Statistics 1 notes

These notes were prepared to help students better understand the course. However, please note that they are not sufficient on their own. It is strongly recommended to practice the suggested questions provided by the instructor to fully grasp the material and prepare well for the exam.

There might be some mistakes in the notes. If you find any, feel free to contact me at the number below. I will review and update them if needed.

I also have the solution for the required Textbook question. If you're interested, feel free to reach out.

📞 My number: [0770693750]

هذه النوتس تم إعدادها لمساعدة الطلاب على فهم المادة بشكل أوضح، لكن يُرجى الانتباه إلى أنها غير كافية وحدها، ويُنصح بشدة بحل أسئلة "السجستد" الخاصة بالدكتور لتثبيت الفهم والاستعداد الجيد للامتحان.

قد تحتوي النوتس على بعض الأخطاء، فإذا لاحظت أي خطأ، يرجى التواصل معي على الرقم أدناه، وسأقوم بمراجعتها وتحديثها إذا لزم الأمر.

كما أن لدي حل الاسئلة المطلوبة من كتاب المادة لهذا الكورس، واللي بحاجة يحصل عليه يمكنه التواصل معي.

📞0770693750

Statistics I

*** statistics:** branch of mathmatics that deals with Colleetion, projeetion & Analyzing of data, to make decisions solve problem & design a product and process.

* types of data:-

① discrete (countable): numerical data representing frequencies or quantities (number of items)

⇒ nominal: data are categorized (=, ≠), Categories with no order
↳ وبتكون أشياء ما بقدر أحكم عليها انو أي وحدة أحسن
ex: colors, male & Female

⇒ ordinal: arranged in order (>, <)
↳ مثل التقييم من خمس نجوم بقدر أحكي إنو ال 5 نجوم أفضل من ثنتين

② continuous

عشان نميز بين ال discrete وال continuous
(اشي بقدر أعده) — بقدر أحكي مترونص مثلا إذا الطول continuous

⇒ Interval : Differenee between values is important
↳ بزبط أعمل عليها عمليات ال (+) وال (−)
ex: wt, tem (درجة الحرارة و الوزن)
↳ there is no real zero in this type of data

⇒ ratio : real zero & there is ratio between numbers (×, ÷)

*** population** : well defined set of objects = ALL

& it could be
- infinite : Hypothecal (افتراضي)
- finite

⇒ the disadvanteges :-
① cost ② time

⇒ Sample : part of the population (subset)

نفس أول حرف
population parameter : term that describes some characteristics of the population
↳ we always use Greek letters for parameter
ex: μ, α, β, ν, nu
for average

Sample statistics : a term that describes some Characteristics of the sample
↳ we use english letters for it:-
p, δ, x

* Data colleetion

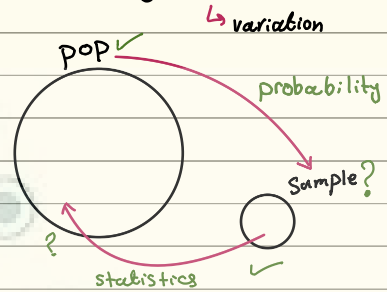
① Retrospective studies : ex: excel sheet ال data بتكون عندي أصلا
② observational (shadowing) : مراقبة بدون تدخل
③ Design of experiment
بعمل التجربة بطريقة و Design معين

---

* statistics
- Discriptive: colleetion / presentation
- inferential : make conclusion

* probability

possibility (chance) that something will occur (happen)
Likelyhood among choices
↳ variation



* before finding the probability we have to:-
① Do experiment (Emperical data)
② Evaluate possible value

* Random experiment

↳ means that each one of the options has the same chance to be selected

→ when we repeat the random experiments, we get different results

*** Sample space:** all possible outcomes

* Event : subset of the sample space
- Simple event : event with one outcome
- compound event : event with more than one outcome

* Relations from set theory

1. union: the union of two events A & B
$A \cup B$
all elements that are in A or in B or in both

2. Intersection
$A \underset{\text{and}}{\cap} B$: all the elements that are in both A & B

& when there is no common elements Between two events
↳ { } , $\varnothing$
لمعتوس فاضي ↳ أو خاي

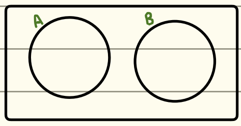
↳ & these events are called
Mutually Exclusive events : when there is no common elements between the elements

* Complement of an event ($\bar{A}$)

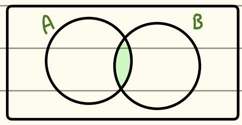
all elements in the sample space (S) but not in A

Examples:-

- $A \cap \bar{A} = \phi$
- $A \cup \bar{A} = S$ (sample space)



$A \cap B = \phi$



$A \cap B$ = the green area

* probability: A Quantity that quantifies the chance that an event (outcome) of Random experiment will occur

$$0 \le p(A) \le 1$$

احتمال ال event (A) بين الصفر والواحد دائمًا

$\Rightarrow p(S) = 1$

S → sample space

* Example:-

$S = \{HH, HT, TH, TT\}$

event A: 2 Heads

event B: one Head

- $p(A) = \frac{1}{4} = 0.25$
- $p(B) = \frac{2}{4} = 0.5$

* counting rules

1. Addition rule

prob of union

$$\Rightarrow \boxed{p(A \cup B) = p(A) + p(B) - p(A \cap B)}$$

Example:-

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

if the events are mutually exclusive (there is no intersection between events)

$$\Rightarrow p(A_1 \cup A_2 \cup A_K) = p(A_1) + p(A_2) + \dots + p(A_K)$$

2. product rule

Multiplication Rule:-

if an operation can be described as a sequence of K steps

Step1: n_1 ways
step2: n_2 ways
Step3: n_3 ways

$$\left.\begin{array}{l} \text{Step1: } n_1 \text{ ways} \\ \text{step2: } n_2 \text{ ways} \\ \text{Step3: } n_3 \text{ ways} \end{array}\right\} n_1 \times n_2 \times n_3$$

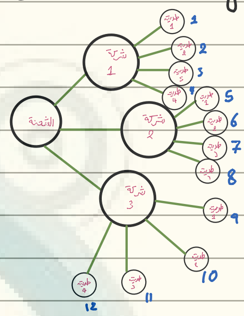
* إذا عندي شحنة بدي أنقلها عن طريق 3 شركات طيران والطرق اللي بتعرضها الطيارة هي 4 طرق

Step1: $n_1 = 3$

step2: $n_2 = 4$

$n = n_1 \times n_2 = 3 \times 4 = 12$

* we also can use tree diagram



هيك بصير عنا 12

3. permentations (we have order & it's important)

we use

$$P^n_K = \frac{n!}{(n-K)!}$$

4. Combinations (order in not important)

$$\frac{n!}{(n-K)!\,K!}$$

له هاي طريقة عشان أعرف كم sequence عندي
مو احتمالات لازم نحل على

$$p(A) = \frac{N(A)}{\text{total \# N ways}}$$

5. permentation of similar objects

ex: $5 \begin{cases} \to 3 \text{ male} \\ \to 2 \text{ Female} \end{cases}$

$$\frac{\overset{\to\ \text{total}}{5!}}{\underset{\downarrow\, n_1 \text{ ways}}{3!}\ \underset{\to\, n_2 \text{ ways}}{2!}}$$

⇒ this is way to count sequences not for probability

* Example: I have 3 types of surjury how many sequences I can have:-

3 Knee

4 hip

5 shoulder

total

$$\Rightarrow \frac{12!}{5!\,3!\,4!} = \frac{12\times 11\times 10\times 9\times 8\times 7\times 6\times \cancel{5!}}{\cancel{5!}\,3!\,4!} = 27720$$

type 1, type 2, type 3

⇒ what is the probability that all hips surjuries done at first?

لا بشيل هدول من حسبة ال total و ال types الي بالمقام

$$\frac{8!}{3!\,5!} = 56$$ ⇒ sequences that all the hips surjuries صارو أولاً

$$P = \frac{56}{27720}$$

* what is prob of $(\bar{A})$

$$S = \bar{A} \cup A$$

$$P(S) = P(\bar{A} \cup A)$$

$$1 = P(A) + P(\bar{A})$$

$$P(\bar{A}) = 1 - P(A)$$

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

* Conditional prob

H: event select high H

M: Male

Contingency table

	H	L	
M	28	35	63 → row هاد مجموع أول
F	42	31	73
	70	66	

$\Rightarrow P(H) = \frac{70}{136}$

$\Rightarrow P(M) = \frac{63}{136}$

$\Rightarrow P(\bar{M}) = 1 - \frac{63}{136} = \frac{73}{136}$

⇒ prob that a male student who scored high is selected?

$P(M \cap H) = \frac{28}{136}$

⇒ prob that the selected student scored high or male?

$P(H \cap M) = P(H) + P(M) + P(H \cap M)$

$= \frac{70}{136} + \frac{63}{136} - \frac{28}{136} = \frac{105}{136}$

⇒ probability that the selected student scored Low or female

$$P(\bar{H} \cup \bar{M}) = P(\bar{H}) + P(\bar{M}) - P(\bar{H} \cap \bar{M}) = \frac{108}{136}$$

⇒ prob that the selected male student scored high?

$\frac{28}{63}$ ⇒ هو حددلي إنو بدو من ال male فأنا حطيت شرط عشان أختار
فأحنا بنسمي هاد الاشي (conditional probability)

given

$$P(H/M) = \frac{P(H \cap M)}{P(M)} = \frac{\frac{28}{136}}{\frac{63}{136}} = \frac{28}{136} \times \frac{136}{63} = \frac{28}{63}$$

So for conditional prob

$$\boxed{P(A/B) = \frac{P(A \cap B)}{P(B)}}\ ,\ P(B) > 0$$

لأن إذا صارت صفر بتصير ال
prob ⇐ ∞ وال P لازم تكون
$0 \le P \le 1$

Example:

	H	L	
M	32	31	63
F	38	35	73
	70	66	136

$P(H) = \frac{70}{136} = 0.514$

$P(H/M) = \frac{32}{63} = 0.5074$

الاحتمالات كثير قريبة عبعض معناها حدوث ال event(M) قبل ال event(H)
ما أثر على احتمالية حدوث event(H)

We call these two events:-

* Independent event

Occurance of one event does not affect the prob of the occurrance of a second event

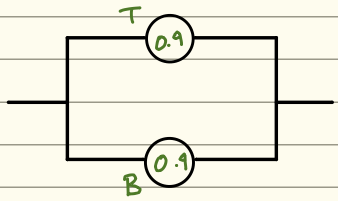
↳ using formula $P(A/B) = \frac{P(A \cap B)}{P(B)}$

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

لأن بطل يأثر ال B عال A

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

Example:- consider A & B independent



$\Rightarrow P(T \cup B) = P(T) + P(B) - P(T \cap B)$

$= 0.9 + 0.9 - (0.9 \times 0.9)$

$= 1.8 - 0.81 = 0.99$

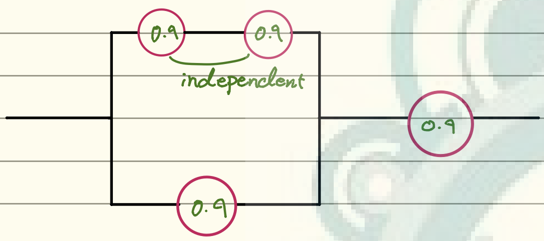
$\Rightarrow P(\bar{T} \cap \bar{B}) = P(\bar{T}) + P(\bar{B}) = 0.01$

or $1 - P(T \cup B)$

$P(T \cup B) \xrightarrow{\text{complement}} P(\bar{T} \cup \bar{B})$

demorgan Law

$P(\bar{T} \cup \bar{B}) = 1 - P(T \cup B)$



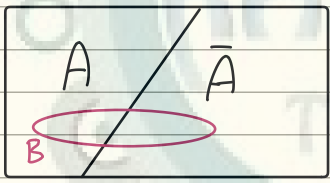
note: if we want to find prob of B

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

so

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

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



notes:-

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

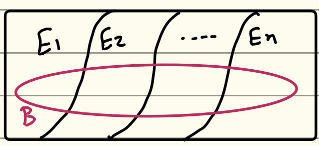
$P(B/\bar{A}) = \frac{P(B \cap \bar{A})}{P(\bar{A})}$

& $P(\bar{A}) = 1 - P(A)$

so that

$$\boxed{P(B) = P(B/A) \cdot P(A) + P(B/\bar{A}) \cdot P(\bar{A})}$$

↳ this is only for 3 events now we have to extend



$P(B) = P(B/E_1) \cdot P(E_1) + P(B/E_2) \cdot P(E_2) + \dots P(B/E_n) \cdot P(E_n)$

where $E_1, E_2, E_3 \dots E_n$ Mutually exclusive

note: complement of (A/B) is $(\bar{A}/B)$

random variable

suppose we toss a coin twice

$S = \{HH, HT, TH, TT\}$

let x represent # of heads in the Random experiment

$x = \{0, 1, 2\}$

here x is a random variable

so Random variable : It is a function that assigns a real number to each outcome in the sample space

↳ these are the only possible values of (x)

Range of x: the set of possible values of R.V.x (Random variable)

types of R.V:-

① discrete: (0, 1, 2) زي تباشري → R.V → finite (countable) / infinite (countable)

② continuous R.V → belongs into interval

↳ set of possible values consists of an intire interval

⇒ نطبق مثال toss a coin الي قبل

$x = \{0, 1, 2\}$

$P(x=0) = \frac{1}{4} = 0.25$

$P(x=1) = \frac{1}{2} = 0.5$

$P(x=2) = \frac{1}{4} = 0.25$

$P(S) = 1$ مجموعهم

prob mass distribution (p.m.f)

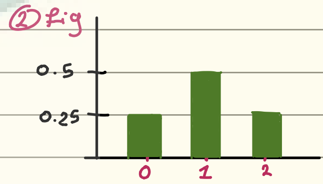
يعني إني أمسك ال (1) تبعت ال P(S) وأوزعها

على $x = (0, 1, 2)$

بربط نظامهم بـ

① table

② fig



مش كل ال Functions بربط أعلمها p.m.f في شروط:-

① $f(x_i) = P(x = x_i) \Rightarrow$ لما أعوض قيمة ال x لازم تعطيني احتمال ال (x)

② $f(x_i) \geq 0 \Rightarrow$ مايزبط تكون سالب

③ $\sum_{i=1}^{\infty} f(x_i) = 1 \Rightarrow$ مجموع احتمالات قيم ال x = 1

Example:-

$\Rightarrow f(x) = \frac{x-2}{2}$ for $x = 1, 2, 3, 4$

we cannot consider it as p.m.f because when we

$f(1) = -0.5 \rightarrow$ أقل من صفر

Example: $f(x) = \frac{x^2}{25}$, for $x = 0, 1, 2, 3, 4$

we can not consider it as p.m.f

$f(0) + f(1) + f(2) + f(3) + f(4) > 1$

لعضا بزبط

* Cummulative distributed function

CDF of a D.R.V (discrete random variable) with p.m.f

capital

$F(x) = p(x \leq X)$

x	$f(x)$	$F(x)$
0	0.04	0.04
1	0.32	0.36
2	0.64	1

التراكمي منجيب كل احتمال الي قبل + احتمال هاد الرقم

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

$F(0) = p(x \leq 0) = 0.04$

$F(1) = p(x \leq 1) = 0.36$

$F(2) = p(x \leq 2) = 1$

$F(3) = p(x \leq 3) = 1$

$F(100) = p(x \leq 100) = 1$

$F(-1) = p(x \leq -1) = 0$

so

$$F(x) = \begin{cases} 0 & , x < 0 \\ 0.04 & , 0 \leq x < 1 \\ 0.36 & , 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Example:

x	$f(x)$
1	0.2
2	0.5
4	0.1
6	a

* what is the value of (a)?

$p(S) = 1$ so $f(6) = 0.2$

* so $F(x)$

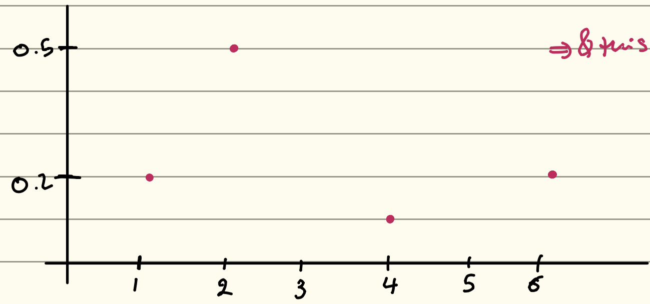
$F(1) = p(x \leq 1) = 0.2$

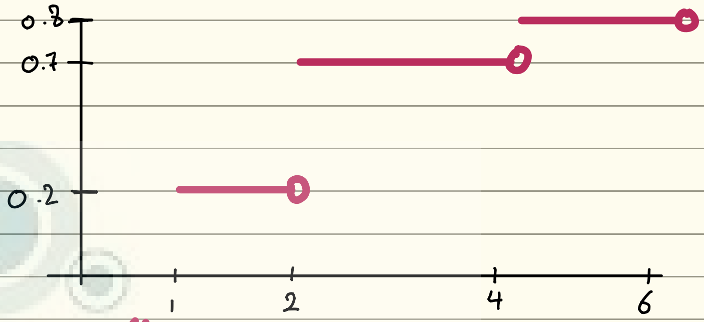
$F(2) = p(x \leq 2) = 0.7$

$F(4) = p(x \leq 4) = 0.8$

$F(6) = p(x \leq 6) = 1$

& we can draw it in the following fig





& this is F(x)

* mean & variance of D.R.V. x

mean = Expected value = E(x)

$E(x) = \mu$ → Greek letter (for population)

⇒ How to calculate E(x)?

$$\boxed{\mu = \frac{\sum x}{n}}$$

μ: Mean , $\sum x$: مجموع القيم , n: عدد القيم

or

$$\boxed{\mu = \sum_{x=i}^{n} x . f(x)}$$

x: the value of the R.V
f(x): prob for this value

& then we calculate the sum of them

Example:

value of x	prob	$E(x) = \mu$
0	0.04	$E(x) = x.f(x) = 0 \times 0.04 = 0$
1	0.32	$E(x) = x.f(x) = 1 \times 0.32 = 0.32$
2	0.64	$E(x) = x.f(x) = 2 \times 0.64 = 1.28$

لما بنجيب ال sum ⇐ the $E(x) = 0.32 + 1.28 = 1.6$

Example:-

x	-2	-1	0	1	2
$f(x)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

\# find $F(1)$, $F(2)$, $p(-1 \le x \le 1)$, $p(x \le -1 \cup x=2)$, $E(x)$

$\Rightarrow F(1) = p(x \le 1) = 1 - p(x > 1) = 1 - \frac{1}{8} = \frac{7}{8}$

$\Rightarrow F(2) = p(x \le 2) = 1$

$\Rightarrow p(-1 \le x \le 1) = \frac{2}{8} + \frac{2}{8} + \frac{2}{8} = \frac{6}{8}$

$\Rightarrow p(x \le -1 \cup x=2) = p(x \le -1) + p(x=2) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8}$

$\Rightarrow E(x) = (-2 \times \frac{1}{8}) + (-1 \times \frac{2}{8}) + (0 \times \frac{2}{8}) + (1 \times \frac{2}{8}) + (2 \times \frac{1}{8}) = 0$

هون طلعتمعي ال $E(x)$ ⇐ واحدة من قيم x (ومش شرط تطلع معي وحدة من القيم)

$$\boxed{V(x) = \sum x^2 f(x) - \mu^2}$$

* شو الفكرة من ال variance ؟ ممكن يكون عندي (2 sets of data) الهم نفس ال Mean بس هم مختلفين مش نفس ال data فلازم أ لاقي طريقة أميز بينهم موجودة من هاي الطرقة هي ال variance وتطبيقة عهاي الفكرة المثال التالي :-

Example:-

x	2	3	4
$f(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

set 1

x	1	2	3	4	5
$f(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

set 2

\# for set (1)

$E(x) = 2(\frac{1}{3}) + (3 \times \frac{1}{3}) + (4 \times \frac{1}{3})$

$E(x) = 3$

$V(x) = \sum x^2 f(x) - \mu^2$

$\{(4 \times \frac{1}{3}) + (9 \times \frac{1}{3}) + (16 \times \frac{1}{3})\} - 9$

$= \frac{29}{3} - 9 = \frac{2}{3}$

so $V(x) = 0.66$

\# for set 2

$E(x) = 3 \Rightarrow$ طلع الهم نفس ال Mean فراح أميز بينهم بال variance

$V(x) = \sum x^2 f(x) - \mu^2 = 2$

so

$\boxed{E(x)=3, \ V(x)=0.66}$ $\boxed{E(x)=3, \ V(x)=2}$

ال variability بال set2 أكبر من set1 مع إنو الهم نفس ال Mean

\# Rules for expected value & variance

$\Rightarrow$ for mean

1. $E(ax) = a E(x)$

   a: constant

Example:-

if $E(x) = 2$

so $E(3x) = 3E(x) = 3 \times 2 = 6$

2. $E(x+b) = E(x) + E(b)$

   the $E(b) = E(\text{constant}) = \text{constant}$

   if we have

   $+b$ (positive) ⇒ shift لليمين

   $-b$ (negative) ⇒ shift لليسار

3. $E(ax+b) = a E(x) + b$

$\Rightarrow$ for variance

$V(x) = \sigma^2$

\# note:- $\sigma^2 \to$ variance

$\sigma \to$ standard deviation

1. $V(ax) = a^2 V(x) = a^2 \sigma^2$
2. $V(x+b) = V(x) + V(b)$

$V(b) = 0$

so $V(x+b) = V(x) + 0$

$V(x+b) = V(x) = \sigma^2$

Chapter 3

\# Discrete uniform distribution

$$\boxed{f(x) = \frac{1}{b-a+1}} \quad \boxed{E(x) = \frac{b+a}{2}} \quad \boxed{V(x) = \frac{n^2-1}{12} = \frac{(b-a+1)^2-1}{12}}$$

b: أكبر value عندي

a: أصغر value

$$\boxed{V(x) = K^2 \left(\frac{n^2-1}{12}\right)}$$

K: هو الفرق بين القيم

هاد القانون لما يكون الفرق بين القيم هو 1

في حالة الفرق بين القيم (مثلاً 0.01) بنستعمل القانون ⇐

\# Bernoulli Distribution

Bernoulli R.V

⇒ bernoulli R.V that has only 2 possible values

⇒ bernoulli experiment: experiment that has only 2 possible outcomes

Example: flipping the coin 1 time

\# note: flipping the coin 2 times is not bernoulli experiment

Example: if we have 2 values of x

$\underset{\text{failure}}{x=0}$, $\underset{\text{success}}{x=1}$

$f(x) = p^x (1-p)^{1-x}$, $p(x=0) = 1-p$

* Binomial distribution

n is fixed bernoulli exp
x here is # of successes

$$p(X=x) = C^n_x \, p^x \, (1-p)^{n-x} \quad , x=0\ldots n$$

Example:-
$n=6$, $p=0.1$, x= # of male students
so if $x=0 \rightarrow$ this means that we have six Female students
if $x=6 \Rightarrow$ this means that we have 6 male students
find $p(x=0)$, $p(x=6)$, $p(x=1)$

$$p(x=0) = C^6_0 \, (0.1)^0 \, (0.9)^6 = \frac{6!}{(6-0)!\,0!} \, (1)\,(0.9)^6 = \frac{6!}{6!} \, (0.9)^6 = 1\times(0.9)^6$$

$$p(x=6) = C^{6!}_{6!} \, (0.1)^6 \, (0.9)^0$$

$$= \frac{6!}{(6-6)!\,6!} \, (0.1)^6 = \frac{\cancel{6!}}{0!\,\cancel{6!}} \, (0.1)^6 = (0.1)^6$$

$$p(x=1) = C^{6!}_{1!} \, (0.1)^1 \, (0.9)^5$$

$$= \frac{6!}{5!\;1!} \, (0.1)^1 (0.9)^5 = 6\,(0.1)(0.9)^5$$

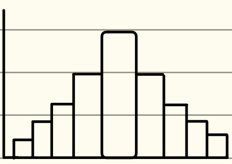
$$= 0.354294$$

* note: $0! = 1$

* note:-
for bernoulli dist

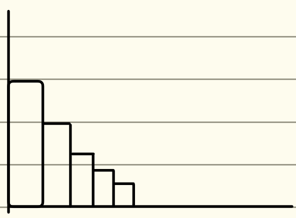
$\boxed{E(x) = np}$ $\boxed{V(x) = n.p.(1-p)}$

* note:-
if $p=0.5$



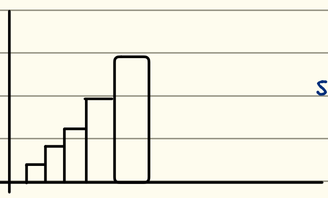
Semitrical

if $p=0.1$



Skewed to the right

if $p=0.9$



Skewed to the left

* Geometric Distribution

n is not fixed

* لو انا واقفة عخط انتاج وسحبت عشر قطع ورحت فحصتهم هاي Binomiall
↳ 2 possible outcomes (working/not working)
↳ وبنفس الوقت في عدد معين من القطع الي راح أفحصهم

بس لو ضليت واقفة عالخط أستنى عين ما أطلع قطعة ما بتشتغل و هون راح يكون
الـ success فعليًا هو إني ألاقي قطعة ما بتشتغل

& here
x: # of trials until the first success

$\boxed{f(x) = (1-p)^{x-1}\,(p)}$ $\boxed{E(x) = \frac{1}{p}}$ $\boxed{V(x) = \frac{1-p}{p^2}}$

* note:-
بالـ Binomiall كنا بنبلش من الصفر لأنو هون الـ n الي عندي ممكن ما يطلعوش ولا
6 success بس بالـ (Geometric) مستحيل يطلع معي success من أول سحبة

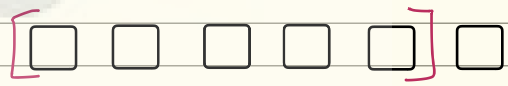
* negative binomial distribution

n is not fixed (until the n^{th} success)

example: It's impossible to get three boys from just the first two draws, because I need three & I have only drawn twice

$$\boxed{f(x) = p(X=x) = C^{x-r}_{r-1} \, (1-p)^{x-r} \, p^r \, , \; x \geq R}$$
($\mapsto$ number of n is not fixed)

example:
$r=3$, $p(x=6) \Rightarrow$ هاي الجملة معناها بمثالنا إنو
الشب الثالث طلع بالسحبة السادسة



among 5 we have 2 males
& I do not care which of them is male
↳ binomial $\Leftarrow$ 5 هون فعليًا عنا

note: $r-1 \Rightarrow$ في مثالنا هو عدد الشباب قبل ما الاقي الشب الثالث
↳ الي هو بمثالنا 2

* Mean & variance

$\boxed{E(x) = \frac{r}{p}}$ $\boxed{V(x) = \frac{r(1-p)}{p^2}}$

Example:

I called a phone 5 times. The person was busy during 2 of those attempts and did not answer the remaining three times

so $p(S_x)=0.4$

↳ prob of success of x

$p(S_y)=0.6$

↳ prob of success of y

$p(x=2)=C_2^5\,(0.4)^2(0.6)^3$

$p(y=3)=C_3^5\,(0.6)^3(0.4)^2$

they are equal ↳ because $\frac{5!}{3!\,2!}$ are the same

وهاد الحكي إحنا حكيناه بال birnoulli dist

now using the Same Question but:-

يعني إذا ضليت أرن لحد ما يردوا علي ⇐ #n is not fixed

↳ وأنا بدي أضل أحاول لحد ما يطلعوا

poisson distribution

of events in an interval

# of defects in a product

# of holes in a Km

defect: المشكلة الي بالشي

defective: الاشي الي فيه العيب نفسه

$$\boxed{p(X=x)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}\ ,x\geq 0}$$

t: unit in time unit (min/sec)

λ

↳ Avg of events in unit of time, $x\geq 0$

where x: # of events in an interval → volume, area, length, time

$\lambda \Rightarrow$ (calls, accidents) in an (hour, day, month, week)

example

$E(x)=\lambda t$ $\qquad V(x)=\lambda t$

example: $\lambda=3$ calls/hr

How to test that I can use poisson to solve the question?

بحسب ال $E(x), V(x)$ إذا متساويين أو متقاربين بعض بزبط أحل عبر poisson

Example: # of telephone calls coming into a switch board average = 4 calls/min. Find the prob, that no calls arrive in a given 1 min period?

$$p(x=0)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}=\frac{(4\times1)^0 e^{-4}}{0!}=e^{-4}=\frac{1}{e^4}$$

2) Find the prob that no calls will arrive in a 5 min period

$$p(x=0)=\frac{(4\times5)^0 e^{-20}}{0!}=e^{-20}=\frac{1}{e^{20}}$$

3) Find the prob that at least 2 calls will delivered

$p(x\geq2)=1-p(x<2)$

$=1-[\,p(x=0)+p(x=1)\,]$

$p(x=0)$ we already calculated it

$$p(x=1)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}=\left(\frac{(4\times1)^1}{1!}\right)e^{-4}=\frac{4}{e^4}$$

4) Find the prob that at most 2 calls will arrive in 30 seconds

$p(x\leq2)=\underbrace{p(x=0)}_{}+p(x=1)+p(x=2)$

we can not use the previous question because we changed t

$$p(x=2)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

& here $t=\frac{1}{2}$

Hypergeometric

pop → N

N → K: # of success ; N−K

sample → n

n → x: # of successes ; n−x: # of failure

ex: N=15, K=5

N−K = 15−5 = 10

if we have group of six people. Some of them are males & some are femles

$$p(x=4)=\frac{C_4^5\cdot C_2^{10}}{C_6^{15}}$$

this means that we have 4 males & 2 Females

so

$$\boxed{p(X=x)=\frac{C_x^K\cdot C_{n-x}^{N-K}}{C_n^N}}$$

$\boxed{E(x)=n.p}$ $\qquad \boxed{V(x)=n.p(1-p)\left(\frac{N-n}{N-1}\right)}$ $\qquad \boxed{p=\frac{K}{N}}$

Finite pop correction factor

{Chapter 4}

Continuous prob Distributions

* if I have 50 batteries

0-1 16

0-2 11

العدد الساعات الي ممكن تستخدم البطارية.

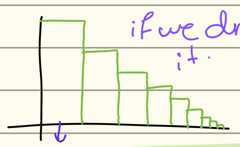
عدد البطاريات

0-1	16	32%
1-2	11	22%
2-3	9	18%
3-4	6	12%
4-5	3	6%
5-6	2	4%
6-7	1	2%
7-8	1	2%
8-9	1	2%

لا تكتبي هيك بالامتحان 0.02 هيك لازم تكتبي.

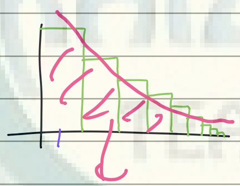
prob = Relative Drag

if we draw it.



we call every single one of them pin & the prob of this = the Area of the rectangulare.

& I can use them & make more & more smaller sections of them & by doing this approaching to zero, بيصير الها نقطة.



$$\boxed{\text{prob} = \text{Area under curve}}$$

$$P = \int f(x).dx$$

↳ prob density function

Area under the Curve.

$f(x)$ is p.d.f

For a continuous R.V. x, a p.d.f is a function such that

1) $f(x) \geq 0, \forall x$ → for all (x) values. ⇒ because we have Area under curve & this can not be zero

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

3) $P(a \leq x \leq b) = \int_a^b f(x).dx$

$$P(x=a) = \int_a^a f(x).d(x) = \text{Zero}$$

because a is a point & the integration, from-to the same point = zero

also the area under a point = Zero

المساواه هون ما بتفرق

$$P(a < x < b) = P(a \leq x \leq b) = P(a \leq x < b)$$

in this topic these are the same but because under a & under b these are points & the integration of them = zero so there is no difference.

⇒ $F(x) = P(x \leq b) = \int_{-\infty}^{b} f(x)\,d(x)$

مش شرط وجودها ما بأثر (اليساوي)

⇒ remember!! for D.R.V (Discrete)

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

$V(x) = \sum x^2 f(x) - \mu^2$

but for continuous

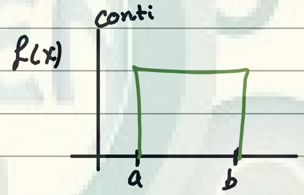
تذكري إنو integration مش sum

$$\mu = E(x) = \int x f(x).dx$$

$$V(x) = \int x^2 f(x)\,dx - \mu^2$$

For continuous

conti



the total area under the curve = $\boxed{P=1}$

$$f(x) = \frac{1}{b-a}$$

$$E(x) = \frac{a+b}{2}$$

$$V(x) = \frac{(b-a)^2}{12}$$

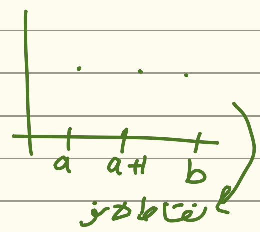
For uniform Distribution

$$f(x) = \frac{1}{n} = \frac{1}{b-a+1}$$

$$E(x) = \frac{a+b}{2}$$ ⇒ the mean in this was the center

$$V(x) = \frac{(b-a+1)^2 - 1}{12} = \frac{n^2-1}{12}$$

Graph



4.5 Normal Dist

$\mu = 170$
يعني الوسط
Mean

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

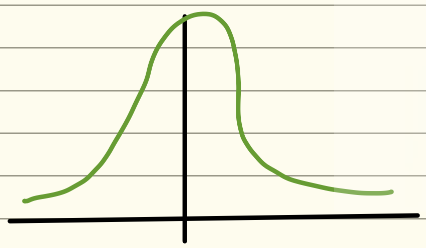
For the Green one variation قليل

for the purple one variation كثير

↳ the mean is the same for both of them.

So if I want to find prob & It's the area under the curve so I must try to find way to find $f(x)$

Ⓕ Normal Dist



normal Dist always semitrical about the mean.

if I want to find prob we will going to integrate.

$\mu = 170cm$

$\sigma = 10cm$

$\mu \pm 1\sigma$

↳ this means that if I do

$(\mu+1\sigma)$ or $(\mu-1\sigma)$ the area of them will be the same.

↳ that means that

68% ⟶ 160 – 180

$\mu \pm 2\sigma$

$(\mu+2\sigma)$ $(\mu-2\sigma)$

150 – 190

95%

$\mu \pm 3\sigma$

$(170+30)$ $(170-30)$

140 – 200

99.73%

ثابت

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

↳ It's a function of $(\sigma / x / \mu)$

$-\infty < x < \infty$

⟹ & this is can not be solve using exact methods

① I have one special type for normal Dist

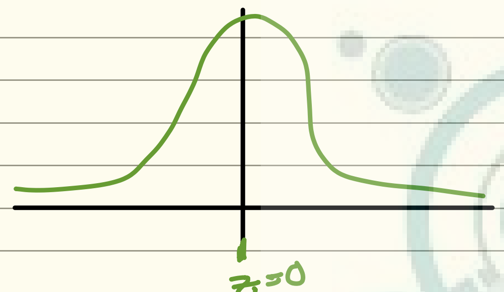
where $\mu = 0$

$\sigma^2 = 1$

& we call this standard normal Dist

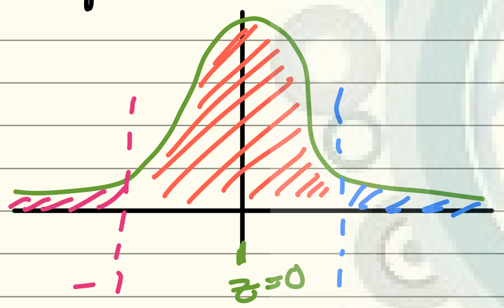
$X \sim N(\mu, \sigma^2)$

this is standard normal Dist $\Leftarrow Z \sim N(\mu = 0, \sigma^2 = 1)$



$p(z<0) = 50\%$

$p(z>0) = 50\%$



$p(z<-1)$ Drawn

$p(z>1)$ Drawn

$\left.\begin{matrix} p(z<1) \\ (z>-1) \end{matrix}\right\}$ these are the same.

$p(-1<z<1) =$ Drawn

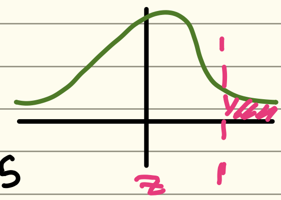
with all of this we use z so we are using standard normal Dist

* using the table for standard normal Distribution (موجود في الصفحة الي بعدها)

① $p(z<3.49) = 0.499767$

from the table calculate following probabilities

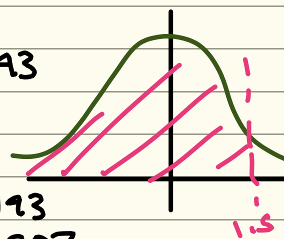
② $p(z>1)$



$1 - 0.841345 = 0.158655$

① $p(z<1) = 0.841345$

③ $p(z<1.5) = 0.933193$

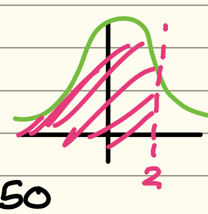


④ $p(z>1.5) = 1 - 0.933193$
$= 0.066807$

⑥ $p(z<0.5) = 0.691462$

⑦ $p(z>0.5) = 1 - 0.691462$
$= 0.308538$

⑧ $p(z<2) = 0.977250$



⑨ $p(z>2) = 1 - 0.977250$

$= 0.22750$

→ for normal distribution not specificly for (standard ND)

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

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

$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

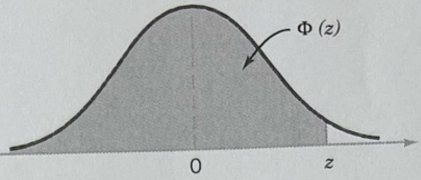


Table II Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.99995
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.99996

* if I have values for $z < 0$ so we will treat it using semitry

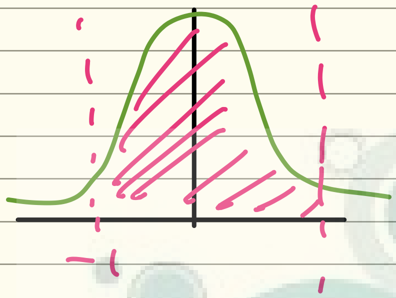
$p(z < -1)$

↳ this is the same area for $p(z > 1)$

$p(-1 < z < 1)$

$p(z < 1) - p(z < -1)$

$1 - p(z > 1) - \ldots$



to get the probability for 2 standard deviation.

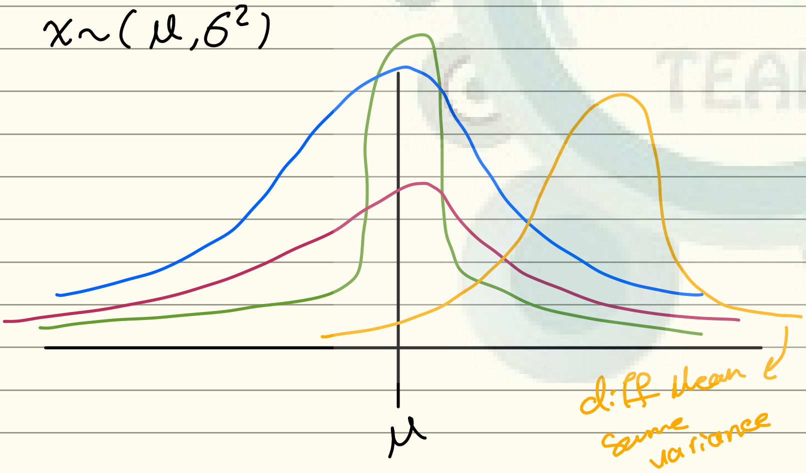
$p(-2 < z < 2) \approx 0.95$

$p(z < 2) - p(z < -2)$

↳ $p(z > 2) = p(z < -2)$

Normal Dist & standard normal Dis

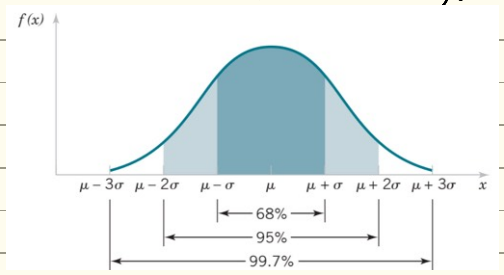
$x \sim (\mu, \sigma^2)$



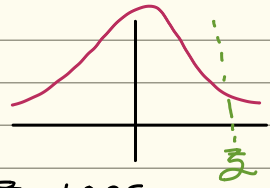
* all of these have the same mean but diff values of variance for each one of them

↑ variance أكبر

Standard Random variable $x \sim N\ (\mu = 0, \sigma^2 = 1)$

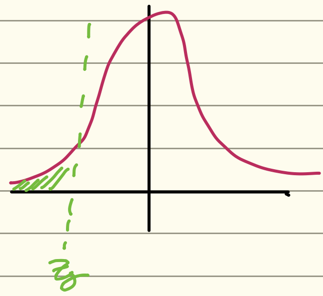


$\Rightarrow p(z < z) = 0.9$



what is the value of $z = 1.285$

$\Rightarrow p(z < z) = 0.1$



so that means that ⇐

$+z > = -z <$

$1 - p(z > z)$

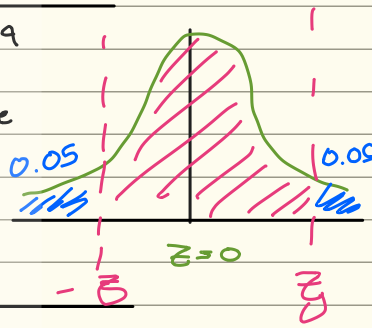
$1 - 0.9 = 0.1$

-1.285

$p(-z < z < z) = 0.9$

so that means that value of $p(z) = 0.95$

$z = 1.645$



$p(z_2 < z < z_1) = 0.9$

مجهولين بمعادلة واحدة (لا يوجد حل) ⇐ ∞ حلول

$x \longrightarrow z \xRightarrow{table} table$

كيف بدنا نخلي $z \Leftarrow x$

$z \sim N(\mu = 0, \sigma^2 = 1)$

$x - N(\mu = 0, \sigma^2 = 1)$

Imp :-

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

$\mu = 170\ cm,\ x = 180$

$\sigma = 5\ cm$ هاد الفرق

value of z

$+2 \rightarrow$ هاد هو كم standard diviation عن الـ mean

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

$= \dfrac{180 - 170}{5} \pm 10\ cm \Rightarrow$ معناها إنه ابتعد عن الـ mean

$x_1 = 180$ if I have $= -5$ (shift) الـ قبل ⇐ 10cm بعد

$p(x > 180)$

$p(z > 2)$

حوسنا

Z- How many standard diviation are you away from the mean

4.5.5 solution

$\mu = 129$

$\sigma = 14$

a) $p(Z>2) = 1 - p(Z<2)$

$= 1 - 0.977250$
$= 0.02275$

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

$z = \frac{100-129}{14} = -2.071428$

$p(z < -2.071428) =$

$1 - 0.980774$

$= 0.019226$

c) $p(Z<z) = 0.95$

$z = 1.645$

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

$1.645 = \frac{x-129}{14}$ $\quad x = 152.03$

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

$z = \frac{199-129}{14} = 5$

السؤال هو كم حجم العمليات يلي عددها

أصلًا الجدول تاعنا بس بوصل لحد 3.99 إذا ال z هون راح نعتبرها (4)

إذا الاحتمالية أو عدد هاي العمليات الي راح توقف فيها وقت ⇐ 0 = الكيترقلية/

Poisson Dist:-

Random variable

λ : # of Events /interval

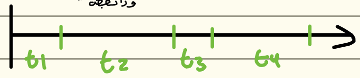
$p(X=x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, x \geq 0$

↗ for continuous

Exponential Dist-

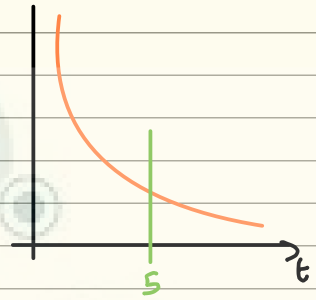
T: time between consecutive events

حدثين متتاليين ورا بعض



$p(t<T) = \int_0^T f(x)$

تمثل المحققات لكل وحدة زمن .



$p(t<T) = F(T)$

$= 1 - e^{-\lambda t}$

we cannot say for this one $p(x=5)$ & we already mention this that area under this point = 0

$p(t>5)$ & this is the complement of $p(t<5)$

so $p(t>5)$
$1 - (1 - e^{-\lambda t})$
$= e^{-\lambda t}$

Ex:-

$\lambda = 3$ cust/hour

find the probability that the next customer arrive with in 10 min

بدي أحول ال min لـ hours

$p(t \leq 10) = 1 - e^{-\lambda t}$

لوحطيت (=) معي بفرق

$= 1 - e^{-3 \cdot \frac{10}{60}}$

$= 1 - e^{-0.5} = 0.3934644$

تذكري $P(X=5)$ هون $= 0$

Ex:

$p(t > 10min) = 1 - (1 - e^{-\lambda t})$

$e^{-\lambda t} = e^{-3 \cdot \frac{10}{60}} = e^{-0.5} = 0.6065$

if I want to discribe it as poisson dist

$p(X=0)$ for $p(t>10min)$

$p(x \geq 1) \Rightarrow$ for $(t < 10min)$

↳ this is $(1 - p(x=0))$

in the next 20 min $\Rightarrow p(t<20)$

in the next 20 min given no body came in the first 10 min

<u>that means</u>

$$p(t>20 \mid t>10) = \frac{p(10<t<20)}{p(t>10)}$$

$$= \frac{p(t<20) - p(t<10)}{p(t>10)}$$

$$p(t<20) = 1 - e^{-\lambda t} = 1 - e^{-3 * \frac{20}{3\,60}}$$

$$= 1 - e^{-1}$$

$$p(t<10) = 1 - e^{-\lambda t} = 1 - e^{-3 * \frac{10}{60}}$$

$$1 - e^{-0.5}$$

$$p(t>10) = e^{-0.5}$$

$$p(t>20 \mid t>10) = 1 - e^{-0.5}$$

↳ & this means Lack of meomory dist that or meomoryless Dis

& that means that

$$p(t>20 \mid t>10) = p(t<10)$$

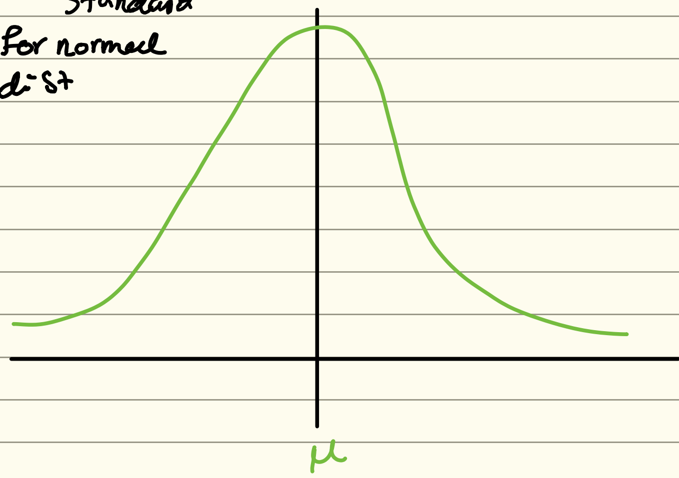
& that means that this dist make an upddate & did not effects with the first (10)

{Chapter 6}

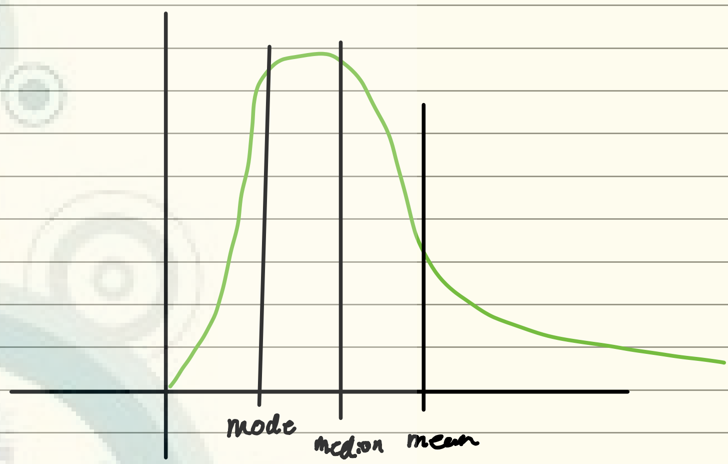
① central tendency | & this discribe the pop

- → mean: $\bar{x}$, μ
- → median: the middle value
- → mode: the most value repeated
  - ↳ & this we can not apply it on continuous data be cause we have to count.

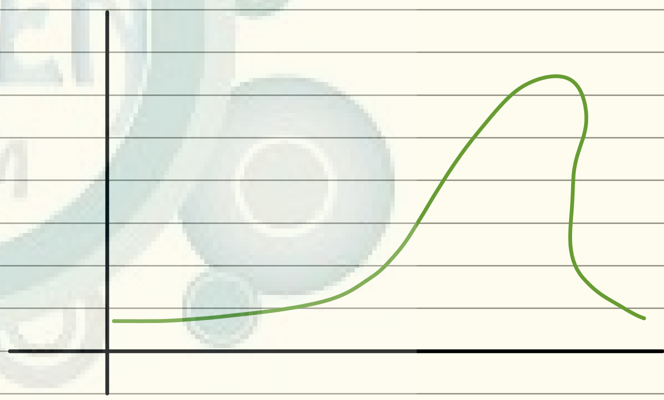
Standard for normal dist



mean = median = mode



Skewed to the right.



Skewed to the left

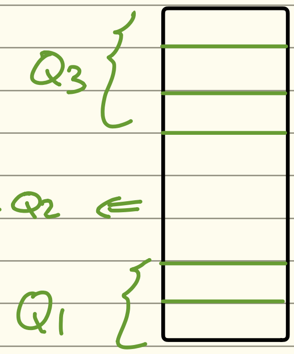
* median is less sensitive to the extreme values.

* so if I calculated mean, median & mode & they are very قريبين for each so this is normal Dist

if I want to find if this Dis non-semetrical (skewed) using the median

\# To find the median position

$\left(\frac{n+1}{2}\right)^{th} \Rightarrow$ position



& this is the median $\Leftarrow Q_2 \Leftarrow$

Q_1 : القيمة اللي ربعهم أقل منها
ثلاث ارباعهم أعلى منها

Q_3: القيمة اللي ثلاث ارباعهم أقل منها وربعهم أعلى منها .

↳ & we can all the previous that we discuss

Measure of central tendency

① mean ② median ③ mode

we consider Q_2 as median

$Q_1 = \frac{n \times 25}{100} = K.q$ K: constant

↓ integer part use K+1

for $n=15 \Rightarrow Q_1 = 3.75$ طلع عندي كسر فبجيب اللي use(K+1)

SO $Q_1 = 4$

$Q_3 = \frac{15 \times 75}{100} = 11.25$ طلع معي كسر بزيدلو 1

$\boxed{Q_3 = 12}$

إذا ما طلع كسر ما بجمع شي بوخذ زي ما هو

Variation

① Range = $x_{max} - x_{min}$

$x_1 \sim B(10, 0.1)$ $x = 0, \dots, 10$ var أكثر تأثيراً

$x_2 \sim B(5, 0.1)$ $x = 0, 1, \dots, 5$

② IQR

inter Quartile Range

$Q_3 - Q_1$

③ variance

$V(x) = \sum x^2 f(x) - \mu^2$

$V(x) = \int x^2 f(x) - \mu^2$

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

& this equals zero

so this equation does not standarize the variation

but for pop $V(x) = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} \Rightarrow$ & we call this average deviation

variance for the sample $S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}$ → Sample mean

④ standard diviation:-

$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$

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

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

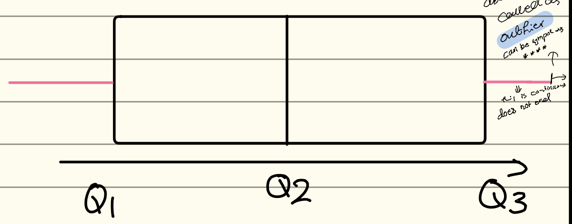
ركيبي الرقم بعدين في المجموع الو

Box-plots

$Q_1 = \frac{n \times 25}{100}$

$Q_2 = \frac{n \times 50}{100}$

$Q_3 = \frac{n \times 75}{100}$

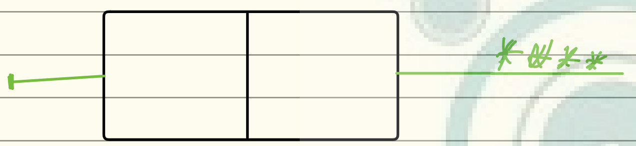


IQR $= Q_3 - Q_1$

$Q_3 + 1.5(IQR) <$ outlier

$Q_1 - 1.5(IQR) >$ outlier

we can have :-



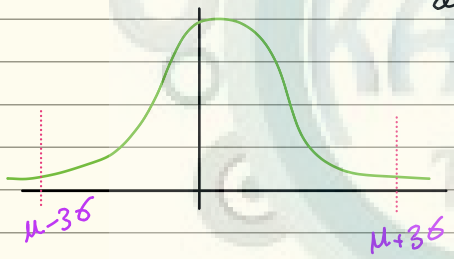
$Q_1 - 1.5\,IQR$

we have other methods to calculate outliers.

$\mu \pm 3\sigma \Rightarrow$ & this is 99.37 from the whole data

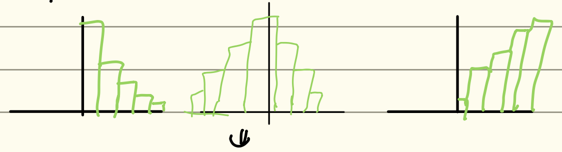
$\mu + 3\sigma$

$\mu - 3\sigma$



the Range can be effect by the outliers but the IQR does not defect by outliers

Chapter 4	chapter 3
cont dist	Discrete
	Binomial $x \sim B(n,p)$



to apply the previos on Binomial Dist it has to be semitry

Use normal dist to approximate Binomial

$\mu = n.p \geq 5$

$V(x) = n(1-p) \geq 5$

it has to be for both of them

if this is not applied the histogram will be skewed to the right or to the left

$z = \frac{x - \mu}{\sigma} \Rightarrow$ & For Binomial

$x \to B$

$\mu = n.p$

$\sigma^2 = n.p(1-p)$

we want σ

$\sigma = \sqrt{n.p(1-p)}$

by adding

$$z = \frac{\bar{x} - (n.p)}{\sqrt{\frac{np(1-p)}{n^2}}}$$

x: number of successes (n) is

p: prob for pop.

applying this.

p Asstimated value

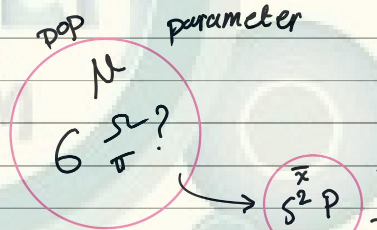
$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

standard normal dist

$\hat{p} = \frac{x}{n}$

{chapter 7}

point Estimation :-



parameter

parameters : N قيمة ثابتة تعبر عن pop

sample statistic

$\bar{x}, s^2, P$ → proportion

average & variance

Example of parameter : central tendency

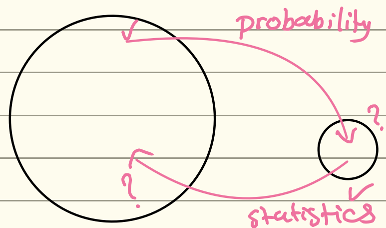
variability

proportion (نسبة كم مثلاً)

usually the pop parameters are unknown so we take a sample statistics to asstimate pop parameters

مثلاً بأخذ عينة منشان اعرف عن الكل.

pop → we need to calculate parameter → we take a sample → sample statistics → to asstimate pop parameter



if I have the sample

π proportion p=0.6

so π=0.6

& this is point single

& these all are variable

$\bar{X}_1 = 175$ cm $\bar{X}_2 = 165$, $\bar{X}_3 = 185$ cm

$\mu = 175$ cm

capital $\bar{X}$ & the random variable change by the prop distribution.

& this is also R.V

small $\bar{x}_1$

prob distribution ⇒ توزيعها

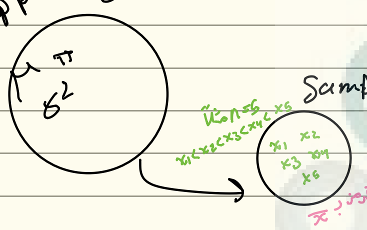
now we need to known the dist of $\bar{X}$

Distribution of $\bar{X}$

$E(\bar{X}) = ?$

$V(\bar{X}) = ??$

pop parameter θ → point Estimation



Sample statistics $\hat{\theta}$ → Estimated

capital Θ

$\bar{X}$ ⇒ sample mean ($\bar{x}$ ⇐ نرمز)

S^2 ⇒ sample variance (σ^2 ←)

$P \Rightarrow \hat{P}$ → Estimated value يعني جاي من sample

لأنو الـ value اللي هاد random variable أو يوصف

variable

to Estimate μ

1) $\hat{\theta}_1$ = sample mean (for n=5) $\bar{x} = \frac{x_1 + x_2 + \dots + x_5}{5}$

2) $\hat{\theta}_2 = \frac{x_1 + x_5}{2}$

3) $\hat{\theta}_3 = x_3$

4) $\hat{\theta}_4 = 1.5x_1$

5) $\hat{\theta}_5 = 0.6x_5$

& these are different point estimators

point Estimation

$\hat{\theta}$ is R.V

Step 1: Find expected value

$E(\hat{\theta}) = E(\bar{x}) = E\left(\frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}\right) =$

Expected value of constant is constant

$= \frac{1}{5} E(x_1 + x_2 + x_3 + x_4 + x_5)$

independent

$= \frac{1}{5}\left[E(x_1) + E(x_2) + E(x_3) + E(x_4) + E(x_5)\right]$

what is the $E(x_1)$?

$E(x_1) = \mu$

$E(\hat{\theta}_1) = E(\bar{X}) = \mu$

$E(\hat{\theta}_2) = \mu$

$E(\hat{\theta}_3) = \mu$

$E(\hat{\theta}_4) = 1.5\mu$ constant ثابت ⇒ over estimation هاد بعطي

$E(\hat{\theta}_5) = 0.67x_5 = 0.67\mu$ → under estimation هاد بعطي

احنا بدنا اللي يعطينا ال μ زي ماهي

baised Estimator: $\hat{\theta}_4$ +positive, $\hat{\theta}_5$ negative

unbiased: $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ (0)

biased $= E(\hat{\theta}_1) = \mu - \mu = 0$

$= 1.5\mu - \mu = 0.5$

biased can be (positive, zero, negative)

now the biased is removed & we will use unbiased value & ⇒ (بنقارن الـ 3 ببعض)

بدي أختار الأكثر بتقرب من sample لـ sample العام بتكون قريبة من بعض.

$V(\hat{\theta}_1) = \frac{\sigma^2}{5}$

$V(\hat{\theta}_2)$ $V\left(\frac{x_1 + x_2}{2}\right) = a^2 V(x_1 + x_5) = \frac{1}{4}(V(x_1) + V(x_5)) = \frac{1}{4} 2\sigma^2 = \frac{\sigma^2}{2}$

$V(\hat{\theta}_3) = Vx_3 = \sigma^2$

$V(\hat{\theta}_1)$ أنو أقل واحد

so when we want to estimate μ $\bar{x}$ the best to be used

$\bar{x}$ is min. var unbiased Estimator

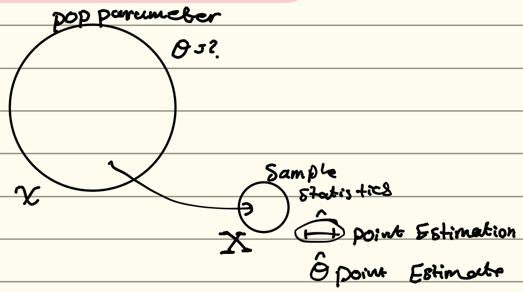
Point Estimate النتيجة اي اوجدناها

$\sigma^2 \Rightarrow S^2$

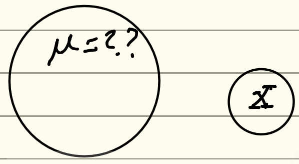
وعشان هيك بنوثق بال μ أكثر من Range

P Estimator الفاصل

Point Estimation



$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$



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

$x = 168$

Poisson :-

$\lambda = 3$ cust/hour

$p(x < 150)$ chapter 4	$p(\bar{x} < 110)$
$p(x < 120)$ ↳ individual element	$\bar{X}$ follows which Distribution ?

& we call this

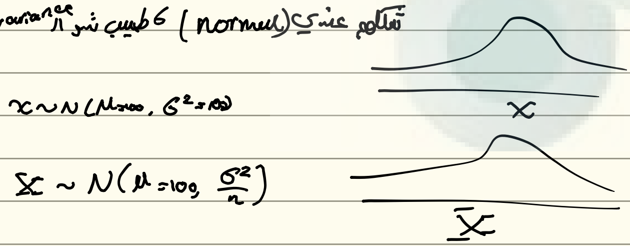
sampling distr of the sample mean

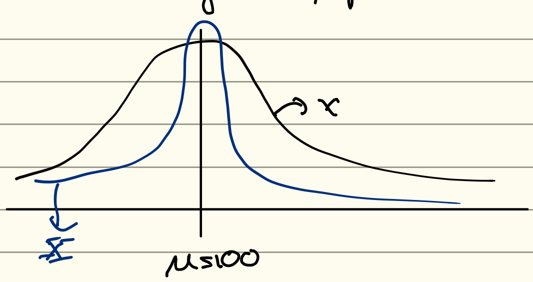
$x \sim N(\mu = 100, S^2 = 100)$

if I want to take 100 Sample of this

شو ال Dist تبع ال $\bar{X}$ يطلع.

تطلع عندي (normal) طبيب شو ال Mean/variance





x & $\bar{X}$ has the same Mean but the variation for $\bar{X}$ is less than it

What if already x do not follow normal Dist

شو ما كانت ال x تبعت ال pop الأصلية وشو ما كانت ال dist تاعها.

دائمًا $\bar{x}$ بيكون ال normal distribution

$\bar{X}$: always follows normal Dist

① $x \sim N(\mu, \sigma^2) \Rightarrow \bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ ↳ تبع ال sample

② $x \sim$ any distribution $\Rightarrow \bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

↳ & this is approved by

Central Limit theorem

if we take random sample (with size n) from a population with Mean $= \mu$ & variance $= \sigma^2$

$x_1, x_2, x_3, \dots, x_n \Rightarrow$ size of n

then the sampling dist of the sample mean ($\bar{X}$), will be normal, with mean $= \mu$ & variance $= (\frac{\sigma^2}{n})$

$$X \Rightarrow Z = \frac{X - \mu}{\sigma}$$

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

Center limit theorem

C.L.T :-

$x_1 - x_2$ from the pop (μ, σ^2) size n

then the sample dist of the sample with the Mean (μ) & variance of (σ^2) then the $\bar{X}$ is normal distributed

$\bar{X}$: normal distribution بتكون if $n \geq 30$

if $n < 30$

للشوفه إذا لأي درجة بعيدة

كمية ال sample عن كمية الناس

الأصلية (N) depends on original pop

$\sigma_{\bar{x}}$ = standard error of $\bar{X}$

$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

S.E

$X \longrightarrow Z = \frac{x-\mu}{\sigma}$

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

S.E

How confident are u in ur Estimation?

169 → this is point & its good but we want to say how much I Estimate?

there is two types of answer → single point / range (L, u)

& that's the topic of chapter 8 confedence interval

chapter 8

$p(L<\mu<u) = 1-\alpha$, where α: probability of error

($1-\alpha$: confidence level)

if I confidence 0.95 or 95%

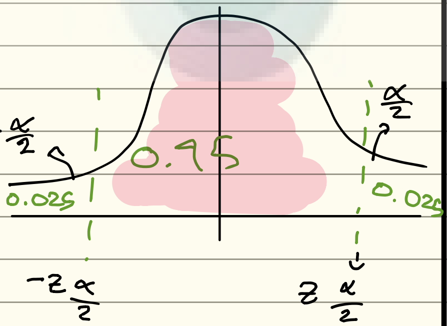
so we are returning to prob

& the best dist to find probabilities for it is standard normal distribution

ل ؟ Area بي

95% تحت الـ

cannot be zero it is an area



$p(L<\mu<u)$

$p(-1.96<z<1.96) = 0.95$

$p(-z\frac{\alpha}{2} < z < z\frac{\alpha}{2}) = 1-\alpha$

$p\left(-z\frac{\alpha}{2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z\frac{\alpha}{2}\right) = 1-\alpha$

بدنا نضرب بهاي أقل قيمة 6 هي مرح ممكن تكون بس مش سالب

$p(-z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}} < \bar{X}-\mu < z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$p(\bar{x}-z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

confidence interval

$-z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}$ — $\bar{X}$ — $z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}$

so that $\bar{x} \pm z\frac{\alpha}{2}\frac{\sigma}{\sqrt{n}}$

#	α	n	σ	C.I (confidence interval)	z
1	0.05	9	5	96.7 , 103.3	1.96
2	0.1	9	5	97.25 102.75	1.645
3	0.01	9	5	95,7 104,3	2.58
4	0.05	16	5	97.55 102.45	
5	0.05	25	5	98.04 101.96	
6	0.05	9	2.5	~~98~~ 98.36 101.63	
7	0.05	9	10	93.46 106.53	

1) $100 \pm 1.96 \frac{5}{\sqrt{9}}$

$100 \pm 1.96 * \frac{5}{3}$

upper = 103.266

Lower = 96.7333

2) Lower = 97.25

upper = 102.75

3) Lower =

upper =

$p(96.7<\mu<103.3) = 0.95$

$\times$ prob

prob لا ما بنحكي انه احتمال يقع μ هي الـ
لا ولكن بتحكي ممكن يحتوي ال ٩٥٪ .

prob that this interval has the mean of the pop inside it 95% true

هادا الي رجع الدكتور شرحو

Confidence interval for pop mean μ

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) =$$

when σ^2 is unknown we use:-

if $n \geq 40$

$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$

if $n < 40$, we use

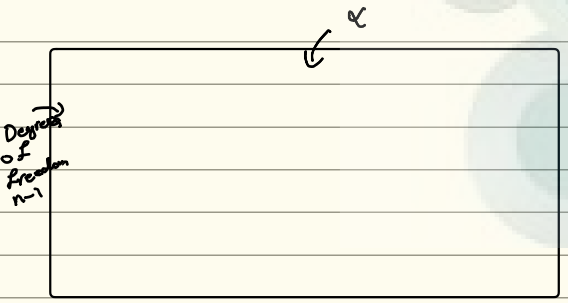
$$\left(\bar{x} - t_{\alpha/2} \quad < \mu < \bar{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}\right)$$

C.I (confidence interval) for pop variance

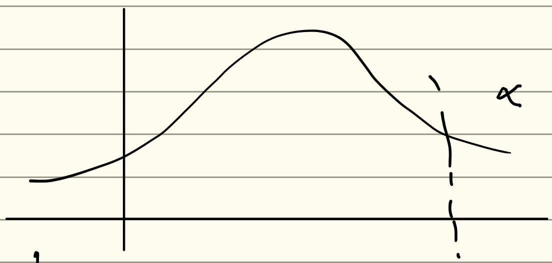
$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-}}\right) = 1-\alpha$$

we have new distribution of t

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



area to the wright ليعطيني المساحة الـ



chi² value D.O.f

Pop proportion

$$P\left(\hat{p} - z_{\frac{\alpha}{2}} \quad < \pi < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

& this over all

Two sided C.I

for one sided C.I

for one for μ

upper يا Lower بوضح الـ

$$\left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}} \quad < \mu\right)$$

لما يكون الـ 6 Known

لو الـ 6 مش معروف

$$\left(\mu < \bar{x} + z_{\alpha}\frac{S}{\sqrt{n}}\right) \quad n > 40$$

$$(\mu < \bar{x}$$

$$\left(\frac{(n-1)S^2}{\chi_{\alpha, n-1}} \quad < \sigma^2\right)$$

$$\left(\pi < \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

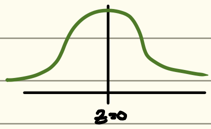
$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < \pi\right)$$

t-distribution

حكينا عن ال z قبل هيك

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

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



$Z \sim N(\mu=0, \sigma^2=1)$

standard normal distribution

$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$

so this distribution will be change between sample & another because it depends on the sample

$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$

كل ما صارت ال sample أكبر كل ما ال variability تأثرت.

α in the table : is the area under curve (prob)

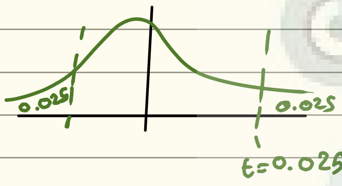
t is different from Z that gives us cumilate to the right

if I want to do confidence

$\le \mu \le \bar{x} + t_{\alpha/2, n-1} \times \frac{S}{\sqrt{n}}$

$S=3$
$n=9$
$\bar{X}=10$

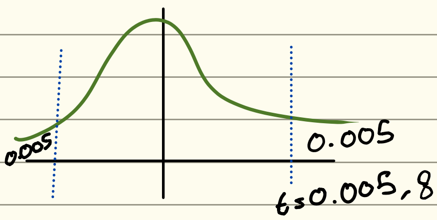
if I want 0.95 confidence interval



$10-2.3 \le \mu \le 12.3$

$(7.7 < \mu \le 12.3) \Rightarrow$ this is 0.95 confindence interval

calculate 99%.
90%.



$\bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$

$10 + 5.408 \frac{3}{\sqrt{9}}$

$10 + 5.408$

$= 15.408$

$10 - 5.408 = 4.592$

$4.592 \le \mu \le 15.408$

upper one side confidence interval

95%

$(\mu < \bar{X} + t_{\alpha, n-1} \frac{S}{\sqrt{n}})$

if we want the lower part

$\bar{X} - t_{\alpha, n-1} \frac{S}{\sqrt{n}}$

Chi distribution

$\chi^2 \Rightarrow$ to measure varianece

which is not semitrical

$\chi^2 = \frac{(n-1) S^2}{\sigma^2}$

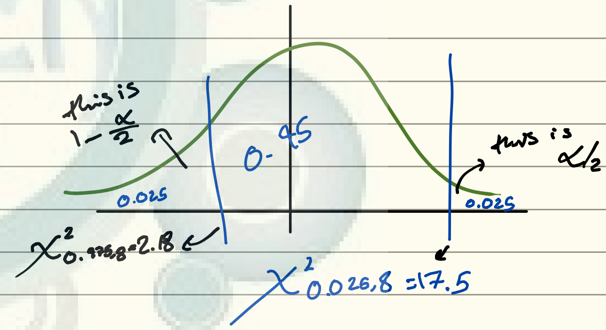
لا مستحيل تكون negative

if we are having $\chi^2 = 0 \Rightarrow$ that means that the area to the right = 1 so [illegible]

χ^2

this is upper

$\left(\frac{(n-1) s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \le \sigma^2 \le \frac{(n-1) S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)$



if I want to find the variance we calculate the variance then take the square root for both sides

& for one side the same but

$\chi^2_{1-\alpha}$

95%

$S = 3$

$n = 9$

↗ L

↘ U

for the upper

$$\left(\frac{8\ 3^2}{17.5} \le \sigma^2 \le \frac{8 \times 9}{2.18}\right)$$

One sided.

$$\frac{8 \times 9}{\chi^2_{0.95,8}} = \left(\sigma^2 \le \frac{8\ 3^2}{2.73}\right)$$

$$\frac{8\ \ 3^2}{\chi^2_{0.05,8}} \le \sigma^2$$

$$\left(\frac{8 \times 9}{15.5} \le \sigma^2\right)$$

⇓

هون لازم نقسم عالرقم

الكبير عشان يكون lower

وهناك نقسم عالرقم

الصغير عشان يكون

upper.

* Calculate 2 side C.I for σ^2

where

$\alpha = 0.95$

$n = 15$

$S^2 = 7$