Probability and Statistics

ishma hafeez
notes

osama.eifaz @nu

1. Walpule
2. Prem S
3. Neil A Weiss

ishma Hafeez notes

mid 1	15%
mid 2	15%
final	50%
Qs/As	20%

- Introduction, variables, momentum of scalars, types of data
- mean, median, mode, quartiles, variants, standard deviation, co.efficience of variant
- Histogram, Bar plot, Box plot, dot plot, frequency, polygon, stan & leaf

$\}$ Mid 1

- Probability, conting techinque, Permutation, combination
- venn digram, addition rule, multiplicative rule
- Conditional Probability and Bayers rule

Statistics

The science of collection, presentation, analysis and interpretation of numerical facts/data

OR

its the collection of procedures and principals for gathering data and analysing information to help people make decisions when found with uncertanity

STATISTICS

Description

- Data summary
- Graphs tables
- Mean, median, mode

Information

- Drawing conclusions about population using samples

Observational study

↳ sample survey

1. researchers observe characeristics
2. take measurements

Designed Eperiment

1. researchers impose treatments and controls
2. then observe characteristics
3. take measures

Population:

The collection of all indivisuals/objects under consideration in a statistical study

Parameter:

characteristics of population

Mean (μ), variant (σ^2) etc

Statistic:

characteristic of sample

Mean ($\bar{x}$), variant (s^2) etc

Sample:

Subset of population

Random sample:

It is a sample selected in such a way that every member of the population has an equal chance of being selected

VARIABLE

→ THIS WHOLE COURCE

Qualitative

LONG STATEMENTS
FEELINGS
JOB INTERVIEWS

- ↳ non numerical
- ↳ categorical
- ↳ Gender
- ↳ Religion
- ↳ Races

Quantitative

QUESTIONARIES

- ↳ numerical
- ↳ A, W, H

Discrete

- ↳ whole numbers
- ↳ data values consists of gaps
- ↳ no of students

Continous

- ↳ w/o gap
- ↳ intervals
- ↳ 64"
  63 — 64.1"

DATA

PRIMARY

- ↳ raw data
- ↳ first hand information
- ↳ Survery questionnaries

SECONDARY

- ↳ alredy collected by someone
- ↳ Published reports of research organisations

cross section data

data collected on different elements at the same Point in time

- ↳ objects are different
- ↳ Time is constant

University	Annual Cost (dollars)
Harvard University	69,600
Princeton University	66,150
Stanford University	69,109
University of California, Berkeley	65,003
University of Chicago	75,735
University of Pennsylvania	71,715
Yale University	71,290
MIT	67,430

Time series data

Data collected on same element for same element at different intervals in time

- ↳ 1 variable
- ↳ different time intervals

Years	Total Population
2015	320,878,310
2016	323,015,995
2017	325,084,756
2018	327,096,265
2019	329,064,917

Panel

- ↳ different cross sections
- ↳ different time Period

MEASUREMENTS OF SCALES

Qualitative

String

if you can't take avg

Quantitative

Nominal scales

- ↳ non numerical data
- ↳ categorial data
- ↳ classification
- ↳ Gender / Blood group
- ↳ Religion
- ↳ Proffesion
- ↳ zip code

Ordinal Scale

- ↳ ranking
- ↳ ordered data
- ↳ customer rating {strongly agree, agree, neutral, disagree}
- ↳ Likert scale
- ↳ Performance of students

Interval scale

- ↳ +, −
- ↳ absolute zero does not exist
- ↳ Temprature
- ↳ IQ scores
- ↳ variables for which '0' isnt meaningful

Ratio Scale

- ↳ +, −, *, ÷
- ↳ absolute zero exist

Show something absence

- ↳ Area / volume / length
- ↳ Distance / Weight / Height
- ↳ Time
- ↳ salary
- ↳ Age

Q)

Variable	Scale
ranking of golfers	ordial
temp	interval
weight	ratio
Salries	ratio
ratings	Ordinal
categories	nominal

Frequency Distribution

- ↳ Histogram
- ↳ Polygon
- ↳ ogive

Ordered Array

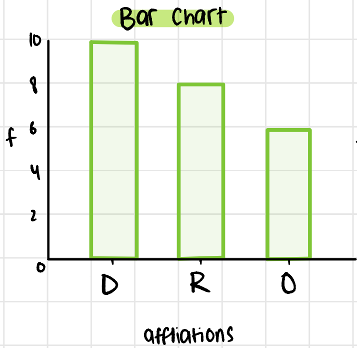
↳ Stem and leaf

Afficiations	Talls	f
D	𝍸 𝍸	10
R	𝍸 III	8
O	𝍸 I	6

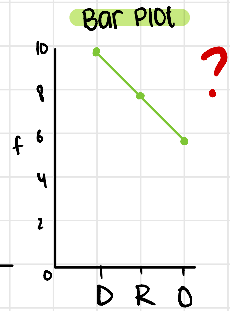
→ qualitative

D	R	O	R
D	D	R	R
O	O	D	D
R	R	R	R
D	D	D	O
O	D	O	D

Bar Chart



Bar Plot



Stem and leaf Plot

↓ leading digits ↘ trailing digits

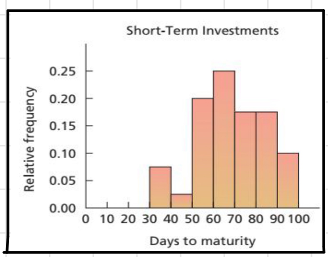
70	64	99	55	64	89	87	65
62	38	67	70	60	69	78	39
75	56	71	51	99	68	95	86
57	53	47	50	55	81	80	98
51	36	63	66	85	79	83	70

→

3	8,6,9
4	7
5	7,1,6,3,5,1,0,5
6	2,4,7,3,6,4,0,7,5
7	0,5,1,0,9,8
8	5,9,1,7,0,3,6
9	9,9,5,8

SORT IT

Stem	leaf								
3	6	8	9						
4	7								
5	0	1	1	3	5	5	6	7	
6	0	2	3	4	4	5	6	7	7
7	0	0	1	5	8	9			
8	0	1	3	5	6	7	9		
9	5	8	9	9					



Class interval, frequency, cumulative frequency, round up, −0.5, +0.5

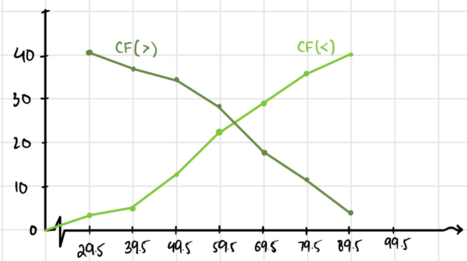
C.I	Tally	f	CF(<)	CF(>)	R.F ($\frac{f}{total}$)	Class boundary
30 – 39		3	3	40	3/40 = 0.075	29.5 – 39.5
40 – 49		1	4	37	1/40	39.5 — 49.5
50 – 59		8	12	36	⋮	49.5 — 59.5
60 – 69		10	22	28		59.5 — 69.5
70 – 79		7	29	18		69.5 — 79.5
80 – 89		7	36	11		79.5 — 89.5
90 – 99		4	40	4	4/40	89.5 — 99.5
Total		40				

Histogram / Polygon

• single peak
• uni model
• normal

Ogive

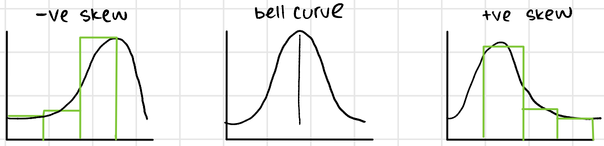
↳ CF



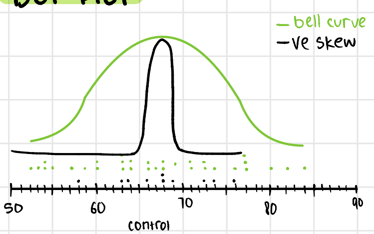
−ve skew

bell curve

+ve skew



Dot plot



Intervention		Control						
68	66	74	52	67	63	77	57	80
74	58	77	53	76	54	73	54	
69	63	60	77	63	60	68	64	
68	73	66	71	66	55	71	84	
64	76	63	73	59	68	64	82	

Pareto Charts

When the variable displayed on the horizontal axis is qualitative or categorical, a *Pareto chart* can also be used to represent the data.

> A **Pareto chart** is used to represent a frequency distribution for a categorical variable, and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

Homeless People

The data shown here consist of the number of homeless people for a sample of selected cities. Construct and analyze a Pareto chart for the data.

City	Number
Atlanta	6832
Baltimore	2904
Chicago	6680
St. Louis	1485
Washington	5518

BAR GRAPH

but arranged by their heights in desending order

Solution

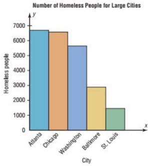
Step 1 Arrange the data from the largest to smallest according to frequency.

City	Number
Atlanta	6832
Chicago	6680
Washington	5518
Baltimore	2904
St. Louis	1485

Step 2 Draw and label the x and y axes.

Step 3 Draw the bars corresponding to the frequencies.

The graph shows that the number of homeless people is about the same for Atlanta and Chicago and a lot less for Baltimore and St. Louis.



Suggestions for Drawing Pareto Charts

1. Make the bars the same width.
2. Arrange the data from largest to smallest according to frequency.
3. Make the units that are used for the frequency equal in size.

The Time Series Graph

When data are collected over a period of time, they can be represented by a time series graph. used to study patterns

> A **time series graph** represents data that occur over a specific period of time.

Workplace Homicides

The number of homicides that occurred in the workplace for the years 2003 to 2008 is shown. Draw and analyze a time series graph for the data.

Year	'03	'04	'05	'06	'07	'08
Number	632	559	567	540	628	517

Source: Bureau of Labor Statistics.

Solution

Step 1 Draw and label the x and y axes.

Step 2 Label the x axis for years and the y axis for the number.

Step 3 Plot each point according to the table.

Step 4 Draw line segments connecting adjacent points. Do not try to fit a smooth curve through the data points.

There was a slight decrease in the years '04, '05, and '06, compared to '03, and again an increase in '07. The largest decrease occurred in '08.

numerical variable

time period

The Pie Graph

Pie graphs are used extensively in statistics. The purpose of the pie graph is to show the relationship of the parts to the whole by visually comparing the sizes of the sections. Percentages or proportions can be used. The variable is nominal or categorical.

> A **pie graph** is a circle that is divided into sections or wedges according to the percentage of frequencies in each category of the distribution.

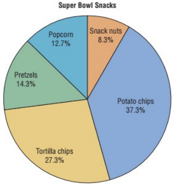
Super Bowl Snack Foods

This frequency distribution shows the number of pounds of each snack food eaten during the Super Bowl. Construct a pie graph for the data.

Snack	Pounds (frequency)	Percentages Step-I	Angles Step-II
Potato chips	11.2 million	(11.2/30)*100 = 37.3	(11.2/30)*360 = 134°
Tortilla chips	8.2 million	(8.2)/30*100 = 27.3	(8.2/30)*360 = 98°
Pretzels	4.3 million	14.3	52
Popcorn	3.8 million	12.7	46
Snack nuts	2.5 million	8.3	30
Total	30.0 million	99.9	360

Step-III

Next, using a protractor and a compass, draw the graph using the appropriate degree measures found in step 2, and label each section with the name and percentages



C.I (Class interval)	Tally	f (frequency)	Mid Point x	fx	C.f
30 – 39		3	$\frac{30+39}{2} = 34.5$	$f \times x =$ 103.5	3
40 – 49		1	44.5		4
50 – 59		8	54.5		12 → Preceding value
60 – 69		10	64.5		22 → closest to 20
70 – 79		7	74.5		29
80 – 89		7	84.5		36
90 – 99		4	94.5	37.8	40
Total		40		2720	

$$\text{mean}(\mu) = \frac{\Sigma x}{n}$$

$$\text{coeff of variance} = \frac{\sigma}{\mu} \times 100$$

$$\text{medium} = \left(\frac{n+1}{2}\right)^{th}$$

Standard deviation (σ)

$$\tilde{x} = \underset{\text{(L.C.B) lower class boundary median class}}{l} + \frac{\overset{\text{class width}}{h}}{f}\left[\frac{\Sigma f}{2} - \underset{\text{Preceding}}{\overset{\text{cumulative frequency preceding medium class}}{C.f}}\right]$$

LCB = −0.5
UCB = +0.5

$$\text{mode} = \underset{\text{(L.C.B) lower class boundary median class}}{l} + \left[\frac{\overset{\text{max frequency}}{f_m} - f_1}{2f_m - \underset{\text{preceding}}{f_1} - \underset{\text{following}}{f_2}}\right] \overset{\text{class size}}{h}$$

grouped ← variance → ungrouped

$$\sigma^2 = \frac{\Sigma[f(x-\mu)^2]}{\Sigma f} \quad \Bigg| \quad \sigma^2 = \frac{\Sigma(x-\mu)^2}{n} = \frac{\text{Sum of squared deviations from mean}}{\text{no of observations}}$$

Group

Q) find mean

$$\mu = \frac{2720}{40} = 68$$

Q) find mode

max frequency 10 at 60-69

$$59.5 + \frac{10-8}{2(10)-8-7} \times 10$$

$$= 63.5$$

Q) $\tilde{x} = \underset{60-0.5}{59.5} + \frac{10}{10}\left[\underset{\to 20}{\frac{40}{2}} - \underset{\text{choose preceding to 22}}{12}\right]$

$= 67.5$

Q) find standard deviation

$\mu = 68$ → found above

$$\sigma^2 = \frac{3(34.5-68)^2 + 1(44.5-68)^2 + 8(54.5-68)^2 + 10(64.5-68)^2 + 7(74.5-68)^2 + 7(84.5-68)^2 + 4(94.5-68)^2}{40}$$

$$= \sqrt{\frac{10510}{40}} = 16.209$$

Q) find coeff of variance

$$= \frac{16.209}{68} \times 100 = 23.8\%$$

68 > 67.5 > 63.5

$\bar{x}$ mean

$\check{x}$ median

$\hat{x}$ mode

$\bar{x} > \check{x} > \hat{x}$

+ve skewed

ungroup

Q) $x = 2, 4, 6, 8$, find mean → no mode as all values occur exactly once

$$\mu = \frac{2+4+6+8}{4} = 5$$

Q) 1, 2, 3, find medium

$$= \frac{3+1}{2} = 2^{nd} \text{ position} \to \text{has } 2$$

Q) 2, 4, 6, 8, find medium

$$\frac{4+1}{2} = 2.5^{th} \to \frac{2^{nd} + 3^{rd}}{2} \to \frac{4+6}{2} = 5$$

Q) 2, 3, 5, 8, find variance

$$\mu = \frac{18}{4} = 4.5$$

$$\sigma^2 = \frac{(2-4.5)^2 + (3-4.5)^2 + (5-4.5)^2 + (8-4.5)^2}{4}$$

$= \sqrt{5.25}$

$= 2.3$

APPLICATIONS

3.49 The following table gives the frequency distribution of the number of hours spent last week on cell phones (making phone calls and texting) by all 100 students in the tenth grade at a school.

Hours	Number of Students (f)	x	fx	cf
0 to less than 4	14	2	28	14
4 to less than 8	18	6	108	32
8 to less than 12	25	10	250	57
12 to less than 16	18	14	252	75
16 to less than 20	16	18	288	91
20 to less than 24	9	22	198	100
	100		1124	

Find the mean, variance, and standard deviation.

$$\mu = \frac{1124}{100} = 11.24$$

$$\text{mode} = 8.5\left[\frac{25-18}{2(25)-18-18}\right] \times 4$$

$$\sigma^2 =$$

Empirical Rule

xbar – mean

s–standard deviation

$\bar{x} \pm s$ → Approx. 68% observations

$\bar{x} \pm 2s$ → Approx. 95% observations

$\bar{x} \pm 3s$ → Approx. 99.7 observations

• Consider the following sample of exam scores, arranged in increasing order. The sample mean and Sample standard deviation of these exam scores are, respectively, 85 & 16.1.

28	57	58	64	69	74
79	80	83	85	85	87
87	89	89	90	92	93
94	94	95	96	96	97
97	97	97	98	100	100

Use the data to obtain the exact percentage of observations that lie within two standard deviations to either side of the mean. Compare your answer.

Empirical rule is fixed if the data is normally distributed or symmetric.

$\bar{x} \pm s$ = 85 ± 16.1 → 68% observation will fall in this interval

$\bar{x} \pm 2s$ = 85 ± 2*16.1 → 95 observation will fall in this interval

$\bar{x} \pm 3s$ = 85 ± 3*16.1 → 99.73 observations will fall in this interval.

(do this if you are asked to check the upper empirical rule otherwise not needed)

Rule	Lower to Upper	Frequency
85 ± 16.1	68.9 to 101.1	
85 ± 2*16.1	52.8 to 117.2	
85 ± 3*16.1	36.7 to 133.3	

Quartiles

5, 15 16 20 21 25 26 27 30 30

31 32 32 34 35 38 38 41 43 66 → outlier outside max

data set represent

no. of observations

25% $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ lower data set $\left(\frac{20+1}{4}\right)^{th} \Rightarrow 5.25^{th} \rightarrow \frac{21+25}{2} = 23$ (5th, 6th)

50% $Q_2 = \left(\frac{2(n+1)}{4}\right)^{th}$ median $2(5.25) \Rightarrow 10.5^{th} \rightarrow \frac{30+31}{2} = 30.5$ (10th, 11th)

75% $Q_3 = \left(\frac{3(n+1)}{4}\right)^{th}$ upper data set $3(5.25) \Rightarrow 15.75 \rightarrow \frac{35+38}{2} = 36.5$ (15th, 16th)

Box and whiskers Plot

↳ Data shape (skew)

↳ Outliers

↳ variation (Box length)(IQR)

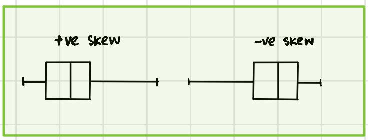
$IQR = Q_3 - Q_1$

min Q1 Q2 Q3 max

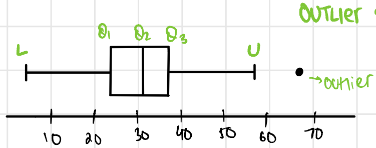
min ↓ L, max ↓ U

$Q_1 - 1.5(IQR) = 2.75$

$56.75 = Q_3 + 1.5(IQR)$



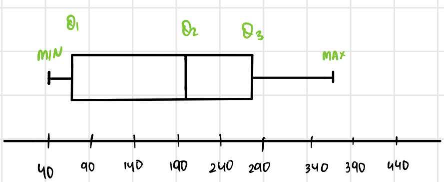
Outlier so use L U



NO outlier so use max min

no outliers

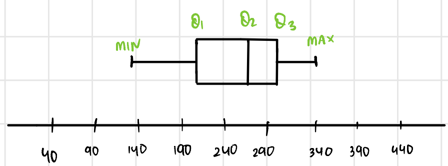
	1	2	3	4	5	6	7	8	Q1	Q2	Q3	IQR	L	U
									2.25	4.50	6.75			
Real cheese $X_1 \rightarrow$	40	45	90	180	220	240	310	420	67.5	200	275	207.5	-243.75	586.25
Cheese substitute $X_2 \rightarrow$	130	180	250	260	270	290	310	340	215	265	300	85	87.5	431.25



REAL CHEESE +ve skew

has more sodium content

greater variation



CHEESE SUBSTITUTE -VE SKEW

Probability

Statistical Experiment

Any activity/process that generates data

↳ Tossing a coin

↳ Rolling a dice

↳ Playing cards

↳ survery/opinion of voters

Sample space

List of all possible outcomes of a statistical experiment

↳ Tossing a coin 2 times: {HH, HT, TH, TT}

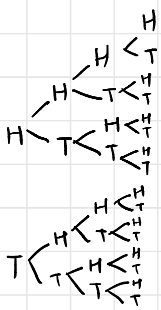
↳ Rolling a device : {1,2,3,4,5,6}

event: Subset of a sample space

✱ Toss a coin 4 times/4 coins tossed together

$2^4 = 6$ (possibilities ↙ 2, ↗ tosses 4)

Tree Digram



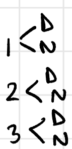
list of elements

S={HHHH, HHHT HHTH
HHTT, HTHH, HTHT
HTTH, HTTT
THHH, THHT, THTH
THTT, TTHH, TTHT
TTTH, TTTT }

✱ 3 items < defected / non defected

$2^3 = 8$

Tree Digram



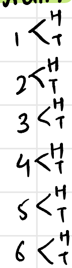
list

(1,D)(1,N)
(2,D)(2,N)
(3,D)(3,N

✱ Tossing a dice & coin together

$6 \times 2 = 12$?

Tree Digram



list of elements

(1,H)(1,T)
(2,H)(2,T)
(3,H)(3,T)
(4,H)(4,T)
(5,H)(5,T)
(6,H)(6,T)

✱ 4 students as either M or F

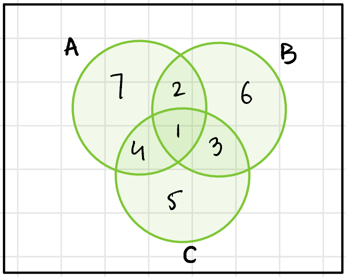
1^{st} sample space

$S_1 = \{MMMM, MMFM \dots\}$

2^{nd} sample space

$S_2 = \{0,1,2,3,4\}$

VENN DIGram



$A \cap B = 1, 2$

$B \cap C = 1, 3$

$A \cup C = 1, 2, 3, 4, 5, 7$

$B' \cap A = 4, 7$

$A \cap B \cap C = 1$

$(A \cup B) \cap C' = 7, 2, 6$

Permutations

↳ order matters

↳ organise

↳ ways

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

Combinations

↳ order dosent matter

↳ different types

↳ select

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

Emperical Formula

Probability

It is a measure of chanel that an uncertain event will occur

Subjective

↳ personal experience

↳ judgement

Objective

↳ classical approach

$$P(\text{Event}=A) = \frac{\text{favourable case of } A}{\text{all possible cases}}$$

$\leq$ Prob = 1 , $0 \leq P(A) \leq 1$

1) A coin is tossed twice. Prob atleast 1 head

$S = \{HH, TH, HT, TT\}$. $3/4$

3) A dice drawn such that an even no is twice likely to occur than odd

$$\begin{Bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ w & +2w+ & w+2w & + & w & +2w \end{Bmatrix}$$

$$P(X<4) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

5) A card is drawn at random from 52 cards

a) $P(J) = \frac{{}^4C_1}{{}^{52}C_1} = \frac{4}{52}$

b) In a poker hand consisting of 3 cards Prob of 2 Aces and 3 Jacks

$$\frac{\overset{\text{2 Aces}}{({}^4C_2)}\overset{\text{3 Jacks}}{({}^4C_3)}}{\text{total} \leftarrow {}^{52}C_5 \rightarrow \text{need how many}} =$$

Q) 10 E 10 M

25 I 8 C

a) $P(I) = 25/53$

b) $P(C \text{ or } E) = P(C) + P(E) = 8/53 + \frac{10}{53}$

Q) Dice rolled. Sum 7 or 11 → exclusive

$P(7) = \begin{matrix} 16 & 25 & 34 \\ 43 & 52 & 61 \end{matrix}$ $P(11) = 65, 56$

$$P(7 \text{ or } 11) = \frac{6}{36} + \frac{2}{36}$$

OR has + → only one event from choice

Mutually exclusive events

↳ nothing in common

$P(A \cup B) = P(A) + P(B)$

$P(A \cap B) = \emptyset$

AND has * → more than 1 events

Non mutually exclusive events

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - P(A \cap B) - P(B \cap C) + P(A \cap B \cap C)$

Independent events

$P(A \cap B) = P(A)\,P(B)$

if $P(A|B) = P(A)$ then independent

else dependent

Dependent events

$P(A \cap B) = P(A)\,P(B|A)$

BAYES RULE

$$P(A_i/B) = \frac{P(B|A_i)\,P(A_i)}{\sum_{i=1}^{k} P(B|A_i)\,P(A_i)}$$

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

A will occur given that B already

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

↳ $P(A|B) = \frac{P(A \cap B)}{P(B)}$ $\quad P(B|A) = \frac{P(A \cap B)}{P(A)}$

↳ $P(A|B)\,P(B) = P(B|A)\,P(A)$

CONDITIONAL PROBABILITY

Q)

	non smokers	moderate smokers	heavy smokers	Total
H	21	36	30	87
NH	48	26	19	93
Total	69	62	49	180

a) $P(H/H_s) = \frac{P(H \cap H_s)}{P(H_s)} = \frac{30}{49}$

b) $P(N_s/NH) = \frac{P(N_s/NH)}{P(NH)} = \frac{48}{93}$

Q)

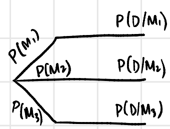
P of choosing M		P of D of M		P of it being M and D
$P(M_1) = 0.30$	x	$P(D/M) = 0.02$	=	0.006
$P(M_2) = 0.45$	x	$P(D/M_2) = 0.03$	=	0.0135
$P(M_3) = 0.25$	x	$P(D/M_3) = 0.02$	=	0.005
$P(D) = ?$				0.0245

$P(D) = P(D \cap M_1) + P(D \cap M_2) + P(D \cap M_3)$

$= P(M_1)P(D/M_1) + P(M_2)(D/M_2) + P(M_3)(D/M_3)$

$= 0.0245$

$P(M_1/D) = \frac{P(M_1 \cap D)}{P(D)} = \frac{0.006}{0.024}$



1)

Station	A	B	C	Total
Electric	2	1	1	
Computers	4	3	2	
Equipment	5	4	2	
Human Error	7	7	5	
Total				

$P(C/HE) = \frac{P(C \cap HE)}{P(HE)} = \frac{5}{19}$

2)

	C	D	
male	10	2	
Female	5	3	
			20

P(diploma holder is a male)

$P(M/D) = \frac{P(M \cap D)}{P(D)} = \frac{2}{5}$

3)

	1	2	3	Total
G	1	3	4	8
Y	2	2	3	2
R	4	5	1	10
Total	7	10	8	25

a) $P(Y) = P(Y \cap B_1) + (Y \cap B_2) + P(Y \cap B_3)$

$\frac{1}{3} \times \frac{2}{7} + \frac{1}{3} \times \frac{2}{10} + \frac{1}{3} \times \frac{3}{8} = \frac{241}{840}$

b) $P(B_2/Y) = \frac{P(B_2 \cap Y)}{P(Y)}$

$= \frac{\frac{1}{3} \times 2/10}{241/840}$

4) M_1 M_2 M_3 → chances M hits

$\frac{1}{6}$ $\frac{1}{4}$ $\frac{1}{3}$

P(target was hit by M_1) =

$P(M_1/H) = \frac{P(M_1 \cap H)}{P(H)} = \frac{\frac{1}{3} \times \frac{1}{6}}{\frac{1}{3}(\frac{1}{6}) + \frac{1}{3}(\frac{1}{4}) + \frac{1}{3}(\frac{1}{3})} = \frac{2}{9}$?

5) H_1 H_2 H_3 → chances

$\frac{1}{6}$ $\frac{1}{4}$ $1/3$

$P(M_1) = \frac{1}{6}(1 - \frac{2}{4})(1 - \frac{1}{3})$

$P(M_2) = (1 - \frac{1}{6})\, 2/4\, (1 - 1/3)$

$P(M_3) = (1 - 1/6)(1 - 2/4)\frac{1}{3}$

P(target was hit by M_1) =

P(   ) =

a) How many diff ways can student check of 1 ans to each question. 5 questions 4 option

$4^5 =$

b) get all wrong

3^5 → 5 questions → 3 wrong

2.37) 4B, 5G

if B, G must alternate

$5 \times 4 \times 4 \times 3 \times 3 \times 2 \times 2 \times 1 \times 1 = 2880$

G B G B G B G B G

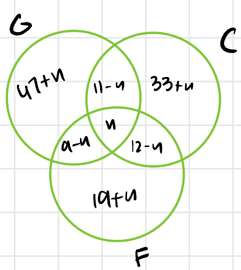
Q) How many ways no 2 students have same bday. 60 student

$365 P_{60}$

days ← → students

Q) In a survey, B=136, G=67, C=56, F=40, $G \cap C = 11$, $C \cap F = 12$, $G \cap F = 9$, each boy played atleast 1

How many played all games



0, 1 2 3 4 5 6 NO REPETIION, in 3 DIGITS

a) $\square \quad \square \quad \square$

$6 \times 6 \times 5 = 180$

↳ for 0

b) How many of these are odd numbers

$5 \times 5 \times 3 = 75$

c) How many > 330

$3 \times 6 \times 5 + \underbrace{1 \times 3}_{300} \times 5 = 105$

4007

a) How many no each of 5 digits constructed

$0 \quad \underline{4 \quad 5} \quad 6 \quad 7$

4 and 5 next to eachother

$3 \times 3! \times 2! = 36$

b) not to be together 4 and 5

5 digits. $4 \times 4! = 96$

$96 - 36 = 60$

APPLICATIONS

2.11 A local gas station collected data from the day's receipts, recording the gallons of gasoline each customer purchased. The following table lists the frequency distribution of the gallons of gas purchased by all customers on this one day at this gas station.

Gallons of Gas	Number of Customers	mid Point n	R.f	P.D
0 to less than 4	31	2	0.084	8.4
4 to less than 8	78	6	0.211	21.1
8 to less than 12	49	10	0.133	13.3
12 to less than 16	81	14	0.220	22
16 to less than 20	117	18	0.317	31.7
20 to less than 24	13	22	0.035	3.5
	369			

a. How many customers were served on this day at this gas station? 369

b. Find the class midpoints. Do all of the classes have the same width? If so, what is this width? If not, what are the different class widths? all have same widths n=4

c. Prepare the relative frequency and percentage distribution columns.

d. What percentage of the customers purchased 12 gallons or more? 57.2

e. Explain why you cannot determine exactly how many customers purchased 10 gallons or less. not specified

f. Prepare the cumulative frequency, cumulative relative frequency, and cumulative percentage distributions using the given table.

Make a dotplot for these data.

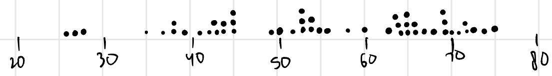
2.35 The following data give the times (in minutes) taken by 50 students to complete a statistics examination that was given a maximum time of 75 minutes to finish.

41	28	45	60	53	69	70	50	63	68
37	44	42	38	74	53	66	65	52	64
26	45	66	35	43	44	39	55	64	54
38	52	58	72	67	65	43	65	68	27
64	49	71	75	45	69	56	73	53	72

20 80

Create a dotplot for these data.

2.36 The following data give the one-way commuting times (in



EMPERICAL?
SETS DIGRAM?
PERMUTATIONS for string?

each stem in increasing order.

2.27 The following data give the times (in minutes) taken by 50 students to complete a statistics examination that was given a maximum time of 75 minutes to finish.

41	28	45	60	53	69	70	50	63	68
37	44	42	38	74	53	66	65	52	64
26	45	66	35	43	44	39	55	64	54
38	52	58	72	67	65	43	65	68	27
64	49	71	75	45	69	56	73	53	72

a. Prepare a stem-and-leaf display for these data. Arrange the leaves for each stem in increasing order.

b. Prepare a split stem-and-leaf display for the data. Split each stem into two parts. The first part should contains the leaves 0,

Cor
for
2.3
boo
univ
56
84
6
84

2 8 6 7
3 7 8 8 5 9
4 1 5 4 5 9 2 3 5 4
5 3 0 2 8 3 3 6 5 2 3 4
6 0 9 3 8 4 6 7 5 9 6 5 5 4 8 4
7 0 1 2 5 4 3 2

RANDOM VARIABLE (RV)

A r.v is a function that associates a real no. with each element in the sample space

Discrete random variable: possible outcomes countable

Continous random variable: valves on a continous scale

Q) let X= no. of Heads in toss of 3 coins

$2^3 = 8$

x		$P(x)$	$F(x)$
0	← TTT	$1/8$	$1/8$
1	← HTT, THT, TTH	$3/8$	$4/8$
2	← HHT, THH, HTH	$3/8$	$7/8$
3	← HHH	$1/8$	$8/8$

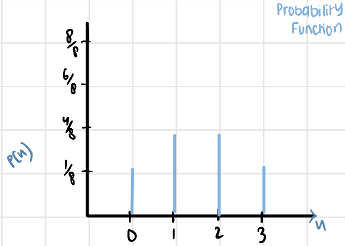
Cumulative Distribution function (CDF)

$F(x) = P(x \leq n)$

$$F(x) = \begin{cases} 0 & n < 0 & P<0 \quad P(n<0) \\ 1/8 & 0 \leq n < 1/8 & P\leq 0 \quad P(n=0) \\ 4/8 & 1/8 \leq n < 4/8 & P\leq 1 \quad P(n=0)+P(n=1) \\ 7/8 & 4/8 \leq n < 7/8 & P\leq 2 \quad P(n=0)+P(n=1)+P(n=2) \\ 1 & 7/8 \leq n < 1 & P\leq 3 \quad P(n=0)+P(n=1)+P(n=2)+P(n=3) \end{cases}$$

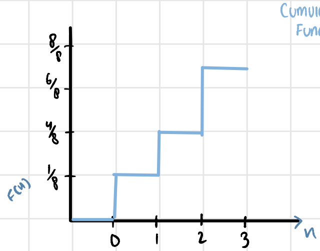
Line Graph → PMF → uses $P(n)$

Probability Mass Function



Line Graph → CDF → uses $F(n)$

Cumulative Distribution Function



Q) let X= no. of Heads in toss of 4 coins

$2^4 = 16$

x		$P(x)$	$F(x)$
0	← TTTT	$1/16$	$1/16$
1	← HTTT, THTT, TTHT, TTTH	$4/16$	$5/16$
2	← HHTT, TTHH, HTHT, THTH	$4/16$	$9/16$
3	← THHH, HTHH, HHTH, HHHT	$4/16$	$13/16$
4	← HHHH	$1/16$	$14/16$

Cumulative Distribution Function (CDF)

↳ $F(x) = P(X \le x)$ ↳ for continous RV

$= \int_{-\infty}^{\infty} f(t)\,dt$

↳ CDF $F(x) = P(X \le x) = \sum P(x)$ PMF

↳ $P(a < x < b) = F(b) - F(a)$

$f(x) = \frac{dF(x)}{dx}$

$F(x) = \sum_{x=\dots}^{x} (x+1)^x$

Probability mass Function (PMF)

1. $f(x) \geq 0$, for all $x \in R$

↳ $\sum f(x)\,dx = 1$ → verify PMF

↳ $P(X=x) = F(x) - F(x-1)$

↳ $P(a < x < b) = F(b) - F(a)$

$P(x) = (x+1)^x$

Example 1

• Consider the following pmf: f(x) = (x/6), x = 1, 2, 3, zero elsewhere

(i) Find distribution function and its graph.

(ii) Calculate P(1.5 < x ≤ 4.5)

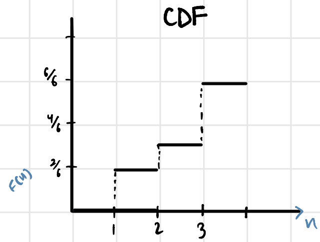
i) $f(x) = x/6$

x	$f(x)$	$F(x)$
1	1/6	1/6
2	2/6	3/6
3	3/6	6/6

→ Distribution function

$F(x)$	0	$x < 1$
	1/6	$1 \le x < 2$
	3/6	$2 \le x < 3$
	6/6	$x \geq 3$

graph



ii) $P(1.5 < x \le 4.5)$

$= F(4.5) - F(1.5)$

$= x \geq 3 - (1 \le x < 2)$

$= 1 - 1/6$

$= 5/6$

Example 3

(3) Consider the following functions

(i) f(x) = (x +2) / 5 for x = 1, 2, 3, 4, 5.

(ii) f(x) = (4Cx) / (2^5), for x = 0, 1, 2, 3, and 4.

and check whether the functions can serve as a pmf ?

i) $f(x) = \frac{(x+2)}{5}$

Since $0 \le x \le 1$ → what

$\sum f(x) \neq 1$

So not PMF

x	$f(x)$
1	3/5
2	4/5
3	1
4	6/5
5	7/5

$\sum f(x) = 5$

ii) $f(x) = \frac{{}^4C_x}{2^5}$

not PMF as $\sum f(x) \neq 1$

x	$f(x)$
1	1/32
2	1/8
3	3/16
4	1/8
5	1/32

$\sum f(x) = 1/2$

Example 4

(4) A coin is biased so that a head occurs 3 times of tail. If the coin is tossed 3 times, find the probability distribution for the number of heads and also find P (1 ≤ x ≤ 3).

H $\frac{3}{4}$ T $\frac{1}{4}$

0 { T T T $\frac{1}{4}\times\frac{1}{4}\times\frac{1}{4}\times 1$

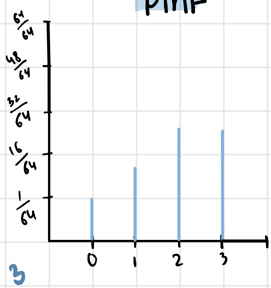
1 { H T T $\frac{3}{4}\times\frac{1}{4}\times\frac{1}{4}\times 3$
T H T
T T H

2 { H H T $\frac{3}{4}\times\frac{3}{4}\times\frac{1}{4}\times 3$
H T H
T H H

3 { H H H $\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}\times 1$

x	P(x)
0	$\frac{1}{64}$
1	$9/64$
2	$27/64$
3	$27/64$

PMF



$$P(1 \le x \le 3) = \overset{1}{9/64} + \overset{2}{27/64} + \overset{3}{27/64}$$
$$= 63/64$$

Example 7

• The distribution function for a discrete random variable x is given as:

F(x) = 1 – (1/2)^(x+1), for x = 0, 1, 2, ...

Find:

(a) P (X = 3)
(b) P (7 ≤ x < 10)
(c) Probability Mass Function.

$$f(x) = 1 - (1/2)^{(x+1)}$$

a) $P(x=3) = F(3) - F(2)$

$\therefore P(X=x) = F(x) - F(x-1)$

$$= \left[1 - \left(\frac{1}{2}\right)^{3+1}\right] - \left[1 - \left(\frac{1}{2}\right)^{2+1}\right]$$
$$= 1/16$$

b) $P(7 \le x < 10) = F(9) - F(6)$ what $F(10) - F(6)$
$= 7/1024$ $= 15/2048$

c) PMF

$$= F(x) - F(x-1)$$
$$= 1 - \frac{1}{2}^{x+1} - 1 + \frac{1}{2}^{x-1+1}$$
$$= \left(-\frac{1}{2}\right)^{x+1} + \left(\frac{1}{2}\right)^{x}$$
$$= 1/2^{x+1}$$ → WHAT

$$\left(-\frac{1}{2}\right)^{x} \cdot \left(-\frac{1}{2}\right)^{1} + \left(\frac{1}{2}\right)^{x}$$
$$-\frac{1}{2}\left(-\frac{1}{2}\right)^{x} + \left(\frac{1}{2}\right)^{x}$$

Example 8

3.5 Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$;

a) $f(x) = c(x^2 + 4)$

$$\sum_{x=0}^{3} f(x) = 1$$

$$\overset{0}{c(4)} + \overset{1}{c(5)} + \overset{2}{c(8)} + \overset{3}{c(3)} = 1$$

$$30c = 1$$

$$c = \frac{1}{30}$$

Q) find $P(n=2) = f(n=2)$ using CDF

$P(2) = F(2) - F(1)$

$= P(n \leq 2) - P(n \leq 1)$

$= 11/16 - 5/16$

$= 3/8$

Q1) A fair coin is tossed until H appears for the first time. Find

a) PMF = $(1/2)^n$ ∴ $n = 1, 2, 3, \ldots$

b) CDF = $\sum_{n=1}^{n=n} (1/2)^n$

c) $F(4) = \sum_{n=1}^{n=4} (1/2)^n$

Q) If two dice are rolled once, find the pmf of the sum of points on dice, CDF and their graph

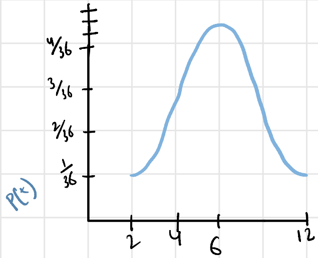
let $t = x + y$

z	2	3	4	5	6	7	8	9	.....	12
$f(t) = P(t) =$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$7/36$	$8/36$	.....	

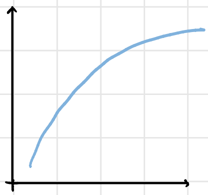
$F(t) = 1/36 \quad 3/36 \quad 6/36 \quad \ldots\ldots \quad 36/36$

$$F(t) = \begin{cases} 0 & z < 2 \\ 1/36 & 2 \leq z \leq 3 \\ 3/36 & 3 \leq z \leq 4 \\ \vdots & \\ 36/36 & z \geq 12 \end{cases}$$

PMF



CDF



✱ Expected Value

let suppose a coin tossed 2 times

X = Head calculate

E(X) = Mass of a row X

X	P(X)	X.P(X)	X^2.P(X)
0	1/4	0	0
1	2/4	2/4	2/4
2	1/4	2/4	4/4
Sum	1	E(X) = 1	$E(X^2) = 3/2$

$E(x) = \sum (x) P(x) \longrightarrow$ Expected value

$\sigma^2 x = V(x) = E(x^2) - [E(x)]^2 \longrightarrow$ Variance

Joint Probability Distribution

1. $f(x, y) \geq 0$
2. $\sum_x \sum_y f(x,y) = 1$
3. $P(X = x, Y = y) = f(x,y)$

$\hookrightarrow P[(X,Y) \in A] = \sum \sum_A f(x,y)$

MATHAMETICAL EXPECTATION

$\sigma^2 \rightarrow$ variance

$V(X) = E(X^2) - [E(X)]^2$

$V(Y) = E(Y^2) - [E(Y)]^2$

mean

$\mu = E(X) = \sum_x x f(x)$

$E(XY) = \sum \sum xy\, f(x,y)$

Covariance $(X,Y) = E(XY) - E(X).E(Y)$

Corelation $(X,Y) = \dfrac{\text{Covariance }(X,Y)}{\sqrt{V(X)}\ \sqrt{V(Y)}}$

$\pm 1 \longrightarrow$ strong
$\pm 0.5 \longrightarrow$ moderate
$0 \longrightarrow$ negligible

Example 9

- Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If *X* is the number of blue pens selected and *Y* is the number of red pens selected, find
- (a) the joint probability function *f(x, y)*
- (b) P[(X, Y) ∈ A], where A is the region {(x, y) | y ≤ 1} → WHAT

$B = 3, R = 2, G = 3 \Rightarrow 8 \rightarrow$ select 2 w/o replacement

JOINT PROB MASS FUNCTION

Blue → x, red → y; Marginal dist $g(x)$

$x \backslash y$	0	1	2	$g(x)$	$x\,g(x)$	$x^2 g(x)$
0	$\frac{{}^3C_0\,{}^2C_0\,{}^3C_2}{{}^8C_2} = \frac{3}{28}$	$\frac{{}^3C_0\,{}^2C_1\,{}^3C_1}{{}^8C_2}$ $\frac{6}{28}$	$\frac{{}^3C_0\,{}^2C_2\,{}^3C_0}{{}^8C_2}$ $\frac{1}{28}$	$\frac{10}{28}$	0	0
1	$\frac{{}^3C_1\,{}^2C_0\,{}^3C_1}{{}^8C_2} = \frac{9}{28}$	$\frac{{}^3C_1\,{}^2C_1\,{}^3C_0}{{}^8C_2}$ $\frac{6}{28}$	—	$\frac{15}{28}$	$\frac{15}{28}$	$\frac{15}{28}$
2	$\frac{{}^3C_2\,{}^2C_0\,{}^3C_0}{{}^8C_2} = \frac{3}{28}$	—	—	$\frac{3}{28}$	$\frac{6}{28}$	$\frac{12}{28}$
$h(y)$	$\frac{15}{28}$	$\frac{12}{28}$	$\frac{1}{28}$	1	$E(X)$ $\frac{21}{28}$	$E(X^2)$ $\frac{27}{28}$
$y\,h(y)$	0	$\frac{12}{28}$	$\frac{2}{28}$	$E(Y)$ $\frac{14}{28}$		
marginal dist of y $y\,h(y^2)$	0	$\frac{12}{28}$	$\frac{4}{28}$	$E(Y^2)$ $\frac{16}{28}$		

$E(x^4)$

$E(\sqrt{x})$

$E[3x] = 3E(x)$

$\sigma^2 x = V(X) = E(X^2) - [E(X)]^2 \longrightarrow$ Variance

$= \frac{27}{28} - \left(\frac{21}{28}\right)^2$

a) $P[X \leq 2, Y = 1] = f(0,1) + f(1,1) + f(2,1)$

$= \frac{6}{28} + \frac{6}{28} + \frac{3}{28}$

$= \frac{12}{28}$

b) $P[X \geq 2, Y \leq 1] = f(2,1) + f(2,0)$

c) $P[X > Y] = f(1,0) + f(2,0)$

d) $P[X + Y = 4] = 0$

Conditional Probability

e) $P[X = 0 / Y = 1] = \dfrac{f(x=0, y=1)}{h(y=1)}$

$= \dfrac{6/28}{12/28} = \dfrac{6}{12}$

f) $P[Y = 1 / X = 0] = \dfrac{f(y=1, x=0)}{g(x=0)} = \dfrac{6/28}{10/28}$

g) $E[XY] = \sum \sum xy f(x,y) = \frac{6}{28}$

0 0 = 0	×	3/28
0 1 = 0	×	6/28
0 2 = 0	×	1/28
1 0 = 0	×	9/28
1 1 = 1	×	6/28
1 2 = 2		
2 0 = 0	×	3/28
2 1 = 2		
2 2 = 4		
		6/28

h) Covariance $(X,Y) = E(XY) - E(X).E(Y)$

$= \frac{6}{28} - \left(\frac{21}{28}\right)\left(\frac{14}{28}\right)$

$= -\frac{9}{56}$

$-1 \leq$ Corelation $(X,Y) = \dfrac{\text{Covariance}(X,Y)}{\sqrt{V(X)}\ \sqrt{V(Y)}} \leq +1$

$= \dfrac{-9/56}{\sqrt{\frac{45}{112}}\ \sqrt{\frac{9}{28}}}$

$= -0.447$ moderate -ve relationship

JPMF

$f(x,y) = \frac{x+y}{30}$ for $x = 0,1,2,3$

$y = 0,1,2$

Calculated

a) $P[X \le 2, Y=1] = F(0,1) + F(1,1) + F(2,1)$

$= \frac{1}{30} + \frac{2}{30} + \frac{3}{30} = \frac{6}{30}$

b) $P[X > 2, Y \le 1] = F(3,1) + F(3,0)$

$= \frac{3}{30} + \frac{4}{30} = \frac{7}{30}$

c) $P[X > Y] = F(3,2) + F(3,1) + F(3,0) + F(2,1) + F(2,0) + F(1,0)$

$\frac{5}{30} + \frac{4}{30} + \frac{3}{30} + \frac{3}{30} + \frac{2}{30} + \frac{1}{30} = \frac{18}{30}$

d) $P[X+Y=4] = F(3,1) + F(2,2)$

$= \frac{4}{30} + \frac{4}{30} = \frac{8}{30}$

e) Calculate

marginal Dist X

x \ y	0	1	2	g(x)	xg(x)	$x^2g(x)$
0	0	1/30	2/30	3/30	0	0
1	1/30	2/30	3/30	6/30	6/30	6/30
2	2/30	3/30	4/30	9/30	18/30	36/30
3	3/30	4/30	5/30	12/30	36/30	108/30
h(y)	6/10	10/30	14/30	1	E(x)= 2	E(x²)= 5
yh(y)	0	10/30	28/30	E(y): 38/30		
$y^2h(y)$	0	10/30	56/30	E(y²): 66/30		

$(X-\mu)^2 f(x)$

$(X-2)^2 f(x)$

$E(x) = \frac{60}{30} = 2$

$E(x^2) = \frac{150}{30} = 5$

$V(x) = E(x^2) - [E(x)]^2 = 5 - (2)^2 = 1$

$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{66}{30} - \left(\frac{38}{30}\right)^2 = 0.59$

$E(xy) = \sum\sum xy\ f(x,y) = 2.4 \longrightarrow E(xy) = \sum\sum xy\ f(x,y) = 2.4$

00	=0	0
01	=0	1/30 × 0
02	=0	2/30 × 0
10	=0	1/30 × 0
11	=1	2/30 × 1
12	=2	3/30 × 2
20	=0	2/30 × 0
21	=2	3/30 × 2
22	=4	4/30 × 4
30	=0	3/30 × 0
31	=3	4/30 × 3
32	=6	5/30 × 6

$\text{Covariance}(x,y) = E(xy) - E(x).E(y) = 2.4 - 2\left(\frac{38}{30}\right) \rightarrow -\frac{2}{15} \rightarrow -0.133$

$\text{Corelation}(x,y) = \frac{\text{Covariance}(x,y)}{\sqrt{V(x)}\ \sqrt{V(y)}} = -0.17277$

↳ weak relation

↳ inverse relation as -ve sign

$= \frac{-0.133}{\sqrt{1}\ \sqrt{0.59}}$

Conditional Distribution of X

$$f(x|y) = \frac{f(x,y)}{h(y)} \quad \therefore h(y) > 0$$

Conditional Distribution of Y

$$f(y|x) = \frac{f(x,y)}{g(x)} \quad \therefore g(x) > 0$$

Slide 20

3.51 Three cards are drawn without replacement from the 12 face cards (jacks, queens, and kings) of an ordinary deck of 52 playing cards. Let X be the number of kings selected and Y the number of jacks. Find

(a) the joint probability distribution of X and Y;

(b) $P[(X,Y) \in A]$, where A is the region given by $\{(x,y) \mid x+y \geq 2\}$.

X = no of Kings Y = no of Jacks total facecards = 12
total cards = 52

a)

y \ x	0	1	2	3	Total g(x) (marginal dist g(x))
0	$\frac{{}^4C_0\,{}^4C_0\,{}^4C_3}{{}^{12}C_3} = \frac{1}{55}$	$\frac{{}^4C_0\,{}^4C_1\,{}^4C_2}{{}^{12}C_3} = \frac{6}{55}$	$\frac{{}^4C_0\,{}^4C_2\,{}^4C_1}{{}^{12}C_3} = \frac{6}{55}$	$\frac{{}^4C_3\,{}^4C_0\,{}^4C_0}{{}^{12}C_3}$	$14/55$
1	$\frac{{}^4C_1\,{}^4C_0\,{}^4C_2}{{}^{12}C_3} = \frac{6}{55}$	$\frac{{}^4C_1\,{}^4C_1\,{}^4C_1}{{}^{12}C_3}$	$\frac{{}^4C_1\,{}^4C_2\,{}^4C_0}{{}^{12}C_3} = \frac{6}{55}$	–	$28/55$
2	$\frac{{}^4C_2\,{}^4C_0\,{}^4C_1}{{}^{12}C_3} = \frac{6}{55}$	$\frac{{}^4C_2\,{}^4C_1\,{}^4C_0}{{}^{12}C_3} = \frac{6}{55}$	–	–	$12/55$
3	$\frac{{}^4C_3\,{}^4C_0\,{}^4C_0}{{}^{12}C_3}$	–	–	–	$1/55$
Total h(y)	$14/55$	$28/55$	$12/55$	$1/55$	1

marginal dist of y

b) $P(X,Y \in A) = P(X+Y \geq 2)$

Total Prob

$P = 1 - P(x+y \leq 1)$

$= 1 - [P(0,0) + P(0,1) + P(1,0)]$

$= 1 - \left[\frac{1}{55} + \frac{6}{55} + \frac{6}{55}\right]$

$$P(x+y \geq 2) = \frac{42}{55}$$

Statistical independence

$f(x,y) = g(x)\,h(y)$

$f(x,y) = f(x|y)\,h(y)$

$$g(x) = f(x|y) = \int_{-\infty}^{\infty} h(y)\,dy = 1$$

.1: Joint Probability Distribution for Example 3.14

	$f(x,y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

show not statistically independent

taking $f(0,1) = 3/14$

$g(0) = 5/14$

$h(1) = 15/28$

$f(0,1) = g(0)\,h(1)$

$3/14 = 5/14 \times 15/28$

$3/14 \neq \frac{75}{392}$

Slide 20

3.52 A coin is tossed twice. Let Z denote the number of heads on the first toss and W the total number of heads on the 2 tosses. If the coin is unbalanced and a head has a 40% chance of occurring, find
(a) the joint probability distribution of W and Z;
(b) the marginal distribution of W;
(c) the marginal distribution of Z;
(d) the probability that at least 1 head occurs.

Marginal Distribution → Discrete

$g(x) = \sum_y f(x,y)\,dy$

$h(y) = \sum_x f(x,y)\,dx$

H = 0.6
T = 0.4

a)

no. of heads on first toss

Total no. of heads on 2nd tosses

y \ x	z0	z1	Total g(x) (marginal dist g(x))
w0	(0.6)(0.6) TT = 0.36	(0.4)(0) TT = 0	0.36
w1	(0.4)(0.6) HT = 0.24	(0.4)(0.6) TH = 0.24	0.48
w2	(0.6)(0) HH = 0	(0.4)(0.4) HH = 0.16	0.16
Total h(w) (marginal dist of y)	0.6	0.4	1

b) Marginal distribution of w

$W = 0 \quad = 0.36$

$W = 1 \quad = 0.46$

$W = 2 \quad = 0.16$

c) Marginal Distribution of z

$z = 0 \quad = 0.6$

$z = 1 \quad = 0.4$

d) P(of atleast 1 H) $= P(W=1 \text{ or } W=2)$

$= P(W=1) + P(W=2)$

$= 0.48 + 0.16$

$= 0.64$

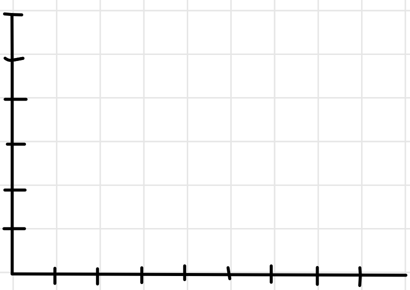
Slide 22

Let the number of phone calls received by a switchboard during a 5-minute interval be a random variable X with probability function

$$f(x) = \frac{e^{-2}2^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots.$$

(a) Determine the probability that X equals 0, 1, 2, 3, 4, 5, and 6.
(b) Graph the probability mass function for these values of x.
(c) Determine the cumulative distribution function for these values of X.

x	f(x)	F(x) → CDF
0	0.1353	0.1353
1	0.2707	0.4060
2	0.2707	0.6767
3	0.1804	0.8571
4	0.0902	0.9473
5	0.0361	0.9834
6	0.1202	0.9954



Probability Density Function (PDF)

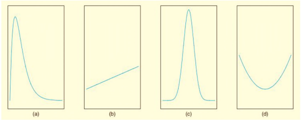
1. $f(x) \geq 0$, for all $x \in R$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ → verify PDF
3. $P(a < X < b) = \int_a^b f(x)\,dx$

Continuous Probability Distribution

↳ $F(x)$ → probability density function (PDF)

↳ Areas used to represent probabilities

↳ can not be given in tabular form



• A probability density function is constructed so that the area under its curve bounded by the *x* axis is equal to 1 when computed over the range of *X*.

Example 1

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

a) $\int_{-1}^{2} \frac{x^2}{3}\,dx = \left| \frac{x^3}{9} \right|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$ → as equal to 1 HENCE VERIFIED

b) $\int_{0}^{1} \frac{x^2}{3}\,dx = \frac{1}{9}$

Example 3

• The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate *b*. The DOE has determined that the density function of the winning (low) bid is:

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

• Find *F*(*y*) and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate *b*.

$F(y) = \int_{\frac{2}{5}b}^{y} \frac{5}{8b}\,dy = \left| \frac{5y}{8b} \right|_{\frac{2b}{5}}^{y} = \frac{5y}{8b} - \frac{10\cancel{b}}{40\cancel{b}} =$

$P(y \leq b) = F(b) = \frac{5\cancel{b}}{8\cancel{b}} - \frac{1}{4} = \frac{3}{8}$

Example 4

The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20,000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

(a) at least 200 days;

(b) anywhere from 80 to 120 days.

Example #04

(a) atleast 200 days.

$\int_{200}^{\infty} f(x)\,dx = \int_{200}^{\infty} \frac{20000}{(x+100)^3}\,dx$

$= \left(-\frac{10000}{(x+100)^2} \right)_{200}^{\infty} =$

$\boxed{\int_{200}^{\infty} f(x)\,dx = \frac{1}{9}}$

(b) $\int_{80}^{120} \frac{20000}{(x+100)^3}\,dx = \left(-\frac{10000}{(x+100)^2} \right|_{80}^{120}$

$= \frac{25}{121} - \frac{25}{\cancel{481}}$

$\boxed{= 0.1020}$

$20000 \int_{200}^{\infty} \frac{1}{(x+100)^3}\,dx$

$u = x+100$

$\frac{du}{dx} = 1 \Rightarrow dx = \frac{du}{1}$

$20000 \int_{200}^{\infty} u^{-3}\,du$

$20000 \left| \frac{u^{-2}}{-2} \right|$

$\cancel{20000}^{1} \left| \frac{(x+100)^{-2}}{-\cancel{2}} \right|$

$\left| \frac{-10000}{(x+100)^2} \right|_{200}^{\infty}$

Example 5

The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours;

(b) between 50 and 100 hours.

Example #05

(a) $x = \frac{120}{100} = 1.2$

$100 \longrightarrow$ hours (units measured)

$P(x < 1.2) = \int_0^1 x\,dx + \int_1^{1.2} (2-x)\,dx$

$= \frac{x^2}{2}\Big|_0^1 + \left(2x - \frac{x^2}{2}\right)_1^{1.2}$

$\boxed{P(x<1.2) = 0.68}$

(b) $x = \frac{50}{100} = 0.5$ and $x = \frac{100}{100} = 1$

$P(0.5 < x < 1) = \int_{0.5}^1 x\,dx = \frac{x^2}{2}\Big|_{0.5}^1$

$\boxed{P(0.5 < x < 1) = 0.375}$

Q2) $f(n) = \begin{cases} 3n^{-4} & ; \; n > 1 \\ 0 & ; \; \text{elsewhere} \end{cases}$

a) Verify that f(x) is a PDF

$= \int_1^\infty 3n^{-4}$

$= \left[-\frac{\cancel{3}}{\cancel{3}} n^{-3}\right]_1^\infty$

$= \left[-n^{-3}\right]_1^\infty$

$\left(-\frac{1}{\infty^3}\right) - \left(-\frac{1}{1^3}\right)$

$= 0 - (-1) = 1$

b) F(n) = CDF

use n as upper limit

$= \int_1^n \frac{3}{n^4}\,dn$

$= \left[-n^{-3}\right]_1^n$

$= -n^{-3} + 1$

$= 1 - n^{-3}$

c) $P(n > 4)$

$\int_4^\infty 3n^{-4}\,dn$

$\left|-n^{-3}\right|_4^\infty$

$-\frac{1}{64} - \frac{1}{\infty^3}$

$-\frac{1}{64}$

Example 7

Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k(3 - x^2), & -1 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.

(b) Find the probability that a random error in measurement is less than 1/2.

(c) For this particular measurement, it is undesirable if the *magnitude* of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

a) $\int_{-1}^{1} k(3 - n^2)\,dn = 1$ → equate to 1 as valid PDF equates to 1

$k\left(3n - \frac{n^3}{3}\right) = 1$

$k\left[3 - \frac{1}{3} - \left(3(-1) - \frac{(-1)^3}{3}\right)\right] = 1$

$k = \frac{3}{16}$

b) $P(n < \frac{1}{2}) = \int_{-1}^{1/2} \frac{3}{16}(3 - n^2)\,dn \Rightarrow \frac{99}{128}$

c) $P[\,|n| < 0.8\,] = P[n < -0.8] + P[n \geq 0.8]$

WHAT

$= F(-0.8) + [1 - F(0.8)]$

$= 0.164$

$-0.8 < n < 0.8$

Joint Density Function (JDF)

↳ for continous variable

1. $f(x,y) \geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$
3. $P[(X,Y)\in A] = \int_A\int f(x,y)\,dx\,dy$

Marginal Distribution

→ for continous variable

$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \rightarrow$ w.r.t. y

$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx \rightarrow$ w.r.t. x

Q3) $f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & 0<x<1 \\ & 0<y<1 \\ & \text{elsewhere} \end{cases}$

a) Verify that $f(x,y)$ is a valid JPDF

$\int_0^1\int_0^1 \frac{2}{5}(2x+3y)\,dx\,dy$

$\frac{2}{5}\int_0^1 (x^2+3xy)\Big|_0^1\,dy$

$\frac{2}{5}\int_0^1 (1+3y)\,dy$

$\frac{2}{5}\left[y+\frac{3y^2}{2}\right]_0^1$

$\frac{2}{5}\left[1+\frac{3}{2}\right]$

$\frac{2}{5}\left[\frac{5}{2}\right]$

1

c) find $g(x)$ & $h(y)$

Marginal $g(x) = \frac{2}{5}\int_0^1 (2x+3y)\,dy$

$= \frac{2}{5}\left(2xy+\frac{3y^2}{2}\right)\Big|_0^1$

$= \frac{2}{5}\left(2x+\frac{3}{2}\right)$

Marginal $h(y) = \frac{2}{5}\int_0^1 (2x+3y)\,dx$

b) $P[0<x<\frac{1}{2},\ \frac{1}{4}<y<\frac{1}{2}]$

$= \frac{2}{5}\int_{1/4}^{1/2}\int_0^{1/2} (2x+3y)\,dx\,dy$

$= \frac{2}{5}\int_{1/4}^{1/2} (x^2+3xy)\Big|_0^{1/2}\,dy$

$= \frac{2}{5}\int_{1/4}^{1/2}\left(\frac{1}{4}+\frac{3y}{2}\right)dy$

$= \frac{2}{5}\left(\frac{y}{4}+\frac{3y^2}{4}\right)\Big|_{1/4}^{1/2}$

$= \frac{2}{5}\left[\frac{1}{8}+\frac{3}{16}-\left(\frac{1}{16}+\frac{3}{64}\right)\right]$

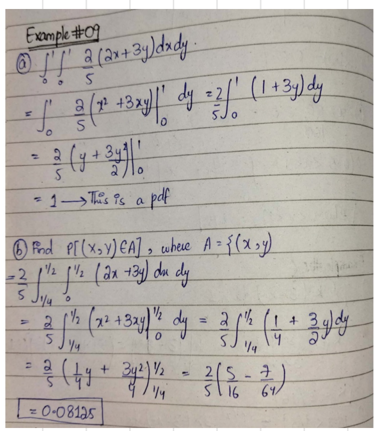
$= 0.08$

Example 9

- A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- (a) Verify for PDF?
(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < 1/2, 1/4 < y < 1/2\}$.



Example #09

ⓐ $\int_0^1\int_0^1 \frac{2}{5}(2x+3y)\,dx\,dy$

$= \int_0^1 \frac{2}{5}(x^2+3xy)\Big|_0^1\,dy = \frac{2}{5}\int_0^1 (1+3y)\,dy$

$= \frac{2}{5}\left(y+\frac{3y^2}{2}\right)\Big|_0^1$

$= 1 \longrightarrow$ This is a pdf

ⓑ Find $P[(X,Y)\in A]$, where $A=\{(x,y)$

$= \frac{2}{5}\int_{1/4}^{1/2}\int_0^{1/2} (2x+3y)\,dx\,dy$

$= \frac{2}{5}\int_{1/4}^{1/2} (x^2+3xy)\Big|_0^{1/2}\,dy = \frac{2}{5}\int_{1/4}^{1/2}\left(\frac{1}{4}+\frac{3}{2}y\right)dy$

$= \frac{2}{5}\left(\frac{1}{4}y+\frac{3y^2}{4}\right)\Big|_{1/4}^{1/2} = \frac{2}{5}\left(\frac{5}{16}-\frac{7}{64}\right)$

$\boxed{= 0.08125}$

Mean of a Random variable

1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x)$$

if X is discrete, and

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

if X is continuous.

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X-1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: By Theorem 4.1, the attendant can expect to receive

$$\begin{aligned} E[g(X)] = E(2X-1) &= \sum_{x=4}^{9}(2x-1)f(x) \\ &= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) \\ &\quad + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67. \end{aligned}$$

: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy$$

if X and Y are continuous.

ample 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X,Y) = XY$. The table is reprinted here for convenience.

	$f(x,y)$	$x=0$	$x=1$	$x=2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	Column Totals	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Solution: By Definition 4.2, we write

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y) \\ &= (0)(0)f(0,0) + (0)(1)f(0,1) \\ &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0) \\ &= f(1,1) = \frac{3}{14}. \end{aligned}$$

when it is clear to which random variable we refer.

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x xf(x)$$

if X is discrete, and

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

if X is continuous.

The reader should note that the way to calculate the expected value, or mean

variance & covariance of Random variables

$\mu = E(x) = \Sigma x f(x)$ → mean

$\sigma^2 = (n-\mu)^2$ → variance

$\sigma^2 = E(x^2) - [E(x)]^2$

μ^2

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \quad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \quad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

MATHEMATICAL EXPECTATION

σ^2 → variance

$V(x) = E(x^2) - [E(x)]^2$

mean → expected

$\mu = E(x) = \sum_x x f(x)$

$V(Y) = E(Y^2) - [E(Y)]^2$

$E(xy) = \Sigma\Sigma\, xy\, f(x,y)$

$\text{Covariance}(X,Y) = E(XY) - E(X).E(Y)$

$$\text{Corelation}(X,Y) = \frac{\text{Covariance}(X,Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}}$$

$\pm 1 \longrightarrow$ strong

$\pm 0.5 \longrightarrow$ moderate

$0 \longrightarrow$ negligible

nple 4.14: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

***Solution*:** We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4\, dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1-y^2)\, dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2\, dx\, dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}.$$

DISCRETE PROBABILITY DISTRIBUTION

1. Binomial Distribution

- The number X of successes in n Bernoulli trials is called a **binomial random variable**.
- The probability distribution of this discrete random variable is called the **binomial distribution.**
- The probability of a success in a binomial experiment can be computed with this formula:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

2. The Bernoulli Process

- ↳ repeated trials $\begin{cases} \text{FAIL} \\ \text{SUCESS} \end{cases}$ TWO POSSIBLE OUTCOMES
- ↳ the P(sucess) remains constant
- ↳ repeated trials are independent
- ↳ $p + q = 1$

Mean $= np$ (p → probability sucess; n → no of trials) $q = 1 - p$ (q → probability fail)

Variance $= npq \rightarrow \sigma^2$

Standard Deviation $= \sqrt{\text{variance}}$

PMF $= f(x) = {}^nC_x \, p^x q^{n-x}$

Q) A coin is tossed 4 times. find mean, variance, s.d of heads that will be obtained

$n = 4$, $p = \frac{1}{2}$ $q = \frac{1}{2}$

1. mean $= np$
   $= 4(\frac{1}{2}) = 2$

2. Variance $= npq$
   $4(\frac{1}{2})(1 - \frac{1}{2}) = 1$

3. SD $= \sqrt{\text{variance}}$
   $= \sqrt{1} = 1$

$2^4 = 16$

X	P(X)	XP(X)	X^2 P(X)
0	$\frac{1}{16}$		
1			
2			
3			
4			

$V(X) = E(X^2) - [E(X)]^2$
$= 5 - 4$
$= 1$

Q) A dice is rolled 480 times. Find the mean, variance of the no of 3s that will be rolled

$n = 480$ $p = \frac{1}{6}$ $q = \frac{5}{6}$

mean $= np$
$= 480(\frac{1}{6}) = 80$

SD $= \sqrt{npq}$
$\sqrt{480(\frac{1}{6})(\frac{5}{6})} = 8.16$

3. Hypergeometric Experiment

- Trial outcomes → FAIL / SUCESS
- P(sucess) changes each trial
- successive trials are dependent
- experiment repeated a fix no. of times

$$P(x) = \frac{{}^{a}C_x \; {}^{b}C_{n-x}}{{}^{a+b}C_n}$$

one kind of items (a), another kind of items (b), n → sample size

selecting with out replacement

a+b = total population

Q) • **Assistant Manager Applicants:** Ten people apply for a job as assistant manager of a restaurant. Five have completed college and five have not. If the manager selects 3 applicants at random, find the probability that all 3 are college graduates. → n

c: a, ${}^{5}C_x$ NC: b, ${}^{5}C_{n-x}$

$a+b = 10$, C_3

P(all 3 colleage graduates)

$$\frac{{}^{5}C_3 \; {}^{5}C_0}{{}^{10}C_3}$$

Q) • **House Insurance:** A recent study found that 2 out of every 10 houses in a neighborhood have no insurance. If 5 houses are selected from 10 houses, find the probability that exactly 1 will be uninsured.

select 5

2 → not insured

8 → insured

P(I will be insured)

$$\frac{\underbrace{{}^{8}C_4}_{\text{insured}} \; \underbrace{{}^{2}C_1}_{\text{not insured}}}{{}^{10}C_5} = \frac{5}{9}$$

I: a, ${}^{8}C_x$ NI: b, ${}^{2}C_{n-x}$ a+b $= 10$, C_5

4. Geometric Experiment

- ↳ Trial outcomes: FAIL / SUCESS
- ↳ P(sucess) remain constant for each trial
- ↳ each trial is independent
- ↳ experiment repeated a variable no. of times, until first sucess is obtained

$$g(x) = pq^{x-1}$$

mean → $\mu = \frac{1}{p}$

variance → $\sigma^2 = \frac{1-p}{p^2}$

Poisson Distribution

$$P(X) = e^{-\lambda}\left(\frac{\lambda^x}{x!}\right)$$

x → no of times event occured

λ → mean of x

- ↳ avg no of sucess
- ↳ time interval
- ↳ mean

Q) • **Radioactive Particles:** During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$\lambda = \underset{ms}{1} \times 4 = 4$

$P(x=6)$

$P(x=6, \lambda=4)$

$$e^{-4}\left(\frac{4^6}{6!}\right) = 0.1042$$

if 6 particles in 2ms

$\lambda = \underset{ms}{2}(4) = 8$

Discrete Random variable

- ↳ PMF
- ↳ CDF
- ↳ mean
- ↳ variance

Continous Random variable

- ↳ PDF
- ↳ CDF
- ↳ mean
- ↳ variance

Normal Distribution

↳ Continous type Distribution

$$\text{PDF} = f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\left(\sqrt{2\pi}\right)} \quad ; -\infty < x < \infty$$

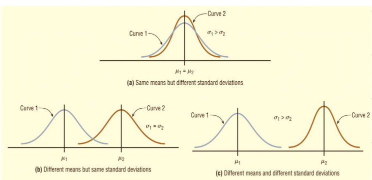
Population mean

Population sd

$\bar{X} = \tilde{X} = \hat{X}$

↓ mean ↓ median ↓ mode

Shapes of Normal Distribution



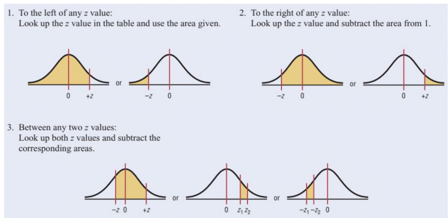
Standard Normal Distribution

↳ $\mu = 0$ and $\sigma = 1$

$$Z\text{-score} = \frac{X-\mu}{\sigma}, \quad y = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$$

$Z \sim SN(\mu = 0, \sigma = 1)$

Areas under the standard Normal Curve



Example 01

a. Find the area to the left of z = 2.06

b. Find the area to the right of z = -1.19

c. Find the area between z = 1.68 and z = -1.37

P(z < 2.06)

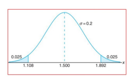
P(z > - 1.19)

P(-1.37 < z < 1.68)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
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
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

Example # 12

- Gauges are used to reject all components for which a certain dimension is not within the specification 1.50 ± d. It is known that this measurement is normally distributed with mean 1.50 and standard deviation 0.2. Determine the value d such that the specifications "cover" 95% of the measurements.



?

- $P(-1.96 < Z < 1.96) = 0.95.$

- $1.96 = \frac{\overset{\mu + d}{(1.50 + d)} - \overset{\mu}{1.50}}{\underset{\sigma}{0.2}}, = Z$

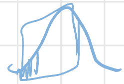
- $d = (0.2)(1.96) = 0.392.$

$\mu = 1.5 \quad \sigma = 0.2 \quad \alpha : 0.025$

AL: 0.95 + 0.025 : 0.975

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−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
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

CHP 6



9.177

6.5 Given a standard normal distribution, find the area under the curve that lies

(a) to the left of $z = -1.39$; $1 - 0.9177 = 0.0823$

(b) to the right of $z = 1.96$; $1 - 0.9750 = 0.0250$

(c) between $z = -2.16$ and $z = -0.65$; $1 - 0.9846 - 1 - 0.7422$

$0.0154 - 0.2578 = 0.2424$

(d) to the left of $z = 1.43$;

(e) to the right of $z = -0.89$;

(f) between $z = -0.48$ and $z = 1.74$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

6.6 Find the value of z if the area under a standard normal curve

(a) to the right of z is 0.3622; $1 - 0.3622 = 0.6378 \rightarrow 0.35$

(b) to the left of z is 0.1131; $1 - 0.1131 = 0.8869 \rightarrow 1.21$

(c) between 0 and z, with $z > 0$, is 0.4838; $0.5 + 0.4838 = 0.9838 \rightarrow 2.14$

(d) between $-z$ and z, with $z > 0$, is 0.9500. $0.25 + 0.95 = 0.975 \rightarrow 1.96$

$$\text{Z-score} = \frac{X - \mu}{\sigma}$$

6.8 Given a normal distribution with $\mu = 30$ and $\sigma = 6$, find

$$\frac{X - 30}{6}$$

(a) the normal curve area to the right of $x = 17$; $\frac{17-30}{6}$

(b) the normal curve area to the left of $x = 22$; $\frac{22-30}{6}$

(c) the normal curve area between $x = 32$ and $x = 41$; $\frac{32-30}{6} - \frac{41-30}{6}$

(d) the value of x that has 80% of the normal curve area to the left; $x = Z \times \sigma + \mu = 35.04$

$0.8 \times 6 + 30$

(e) the two values of x that contain the middle 75% of the normal curve area.

6.9 Given the normally distributed variable X with mean 18 and standard deviation 2.5, find

$\mu = 18 \quad \sigma = 2.5$ $\qquad \frac{X - 18}{2.5}$

(a) $P(X < 15)$; $\frac{15-18}{2.5} = -1.2$

(b) the value of k such that $P(X < k) = 0.2236$;

(c) the value of k such that $P(X > k) = 0.1814$;

(d) $P(17 < X < 21)$.

6.9 (a) $z = (15 - 18)/2.5 = -1.2$; $P(X < 15) = P(Z < -1.2) = 0.1151$.

(b) $z = -0.76$, $k = (2.5)(-0.76) + 18 = 16.1$.

(c) $z = 0.91$, $k = (2.5)(0.91) + 18 = 20.275$.

(d) $z_1 = (17 - 18)/2.5 = -0.4$, $z_2 = (21 - 18)/2.5 = 1.2$;
$P(17 < X < 21) = P(-0.4 < Z < 1.2) = 0.8849 - 0.3446 = 0.5403$.

centeral limit theorem

z test

$$Z = \frac{\bar{X} - \mu}{\underbrace{\sigma\sqrt{n}}_{\text{standard error of mean}}}$$

Formula	Use
1. $z = \frac{X - \mu}{\sigma}$	Used to gain information about an individual data value when the variable is normally distributed.
2. $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	Used to gain information when applying the central limit theorem about a sample mean when the variable is normally distributed or when the sample size is 30 or more.

Example 3

An upgrade of a certain software package requires the installation of 68 new files. Each file is installed sequentially. On average, it takes 15 seconds to install one file, with a variance of 11 sec^2.

(a) How likely is it that the whole package will be updated in less than 16 minutes?
(b) There is a new version of the package. It requires only N new files to be installed, and it is premised that 95% of the time upgrading takes less than 10 minutes. Based on this information, calculate N.

Solution:

(a) $n = 68$, $\mu = 15sec$ $Var(X) = 11sec^2$, $x = 16min = 960sec$

$$P(S_n \leq 960) \sim P[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{960 - 1020}{\sqrt{748}}] = P[Z \leq -2.20] = 0.0143$$

(b) $N = ?$, $\mu = 15sec$ $Var(X) = 11sec^2$, $x = 10min = 600sec$

$$P(S_n \leq 600) \sim P[\frac{S_n - N\mu}{\sigma\sqrt{N}} \leq \frac{600 - 15N}{\sqrt{11N}}] = P[Z \leq z] = 0.95$$

So for $N = 37$ you get the correct probability more than 95%.

Finite Population correction Factor

$$\sqrt{\frac{\overset{\text{population size}}{N} - \overset{\text{sample size}}{n}}{N - 1}}$$

ishma hafeez notes

↳ Estimation
- point estimate : $\tilde{x}, s^2, p$ — represent these points at — 0.159 → 0.16
- interval estimate — $18 < \mu < 25$ — attach probability — CONFIDENCE INTERVAL

↳ Hypothesis testing

↳ ANOVA

↳ Regression and corruption — Gallub, Pulse

Properties of good estimators

1. Unbiased estimator
   ↳ $\bar{n}$ = parameter being estimated
2. consistent estimator
3. Relatively efficient estimator
   ↳ has the smallest variance

$\bar{x} \equiv \mu$

unbiased estimator of population mean (μ)

$\bar{x}$ → statistic

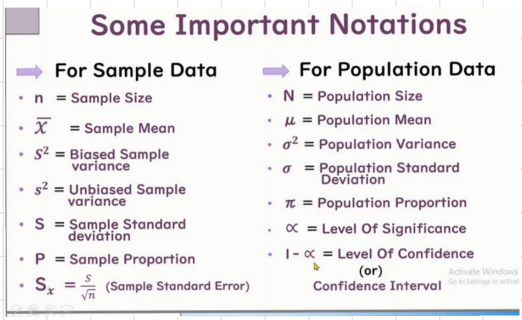
μ → parameter

$\boxed{\text{sampling error} = \bar{x} - \mu}$

Q) $90\% \rightarrow z_{\frac{(1-0.9)}{2}} = 1.65$

$95\% \rightarrow z_{\frac{(1-0.95)}{2}} = 1.96$

$99\% \rightarrow z_{\frac{(1-0.99)}{2}} = 2.58$



eg $n = 50, \bar{x} = 54, \sigma = 6$

95% of CI mean?

$54 \pm (1.96)\left(6/\sqrt{50}\right)$

$52.3 < \mu < 55.7$

95% chance that the above interval contains true population mean

PROB ASSIGNMENT K214688

Q1)

a) $f(x,y,z) = \{ ky^2z,\quad \begin{matrix} 0<x \\ y<1 \\ 0<z<2 \end{matrix}$

$$\int_0^2\int_0^1\int_0^1 ky^2z\,dx\,dy\,dz = 1$$

$$\int_0^2\int_0^1 \left|\frac{kx^2y^2z}{2}\right|_0^1 = 1$$

$$\int_0^2\int_0^1 \frac{ky^2z}{2} = 1$$

$$\int_0^2 \left|\frac{ky^3z}{6}\right|_0^1 = 1$$

$$\int_0^2 \frac{kz}{6} = 1$$

$$\left|\frac{kz^2}{12}\right|_0^2 = 1$$

$$\frac{k}{3} = 1$$

$$\boxed{k=3}$$

b) $P(x<1/4,\ y>1/2,\ 1<z<2)$

$$3\int_1^2\int_{1/2}^1\int_0^{1/4} xy^2z\,dx\,dy\,dz$$

$$3\int_1^2\int_{1/2}^1 \left|\frac{x^2y^2z}{2}\right|_0^{1/4}$$

$$3\int_1^2\int_{1/2}^1 \frac{y^2z}{32}$$

$$3\int_1^2 \left|\frac{y^3z}{96}\right|_{1/2}^1$$

$$3\int_1^2 \frac{7z}{768}$$

$$3\left|\frac{7z^2}{1536}\right|_1^2$$

$$\frac{21}{512}$$

c) Marginal

$$f(y)\ 3\int_0^2\int_0^1 xy^2z\,dx\,dz$$

$$3\int_0^2 \left|\frac{x^2y^2z}{2}\right|_0^1$$

$$3\int_0^2 \frac{y^2z}{2}$$

$$3\left|\frac{y^2z^2}{4}\right|_0^2$$

$$= 3y^2$$

$$f(z) = 3\int_0^1\int_0^1 xy^2z\,dx\,dy$$

$$3\int_0^1 \frac{y^2z}{2}$$

$$3\left|\frac{y^3z}{6}\right|_0^1$$

$$= \frac{z}{2}$$

$$E(y) = \int_0^1 3y^3$$

$$= 3/4$$

$$E(z) = \int_0^2 \frac{z^2}{2}$$

$$= 4/3$$

$$E(yz) = 3/2\int_0^2\int_0^1 yz(y^2z)$$

$$= 3/2\int_0^2\int_0^1 y^3z^2$$

$$= 3/2\int_0^2 \left|\frac{y^4z^2}{4}\right|_0^1$$

$$= 3/2\int_0^2 \frac{z^2}{4}$$

$$= 3/2\left|\frac{z^3}{12}\right|_0^2$$

$$= 1/2$$

$$E(x^2) = \int_0^1 3y^4$$

$$\left|\frac{3y^5}{5}\right|_0^1$$

$$= 3/5$$

$$E(z^2) = \int_0^2 \frac{z^3}{2}$$

$$= \left|\frac{z^4}{8}\right|_0^2$$

$$= 2$$

$$var(y) = 3/5 - (3/4)^2$$

$$= 3/80$$

$$var(z) = 2 - (4/3)^2$$

$$= 2/9$$

$$covariance = \frac{1}{2} - \frac{3}{80}\left(\frac{2}{9}\right)$$

$$= \frac{59}{120}$$

$$correlation = \frac{59/120}{\sqrt{3/80}\sqrt{2/9}}$$

$$= 11.738$$

ESTIMATION

Point estimate

sample mean

1. $\bar{x} = \frac{\sum x}{n}$ sample size

S.D

2. $S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$

standard error mean

3. $S_{\bar{x}} = \frac{s}{\sqrt{n}}$ S.D

margin of error

4. $E = \underbrace{Z_{\frac{\alpha}{2}}}_{\text{critical region of z-distribution}} \underbrace{\left(\frac{\sigma}{\sqrt{n}}\right)}_{\text{standard error of mean}}$

critical region of z-distribution

5. $n = \left(\frac{Z_{\frac{\alpha}{2}}\,\sigma}{E}\right)^2$

Interval estimate

Example # 04

- **Depth of a River:** A scientist wishes to estimate the average depth of a river. He wants to be 99% confident that the estimate is accurate within 2 feet. From a previous study, the standard deviation of the depths measured was 4.33 feet.

$$n = \left(\frac{z_{\alpha/2}\cdot\sigma}{E}\right)^2 = \left[\frac{(2.58)(4.33)}{2}\right]^2 = 31.2$$

Confidence interval for μ

1 distribution

1. $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ population S.D

↳ σ known

2. $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ sample S.D

↳ σ unknown

↳ $n \geq 30$

3. $\bar{x} \pm t_{\frac{\alpha}{2}(v)} \frac{s}{\sqrt{n}}$ sample S.D

critical value of t distribution

n-1

↳ σ unknown

↳ $n < 30$

Confidence interval for $\mu_1 - \mu_2$

2 distributions

1. $(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

↳ σ_1, σ_2 known

2. $(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

↳ σ_1, σ_2 unknown

↳ $n \geq 30$

3. $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

↳ σ_1, σ_2 unknown ?

↳ $n < 30$

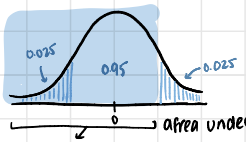
z-distribution

Question # 1

An electrical firm manufactures light bulbs that have a length of life with mean μ and a standard deviation of 40 hours. If a sample of 100 bulbs has an average life of 780 hours, Find a 95% Confidence Interval for the population mean of all bulbs produced by this firm.

$\sigma = 40$ hours $n = 100$ $\bar{x} = 780$ hours $CI = 95\%$ $\mu = ?$

① as $n > 30$ we use z test



afrea under left curve

$A_L =$ confidence level $+ \alpha$

$A_L = 0.95 + 0.025 = 0.975 \rightarrow$ find on Table

$Z_{\frac{\alpha}{2}} = 1.96$

② $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$= 780 \pm 1.96 \left(\frac{40}{\sqrt{100}}\right)$

$= 780 \pm 7.84$

$= 772.16 \; , \; 787.84$

$772.16 < \mu \leq 787.84$

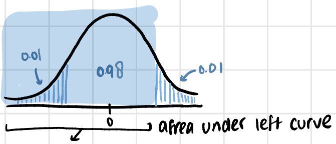
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

Question # 2

The height of a random sample of 50 college students showed a mean of 174.5 cm and a standard deviation of 6.9cm. Construct a 98% Confidence interval for the mean height of all college students.

$n = 50$, $\bar{x} = 174.5$ cm , $s = 6.9$ cm , $CI = 98\%$, $\mu = ?$

① $n > 50$ so z test



afrea under left curve

$A_L =$ confidence level $+ \alpha$

$A_L = 0.98 + 0.01 = 0.99 \rightarrow$ find on Table

$Z_{\frac{\alpha}{2}} = 2.3263$

② $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

$= 174.5 \pm 2.3263 \left(6.9 / \sqrt{50}\right)$

$= 174.5 \pm 2.27$

$= 172.23 \; , \; 176.77$

$172.33 < \mu \leq 176.77$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

→ closest to 0.99
→ take avg
→ $\frac{0.02 + 0.03}{2} = 0.025$
→ 2.325

Question # 3

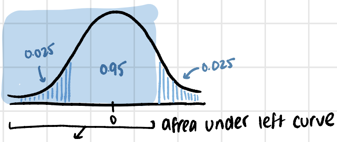
A random sample of 100 farms in a certain year gives an average yield of barley of 2100 lbs. Per acre. A random sample of 100 farms in the following years given an average yield of 2000 lbs. per acre. The S.D for two populations are 224 and 192 respectively. Compute a 95% C-I for the difference between two population means. Assuming data follows normal distribution

$n_1 = 100 \quad \bar{x}_1 = 2100 \quad CI = 95\%$

$n_2 = 100 \quad \bar{x}_2 = 2000 \quad \mu_1 - \mu_2 = ?$

$\sigma_1 = 224 \quad \sigma_2 = 192$

① as $n \geq 30$ so z test



A_L = confidence level + α

$A_L = 0.95 + 0.025 = 0.975 \rightarrow$ find on Table

$Z_{\frac{\alpha}{2}} = 1.96$

② $(\bar{x}_1 - \bar{x}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$= (2100 - 2000) \pm 1.96 \sqrt{\frac{224^2}{100} + \frac{192^2}{100}}$

$= 100 \pm 57.82$

$= 42.18 \, , \, 157.82$

$= 42.18 < \mu \leq 157.82$

Confidence Inteval for Propotions

1. $\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

sample proportion

$\frac{x}{n} \rightarrow$ no of sample units

$\hat{q} \rightarrow 1 - \hat{p}$

Example # 09

• **Covering College Costs:** A survey conducted by Sallie Mae and Gallup of 1404 respondents found that 323 students paid for their education by student loans. Find the 90% confidence of the true proportion of students who paid for their education by student loans.

$n = 1404 \, , \, x = 323 \, , \, CI = 90\%$

$P = \frac{323}{1404} = 0.23 \qquad q = 1 - 0.23 = 0.77$

$0.23 - 1.65 \sqrt{\frac{(0.23)(0.77)}{1404}}$

$= 21.1\% < p < 24.9\%$

Example #10

• **Religious Books:** A survey of 1721 people found that 15.9% of individuals purchase religious books at a Christian bookstore. Find the 95% confidence interval of the true proportion of people who purchase their religious books at a Christian bookstore.

$n = 1721 \, , \, \hat{p} = 15.9\% \, , \, CI = 95\%$?

$\hat{q} = 1 - 0.159 = 0.841$

$= 0.159 \pm 1.96 \sqrt{\frac{(0.159)(0.841)}{1721}}$

$= 0.142 < p < 0.176$

Minimum 'n' needed Interval Estimate of Population Porportion

$$\boxed{n = \hat{p}\hat{q}\left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2}$$

→ rearrange confidence interval for porportions

Example # 11

- **Home Computers:** A researcher wishes to estimate, with 95% confidence, the proportion of people who own a home computer. A previous study shows that 40% of those interviewed had a computer at home. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

CI = 95%, $\hat{p}$ = 40%, E = 2%

$\hat{q}$ = 1 − 40 = 60%

$n = (0.4)(0.6)\left(1.96/0.02\right)^2$

$= 2304.96$

Example # 12

- **M&M Colors:** A researcher wishes to estimate the percentage of M&M's that are brown. He wants to be 95% confident and be accurate within 3% of the true proportion. How large a sample size would be necessary?

CI = 95%, E = 3%

$\hat{p}$ = 5% $\hat{q}$ = 5%

$n = (0.5)(0.5)\left(1.96/0.03\right)$

$= 1067.1$

Chi square Distribution

↳ random sample

↳ normal distribution

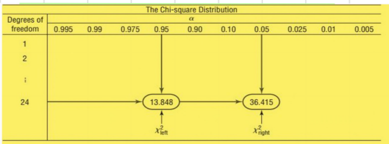
1. $\dfrac{(n-1)s^2}{\chi^2_{right}} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2_{left}}$

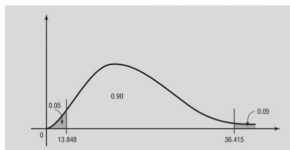
2. $df = n-1$

Example # 14

Find the values for χ^2_{right} and χ^2_{left} for a 90% confidence interval when $n = 25$.

$X_{right} = \dfrac{1-0.90}{2} = 0.05$, $X_{left} = 1-0.05 = 0.95$, $df = 25-1 = 24$





$\chi^2_{left} = 13.848$ $\chi^2_{right} = 36.415$

Example # 16

• **Cost of Ski Lift Tickets** : Find the 90% confidence interval for the variance and standard deviation for the price in dollars of an adult single-day ski lift ticket. The data represent a selected sample of nationwide ski resorts. Assume the variable is normally distributed.

59	54	53	52	51
39	49	46	49	48

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

.

$n = 10$, CI: 90

$X_{right} = \dfrac{1-0.90}{2} = 0.05$, $X_{left} = 1-0.05 = 0.95$, $df = 9$

$\chi^2_{left} = 16.919$ $\chi^2_{right} = 3.325$?

$\dfrac{(10-1)(28.2)^2}{16.919} < \sigma^2 < \dfrac{(10-1)(28.2)^2}{3.325}$

$15.0 < \sigma^2 < 76.3$

S.D

$S = \sqrt{\dfrac{\Sigma x^2}{n} - \left(\dfrac{\Sigma x}{n}\right)^2}$

$= \sqrt{\dfrac{25254}{10} - \left(\dfrac{500}{10}\right)^2}$

$= 5.039$

Example # 15

- **Nicotine Content:** Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams.

CI = 95% , n = 20 , s.d = 1.6

$X_{right} = \frac{1-0.95}{2} = 0.025$, $X_{left} = 1-0.025 = 0.975$, df = 19

$X^2_{right} = 32.852$, $X^2_{left} = 8.907$

$$\frac{(20-1)(1.6)^2}{32.852} < \sigma^2 < \frac{(20-1)(1.6)^2}{8.907}$$

$$1.5 < \sigma^2 < 5.5$$

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

CHP 9

$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

9.3 Many cardiac patients wear an implanted pacemaker to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 inch and an approximately normal distribution, find a 95% confidence interval for the mean of the depths of all connector modules made by a certain manufacturing company. A random sample of 75 modules has an average depth of 0.310 inch.

$\sigma = 0.0015$ $CI = 95\%$

$n = 75$ $\bar{x} = 0.310$

$\frac{1-0.95}{2} = 0.025$

$A_L = 0.025 + 0.95$

$= 0.975$

$P \sim 1.96$

$0.310 \pm 1.96\left(\frac{0.0015}{\sqrt{75}}\right)$

$0.309 < \mu < 0.3103$

9.6 How large a sample is needed in Exercise 9.2 if we wish to be 96% confident that our sample mean will be within 10 hours of the true mean?

$n = ?$ $CI = 96$ $\bar{x} = 10$ $\sigma = 40$ $E = 10$

$\frac{1-0.96}{2} = 0.02$

$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$A_L = 0.98$

$P = 2.055$

$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{E}\right)^2$

$n\left(\frac{2.055 \times 40}{10}\right)^2 = 68$ round up

9.35 A random sample of size $n_1 = 25$, taken from a normal population with a standard deviation $\sigma_1 = 5$, has a mean $\bar{x}_1 = 80$. A second random sample of size $n_2 = 36$, taken from a different normal population with a standard deviation $\sigma_2 = 3$, has a mean $\bar{x}_2 = 75$. Find a 94% confidence interval for $\mu_1 - \mu_2$.

$n_1 = 25$ $\bar{x}_1 = 80$ $\sigma_1 = 5$ $CI = 94$

$n_2 = 36$ $\bar{x}_2 = 75$ $\sigma_2 = 3$

$\frac{1-0.94}{2} = 0.03$

$A_L : 0.97$

$P = 1.881$

$80 - 75 \pm 1.885\sqrt{\frac{5^2}{25} + \frac{3^2}{75}}$

$2.9 < \mu_1 - \mu_2 < 7.1$

ishma hafeez notes

repsht