Statistics Notes

إعداد :

zenon

مجموعة من طلبة الهندسة - الجامعة الأردنية

بالتعاون.. نمضي..

Subject Statistics Date 22/9. No.

⟹ Number of classes: from 5 to 15

when all classes has the same length,

$$\text{class length} = \frac{\text{Maximum observation} - \text{Minimum obs}}{\text{no. of classes}}$$

تقريب للأعلى

rounded from above to the nearest number that has the accuracy unit of the data.

Example:-

Find the class length in the following cases if no. of classes equales 7 obs = مشاهدة

[a] Max. obs. = 37 , Min. Obs = 11

[b] " " = 37.0 , " " = 11.0

[c] " " = 37.00 , " " = 11.00

Solution:-

a- $\frac{37-11}{7} = 3.714$

∴ class length 4

b- $\frac{37-11}{7} = 3.71$

∴ class length = 3.8

c- $\frac{37-11}{7} = 3.7142$

∴ class length = 3.72

Constructing the Classes:-

exm.:- construct the first 3 classes in each of the following cases:-

(a) Min. obs. = 6, class length = 4

(b) Min. obs = 6.0, " " = 4.0

(c) Min. obs = 6.00, class length = 4.00

Solu:-

[a] 1st class: Min. obs To Min. obs + class length − accuracy unit

Secound 6 − 6 + 4 − 1 = 9

T 10 − 10 + 4 − 1 = 13

Third → 14 − 14 + 4 − 1 = 17

[b] 1st class: 6.0 — 6.0 + 4.0 − 0.1 = 9.9

2nd :- 10.0 − 10.0 + 4.0 − 0.1 = 13.9

3rd :- 14.0 − 14.0 + 4.0 − 0.1 = 17.9

h.w [c]

Example:- Find the length of the class 4.3 — 8.7

Sol:- class length = Right-handside − left-handside + Accuracy unit

طول الفترة = 8.7 − 4.3 + 0.1 = 4.5

A typical of Frequency Table:

class	Frequancy	Relative Frequancy	Midpoind	Cumulative Freq	التراكم النسبي
4 − 8	3 التكرار	3/10 = 0.3	$\frac{8+4}{2} = 6$	3	3/10
9 − 13	4	4/10 = 0.4	6 + 5 = 11	4 + 3 = 7	⇒
14 − 18	2	2/10 = 0.2	11 + 5 = 16	2 + 7 = 9	
19 − 23	1	1/10 = 0.1	16 + 5 = 21	1 + 9 = 10 = Total.	
Total	10	1.0		Freqnany	

Subject Date No.

الحدود الفعلية Actual limits

$4 - \frac{1}{2} * \text{accuracy unit} = 3.5$ — (less than) $8 + \frac{1}{2} * \text{accuracy} = 8.5$

8.5 - 13.5

13.5 - 18.5

18.5 - 23.5

Example:

Consider the following table

class	Cum. Freq
4.2 - 8.3	3
8.4 - 12.5	10
12.6 - 16.7	15
16.8 - 20.9	25

[a] write down the actual limits of the first class

Ans. $4.2 - \frac{1}{2} * 0.1 = 8.3 + \frac{1}{2} * 0.1$

4.15 - 8.35

[b] Find the relative freq. of the 3rd class

Ans: $\frac{15-10}{25} = \frac{5}{25} = 0.2$

[c] Find the proportion of observation that are greater than or equal to 8.4 and less than or equal to 16.7

Ans. $\frac{(10-3)+(15-10)}{25} = \frac{12}{25} = 0.48$

Tunis

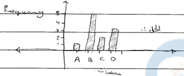
14-1-2011

1 Bar charts :-

Bar charts can be used to represent all types of data

Example:- sketch the bar chart of the following table

X	Frequency
A	1
B	5
C	2
D	3



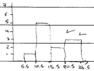
2 Histograms & Polygons :-

can be used to represent grouped data.

Example:

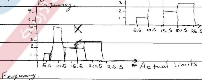
sketch the histogram and the polygons of the following table

class	Frequency
6 - 10	1
11 - 15	5
16 - 20	2
21 - 25	3
Total.	11

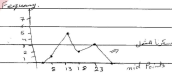


Solution:-

1) histograms



2) Polygons:



Subject Date 9/2010 No.

Measures of Central Tendancy. مقاييس النزعة المركزية

[1] Arithmetic Mean :-

⇒ The mean of a sample is denoted by $\bar{X}$

⇒ The mean of population is denoted by M

Note :- $\bar{X}$ is a random variable but M is not

Rule :(1) $\text{Mean} = \dfrac{\text{Sum of data}}{\text{no. of observation}}$

For Frequency tables

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$$

x_i = Observation (or midpoint)

f_i = its frequency (or class frequency).

Example:-

Find the mean for each of the following.

[1] 3, 2, 5, 6

[2]

x	Freq	
3	2	6
5	1	5
6	2	12
8	5	40

[3]

class	Freq	mid
0 - 4	4	2
5 - 9	5	7
10 - 14	1	12

⇒ [3]

x_i = midpoint	f_i	$x_i f_i$
2	4	8
7	5	35
12	1	12
Total	10	55

$\text{mean} = \frac{55}{10} = 5.5$

Solu :- [1] $\bar{X} = 3+2+5+6/4 = 4$

[2]

x_i	f_i	$x_i f_i$
3	2	6
5	1	5
6	2	12
8	5	40
Total	10	63

$\text{mean} = \frac{63}{10} = 6.3$

Example:- Consider the following data:-

X	freq
3	2
5	1
6	a
8	5

$\Rightarrow$ Is the mean is 6.3, find a

Solu:-

X_i	f_i	$X_i f_i$
3	2	6
5	1	5
6	a	6a
8	5	40
Total	8+a	51+6a

mean = 6.3

$\therefore \frac{51+6a}{8+a} = 6.3$

$(8+a)6.3 = 51 + 6a$

$\boxed{a = 2}$ ※.

Ex :- A student at the UJ finished 96 credit hours with cumulative average 3.2 he wishes to raise his " " To 3.3 through 15 credit hours next semester. Is this possible? If yes, find the average that he should get next semester.

Solu:- total points so far is $3.2 \times 96 = 307.2$

If next semester average is A, then

$15 \times a + 307.2 = (3.30)(96 + 15) = 3.30 \times 111 = 366.3$

$\therefore a = \frac{366.3 - 307.2}{15} = 3.94$ So, yes it is possible with 3.94 average.
at least

الوسيط The median is the number that is located in the middle of the data after sorting

Thus (about) 50% of the Observation are less than it

⇒ The rule :- suppose that $x_1 \le x_2 \le \dots \le x_n$ represent our sorted data. Then

$$\text{median} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \quad \text{فردي} \\ \dfrac{x_{\frac{n}{2}} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is even} \quad \text{زوجي} \end{cases}$$

Example :-

Find the median for each of the following :-

[1] 3, 2, 3, 3, 1, 5, 7

[2] 5, 2, 7, 2, 2, 4

[3]

x	Frequency	Cu.Fre
1	3	3
2	6	9
5	11	20
7	8	28
10	4	32
11	3	35
13	3	38
Total	38	

$\dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

$\dfrac{x_{19} + x_{20}}{2}$

$\dfrac{5+5}{2}$

Solu :- [1] sort the data :- $\underset{x_1}{1}, \underset{x_2}{2}, \underset{x_3}{3}, \underset{x_4}{\boxed{3}}, \underset{x_5}{3}, \underset{x_6}{5}, \underset{x_7}{7}$ (median)

$n = 7$ odd ⇒ median $= x_{\left(\frac{7+1}{2}\right)} = x_4 = 3$

[2] sort ... $2, 2, \boxed{2, 4}, 5, 7$

median $= \frac{2+4}{2} = 3$

$n = 6$, even So median $= \dfrac{x_3 + x_4}{2} = \dfrac{2+4}{2} = 3$

[3] No need for sorting Data

$n = 38$ Observation, which is even

$\therefore$ median $= \dfrac{\frac{X_{38}}{2} + X_{\frac{38}{2}+1}}{2} = \dfrac{X_{19} + X_{20}}{2} = \dfrac{5+5}{2} = 5$

$\Rightarrow$ Mode :- المنوال For qualititative

A mode is an Observation (or mid point) with the highest Frequency.

Example:- In [3] at the previous example, 5 is the only mode.

$\Rightarrow$ Percentiles :- المئينات

let $0 < P < 1$ Then by P_{100P} we denote the $100p^{th}$ percentile.

It is the number such that (about) 100 p% Observation are less than it

Example:- $P_{25} = 70$ mean that about 25% of the Observations are less than 70.

$\Rightarrow P_{40} = 30 \neq P_{30} = 40$

$\Rightarrow$ Quartiles الربيعيات 1st quartile $Q_1 = P_{25}$

structure :- 2nd " $Q_2 = P_{50}$ = median

3rd " $Q_3 = P_{75} =$

Finding Percentiles:

For Frequency tables without classes:

⇒ The rule is:

$$P_{100} = \begin{cases} \lceil np \rceil \text{ if } np \text{ is not an integer} \quad \text{ceiling} \\ \dfrac{X_{(np)} + X_{(np)+1}}{2}, \text{ if } np \text{ is an integer (عدد صحيح)} \end{cases}$$

Ex: $\lceil 3.1 \rceil = 4$, $\lceil 3.9 \rceil = 4$, $\lceil 4 \rceil = 4$ ceiling.

Example: consider the following data;

X	Freq	
3	4	4
5	2	6
7	6	12
9	2	14
Total	14	

Find: [1] Median = P_{50}

[2] Q_1

x_4 $\frac{x_7 + x_8}{2}$ P_{25}

$14 \times 0.25 = 3.5$

x_4

Solu: [1] median = P_{50}

$P = 0.5 \Rightarrow P = \frac{50}{100}$

$np = 0.5 \times 14 = 7$, integer

$\therefore$ median $= \frac{X_7 + X_8}{2} = \frac{7 + 7}{2} = 7$

[2] $P_{25} = ?$

$P = \frac{25}{100} = 0.25$

$np = 14 \times 0.25 = 3.5$ not an integer

$\therefore P_{25} = X_{\lceil 3.5 \rceil} = X_4 = 3$.

We use "linear interpolation" to estimate percentiles for grouped data.

Main idea:- think of $n \times p$ to be the cumulative frequency of P_{100p}

Example:- Consider the following data

Find:-

1. the median Q_{50}
2. Q_1
3. the proportion of observations that are less than 16.

class	Frequency	
4 - 8	3	3
9 - 13	5	8
14 - 18	4	12
19 - 23	1	13
Total	13	

Solu: 1. n = no. of observations = 13

$\Rightarrow$ step 1: Find the percentile class.

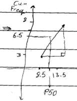
It's the first class that has cumulative frequency $\geq np$

In our question, median = P_{50}, 50 $\quad P = \frac{50}{100}$

$np = 13 \times \frac{50}{100} = 6.5$

The perc. class of P_{50} is 9 - 13

$\Rightarrow \dfrac{8-3}{13.5-8.5} = \dfrac{6.5-3}{P_{50}-8.5}$ $\Rightarrow$ step 2 $\Rightarrow$



Cumulative

Actual limit

$1 = \dfrac{3.5}{P_{50}-8.5}$

$P_{50} = 8.5 + 3.5 = \boxed{12}$

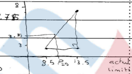
$\rightarrow$ Actual values at perc. class

[2] 1- Percentile class at $Q_1 = P_{25}$ H.W → P_{80}

$P = \frac{25}{100}$, $n = 13$ $np = \frac{25}{100} \times 13 = 3.25$

So perc. class is 9-13

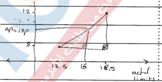
$\frac{5}{5} = \frac{3.25-3}{P_{25}-8.5} \Rightarrow 1 = \frac{0.25}{P_{25}-8.5} \Rightarrow P_{25} = 8.75$



[3] To find p such that $16 = P_{100p}$

Then, The required proportion is p.

16 belongs to the class 14-18



$\frac{4}{5} = \frac{13p-8}{2.5}$

$10 = 5(13p-8)$

$p = \frac{10}{13} = 0.77$

⟹ Measures of Dispersion 80   مقاييس التشتت

(1) Range :-

Range = Max.obs − Min.obs. $Q_3 - Q_1$

(2) Interquartile range. (IQR) :-   المدى الربيعي

$IQR = Q_3 - Q_1$

(3) ⇒ Notice that $[0 \le IQR \le Range]$.

(3) Variance.   التباين

For population it's denoted by σ^2

For a sample " " " S^2

Subject Date 4/10/2010 No.

$$\Rightarrow \sigma^2 = \frac{\sum_{i=1}^{N}(x_i - M)^2}{N}$$

$$\Rightarrow S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

For knowledge.

For computitional purposes :-

$$\Rightarrow S^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{X}^2}{n-1}$$

For Frequancy tables :-

$$\Rightarrow S^2 = \frac{\sum_{i=1}^{k} x_i^2 f_i - (\sum_{i=1}^{k} f_i)\bar{X}^2}{(\sum_{i=1}^{k} f_i) - 1}$$

$$\Rightarrow \bar{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

Example :-

Find S^2 for each of the followings :-

[1] 3, 5, 4, 2

[2]

class	Freq	mid x_i
0-4	3	2
5-9	2	7
10-14	1	12
	6	

[1] Solu:-

X	X^2
3	9
5	25
4	16
2	4
Total 14	54

$$S^2 = \frac{54 - 4\left(\frac{14}{4}\right)^2}{4-1} = \boxed{} \ \#$$

[2]

X_i = midpoint	f_i	$x_i f_i$	X_i^2	$X_i^2 f_i$
2	3	6	4	12
7	2	14	49	98
12	1	12	144	144
Total	6	32		254

$$S^2 = \frac{254 - 6 * \left(\frac{32}{6}\right)^2}{6-1}$$

X | Rel freq

Median

X | freq

Standard Deviation:-

$= \sqrt{\text{Variance}}$

$S = \sqrt{S^2}$

$\sigma = \sqrt{\sigma^2}$

$S = \sqrt{S^2}$

$\Rightarrow$ Chebyshev's Rule 80 6/10/2010

[1] At least $1 - \frac{1}{2^2} = \frac{3}{4}$ of the Observations are between $\bar{X} - 2S$ and $\bar{X} + 2S$

[2] At least $1 - \frac{1}{3^2} = \frac{8}{9}$ of the Observation are between $\bar{X} - 3S$ and $\bar{X} + 3S$

[3] In general for any $K > 1$, at least $1 - \frac{1}{K^2}$ of the Observ's are between $\bar{X} - KS$ and $\bar{X} + KS$.

Example: The grades of 1000 students have mean $\bar{X} = 60$ and $S = 10$.

[1] At least how many students got grades between 40 and 80.

[2] At most how many students got grades greater than 90 or less than 30.

[3] Find an Interval that contains ~~at least~~ the grades of at least 70% of the students.

Sol.: [1] $40 = \bar{X} - kS = 60 - k*(10) \Rightarrow k = 2$

$1 - \frac{1}{2^2} = \frac{3}{4}$ of 1000 students got grades between 40 and 80

$\frac{3}{4} * 1000 = 750$ students.

[2] $30 = \bar{X} - kS = 60 - k(10) \Rightarrow k = 3$

$\therefore$ At least $1 - \frac{1}{3^2} = \frac{8}{9}$ of the students got grades between 30 and 90. $\therefore$ At most $\frac{1}{9}$ of the 1000 students got grades outside this interval.

$\frac{1}{9} * 1000 \approx 111$ students

[3] $1 - \frac{1}{k^2} = \frac{70}{100} \Rightarrow \frac{1}{k^2} = \frac{30}{100} \Rightarrow k^2 = \frac{10}{3}$

$\overset{k>0}{\Rightarrow} k = \sqrt{\frac{100}{3}} = 1.82$

$\therefore$ the required interval is $(\bar{X} - kS, \bar{X} + kS) =$

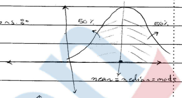
$= (60 - (1.82)*10), 60 + 1.82*10)$

$= (41.8, 78.2)$

Subject Follow:- Date 6/10/2010 No.

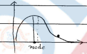
Bell-Shaped Distribution, Emperical Rules:-

⇒ Bell-shaped Distributions:



Skewed to the right:

⇒ mode < median < mean ⇒

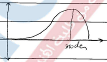


2:20

Skewed to the left:

mean < median < mode ⇒

left → mode
left mean

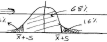


Omniah

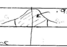
Emperical Rule:

For bell-shaped distribution the following are true:

[1] about 68% of the Observations are between $\bar{X}-S$ and $\bar{X}+S$.



[2] About 95% of the Observations are between $\bar{X}-2S$ and $\bar{X}+2S$



[3] About 99% of the Obs. are between $\bar{X}-3S$ and $\bar{X}+3S$

Remark: We deduce that for bell-shaped data,

$P_{84} = \bar{X} + S$

$P_{16} = \bar{X} - S$

Example:

The weights of children at a certain age are bell-shaped with mean $\bar{X} = 30$ kg, and standard deviation 4 kg

[1] Estimate the range of the weights.

[2] Find the percentage of children with weights between 26 and 38. $26 = 30 - 4$

[3] Find P_{84} for the weights.

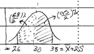
[4] Find the proportion of children with weights less than 18 kg $3S$ $P_{1000} = 18$

$18 = 30 - 12$ $\bar{X} - 3S$

Solu: [1] 99% of the children have weights between $\bar{X} - 3S$ and $\bar{X} + 3S$ Thus

$Range \approx \bar{X} + 3S - (\bar{X} - 3S) = 6S = 6 * 4 = 24$

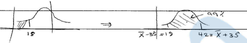
[2] $(\frac{68}{2})\% + (\frac{95}{2})\% = 81.5\%$ are between 26 & 38



[3] $P_{84} = \bar{X} + S = 30 + 4 = 34$.

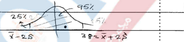
↘ [4] $18 = \bar{X} - 3S$.

∴ Proportion of children with weights < 18 = $\boxed{0.5\%}$



Extra question:

what is the percentage of children with weights < 38 kg?



As: 95% + 2.5% = <u>97.5%</u>

Examples:

Imp: The grades are bell-shaped with $Q_1 = 30$, $IQR = 50$

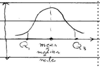
Find the <u>mean</u> of the grades?

Solu: $mean = \frac{Q_3 + Q_1}{2}$ (For bell-shaped)

$Q_3 = IQR + Q_1$

$= 50 + 30 = 80$

Thus $mean = \frac{80+30}{2} = 55$



Mean 4 Bellshaped $\frac{Q_3+Q_1}{2}$

Recall:-

[1] $\bar{X} = \frac{\sum X}{n} \Rightarrow \sum X = n * \bar{X}$.

[2] $S^2 = \frac{\sum x^2 - n\bar{X}^2}{n-1}$

$\Rightarrow \sum X^2 = (n-1)S^2 + n\bar{X}^2$ #

Example:-

suppose that we have two samples with the following measures.

	Size	Mean	Variance
sample A	$n_A = 50$	$\bar{X}_A = 60$	$S^2_A = 200$
sample B	$n_B = 20$	$\bar{X}_B = 35$	$S^2_B = 150$

[1] The two samples are combined to produce a new sample, say sample C. Find $\bar{X}_C$, S^2_C

[2] An Observation $X = 50$ is deleted from sample A. Find the new values of [1] (i) n_A 49, (ii) $(\sum X)_A$, iii $(\sum x^2)_A$

[3] An observation $X = 50$ is added to sample A, Find new $(\sum X^2)$

[4] " " $X = 40$ in sample B is corrected to 100, Find ① new(n_B) ② new$(\sum x)$, new$(\sum X^2)$

Solu:- [1] $\bar{X}_C = \frac{(\sum x)_C}{n_C} = \frac{(\sum X)_A + (\sum X)_B}{n_A + n_B}$

$= \frac{n_A * \bar{X}_A + n_B * \bar{X}_B}{n_A + n_B} = \frac{50 * 60 + 20 * 35}{50 + 20} = 52.8$

$S^2_C = \frac{(\sum X^2)_C - n_C (\bar{X}_C)^2}{n_C - 1} \Rightarrow$

$\Rightarrow (\Sigma x^2)_C = (\Sigma x^2)_A + (\Sigma x^2)_B$

$(\Sigma x^2)_A = (n_A - 1) S_A^2 + n_A (\bar{x}_A)^2$

$= 49 * 200 + 50 * 60^2 = 12800$

$(\Sigma x^2)_B = 19 * 150 + (20 * (35)^2 = 4095$

substitute $(\Sigma x^2)_C$ in S_C^2 and get the answer

$(\Sigma x^2)_C = \cancel{16}$ ----

$\boxed{2}$ New n_A = old $n_A - 1 = 50 - 1 = 49$

(ii) new $(\Sigma x)_A$ = old $(\Sigma x)_A - 50$

iii new $(\Sigma x^2)_A$ = old $(\Sigma x^2)_A - 50^2$

$\boxed{3}$ new $(\Sigma x^2)_A$ = old $(\Sigma x^2)_A + 50^2$

$\boxed{4}$ ~~new~~ Correct (n_B) = old (n_B)

$= 20$

ARASNA #1

Correct $(\Sigma x)_B$ = old $(\Sigma x)_B - 40 + 100$

correct $(\Sigma x^2)_B$ = old $(\Sigma x^2)_B - 40^2 + 100^2$

$\frac{\Sigma x_c^2 - n_c \bar{x}_c^2}{n_c - 1}$

let $y = ax + b$, $a \neq 0$

Then:

⇒ [1] $\bar{y} = a\bar{x} + b$.

⇒ [2] $S_y^2 = a^2 S_x^2$.

⇒ [3] $S_y = |a| S_x$

⇒ [4] Range of $y = |a| *$ Range of X

⇒ [5] IQR of $y = |a| *$ IQR of X.

⇒ [6] P_{100p} of $y = \begin{cases} a * P_{100p} \text{ of } X + b, & \text{if } a > 0 \\ a * P_{100-100p} \text{ of } X + b, & \text{if } a < 0 \end{cases}$

Example:

let $y = ax + b$, $a > 0$

Suppose that $\bar{y} = 30$, $S_y = 20$, $S_x = 10$, $\bar{X} = 10$

Find [a] and [b]

Solu:

$\bar{y} = a\bar{x} + b \Rightarrow 30 = a(10) + b \rightarrow ①$

$S_y = |a| S_x = a S_x$

$\Rightarrow 20 = a * 10 \Rightarrow a = 2$ ②

Subs in ① to get $20 + b = 30 \Rightarrow b = 10$

Example:

let $y = 1 - 3X$, suppose that $\bar{y} = 50$, Q_1 of $X = -8$, $S_y = 10$

IQR of $X = 20$. find [1] $\bar{X}$ [2] Q_1 of y.

Solu:-

[1] $\bar{y} = 1 - 3\bar{x} \Rightarrow 50 = 1 - 3\bar{x} \Rightarrow \bar{x} = -\frac{49}{3}$

[2] Q_1 of $y = 1 - 3(Q_3$ of $x)$ $\Rightarrow Q_1 = Q_3$

note :- [$\Rightarrow$ IP + 3 Q_1]

Q_3 of x - IQR of x + Q_1 of x = $20 - 8 = 12$

$\therefore Q_1$ of $y = 1 - 3 * 12 = -35$ #

Ch. 2 // Elements of Probability :-

RANDOM Experiments :-

A Random Eexperiments is an experment whose outcome is a random variable.

Sample space:- The sample space S or (Ω) is the collection of all possible values of the outcome of the random experiments.

Events :- A subcollection of S is called an event.

Example:-

Consider the experiment of rolling a dice two times

[1] the sample space of this experiments is $S = \{(1,1)(1,2), \cdots, 6.6\}$

$= \{(a,b) : 1 \leq a \leq 6, 1 \leq b \leq 6\}$

[2] the following are events in this experiments:

[1] S (This is called the certin event)

[2] $\varnothing$ (" " " impossible ")

[3] $A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\} = \{(a,b): a=b\}$

(4) $B = \{(a,a): 1 \leq a \leq 6\}$

[4] $B = \{(a,b): a+b \leq 3\} = \{(1,1),(1,2),(2,1)\}$

Equally-likely Experiments:

An experiment is equally-likely if:

[1] S is finite, i.e.

$$S = \{x_1, x_2, \dots, x_n\}.$$

[2] Every element in S has the same "chance" to occure. نسبة الحدوث

⇒ Probability of an event:

The Probability of an event E is denoted by $P[E]$

It is a number between $[0,1]$ That measures the likelyhood of the event E.

Finding $P(E)$:

If the experiment is equally-likely then

$$P(E) = \frac{\text{no. of elements in } E}{\text{no. of } \quad " \quad " \ S} = \frac{|E|}{|S|}$$

Example:- One ball is drawn from a box that contains 3 red and 1 black balls.

let E = drawing a red ball.

Find $P(E)$

Solu:- $S = \{Red1, Red2, Red3, Red4, Black\}$

$E = red = \{Red1, Red2, Red3\}$

$P(E) = \frac{|E|}{|S|} = \frac{3}{4}$

Example:- A pair of fair dice is rolled by time. Find the probability that the total number of dots is ≤ 3.

Soluso

$S = \{(1,1)(1,2) \ldots (6,6)\}$

$= |S| = 36$

$E = \{(1,1),(1,2)(2,1)\}$

$|E| = 3$

$P(E) = \frac{3}{36} = \frac{1}{12}$

Rules of Counting:-

[1] The m×n Rule:-

suppose that a project consists of k stages and that

stage 1 can be performed or completed in m_1 ways

stage 2 " " " " " m_2 "

stage k " " " " m_k "

⇒ then the whole project can be completed in $m_1 * m_2 * \ldots * m_k$ ways

Ex:- In how many ways can we roll a fair die 4 times?

Ans:- $\boxed{6} \times \boxed{6} \times \boxed{6} \times \boxed{6} = 6^4$

Ex:- In how many ways no two rolls show the same number of dot? $\boxed{6} \times \boxed{5} \times \boxed{4} \times \boxed{3}$

Subject Date 18/10/2010 No.

[3] A fair die is rolled 4 times, Find the Probability that no two rolls show the same number of dots.

Ans: $\frac{6\times5\times4\times3}{6^4} = \frac{5}{18}$

$\Rightarrow$ if it's 7 times $= 0$

Example:

Suppose that there are 4 bus lines From $A \to B$ And 2 Bus Lines From B To C. A man travelles the trip $A \to B \to C \to B \to A$

Find the Probability that the man did not use the Same bus line twice

Ans $\frac{4\times2\times1\times3}{\boxed{4}\times\boxed{2}\times\boxed{2}\times\boxed{4}} = \frac{3}{8}$ #

Example:

A meal consists of a main dish, a soup and a dessert. Suppose that there are 5 choices for the 5 main dishes, namely M_1, M_2, M_3, M_4, M_5

2 choices for the soup, namely S_1, S_2 and 3 choices for the dessert, namely D_1, D_2, D_3

[1] How many different meals can we have?

$\Rightarrow 5\times2\times3 = 30$

[2] A person ordered A meal at random what is the Probability that she ordered M_1, S_2, D_3?

$\frac{1}{30}$ #

N O T E B O O K

⇒ Permutations : التباديل : line - shelf

To Permute mean to organize taking order into account.

The Rules :

[1] the number of permutation of n distinct (مختلفين) Objects is $n! = n \times (n-1) \times \cdots \times 2 \times 1$.

⇒ Example :

How many Permutation are there for the letters a, b, c, d?

Ans $4! = 4 \times 3 \times 2 \times 1 = 24$.

[2] the number of permutation of n objects such that $n_1, n_2 \cdots n_k$ objects are alike is :

$$\frac{n!}{n_1!\, n_2! \cdots n_k!}$$

Example :

How many Permutation are there for a, a, a, b, b, c, c, c, c, d, e, f.

Ans : $\frac{12!}{3!\,2!\,4!}$

[3] The number of permutation of n distinct objects taken r at atime is

$${}_rP^n = n \times (n-1) \times \cdots (n-r+1).$$

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

Subject الدرس الرابع Date 18/10/2010 No.

Example:

How many 3 distinct - letter words can we Form From the letters a, b, c, d, e, f, g, h?

Ans: $^{8}P_{3} = 8 \times 7 \times 6 = 336$

without distinct $\Rightarrow 8 \times 8 \times 8 = \boxed{}$

Example:

A bus contains 5 white, 3 black, 4 red and 2 blue balls. 4 balles are drawn from the box One at a time without replacement. Find the probability that the first ball is black and the second is red.

Ans. $\dfrac{3 \times 4 \times 12 \times 11}{\boxed{14} \times \boxed{13} \times \boxed{12} \times \boxed{11}} = \dfrac{6}{91}$ # سحب بدون إرجاع

Example:

Suppose that we want to put the books, maths 1, maths 2, maths 3, physic 1, physic 2 On a shelf, Find the probability that Only the math books will come next to each other?

Sol $\dfrac{3!\ 2!}{5!} = \dfrac{1}{10}$ without Only $\Rightarrow \dfrac{3!\ 3!}{5!}$

N O T E B O O K

Subject: Date 18/10/2010 No.

Combination: توافيق

Combination means "to organize without taking order into account

Rule:

The number of combinations of n distinct Objects taken r at a time is $C_r^n = \binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{(n-r)!\,r!}$

Example:

In how many ways can we select a team of 3 students from 10 students?

Ans: $C_3^{10} = \frac{10!}{3!\,7!} = \frac{10 \times 9 \times 8 \times 7!}{7! \times 3 \times 2} = 120$ ways.

Example:

A box contains 5 white, 4 black, 3 red and 2 blue balls. 6 balls are drawn from the box.

Find the Probability of having 2 white balls?

Ans: $\frac{C_2^5 \, C_4^9}{C_6^{14}}$

Probability Rules:

1) $P(\text{not } A) = P(\bar{A}) = 1 - P(A)$ حادث مع متممة

2) $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B) = P(A) + P(B) - P(A \cap B)$

3) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

4) $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$

5) $P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$

NOTEBOOK

Conditioned Probability :- ⇒

When we write $P(A|B)$ we read "Probability of A given "B" it means the Probability that A will occur if we know that B has occured.

Example:

In A class students are distributed according to the following table :-

⇒ A student is randomly selected from this class. Find

	Math Major	Physics major	
Female	30	20	50
Male	10	8	18
total	40	28	68

[1] P(student is a male.)

Ans: $\frac{18}{68}$ *

[2] P(" " Female) ⇒ Ans $\frac{50}{68}$

[3] P(" " a male or major is math)

Ans $P(male) + P(math) - P(male \cap math)$

$= \frac{18}{68} + \frac{40}{68} - \frac{10}{68} = \frac{48}{68}$ #

[4] P(male and not math) ⇒ Ans :- $P(male \cap \overline{math})$

$= P(male) - P(male \cap math)$

$= \frac{18}{68} - \frac{10}{68} = \frac{8}{68}$

Or directly $P(male \cap \overline{math}) = P(male \cap Physics) = \frac{8}{68}$

[5] $P(male | math) = \frac{10}{40}$

[6] $(math | male) = 10/18$

More Rules :

$\boxed{1}$ If $A \subseteq B$ then $P(A) \leq P(B)$ ⇐ اذا حصل A يعني B حاصل

$\boxed{2}$ $P(\bar{A}|B) = 1 - P(A|B)$

$\boxed{3}$ $P(\bar{A} \cap \bar{B})(C) = P(A|C) - P((A \cap B)|C)$

$\boxed{4}$ $P((A \cup B)|C) = P(A|C) + P(B|C) - P((A \cap B)|C)$.

$\boxed{5}$ $P(A \cap B) = P(A)\, P(B|A)$.

$P(B|A) = \dfrac{P(A \cap B)}{P(A)}$, Provided $P(A) > 0$.

Example :

suppose that $P(A) = 0.4$, $P(B) = 0.3$

$P(A|B) = 0.03$ $P(B|A)$ $A \cap B$

Find : $\boxed{1}$ $P(A \cap \bar{B})$. $P(\bar{A}|B) = 1 - P(A|B)$

$\boxed{2}$ $P(B|A)$.

$\boxed{3}$ $P(A \cup \bar{B})$.

Solu :

	A	$\bar{A}$	Total
B	$P(A \cap B) = 0.009$	$P(\bar{A} \cap B) = 0.291$	$P(B) = 0.3$
$\bar{B}$	$P(A \cap \bar{B}) = 0.391$	$P(\bar{A} \cap \bar{B}) = .309$	0.7
Total	0.4	0.6	1

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

$= 0.3 * 0.03 = 0.009$

Ans : $\boxed{1}$ 0.391

$\boxed{2}$ $\dfrac{P(B \cap A)}{P(A)} = \dfrac{0.009}{0.4}$.

$\boxed{3}$ $P(A \cup \bar{B}) = P(A) + P(\bar{B}) - P(A \cap \bar{B})$

$0.4 + 0.7 - 0.391 = \boxed{}$

Mutually Exclusive Events ; Independent Events :

المستقلة

[1] A and B are mutually exclusive.

if $A \cap B = \phi$

(i.e. $P(A \cap B) = 0$, i.e. $P(A \cup B) = P(A) + P(B)$.

[2] A and B are Independent if $P(A \cap B) = P(A)\,P(B)$

$\Rightarrow$ when Probability $(P(A) > 0)$, A and B are indep. if $P(B|A) = P(B)$

$\Rightarrow$ " " $P(B) > 0$, " " " " $P(A|B) = P(A)$.

Fact :

If A, B are independent then each of the following are also independent :-

[1] A and $\bar{B}$

[2] $\bar{A}$ and B

[3] $\bar{A}$ and $\bar{B}$

Example :

Suppose that A and B are independent and $P(A) = 0.4$, $P(B) = 0.3$

Find $P(\bar{A} \cup \bar{B})$

Sol : A, B are independent so $\bar{A}, \bar{B}$ are also indep.

$\therefore P(\bar{A} \cup \bar{B}) = P(\bar{A}) + P(\bar{B}) - P(\bar{A})\,P(\bar{B})$

$= 0.6 + 0.7 - 0.6\,(0.7)$

$1.3 - 0.42 = 0.88$

⇒ Law of Total Probability : Bayes' Rule :.

Let $A, B_1, B_2, \dots, B_n$ be events such that

① For each $i \neq j$, $B_i \cap B_j = \varnothing$

② $B_1 \cup B_2 \cup \dots \cup B_n = S$

Then $P(A) = P(A \cap B_1) + (P(A \cap B_2) + \dots + P(A \cap B_n)$

$= P(A|B_1)\, P(B_1) + P(A|B_2)\, P(B_2) + \dots + P(A|B_n)\, P(B_n)$

This law is called the law of total Probability.

Baye's Rule :

$P(B|A) = \dfrac{P(A|B)\; P(B)}{P(A)}$

Examples:-

One box is randomly selected

then One ball is randomly selected, Then one ball is randomly

Find ① P(The ball is black). 2

② P(box2 | black).

if the Probability of selecting box1 is 0.4, box2 0.3, box3 = 0.3

Solu:- ① $P(\text{black}) = P(\text{black} \cap \text{box1}) + P(\text{black} \cap \text{box2}) + P(\text{black} \cap \text{box3})$

$= P(\text{black}|\text{box1})\, P(\text{box1}) + P(\text{black}|\text{box2})\, P(\text{box2}) + P(\text{black}|\text{box3})\, P(\text{box3})$

$= \frac{2}{5} \times \frac{4}{10} + \frac{1}{6} \times \frac{3}{10} + \frac{3}{6} \times \frac{3}{10} = 0.36$

② $P(\text{box2}|\text{black}) = \dfrac{P(\text{black}|\text{box2})\, P(\text{box2})}{P(\text{black})} = \dfrac{\frac{1}{6} \times \frac{3}{10}}{0.36} = 0.138$

$= \dfrac{P(\text{bla}|\text{box2})\, P(\text{box2})}{P(\text{black})}$

Lock 80

Subject Excs- 8, 9 page 98 * Date 25-10-2010 No.

Ex 8:- $P(male) = 0.75$, $P(Female) = 0.25$

$P(accounting \mid male) = 0.12$

$P(accounting \mid female) = 0.20 \Rightarrow (\overline{acc} \mid female) = 0.8$

Find:- [1] $P(accounting)$ [2] $P(male \mid accounting)$

Solu:-

[1] $P(accounting) = P(accounting \cap male) + P(accounting \cap female)$

[2] $P(male \mid accounting) = \dfrac{\overset{0.12}{P(acc \mid male)}\ \overset{0.75}{P(male)}}{P(accounting)}$ from part [1]

Univariate Random Variables:-

The distribution of a random variable is a table or a graph or a formula that gives all possible values of the random variable and the Probability of each of these values.

A formula that is used to find such Probability is called the Probability density function (P.d.F) of the distribution

Example:- The following is a distrib

x	f(x)
0	0.1
1	0.3
2	0.1
3	0.5
	1

[1] what are the possible values of x?

Ans. $\{0, 1, 2, 3\}$

[2] Find $P(X \text{ is odd})$? $= P(X=1) + P(X=3)$

$= 0.3 + 0.5 = 0.8$

[3] Find $P(X \text{ is odd} \mid X>1)$?

Ans:- $P(X \text{ is odd} \mid X>1) = \dfrac{P(X \text{ is odd} \cap X>1)}{P(X>1)} = \dfrac{0.5}{\underset{x=2\ +\ x=3}{0.1 + 0.5}} = \dfrac{0.5}{0.6} = \dfrac{5}{6}$

N O T E B O O K

Examples:-

A box contain 3 white and 2 black balls, 3 balls are drawn from this box. let X be the number of black ball within the drawn balls.

[1] construct the dist of X.

[2] what is the P. d. F of X.

Solu..

[1]

X	$P(X)$
0	$P(X=0) = \frac{C_0^2 C_3^3}{C_3^5} = \frac{1}{10}$
1	$P(X=1)$ $\frac{C_1^2 C_2^3}{C_3^5} = \frac{2\times 3}{10} = 0.6$
2	0.3

[2] $P(X=k) = \frac{C_k^2 \, C_{3-k}^3}{C_3^5}$, $k = 0, 1, 2$.

Ex:- X is a variable r.v taking the values $-1, 1, 2, 3$ where

$P(-1) = P(2)$, $P(1) = 0.1$, $E[X] = 1.8$

Find : [1] The P. d. F of x.? .....

$(-1)(P(-1)) + P(1) + 2P(2) + 3P(3) = 1.8$

$P(-1) + P(1) + P(2) + P(3) = 1$

حذف وتعويض

$-P(-1) + 0.1 + 2P(-1) + 3P(3) = 1.8$

$P(-1) + 0.1 + P(-1) + P(3) = 1$ ✱

Variance:-

For a random variable X, $Var(X) = E(X - M_X)^2$

$= E(X^2) - (E(X))^2$ ✱

$\Rightarrow E(X^2) = \sum_{\text{over all } x} x^2 \, P(x)$ if X is discrete.

Examples:-

Suppose that $var(X) = 10$, $E(X) = 20$ Find $E(X^2)$.

Solu:- $Var(X) = E(X^2) - (E(X))^2$

$E(X^2) = 10 + 20^2 = 410.$

Example:-

Consider the following distribution:-

relative Pdf

X	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
-1	0.1	-0.1	1	0.1
1	0.3	0.3	1	0.3
2	0.2	0.4	4	0.8
3	0.4	1.2	9	3.6
Total	1	1.8		4.8

Find [1] $E(X)$ [2] $E(X^2)$ [3] $Var(X)$

[1] $E(X) = 1.8$

[2] $E(X^2) = 4.8$

[3] $Var(X) = 4.8 - (1.8)^2 = 1.56$

Facts:-

[1] $E(aX+b) = aE(X) + b$

[2] $Var(aX+b) = a^2 Var(X)$

[3] $\sigma_{aX+b} = \sqrt{Var(aX+b)} = |a| \sigma_X = |a| \sqrt{Var(X)}$.

Bivariate Distribution:-

A discrete bivarite dis. looks like the following:-

$X \backslash y$	y_1	y_2	y_3	----	Total
X_1	$P(X=x_1 \text{ and } y=y_1)$	$P(X=x_1 \text{ and } y=y_2)$	$P(X=x_1 \text{ and } y=y_3)$		$P(X=x_1)$
X_2	$P(X=x_2 \text{ and } y=y_1)$	--	--		$P(X=x_2)$
X_3			$P(X=x_3 \text{ and } y=y_3)$		
⋮					
Total	$P(y=y_1)$	$P(y=y_2)$			1

Example:-

Consider the following bivariate distn.

$x \backslash y$	-1	1	2	Total
-1	0.1	0.2	0.1	0.4
1	0.3	0.1	0.2	0.6
Total	0.4	0.3	0.3	1

Find :-
[1] $E(x)$
[2] $E(y)$
[3] $P(y \geq 1 \mid x = -1)$
[4] $P(x + y > 0)$
[5] $P(x^2 + y^2 = 2)$

Sol :-

[1] $E(x) = (-1)(0.4) + (1)(0.6) = 0.2$

[2] $E(y) = (-1)(0.4) + (1)(0.3) + (2)(0.3) = 0.5$

[3] $P(y \geq 1 \mid x = -1) = \dfrac{P(y \geq 1 \cap x = -1)}{P(x = -1)} = \dfrac{0.2 + 0.1}{0.4} = 0.75$

[4] $P(x + y > 0) = 0.1 + 0.1 + 0.2 = 0.4$

[5] $P(x^2 + y^2 = 2) = 0.1 + 0.2 + 0.3 + 0.1 = 0.7$

Independance :-

The two random variables X and Y are independant if for all i and j

$$P(X = x_i \text{ and } Y = y_j) = P(X = x_i)\,P(Y = y_j)$$

Example:- The variables in the following distribution are Indepen.

$x \backslash y$	-1	1	2	Total.
1	0.08	0.12	0.2	0.4
2	0.12	0.18	0.3	0.6
Total	0.2	0.3	0.5	1

الليل جسر نحو الصباح
والشمس دارت واستدارت
للشروق ...

Example:-

The variables in the following dis. are not indep.

$X \backslash y$	a	b	c	Total
1	0.1	0.1	0.2	0.4
2	0.3	0.1	0.2	0.6
Total	0.4	0.2	0.4	1

$\Rightarrow$ because $P(X=1 \text{ and } y=a) \neq P(X=1)\,P(y=a)$.

Covariance :-

The covariance of X and y is given by $cov(X,y)$ $\sigma_{X,y}$

$$cov(X,y) = E((X-M_X)(y-M_y))$$

$$= E(Xy) - E(X)E(y)$$

$$E(Xy) = \sum_{\substack{\text{over all}\\(i,j)}} x_i y_j \, P(X=x_i \text{ and } y=y_j)$$

Example:-

Consider the following distribution

Find:- [1] $E(X)$ [2] $E(y)$ [3] $E(Xy)$

[4] $Cov(X,y)$

$X \backslash y$	1	2	Total
-1	0.1	0.3	0.4
0	0.2	0.1	0.3
1	0.1	0.2	0.3
Total	0.4	0.6	1

Solu:-

[1] $E(X) = -1(0.4) + 0 + (1)(0.3) = -0.1$

[2] $E(y) = 1(0.4) + 2(0.6) = 1.6$

[3] $E(Xy) = -1(1)(0.1) + -1(2)(0.3) + 0 + (1)(1)(0.1) + (1)(2)(0.2) = -0.2$

[4] $cov(X,y) = E(Xy) - E(X)E(y)$

$$= -0.2 - (-0.1)(1.6) = -0.04$$

$\Rightarrow$ covariance can have negative values.

variance if cant [illegible]

Properties of cov :-

1 If x and y are Independent then $Cov(x,y) = 0$

If $cov(x,y) = 0$ then we don't conclude that x and y are independent

2 $Cov(ax+b, cy+d) = ac\ Cov(x,y)$

3 $Var(ax+by+c) = a^2 Var(x) + b^2 var(y) + 2ab\ Cov(x,y)$

Correlation Coefficient :-

The correlation coefficient of x and y is denoted by :-

$Corr(x,y)$ (or ρ) ρ row

It is given by

$$Corr(x,y) = \frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}} = \frac{E(xy) - E(x)E(y)}{\sqrt{(E(x^2) - (E(x))^2)(E(y^2) - (E(y))^2)}}$$

Example :-

Compute $Corr(x,y)$ for the variables given in the previous ex.

Solu:- Remember that $E(x) = -0.1$, $E(y) = 1.6$, $Cov(x,y) = -0.04$

we need $E(x^2)$, $E(y^2)$

$E(x^2) = (-1)^2(0.4) + 0 + (1)^2(0.3) = 0.7$

$E(y^2) = 1^2(0.4) + 2^2(0.6) = 2.8$

$\therefore\ var(x) = 0.7 - (-0.1)^2 = 0.69$

$Var(y) = 2.8 - (1.6)^2 = 0.24$

$$\therefore\ Corr(x,y) = \frac{-0.04}{\sqrt{(0.69)(0.24)}} = -0.098 \approx -0.1$$

Subject Date 1/11/2010 No.

Properties:-

[1] $-1 \leq corr(X,Y) \leq 1$.

* [2] If $corr(X,Y) = 0$ then there is no linear relation between X and Y (there could be possibly be another kind of relation)

[3] If X and Y are Independent, then $corr(X,Y) = 0$

[4] $corr(aX+b, cY+d) = \begin{cases} corr(X,Y) & \text{if } ac>0 \\ -corr(X,Y) & \text{if } ac<0 \end{cases}$

Example:

Suppose that $cov(X,Y) = 3$; $var(X) = 16$, $var(Y) = 25$

$E(X) = 5$, $E(Y) = 8$

[1] Find $E(XY)$ [2] $corr(1-3X, 4Y+1)$? [3] $cov(2X+5, 1-7Y)$.

Solu:

[1] $E(XY) = cov(X,Y) + E(X)E(Y) = 3 + 5 \times 8 = 43$

[2] $corr(1-3X, 4Y+1) = -corr(X,Y) = -\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}}$

$= \frac{-3}{4 \times 5} = -0.15$

[3] $cov(2X+5, 1-7Y) = 2(-7)\,cov(X,Y) = -14 \times 3 = -42$.

[4] Find $var(X-Y)$

$var(X-Y) = var(X) + var(Y) - 2cov(X,Y)$

$= 25 + 16 - 2 \times 3 = 35$

N O T E B O O K

Some Discrete Distribution:

⇒ The Binomial Distribution:

⇒ Bernoulli Experiment:

A random experiment is called a Bernoulli (or binomial) experiment. if the following conditions are satisfied:

1. the experiment can be repeated as many times as we wish [illegible] then no. of success.
2. each trial of the experiment has only two Possible values, one is called "success" and the other is called "failure"
3. The Probability of success in each trial is constant: i.e. all trials are independent.

⇒ Binomial Random variables:

Repeat a Bernoulli experiment n times and let X be the number of success within these n trials. Then X is called a Binomial Random variable.

The distribution of X is called a binomial dist. with parameters n and p, where p is the probability of success in each trial.

This distribution is denoted by $B(n,p)$ (or $X \overset{has}{\sim} B(n,p)$).

Fact: let $X \sim B(n,p)$. Then

1. the possible values of X are $0, 1, 2, \dots, n$.
2. the P.d.f of X is given by

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \Rightarrow k = 0, 1, \dots, n.$$

3. $E(X) = np$ Just 4 Binomial =7
4. $Var(X) = np(1-p)$

Example:

let $X \sim B(\overset{n}{10}, \overset{p}{0.3})$. Find,

1. $P(X=3)$ 2. $E(X^2)$

Solu:- 1. $P(X=3) = \binom{10}{3}(0.3)^3(0.7)^7$

2. $E(X^2) = Var(X) + (E(X))^2$

$$= 10 * 0.3 * 0.7 + (10 * 0.3)^2 = 11.1$$

Example:-

let $X \sim B(n,P)$, suppose that $E(X) = 20$, $Var(X) = 10$

Find n and p!?

Solu:-

$E(X) = np = 20$

$Var(X) = np(1-P) = 10$

$\therefore 1 - P = \frac{10}{20} = 0.5 \quad \Rightarrow \therefore p = 0.5$

$E(X) = np = 20 \quad \Rightarrow \quad n = 40.$

Example:-

Suppose that 20% of all Jordanian are smokers.
In a random sample of size 10, selected from all Jordanian, find:

[1] expected number of smokers

[2] Probability of having at least one smoker

[3] Probability of having at ~~least~~ most one smoker

Sol.:-

Let X be the no. of smokers in our sample. Then

$X \sim B(10, 0.20)$

[1] $E(x) = 10 * 0.20 = 2$

[2] $P(X \geq 1) = 1 - P(X=0) = 1 - \binom{10}{0}(0.2)^0(0.8)^{10} = 1 - (0.8)^{10}$

[3] $P(X \leq 1) = P(X=0) + P(X=1) = \binom{10}{0}(0.2)^0(0.8)^{10} + \binom{10}{1}(0.2)^1(0.8)^{10}$

$= (0.8)^{10} + 2 * (0.8)^9 =$

(2) at least 2

$\therefore 1 - (P(X=0) + P(X=1))$

never!

Binomial Tables:- الجداول الثنائية

Binomail tables gives $P(X \le k)$ when $X \sim B(n,p)$, $k = 0, 1, 2, \ldots, n$. For several values of n and p.

An example of abinomail table is the following:

$n = 10$

k	0.01	0.05	0.10	0.20	0.30	...	0.99
0							
1							
2			$P(X \le 2)$;				
3			$X \sim B(10, 0.10)$				
4							
5					0.953		
6					$P(X \le 5)$; $X \sim B(10, 0.3)$		
⋮							
10							

Q:- How can we use abinomial table for finding each of the following?

1) $P(X = k)$?

Ans:- $P(X = k) = P(X \le k) - P(X \le k-1)$

2) $P(X < k)$? $\quad k-3 \quad k-2 \quad k-1$

Ans:- $P(X < k) = P(X \le k-1)$

3) $P(X \ge k)$? $\quad k \ldots n$

Ans:- $P(X \ge k) = 1 - P(X \le k-1)$

4) $P(X > k)$? $\quad k+1 \quad k+2 \ldots$

Ans: $P(X > k) = 1 - P(X \le k)$ $\quad 1 - P(X \le k+1$

5) $P(k \le X \le L)$? $\quad k \ldots L$

Ans:- $P(k \le X \le L) = P(X \le L) - P(X \le k-1)$

6) $P(k < X \le L)$? $\quad k+1 \quad L$

$P(k < X \le L) = P(X \le L) - P(X \le k)$

Subject Date 8-11-2010 No.

[7] $P(k \le X < L)$? k L-1

$P(k \le X < L) = P(X \le L-1) - P(X \le k-1)$

[8] $P(k < X < L)$? k+1 L-1

$= P = (X \le L-1) - P(X \le k)$

Examples.

A box contains two white and 8 black balls.
One ball is drawn from the box $\frac{20}{n}$ times, with replacement.

[1] Find the probability that a white ball will show up at least 10 times!

[2] the expected number of times a black ball shows up?

[1] let X be the no. of times a white ball shows up
Then $X \sim B(20, P)$, P = Propability that a white ball shows up each time.

$= \frac{2}{10} = 0.2$

$\therefore X \sim B(10, 0.20)$. To find $P(X \geq 10)$

$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.997 = 0.003$.

[2] let y be no. of times a black ball shows up. Then
$y \sim B(20, 0.8)$

$E(y) = np = 20 \times 0.8 = 16$.

Subject Date 8-11-2010 No.

Poisson Distribution :- حدث نادر

Let X be the no. of occurances of an event in time or space unit. such that:-

[1] The Probability that X=1 is directly Proportional with time or space unit.

(2) $P(X>1)$ in "a small" or "short" time is or space unit is "almost" Zero.

[3] occurences are independent then X is said to be a Poisson random variable.

It's distribution is called a Poisson dist.

Notation :-

Let X be a Poisson random variables, when we write $X \sim$ Poisson (λ) .... we understand that X has a Poisson distribution with mean λ occurences per times of space unit

Fact :-

Let $X \sim Poisson(\lambda)$ Then:

[1] The Possible values of X are $0, 1, 2, 3, \ldots$

[2] The P.d.f of X is given by $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, 2 \ldots$

[3] $E(X) = Var(X) = \lambda$ $e \approx 2.71$

Subject Date 10/11/2010 No.

Example:-

let $X \sim$ Poisson (3). Find:-

[1] $P(X \geq 1)$.

[2] $E(x^2)$.

Solu:- [1] $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$

$= 1 - \frac{e^{-3} \, 3^0}{0!} = 1 - \frac{1}{e^3} = \approx 0.96$

$0! = (1)$

[2] $E(x^2) = Var(x) + (E(x))^2 = 3 + 3^2 = 12$.

Poisson tables:-

Poisson tables give $P(X \leq k)$

when $X \sim Poisson(M)$ for several values of M

k	.1	.2	.3	.4
1				
2				
3				
4				

$P(X \leq 3)$; $X \sim Possion(0.3)$.

Examples:

The number of People arriving at acertain store is distributed according to a Possion dist. with mean 3 per hour. Find the Probability that at most 16 people arrive at this store in aperiod of 4 hour.?!

Solu:- Let X be the number of Peoples arriving in this store per 4 hours. Then $X \sim Poisson(4 \times 3) = Poisson(12)$

To find $P(X \leq 16) = 0.899$

⟹ Geometric Distribution:

Suppose that we have a Bernoulli experiment with Probability of success P.
Let X be the number of trials of this experiment Until the First success Then X is a random variables called a Geometric Random variable with parameter P and written as $X \sim (Geometric(P))$.

Facts:-

let $X \sim Geometric\ (P)$ Then:-

1- the Possible values of X are 1, 2, 3, ...

2- the P.d.f of X is

$$P(X=k) = (1-P)^{k-1} P$$

3- $E(x) = \frac{1}{P}$, $Var(x) = \frac{1-P}{P^2}$

Examples:

⟹ A box contain 2 white and 3 black balls, One ball is drawn from the box with Replacement, till a white ball shows up

Let X be the number of trials at this experiment.

Find [1] $P(X=4)$ [2] $E(X^2)$.

$X \sim Geometric\left(\frac{2}{5}\right) = Geometric\ (0.4)$

[1] $P(X=4) = (0.6)^3 (0.4) = 0.0864$.

[2] $E(X^2) = Var(x) + (E(X))^2$

$$= \frac{1-0.4}{(0.4)^2} + \left(\frac{1}{0.4}\right)^2 = 3.75 + 6.25 = 10$$

Hypergeometric Distribution:-

type 1 items: $\underset{\text{items}}{M}$ $\longrightarrow$ n items are selected without replacement (or together).

type 2 items: $\underset{\text{items}}{N-M}$

N items!

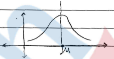
let x be the no. of type 1 items in our selected sample. Then X is a random variable it's distribution is called a hyper geometric Distribution.

The P.d.f of x is given by:-

$$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

$P(X=k) =$

Normal Probability Distribution:- (Bell-shaped).
The random variables that have the normal distribution are continous.
A typical normal disn. is



The P.d.f of the normal disn with mean μ and variance σ^2 - is $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$

Probability of a continuous random variable:-
Let $f(x)$ be the P.d.f of a continuous random variable.
Let a, b be numbers selected from the Possible values of X such that $a<b$.
Then:
[1] $P(X=a) \overset{\text{by def.}}{=\!=} 0$
[2] $P(a<X<b) \overset{\text{by def}}{=\!=} \int_a^b f(x)\,d(x)$

$\Rightarrow$ Notation:- When we write $X \sim N(\mu, \sigma^2)$ we understand that X has the normal disn. with Mean μ and variance σ^2.

Standard Normal dis. and tables :-

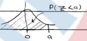
The standard Normal disn is denoted by Z

It's nothing but $N(0,1)$ dis.

Fact:-

If $X \sim N(\mu, \sigma^2)$

then $\frac{X-\mu}{\sigma} \sim N(0,1)$



a	.01	.02	.03	.04	.08 - -	.09	a
0.0							
0.1							
0.2							
0.3							
0.4							
1.2					0.8997		
1.3							
03.4							

$P(Z<1.32) = 0.9066$ *

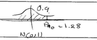
Example:- Find $P(Z<1.32121)$

Solu:-

$P(Z<1.32121) \simeq P(Z<1.32)$

Example :-

Find P_{90} at $N(0,1)$



Solu:-

So, P_{90} at $N(0,1) = 1.28$

$(Z<0.9)$

Examples:

Suppose that the grades of a statistics couse have $N(60, 15^2)$

1) If a student is randomly selected, find the probability that her grade is greater than 90

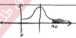
2) Find the proportion of student who got grades between 50 and 80.

3) If 30% of the student got grades between 50 and a, find a.

4) If "A" will be given to ten percent of the student, find the least grade that will correspond to an "A"

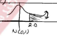
⇒ Solution

1) to find $P(\text{Grade} > 90)$



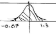
$P(\text{Grade} > 90) = P(Z > \frac{90-60}{15}) = P(Z > 2.00)$

$= P(Z < -2.00) = 0.0228.$



2) To find $P(50 < \text{Grade} < 80)$

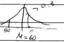
$P(50 < \text{grade} < 80) = P(\frac{50-60}{15} < Z < \frac{80-60}{15})$



$= P(-0.67 < Z < 1.3)$

$= P(Z < 1.3) - P(Z < -0.67) = 0.9032 - 0.2514 = 0.6518$

3) $P(50 < \text{Grade} < a) = 0.30$



Notice that $P(\text{Grade} < a) = 0.30 + P(\text{Grade} < 50)$

So: $P(Z < \frac{a-60}{15}) = 0.30 + P(Z < \frac{50-60}{15})$. Thus,

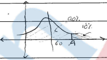
$P(Z < \frac{a-60}{15}) = 0.3 + P(Z < -0.67)$

$0.3 + 0.2514 = 0.5514$

$\Rightarrow$ $\therefore P\left(Z < \frac{a-60}{15}\right) = 0.5514$

Using tables: $\frac{a-60}{15} = 0.13$

$a = 0.13 * 15 + 60 = 61.95$



$\boxed{4}$ A is nothing but P_{90} of the grades

P_{90} of the grades = ??

$P(\text{Grade} < P_{90}) = 0.9$

So $P\left(Z < \frac{P_{90}-60}{15}\right) = 0.9$

$\therefore \frac{P_{90} - 60}{15} = 1.28$

$P_{90} = 1.28 * 15 + 60 = 79.2 \approx 79$

$\Rightarrow$ In general P_{100p} of $N(\mu, \sigma^2)$ is $(P_{100p}$ of $N(0,1))\sigma + \mu$.

Example 80

Let $X \sim N(10, 25)$, Find P_{95} of X.

Sol: P_{95} of X is P_{95} of $N(10, 25)$

$= P_{95}$ of $N(0,1) * \sigma + \mu = 1.65 * 5 + 10 = 18.25$

Example: Suppose that the grades have normal $N(\mu, 15^2)$ and that 10% of the students got grades greater than 80. Find P_{75} of the grades.

Solu:

P_{75} at $N(\mu, 15^2) = (P_{75}$ at $N(0,1)) * 15 + \mu$

$= 0.66 * 15 + \mu.$

To Find μ.

Now, P_{90} at $N(\mu, 15^2) = 80 \Rightarrow 15 * 1.28 + \mu = 80$

$\therefore \mu = -19.2 + 80$

$\mu = 60.8$

$\therefore P_{75}$ at Grades $= 0.60 * 15 + 60.8 = 70.7$.

$\Longrightarrow$ The Normal Approximation to the Binomial Distribution.

Let $X \sim B(n,p)$. Suppose that $np > 5$ and $n(1-p) > 5$ then, the dist. of X, namely the binomial dist. can be approximated by $N(np, np(1-p))$.

$\Rightarrow$ To reduce the errors that comes out though Approximating a discrete dist. by a continuous dist. we use the so called "continuity correction"

التحليل

$P(V) = 0.53$

$P(A) = 0.05$

$\Rightarrow P(A \cap V) = P(V|A)\ P(A)$

$= 0.73 * 0.05$

	A	$\bar{A}$	
V	*		0.53
$\bar{V}$	*	$\square$	$\square$
	0.05	0.95	1

(2) $P(A|\bar{V}) = \dfrac{P(A \cap \bar{V})}{P(\bar{V})}$ (1) $P(V|\bar{A}) = \dfrac{P(V \cap \bar{A})}{P(\bar{A})}$

Continuity corrections:-

Before we approximate $B(n,p)$ by $N(np, np(1-p))$ we must perform continuity correction

$P(X=a)$ in the binomial is corrected to $P(a-\frac{1}{2} < X < a+\frac{1}{2})$ in the normal.

$P(a \leq X \leq b) \Rightarrow$ in the binomial is corrected to $P(a-\frac{1}{2} < X < b+\frac{1}{2})$.

Example:- Let $X \sim B(100, 0.5)$, Use a normal dist to approximate each of the following:-

[1] $P(X=60)$ [6] $P(40 \leq X \leq 60)$

[2] $P(X<60)$ [7] $P(40 < X \leq 60)$

[3] $P(X \leq 60)$ [8] $P(40 \leq X < 60)$

[4] $P(X>60)$ [9] $P(40 < X < 60)$

[5] $P(X \geq 60)$

Solu:- Binomial $B(100, 0.5)$ is approximated by $N(100 * 0.5,\ 100 * 0.5(1-0.5))$

(1) $P(X=60) \approx P(59.5 < X < 60.5)$ $= N(50, 5^2)$

$= P(\frac{59.5-50}{5} < Z < \frac{60.5-50}{5}) = P(1.9 < Z < 2.1)$

$= P(Z<2.1) - P(Z<1.9) = 0.9821 - 0.9713 = 0.0108$

(2) $P(X<60) = P(X \leq 59) \approx P(X < 59.5) = P(Z < \frac{59.5-50}{5})$

$P(Z<1.9) = 0.9713$

[3] $P(X \leq 60) \approx P(X < 60.5) = P(Z < \frac{60.5-50}{5})$

$= P(Z<2.1) = 0.9821$

[4] $P(X>60) = P(X \geq 61) \approx P(X > 60.5) = 1 - 0.9821 = 0.0179$

$(P\ X > \frac{60.5-50}{5})$

$= \binom{100}{x}(0.5)^{x}(0.5)^{100-x}$

5) $P(X \geq 60) \approx P(X > 59.5) = 1 - 0.9713 = 0.0287$

6) $P(40 \leq X \leq 60) \approx P(39.5 < X < 60.5) = P(\frac{39.5-50}{5} < Z < \frac{60.5-50}{5}) = ex$

7) $P(40 < X \leq 60) = P(41 \leq X \leq 60) \approx P(40.5 < X < 60.5) \ldots$

8) $P(40 \leq X < 60) = P(40 \leq X \leq 59) \approx P(39.5 < X < 59.5) \ldots$

9) $P(40 < X < 60) = P(41 \leq X \leq 59) \approx P(40.5 < X < 59.5) \ldots$

→ The Central Limit Theorem

let $\bar{X}$ be the mean of a random sample with size n selected (with replacement) from a population that has mean μ_x and variance σ_x^2. Then $\bar{X}$ is a random variable.

→ If n is "large" ($n \geq 30$).

Then $\bar{X}$ almost has $N(\mu_x, \frac{\sigma_x^2}{n})$.

Example: Suppose that the ages of JU students have mean $\mu = 21$ and variance $\sigma^2 = 36$ $N(21, 6)$

variance σ^2

A random sample of size $n = 49$ is selected from JU students. let the mean age of this sample.

1) what is the distn of $\bar{X}$. ?

2) Find the $P(\bar{X} > 22)$. ?

3) Find the Percentile P_{95} of $\bar{X}$. ?

Solution: 1) Since $n = 49 > 3$. $\bar{X}$ (almost) has $N(21, \frac{36}{49})$

$= N(21, (\frac{6}{7})^2)$.

Normal 2) $P(\bar{X} > 22) = P(Z > \frac{22-21}{\frac{6}{7}}) = P(Z > \frac{7}{6}) = P(Z > 1.16)$

$= P(Z < -1.16) = \cancel{0.8770}$ 0.123

3) P_{95} of $\bar{X} = P_{95}$ of $N(21, (\frac{6}{7})^2) = P_{95}$ of $N(0,1)) * \frac{6}{7} + 21$

$= 1.65 * \frac{6}{7} + 21 = 22.4$

Fact :-

The condition that $n \geq 30$ is not necessary if X itself is normally distributed.

In fact, if $X \sim N(\mu, \sigma^2)$, then $\bar{X} \overset{\text{exactly normal}}{\sim} N(\mu, \frac{\sigma^2}{n})$

Example :-

Suppose that the weights of Jordians have $N(70, (20)^2)$ distn.

A random sample of six Jordanians will use a lift that has a threshold 400 kg. Find the probability that their total weight exceeds the threshold of the lift.

Solu :- To find $P(\sum_{i=1}^{6} X_i > 400)$, where $\sum_{i=1}^{6} X_i$ is the total weigh of the sample.

$P(\sum_{i=1}^{6} X_i > 400) = P(\frac{\sum_{i=1}^{6} X_i}{6} > \frac{400}{6}) = P(\bar{X} > 66.67)$

Since $X \sim N(70, (20)^2)$, $\bar{X} \sim N(70, (\frac{20}{\sqrt{6}})^2) = N(70, (8.16)^2)$

$P(\bar{X} > 66.67) = P(Z > \frac{66.67 - 70}{8.16}) = P(Z > -0.4)$

$= P(Z < 0.4) = 0.6554.$

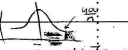
⇒ Example :-

In Previous Example, How many people should use the lift So that the Probability of exceeding the threshold is 0.05

Solu :- To find n such that $P(\sum_{i=1}^{n} X_i > 400) = 0.05$

$P(\sum_{i=1}^{n} X_i > 400) = 0.05 \Rightarrow P(\frac{\sum_{i=1}^{n} X}{n} > \frac{400}{n}) = 0.05$

$\Rightarrow (\bar{X} > \frac{400}{n}) = 0.05$



$\therefore \frac{400}{n} = P_{95}$ of $\bar{X} = P_{95}$ of $N(0,1) \frac{20}{\sqrt{n}} + 70$ $\qquad \bar{X} \sim N(70, (\frac{20}{\sqrt{n}})^2)$

$\frac{400}{n} = (1.65)\frac{20}{\sqrt{n}} + 70$ $\qquad \frac{\frac{400}{n} - 70}{\frac{20}{\sqrt{n}}} = -2.52$

$400 = (1.65)\ 33\sqrt{n} + 70$, Put $\sqrt{n} = x$ $\qquad \frac{400}{n}$

$70x^2 + 33x - 400 = 0$

$x = \frac{-33 \mp \sqrt{133^2 + 4(70)(400)}}{2 \times 70} = 2.16$

So $n = x^2 = (2.16)^2 = 4.665 \approx 4.0$ peoples

→ The t-distribution

If $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Thus $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Qu: If we use S instead of σ, will we have $N(0,1)$?

i.e.: Does $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ have $N(0,1)$?

Ans: No it does not.

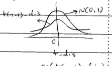
what is the distribution of $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$?

The answer was given by Gosset.

$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ has the called t-distribution with $n-1$ degrees of freedom

→ It is denoted by $t(n-1)$-dis

A typical t-dis looks like the following.



Facts:

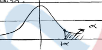
As n gets larger, the t-dis gets close to $N(0,1)$

For $n \geq 30$, we will consider t-dis $= N(0,1)$

Notation:-

Percentiles of t-disn. are denoted by the following

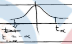
Notice that $t_{\alpha} = P_{(1-\alpha)}$ of the T-disn.



Remarks:

Since t-disn is symmetrical about the line $X=0$.

$t_{1-\alpha} = -t_{\alpha}$



degrees of freedom

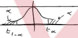
t- tables.

d.F	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$
1			
2			
3			
4			
5		$t_{0.05}(4)$	
6			
7			
⋮			

Recall:- ① the random variable $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ has $t(n-1)$-disn

Provided $X \sim N(\mu, \sigma^2)$ ⇐ توزيع $\bar{X}$ Normal عليه T.

② $P_{(1-\alpha)100}$ of $t(n-1)$-disn is denoted by $t_{\alpha}(n-1)$

③ 

} $t_{1-\alpha} = -t_{\alpha}$

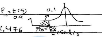
⇒ [1] Find the Percentile P_{10} of $t(5)$-disn. t_{n-1}

[2] Find P_{90} of $\frac{\bar{X}-\mu}{\frac{S}{\text{root-}n}}$, where $n=6$ $t^{6}_{0.1}$ $t^{5}_{0.1}$

[3] Find P_{70} of $\bar{X}$, where $\bar{X}$ is the mean of a sample with size 6

selected from $N(60, \sigma^2)$ and has variance $S^2 = 144$

Solu:



[1] To find $t_{0.9}^{(5)}$ which equels $-t_{0.1}(5) = 1.476$

[2] P_{90} for $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ is $t_{0.1}(n-1) \Rightarrow P_{90}$ for $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{6}}}$ is $t_{0.1}^{(5)} = 1.476$

[3] P_{90} of $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ is $t_{0.1}$, thus P_{90} of $\bar{X}$ is $t_{0.1} * \frac{S}{\sqrt{n}} + \mu$

$= 1.476 * \frac{12}{\sqrt{6}} + 60 = 67.2$.

Distribution of $\hat{P}$: "Sample Proportion":

Let $X \sim B(n, P)$. Let $\hat{P} = \frac{X}{n}$

Then $E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} * np = P$

$Var(\hat{P}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} * nP(1-P) = \frac{P(1-P)}{n}$.

Fact: If $np \geq 5$ and $n(1-P) \geq 5$,

then $\hat{P}$ (almost) has $N\left(P, \frac{P(1-P)}{n}\right)$.

Example:

⇒ Suppose that 30% of all Jordanian are smokers.

let $\hat{P}$ be the Proportion of smokers in a random sample of 100 Jordanians.

[1] Find the disn of $\hat{P}$ [2] $P(\hat{P} > 0.35)$

[3] P_{95} of $\hat{P}$.

Solu: [1] $\hat{P}$ (almost) $\sim N\left(0.30, \frac{(0.30)(1-0.30)}{100}\right) = N\left(0.30, \frac{0.21}{100}\right)$

$= N(0.30, (0.046)^2)$

[2] $P(\hat{P} > 0.35) = P(Z > \frac{0.35 - 0.30}{0.046}) = P(Z > 1.09)$

$= P(Z < -1.09) = 0.1379$

[3] P_{95} of $\hat{P}$ is P_{95} of $N(0.30, (0.046)^2)$

P_{95} of $(N\ 0.30, (0.046)^2) = (P_{95}$ of $N(0,1))\,\sigma + \mu$

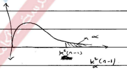
$= 1.65 * 0.046 + 0.30 = 0.375$

Distribution of S^2 so

$\Rightarrow$ Let S^2 be the variance of a random sample with size(n) selected from $N(\mu, \sigma^2)$. Then $\frac{(n-1)S^2}{\sigma^2}$ is a random variable. Its distribution is called a Chi-square distn with $(n-1)$ degrees of freedom. It's denoted by $\chi^2(n-1)$ - distn.

A typical χ^2 - distn looks like the following :-

$P(1-\alpha)_{100}$ of $\chi^2(n-1)$ is denoted by $\chi^2_\alpha(n-1)$



$\chi^2_\alpha(n-1)$

Example:-

let S^2 be the variance of a random sample with size 6 selected from $N(\mu, (10)^2)$. Find:-

[1] The distn of $\frac{5S^2}{100} = (0.05) S^2$

$\rightarrow$ [2] C such that $P(S^2 > C) = 0.05$

Solu: [1] $\frac{5S^2}{100} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(5)$ - distn

[2] To find P_{95} of S^2. Now P_{95} of $\frac{(n-1)S^2}{\sigma^2}$ is $\chi^2_{0.05}$

thus, P_{95} of $S^2 = \frac{\sigma^2 \chi^2_{0.05}(n-1)}{n-1} = \frac{100 * \chi^2_{0.05}(5)}{6-1} = \frac{100 * \boxed{3.84}}{5} = 76.8$.

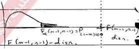
The Rules:-

$$P_{(1-\alpha)100} \text{ of } S^2 = \frac{\sigma^2 \chi^2_\alpha (n-1)}{n-1}$$

Distribution of $\frac{S_x^2}{S_y^2}$: Indep. samples:-

Let S_x^2 be the variance of a sample with size m selected from $N(\mu_x, \sigma_x^2)$.
Let S_y^2 be " " " " " n " " $N(\mu_y, \sigma_y^2)$. Then $\frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2}$ is a random variables. Its distribution is called an F-distribution with $(m-1, n-1)$ degrees of Freedom. It is denoted by $F(m-1, n-1)$-disn

$F_{0.1}(m-1, n-1)$



d.f.1 (m-1)

d.f.2 (n-1) | 1 2 3 4
1
2
3 — $F_{0.1}(2,3) = P_{90}$ of $F(2,3)$-disn

Fact:- $F_{1-\alpha}(k, l) = \frac{1}{F_\alpha(l, k)}$

Example:-

Find P_{10} of $F(3,5)$. - $F_{0.9}(3,5) = \frac{1}{F_{0.1}(5,3)}$

Solu:-

$$P_{10} \text{ of } F(3,5) = F_{0.9}(3,5) = \frac{1}{F_{0.1}(5,3)} = \frac{1}{5.31} = 0.188$$

Example: let S_x^2 be the variance of a random sample with size 6, selected from $N(\mu_x, 200)$. let S_y^2 be the variance of a random sample with size 4, selected from $N(\mu_y, 160)$

Find P_{90} of $\frac{S_x^2}{S_y^2}$??

Solu:-

P_{90} of $\frac{S_x^2}{S_y^2} \cdot \frac{\sigma_y^2}{\sigma_x^2} = F_{0.1}(6-1, 4-1)$ $\therefore P_{90}$ of $\frac{S_x^2}{S_y^2} = \frac{\sigma_x^2}{\sigma_y^2} \cdot F_{0.1}(5,3)$

$= \frac{200}{160} \cdot 5.31 = 6.6$

$\Rightarrow$ Distribution of $\hat{P}-\hat{q}$, indep. sample:-

$\hat{P} \sim N(P, \frac{P(1-P)}{m})$

$\hat{q} \sim N(q, \frac{q(1-q)}{n})$

$\hat{P}-\hat{q} \sim N(P-q, \frac{P(1-P)}{m} + \frac{q(1-q)}{n})$

Examples:- Suppose that 40% of all men and 30% of all women ~~Consider the following Information about two independent sample~~ in Jordan are smokers, Two independent random samples, each of size 100 are selected from men and women respectively. let $\hat{P}_m$, $\hat{P}_w$ be the proportions of smokers in the men's and women's samples, respectively.

[1] Find $P(\hat{P}_m > \hat{P}_w)$ [2] Find P_{90} of $\hat{P}_m - \hat{P}_w$.

[1] $\hat{P}_m - \hat{P}_w \sim N(0.40-0.30, \frac{(0.4)(0.6)}{100} + \frac{(0.3)(0.7)}{100})$

$= N(0.1, \frac{0.45}{100}) = N(0.1, (0.067)^2)$

[1] $P(\hat{P}_m - \hat{P}_w > 0) = P(Z > \frac{0-0.1}{0.067}) = P(Z > -1.49) = P(Z < 1.49) = 0.9314$

[2] P_{90} of $\hat{P}_m - \hat{P}_w = P_{90}$ of $N(0.1, (0.067)^2) = 1.28 \cdot (0.067) + 0.1 = 0.185$

Subject Date 15-12-2010 No.

$N \ \hat{P} = (P, P(1-P))$

$\frac{S_x^2}{S_y^2} = \frac{\sigma_x^2}{\sigma_y^2} F^n_{(M,L)}$

$\frac{n\sigma_x^2}{P-\bar{y}}$

$\Rightarrow$ Distribution of $\bar{x}-\bar{y}$: independent samples:

let $\bar{x}$ be the mean of a random sample with size m, selected from a population that has mean μ_x and variance σ_x^2.

let $\bar{y}$ be the mean of an independent sample with size n, " " " " " μ_y " " " σ_y^2

then $\bar{x}-\bar{y}$ is a random variable. $N(\mu_x-\mu_y, \frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n})$

Fact:-

If $m \geq 30$ and $n \geq 30$, then $\bar{x}-\bar{y}$ (almost) has

$N(\mu_x-\mu_y, \frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n})$ The conditions that $m \geq 30$ and $n \geq 30$ are not necessary if the respective populations are normal.

Remarks:

[1] $\frac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n}}} \sim N(0,1)$

[2] If $m \geq 30$ and $n \geq 30$ then:

$\frac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\frac{S_x^2}{m}+\frac{S_y^2}{n}}}$ (almost) has $N(0,1)$

N O T E B O O K

Fact:- If $m<30$ and $n<30$ and $\sigma_y^2=\sigma_x^2$

then $\dfrac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\dfrac{S_p^2}{m}+\dfrac{S_p^2}{n}}}$

$S_p^2=\dfrac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}$; is the pooled variance of the two samples.

Example:-

Let $\bar{x}, \bar{y}$ be the means of two independent random samples with size $m=50, n=60$ and variances $\sigma_x^2=200$
$S_y^2=\sigma_y^2=360$, selected from two populations with means
$\mu_x=70$ and $\mu_y=62$.

[1] what is the dis'n of $\dfrac{\bar{x}-\bar{y}-\overset{70-62}{8}}{\sqrt{\frac{200}{50}+\frac{360}{60}}}$? $\dfrac{(m-1)S_x^2+(n-1)S_y^2}{(m+n-2)}$

[2] Find P_{90} of $\bar{x}-\bar{y}$, [3] Find $P(\bar{x}>\bar{y})$

Solu [1] Since $m>30$ and $n>30$

$\bar{x}-\bar{y}$ (almost) $\sim N(\mu_x-\mu_y, \frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n}) \approx N(\mu_x-\mu_y, \frac{S_x^2}{m}+\frac{S_y^2}{n})$

$= N(8, \frac{200}{50}+\frac{360}{60})$

$\dfrac{\bar{x}-\bar{y}-8}{\sqrt{\frac{200}{50}+\frac{360}{60}}} \sim N(0,1)$

[2] P_{90} of $\bar{x}-\bar{y}$ = P_{90} of $N(8, \frac{200}{50}+\frac{360}{60}) = N(8, 10)$

$= N(8, (3.16)^2)$

So P_{90} of $\bar{x}-\bar{y} = (P_{90}$ of $N(0,1)) * (3.16) + 8 = (1.28)(3.16)+8 = 12$

[3] $P(\bar{x}>\bar{y}) = P(\bar{x}-\bar{y}>0) = P(Z>\frac{0-8}{3.16}) = P(Z>\frac{-8}{3.16})$

$= P(Z>-2.5) = P(Z<2.5) = 0.9938$

Example:- In the previous example assume that $m=5$, $n=6$ the populations are normal and σ's are equal. Find P_{90} of $\bar{x}-\bar{y}$.

Solu:-

$$P_{90} \text{ of } \frac{\bar{x}-\bar{y}-8}{\sqrt{\frac{S_p^2}{5}+\frac{S_p^2}{6}}} = P_{90} \text{ of } t(5+6-2)$$

$$= P_{90} \text{ of } t(9) = t_{0.1}(9) = 1.383$$

$$S_p^2 = \frac{(5-1)200+(6-1)360}{9} = \frac{2600}{9} = 288.89$$

Thus, P_{90} of $\dfrac{\bar{x}-\bar{y}-8}{\sqrt{\frac{288.89}{5}+\frac{288.89}{6}}} = 1.383$

So,

$$P_{90} \text{ of } \bar{x}-\bar{y} = (1.383)\sqrt{\frac{288.89}{5}+\frac{288.89}{6}} + 8 = 22.23$$

$\Rightarrow$ Chapter 6, 7, 8.

Concepts of statistical Inference:-

$$\text{Inference} \nearrow \text{Estimation}, \quad \text{Inference} \rightarrow \text{Hypothesis Testing}$$

Estimation & statistical

Definition:- A descriptive measure for a population is called a parameter. ليس متغير

A descriptive statistical measure for a sample is called a statistic. متغير عشوائي

In general notation:- For a parameter we use θ σ^2

" a statistic " " $\hat{\theta}$ S^2

Example:- مجتمع عينة

Parameter θ	Corresponding statistic $\hat{\theta}$
$\Rightarrow$ μ	$\bar{x}$
σ^2	S^2
P	$\hat{P}$
$\mu_x - \mu_y$	$\bar{x} - \bar{y}$

Estimation $\nearrow$ Point estimation — قيمة بحد ذاتها
Estimation $\searrow$ $(1-\alpha)100\%$ confidence Intervals

Point Estimation:-

Use $\hat{\theta}$ directly to approximate θ

i.e $\theta \approx \hat{\theta}$.

$(1-\alpha)100\%$ C.I:-

Use $\hat{\theta}$ and the random sample to give an interval

$(L(\hat{\theta}), U(\hat{\theta}))$ such that

$P(L(\hat{\theta}) < \theta < U(\hat{\theta})) = 1-\alpha$

كلما زادت الثقة ~~كلما~~ زادت ال interval

$(1-\alpha)100\%$ C.I (confident Interval) for θ, when $\hat{\theta}$ is standardized

to $N(0,1)$ or t-disn.

Notation:-

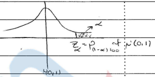
the Percentile $P_{(1-\alpha)100}$ at $N(0,1)$ is denoted by Z_α

Example:-

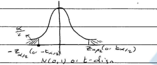
$Z_{0.1} = P_{90}$ at $N(0,1) = 1.28$

$Z_{0.05} = P_{95}$ at $" = 1.65$

$Z_{0.025} = P_{0.975}$ of $N(0,1) = 1.96$



$\Rightarrow$



(the standard disn of $\hat{\theta}$, let us say $\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$)

Thus:- $P\left(-Z_{\frac{\alpha}{2}} (\text{or } t_{\alpha/2}) < \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} < Z_{\frac{\alpha}{2}} (\text{or } t_{\alpha/2})\right) = 1-\alpha$

So:- $P\left(\hat{\theta} - Z_{\frac{\alpha}{2}} (\text{or } t_{\alpha/2}) \sigma_{\hat{\theta}} < \theta < \hat{\theta} + Z_{\frac{\alpha}{2}} (\text{or } t_{\frac{\alpha}{2}}) \sigma_{\hat{\theta}}\right) = 1-\alpha$

Therefore:-

a $(1-\alpha)100\%$ C.I. for θ is

$$\hat{\theta} \pm \underset{\text{or } t_{\alpha/2}}{Z_{\alpha/2}} \sigma_{\hat{\theta}}$$

let us see how to construct

$(1-\alpha)100\%$ C.I. for θ when $\theta = \mu, \mu_x - \mu_y, P, P-q$

[1] $\theta = \mu$

$(1-\alpha)100\%$ C.I. for μ

is $\bar{X} \pm \underset{\text{or } t_{\alpha/2}}{Z_{\alpha/2}} * \underset{\text{or } \frac{S}{\sqrt{n}}}{\frac{\sigma}{\sqrt{n}}}$

[2] $(1-\alpha)100\%$ C.I. for $\mu_x - \mu_y$:-

$(\bar{x}-\bar{y}) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n}}$

or

$Z_{\alpha/2}\sqrt{\frac{S_x^2}{m}+\frac{S_y^2}{n}} \rightarrow$ or $\left(t_{\alpha/2}\sqrt{\frac{S_p^2}{m}+\frac{S_p^2}{n}}\right)$

In case $\sigma_x^2 = \sigma_y^2$ = unknown ; small samples.

(3) $(1-\alpha)100\%$ C.I for p :-

$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(4) $(1-\alpha)100\%$ C.I for $p-q$:-

$$\hat{p}-\hat{q} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{m} + \frac{\hat{q}(1-\hat{q})}{n}} = 1-\alpha$$

Example:-

A random sample has size $n=10$, mean $\bar{x}=40$, Variance $S^2=100$ is selected from $N(\mu, \sigma^2)$. Use this sample to construct a 90% C.I. for μ.

Solution:-

The required C.I is

$$\bar{x} \pm \cancel{Z}_{\alpha/2}\, t_{(n-1)} \frac{S}{\sqrt{n}} = 40 \pm t_{0.05}(9) * \frac{10}{\sqrt{10}}$$

$$= 40 \pm 1.833 * \frac{10}{\sqrt{10}} = 40 \pm 5.79 \rightarrow (34.21, 45.79)$$

Example:- A random sample with size $n=100$ contains 30 smokers, use this sample to construct a 95% C.I for the proportion P of smokers in the population.

Solu:-

the required C.I is $\hat{P} \mp Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

$\alpha = 0.05 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$

So, the required C.I. is $0.30 \pm 1.96\sqrt{\frac{(0.30)(0.70)}{100}}$

$= 0.30 \mp 1.96\sqrt{\frac{0.21}{100}} = 0.30 \mp 1.96 * 0.046$

$= 0.30 \pm 0.090 = (0.21, 0.39)$

Example:-

The following table gives information about two independent samples.

	size	No. of smokers
Femals sample	100	25
males sample	100	30

Use these two samples to construct a 95% C.I. for $P_m - P_F$ where P_M (or P_F) is the proportion of smokers in the males (or Females) population.

Solu:-

The required C.I is $\hat{P}_M - \hat{P}_F \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}_M(1-\hat{P}_M)}{n} + \frac{\hat{P}_F(1-\hat{P}_F)}{n}}$

$(0.30 - 0.25) \pm 1.96\sqrt{\frac{0.30(0.70)}{100} + \frac{(0.25)(0.75)}{100}} \Rightarrow \boxed{}$

Example:-

the following table gives information about two indep. random samples, selected from two normally distd. populations with equal variances.

Subject Follow :- Date 27-12/2010 No. Monday

$\Rightarrow$ Use two these samples to construct a 90% C.I for $\mu_x - \mu_y$.

	size	Mean	variance
x-sample	$10 = m$	$60 = \bar{x}$	$100 = S_x^2$
y-sample	$5 = n$	$50 = \bar{y}$	$70 = S_y^2$

Solution:- عينات صغيرة σ مجهولة $\therefore$ t وليس $N(0,1)$

the required C.I is

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2}(m+n-2)\sqrt{S_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$S_p^2 = \frac{(10-1)*100+(5-1)70}{10+5-2} = \frac{1180}{13} = 90.8$$

So, the required C.I is $(60-50) \pm t_{0.05}(13)\sqrt{90.8\left(\frac{1}{10}+\frac{1}{5}\right)}$

$10 \pm (\approx 1.771)$

Exercise :- Solve the problem in Previous example with the Following modification:-

1) Sample sizes $m = 100$, $n = 50$ نستخدم $N(0,1)$ ونستبدل s_p^2

2) $m = 10$, $n = 5$, $\sigma_x^2 = 100$, $\sigma_y^2 = 70$ نستخدم σ و $N(0,1)$

$\bar{x}-\bar{y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n}}$

$\Rightarrow$ Sample Size :-

Definition:- the quantity $Z_{\alpha/2}\,\sigma_{\hat{\theta}}$ in the C.I is called estimation.

Remarks :- 1) If $\theta = \mu$, then error $= Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

2) If $\theta = p$, " " $= Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

3) Error $= \frac{1}{2} \cdot$ lenght of the C.I.

Example(6):-

Suppose that we want to estimate μ use a 95% C.I of length 4. Find the sample size if σ is estimated previously around 15.

Solu:-

$Error = \frac{1}{2}\ length = \frac{1}{2}\ 4 = 2$

$Error = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow 2 = Z_{0.025} * \frac{15}{\sqrt{n}} \Rightarrow 2 = 1.96 * \frac{15}{\sqrt{n}}$

$\sqrt{n} = \frac{1.96 * 15}{2} \Rightarrow \boxed{n = 217}$

$\Rightarrow$ Fact :: Maximum of $P(1-P) = 0.25$ $\quad 0 \le P \le 1$

Thus, if no previous information about P are available we use 0.25 in place of $P(1-P)$ in the error formula.

Examples

Suppose that we want to estimate $\hat{P}$ using a 95% C.I that is True to within 0.1 from the correct value. Find the sample size if:

(1) No previous Information about P is available.

(2) P is previously estimated around 0.1.

(*) $Error = 0.1 = Z_{0.025} \sqrt{\frac{P(1-P)}{n}}$

$0.1 = 1.96 \sqrt{\frac{P(1-P)}{n}}$

(1) $0.1 = 1.96 \sqrt{\frac{0.25}{n}}$

$\sqrt{n} = \frac{1.96 * 0.5}{0.1} \Rightarrow \boxed{n = 96}$

(2) $\sqrt{n} = \frac{1.96 * 0.3}{0.1} = 5.88 \quad \Leftarrow \sqrt{\frac{0.1(0.9)}{n}}$

$n = 35.$

$H_0: \mu \leq 30$
$H_1: \mu > 30$

Subject Date 29-12-2010 No.

→ Hypothesis Testing :-

Main goal :-

To use a random sample for testing an alternative hypothesis H_1 against a null hypothesis H_0 about population parameter θ.

Steps for Hypothesis testing :-

(1) write down H_0 and H_1.

(2) select a level of significance α for your test

Usually we use $\alpha = 0.05$ or 0.10 or 0.01 (عينة صغيرة ، كبيرة)

α is called the Probability of type 1 error

⇒ type 1 error :- is to reject H_0 given that H_0 is True.

The Probability of type 2 error is denoted By β.

⇒ Type 2 error :- is to reject H_1 given that H_1 is True.

H_1 عكس المساواة

(3) sketch the acceptance and rejection regions of H_0.

(4) compute a quantity called test statistic

If the test statistic is located in the acceptance region of H_0, then we reject H_1, otherwise we accept H_1

Examples :- (testing about μ) :-

A random sample with size 16 is selected from $N(\mu, 25)$ and has mean $\bar{X} = 10$. Can we deduce from this sample that :-

(1) $\mu > 8$? test at $\alpha = 0.05$ (right-sided)

(2) $\mu < 13$? " " $\alpha = 0.05$ (left-sided)

③ $\mu \neq 8$? test at $\alpha = 0.05$ (two sided)

① $H_0 : \mu = 8$

$H_1 : \mu > 8$

دائماً =

Rejection and acceptance regions:- // إذا كان الفحص عن μ يعني ⇐

Rej. and acce region for tests // تحت $N(0,1)$ أو T ←

about μ are located under $N(0,1)$ or t-dist.

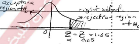
In our example is under $N(0,1)$ dist because σ is known

If we don't know σ then we use t-dist.

For right-sided tests (i.e $H_1 ; \Theta > \Theta_0$)

test statistic for test about μ :-

$$\text{test statistic} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$



if σ is not known we use S, In our example

$$\text{test statistic} = \frac{10 - 8}{\frac{5}{\sqrt{16}}} = \frac{8}{5} = 1.6$$

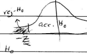
Since 1.6 is located in the acceptance region of H_0 we reject H_1. So we can't deduce that $\mu > 8$

②:] $H_0 : \mu = 13$, $\alpha = 0.05$

$H_1 : \mu < 13$

Our test is left sided, so the rejection region of H_0 is left-sided.

$$\text{test statistic} = \frac{10 - 13}{\frac{5}{\sqrt{16}}} = -2.4.$$



which belongs to the rejection region of H_0

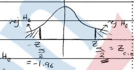
So, we accept H_1, So we can deduce $\mu > 130$

(3) $H_0 = 8$

$H_1 \neq 8$

our test is two sided, So the rejection region is two sided.

test statistic $= \frac{10-8}{\frac{5}{\sqrt{16}}} = 1.6$



located in the acceptance region of H_0

So we reject H_1, we can't deduce

Exercise:- Solve the Problem in the Previous example assuming that σ^2 is unknown and $S^2 = 25$

→ (the only difference is that we use t-dis) $t_{\alpha}^{(n-1)}$

Examples:- (testing about P):-

A random sample with size 100 contains 30 smokers. Can we deduce from this sample that the Proportion of smokers in the Population is Greater than 0.23. test at $\alpha = 0.1$. $\frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ $\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ $\frac{\hat{P}-P_0}{\sqrt{\cdots}}$

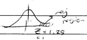
$H_0 :- P = 0.23$

$H_1 :- P > 0.23$ $\alpha = 0.1$

Rejection region :- right-sided. under $N(0,1)$

test statistic for test about P

$= \frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.30 - 0.23}{\sqrt{\frac{0.23(0.77)}{100}}} = \frac{0.07}{0.042} = 1.66$



So, we accept H_1

Example:- (testing about $\mu_x - \mu_y$):-

$H_0: \mu_x - \mu_y = (\mu_x - \mu_y)_0$

$H_1: \mu_x - \mu_y > (\mu_x - \mu_y)_0$

$< \text{ or } \neq$

$\frac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}}$

test statistic $= \frac{(\bar{X}-\bar{y}) - (\mu_x - \mu_y)_0}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$

if σ_x^2, σ_y^2 are not known and $m \geq 30$, $n \geq 30$ then use $S_x^2, S_y^2 \Rightarrow t$ or N

if $\sigma_x^2 = \sigma_y^2$ and " " " $m < 30$, $n < 30$

" $S_p^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}$ and $t(m+n-2)$ - disn

$\bar{x}-\bar{y}$ - $m+n-2$

Example:-

the following table gives information about two indep random samples, selected from two normal populations with equal variances.

Use two these samples to test $H_1: \mu_x - \mu_y > 3$ use $\alpha = 0.05$

	Sample size	Sample mean	Sample variance
X - Sample	10	$\bar{x} = 30$	$S_x^2 = 100$
y - Sample	5	$\bar{y} = 22$	$S_y^2 = 80$

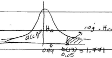
E-dis

Solu:-

$H_0: \mu_x - \mu_y = 3$

$H_1: \mu_x - \mu_y > 3$

test statistic $= \frac{\bar{X} - \bar{y} - 3}{\sqrt{S_p^2(\frac{1}{m} + \frac{1}{n})}} = \frac{(30-22)-3}{\sqrt{S_p^2(\frac{1}{10}+\frac{1}{5})}}$



$S_p^2 = \frac{(10-1)100 + (5-1)80}{13} = \frac{1220}{13} = 93.8$ $t(10+5-2)$ - disn.

So, test statistic $= \frac{5}{\sqrt{93.3(0.3)}} = \frac{5}{5.3} = 0.94$

this No is located in the acceptance region of H_0 So we reject H_1

Exercise :- Solve the Problem in the Previous example with the Following modification:

① $m=100$, $n=50$, every thing else is the same (S_p^2)

② $\sigma_x^2=100$, $\sigma_y^2=80$ " " "

Example :- testing about $P-q$, indep-sample.

$H_0: P-q=0$

$H_1: P-q>0$ or $H_1: P-q<0$ or $H_1: P-q\neq 0$

Rejection Region at H_0 :- Under $N(0,1)$

test statistic $\dfrac{\hat{P}-\hat{q}}{\sqrt{P^*(1-P^*)(\frac{1}{m}+\frac{1}{n})}}$

$P^*=\dfrac{m*\hat{P}+n*\hat{q}}{m+n}$

Example :- the following table gives information about two independent sample Can we conclude From this Sample that the Porportion of smokes in the males Population is greater than that in the Females Population test at $\alpha=0.05$

	Sample size	No of smoker
Female sample	150	10
Males sample	100	28

$P>F$

$H_0: P_m-P_F=0$

$H_1: P_m-P_F>0$ $\alpha=0.05$

acc H_0 rej. H_0 $Z_{0.05}=1.65$

test stat :- $\dfrac{\frac{28}{100}-\frac{10}{150}}{\sqrt{P^*(1-P^*)(\frac{1}{100}+\frac{1}{50})}}$, $P^*=\dfrac{10+28}{100+50}=0.25$

$=\dfrac{0.08}{\sqrt{0.25\cdot(0.75)(0.03)}}=\dfrac{0.08}{0.075}=1.06$ we reject H_1

Subject ✗ Date 3/1/2011 No.

Paired Samples:-

A paired sample consist of n subjects or experimental units each subject is associated with two related observations A (after or with) and B (before or without)

we are interested in drawing inference about μ_D

where $D = A - B$ (or $B - A$)

$A(1-\alpha)100\%$ C.I. for

μ_D is $\bar{d} \pm t_{\alpha/2}(n-1) \frac{S_d}{\sqrt{n}}$

$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$, $S_d = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - n(\bar{d})^2}{n-1}}$

where $d_1, d_2, \dots, d_n$ are the differences in our paired sample

$(d_i = a_i - b_i)$ $D = A - B$ / $B - A$

$H_0: \mu_D = 0$

رح نأخذ $H_1: \mu_D > 0$ or < 0 or $\neq 0$

Example:-

The diastolic Blood Pressure ~~for~~ of 4 patients after and before a certain medicine are given in the following table.

Patient	After A	Before B
1	80	88
2	85	90
3	78	85
4	80	85

Let $D = A - B$

① construct a 90% C.I for μ_d?

② can we conclude that the medicine is efficient? test at $\alpha = 0.05$

يعني سالب

N O T E B O O K

Solu:-

$d_i = a_i - b_i$	d_i^2
-8	64
-5	25
-7	49
-5	25
total -25	163

$\therefore \bar{d} = \frac{-25}{4} = -6.25$

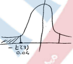
$S_d = \sqrt{\frac{163 - 4(-6.25)^2}{3}} = 1.5$

$\therefore$ ① the required C.I is $\bar{d} \pm t_{\frac{0.1}{2}}(3) * \frac{1.5}{\sqrt{4}}$

$-6.25 \pm (2.353)\frac{1.5}{2}$

② $H_0: \mu_D = 0$

$H_1: \mu_D < 0 \qquad \alpha = 0.05$



دالتي $\Rightarrow$ t-test st $= \frac{\bar{d} - 0}{\frac{S_d}{\sqrt{n}}} = \frac{-6.25}{\frac{1.5}{2}} = \frac{-12.5}{1.5} = -8.3$

So: we accept H_1.

العواد منح.