Module 3 :
Random Variables

Basic Probability

- Random Variable Definitions
- Mean, Variance, and Standard Deviation of a Random Variable
- Mean, Variance, and Standard Deviation for Sum of Random Variables
- Binomial Random Variable
- Poisson Random Variable
- Normal Random Variable
- Central Limit Theorem
- Z Scores

Random Variable Definitions

- a set of possible values from a random experiment
- Example:

Flipping a coin

Rolling a dice etc.

Discrete & Continuous

- Discrete: can take on only certain values within a range

  example: gender of random person, H(0) or T(1) when flipped coin
- Continuous: can take on any value in a range of values

  example: height or random person

Discrete & Continuous (Example)

Variables	D / C ?
Number of aces drawn if 5 cards are drawn from a deck of cards	
Number of games won by Manchester United during 2018 Premier League Season	
Market share during its first year for a new statin drug	
Time you wait in line at ICA security	

Discrete & Continuous (Example)

Variables	D / C ?
Number of aces drawn if 5 cards are drawn from a deck of cards	D
Number of games won by Manchester United during 2018 Premier League Season	D
Market share during its first year for a new statin drug	C
Time you wait in line at ICA security	C

Random Variables Independent/NOT

Variables	I / N ?
Price charged for a candy bar and number of candy bars sold	N
Return on Microsoft and RedHat stock during 2018	N
Games won by Manchester United and the Golden State Warriors in 2018	I
2018 return on the Dow Index and your child's score on an IQ test	I

Mean, Variance, and Standard Deviation of a Random Variable

- Mean: Average or Expected Value

  =AVERAGE()
- Median: value that divide data/sample into 2

  =MEDIAN()
- Variance: average squared deviation of values from mean

  =VAR.P()
- Standard Deviation: square root of the variance

  =STDEV.P()

Mean, Variance, and Standard Deviation of a Random Variable

- The mean of the sum of two random variables, X and Y, is simply the sum of their means
- Mean = Expected Value

look at sample in Excel

Binomial Probability Distribution

- The experiment consists of **n** repeated trials.
- Each trial can result in just two possible (**Binomial**) outcomes. We call one of these outcomes a **success** and the other, a **failure**.
- The probability of success, denoted by P, is the **same on every trial**.
- The trials are **independent**; that is, the outcome on one trial does not affect the outcome on other trials.

Binomial Random Variable

- a specific type of discrete random variable that counts how often a particular event occurs in a fixed number of trials
- Requirements:

1. Fixed number of trials (n)
2. Fixed sample size
3. Trials are independent of one another

- Example:

1. Tossing a coin 6 times
2. Rolling 2 dices 5 times, etc.

Binomial Random Variable

- BINOM.DIST.RANGE(trials, probability_s, number_s, [number_s])
- Sample: Home Work

  M3L4HW1

  M3L4HW2

  M3L4HW3

Binomial Random Variable

M3L4HW1

Suppose that, on average, 4% of all CD drives received by a computer company are defective. The company has adopted the following policy: Sample 50 CD drives in each shipment, and accept the shipment if none are defective. Using this information, determine the following:

What percentage of shipments will be accepted?

Binomial Random Variable

M3L4HW1

- What percentage of shipments will be accepted?

13%
=BINOM.DIST.RANGE(50,0.04,0)

- Suppose the policy changes so that a shipment is accepted if up to one CD drive in the sample is defective. What percentage of shipments will be accepted?

40%
=BINOM.DIST.RANGE(50,0.04,0,1)

Binomial Random Variable

M3L4HW1

- What is the probability that a sample size of 50 will contain at least 10 (i.e., 10 or more) defective CD drives?

0.0025%
=BINOM.DIST.RANGE(50,0.04,10,50)

Binomial Random Variable

M3L4HW2

Airline overbooking data tells us that 95% of booked passengers show up for their flight. On a flight with 100 tickets available, determine how the probability of overbooking varies as the number of tickets sold varies from 100 through 115.

What is the probability of overbooking when 101 tickets are sold?

What is the probability of overbooking when 105 tickets are sold?

What is the probability of overbooking when 112 tickets are sold?

Binomial Random Variable

M3L4HW2

What is the probability of overbooking when 101 tickets are sold?

What is the probability of overbooking when 105 tickets are sold?

What is the probability of overbooking when 112 tickets are sold?

1%	=BINOM.DIST.RANGE(101,0.95,101,101)
39%	=BINOM.DIST.RANGE(105,0.95,101,105)
99%	=BINOM.DIST.RANGE(112,0.95,101,112)

Binomial Random Variable

M3L4HW3

Suppose that during any given year, a given mutual fund has a 50% chance of beating the Standard and Poor's 500 Stock Index.

What is the probability that one fund will beat the Standard and Poor's 500 Stock Index during at least 8 out of 10 years?

In a group of 100 mutual funds, what is the probability that at least 10 funds will beat the Standard and Poor's 500 Stock Index during at least 8 out of 10 years?

Binomial Random Variable

M3L4HW3

5.47% =BINOM.DIST.RANGE(10,0.5,8,10)

4.72% =BINOM.DIST.RANGE(100,B72,10,100)

Binomial Random Variable

The binomial distribution formula is:

$$b(x; n, P) = {}_nC_x * P^x * (1 - P)^{n-x}$$

Where:

b = binomial probability

x = total number of "successes" (pass or fail, heads or tails etc.)

P = probability of a success on an individual trial

n = number of trials

Sample binomial probability problem

- **Exercise 4.5 answer**
- Ten percent of the global population is left-handed.
  You conduct a survey in which you stop three people at random.
- What is the probability that two of the people you stop are left-handed, and the other is not?
- ANSWER: 0.027

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

Sample binomial probability problem

- **Exercise 4.5 answer**
- The sweet factory that produces your favourite candy fills the bags randomly, with a 40% probability that any single bag will contain a Sour Plum.
- You buy five bags of candy, what is probability 3 bags contains sour plum)
- ANSWER: 23.04%

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

Binomial Random Variable

$$P(X) = \frac{n!}{(n-X)!\,X!} \cdot (p)^X \cdot (q)^{n-X}$$

Poisson Random Variable

- a discrete random variable that counts the number of times a certain event occurs in a specific interval
- Is Binomial
- Example of usage:

1. to estimate the probability of the number of deaths by horse-kicking in the Prussian army in a given year
2. To count the number of car accidents at a busy intersection between 6 and 9 AM
3. To count The number of typing errors on a page
4. To count the defect rate of a machine in one day or a month

Poisson Random Variable

- Requirements:

1. the event is something that can be counted in whole numbers (NOT in Fraction)
2. The occurrences are independent
3. The average frequency of occurrences for a time period in question is known

Poisson Random Variable in Excel

- POISSON.DIST(x, mean, cumulative)
- Example: M3L5HW1

A math model predicts that on average a soccer team should score 2.1 goals in today's game. What is the chance they score >=2 goals?

Answer: PTO

Poisson Random Variable

- Example: M3L5HW1

A math model predicts that on average a soccer team should score 2.1 goals in today's game. What is the chance they score >=2 goals?

Answer: =1-POISSON.DIST(1,2.1,TRUE) = 62%

Poisson Random Variable

- Example: M3L5HW2

A book has an average of 1 misprint per 20 pages. What is the chance that a 200 page book has $<=5$ misprints?

Answer: PTO

Poisson Random Variable

- Example: M3L5HW2

A book has an average of 1 misprint per 20 pages. What is the chance that a 200 page book has <=5 misprints?

Answer: =POISSON.DIST(5,10,TRUE) = 7%

Normal Random Variable

- Measurements collected of continuous variables, such as height, weight, temperature, distance traveled, are considered continuous random variables because these variables can take on any value across a wide, possibly infinite range, and can vary randomly
- The distribution is bell-shaped or normal.

Normal Random Variable

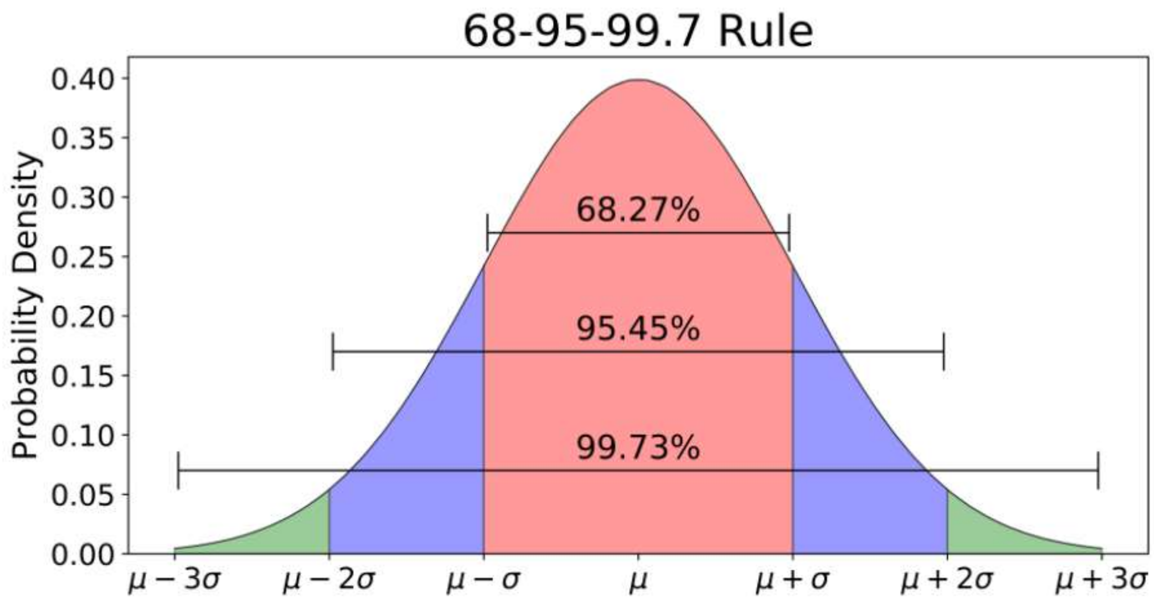
- Z Tables (demo)
- NORM.DIST(probability, mean, standard deviation, cumulative)

Gives the probability a normal random variable with given mean and sigma is basically less than or equal to x

- NORM.INV(probability, mean, standard deviation)

Gives the p^{th} percentile

68-95-99.7 rule for Normal Distribution



68% of the data is within 1 standard deviation, 95% is within 2 standard deviation, 99.7% is within 3 standard deviations

Normal Random Variable

- Example

IQs have a mean of 100 and a standard deviation of 15. What's the probability somebody's IQ is between 90 and 120?

Answer:

Hints: Substract area on the left of 120 with the area on the left of 90

What is Probability someone's IQ is >120

Answer: pto

Normal Random Variable

- Answer:

	A	B	C
76	IQ<=120	90.88%	=NORM.DIST(120,100,15,TRUE)
77	IQ<=90	25.25%	=NORM.DIST(90,100,15,TRUE)
78	IQ between 90-120	65.63%	=B76-B77
79	IQ >120	9.12%	=1-B76

Exercise 17

- Assume the average weight of a loaf of bread is a normal random variable with mean = 1 pound and standard deviation .05 pounds. What fraction of the loaves weigh between 0.98 and 1.04 pounds?
- A. 0.44
- B. 0.46
- C. 0.48
- D. 0.50

Answer: PTO

Exercise 17 - Answer

	A	B	C	D	E
1	**Final 17**				
2					
3	mean	1			
4	std. dev	0.05			
5	x1	1.04			
6	x2	0.98			
7	P(<= 1.04)	0.78814	=NORM.DIST(B5,B3,B4,TRUE)		
8	P(<= 0.98)	0.34458	=NORM.DIST(B6,B3,B4,TRUE)		
9	P(0.98 - 1.04)	0.44357	=B7-B8		

Central Limit Theorem

- The sample means that could be obtained from various samples tend to be normally distributed around the population mean.
- the bigger the sample, the more likely the distribution is to be normal
- Sample variance shrinks in size as n increases
- the mean of the sample means should be approximately equal to the mean of the population
- the variance approximately equal to the population variance, divided by the sample size

Central Limit Theorem (in Excel)

- NORM.DIST

Returns the normal distribution for the specified mean and standard deviation. This function has a very wide range of applications in statistics, including hypothesis testing.

- NORM.INV

calculates the inverse of the Cumulative Normal Distribution Function for a supplied value of x, and a supplied distribution mean & standard deviation

Central Limit Theorem

- you're running a pizza shop in your town and let's suppose you know on average each day in a 30 day month you'll sell 45 pizzas and the standard deviation is 12
- if we have a 30 day month, what's the chance we'll sell more than1,400 pizzas

Answer: **PTO**

Hints: Calculate Mean per month, standard deviation per month, then use excel function NORM.DIST(1400.5,1350,65.72,TRUE)

Central Limit Theorem

- Answer: **22%**

Mean in one month = 45 * 30 = 1,350

Std. dev in one month = sqrt(12^2 * 30) = 65.72

	A	B	C	D	E
1	Mean/month =	1350	=45*30		
2	std. dev/month =	65.72671	=SQRT(12^2 * 30)		
3	Probability os selling pizzas <= 1,400	78%	=NORM.DIST(1400.5,B1,B2,TRUE)		
4	Probability os selling pizzas > 1,400	22%	=1-B3		

Z Scores

- Z score = STANDARD score
- indicates how many standard deviations a data point is from the

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

where:

μ is the mean of the population.

σ is the standard deviation of the population.

Z Scores

- Provide information about how a value compares to a population mean
- Z Score < -2 OR Z Score $> +2$ is Outliers OR 5% of the sample/data
- $-1 <$ Z Score $< +1$ is 84.13%
- See the Area Under the Normal Distribution Table

Z Scores

- Example:

a product was rated by customers as a 70, and the average is 55.

Does it mean the product was highly rated?

If standard deviation 10 ; Z score is 15/10 = 1.5 => High

If standard deviation 20 ; Z score is 15/20 = 0.75 => Not great

Homework & Quiz

- JulyNormalHW.xlsx
- JulyCentralLimitHW.xlsx
- JulyZscoresHW.xlsx