

WEEKLY NOTEBOOK

statistics I

summer semester
2021



* Introduction.

(天)

* retrospective $\therefore$ historical data

* Observational :: presently collected

↓ * designed experiment is to process

Passive slower input changes.

k mechanistic model

Phy
Now

(Population)

* type of

statistical

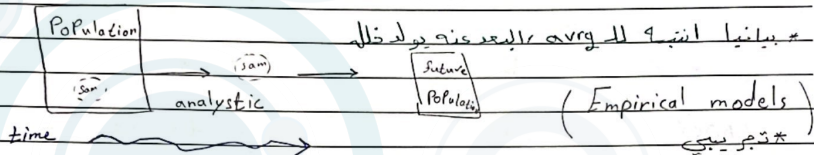
Reasoning

inference

Product designs

Sample

احتمالات



Controlled variables

٢٢٢ #

کالہ $\sqrt{s}$: سو ذرات

- فنہ امتحان

Temperature -

input

System

output

Noise variables

* experiment: operation or procedure carried out under controlled conditions, to discover unknown result.

* (E) Event: is subset of the sample space.

Random experiment ← تجربة الزرد

different outcomes

All possible outcomes,

(S) space ←

→ discrete

Continuous

finite

countable

infinite

Time

$$* S = P^+ = \{x \mid x > 0\}$$

continuous : متصلة *

٢٠٢١ / ١٢ / ١٠

(محدد)

۱. عمر مہر

Five Apple

11. 10/10/10

طويل الكف معدود.

* Ch: 2: Probability

Event Combination

1- Union ($A \cup B$): (or) (+)

2- Intersection ($A \cap B$): (both) (و)

3- Complement: (A', B'): (ليس)

4. $\emptyset$

* Rules:

$$* (A \cap B) = (B \cap A)$$

$$* (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$* (A \cup B)' = A' \cap B': \text{complement union} = \text{intersection of complement.}$$

$$* (A \cap B)' = A' \cup B': \text{(و ليس) (و ليس)}$$

* Multiplication rule: $= n_1 * n_2 * \dots * n_k$

* The total number of ways or outcomes are:

* Combination: (order doesn't matter). * identical.

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

لترتيب غير مهم

* Excel Combin(n,r).

* (order) : ترتيب

* Graphical represent

Tree diagrams.

ex done with several steps

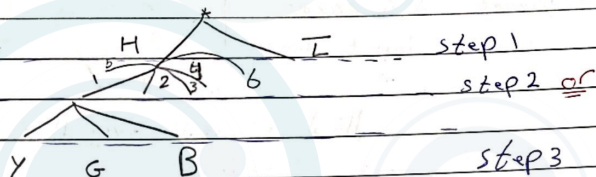
n1 step
n2 step

* prof

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

* tossing a coin, rolling a die, choose a ball
 $\{Y, G, B\}$.

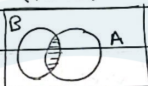
* مثال ١٥
 في المثال



* $(A \cap C)$

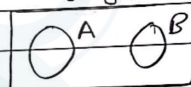


$(A \cap B)$



disjoint

⊕



exclusive.

$$P_8^{12} \left(\begin{matrix} 8 \text{ choices} \\ 12 \text{ items} \end{matrix} \right)$$

(order) matter: $GR \neq RG$ الترتيب مهم

* Permutation: how many ways can we select r from n .

$$1^{\text{st}} \quad 2 \quad 3 \quad \dots \quad k$$

$$(n) \quad (n-1) \quad (n-2) \quad \dots \quad (n-k+1)$$

Excel: $\text{permut}(n, r)$

$n > r$

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

Ex: 5 student to choose 3 students

$$\frac{5 \times 4 \times 3}{1 \times 1 \times 1} = 60$$

$$P_3^5 = \frac{5!}{(5-3)!} = \frac{120}{2!} = 60$$

* Similar items Permutations: (groups)

$$n = n_1 + n_2 + n_3 + \dots + n_r$$

قد يكون في
تأثير في الترتيب الواحد

$$* \text{calculated} = \frac{n!}{n_1! \times n_2! \times \dots \times n_r!} \quad (\text{without replacement})$$

* Probability: $P()$: chance that an outcome will occur. get it.

$$* P(S) = 1$$

* $A_1, A_2, \dots$ disjoint

$$* 1 = P(A) + P(A')$$

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n) = P(A_1) + P(A_2) + \dots$$

* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ * قد توقع موقع قانونه

$$\textcircled{*} \text{ if } E_1 \text{ contained in } E_2 \quad P_1(E_1) \leq P(E_2)$$

$\#$ if E

* Some events are exclusive so $P = E_1 + E_2 + E_3$

$$= P(E_1) + P(E_2) + P(E_3)$$

$$= P(S)$$

$$* P_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n! \quad \text{ordering } n \text{ items.}$$

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



$$* P(A \cap B) = \emptyset \quad \text{so} \quad P(A \cup B) = P(A) + P(B). \quad \text{exclusive.}$$

$$* P(E) = \text{all sample contain } (f) \quad (\text{الكل})$$

independent. # possible sample (الكل) التي تبدأ في أول stop (يكون مجموع ال (1=P)). (*)

$$* \text{exactly one:} = P(A) + P(B) - 2P(A \cap B).$$

one not both

$$* \{ \text{at most } x \} = \{ x \text{ or less} \}$$

$$P^* \text{ } \textcircled{*} P(T \cup B) = 1 - P(T' \cap B') = 1 - P(T') * P(B').$$

$$P(T' \cap B') = 1 - P(A \cup B) \quad \# \text{ } \# \text{ } \#$$

$$= 1 - P(T) + P(B) - P(T \cap B)$$



$$* P(T|B) * P(B)$$

(Tree) شجرة الاحتمالات

* Conditional Probability: when the prob change as a distri additional information because available. (غير مستقل)

$P(B|A)$: Prob B occurring given that event A occurred

* Knowledge of the occurrence of an event A will change the prob of event B $P(B|A)$ given

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

$P(B) = P(B|A) * P(A) + P(B|\bar{A}) * P(\bar{A})$

* Independence: لا تأثير لأحد على الآخر * conditionally independent

* $P(B|A) = P(B)$

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

* $P(A|B) = P(A)$

* $P\left(\bigcap_{i=1}^n E_i\right) = \prod_{i=1}^n P(E_i) = P(E_1) * P(E_2) * \dots$ مستقلين

* Bayes theorem: النسبة

$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$ * $P(B) > 0$

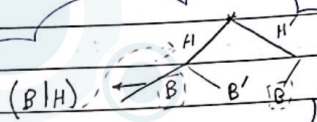
$P(E_i|B) = \frac{P(B|E_i) * P(E_i)}{\sum_{j=1}^n P(B|E_j) * P(E_j)}$

	(A)	(A')	
(B)	$P(A \cap B)$	$P(A' \cap B)$	$P(B)$
(B')	$P(A \cap B')$	$P(A' \cap B')$	$P(B')$
	$P(A)$	$P(A')$	

if $P(B) = 0.9$, doesn't conditional probability
 $P(B|A) = 0.9$ من صحت A انصرفت في لا يوجد

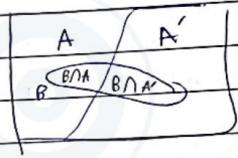
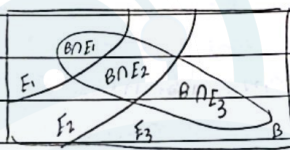
$$P(B) = P(H|B) + P(H'|B)$$

$$= P(H) \cdot P(B|H) + P(H') \cdot P(B|H')$$

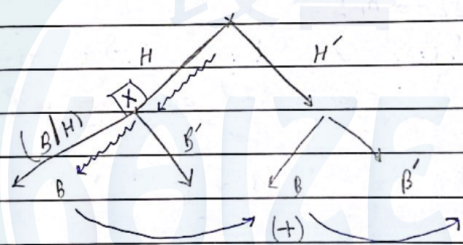


(and) $\Rightarrow \cap$
 meet specification

Important Rule:-



* $P(B) = \sum_{i=1}^n P(B|E_i) \cdot P(E_i)$ (Total Probability Rule).



* Series system: a 1 2 3 b
 • each components works/fails independently
 • system work if there is a path.

$P(\text{system to work}) = P(A_1 \cap A_2 \cap A_3 \dots)$
 $= P_1 \cdot P_2 \cdot P_3 \dots$

* Parallel system: a 1 2 3 b

* $P(\text{system to work}) = P(A_1 \cup A_2 \cup A_3 \dots)$

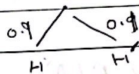
$= 1 - (\text{system doesn't work})$

* $P(A'_1 \cap A'_2 \cap \dots)$

H

quick ex: $P(\text{high level contamination}) = 0.1$, 5 samples checked and the sample are independent.

(a) $P \rightarrow$ none contain high level of contamination.



$$P(B) = \prod_{i=1}^n P_i = P(H_1) * P(H_2) * P(H_3) * P(H_4) * P(H_5) = 0.9^5 = 0.59$$

(b) $P \rightarrow$ one exactly contain high contamination.

$$= P(H_1, H_2) * P(H') = (0.9)^4 * (0.1) = 0.06561$$

(c) $P \rightarrow$ at least one contain high contamination.

$$P(B') = 1 - P(B) = 1 - 0.59 = 0.41$$



* random variable: $\{P(X=x)\}$ specific value.
(R.V.) $P(X=6)$

continuous. $\rightarrow$ discrete

ch2: استاذة

ch3: اللي دة لقا

في كذا الايام

المتغير العشوائي والتوزيع الاحتمالي

Ch3:- Discrete Random Variables & Probability distribution.

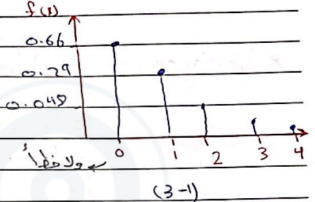
function

لا يتم من نفسه نتيجة بل من الذي ينجح في نفسه

* Figure (3-1) Probability distribution for bits in error.

$P(X=0) = 0.6561$
$P(X=1) = 0.2916$
$P(X=2) = 0.0486$
$P(X=3) = 0.0036$
$P(X=4) = 0.0001$

Total = 1.0000



* $P(X=r_i)$ probability mass (ind) $\rightarrow$ (PMF) function.

$$f(x) = P(X=x)$$

$$① f(x) \geq 0$$

$$② \sum_{i=1}^n f(x_i) = 1 \quad (1)$$

* تعريف (variable)

Ex *3-5 | X denote the number of wafers

* $P(\text{wafer contains large particle}) = 0.1$
present, absent $P(a) = 0.9$

$$S = \{p, ap, aap, aaap, \dots\}$$

$P(X=1)$	0.1	0.1
$P(X=2)$	0.1×0.9	0.09
$P(X=3)$	$0.1 \times (0.9)^2$	0.081
$P(X=4)$	$0.1 \times (0.9)^3$	0.0729
		0.3439

③ $X=5$ or more $\rightarrow$

$$P(X \geq 5)$$

$$P(X \geq 5) = 1 - P(X < 5) = 1 - 0.3439$$

$$* P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 0.9999$$

PMF CDF - CDF

$$* P(X=3) = P(X \leq 3) - P(X \leq 2) = 0.0036$$



المجال

Five Apple

* Cumulative distribution Function (CDF)

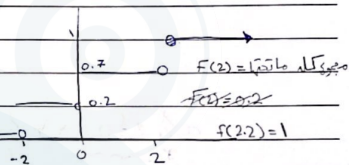
$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

$$1 \geq F(x) \geq 0$$

$$x \leq y, F(x) \leq F(y).$$

ex 1)

$F(x) =$	0.0	$x < -2$
	0.2	$-2 \leq x < 0$
	0.7	$0 \leq x < 2$
	1	$2 \leq x$



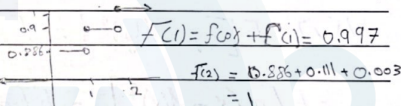
ex 2 850 parts, 50 defective, Two selected without replacement.

$x = \text{number of D}$

$$P(X=0) = \frac{800}{850} \cdot \frac{799}{849} = 0.886$$

$$P(X=2) = 0.003$$

$$P(X=1) = \frac{2 \cdot 800}{850} \cdot \frac{50}{849} = 0.111$$

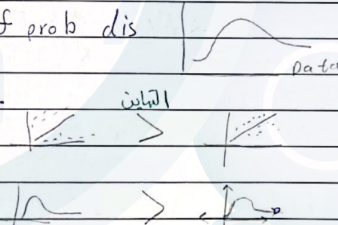


* Weight avg: $\sum x_i \cdot f(x_i) + \sum x_j \cdot f(x_j)$

* Mean: measure of the center probability dis

* Variance: measure of dispersion of prob dis

* Standard deviation: $\sqrt{\text{variance}}$



(4)
Mean or expected value = $E(X) = \sum_{x=-\infty}^{\infty} x \cdot f(x)$
 x value
 $f(x)$ PMF
 Weight (Probability)
 . avg



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

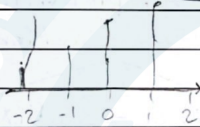
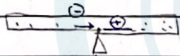
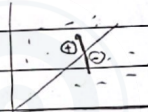
$$= (1.25) \times 1.20$$

* Variance (σ^2) = $V(x) = E[(X - \mu)^2]$

$$= \sum_x (x - \mu)^2 \cdot f(x) \text{ prob}$$

$$= \sum_x x^2 \cdot f(x) - \mu^2$$

relate



① CDF $F(x)$

$$P(2 < x < 6) = P(3 \leq x \leq 5)$$

$$E[X^2] = \sum x^2 \cdot f(x)$$

probability

function

$$\mu = (E[X])^2$$

$$E(X^2) \neq (E(X))^2$$

Ans: 1

new function

$$\mu^2 = 5^2 = 25$$

* Function Random variable. (X discrete random variable with probability mass function $f(x)$)

$$E[h(x)] = \sum_x h(x) \cdot f(x)$$

ex: X is number of error. what is the expected value of the square of the number of bits in error?

x	$f(x)$	$x^2 \cdot f(x)$
0	0.651	0
1	0.2916	0.2916
2	0.0486	0.1944
3	0.0036	0.0324
4	0.0001	0.0016

$$E(x^2) = 0.52$$

+ Discrete

* Distribution التوزيع

Uniform ①

↓
equal like

$$f(x) = \frac{1}{n}, (1-p) \text{ geo}$$

Probability distribution

0.2 0.2 0.2

1 1 1 1 1
0 1 2 3 4

General

$\min a < b \max$

$$* f(x) = 1$$

Probability $(b-a+1)$

$$* E(x) = \frac{\max + \min}{2}$$

$$* \sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

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

Five Apple

$$\sigma^2 = \frac{\mu^2}{12}$$

* Rule for $E(x)$, $V(x)$

$$\sum_1^{\infty} f(x) = 1$$

$$y = ax + b$$

* a, b constant

$$\mu_y = E[y] = E[ax + b]$$

لا يؤثر في القيمة

shifting *

$$= \sum_1^{\infty} (ax + b) \cdot f(x)$$

$$= \sum_x ax f(x) + \sum_x b f(x)$$

$$= a \sum_x x f(x) + b \sum_x f(x)$$

$$= a E[x] + b$$

$$\mu_y = a \mu_x + b$$

$$V[y] = E[(y - \mu_y)^2] = E[(ax + b) - (a\mu_x + b)]^2$$

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

$$\downarrow$$

$$V[x]$$

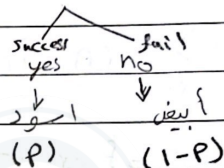
$$* V[y] = a^2 V[x]$$

لا يـ

$$* E[y] = a E[x] + b$$

②A) * Bernoulli trial: exp with two outcome^①

* Probability of success in each trial is constant (p).



* out comes of successive trial are independent.

$$P = P.P.P.P \dots_k$$

النتائج متباعدة *

a. ✓ 0.25

كل محاولة مستقلة *

b. x

ind = النسيء

c. x

لا تتأثر بالنتيجة السابقة

d. x

0.75

{ n = number of trial.
x = successes }
العدد الكلي للمحاولات

② Binomial Random Variables. (Bernoulli trials)

* Flip a coin 10 times $X = \# \text{ head}$

(series)

→ count the number of success (head)

* A worn tool produces 1% defective parts $X = \# \text{ defective parts}$
 : successful for me to find defective.

$$* f(x) = C_n^x p^x (1-p)^{n-x}$$

سلسلة

for $x = 0, 1, \dots, n$

* BINOMDIST(x, n, p, TRUE)

* $0 < p \leq 1$

$n = 0, 1, \dots$

ex: error = 0.1 let $X = \text{number of bits error in the next 4 bits transmitted.}$

① Find $P(X \neq 2)$ $C_2^4 * (0.1)^2 (1-0.1)^{4-2}$
 $C_2^4 * (0.1)^2 (0.9)^2$

② constant at least 4 sample pollution in next 18 sample
 $P(\text{Pollution}) = 0.10$

$$* P(X \geq 4) = 1 - P(X \leq 3)$$

$$1 - \sum_{x=0}^3 C_x^{18} (0.1)^x (0.9)^{18-x}$$

③ $3 \leq X \leq 7$:

$$P(3 \leq X \leq 7) = \sum_{x=3}^7 C_x^{18} (0.1)^x (0.9)^{18-x}$$

* Binomial Mean and Variance.

$$\mu = E(X) = n * p$$

$$\sigma^2 = V(X) = np * (1-p)$$

$$= \mu (1-p)$$

(x, p)

(4). Geometric Distribution (keyword: until), (detected) نجاح أول

* number of trial = unknown (غير معروف)

* $P(X=5) = (0.0000 F)$ ∴ First 4 are correct. ∴ $P(F) = 0.1$

$$\therefore P(X=5) = 0.9^4 \times 0.1 = 0.0656$$

* $X =$ The number of trial until the first success → you choose it.

R * $f(x) = p(1-p)^{x-1}$

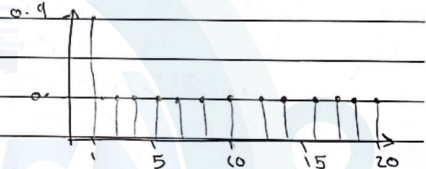
probability of success
∴ one parameter.

$x = 1, 2, \dots, \infty$ (أولاً وليس آخر)

* كيف جينا القافون: $1-p, 1-p, 1-p, 1-p, p$

Geometric distribution تحتاج

you need one trial to *
the first success.



* Mean $E(X) = \frac{1}{p}$

* Variance $E(V) = \sigma^2 = \frac{1-p}{p^2}$

* SD = $\sigma = \frac{\sqrt{1-p}}{p}$

$\mu = 1 = \frac{1}{0.1}$ متوسط (العدد المتوقع لنجاح أول)
trial → 1st success

* $\frac{1}{0.25} = 4$ * الحاجة لعدد القوية
عمرات لأول نجاح

* lack of Memory Property

II Independent.

$P(F) = 0.1$, Suppose 50 bits transmitted.

100 → 106
0 → 6 Same.

what is the mean number of bits transmitted until the next error? (the mean number - next error)

$\frac{1}{0.1} = 10.$

$\bar{q}_i, \bar{q}_0 \neq$

① Bin

- * Fixed n of trials
- * Random n of success

② Geo

- Random
- Fixed (1)

③ Neg Bin

- Random
- Fixed (r)

Ex

ex: X binomial random variable with $p=0.1$ and $n=10$

calculate the following from binomial probability mass function and also binomial table.

①

$$(a) P(X \leq 2) = P(X=2) + P(X=1) + P(X=0)$$

$$= f(2) + f(1) + f(0) \\ = \sum_{x=0}^2 C_x^n p^x (1-p)^{n-x} = \frac{10!}{2!} (0.1)^2 (0.9)^8 + \frac{10!}{1!} (0.1)^1 (0.9)^9 + \frac{10!}{0!} (0.1)^0 (0.9)^{10}$$

$$(b) P(X > 8) = P(X \geq 9) = P(X=9) + P(X=10) = \sum_{x=9}^{10} C_x^n p^x (1-p)^{n-x} \\ = 1 - P(X \leq 8)$$

$$(c) P(X=5) = C_5^{10} (0.1)^5 (0.9)^5 = 0.00149.$$

$$(d) P(5 \leq X \leq 7) = P(X=5) + P(X=6) + P(X=7)$$

② Table

$$(a) P(X > 8) = 1 - P(X \leq 8)$$

$n=10, p=0.1$

(1) = 2.90

5 * Negative Binomial definition. (x, p, r)

ex: $P(F) = 0.1$, Assume that the transmissions are independent,

Let $(X = \# \text{ of bits transmitted until the } 4^{\text{th}} \text{ error.})$

rs * given # of success (r).

* required # of trial.

1st 2nd 3rd 4th
F 0 0 0 0 0 F 0 0 F 0 0 0 0 0 F

$$* f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r}$$

* $x = \# \text{ trial } (x = r, r+1, r+2, \dots)$

* $r = \# \text{ success}$

تجارب النجاح x

متطلبات نجاح $r > x$ لازم بقا الفاشل

* geometric $r=1$

(r/p) نجاح *

$$P(X < r) = 0$$

$$* \mu = E(X) = \frac{r}{p}$$

$$\sigma^2 = V(X) = \frac{r(1-p)}{p^2}$$

$$SD(X) = \sqrt{\frac{r(1-p)}{p}}$$

* الفرق عن (geometric) هو $(*r)$ اوجد اكل من Success

x r p

until number $\leq 7 * \ln$ Excel = NEG-BI NoMdist (2, 3, 0.005), True False

قسط #

2. Find the (P) that in next 18 sample, exactly 2 contain the pollutant.

* Binomial

→ Independent

2- What the (P) that exactly 125 wafers need to be analyzed before a particle is detected.

* Geometric

3- The (P) that B errors over the first 9 trials

* Negative Binomial

F	S	S	S	F	F	F	F	$\sqrt{S}$
1	2	3	4	5	6	7	8	9

$$p^{r-1} (1-p)^{(x-1)-r-1}$$

$$p^{r-1} (1-p)^{x-r} = p \text{ success}$$

(x) 9-1 trials

(r) 4-1 success

كيف

بين القانون

كم طريقه بترتيبوا فيها

$$\binom{x-1}{r-1} p^r (1-p)^{x-r}$$

6. Hypergeometric Distribution:

* Trial not independent, Tree diagram used.

* $N \rightarrow$ objects contains 850

* $k \rightarrow$ objects classified as success 50

* $N-k \rightarrow$ (Failure) = $850-50=800$.

* $N \geq k$

* $N \geq n \rightarrow$ from then

(n): sample of size (n) is selected without replacement. (as 850)

$$f(x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

combination

x : denote the number of success.

* minimum $= x = \max(0, n+k-N)$
 * maximum $= \min(k, n)$

Ex: 50 parts defective from 850, Two sampled, $X = \#$ of defective.

* Find hypergeometric distribution?

* $k=50, N=850$

* $x=0, n=2$

$$* P(X=0) = \frac{\binom{50}{0} \binom{800}{2}}{\binom{850}{2}} = 0.886$$

$$* P(X=1) = \frac{\binom{50}{1} \binom{800}{1}}{\binom{850}{2}} = 0.111$$

* Needs 3 parameter. (N, k, n) .

$$\mu = E(X) = np, \quad \sigma^2 = V(X) = np(1-p) \frac{(N-n)}{N-1}$$

where $p = \frac{k}{N}$

(X n K N)
 * In Excel HypGeoMIST (6, 2, 50, 850) True False

* Some time $N = A + B = 200 + 100 = 300$
 .Allis

smult. * $P(X \geq 1) = 1 - P(X = 0)$

$X = \begin{cases} \text{Minimum} = [\max(0, n+k-N)] \\ \text{Maximum} = [\min(k, n)] \end{cases}$

maximum و Minimum بين X

(i) $n=4$
 $\max = \min(8, 4)$
 $n+k = 12 < 16$
 $\min = \max(0, -4)$
 $= 4$
 Min value of $X = 0$ *

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

$N=16$

$k=8$

*

$n+k > N$ sample might cover all failure + some success

$n=10$

$n+k > N$

$10+8 > 16$

8 failure و 8 success

2 success →

$\min = \max(0, 18-16)$

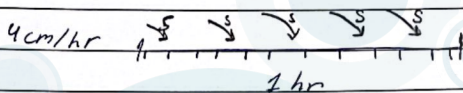
$(0, 2)$ $\min=2$

$\max = \min(8, 10) = 8 \rightarrow X$

$P(\text{at least of one}) = P(X \leq 2) = 1 - P(X = 3) =$

incl $\nearrow$ Per $\nearrow$ specific situation

7. Poisson distribution:- $\left[\frac{\# \text{ cm/hr, time, volume, area, ...}}{\text{per unit time}} \right]$



λ = rate of (np)

$\hookrightarrow$ Probability of successes.

$$* PMF = f(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

$$* E(x) = V(x) = \lambda$$

$$* \lambda > 0$$

$$* X = 0, 1, 2, 3, \dots$$

$$* e^{-\lambda} \text{ constant.}$$

* Excel: Poisson = (x, λ , F)

$$* \lambda = np \quad \left(\rightarrow p = \frac{\lambda}{n} \right)$$

* unit length.

ex: # of flaws follows a Poisson distribution of 2.3 flaw per mm. Let X denote the number of flaws in mm

Find: ① P of exactly 2 flaws in 2mm?

$$P(X=2) = \frac{e^{-2.3} 2.3^2}{2!} = 0.265$$

② P of 10 flaws in 5mm?

$$* E(x) = \lambda_{\text{new}} = 2.3 * 5 = 11.5 \text{ flaws/5mm}$$

$$* P(X=10) = \frac{e^{-11.5} 11.5^{10}}{10!} = 0.1129$$

③ P of at least 1 flaws in 2mm of wire?

$$* \lambda_{\text{new}} = 2.3 * 2 = 4.6 \text{ flaws/2mm}$$

$$* P(X \geq 1) = 1 - P(X=0) = 1 - \frac{e^{-4.6} 4.6^0}{0!} = 0.9899.$$

$$④ \text{ P of 12 or fewer? } P(X \leq 12) = \sum_{x=0}^{12} \frac{e^{-10} 10^x}{x!} = P(X=0) + P(X=1) + \dots + P(X=12).$$

if $\lambda = 10$ in 1mm

Five Apple

* example of Poisson Processes:

- 1- Particle of contamination per wafer. جسيمات التلوث لكل رقاقة
- 2- flaw per roll of textile. عيوب في لفافة النسيج
- 3- calls at a telephone exchange per hour. المكالمات في المحطة بالساعة
- 4- Power outages per year. انقطاع التيار الكهربائي سنويا

$\lambda = 3.4$ per mm

if Avg per 10-mm

$\lambda_{new} = \frac{34 \text{ flaws}}{10 \text{ mm}}$

$$\lambda_{new} = \lambda_{old} \times t$$

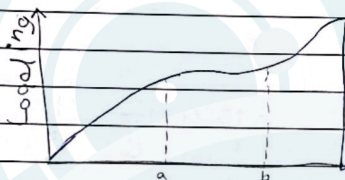
Chapter 4: Continuous distribution:

Continuous R.V → finite $[1, 3]$

البداية والنهاية معروفة
فترة محددة

→ length, time, Temperature. → Infinite $[0, \infty), (-\infty, 1]$

$X = \dots$ continuous random variable



$[0.98, 101.5]$ infinite numbers.

* loading beam between a, b
is integration between a and b.

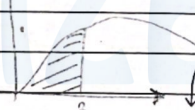
$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \left(\text{مجموع القيم الموزعة على 1} \right), \quad f(x) \geq 0 \quad \int_a^b f(x) dx = 1 \quad \text{الفترة المبرر عليها}$$

$$* P(a < X < b) = \int_a^b f(x) dx \quad * f(x) = 0 \text{ there is no area at } x$$

$$P(X_1 < X < X_2) = P(X_1 < X < X_2) = P(X_1 < X < X_2) = P(X_1 < X < X_2)$$

which $P(X=X) = 0$

مقدار (0) غير مهم



في ال continuous d's

* الهندسة الواسطة يخدم

ال سوال عزاحة ال قيمة معينة (c) $P(x=2) = 0$ Zero
بحد ذاتها = صفر

* The curve

$$P(x=3) = \int_3^3 f(x) dx = 0$$

(1) عند أكبر قيمة

$$P(X < 4) = \int_0^4 f(x) dx$$

$$P(1 < X < 2.5) = \int_1^{2.5} f(x) dx$$

$$P(3 \leq X) = \int_3^{\infty} f(x) dx$$

$$P(X < X) = 0.20$$

$\int_{-\infty}^{\infty} f(x) dx = 1$
Five Apple
x = number.

* The cumulative distribution function (CDF):

$$y = 5x + 2 \quad (\text{PDF}) (f(x))$$

$$\int y = \frac{5}{2} x^2 + 2x + C \quad (\text{CDF}) (F(x))$$

$$\int \text{PDF} = \text{CDF} \rightarrow \text{متكامل}$$

* CDF

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad -\infty < x < \infty$$

as function $\int_{-12.5, 20}$

continuous.

$$\text{Mean } (\mu) = \int_{-\infty}^{\infty} x f(x) dx$$

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

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \underbrace{(\mu^2)}_{\text{مربع المتوسط}}$$

$$* E[h(x)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

* Squared = (x^2) $\rightarrow$ $h(x)$ function

Continuous uniform distribution.

$$\textcircled{1} \text{Mean} = \frac{(a+b)}{2}$$

$$\textcircled{2} \sigma^2 = \frac{(b-a)^2}{12}$$

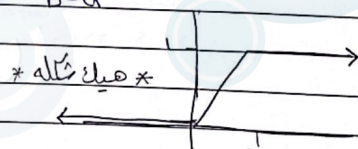


(a) $f(x) = 0$
(b) $f(x) = 1$

* Continuous uniform CDF.

$$F(x) = \int_a^x \frac{1}{(b-a)} du = \frac{x-a}{b-a}$$

$$f(x) = \begin{cases} 0, & x < a \\ \frac{(x-a)}{(b-a)}, & a \leq x < b \\ 1, & b \leq x \end{cases}$$



$$* \quad \underbrace{F(x)}_{(-\infty, \infty)} = \int_a^x f(x) dx \quad * \quad f(x) = \frac{dF(x)}{dx}$$

e: $f(x) = 0.25$, $0 < x < 4$

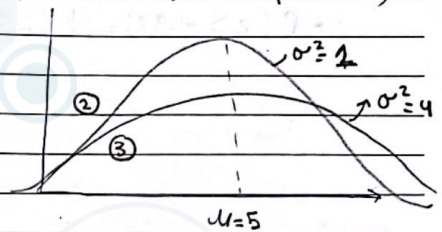
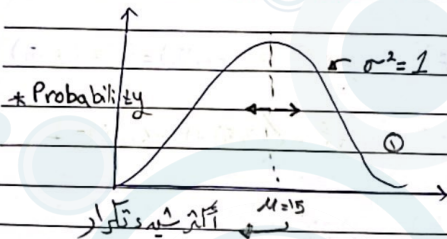
$$\mu = \int_0^4 x f(x) dx = \int_0^4 0.25 x = \frac{x^2}{2} \cdot 0.25 = 2$$

$$\sigma^2 = \int_0^4 f(x) (x - \mu)^2 = \int_0^4 0.25 (x - 2)^2 = \frac{0.25 (x - 2)^3}{3} \Big|_0^4 = \frac{4}{3}$$

(ملاحظات من إله)

(4-6) Normal distribution:

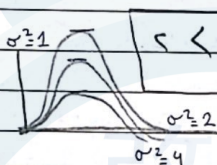
* يستخدم في قياس الانحراف في التجارب
(Normal for experiment)



* location = Mean

* Spread data = Variance.

location 3 = location 2
variance 3 > variance 2

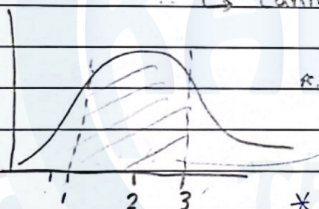


* كلما قل Variance كلما ازداد y-axis

2 parameter (μ, σ)

* PDF of Normal distribution: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$
 $\therefore CDF = F(x) = \int f(x)$ $\sigma > 0$

(cannot be integrated, so use table III)
(742)-(743)



* $\mu=2$

* $F(3)$ = area under pdf to the left of 3

$$* F(3) - F(1) = \int_1^3 PDF$$

$$* F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

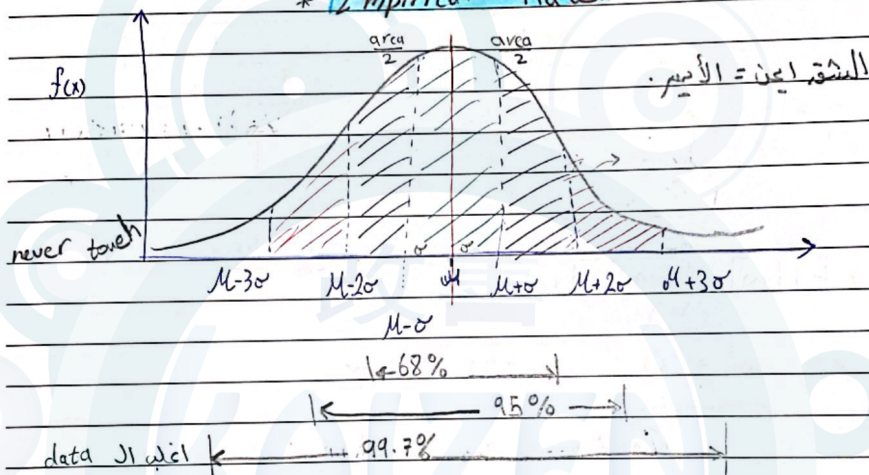
* $N(\mu, \sigma^2)$

* Excel = NORMDIST (x, μ, σ, F).

key word for normal (excessed)

$$* \int_{-\infty}^{\infty} pdf = 1 - \int_{-\infty}^{\infty} pdf$$

* Empirical Rule.



$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545 \quad \mu + 3\sigma$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973 \rightarrow \int_{\mu - 3\sigma}^{\mu + 3\sigma} f(x) dx = 0.9973$$

$$\textcircled{+} P(X < \mu) = P(X > \mu) = 0.5$$

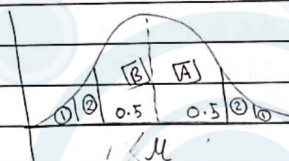
$$\mu = 75, \sigma = 15$$

$$[75 - 15, 75 + 15] = [60, 90] = 68\%$$

Standard Normal Distribution. (Z) Table

$\mu = 0$ and $\sigma^2 = 1$

$\Phi(z) = P(Z \leq z) = F(z)$



σ مقياس μ عن x $\rightarrow$ مقياس x

as a mirror.

standard normal random * $Z = \frac{(x - \mu)}{\sigma}$

* جدول z مقياس x

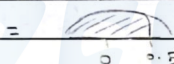
* x مقياس z

* $P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z)$

ex: $\mu = 10$ $\sigma = 2$ what probability the current between 9, 11?

$P(9 < x < 11) = P\left(\frac{9 - \mu}{\sigma} < \frac{x - \mu}{\sigma} < \frac{11 - \mu}{\sigma}\right) = P\left(\frac{9 - 10}{2} < z < \frac{11 - 10}{2}\right)$

$= P(-0.5 < z < 0.5) = P(z < 0.5) - P(z < -0.5)$



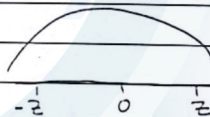
* Symmetric about (0) include 99% of all readings.

$P(-z < Z < z) = 0.99$

$P(-x < N < x) = 0.99$

$\rightarrow$ جدول z

$P(-2.58 < z < +2.58)$



* اقل قيمة في الجدول -3.9

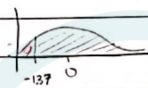
$X = 0.2500 \pm 0.0015$

$z_{0.99} = -3.9$ in الجدول

* $P(z < -4.6) = 0.000$, * $0.2485 < x < 0.2515$

$$* P(Z > 1.26) = 1 - P(Z < 1.26) = 0.1038$$

$$* P(Z > -1.37) = P(Z < 1.37) \quad , * P(Z > -2.15) = P(Z < 2.15)$$



$$* P(Z < -2.15) + P(Z < 2.15) = 1$$

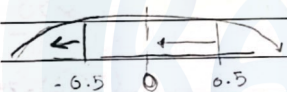
* قواعد بخصوص z

① الوفاة (الطبيعي) * التعميم جافو $P(Z < 1.2)$

② في حالة $(z < 1.2)$: الحل $P(Z > 1.2) = 1 - P(Z < 1.2)$
 z أكبر من z : z

③ z بين فترتين : الحل : الفرق الأكبر من الأصغر
 $* P(Z \leq 1.5) + P(Z \leq -0.5)$

$$P(Z > 0.5) = P(Z < -0.5) \quad ④$$



$$1 = P(Z < 0.5) + P(Z < -0.5) \quad ⑤$$

excel للتحليل العكسي

$$Z = \text{NORMINV}(0.97, \mu, \sigma)$$

(العكس)

$$0.97(0.97) = Z$$

اللي جافو $[X]$

$$* P(X \leq x) = 0.5$$

$$[Z=0]$$

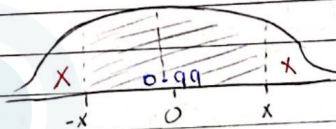
$$\text{so } x = \mu$$

* تكون x من μ

$$0.5 = 0$$

$$z=0 = z$$

* Symmetric about (0)



* $P(-z < Z < z) \Rightarrow 0.99$

$p(Z < -z) = p(Z > z)$ symmetric.

$1 - 2(P(Z < -z)) = 0.99$

$P(Z < -z) = \frac{1 - 0.99}{2} = 0.005$ جدول $\rightarrow$ $z_1 = -z$
 $z_2 = z$

4 ← عند z في استطيع اخراج x من طاوله
 0 ←

② Binomial , Possion = discrete , approximate) ليست نرمال (continuous) Normal (توزيع الاحتمال) (توزيع الاحتمال)

Keywords: التقريب > العادي

Normal Approximations:

* $n =$ # of object contains.

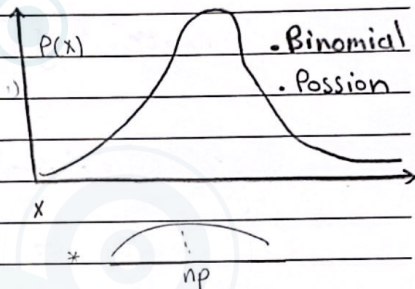
* $p =$ probability

• is good for the:

n very large.

- Binomial if $np > 5$, and $n(1-p) > 5$

- Poisson if $\lambda > 5$.



$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - \mu}{\sigma}$$

* Continuity correction: $P(X \leq x) = P(X \leq x + 0.5) = P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$

التي بها
ابديتها
بمقدار 1/2

(مدين)

$$* P(X \leq x) = x + 0.5$$

$$* P(X \geq x) = x - 0.5$$

$$* P(X \leq 150) = P(X \leq 150.5)$$

$$* P(a \leq X \leq b) = P(a - \frac{1}{2} \leq X \leq b + \frac{1}{2})$$

(مدين)

PX: $n=200$, $p=0.3$

لا يمكن بين مدين - ليس نقطة محددة

Approximate $P(X \leq 70)$

$P(X=80)$ continuity correction

$$P(79.5 \leq X \leq 80.5)$$

نفس

table

$$P(Z \leq \frac{70 + 0.5 - 60}{\sqrt{60(1-0.3)}}) \Rightarrow P(Z \leq 1.62) = 0.94738$$

$$P(70 < X < 90) \Rightarrow P(71 \leq X \leq 89) \Rightarrow P(\frac{71 - 0.5 - 60}{\sqrt{42}} \leq Z \leq \frac{89 + 0.5 - 60}{\sqrt{42}})$$

$$P(1.620 \leq Z \leq 4.55) = P(Z \leq 4.55) - P(Z \leq 1.62)$$

$$= 1 - 0.94738 = 0.05262$$

أكبر من 3.99 قيمة Z في جدول
1 =

Five Apple

• Poisson ($\lambda > 5$) $\sim$ (np > 5) Binomial $\approx$ Poisson $\neq$

So can be $\downarrow$

$$\leadsto Z = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0,1)$$

normal approximate

Ex: $X \sim \text{Pois}(6) = \mathbb{P}$

a) Exact $P(X < 4) = \sum_{i=0}^3 \frac{e^{-\lambda} \lambda^i}{i!} = \sum_{i=0}^3 \frac{e^{-6} 6^i}{i!} = \frac{e^{-6} 6^0}{0!} + \frac{e^{-6} 6^1}{1!} + \frac{e^{-6} 6^2}{2!} + \frac{e^{-6} 6^3}{3!}$
 $= 0.1512 \quad \cdot (CH3)$

b) $P(X < 4) = P(Z < \frac{4-6}{\sqrt{6}}) = P(Z < -0.82) = 0.206108$

if we use continuity correction:

$$P(X < 4) = P(X \leq 3) = P(Z \leq \frac{3+0.5-6}{\sqrt{6}}) = P(Z \leq -1.02) = 0.1539$$

c) Approximate the $P(8 < X < 12)$

$$P(8-6 < Z < 12-6) = P(0.82 < Z < 2.45) = 0.1999 \quad \text{"Z table"}$$

with continuity

$$P(8 \leq X \leq 11) = P(8-6 < Z < 10.5-6) = P(1.02 \leq Z \leq 2.25) \\ = 0.1416 \quad \checkmark \quad \text{Z table}$$

approximate discrete

* Percent error approximation = $\frac{\text{continuity} - \text{Exact}}{\text{Exact}} \times 100\% = \frac{0.1539 - 0.1512}{0.1512} = 0.017$
 $= 1.7\%$

* Exact = no normal approximate.

↳ back to ch3.

or equal

$$70 < X < 90$$

$$71 < X < 89$$

$$70.5 < X < 89.5$$

* approximate (6, 20) ...

* Normal approximation

to Hypergeometric

$$p = \frac{k}{N}$$

* sample size

$$n < 0.1 N$$

are small relative to population size

• Binomial = # of successes in specific number of trials

Ex. 5.7

• Poisson = # arrivals in specific time interval.

trials go to ∞ become Poisson.

$$(X \leq 4) \rightarrow \text{we 2 jobs } (1) \quad (1)$$

or 1 job (2)

$$(X \leq 3) \rightarrow X \leq 3 + 0.5 \dots (1)$$

1st
* Exponential distribution. (To model a time until an event)

- Time إلى غير
- time until next customer arrival
 - " " " failure.
 - " " " car crossing traffic light
- uses

* In Poisson, interested in # of arrivals

in t time unit: (4) إلى 0

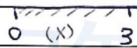
* Exponential: interst time until next arrivals.

متوقع بعد دقيقتين، فلو لم يحدث
في [0.3] إلى

* (λ) rate of arrivals

event per time unit

λ = 5, customer/hour



$X \sim$ time until an arrival from any point.

$$* P(X > x) = P(N=0) = e^{-\lambda x} (\lambda x)^0 = e^{-\lambda x}$$

$$\Rightarrow P(X < x) = 1 - e^{-\lambda x} \quad (\text{CDF}) \text{ area to the left.}$$

$$= (x < 0.5) = F(0.5), \quad x \geq 0$$

$$* f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \quad (\text{PDF})$$

$$* F(x) = 1 - e^{-\lambda x}$$

* represent time.

$$E(x) = \frac{1}{\lambda}, \quad V(x) = \frac{1}{\lambda^2}$$

$$SD = \frac{1}{\lambda}$$

$$* P(X > x) = \int_x^{\infty} \lambda e^{-\lambda t} dt = e^{-\lambda x}$$

$$* = 1 - \lambda e^{-\lambda x}$$

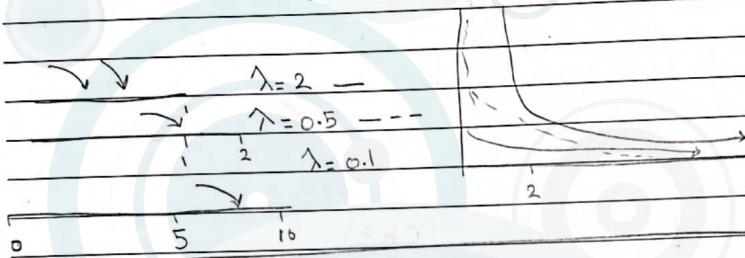
$$x \text{ etel: } \text{EXPONDist}(X, \lambda, \frac{1}{\lambda})$$

$$1 - F(x)$$

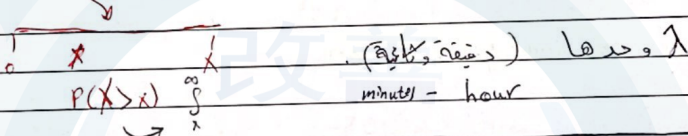
Five Apple

(time until next arrival) should be greater than t

$$\{N(t)=0\} = \{X > t\}$$



* انتبه لتعريف X وارتباطه بـ λ و μ



lack of memory

$P=0.082$

1st log

* الملاحظة التي يمكن استنتاجها



$P=0.082$

ex: X - the time until the arrival of taxi

* $\mu = E(x) = 10 \text{ min}$ $\lambda = \frac{1}{\mu} = \frac{1}{10} = 0.1 \text{ arrival/min}$

$$P(X > 60) = \int_{60}^{\infty} 0.1 e^{-0.1x} dx = e^{-6} = 0.0025$$

$$P(X < 10) = \int_0^{10} 0.1 e^{-0.1x} dx = 1 - e^{-1} = 0.6321$$

$$P(X > x) = 0.15 \quad \int_x^{\infty} 0.1 e^{-0.1x} dx = 0.15$$

$$0.15 = e^{-0.1x}$$

$$x = \frac{\ln 0.15}{-0.1} = 11.97$$

Five Apple

ex: $X \sim \exp(2)$

(a) $P(X \leq 0) = \int_0^0 \lambda e^{-\lambda x} = 0$
 time $\leftarrow$

can use CDF
 equation.

(b) $P(X \geq 2) = \int_2^{\infty} 2e^{-2x} = -e^{-2x} \Big|_2^{\infty} = e^{-4} = 0.0183$

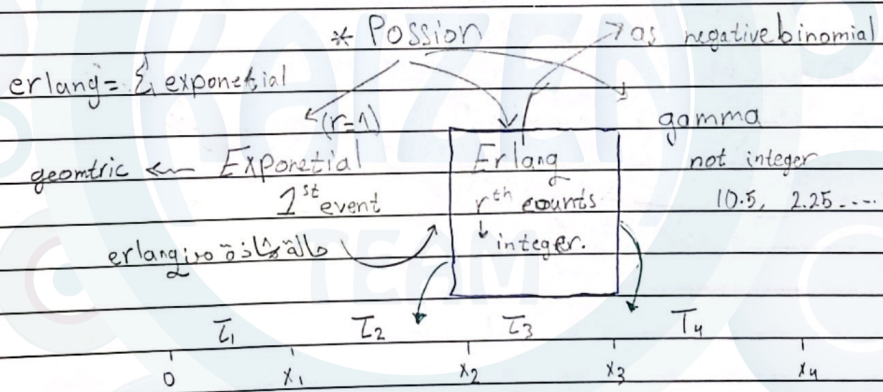
(c) $P(X \leq 1) = \int_0^1 2e^{-2x} = -e^{-2x} \Big|_0^1 = 1 - e^{-2} = 0.8647$

(d) $P(1 < X < 2) = \int_1^2 2e^{-2x} = -e^{-2x} \Big|_1^2 = e^{-2} - e^{-4} = 0.117$

(f) $P(X \leq x) = 0.05 \sim \int_0^x 2e^{-2u} = 0.05$

$0.05 = -e^{-2u} \Big|_0^x \sim 0.05 = 1 - e^{-2x}, \quad 0.95 = e^{-2x}$

$x = \ln 0.95 / -2$



$x_1 = T_1$

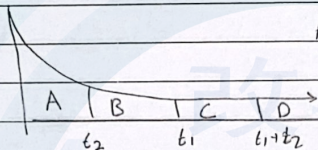
$x_2 = T_1 + T_2$

$x_3 = T_1 + T_2 + T_3$

$x_4 = T_1 + T_2 + T_3 + T_4$

continuous $\leftarrow$ rth

gamma > erlang > exp



$$\text{Areas} = A + B + C + D = 1$$

$$A = P(X < t_2)$$

$$A + B + C = P(X < t_1 + t_2)$$

$$C = P(X < t_1 + t_2 \cap X > t_1)$$

$$C + D = P(X > t_1)$$

$$* C / (C + D) = P(X < t_1 + t_2 | X > t_1)$$

$$* A = \frac{C}{(C + D)}$$

$r=1$, exponential

* Erlang distribution (r, λ) keyword: y^{th} arrives

• Erlang generalization from Exponential $\sim$ continuous (until r^{th} arrival) as

• Negative Binomial generalization from geometry discrete (first)

$$* \{X_r > t\} = \{N(t) \leq r-1\}$$

$X_r = \#$ of failure is time of the r^{th} failure

$$\textcircled{1} P(N(t) \leq r-1) = \sum_{n=0}^{r-1} \frac{\lambda^n t^n}{n!} e^{-\lambda t}$$

$$\textcircled{2} P(X_r \leq t) = 1 - P(X_r > t)$$

$$\textcircled{3} F(t) = P(X_r \leq t) = 1 - \sum_{n=0}^{r-1} \frac{\lambda^n t^n}{n!} e^{-\lambda t}$$

$$\textcircled{4} f(t) = \frac{d}{dt} F(t) = \frac{\lambda t^{r-1}}{(r-1)!} e^{-\lambda t}$$

$$* E[X] = \frac{r}{\lambda}, \quad V[X] = \frac{r}{\lambda^2}, \quad SD[X] = \sqrt{\frac{r}{\lambda^2}} \quad \left. \begin{array}{l} \text{gamma} \\ \text{erlang} \end{array} \right\}$$

$$P(X_r > x) = \sum_{k=0}^{r-1} \frac{\lambda^k x^k}{k!} e^{-\lambda x}$$

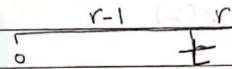
$$= 1 - F(x)$$

poisson

exponential

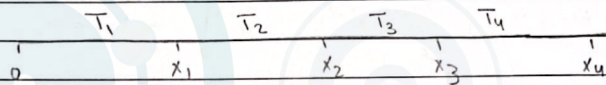
erlang

$x_r = \text{time of the } r^{\text{th}} \text{ failure}$



* $P(x_r \leq t) = 1 - P(x_r > t)$

$N(t) = \# \text{ of events in some interval } \underline{t}$



* $T_1 \sim \text{exponential} = x_1$

* $T_2 = x_2 - x_1$

$T_3 = x_3 - x_2$

$x_4 = T_1 + T_2 + T_3 + T_4$

* $P(x_4 > 40,000) = P(N(40,000) \leq 3)$

$$\sum_{n=0}^3 \frac{e^{-\lambda t} (\lambda t)^n}{n!} = 0.433$$

* $P(x_4 > 40,000) = 0.433$

elong 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

ex: Assume the number of arrivals to a fast food restaurant is governed by Poisson process with mean $\lambda = 3$ arrival per hour:

[1] What is the distribution and parameter for the random variable that represent the time until 4th customer arrives?

* time between events is exp with mean $\frac{1}{\lambda} = \frac{1}{3} = 20 \text{ min}$

* time for 4th: erlang ($r=4, \lambda=3$)

[2] What is the mean time until 4th arrives?

$$E[X] = \frac{r}{\lambda} = \frac{4}{3} \text{ hours} = 80 \text{ mins}$$

[3] What is the time between the 3rd and 4th customers?

* lack of memory: exp = mean $\frac{1}{\lambda}$

$$\frac{1}{3} = 20 \text{ min}$$

events $\rightarrow$ Poisson

[4] What is the probability the exactly 2 custome arrive in 1 hours?

$$* P(N(t)=2) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-3} 3^2}{2!} = 0.22404$$

[5] What is the probability that the 2nd arrives after 30 min?

$$P(t_2 > 0.5) = \int_{0.5}^{\infty} \lambda e^{-\lambda t} dt$$

excel = poisson(1, 3*0.5, T)

log normal distribution:

μ = mean for (w)

$$X = e^w$$

$$\ln X = w$$

$$\mu = 0$$

$$F(X) = e^{\frac{\mu + \sigma^2}{2}}$$

$$V(X) = e^{2\mu + \sigma^2} \cdot (e^{\sigma^2} - 1)$$

$$F(X) (X \leq x) = P(e^w \leq x) = P(w \leq \ln x)$$

$$= P(Z \leq \frac{\ln x - \mu}{\sigma})$$

* مربوط في (Z)

$$f(x) = \frac{1}{xw\sqrt{2\pi}} e^{-\frac{[\ln x - \mu]^2}{2w^2}}$$

الجدول

* Beta Distribution: $[0, 1]$

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha-1} \cdot (1-x)^{\beta-1} \quad 0 < x < 1$$

* $\alpha > 0, \beta > 0$

(0.5) * $\alpha = \beta$ symmetrical

* $\alpha = \beta = 1$ uniform

* $\alpha \neq \beta$ skewed

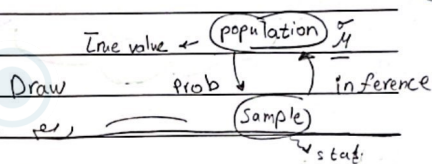
* $\alpha = \beta < 1$ sym - U shaped.

* $\alpha = \beta > 1$ sym - mound shaped.

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)}$$

CH6:- Discriptive Statistics. الإحصاء الوصفي *

Intro:



* Center measured = mean μ

* spread measured = variance σ

* CH6 ملخص

I. * Experiment $\rightarrow$ sample $\rightarrow$ events $\rightarrow$ counting techniques

$\rightarrow$ calculate probability $\rightarrow$ r.v $\rightarrow$ present data by dist $\leftarrow$ count
 Discrete

II. Descriptive statistics: Data (number) (CH6)

Summarize the data

Graphical $\rightarrow$ A

B. numerical - μ , σ , σ^2 , σ^2

III Data analyses:- I + II

* التوضيف (analysis)

process of drawing conclusions about problems based on the sample

* variable: characteristics population: length

* Data: observed values of variable. 1m 2m 3m

$\rightarrow$ ordinal 1-5

interval Nominal Ratio

1 $\rightarrow$ 2 $\rightarrow$ 3

no: 200

$\downarrow$ 0-11

* when variable = 0.0 it means that there is no of this at the variables.

step 1 ~ Minitab ~ help

ط قبل CH16 كنا نأخذ عينات لسلسلة الحساب
CH16 نحسب للبحث كالمسألة.

① ~ excel : Data analysis

: Histogram

histogram ~

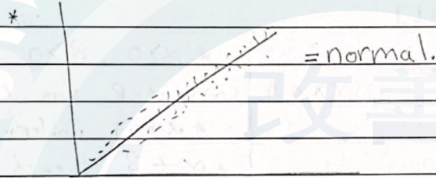


Bin

Data

* Summary Statistics.

ما نعلمه بالأرقام



Stat Analysis

Descriptive

~ Summarize data

Inferential

point estimation

Interval estimation. (PPC)

x (range)

Hop
testing

Compare
mean

Mean associated

* (6-1) Numerical Summaries of Data:

III * Sample mean $(\bar{X}) = \sum_{i=1}^n X_i$

Location.

left weight = weight right

* $\bar{X}$ is a reasonable estimate of μ .

$\mu, \sigma^2 \rightarrow \bar{X}, S^2$

[A] Module: (nominal data) (yes, no).

[B] Medical: (ordinal / interval).

[2] Variability: $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$

avg. deviation

[3] Range (r) = max - min.

[4] Coefficient of variation $CV = \frac{S}{\bar{X}}$ → standard deviation

[3.A] Order statistics: $x_1, x_2, x_3, \dots, x_n$

a-z

Sort: ascending = $X_{(1)} X_{(2)} X_{(3)} \dots X_{(n)}$ → largest

or z-a

smallest

descending.

* min = $X_{(1)}$

* r = $X_{(n)} - X_{(1)}$

* max = $X_{(n)}$

[3.B] Median: if n (odd)

$\bar{X}$

$\bar{X} = X_{(\frac{n+1}{2})}$

الوسيط الحسابي

2, 4, 5, 9, 10 n=5

$X_{(\frac{5+1}{2})} = X_{(3)} = 5$

[n=6]

* if n (even) $\bar{X} = \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$

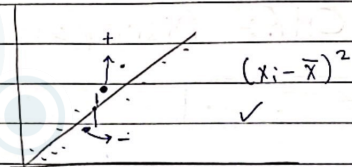
2, 4, 5, 9, 10, 12

$\frac{X_{(\frac{6}{2})} + X_{(\frac{6}{2}+1)}}{2} = \frac{X_{(3)} + X_{(4)}}{2}$

$= \frac{5+9}{2} = 7$

Five Apple

$$* S^2 = \frac{\sum X_i^2 - (\sum \bar{X})^2}{n-1}$$



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

درجات الحرية (n-1) → degrees of freedom.

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

* observing Can be freely determined.

$$ex = \frac{y_1 + y_2 + y_3 + y_4}{n=4} = 7$$

Free change.

ex:

X: 30,000, 38,000, 42,000, 57,000, 73,000

Y: 30,000, 38,000, 42,000, 57,000, 244,000

X: 48,000 $\bar{Y} = 82,200$

$\bar{X} = 42,000$ $\bar{Y} = 42,000$

data is close
to each other

data is skewed



mean is good
to describe.

Median is better
to describe.

* Single model bars = uni-model
2 Bi mode



* more model bars =

Multi



$Q_1 = \text{observation at position} = \left(\frac{n+1}{4} \right)$

$$Q_3 = \frac{3(n+1)}{4}$$

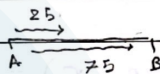
* $n_1 = 10 \therefore Q_1 = \frac{11}{4} = 2.75$ position of Q_1

$2^{\text{nd}} < Q_1 < 3^{\text{rd}}$ observation $\frac{1.75}{2}$ first $\frac{1}{3}$ بعد الترتيب.

$$* Q_1 = X_{(2)} + 0.75(X_{(3)} - X_{(2)})$$

$$* Q_1 = X_{(A)} + 0.25(X_{(B)} - X_{(A)})$$

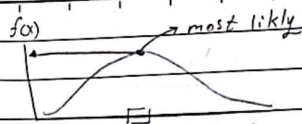
0.75



* $Q_2 = \text{Median}$

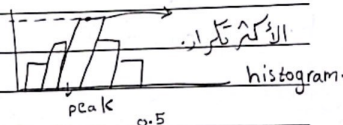
3.C Mode: most likely values.

* الأكثر تكراراً لها أكثر Prob. تحدث في التوزيع



5 Quartiles:

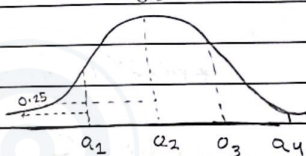
* Sort the data. ترتيب



Q_1 : 25% data point

First quartile $\rightarrow$ of data less

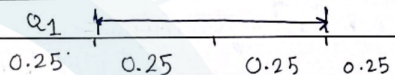
Q_2 : 50% of data less (Median)



Q_3 : 75% of data less

$Q_4 = X_{(n)} (\max)$

* IQR: $Q_3 - Q_1$



* $Q_1 = X_{(2)} + 0.75(X_{(3)} - X_{(2)})$

(ex) 6, 9, 12, 13, 13, 13, 14, 17, 20, 23

① ترتيب البيانات

$$Q_1 = 9 + 0.75(12 - 9) = 9 + 2.25 = 11.25$$

$$Q_3 = \frac{3(16+1)}{4} = 8.25$$



$$Q_3 = X_{(8)} + 0.25(X_{(9)} - X_{(8)}) = 17 + (20 - 17)0.25 = 17 + 0.75 = 17.75$$

* ليس شرطاً وجود $\bar{x}$ في sample.

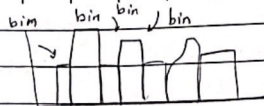
can $S=0$? *

yes when $\{3, 3, 3, 3, 3\}$



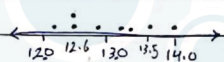
Graphical Summary:

(6-3) Histograms:-



I. Dot plot:

very small data:



* shape data
* center

II. Stem & leaf: (visually, the data)

Count stem leaf gap

smallest
largest

ex. 123, 154, 150, 161, 189, 165, 168, 191, 189, 190

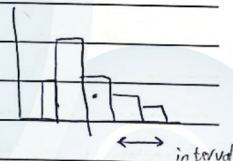
count		stem	leaf
1	upward	12	3
still 1		13, 14	—
3		15	0 4
(3)		16	1 5 8
still 3	backward	17	—
3		18	9
2		19	0 1

median

III. Histograms: 1- divide range into class / bins

2- # of bin = $\sqrt{n}$ [5-20]

3- equal width = $\frac{\text{range}}{\text{\# of bin}}$ (max - min)



$[b_0, b_1)$, $[b_1, b_2)$, $[b_2, b_3)$

bin	$60 \leq x < 80$	$80 \leq x < 100$	$100 \leq x < 120$
frequency:	2	3	6
relt freq:	2/11	3/11	6/11
cumulative:	2/11	5/11	11/11

$[b_0, b_1), [b_1, b_2), [b_2, b_3),$

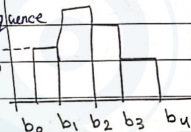
interval



$[b_{k-1}, b_k]$ C_j

data point data

frequency



$$b_1 = b_0 + \text{width}$$

$$b_0 \leq x_i < b_1 \Rightarrow C_j \text{ frequency of interval}$$

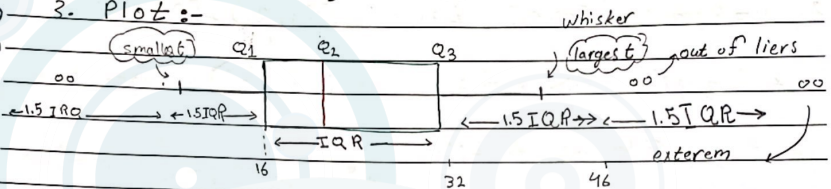
* relative frequency

$$h_j = \frac{C_j}{n}$$

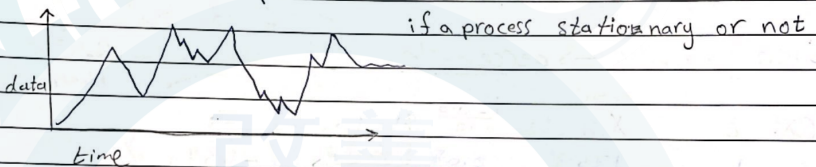
$$\text{rectangle height} = \frac{\text{bin frequency}}{\text{bin width}}$$

TV. Box Plot: (information Summary) outliers unusually observation

1. orders data:
2. Find Q_1, Q_2, Q_3
3. Plot:-



V. Time series plots.



VI. Probability plots: "normal" "check data: if it normal"

* Which distribution is a reasonable model for the data?

* p-p plot
* Q-Q plot
) Visual.

↓

$x_{(1)}, x_{(2)}, \dots, x_{(n)}$	$\frac{1}{10}, \frac{2}{10}, \dots, \frac{10}{10}$	$n=10$
$F(x_{(i)}) = \frac{i}{n}$	$x_{(1)}, x_{(2)}, \dots, x_{(n)}$	
$F_n(X \leq x_{(i)}) = \frac{i}{n}$	1. sort 2. $q_i = i - 0.5$ 3. $X_{q_i} = P(X \leq q_i) = \frac{i - 0.5}{n}$	* continuity adjustment
	$= \frac{i - 0.5}{n}$	$\frac{i}{n+1}$

4. plot q_i vs $x_{(i)}$ = z_i vs x_{achal}
 z_i vs z

Five Apple

* Box Plot

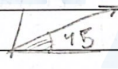


$$* q_i = \frac{i-0.5}{n}$$

$$X_{q_i} = F^{-1}(q_i)$$

* z-table

⊕ Probability plots:- to check that the data come from normal distribution or not

from drawing a graph 

مركز أطراف

* Q-Q plot (Quantile - Quantile)

* testing that the data normal or not.

$F(x) = P(X \leq x) :=$ cumulatively.

$$* F_i(x_{(i)}) = \frac{i}{n}$$

$x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$
1 2 3

(1) ordering
(2) Calculate co

* proportion of $X_i \leq$ to the i^{th} order.

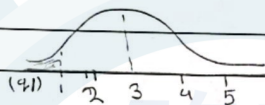
$$F_n(X \leq x_{(i)}) = \frac{i}{n}$$

continuous data or discrete data go to test

$$q_i = \frac{i - 0.5}{n}$$

$$\frac{i - 0.5}{n} \text{ (continuity)}$$

الملاحظة
التي



$$* F_n(x_{(i)}) = q_i$$

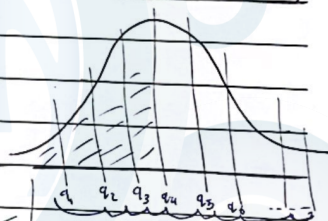
* القيمة التي تقابل الملاحظة في التوزيع (distribution)

$$* X_{q_i} = F^{-1}(q_i)$$

$$\Phi^{-1}(q_i)$$

$$\Phi(1.96) = 0.97 \quad \therefore 1.96 = \Phi^{-1}(0.97)$$

* $x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)}$



$$q_2 = \frac{2 - 0.5}{n} = 0.2 \quad \text{أول 20\%}$$

* Sort data

* Calculate cumulative freq: $q_i = \frac{i - 0.5}{n}$

$$* X_{q_i} \Rightarrow P(X \leq x_{q_i}) = \frac{i - 0.5}{n}$$

actual

(q_i, z) Areas

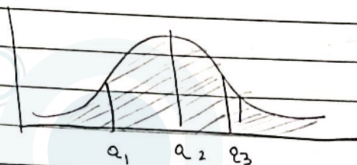
* Plot either (q_i) versus $x_{(i)}$ on the normal plots. (log scale)

or z_i versus $x_{(i)}$

* z_i versus x_{actual}
 z_i " z_{actual}
 q_i " x_{actual}

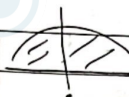
Five Apple

* quartiles.



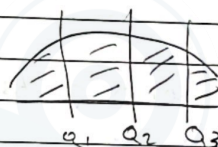
3 divids = 4 groups $(-\infty, a_1], [a_1, a_2), [a_2, a_3), [a_3, \infty)$

* 2 quartiles $\Rightarrow$ media



cutpoint.

3 quartiles $\Rightarrow$ terbles

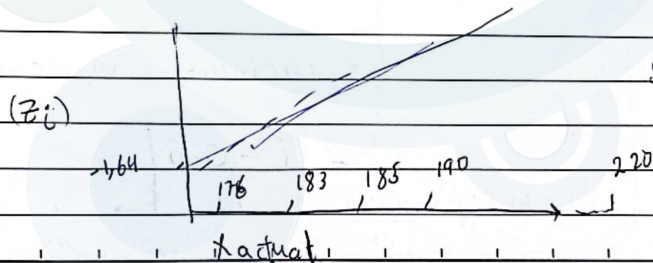


* 4 - quartiles $\Rightarrow$ qurtiles

a_1, a_2, a_3

* 100 quartiles $\Rightarrow$ percentile

EX: i	actual $X_{(i)}$	$i - \frac{0.5}{10}$ (4) $\frac{0.5}{10}$ $\frac{0.5}{10}$	Z_i (table).
1	176	0.05	-1.64
2	183	0.15	-1.04
3	185	0.25	-0.67
4	190	0.35	-0.39
5	191	0.45	-0.13
6	192	0.55	0.13
7	201	0.65	0.39
8	205	0.75	0.67
9	214	0.85	1.05
10	220	0.95	1.64



So this data is normal

✓
come from normal

Five Apple

توزيع العينات و تقديرات نقطية

CH 7: Sampling distributions and Point Estimation of parameter

* Observation random $X_1, X_2, \dots, X_n$

* Statistic: function of the observation

avg ↓

Sample

* Random Sample: X_i are independent.

every X use the same probability

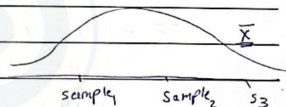
$\bar{X}, S^2$ random.

also have prob distribution

* μ, σ^2 :- parameters

Sampling

distribution



↓ good guess of true value

of population parameter is called: Point estimate. تقدير

* $\hat{\mu}$: estimate = $\bar{X}$

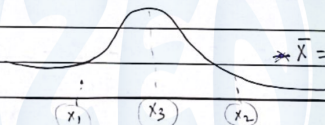
, $\hat{\sigma}^2 = S^2$ → good point estimate to population variance

* Sampling distribution of sample mean $\bar{X}$:

$X_1, X_2, \dots, X_n$ random from normal prob with μ, σ^2

* each observation

$X_i \sim N(\mu, \sigma^2)$



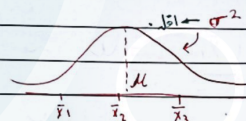
* $\bar{X} = X_1 + X_2 + X_3 + \dots + X_n$

n

avg for a sample chosen.

* $\mu_{\bar{X}} = \mu + \mu + \dots + \mu = n\mu = \mu$

* $\sigma_{\bar{X}}^2 = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$



$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ → original data. = std error

valid of any population prob distribution with n large

center limit Theorem.

parameter for sample distrib $\mu_{\bar{X}}, \sigma_{\bar{X}}^2, \sigma_{\bar{X}}$

استدلال كمي

data curves

* Sample

وقت

pop

σ, μ, σ^2

parameter

prob

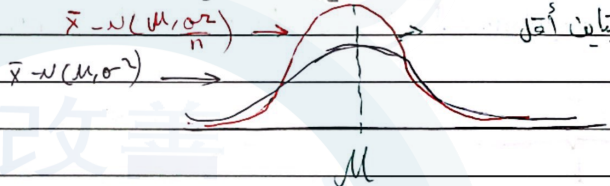
Inference

sample

تقدير

* افت لا تعرف القيمة الحقيقية (σ, σ^2, μ) لذلك تقدرهم

* البيان أقل



* Center limit Theorem: نظرية حدود المركز *

(C.I.T)

For $x_1, x_2, \dots, x_n$ random sample from a population with μ, σ^2 . * for large $n \Rightarrow n \geq 30$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (\text{table for } Z \text{ bar})$$

example: $\mu = 2.565 \text{ cm}$, $n = 9$

① $\sigma = 0.008 \text{ cm}$

$$P(2.56 < \bar{X} < 2.57)$$

$$= P\left(\frac{2.56 - 2.565}{0.008/\sqrt{9}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{2.57 - 2.565}{0.008/\sqrt{9}}\right)$$

$$= P(-1.875 < Z < 1.875)$$

$$= \Phi(1.875) - \Phi(-1.875)$$

$$1 - \{1 - 2\Phi(1.875)\}$$

↑ طريقة اخرى

② $\mu = 520 \text{ kN/m}^2$

$\sigma = 25 \text{ kN/m}^2$ ($\bar{X} > 525$)^o

$n = 6$

$$* P(\bar{X} > 525) = P\left(Z > \frac{525 - 520}{25/\sqrt{6}}\right) = P(Z > 0.4899)$$

$$= 1 - P(Z \leq 0.4899)$$



③ $\mu = 127$

$\sigma_x = 1.2$

$\sigma^2 = 36$

↳ std error of sample average.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \Rightarrow 1.2 = \frac{6}{\sqrt{n}}$$

$n = ?$

$n = 25$

* $n < 30$ for dist which is not evenly non-normal

* Approximate Sampling dist of a difference?

* Two population

$$Z = \bar{X}_1 - \bar{X}_2$$

نوع اول
نوع دوم

$$* Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

ex: $n_1 = 16$

$n_2 = 9$

$\mu_1 = 75$

$\mu_2 = 70$

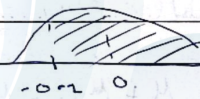
$\sigma_1 = 8$

$\sigma_2 = 1.2$

$$\textcircled{1} P(\bar{X}_1 - \bar{X}_2 > 4) = P\left(\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \geq \frac{4 - (75 - 70)}{\sqrt{\frac{8^2}{16} + \frac{1.2^2}{9}}}\right)$$

$$P(Z \geq -0.2236) = 1 - \Phi(-0.2236) = 0.5885$$

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$



* $\frac{50^2}{2(0.5)}$

$$\boxed{\bar{X}_1 - \bar{X}_2 = W} \sim N(75 - 70, \frac{8^2}{16} + \frac{1.2^2}{9})$$

نوع

* $W \sim N(5, 20)$

CH 8:- Statistical Intervals: الفاصل الإحصائي

Estimates a parameter: include error due to variability

→ $\hat{\mu} \rightarrow \bar{x}$, $(\hat{\mu} \neq \bar{x})$ # How closed $\hat{\mu}$ is to the true mean μ ?

estimate μ by provides an interval that quantifies our confidence.

I. Confidence intervals (C.I.): μ of normal dist

* confidence level

σ^2 known.

$$L \leq \mu \leq U$$

$$P\{L \leq \mu \leq U\} = 1 - \alpha$$

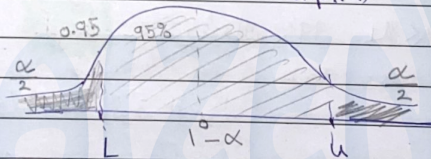
* L and U computed from the (r.v) sample

* $P(1-\alpha)$ of selecting sample

for which the C.I will contain the true mean μ .

* α number $0 \leq \alpha \leq 1$ (alpha)

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$



$$L = -Z_{\alpha/2}$$

$$U = 1 - Z_{\alpha/2}$$

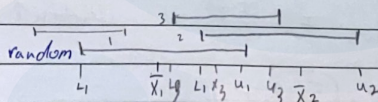
if $L = -1.96$, $U = 1.96$

$$\text{So } P\left\{-Z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right\} = 1 - \alpha$$

$$= P\left\{\bar{x} - \left(Z_{\alpha/2}\right) \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \left(Z_{\alpha/2}\right) \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

* $100(1-\alpha)\% = \text{C.I on } \mu$

C.I
 α
 α



infinite of r. sample, $100(1-\alpha)\%$ for μ from each sample.

* $100(1-\alpha)\%$ of these intervals will contain true value of μ .

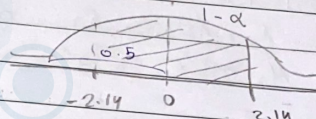
(exp: μ)

Five Apple

example: ① $\bar{x} - 2.14 \frac{\sigma}{\sqrt{n}} \leq M \leq \bar{x} + 2.14 \frac{\sigma}{\sqrt{n}}$

$\frac{z}{2} = 2.14$

$\phi(2.14) = P(Z \leq 2.14) = 0.9838$



$P(-2.14 \leq z \leq 2.14) = (0.9838 - 0.5) \times 2 = 0.9676$

② $z = 2.49$, $\phi(2.49) = 0.9936$

Count level = $2 \times (\phi(\frac{z}{2}) - 0.5)$

$\phi(\frac{z}{2})$

$= 2(0.9936 - 0.5) = 98.72\%$

$1 - \alpha = 98.72\%$

$\alpha = 1.28\%$

CH8 : $\delta \bar{x}$ of M go to table *

CH7 : $\bar{x}$ of M " " " "

* $1 - \alpha = \text{Confident level}$

* $\alpha = \text{Significance}$

$[4, 8]$

$\bar{x}_1 = 6$

4 6 8

$\leftarrow 2 \quad 2 \rightarrow$

$\leftarrow CI = \text{width} \rightarrow$
 $L \quad \bar{x} \quad u$

$(\frac{z}{2}) \frac{\sigma}{\sqrt{n}}$

* $u - L \Rightarrow \text{width} = 4$

half width (h) = $\frac{z}{2} \frac{\sigma}{\sqrt{n}}$

* $2h = w = u - L$

* when σ is known

* Confidence Level and Precision of Error:-

$$* E = |\bar{x} - \mu|$$

$$= |\bar{x} - \mu|$$

measure - Theo

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

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

* maximum value of

the error = "h"

$$(\bar{x} - h)$$

$$(\bar{x} + h)$$

$$* z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$Error = |\bar{x} - \mu|$$

$$L < \mu < U$$

maximum value to μ

* E أقل كلما كان α أكبر

* Sample size:- sample size that give you this Error. (n)

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

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

Two sides

* As $2E \downarrow = n \uparrow$ and σ fixed.

* As $\sigma \uparrow = n \uparrow$ and $2E, CI$ fixed.

* confidence level $\times 100 (1 - \alpha) \uparrow = n \uparrow$ fixed $\sigma, 2E$.

* قانون

$$\text{one side} := \mu \leq \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} = L \leq \mu$$

* large sample size: $n \geq 40$

replace σ by S

$$\bar{x} - z_{\alpha} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha} \frac{S}{\sqrt{n}}$$

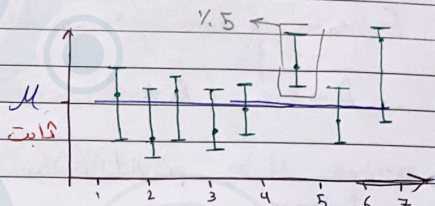
$$\sigma \rightarrow z$$

$$S \xrightarrow{n \geq 40} t$$

$$n = 15.37 \approx n = 16$$

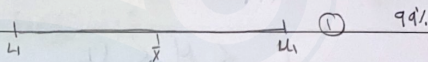
when $63.4 \leq \mu \leq 65$) 95% of samples will give you C.I. : μ بين هاتين القيمتين
 (will contain μ)

* $1 - \alpha = 95\%$



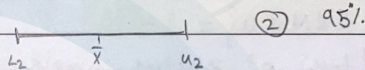
Interval number

• width ↑, Precision ↓. 8:30 - 9:30



• width ↓, Precision ↑. 9:00 - 9:15

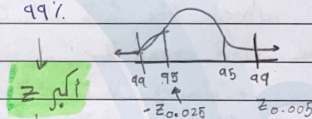
L, u Error ↓



* Confidence level = $1 - \alpha = 95\%$ → 99%

$L_1 \leq \mu \leq u_1$ 99% $z_{0.005}$

$L_2 \leq \mu \leq u_2$ 95% $z_{0.025}$



Confident ↑, Precision ↓
 دقة التقدير ↓

* Precision of estimate: to be more confident you should give your self more range ↑

with 95% CI $\Rightarrow \alpha = 0.05$ $Z = 1.96$

width = $2h = (2 \times 1.96 \frac{\sigma}{\sqrt{n}}) = \frac{3.92 \sigma}{\sqrt{n}}$

* count level ↑ = Precision ↓

99% $\alpha = 0.01$ $z_{0.005} = 2.58$

width ↑ = $\Rightarrow$ ↓

width = $2h (2 \times 2.58 \frac{\sigma}{\sqrt{n}}) = \frac{5.16 \sigma}{\sqrt{n}}$

غير معروف

ex: Circuit is $N(\mu, 20^2)$

$$\sigma = 20$$

① 95% CI for μ , $n=10$, $\bar{x}=1000$

↓

$$Z_{\frac{0.05}{2}} = Z_{0.025} = 1.96$$

$$\text{So } \alpha = 0.05$$

$$\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$1000 - 1.96 \frac{20}{\sqrt{10}} < \mu < 1000 + 1.96 \frac{20}{\sqrt{10}}$$

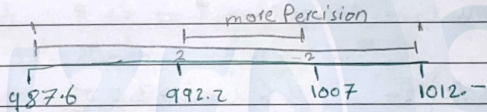
$$987.6 < \mu < 1012.4$$

② 95% $n=25$, $\bar{x}=1000$

error ↓

$n \uparrow$

$$* 992.2 < \mu < 1007.8$$



③ 99% $n=10$ $\bar{x}=1000$

$$1000 - 2.58 \frac{20}{\sqrt{10}} < \mu < 1000 + 2.58 \frac{20}{\sqrt{10}}$$

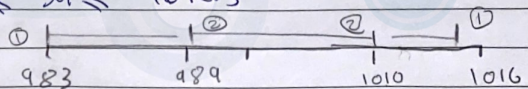
$Z_{0.005}$

$$983.7 < \mu < 1016$$

② 99% $n=25$, $\bar{x}=1000$

$$* n \uparrow \Rightarrow CI \downarrow \quad 989.7 < \mu < 1010.3$$

والتي



(8-7) $n?$ to achieve 95% CI of 40? $\sigma=20$

$x \text{ width} = 40$

$Z_{\alpha/2} = 1.96$

①

$$h = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \rightarrow \frac{1}{2}(40) = 1.96 \frac{(20)}{\sqrt{n}} = (n) = \left(\frac{39.2}{20}\right) = 3.8$$

$n = 4$

② $n?$ 99% CI of 40?

$Z_{0.005} = 2.58$

So $n = 6.66 \approx n = 7$

* random sample

* normal $\rightarrow$ * CI of μ of a normal distribution

σ^2 unknown.

large n $\rightarrow$ $n > 40$ أو غير

$t\text{-distribution} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$

* heavier *

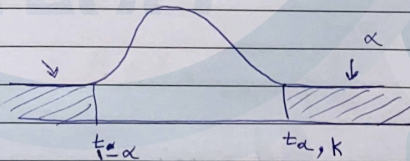
$K = \text{degrees of freedom}$

$K = n - 1$

* normal distribution samples n $\rightarrow$ $n > 40$ أو غير

$t\text{-dist} \rightarrow \text{normal } K = \infty$

$K=1$
 $K=10$



(Z 11, 12)

2 parameter

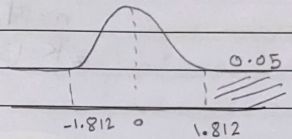
* t_{α} to right.

* $t_{\alpha, k}$ is upper-tail 100 α percentage going

ex: $n=11 \rightarrow k=10$
 $\alpha=0.05$

$P(T_{10} > t_{0.05, 10})$
 $P(T_{10} > 1.812)$

mirror image



* $t_{1-\alpha, k} = -t_{\alpha, k}$

$P\left\{-t_{\frac{\alpha}{2}, n-1} \leq T \leq t_{\frac{\alpha}{2}, n-1}\right\} = 1-\alpha$
 * C.I

$\bar{x} - \mu$
 $S/\sqrt{n}$

$P\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}\right) = 1-\alpha$

ex: Confident level = 95% , $df=14$;

* $t_{0.05, 14} = 1.761 \Rightarrow$ default twosided , $t_{0.025, 14} =$ from table.

* 99% , $df=19$: $t_{0.01, 19} = 2.539$

* 99.9% $df=24$: $t_{0.001, 24} = 3.467$

8-26 | SF mean = $\frac{S}{\sqrt{n}} \rightarrow \frac{6.11}{\sqrt{n}} = 1.58$ * $n=14.9 \approx 15$

variance = $(6.11)^2 = 37.3321$

C.I 99% in the db P $\alpha = 2 * 0.005$

$\mu \leq \bar{x} + t_{0.005, 14} \frac{S}{\sqrt{n}}$

$\uparrow \uparrow$
 1.761 1.761

$$* \mu = \frac{\sum}{n} \quad \text{and } \bar{x} \text{ is}$$

$$* SE = \frac{sd}{\sqrt{n}}$$

mean

* SE = standard error.

sd = standard deviation.

* N large : can use z & μ & σ

In excel: T = give you Area of T

T.INV.2T = give you value of t

$$\text{two side} = \frac{\alpha}{2}$$

$$\text{one side} = \alpha$$

8-31

Avg BP = 116.3

$\bar{x} = 116.3$

$s = 9.9$

$n = 14$

$* K = 13$

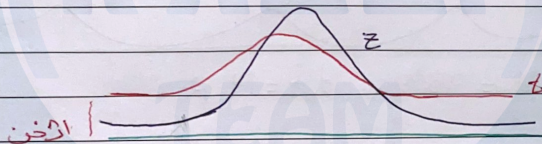
$$\mu \leq \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

$$\mu \leq 116.3 + t_{0.05, 13} \frac{9.9}{\sqrt{14}}$$

$$\mu \leq 116.3 + 1.771 \left(\frac{9.9}{\sqrt{14}} \right)$$

$$* \mu \leq 113.614$$

* كلما ما كبر sample كلما كان $\bar{x}$ قريب من μ



ex: 8-43 ① 95%, $df = 24$ one side upper.

* CI on σ^2 of normal distribution:

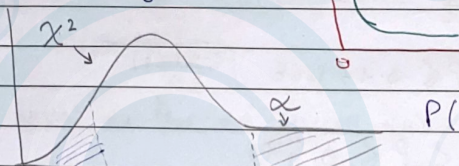
• Chi- χ^2 , with $(n-1)$ degrees of freedom

χ^2
↳ $k=1$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad (\text{all positive, no negative})$$

* not normal

So not symmetric
(not mirror image)



$$P(\chi^2_{1-\alpha/2, k} \leq \chi^2 < \chi^2_{\alpha/2, k}) = 1-\alpha$$

$$\chi^2_{1-\alpha, k} \quad \chi^2_{\alpha, k}$$

$$P(\chi^2_{1-\alpha, k} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha, k}) = 1-\alpha$$

$$* \frac{(n-1)S^2}{\chi^2_{\alpha/2, k}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2, k}}$$

نريد (σ^2) ثنائي الجانب

two sided

* one side

$$* \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha, n-1}} \quad \text{upper}$$

$$* \sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{\alpha, n-1}} \quad \text{lower}$$

* not symmetric so $\chi^2_{1-\alpha, k} \neq \chi^2_{\alpha, k}$

8-51 95% CI for σ , $n=15$, $S=0.00831$
upper?

$$\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha, k}}$$

$$\sigma^2 \leq \frac{14(0.00831)^2}{\chi^2_{0.05, 14}}, \quad \sigma^2 \leq \frac{14(0.00831)^2}{6.53}$$

$$\sigma^2 \leq 0.000148$$

$$\sigma \leq 0.0122$$

not normal
Five Apple

ex: 8-43 ① 95%, $df=24$ one side upper.

$$\chi^2_{1-\alpha, df} = \chi^2_{0.95, 24} = 18.31$$

② 99%, lower CI, $df=9$

$$\chi^2_{\alpha, df} = \chi^2_{0.01, 9} = 13.85$$

③ 90% CI and $df=25$

$$\chi^2_{\frac{\alpha}{2}, df} = \chi^2_{0.05, 25} = 37.65$$

$$\chi^2_{1-\frac{\alpha}{2}, df} = \chi^2_{0.95, 25} = 14.61$$

Five Apple

* population proportion C.I with large sample :-

$$X \sim \text{Binomial}(n, p)$$

$$* np \geq 5$$

$$* n(1-p) \geq 5$$

standardized normal

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$$

(normal distribution)

X: total number of success out of n observation

* $\hat{p} = \frac{X}{n}$, So $Z = \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$

hazrat qur'an

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha$$

$\hat{p}$ hat p

$$P\left(\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

(lower) (upper)

Sample Size :-

$$* n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot p(1-p)$$

margin of error

$$\text{or } n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p})$$

* prediction interval for a future observation

$$\bar{x} - t_{(\alpha/2, n-1)} \sqrt{1 + \frac{1}{n}} \leq \underline{\underline{x_{n+1}}} \leq \bar{x} + t_{(\alpha/2, n-1)} \sqrt{1 + \frac{1}{n}}$$

* worst case

* $p = 0.5$ * maximum

margin of error

* no estimate or estimate

* $\hat{p}$ على 131 * $\hat{p}$ على 131 *

* 8-57: 99% CI on proportion of under,

$$\textcircled{1} \quad \hat{p} = \frac{x}{n} = \frac{12}{30} = 0.4$$

$$n=30 \quad z_{\frac{\alpha}{2}} = 1.65$$

$$0.4 - 1.65 \sqrt{\frac{0.4(0.6)}{30}} < P < 0.4 + 1.65 \sqrt{\frac{0.4(0.6)}{30}}$$

$$0.252 < P < 0.548$$

$$\textcircled{2} \quad E=0.02, \alpha=0.1$$

$$z_{\frac{\alpha}{2}} = z_{0.05} = 1.65$$

$$\hat{p} = 0.4$$

$$n = \left(\frac{1.65}{0.02} \right)^2 \times (0.4)(0.6) = 1633.5$$

$$n \approx 1634$$

* lower of P

$$\text{one side: } \hat{p} - z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P$$

* upper of P

$$* P < \hat{p} + z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{one side:}$$