0.5)^2(0.32805) + (2-0.5)^2$$
$$(0.0729) + (3-0.5)^2(0.0081) + (4-0.5)^2(0.00045)$$
$$+ (5-0.5)^2(0.00001)$$
$$= 0.45$$

$$\rightarrow \sigma = 0.6708$$

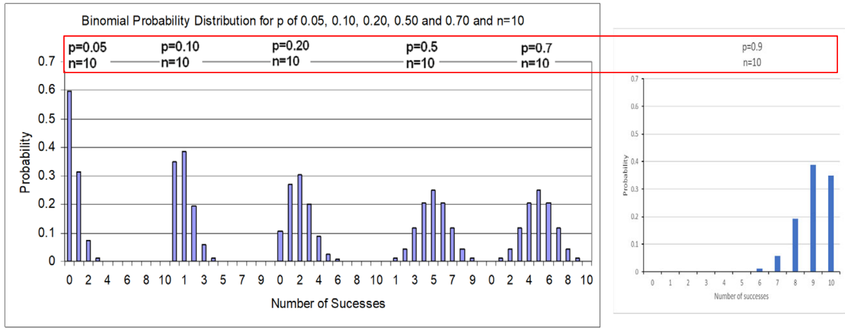
$\mu = np = 5(0.1) = 0.5$

$\sigma^2 = np(1-p)$
$= 5(0.1)(0.9) = 0.45$

$\sigma = \sqrt{np(1-p)}$=0.6708

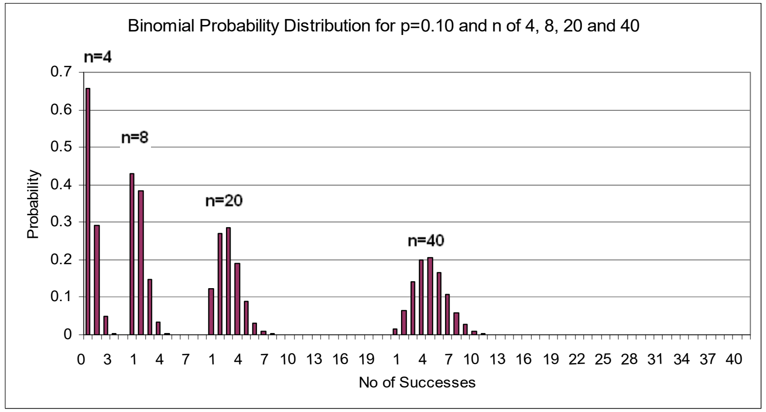
Shape of Binomial Distribution

- The shape of a binomial distribution depends on the values of n and p
- Whenever $p = 0.5$, the distribution is symmetrical, regardless of how large or small the value of n
  - Whenever $p \neq 0.5$, the distribution is skewed
    - $p < 0.5$, right-skewed; $p > 0.5$, left-skewed



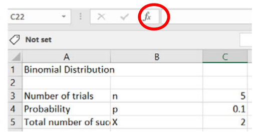
Shape of Binomial Distribution

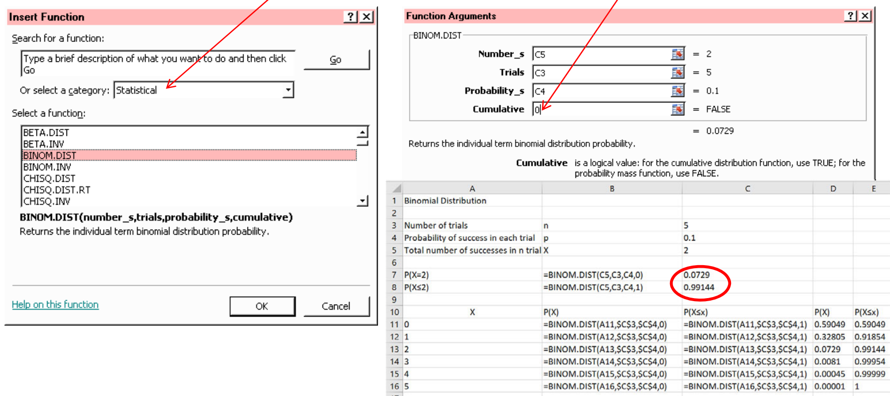
- The shape of a binomial distribution depends on the values of n and p
- Whenever $p \neq 0.5$, the distribution is skewed. If n increases, the distribution becomes more and more symmetrical.



Binomial Distribution in Excel

- Invoice payment example: n=5,p=0.1, X= number of late invoice payments. Find
  - P(X=2).
  - P(X≤2)
  - Probability distribution of X
- Click *fx* in the Menu bar and select the Statistical category and the BINOM.DIST Function name. Click OK.
- Complete the BINOM.DIST dialog box as shown below. To find P(X=2),we type 0 or FALSE in cumulative box. To find P(X≤2),we type 1 or TRUE in cumulative box.





	A	B	C	D	E
1	Binomial Distribution				
2					
3	Number of trials	n	5		
4	Probability of success in each trial	p	0.1		
5	Total number of successes in n trial	X	2		
6					
7	P(X=2)	=BINOM.DIST(C5,C3,C4,0)	0.0729		
8	P(X≤2)	=BINOM.DIST(C5,C3,C4,1)	0.99144		
9					
10	X	P(X)	P(X≤x)	P(X)	P(X≤x)
11	0	=BINOM.DIST(A11,C3,C4,0)	=BINOM.DIST(A11,C3,C4,1)	0.59049	0.59049
12	1	=BINOM.DIST(A12,C3,C4,0)	=BINOM.DIST(A12,C3,C4,1)	0.32805	0.91854
13	2	=BINOM.DIST(A13,C3,C4,0)	=BINOM.DIST(A13,C3,C4,1)	0.0729	0.99144
14	3	=BINOM.DIST(A14,C3,C4,0)	=BINOM.DIST(A14,C3,C4,1)	0.0081	0.99954
15	4	=BINOM.DIST(A15,C3,C4,0)	=BINOM.DIST(A15,C3,C4,1)	0.00045	0.99999
16	5	=BINOM.DIST(A16,C3,C4,0)	=BINOM.DIST(A16,C3,C4,1)	0.00001	1

Part 3

- Discrete Probability Distribution
- Binomial Distribution
- Continuous Probability Distribution
- Normal Distribution

Continuous Random Variable and Continuous Probability Distribution

- A continuous random variable can assume an infinite number of values corresponding to the points on a line interval.
  - It can potentially take on any value depending only on the ability to precisely and accurately measure
  - In practice, a discrete numerical variable with large number of values is often considered as a continuous variable
- Examples
  - Height of students in cm
  - Survival time of pigs in days
  - Customer service time in a bank in mins
- The probability distribution of a continuous random variable X is also called probability density function *f(x)*. The graph is called density curve.
- Continuous probability distribution plays a major role in statistics
  - most quantitative variables are measured on a continuous scale.
  - often provide very good approximations for discrete random variables.

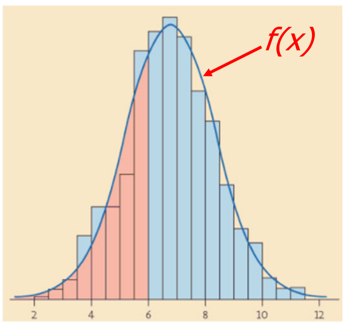
Properties of Continuous Probability Distribution

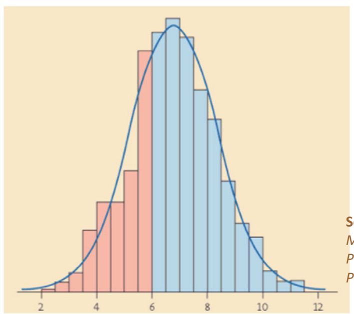
$$f(x) \geq 0 \qquad \text{for all } x$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Properties of Discrete Probability Distribution
1) $0 \leq P(X=x) \leq 1$
2) $\Sigma P(X=x) = 1$ where summation is over all possible values of *X* with non-zero probability

- The probability density function is always on or above the horizontal axis
- It has area exactly 1 underneath the density curve





Source
Moore D S, The Basic Practice of Statistics, Palgrave Macmillan

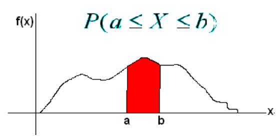
Computing Probabilities of Continuous Probability Distribution

- For a continuous random variable *X*, the probability that *X* takes any particular value is zero

  $P(X=x)=0$ for all x

- This implies that, for any two numbers a and b with $a<b$,

  $P(a \leq X \leq b) = P(a<X\leq b)$
  $= P(a \leq X<b)$
  $= P(a<X<b)$
- For any two numbers a and b with $a<b$, the probability that the random variable *X* takes on a value in the interval $[a,b]$ is the area under the graph of the density function.

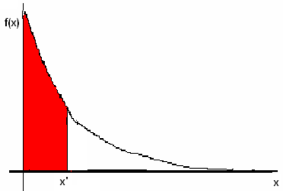


$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Cumulative Distribution Function of Continuous Random Variable

- The cumulative distribution function of a continuous random variable X with probability density function $f(x)$ is denoted by $F(x)$ and defined for every x by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy \quad \text{for } -\infty < x < \infty$$



Mean and Variance of Continuous Random Variable

- The mean value (or expected value) of a continuous random variable X is

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

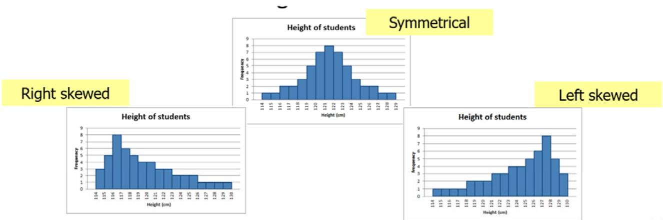
- The variance of a continuous random variable X is

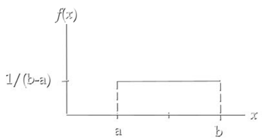
$$\sigma^2 = Var[X]$$
$$= E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$
$$= E[X^2] - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

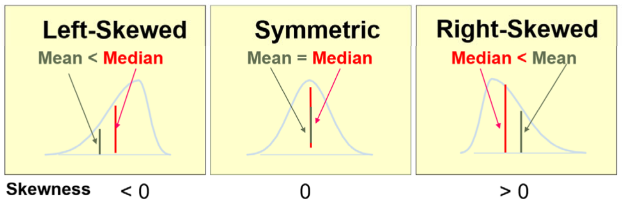
- The standard deviation of a continuous random variable X is

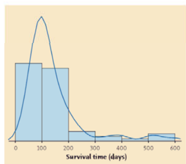
$$\sigma = \sqrt{Var[X]}$$

Examples of Density Curves









Part 4

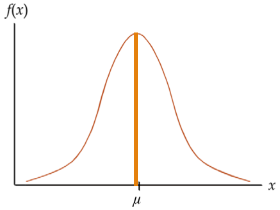
- Discrete Probability Distribution
- Binomial Distribution
- Continuous Probability Distribution
- Normal Distribution

Normal Distribution

- If a continuous random variable X has the following density function,

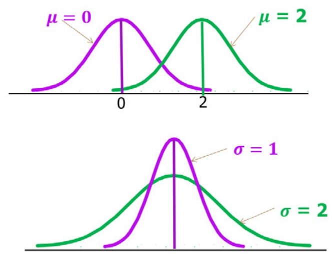
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{1}{2}\right)\left[\frac{x-\mu}{\sigma}\right]^2}$$

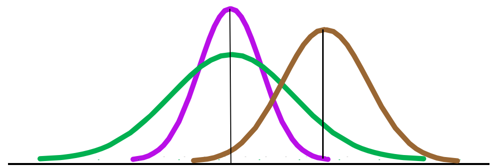
where x = any value in the range of $-\infty$ to $+\infty$
μ = mean
σ = standard deviation
e = 2.71828
π = 3.14159



- *X* is said to be a normal random variable and has a normal distribution with mean μ and variance σ^2. We write $X \sim N(\mu, \sigma^2)$
- Shape of the normal distribution
  - Symmetrical about the mean and bell shaped
  - The highest point on the normal curve is at the mean, which is also the median and mode.
  - The total area under the curve is 1
    - .5 to the left of the mean and .5 to the right

Normal Distribution with Different Means and Standard Deviations





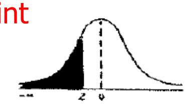
When μ = 0 and σ = 1, the normal random variable *X* is called a standard normal random variable (denoted Z) and the normal distribution is called a standard (standardized) normal distribution. We write $Z \sim N(0,1)$ where $-\infty < Z < +\infty$

The Standard Normal Table

This standard table gives P(-∞ ≤*Z*≤*z*)

The column gives the value of Z to the second decimal point

The Cumulative Standardized Normal Distribution
Entry represents area under the cumulative standardized normal distribution from -∞ to Z



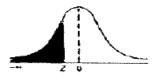
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-6.0	0.000000001									
-5.5	0.000000019									
-5.0	0.000000287									
-4.5	0.000003398									
-4.0	0.000031671									
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024

The row shows the value of Z to the first decimal point

The value within the table gives the probability from $Z = -\infty$ up to the desired Z value
$P(Z < -3.45) = 0.00028$

Example: Find Normal Probabilities

The Cumulative Standardized Normal Distribution
Entry represents area under the cumulative standardized normal distribution from -∞ to Z



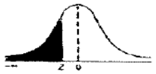
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2388	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2482	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

- *P(Z≤-0.75)=0.2266*
- *P(Z≤1.25) = 0.8944*
- *P(Z≥0.59)=1-P(Z<0.59)=1-0.7224=0.2776*
- *P(-0.36≤Z≤1.24)*
  =P(Z ≤1.24)-P(Z<-0.36)
  =0.8925-0.3594=0.5331
- *P(|Z|≥0.82) =P(Z≤-0.82) + P(Z≥0.82)*
  *=0.2061*2=0.4122*

Example: Find z value from Known Probabilities

The Cumulative Standardized Normal Distribution
Entry represents area under the cumulative standardized normal distribution from -∞ to Z



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

- Find *z* such that $P(Z>z)=0.025$
  - $P(Z\le z)=0.975 \Rightarrow z=1.96$
- Find *z* such that $P(-z<Z<z)=0.9$
  - $P(-z<Z<z)=0.9 \Rightarrow P(Z\le z)=0.95 \Rightarrow z=1.645$

Compute Normal Probabilities

- If X is a normal random variable $X \sim N(\mu, \sigma^2)$, it can be standardized to a standard normal variable $Z \sim N(0,1)$ by $Z = \frac{X-\mu}{\sigma}$
- Then we can use standard normal table to find normal probabilities for all normal distributions

$$P(a \le X \le b) = P(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma})$$

$$= P(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma})$$

Example

$$X \sim N(25, 0.5^2)$$

$$P(X \le 24) = P(\frac{X-25}{0.5} \le \frac{24-25}{0.5}) = P(Z \le -2) = 0.0228$$

$$P(24 \le X \le 25.5) = P(\frac{24-25}{0.5} \le \frac{X-25}{0.5} \le \frac{25.5-25}{0.5})$$

$$= P(-2 \le Z \le 1) = P(Z \le 1) - P(<-2) = 0.8413 - 0.0228 = 0.8185$$

The Cumulative Standardiz
Entry represents area unde
distribution from -∞ to Z

Z	0.00	0.01
-3.0	0.00135	0.00131
-2.9	0.0019	0.0018
-2.8	0.0026	0.0025
-2.7	0.0035	0.0034
-2.6	0.0047	0.0045
-2.5	0.0062	0.0060
-2.4	0.0082	0.0080
-2.3	0.0107	0.0104
-2.2	0.0139	0.0136
-2.1	0.0179	0.0174
-2.0	0.0228	0.0222
-1.9	0.0287	0.0281
-1.8	0.0359	0.0351
-1.7	0.0446	0.0436
-1.6	0.0548	0.0537
-1.5	0.0668	0.0655
-1.4	0.0808	0.0793
-1.3	0.0968	0.0951
-1.2	0.1151	0.1131
-1.1	0.1357	0.1335
-1.0	0.1587	0.1562

The Cumulative Standardiz
Entry represents area unde
distribution from -∞ to Z

Z	0.00	0.01
1.0	0.8413	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207
1.5	0.9332	0.9345
1.6	0.9452	0.9463
1.7	0.9554	0.9564
1.8	0.9641	0.9649
1.9	0.9713	0.9719
2.0	0.9772	0.9778
2.1	0.9821	0.9826
2.2	0.9861	0.9864
2.3	0.9893	0.9896
2.4	0.9918	0.9920
2.5	0.9938	0.9940
2.6	0.9953	0.9955
2.7	0.9965	0.9966
2.8	0.9974	0.9975
2.9	0.9981	0.9982
3.0	0.99865	0.99869

Compute Normal Probabilities

The Cumulative Standardiz
Entry represents area unde
distribution from -∞ to Z

Z	0.00	0.01
-3.0	~~0.00135~~	0.00131
-2.9	0.0019	0.0018
-2.8	0.0026	0.0025
-2.7	0.0035	0.0034
-2.6	0.0047	0.0045
-2.5	0.0062	0.0060
-2.4	0.0082	0.0080
-2.3	0.0107	0.0104
-2.2	0.0139	0.0136
-2.1	0.0179	0.0174
-2.0	~~0.0228~~	0.0222
-1.9	0.0287	0.0281
-1.8	0.0359	0.0351
-1.7	0.0446	0.0436
-1.6	0.0548	0.0537
-1.5	0.0668	0.0655
-1.4	0.0808	0.0793
-1.3	0.0968	0.0951
-1.2	0.1151	0.1131
-1.1	0.1357	0.1335
-1.0	~~0.1587~~	0.1562

The Cumulative Standardiz
Entry represents area unde
distribution from -∞ to Z

Z	0.00	0.01
1.0	~~0.8413~~	0.8438
1.1	0.8643	0.8665
1.2	0.8849	0.8869
1.3	0.9032	0.9049
1.4	0.9192	0.9207
1.5	0.9332	0.9345
1.6	0.9452	0.9463
1.7	0.9554	0.9564
1.8	0.9641	0.9649
1.9	0.9713	0.9719
2.0	~~0.9772~~	0.9778
2.1	0.9821	0.9826
2.2	0.9861	0.9864
2.3	0.9893	0.9896
2.4	0.9918	0.9920
2.5	0.9938	0.9940
2.6	0.9953	0.9955
2.7	0.9965	0.9966
2.8	0.9974	0.9975
2.9	0.9981	0.9982
3.0	~~0.99865~~	0.99869

$$P(\mu-\sigma \le X \le \mu+\sigma)$$

$$= P(\frac{\mu-\sigma-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\mu+\sigma-\mu}{\sigma})$$

$$= P(-1 \le Z \le 1) = P(Z \le 1) - P(Z < -1)$$

$$= 0.8413 - 0.1587 = 0.6826$$

$$P(\mu-2\sigma \le X \le \mu+2\sigma)$$

$$= P(\frac{\mu-2\sigma-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\mu+2\sigma-\mu}{\sigma})$$

$$= P(-2 \le Z \le 2) = P(Z \le 2) - P(Z < -2)$$

$$= 0.9772 - 0.0228 = 0.9544$$

$$P(\mu-3\sigma \le X \le \mu+3\sigma)$$

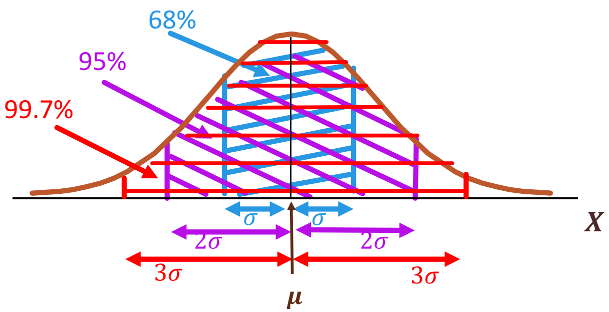
$$= P(\frac{\mu-3\sigma-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\mu+3\sigma-\mu}{\sigma})$$

$$= P(-3 \le Z \le 3) = P(Z \le 3) - P(Z < -3)$$

$$= 0.99865 - 0.00135 = 0.9973$$

The Empirical Rule

- The Empirical Rule said that
  - Area within $\mu \pm \sigma$ equals 68% approximately
  - Area within $\mu \pm 2\sigma$ equals 95% approximately
  - Area within $\mu \pm 3\sigma$ equals 99.7% approximately

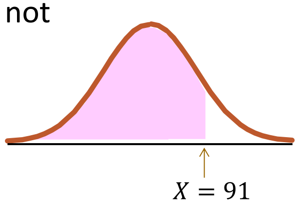


Example: Student Scores

- A set of final exam scores was normally distributed with a population mean 73 and population standard deviation 8

2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890

1. What is the probability of getting a score not higher than 91 on this exam?



Let the score be X, and $X \sim N(73, 8^2)$

$$P(X \le 91)$$
$$= P\left(\frac{X-73}{8} \le \frac{91-73}{8}\right) = P(Z \le 2.25) = 0.9878$$

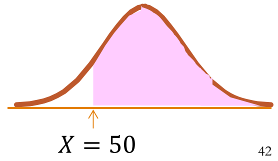
Example: Student Scores

$X \sim N(73, 8^2)$

2. If the passing score is 50, what is the chance that a student can pass the exam?

$$P(X \ge 50) = P\left(\frac{X-73}{8} \ge \frac{50-73}{8}\right) = P(Z \ge -2.875)$$
$$= 1 - P(Z < -2.875) \cong 1 - 0.00205 = 0.99795$$

-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026

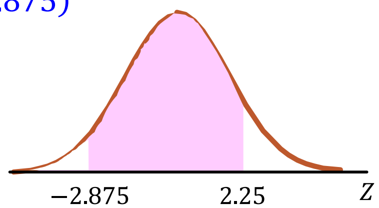


Example: Student Scores

$X \sim N(73, 8^2)$

3. What percentage of students scored between 50 and 91?

$$P(50 \le X \le 91)$$
$$= P(X \le 91) - P(X < 50)$$
$$= P\left(\frac{X-73}{8} \le \frac{91-73}{8}\right) - P\left(\frac{X-73}{8} < \frac{50-73}{8}\right)$$
$$= P(Z \le 2.25) - P(Z < -2.875)$$
$$= 0.9878 - 0.00205$$
$$= 0.98575$$

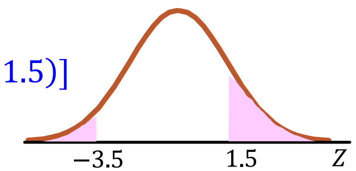


Example: Student Scores

4. What percentage of students scored below 45 or above 85?

$X \sim N(73, 8^2)$

$$P(X < 45) + P(X > 85)$$
$$= P\left(\frac{X-73}{8} < \frac{45-73}{8}\right) + P\left(\frac{X-73}{8} > \frac{85-73}{8}\right)$$
$$= P(Z < -3.5) + P(Z > 1.5)$$
$$= P(Z < -3.5) + [1 - P(Z \le 1.5)]$$
$$= 0.00023 + [1 - 0.9332]$$
$$= 0.06703$$



-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024

1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

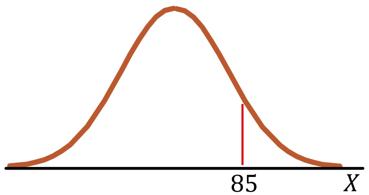
Example: Student Scores

5. What is the probability for a student to score exactly 85?

$P(X = 85)$
$= 0$

Not an area,

but just a line!!!



Example: Student Scores

6. What is the minimum score a student needs in order to be in the top 5% of the class?

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

$P(X \geq a) = 0.05$

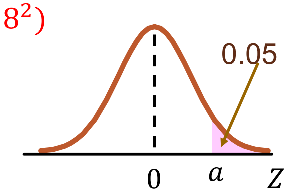
$P\left(\frac{X-73}{8} \geq \frac{a-73}{8}\right)$=0.05

$P\left(Z \geq \frac{a-73}{8}\right)$=0.05

$\frac{a-73}{8}$=1.645

$\Rightarrow$ a $= 73 + 1.645 \times 8$
$= 86.16$

$X \sim N(73, 8^2)$



Example: Student Scores

$X \sim N(73, 8^2)$

7. The middle 50% of the students scored between what two scores?

$P(a \leq X \leq b) = 0.5$

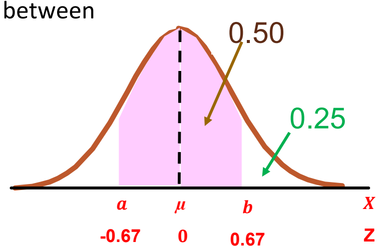
$P\left(\frac{a-73}{8} \leq \frac{X-73}{8} \leq \frac{b-73}{8}\right)$=0.5

$P\left(\frac{a-73}{8} \leq Z \leq \frac{b-73}{8}\right)$=0.5

$P(Z \leq 0.67) = 0.75$

$\frac{a-73}{8} = -0.67 \Rightarrow a = 73 + (-0.67) \times 8 = 67.64$

$\frac{b-73}{8} = 0.67 \Rightarrow b = 73 + (0.67) \times 8 = 78.36$



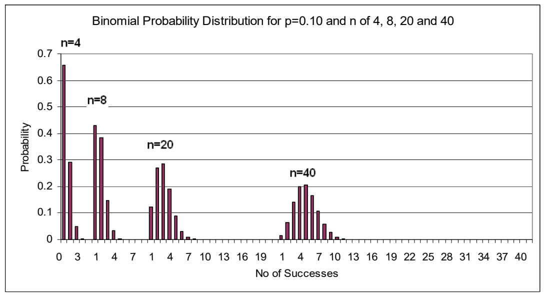
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549

Normal Approximation for the Binomial Distribution

- Consider a binomial random variable X ~ Bin (n, p). If *n* is large, it may be cumbersome to compute $P(a \leq X \leq b)$
- Recall binomial distribution becomes more and more symmetrical when *n* increases or *p* is close to 0.5.
- If *np≥5 and nq≥5*, we can approximate

a binomial distribution X ~ Bin (n, p)

by a normal distribution $X \sim N(\mu, \sigma^2)$ with $\mu=np$, $\sigma^2=npq$



Normal Approximation for the Binomial Distribution

X ~ Bin (50, 0.25)

np=50*0.25 =12.5 >5

nq=50*0.75 >5

X ~ app. N(np, npq)

X ~ app. N(12.5, 9.375)

$$P(X \le 10) \approx P(X \le 10.5)$$

$$= P\left(\frac{X-\mu}{\sigma} \le \frac{10.5-12.5}{3.06}\right)$$

$$= P(Z \le -0.65) = 0.2578$$

$$P(5 \le X \le 15) \approx P(4.5 \le X \le 15.5)$$

$$= P\left(\frac{4.5-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{15.5-\mu}{\sigma}\right)$$

$$= P\left(\frac{4.5-12.5}{3.06} \le Z \le \frac{15.5-12.5}{3.06}\right)$$

$$= P(-2.61 \le Z \le 0.98)$$

$$= P(Z \le 0.98) - P(Z < -2.61)$$

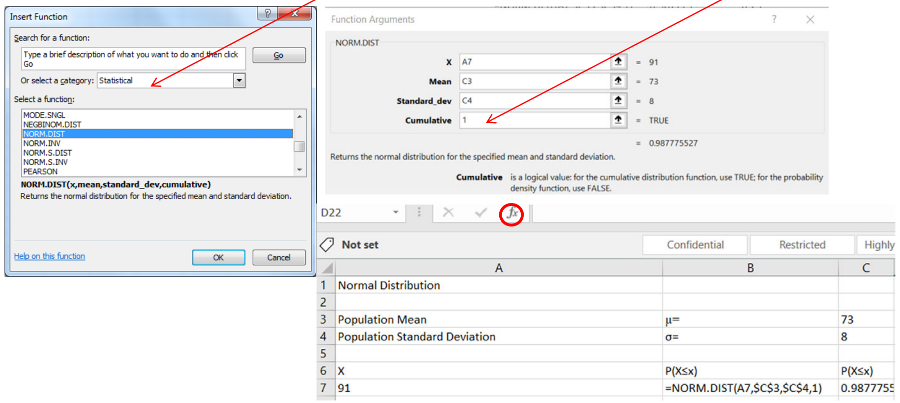
$$= 0.8365 - (0.0045)$$

$$= 0.832$$

To use a continuous distribution to approximate a discrete distribution, we add or subtract 0.5 to a discrete value (called continuity correction) for better approximation

Find Normal Probability in Excel

- Student score example: $X \sim N(73, 8^2)$. Find
  - P(X ≤ 91)
  - Find z and a such that P(Z≤ z)=0.25, P(X≤ a)=0.25
- Click *fx* in the Menu Bar and select the Statistical category and the NORM.DIST Function name. Click OK.
- Complete the NORM.DIST dialog box as shown below. To find P(X ≤ 91), we type 1 or TRUE in cumulative box.



Find Z and X Values from Known Probabilities in Excel

- Student score example: $X \sim N(73, 8^2)$. Find
  - P(X ≤ 91)
  - Find z and a such that P(Z≤ z)=0.25, P(X≤ a)=0.25
- Click *fx* in the Menu Bar and select the Statistical Function category. Click the NORM.S.INV Function name to find Z value and NORM.INV Function name to find X value. Click OK.
- Complete the NORM.S.INV and NORM.INV dialog box as shown below.