*Ch 8: The normal distribution

التوزيع الطبيعي

• Re.call that.. a probability distribution determins the probability of the values of a random varibles X.

السوزيعات الاحتمالية تكون للمتغيرات العشوائية التي من الممكن أن تكون متصلة أو منفصلة.

متصلة continuous — منفصلة discrete.

• Three probability distributions: — لدينا ٣ توزيعات نوزعها دراستها

① The Normal distribution. «continuous random variable»
التوزيع الطبيعي خاص بالمتصلة «Ch8»

② The Binomial distribution. «discrete random variable»
توزيع ذو الحدين خاص بالمنفصلة «Ch9»

③ Poisson distribution. «discrete random variable»
توزيع بواسون في Ch10 خاص بالمنفصلة.

*Normal distribution:

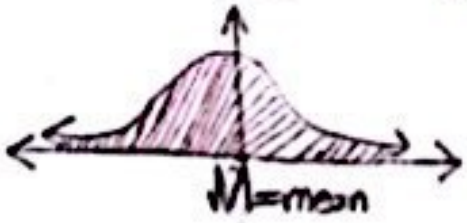
The most important probability for describing a continuous random variable. أهم توزيع للمتغيرات المتصلة هو التوزيع الطبيعي.

* The normal distribution has been used in a wide variety of practical applications in which the random variables are heights, weights, test scores, ...

• تُستخدم التوزيع الطبيعي في تطبيقات كثيرة منها الأوزان، العلامات،

*Normal curve: منحنى التوزيع الطبيعي

① It has a bell-shaped. شكله مثل الجرس



② The normal curve has 2 parameters M and σ, they determine the location and shape of the normal distribution.

يتم تحديده وتحديد شكله من خلال المتوسط mean و الانحراف المعياري Standard deviation

③ Mean = Median = Mode.

→ the mean can be positive, negative, or 0.

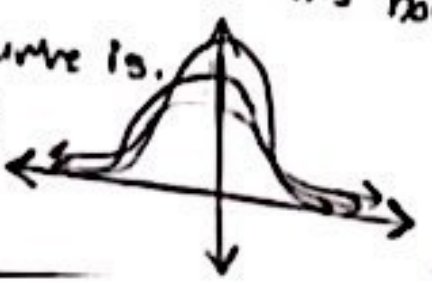
يكون فيه الوسط = الوسيط = المنوال ويمكن أن يكون + أو - أو 0

④ The normal distribution is symmetric about the mean.

التوزيع الطبيعي متماثل حول الوسط

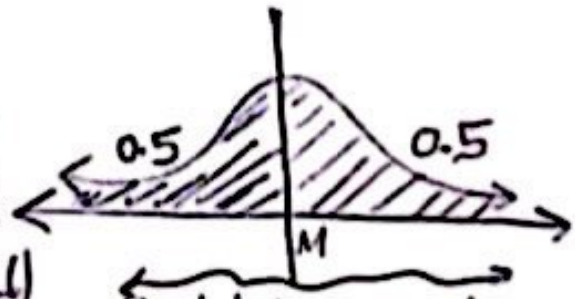
⑤ The standard deviation σ or δ determines how flat and wide the normal curve is.

الانحراف المعياري Standard deviation σ or δ تحدد مدى اتساع المنحنى الطبيعي.



⑥ The total area under the curve is 1.

→ Because the distribution is symmetric, the area under the curve to left of the mean is 0.5, and the area under the curve to the right of the mean is 0.5.



المساحة الكلية تحت المنحنى قيمتها 1 لذا فإن المساحة على يمين الصفر = 0.5 وعلى يساره = 0.5 لأنه متماثل.

*Standard normal distribution: التوزيع الطبيعي المعياري

A random variable has a normal distribution with a mean of 0 and a standard deviation of 1.

The letter Z is commonly used to designate this particular normal random variable.

هو توزيع طبيعي يكون فيه قيمة الوسط الحسابي "M" = 0 والانحراف المعياري "σ" = 1 وسنستخدم Z للتعبير عنه.

• To find the probability that a normal random variable is within any interval, we must compute the area under the curve. المساحة تحت المنحنى هي تمثيل الاحتمال

- For the standard normal distribution, areas under the curve have been computed and are available in tables that can be used to compute probabilities. لإيجاد المساحة في التوزيع الطبيعي المعياري سنستخدم جدول.

- Z-table: جدول Z
a table that is used to find probability for the normal distribution.
جدول يحدد لنا قيمة الاحتمال والمساحة في التوزيع الطبيعي.

* Computing probabilities for any normal distribution:-
the probability for all normal distributions are computed by using the standard normal distribution.

→ Converting to the standard normal random variable

$$Z = \frac{X - \mu}{\sigma} \quad ; \quad Z: +ve, 0, -ve.$$

سنحوّل أي normal إلى standard normal باستخدام العلاقة $Z = \frac{X-\mu}{\sigma}$

- Ex: The weights of BZU students have a normal distribution with a mean of 60 Kg and standard deviation of 10 Kg. Find the Z-scores.

If $X = 60 \rightarrow Z = \frac{X-\mu}{\sigma} = \frac{60-60}{10} = 0$

If $x=70 \rightarrow z=\frac{70-60}{10}=1$

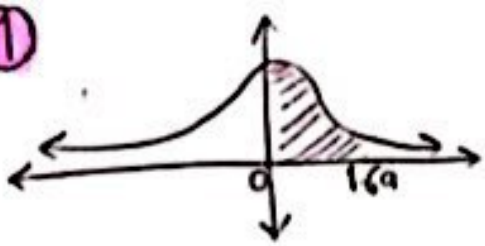
If $x=48 \rightarrow z=\frac{48-60}{10}=-1.2$

* Z- table:

The Z-table gives only the area between a positive Z-value and $z=0$. Symmetry helps us find other areas.

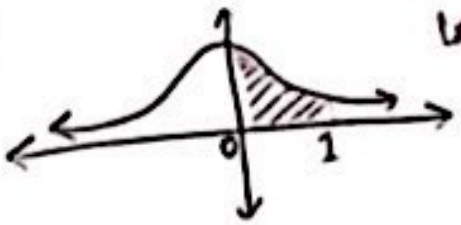
نستخدم جدول Z الذي يحدد لنا المساحة من الصفر الى اي عدد موجب والتماثل.

- Ex:

① 

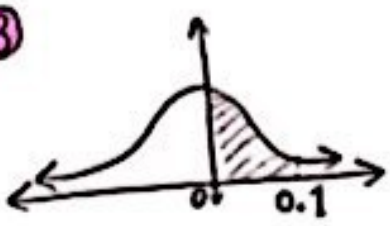
نجدها من الجدول مباشرة

1.6 | 0.0 | 0.4545

② 

نجدها من الجدول، نعين 1.00

1.0 | 0.00 | 0.3413

③ 

نجدها من الجدول، نعين a10

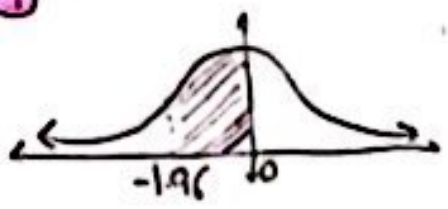
0.1 | 0.00 | 0.0398

④

نستخدم التماثل

0.1 | 0.0398

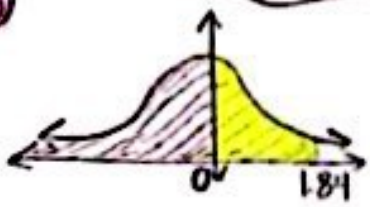
④



نستخدم التماثل المساحة على يسار الصفر هي نفسها التي على يمين الصفر كما ان نعين على الجدول:

0.06 | 0.4750
1.9

⑤

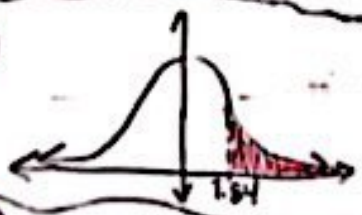


الجدول يجد لنا المساحة من الصفر لعدد موجب اي المنطقة الملونة باصفر

0.09 | 0.4671
1.8

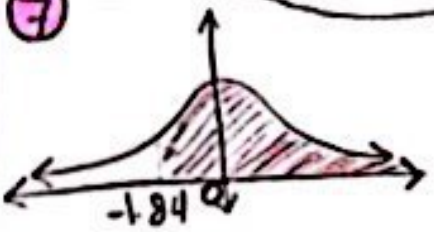
ونعلم ان المساحة على يسار الصفر = 0.5
مساحة المنطقة كاملة = 0.5 + 0.4671 = 0.9671

⑥



الجدول يجد المساحة من الصفر لـ 1.84 ونعلم ان المساحة على يمين الصفر كاملة = 0.5 اذن مساحة المنطقة المطلوبة = 0.5 - 0.4671
= 0.329 ← نجدها الارقام من الجدول

⑦

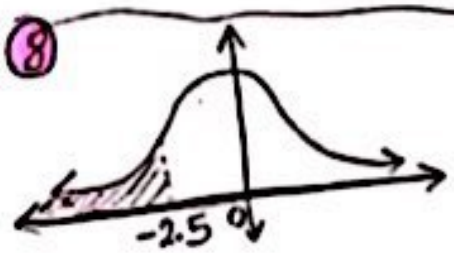


نجد من الجدول المساحة من صفر الى 1.84- :

0.04 | 0.4671
1.8

نضيف لها المساحة التي على يمين الصفر وتساوي 0.5
← المساحة المنطقة المطلوبة = 0.5 + 0.4671 = 0.9671

⑧

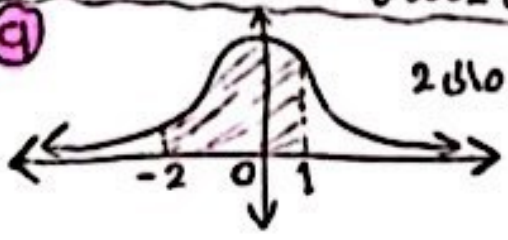


نجد المساحة من صفر الى 2.5- من الجدول ونعلم ان المساحة على يسار الصفر = 0.5 نطرح

0.00 | 0.4938
2.5

← مساحة المنطقة المطلوبة = 0.5 - 0.4938 وتساوي 0.0062

⑨



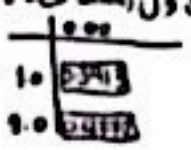
نجد من الجدول المساحة من صفر الى 1 وايضاً من 0 الى 2 ونجمعها:

0.00 | 0.3413
1.0 | 0.4772

المساحة الكلية = 0.3413 + 0.4772
= 0.8185

⑩

نجد المساحة من صفر الى 1.32 ونطرح منها المساحة من صفر الى 0.8

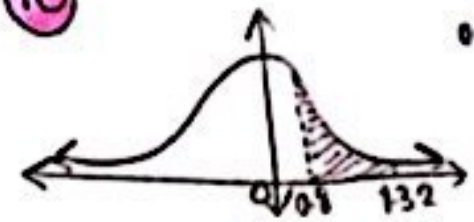


= المساحة الكلية
0.3413 + 0.4772
= 0.8185

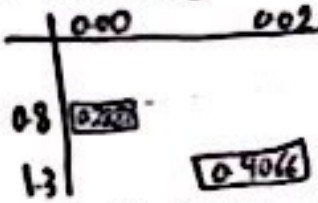
بنخصها:

---

(10)

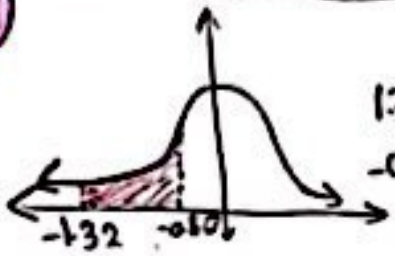


نجد المساحة من صفر الى 1.32 ونطرح منها المساحة من صفر الى 0.8
يتبقى المنطقة المظللة



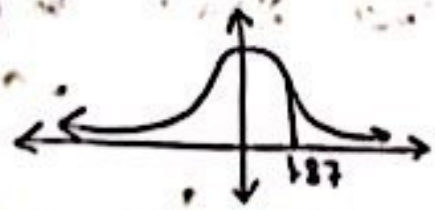
اذن مساحة المنطقة المطلوبة = 0.4066 - 0.2881
= 0.1185

(11)



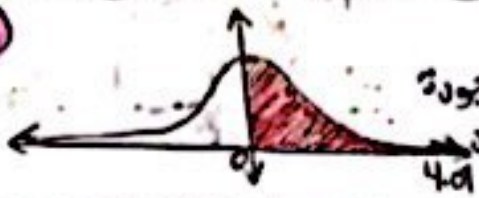
نجد المساحة من 0 الى 1.32- وهي نفس المساحة من 0 الى 1.32
اي نجدها من الجدول مباشرة ونجد المساحة من 0 الى 0.8-
من الجدول مباشرة ونطرحها فيتبقى مساحة المنطقة
المطلوبة: 0.4066 - 0.2881
= 0.1185

(12)



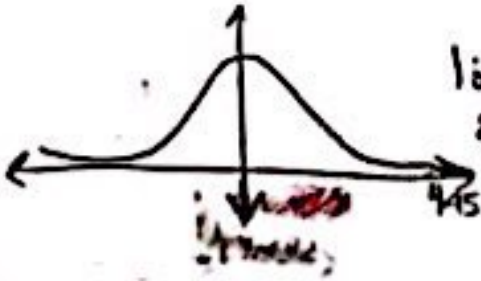
اذا اردنا ايجاد المساحة عند نقطة وليس
على منطقة معين لو اردنا ايجاد المساحة عند
1.87 مثلاً فتكون المساحة صفر دائماً.

(13)



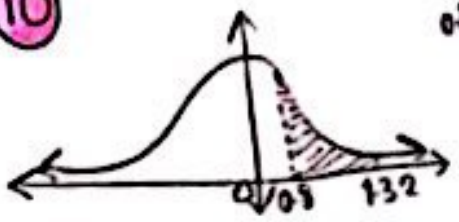
نريد ايجاد المساحة من صفر الى 4.01
نجدها من الجدول مباشرة ولكن 4.01 غير موجودة
في الجدول فاذن تعتبر المساحة لاي عدد غير موجود
تقريباً 0.5 فالمساحة المطلوبة = 0.5

(14)

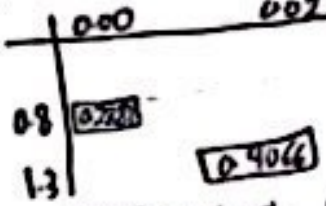


نريد ايجاد المساحة على يمينه 4.95 ولكن هذا
العدد غير موجود في الجدول ونعلم ان المساحة
قبله اي من صفر الى 4.95 تساوي 0.5
والمساحة الكلية على يمين الصفر = 0.5
اذن مساحة المنطقة المطلوبة = صفر.

(10)

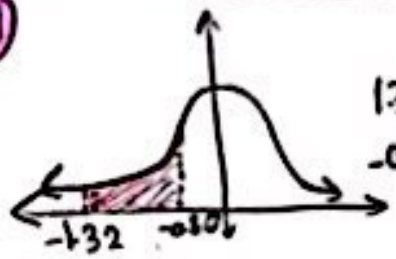


نجد المساحة من صفر الى 1.32 ونطرح منها المساحة من صفر الى 0.8 يتبقى المنطقة المظللة



اذن مساحة المنطقة المطلوبة = 0.4066 - 0.2881
= 0.1185

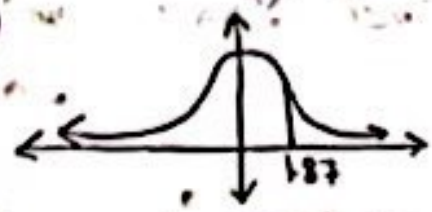
(11)



نجد المساحة من 0 الى 1.32- وهي نفس المساحة من 0 الى 1.32 اي نجدها من الجدول مباشرة ونجد المساحة من 0 الى 0.8- من الجدول مباشرة ثم نطرحهما فيتبقى مساحة المنطقة المطلوبة:

0.4066 - 0.2881
= 0.1185

(12)

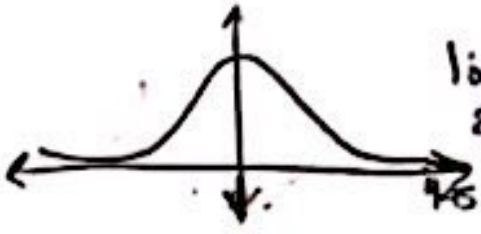


• اذا أردنا ايجاد المساحة عند نقطة وليس على منطقة معين لو اردنا ايجاد المساحة عند 1.87 مثلاً تكون المساحة صفر دائماً.

(13)

نريد ايجاد المساحة من صفر الى 4.01
نجدها من الجدول مباشرة ولكن 4.01 غير موجودة في الجدول لماذن نعتبر المساحة لأن العدد غير موجود تقريباً 0.5 فالمساحة المطلوبة = 0.5

(14)



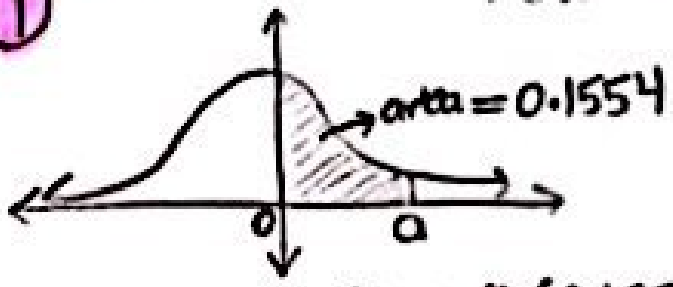
نريد ايجاد المساحة على يمينه 4.95 ولكن هذا العدد غير موجود في الجدول ونعلم أن المساحة قبله اي من صفر الى 4.95 تساوي 0.5
والمساحة الكلية على يمين الصفر = 0.5
إذن مساحة المنطقة المطلوبة = صفر.

* Now, we ask about finding a Z-value if the area is given:

كيفية ايجاد العدد a اذا عرفنا المساحة اي المساحة معطاة ونريد ايجاد قيمة a المقابلة لها.

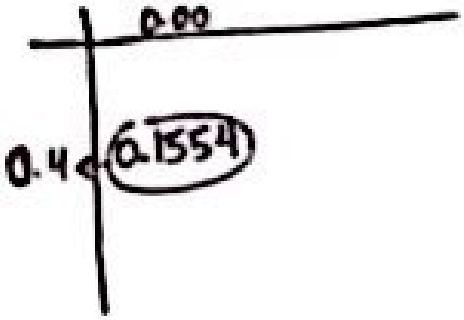
Ex:

①



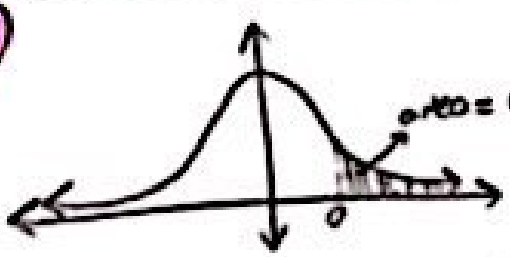
← مساحة المنطقة المظللة = 0.1554 المساحة كانت معطاة من السؤال ونريد ايجاد a:

نبحث عن المساحة في الجدول:



∴ a = 0.4

②



← مساحة المنطقة المظللة = 0.0485 نريد ايجاد العدد a الذي يجعل المساحة على يمينه 0.0485 حتى نتمكن من البحث عن هذه المساحة في الجدول، يجب ايجاد المساحة من 0 الى a وذلك بطرح 0.5 - 0.0485

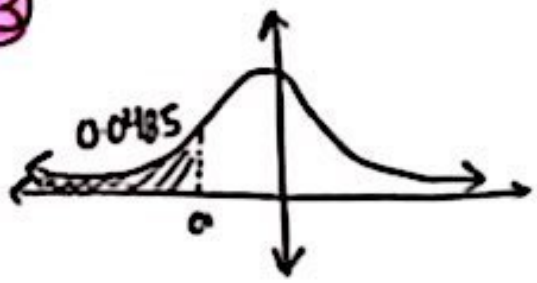
لاننا نعلم ان المساحة على يسار الصفر 0.5، نطرح منها مساحة المنطقة المعطاة، ينتج المساحة من 0 الى a وتساوي

0.5 - 0.0485 = 0.4515

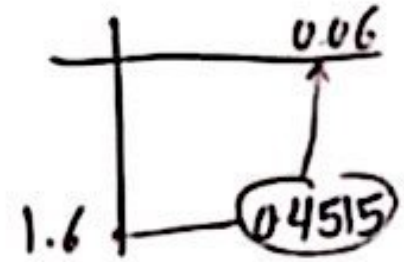
نبحث عن العدد في الجدول

∴ a = 1.66

(3)



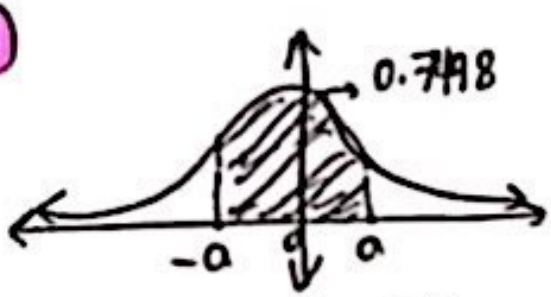
لمعرفة العدد a فنجد المساحة من 0 الى a وذلك بطرح 0.5 - 0.0485
لان المساحة على يسار الصفر = 0.5 فلو طرحنا منها المنطقة المظللة يتبقى المساحة من 0 الى a وتساوي 0.5 - 0.0485
= (0.4515) ← نبحث عن هذه المساحة من الجدول



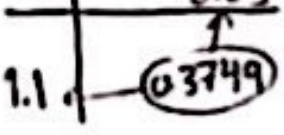
∴ a = -1.66

وضعنا للعدد a إشارة سالب لأنه على يسار الصفر فإذن فهو سالب.

(4)

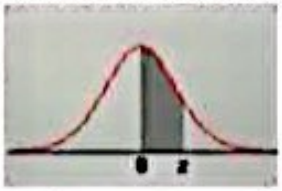


لإيجاد العدد a يجب معرفة المساحة من 0 الى a ومن ثم نبحث عنها في الجدول، المساحة من 0 - الى a تساوي 0.7498 نقسمها على 2
لمعرفة المساحة من 0 الى a = $\frac{0.7498}{2}$ = 0.3749 نبحث عن هذه المساحة في الجدول



∴ a = 1.15
-a = -1.15

Standard Normal (Z) Table
Area between 0 and z



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

- If x is normally distributed, we will transform x to the standard normal Z, then use the Z-table to find areas. (probabilities).

- Ex: The blood pressure of men is normally distributed with a mean of 120 and standard deviation of 8. $M = 120$, $\delta = 8$, normal.

① What is the probability that a man's blood pressure is between 100 and 130 ?

$\longrightarrow P(100 < x < 130)$

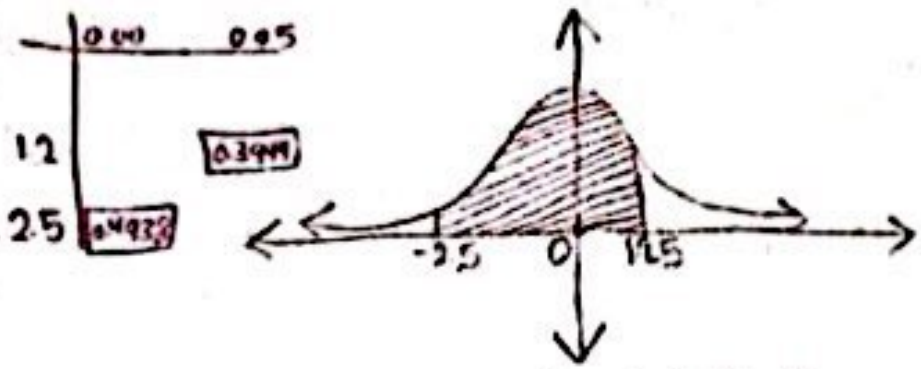
Z-score لكل x نحول Z-table حتى نستطيع استخدام

$$Z = \frac{x - M}{\delta}$$

if $x = 100 \longrightarrow Z = \frac{100 - 120}{8} = -2.5$

$x = 130 \longrightarrow Z = \frac{130 - 120}{8} = 1.25$

$\longrightarrow P(-2.5 < Z < 1.25)$ نرسم المنطقة



نجد هاي العمود — نجد المساحة من الجدول

$$P(-2.5<Z<1.25) = P(-2.5<Z<0) + P(0<Z<1.25)$$

$$= 0.4938 + 0.3944$$

$$= 0.8882$$

② What is the percentage of men whose blood pressure is at least 135 ?
على الأقل

$$\longrightarrow P(X \geq 135)$$

نحوّل من X إلى Z .

$$Z = \frac{X - M}{\delta}$$

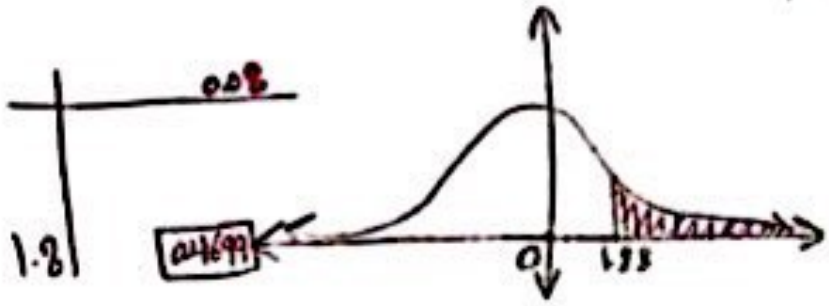
$$\text{if } X=135 \text{ ; } Z = \frac{135-120}{8} = 1.875 \cong 1.88$$

نقرّبها ، نأخذ فقط أول منزلتين بعد الفاصلة ونقرّب الثالث فيصبح 1.88.

ـ علماً في حال كان ناتج Z عدد يحتوي على أكثر من منزلتين بعد

العاصلة 6 نقريبه .

→ $P(Z \geq 1.88)$ نرسم المنطقة



لايجاد مساحة هذه المنطقة نجد من الجدول من 0 الى 1.88 ثم نطرحها من 0.5

$\therefore P(Z \geq 1.88) = 0.5 - 0.4699 = 0.0301$

$\text{Percentage} = 0.0301 \times 100\% = 3.01\%$

(3) What is the percentage of men whose blood pressure is less than 135?

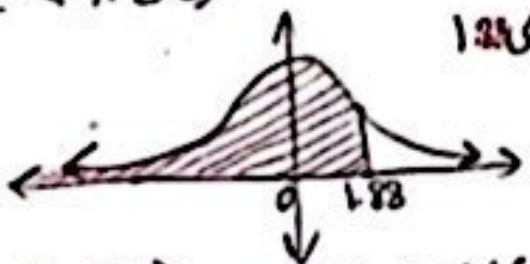
أقل من

→ $P(X < 135)$

$$Z = \frac{X - M}{\sigma}$$

$X = 135 \rightarrow Z = \frac{135 - 120}{8} = \frac{15}{8} = 1.88$

→ $P(Z < 1.88)$



نجد المساحة من 0 الى 1.88 ثم نضيف لها 0.5

$\therefore P(Z < 1.88) = 0.5 + 0.4699 = 0.9699 \rightarrow$ percentage $= 96.99\%$

$\therefore P(Z<1.88) = 0.5 + 0.4699 = 0.9699$
$= 96.99\%$

• ملاحظة هامة: اشارة المساواة عند Z والاشوتر على اتجاه المساحة (الاحتمال).

- Note:
In the normal distribution, the area at a single value is 0. Therefore, $P(Z_1 < Z < Z_2) = P(Z_1 \leq Z \leq Z_2)$
يعني أن المساواة لا تؤثر لان المساحة عند نقطة تساوي صفر.

(4) What is the probability that a man's blood pressure is between 130 and 140.

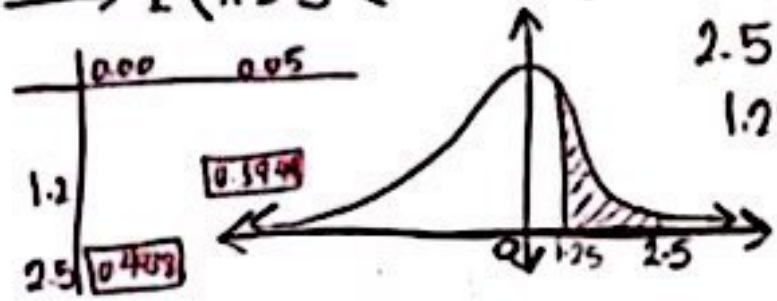
$\rightarrow P(130 < X < 140)$

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

if $X = 130$: $Z = \frac{130 - 120}{8} = 1.25$

$X = 140$: $Z = \frac{140 - 120}{8} = 2.5$

$\rightarrow P(1.25 < Z < 2.5)$



نجد المساحة من 0 الى 2.5
ونجد ايضا من 0 الى 1.25
ونطرحهما.

Partner 2:05 PM 19%
ritaj.birzeit.edu

CamScanner الممسوحة ضوئيا بـ

$\therefore P(1.25 < Z < 2.5) = 0.4938 - 0.3944$

$= 0.0994$

أي أن كان المطلوب percentage نضربها في 100%.

(5) What is the probability of men whose blood bressure is at most 110.
على الأكثر

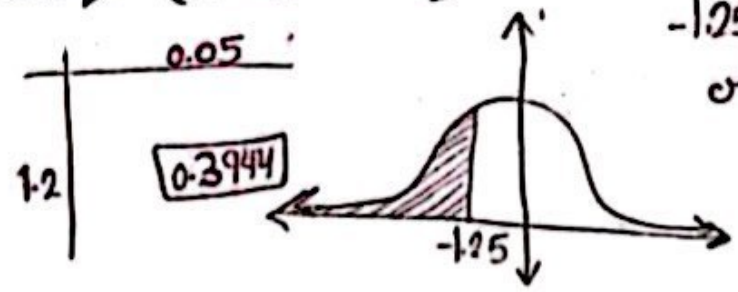
$\longrightarrow P(X \leq 110)$

$Z = \frac{X - M}{\delta}$

if $X = 110$: $Z = \frac{110 - 120}{8} = -1.25$

$\longrightarrow P(Z \leq -1.25)$

نجد المساحة من الى -1.25 من الجدول، ونطرحها من 0.5.

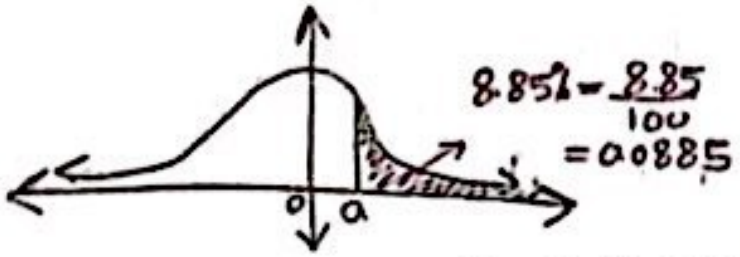


$\therefore P(Z < -1.25) = 0.5 - 0.3944 = 0.1056$

CamScanner الممسوحة ضوئيا بـ

⑥ The blood pressure of a man is considered too high if it is in the top 8.85%. What is the too high blood pressure?

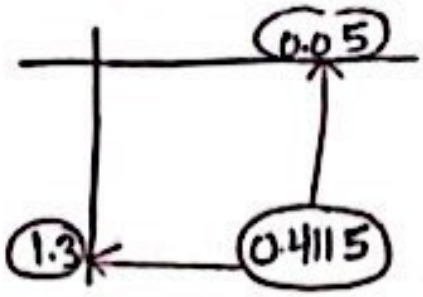
نعتبر ضغط الدم عالي اذا كانت نسبته من أعلى 8.85% عامة الضغط الذي نعتبره مرتفع .



حتى نستطيع ايجاد a نجد المساحة من 0 الى a:

$0.5 - 0.0885 = 0.4115$

نبحث عن هذا العدد من الجدول:



$\therefore a = 1.35$

هذه القيمة تمثل Z لإيجاد الضغط يجب أن نجد X.

$$\rightarrow Z = \frac{X - M}{\sigma}$$

$$1.35 = \frac{X - 120}{8}$$

$$1.35(8) = X - 120$$

$$10.8 = X - 120$$

$\therefore \boxed{X = 130.8}$ or more

$1.35(8) = X - 120$

$10.8 = X - 120$
$+120 \quad +120$

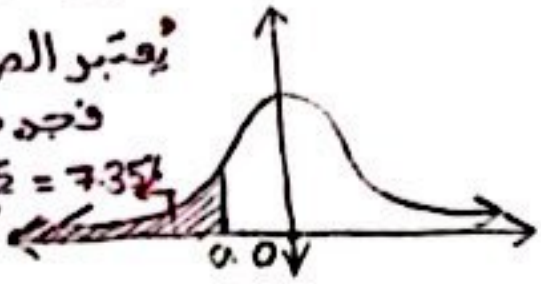
$\therefore \boxed{X = 130.8}$ or more

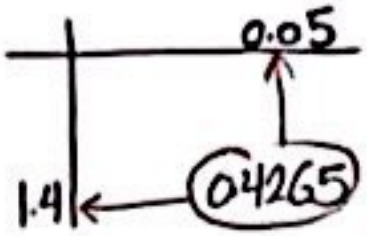
اذا الضغط 130.8 او اكثر يعتبر مرتفع .

(7) The blood pressure of a man is considerd too low if it is in the bottom 7.35%. What is the too low blood pressure?

يعتبر الضغط منخفض اذا كان من أقل 7.35%
فنجد قيمة الضغط المنخفض.

$0.0735 = \frac{7.35}{100} = 7.35\%$





نجد المساحة من 0 الى a
ونبحث عنها في الجدول
$0.5 - 0.0735$
$= 0.4265$

$\therefore 0 = -1.45$
$\therefore$ نجد قيمة X

$$Z = \frac{X - M}{\sigma}$$

$$-1.45 = \frac{X - 120}{8}$$

$$-1.45(8) = X - 120$$

$$-11.6 = X - 120$$
$$+120 \quad +120$$

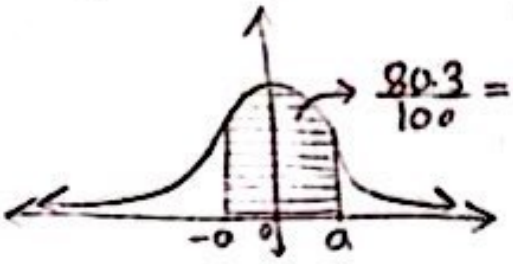
$\therefore \boxed{X = 108.4}$ or less

Partner 2:05 PM 19%
ritaj.birzeit.edu

$-11.6 = X-120$ $\quad +120 \quad +120$ $\quad \therefore \boxed{X = 108.4}$ or less

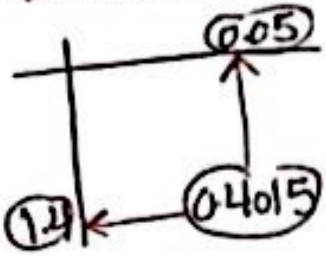
(8) For a medical study, men with blood pressure in the middle 80.3% are needed. Which men would fit this study?

يريده أن تكون المساحة في المنتصف أي الاحتمال يساوي 80.3%.



لإيجاد المساحة من الجدول نجد من 0 الى a وذلك بقسمة المساحة المعطاة على 2.

$$\frac{0.803}{2} = 0.4015$$



$\therefore a = 1.45$, $-a = -1.45$

نحول هذه القيم الى X:

$$\rightarrow Z = \frac{X - M}{\sigma}$$

$$1.45 = \frac{X-120}{8}$$

$$1.45(8) = X - 120$$

$$11.6 = X - 120$$
$$+120 \quad +120$$

$\therefore \boxed{X = 131.6}$

+120 +120

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

$-1.45 = \frac{X-120}{8}$

$-145(8) = X-120$

$-11.6 = X-120$

$+120 \quad +120$

$\therefore \boxed{X = 108.4}$

∴ Men with blood pressure between 108.4 and 130.32

Ex) BZU students consume on average 1.5 cups of coffee per day. Assume the variable has a normal distribution with a standard deviation of 0.4 cups. If 1000 students are selected, how many will drink more than one cup every day?

$\mu = 1.5, \ \sigma = 0.4$, normal, $n = 1000$

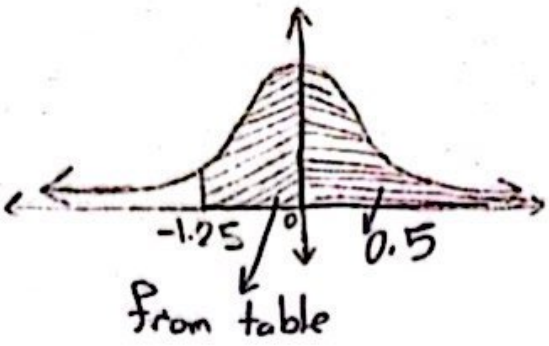
$\longrightarrow P(X > 1)$

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

$X = 1: \quad Z = \frac{1-1.5}{0.4} = -1.25$

$\longrightarrow P(Z > 1.25)$ نرسم المنطقة

Partner 2:05 PM @ 19%
ritaj.birzeit.edu



from table

$\therefore P(Z > 1.25) = 0.5 + 0.3944$

$= 0.8944$

# of students $= (1000)(0.8944)$

$= 894.4 \cong 894$ students.

Ex: The monthly income for palestinian families are normally distributed with mean equal to 5000 and a standard deviation of 1500.

$\longrightarrow M = 5000, \delta = 1500$, normal.

(a) What is the percentage of palestinian families whose monthly income are more than 8000? أكثر من

$\longrightarrow P(x > 8000)$

$$Z = \frac{x - M}{\delta} = \frac{8000 - 5000}{1500} = 2$$

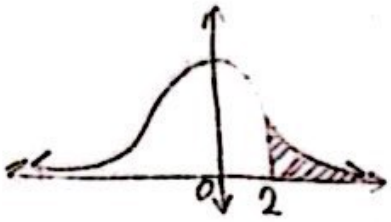
$\longrightarrow P(Z > 2)$

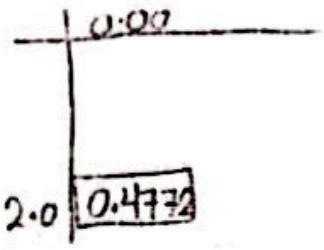
Partner 2:06 PM 19%
ritaj.birzeit.edu

$$Z = \frac{X - M}{\sigma} = \frac{8000 - 5000}{1500} = 2$$

CamScanner

$\rightarrow P(Z > 2)$





$\therefore P(Z > 2) = 0.5 - 0.4772 = 0.0228$

Percentage $= 0.0228 \times 100\% = 2.28\%$

b) What is the probability that a randomly selete palestnion family has a monthly income between 6500 and 8000?

$\rightarrow P(6500 < X < 8000)$

$$Z = \frac{X - M}{\sigma}$$

if $X = 6500$: $Z = \frac{6500 - 5000}{1500} = 1$

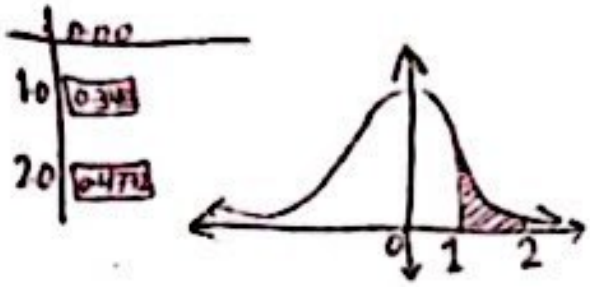
$X = 8000$: $Z = \frac{8000 - 5000}{1500} = 2$

$\rightarrow P(1 < Z < 2)$

CamScanner

1500

12 of 16 $<Z<2)$



نجد المساحة من 0 الى 2
ونجد المساحة من 0 الى 1
ونطرحهما.

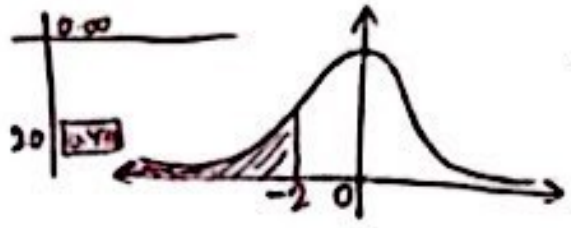
$P(1<Z<2) = 0.4772 - 0.3413 = 0.1359$

c) A family is considered poor if its monthly income is less than 2000. What is the percentage of poor palestinian families?

$\longrightarrow P(X<2000)$

$X = 2000: Z = \frac{X-\mu}{\sigma} = \frac{2000-5000}{1500} = -2$

$\longrightarrow P(Z<-2)$ نرسم المنطقة



نجد المساحة من -2 الى 0
ونطرحها من 0.5

$\therefore P(Z<-2) = 0.5 - 0.4772 = 0.0228$

The percentage of poor families $= 0.0228(100)\%$
$= 2.28\%$

$= 2.28\%$

- Ex: The weights of newborn babies are normally distributed with a mean of 4.2Kg and a standard deviation of 1.2Kg.

$M = 4.2Kg,\ \sigma = 1.2Kg$, normal.

a) What is the probability that a newborn baby will weigh more than 5Kg.

$\longrightarrow P(x > 5)$

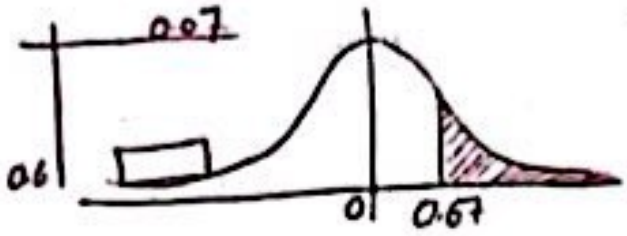
$$Z = \frac{x - M}{\sigma}$$

$$x = 5: \quad Z = \frac{5 - 4.2}{1.2} = 0.666\ldots \cong 0.67$$

$\longrightarrow P(Z > 0.67)$.

نرسم المنطقة

نجد المساحة من 0 الى 0.67 ونطرحها من 0.5



$$\therefore P(Z > 0.67) = 0.5 - 0.2486 = 0.2514$$

Partner 2:06 PM 19%
ritaj.birzeit.edu

$\therefore P(Z > 0.67) = 0.5 - 0.2486 = 0.2514$

---

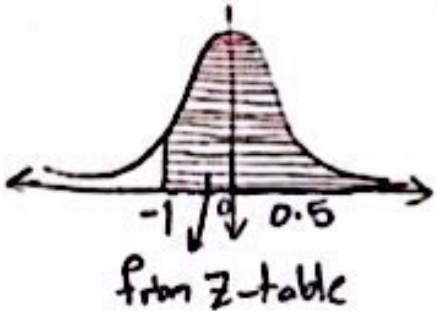
b) What is the percentage of newborn babies who weighs at least 3 kg ?

على الأقل

$\rightarrow P(X \geq 3)$

$X = 3: \quad Z = \frac{X - M}{\sigma} = \frac{3 - 4.2}{1.2} = -1$

$\rightarrow P(Z \geq -1)$ نرسم المنطقة



نجد المساحة من 0 الى -1 ونجمعها مع 0.5

$\therefore P(Z > -1) = 0.5 + 0.3413 = 0.8413$

$\text{Percentage} = 0.8413 \times 100\% = 84.13\%$

c) last week 20 newborn babies were at Ramallah's hospital. How many of them do you expect to weigh at least 3 kg?

$\longrightarrow P(X \geq 3) = 0.8413$

وجدناها من الفرع السابق فلا داعي لإيجادها مرة أخرى

$\longrightarrow n = 20$:

# of newborn babies weigh at least 3kg =

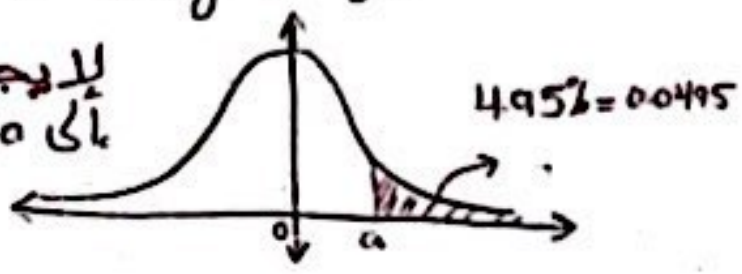
$(20)(0.8413) = 16.826 \cong 17$

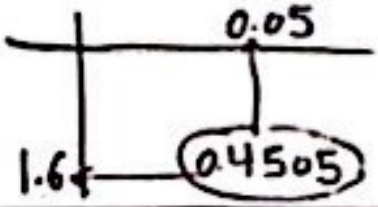
d) A newborn is considered heavy baby if his/her wight is in the top 4.95%. How much a newborn must weight to be a heavy baby?

لإيجاد a نجد المساحة من 0 إلى a:

$0.5 - 0.0495 = 0.4505$

نبحث عن الرقم من الجدول





$\therefore a = 1.65$

نجد X.

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

$$1.65 = \frac{x - 4.2}{1.2}$$

$$1.65(1.2) = x - 4.2$$

$$1.98 = x - 4.2$$

$$+4.2 \qquad +4.2$$

$\therefore$ $\boxed{x = 6.18}$ or more.

ritaj.birzeit.edu

*Ch9: The binomial distribution توزيع ذو الحدين

- The **binomial probability** distrebution is a **discrete** probabity distribution that prorides many applications.
يكون التوزيع ذو الحدين للمتغيرات المنفصلة

- **Binomial experiment**: A probability experiment that satisfies the following conditions: تجربة ذو الحدين

1. There is a fixed number (n) of identical trials.
يجب أن يكون لدينا عدد من المحاولات.
2. 2 outcomes are possible on each trial. في كل محاولة خيارين:
   The outcome we study: S « Success» الخيار الذي نريد دراسته هو النجاح
   The complement: F « Failure». والآخر الفشل
3. The trials are independent. تكون المحاولات مستقلة للا شيء تركل واحدة على الأخرى
4. The probability of a success denoted by p, doesn't change. (That is, the probability of success remains the same). احتمال النجاح p يكون ثابت في كل المحاولات.

٧:٤٨ م JAWWAL
ritaj.birzeit.edu

- Ex: In tossing a coin 3 times.
let X be the number of heads.

→ X = 0, 1, 2, 3 عدد محاولات النجاح أي الحصول على H مثل 0, 1, ...

We have 3 trials. ∴ n = 3
The trials are independent: كل محاولة لا تتأثر على الأخرى
2 outcomes: H, T
H: Success احتمال النجاح أي الإسناد
T: Failure. احتمال الفشل يساوي $\frac{1}{2}$ واحتمال الفشل يساوي $\frac{1}{2}$

∴ the experiment is called a binomial experiment.

- Ex: If 40% of BZU students are smokers, and we took a sample of 25 students. let X: the number of smokers.

→ X = 0, 1, 2, ..., 25. عدد المدخنين ممكن 0 أو 1 أو 2, ... 25

We have 25 trials. ∴ n = 25
The trials are independent.
2 outcomes: Yes, No. كل شخص ممكن أن يجيب بنعم أي أنه مدخن أو لا.
Smoker: Success احتمال أن يكون الشخص مدخن في كل محاولة 40%
nonsmoker: Failure

- Binomial probability formula:

JAWWAL
ritaj.birzeit.edu

Smoker: Success
nonsmoker: Failure

- Binomial probability formula:

① $P(X)=\binom{n}{x}p^x(1-p)^{n-x}$;

X: # of successes.

n: # of trials.

$\binom{n}{x}=\frac{n!}{x!(n-x)!}=n\boxed{nCr}x$ "X successes choosen from n trials".

p: probability of success.

② The expected value "Mean":

$E(X)=M=np$ الوسط أو التوقع = np

③ The standard deviation:

$\sigma=\sqrt{np(1-p)}$ الانحراف المعياري

④ The variance:

$\sigma^2=np(1-p)$. التباين

⑤ $P(0)+P(1)+\cdots+P(n)=1=\sum_{x=0}^{n}P(x)$ مجموع الاحتمالات = 1

- Ex: 30% of workers take public transportation

ritaj.birzeit.edu

مجموع الاحتمالات $\sum_{x=1} f(x) = 1 =$

CamScanner

- **Ex:** 30% of workers take public transportation daily, in a sample of 10 independent workers.

Binomial, $n=10$, p(take public transportation) $= 30\% = \frac{30}{100} = 0.3$

a) What is the probability that exactly 3 workers take public transportation?

$$\rightarrow P(X=3) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x=3, n=10, p=0.3$$

$$= \binom{10}{3} (0.3)^3 (0.7)^7$$

from the calculator.

$$= 10 \boxed{nCr} 3 \boxed{X} 0.3 \boxed{\wedge} 3 \boxed{X} 0.7 \boxed{\wedge} 7 \boxed{=}$$

$$= 0.2668 \cong 0.27$$

b) What is the probability that exactly 6 workers take public transportation?

$$\rightarrow P(X=6) = \binom{10}{6} (0.3)^6 (0.7)^4$$

$$= 10 \boxed{nCr} 6 \boxed{X} 0.3 \boxed{\wedge} 6 \boxed{X} 0.7 \boxed{\wedge} 4 \boxed{=}$$

$$= 0.03667 \cong 0.037$$

c) What is the probability that exactly 6 workers don't take public transportation?

CamScanner

p(don't take public transportation) $= 70\% = \frac{70}{100} = 0.7$

ritaj.birzeit.edu

don't take public transportation?

9 من 5

P(don't take public transportation) = 70% = $\frac{70}{100}$

$\rightarrow P(X=6) = \binom{10}{6}(0.7)^6(1-0.7)^4$

$= 10\ nCr\ 6\ X\ 0.7\ \wedge\ 6\ X\ 0.3\ \wedge\ 4\ =$

$= 0.2001$

d) What is the probability that all of them take public transportation? $X = 10$

$\rightarrow P(X=10) = \binom{10}{10}(0.3)^{10}(0.7)^0$

$= 5.9049 \times 10^{-4} = 0.00059049$

$\cong 0.00059$

e) What is the probability that at most 2 workers take public transportation?

2 على الأكثر

$\rightarrow P(0) + P(1) + P(2)$

$= \binom{10}{0}(0.3)^0(0.7)^{10} + \binom{10}{1}(0.3)^1(0.7)^9 + \binom{10}{2}(0.3)^2(0.7)^8$

$= 0.3827 \cong 0.383$

٧:٤٨ م JAWWAL
ritaj.birzeit.edu

CamScanner

e) What is the probability that at least 3 workers take public transportation daily? ↓ على الأقل 3

$\rightarrow P(3)+P(4)+P(5)+\cdots+P(10)$

$= 1-(P(0)+P(1)+P(2))$

$= 1-0.383 = 0.617$

J) What is the expected number of workers that take a public transportation?

$\rightarrow E(X)=M=np$

$=10(0.3)=3$

f) What is the expected number of workers that don't take a public transportation?

$\rightarrow E(X)=np$

$=10(0.7)=7$

CamScanner الممسوحة ضوئيا بـ

9) What is the variance of workers who take a public transportation?

$$\sigma^2 = np(1-p)$$
$$= 10(0.3)(0.7) = 2.1$$

- Ex: A survey found that one out of five palestinians visited a dentist last month. If 12 people selected at random, what is the probability that none of them visited a dentist.

p(visited a dentist) = 1 out of five $= \frac{1}{5} = 0.2$

$n = 12$

$X = 0$ (none of them)

$$\rightarrow P(X=0) = \binom{12}{0}(0.2)^0(0.8)^{12}$$
$$= 0.0687$$

- Ex: A coin is tossed 8 times. Find the mean, variance, and standard deviation of the # of heads.

$\rightarrow n = 8$, p(head) $= \frac{1}{2} = 0.5$

$\therefore \mu = np = 8(0.5) = 4$

$\sigma^2 = np(1-p) = 8(0.5)(0.5) = 2$

$\sigma = \sqrt{np(1-p)} = \sqrt{2} = 1.41$

CamScanner الممسوحة ضوئيا بـ

$2-\sqrt{np(1-p)} = \sqrt{2} = 1.41$

CamScanner الممسوحة ضوئيا بـ

-Ex: Suppose that 55% of BZU employees own private cars. We took a random sample of 10 independent employees.

a) What is the probability that 9 employees own private cars?

$$\longrightarrow P(9) = \binom{10}{9}(0.55)^9(0.45)^1$$

$$= 0.0207$$

b) What is the probability that more than 8 employees own private cars? (more than 8 = 9 or 10)

$$\longrightarrow P(9)+P(10) = \binom{10}{9}(0.55)^9(0.45)^1 + \binom{10}{10}(0.55)^{10}(0.45)^0$$

$$= 0.023$$

c) What is the probability that less than 9 employees own private cars? (less than 9 = 0 or 1 or ... or 8)

$$\longrightarrow P(0)+P(1)+\cdots+P(8)$$

$$= 1-(P(9)+P(10))$$

$$= 1-0.023 = 0.977$$

CamScanner الممسوحة ضوئيا بـ

d) What is the probability that less than 2 employees own private cars?

$= 0.023$

9 من 9 the probability that less than 9 employs own private cars?
9 or less

$\rightarrow P(0)+P(1)+\cdots+P(8)$

$= 1-(P(9)+P(10))$

$= 1-0.023 = 0.977$

d) What is the probability that less than 2 employees don't own private cars?
2 or less

$\rightarrow P(0)+P(1)$

$= \binom{10}{0}(0.45)^0(0.55)^{10}+\binom{10}{1}(0.45)^1(0.55)^9$

$= 0.023$

*Ch11: Introduction to inference
مقدمة في الإحصاء الاستدلالي.

• Recall that **Inferential statistics** means using samples to predict characteristics of the population. دراسة مجتمع باستخدام عينة

***Areas of inferential statistics:**

① Point Estimation. } Ch11
② Interval Estimation. }
③ Hypothese Testing } Ch12

***Point Estimation:** التقريب بنقطة

Usually, it is hard (or impossible) to find the population parameters μ and σ.

① $\bar{X}$ is the point estimator (best estimate) of μ.
② S is the point estimator (best estimate) of σ.

⟶ Statistics (point estimator) will vary from sample to sample.

ritaj.birzeit.edu

② $\bar{X}$ is the point estimator (best estimate) of μ ... statistics (point estimator) will vary from sample to sample.

-Ex: The following are the ages of a sample of BZU employees: 26, 45, 32, 38, 55, 60

① Find a point estimator of the ages mean.

→ $\bar{X} = 42.8$

by calculator: Mode 2 26 M+ ....60 M+ ON
Shift 2 1 =

② Find a point estimator of the ages standard deviation

→ $S = 12.95$

by calculator: Shift 2 3 =

③ Find a point estimator of the ages variance.

→ $S^2 = 167.7$

by calculator: 12.95 x^2 =

* Standard error of the mean (SE): الخطأ المعياري
measures how likely $\bar{X}$ is close to μ. يقيس مدى قرب $\bar{X}$ من μ

$$SE = \frac{S}{\sqrt{n}}$$; S: standard deviation.
n: sample size.

ritaj.birzeit.edu

SE = $\frac{S}{\sqrt{n}}$; S : standard deviation.
n: sample size.

→ Smaller value of standard errors (SE) indicate that $\bar{X}$ is closer to M. كلما كان SE أصغر يكون $\bar{X}$ أقرب على M.

Ex: Given the sample: 8, 5, 4, 10, 8. Find the standard error of the mean.

$n = 5$, $S = 2.45$ by calculator

→ $SE = \frac{S}{\sqrt{n}} = \frac{2.45}{\sqrt{5}} = 1.095 \cong 1.1$

- Note:
When n increases, SE decreases. كلما زادت n حجم العينة يقل SE.

* Interval Estimation of M: التقدير بفترة
We will use the sample to construct a special interval estimate for M called the confidence interval.

When S is used to estimate σ, the margin of error and the interval estimate are based on a probability distribution Known as t-distribution. عند تقريب σ بـ S فإن يوجد هامش خطأ يعتمد على توزيع يسمى t-distribution.

- The t-distribution is a family of similar probability distributions, with a specific t distribution depending on a parameter Known as the degrees of freedom. يعتمد توزيع t على درجات الحرية d.f.

ritaj.birzeit.edu

يعتمد t-distribution على درجات الحرية "df".

CamScanner

df \ α	0.1	0.05	0.025	0.01	.....
1					
2					
3					
⋮					
∞					

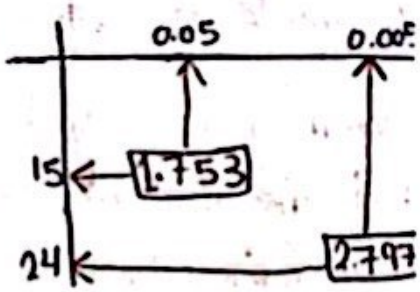
t-table

توزيع t يعتمد على parameter يمثل درجات الحرية "df".

Ex: Find:

① $t_{0.05,15} = 1.753$

② $t_{0.05,24} = 2.797$



CamScanner

- Each confidence interval has: لكل فترة ثقة:

• Each confidence interval has: لكل فترة ثقة:

① Confidence level $(1-\alpha)\%$: the probability that the confidence interval contains M. مستوى الثقة: احتمال ان يكون M موجود في الفترة.

→ The common confidence level are 90%, 95%, 99%

② Significance level α: The risk that the confidence interval doesn't contain M. مستوى الدلالة α: نسبة الخطر بكون M لا تنتمي الفترة

Ex: If a 95% confidence interval for the daily mean of car accidents in Palestine is (12, 20), then this means that we are 95% confident that the the population mean of car accidents is between 12 and 20.
نثق بنسبة 95% ان M تقع بين الفترة.

*To construct a $(1-\alpha)\%$ confidence interval ; C.I طريقة لايجاد فترة الثقة

① determine α. ← the risk that the confidence interval doesn't contain M.

② determine. margin of error: هامش الخطأ

$$E = t_{\frac{\alpha}{2},df} \cdot S_E = t_{\frac{\alpha}{2},df} \cdot \frac{S}{\sqrt{n}} \;;$$

$df = n - 1$; n sample size.

S = standard deviation.

$t_{\frac{\alpha}{2},df}$ ← نجدها من الجدول ← df | $\frac{\alpha}{2}$

ritaj.birzeit.edu

S = standard deviation.

11 من 6 ,df ← ابحث في الجدول عن df

CamScanner

③ The interval estimate of a population mean:

$(\bar{X} - E, \bar{X} + E)$; $\bar{X}$: sample mean.
E: margin of error.

Ex: To estimate the mean price of iphones, we took a sample of 20 iphones. This sample gave us a mean of 1400 NIS and standard deviation of 260 NIS.

a) Construct a **90% CI** for the true mean price, then interpret your answer.

$\longrightarrow n = 20, \bar{X} = 1400, S = 260$

confidence level = 90% $\longrightarrow$ significance level = 10%

$\therefore \alpha = \frac{10}{100} = 0.1$

$$S_E = \frac{S}{\sqrt{n}} = \frac{260}{\sqrt{20}} = 58.14$$

$E = t_{\frac{\alpha}{2}, df} \cdot S_E$; $\alpha = 0.1 \longrightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

$= 1.729 \times 58.14$

$= 100.52$

$df = n - 1 = 20 - 1 = 19$

	0.05
19	1.729

CamScanner

ritaj.birzeit.edu

19 | 1.729

A 90% C.I:

$$(\bar{X}-E\ ,\ \bar{X}+E)=(1400-100.52,\ 1400+100.52)$$
$$=(1299.48,\ 1500,52)$$

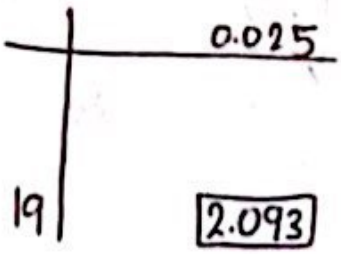
We are 90% confident that the true mean price is between 1299.5 NIS and 1500.5NIS.

b) Construct a 95% confidence interval for the true mean price, then interpret your answer.

$n=20$, $\bar{X}=1400$, $S=260$

→ significance level "α" $= 5\% = 0.05$

∴ Margin of error "E" $= t_{\frac{\alpha}{2},df} \cdot SE$



$df = n-1 = 19$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

$\therefore t_{\frac{\alpha}{2},df} = 2.093$

$$E = 2.093\left(\frac{S}{\sqrt{n}}\right)$$
$$= 2.093\left(\frac{260}{\sqrt{20}}\right) = 121.68$$

ritaj.birzeit.edu

A 95% C.I is: $(\bar{x}-E, \bar{x}+E)$

$= (1400-121.68, 1400+121.68)$

$= (1278.32, 1521.68)$

We are 95% confident that the true mean is between 1278.32 and 1521.68 .

Ex: The following are the weekly time (hours) spent on Pubg of a sample of BZU students.

45, 62, 20, 48, 55, 0, 38, 80

Find a 99% CI for the population mean time, then interpret your result.

$\longrightarrow n = 8$, $\bar{x} = 43.5$, $s = 24.8$

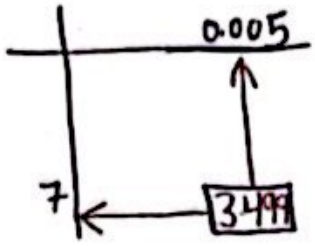
$\bar{x}$: نجدها من الآلة الحاسبة

s: نجدها من الآلة

$\alpha = 1\% = \frac{1}{100} = 0.01 \longrightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

$df = n-1 = 8-1 = 7$

$\therefore t_{\frac{\alpha}{2}, df} = 3.499$



$\longrightarrow$ margin of error "E" $= t_{\frac{\alpha}{2}, df} \cdot SE$

$= 3.499 \left(\frac{s}{\sqrt{n}}\right)$

→ margin of error "E" $= t_{\frac{\alpha}{2}, df} \cdot SE$

$= 3.499\left(\frac{S}{\sqrt{n}}\right)$

$= 3.499\left(\frac{24.8}{\sqrt{8}}\right)$

$= 30.68$

→ A 99% C.I: $(\bar{x} - E, \bar{x} + E)$

$= (43.5 - 30.68, 43.5 + 30.68)$

$= (12.82, 74.18)$

We are 99% confident that the true mean is between 12.82 and 74.18.

* Remarks:

(1) When n increases, the CI becomes shorter.
عندما نزيد حجم العينة فتصبح فترة الثقة اصغر.

(2) When confidence level increases, the CI becomes longer.
كلما زاد مستوى الثقة فتكون فترة الثقة أطول.

(3) When significance level increases, the CI becomes shorter.
كلما زاد مستوى الدلالة فتكون فترة الثقة أقصر

(4) When margin of error increases, the CI becomes longer.
كلما زاد هامش الخطأ فتكون فترة الثقة أطول

Ex: If the confidence interval of estimation of a population mean is (40, 48).

في هذا السؤال كانت فترة الثقة معروفة

a) Find the sample mean.

$$\rightarrow \overline{X} = \frac{40+48}{2} = 44$$

لإيجاد الوسط الحسابي للعينة نأخذ متوسط أطراف الفترة

b) Find the margin of error.

$$\overline{X} - E = 40$$
$$44 - E = 40 \rightarrow \therefore \boxed{E = 4}$$

*** Determining the sample size:**

To find the minimum sample size needed for a C.I to be within a specific margin of error E, we use the following formula:

$$\rightarrow n' = \left(\frac{t_{\frac{\alpha}{2},df} \cdot S}{E}\right)^2$$

; S: sample standard deviation. (previous study).

E: margin of error.

df: ∞ "inf"

لتحديد حجم العينة التي يجب أخذها للحصول على ما مش حظا معين

ritaj.birzeit.edu

E: margin of error.

df: ∞ infi

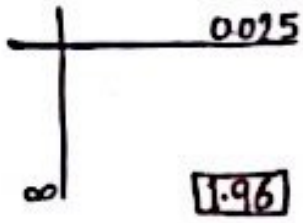
لتحديد حجم العينة اللي يجب أخذها للحصول على هامش خطأ معين

Ex: What **sample size** needed to get a 95% CI within 20 centimeters for the mean students height. Assume $S = 50$.

$$\rightarrow n = \left(\frac{t_{\frac{\alpha}{2},df} \cdot S}{E}\right)^2$$

$$\alpha = 5\% = \frac{5}{100} = 0.05 \rightarrow \therefore \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

$df = \infty$



$S = 50$, $E = 20$

$$\therefore n = \left(\frac{1.96(50)}{20}\right)^2 = 24.01 \rightarrow 25$$

نزد للعدد اللي بعده بغض النظر عن الأعشار

We need a sample of 25 students or more

*Ch12: Hypothesis testing. اختبار الفرضيات

• In this chapter, we will use the sample to test hypotheses about the population mean μ.
سنستخدم عينة لدراسة فرضية معينة عن الوسط الحسابي للمجتمع μ.

• Hypothesis: A claim or conjecture about the parameter μ.
الفرضية: ادعاء أو تخمين عن الوسط الحسابي للمجتمع.

*In any hypothesis testing, there are 2 opposite hypotheses:
في اي اختبار فرضيات يكون لدينا فرضيتين متعاكستين

① The null hypothesis: H_0 الفرضية الصفرية

② The alternative hypothesis: H_1 الفرضية البديلة

*Three types of hypotheses tests:- 3 أنواع من اختبار الفرضيات

① Upper-tailed test:
$H_0: \mu \leq \mu_0$
$H_1: \mu > \mu_0$
دائماً المساواة تكون مع الفرضية الصفرية (H_0) والفرضية البديلة تكون العكس.

② Lower-tailed test:
$H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$

③ Two-tailed test:
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

③ Two-tailed test:

$H_0: M = M_0$

$H_1: M \neq M_0$

- Ex: The mean birth weight of babies is 5.6 Kg. A doctor thinks that using vitamin pills will increase the mean birth weight. ($M > 5.6$)

→ $H_0: M \leq 5.6$ vs $H_1: M > 5.6$ «Upper-tailed test»

- Ex: The mean price of apartments in Ramalloh is \$1000 Due to inflation, a researcher believes that the mean price will decrease. ($M < 100000$.)

→ $H_0: M \geq 100000$ vs $H_1: M < 100000$.

«lower tailed test»

- Ex: The mean grades of BZU students in PHED120 is 72. I claim that playing classical music during the exames will change the mean grades. ($M \neq 72$)

→ $H_0: M = 72$ vs $M \neq 72$

«2-tailed test»

* Test Conclusion (Decision). قرار الاختبار

Using the sample, Our test conclusion is either reject or not reject (accept) H_0.

باستخدام العينة، يكون قرارنا لأي اختبار إما رفض H_0 أو قبولها.

→ If H_0 is rejected (رفض H_0), this mean that H_1 is accepted (قبول H_1).

→ If H_0 is accepted (قبول H_0), this mean that H_1 is rejected (رفض H_0).

* Type I and II Errors: الخطأ من النوع الأول والثاني

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct Decision
Accept H_0	Correct Decision	Type II Error

• Type I error: Rejecting H_0 when it is true. رفض H_0 عندما تكون صحيحة

• Type II error: Accepting H_0 when it is false. قبول H_0 عندما تكون خاطئة.

* In hypothesis testing, we use significance level (α).

α is the probability of committing type I error. For example if $\alpha = 10\%$, then there is a 10% chance that we will reject a true H_0.

مستوى الدلالة "α": احتمال حدوث خطأ من النوع الأول، نقوم دائماً قبل عمل أي اختبار بتحديد قيمة α.

*Critical value(s): القيم الحرجة

• **Critical value:** A t-value that is found based on α and the type of the test. قيمة نجدها من الجدول بناء على نوع الاختبار وقيمة α.

① Upper-tailed test:

critical value: t_α ; $df = n-1$.

② Lower-tailed test:

critical value: $-t_\alpha$; $df = n-1$.

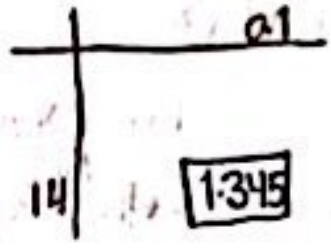
③ Two-tailed test:

critical value(s): $t_{\frac{\alpha}{2}}, -t_{\frac{\alpha}{2}}$; $df = n-1$

- **Ex:** ① If we have an upper tailed test and $\alpha = 0.$

$n = 15$.

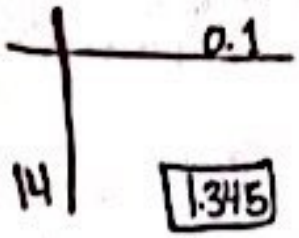
→ Critical value: $t_{\alpha.} = t_{0.1}$; $df = n-1 = 14$



$\therefore t_\alpha = 1.345$

② lower-tailed test and $\alpha = 0.1$, $n = 15$.

→ Critical value: $-t_\alpha = -t_{0.1}$; $df = n-1 = 14$

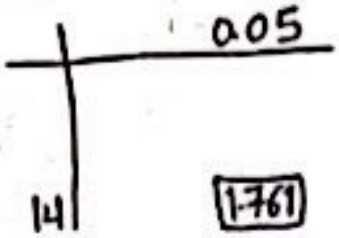


$\therefore -t_\alpha = -1.345$

③ Two-tailed test and $\alpha = 0.1$, $n = 15$.

→ Critical value: $t_{\frac{\alpha}{2}}$, $-t_{\frac{\alpha}{2}}$; $df = n-1 = 14$

$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$



$\therefore$ critical values = 1.761 and -1.761

***Test Statistic:** قيمة t المحسوبة

Using the sample, we compute a special t-value called t_{test}, using the following formula:- باستخدام العينة نجد قيمة t المحسوبة

$$t_{test} = \frac{\bar{X} - M_0}{SE}$$

$\bar{X}$: sample mean الوسط الحسابي للعينة

M_0: the current population mean. القيمة المفروضة في المجتمع

$SE = \frac{S}{\sqrt{n}}$; S: standard deviation

n: sample size

*Rejection and Acceptance Regions of H_0:- منطقة الرفض والقبول لـ H_0

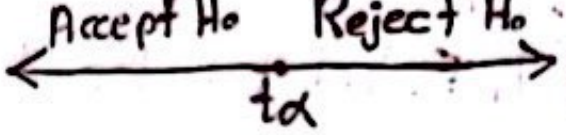
To reject or accept H_0, we compare the t_{test} with the critical value(s):

حتى نستطيع الرفض أو القبول لـ H_0 نقارن قيمة t_{test} مع القيمة الحرجة

① Upper-tailed test:

Reject H_0 if $t_{test} \geq t_\alpha$

Accept H_0 if $t_{test} < t_\alpha$

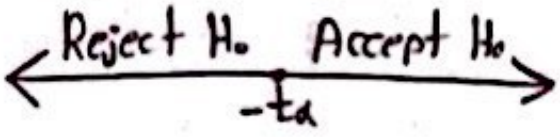


② lower-tailed test:

نعين t_α على الخط وبعدها نعين t_{test} إذا كان في منطقة القبول نقبل H_0 وإذا كان في منطقة الرفض نرفض H_0.

Reject H_0 if $t_{test} \leq -t_\alpha$

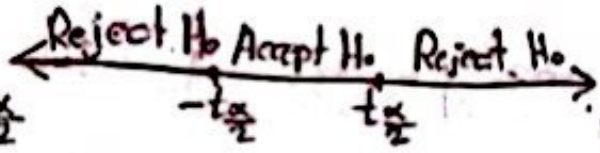
Accept H_0 if $t_{test} > -t_\alpha$



③ 2-tailed test:

Reject H_0: $t_{test} \leq -t_{\frac{\alpha}{2}}$ or t_{test}

Accept H_0: $-t_{\frac{\alpha}{2}} < t_{test} < t_{\frac{\alpha}{2}}$



*Hypothesis testing procedure :-

① Set up the hypotheses H_0 and H_1. نكتب H_0 ، H_1

② Determine the type of the test: (Upper, lower, 2-tailed test) نحدد نوع الاختبار

③ Find the test statistic: ($t_{test} = \frac{\bar{X} - \mu_0}{SE}$; $SE = \frac{S}{\sqrt{n}}$). نجد قيمة t_{test}

④ Find the critical value(s). نجد القيمة الحرجة بناءً على نوع الاختبار

Upper-tailed test: t_α , lower-tailed: $-t_\alpha$,

2-tailed test: $-t_{\frac{\alpha}{2}}$, $t_{\frac{\alpha}{2}}$.

⑤ Determin the rejection regions نحدد منطقة الرفض والقبول

⑥ Make your conclusion. نكتب القرار

Ex: The mean pulse rate of people per minute is 82. A researcher claims that getting married will increase the mean pulse rate. A sample of 20 people was taken. The sample had mean of 86 and standard deviation of 6. Test the claim using $\alpha = 0.01$.

$$M_0 = 82, n = 20, \bar{X} = 86, S = 6, \alpha = 0.01$$

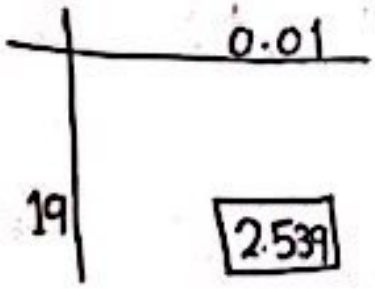
$$H_0: M \leq 82 \quad \text{vs} \quad H_1: M > 82$$

« Upper-tailed test ».

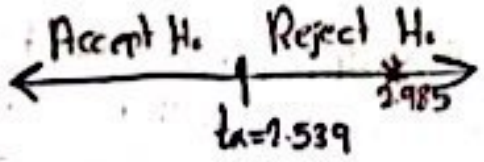
$$t_{test} = \frac{\bar{X} - M_0}{SE} \; ; \; SE = \frac{S}{\sqrt{n}} = \frac{6}{\sqrt{20}} = 1.34$$

$$= \frac{86 - 82}{1.34} = 2.985$$

critical value: t_α ; $df = n-1 = 19$



∴ critical value $= t_\alpha = 2.539$



∴ Conclusion: Reject H_0, Accept H_1.

- Ex: The mean time of washing cars in a car wash is 20 minutes. A client thinks buying new equipments will reduce the mean time. A sample of 25 cars was selected. If $\bar{X}=18$ and $S=8$, test the clients claim using a significance level of 10%.

$\mu_0 = 20$, $n=25$, $\bar{X}=18$, $S=8$, $\alpha = \frac{10}{100} = 0.1$

→ $H_0: \mu \geq 20$ vs $H_1: \mu < 20$

a lower-tailed test

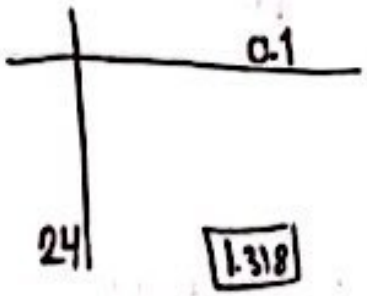
$$t_{test} = \frac{\bar{X} - \mu_0}{SE} \; ; \; SE = \frac{S}{\sqrt{n}} = \frac{8}{\sqrt{25}} = 1.6$$

$$= \frac{18-20}{1.6} = -1.25$$

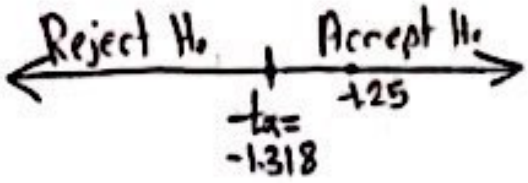
---

critical value: $-t_\alpha$; $df = n-1 = 24$

Critical value: $-t_\alpha$; $df = n-1 = 24$



$\therefore$ critical value $= -t_\alpha = -1.318$



$\therefore$ Conclusion: Don't reject H_0, (Client's claim is wrong).

Ex: The mean of grades in Stat 2371 is 70. Some teacher is convinced that adopting online exams will change this mean. To test the teacher's belief, 10 students were selected randomly. Their grades were 55, 90, 78, 80, 68, 70, 100, 92, 40, 52. Test using $\alpha = 0.05$.

$M_0 = 70$, $n = 10$, $\alpha = 0.05$

Using calculator, mode 2 55 M+ 90 M+ .... 52 M+

- shift 2 1 = $\quad \therefore \bar{X} = 72.5$
- shift 2 3 = $\quad \therefore S = 19.3$

$H_0: M = 70$ vs $H_1: M \neq 70$

«2-tailed test»

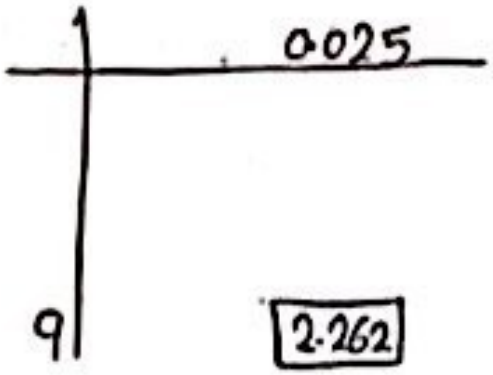
$$t_{test} = \frac{\bar{X} - M_0}{SE} \; ; \; SE = \frac{S}{\sqrt{n}} = \frac{19.3}{\sqrt{10}} = 6.1$$

$$t_{test} = \frac{\bar{X} - M_0}{SE} \; ; \; SE = \frac{s}{\sqrt{n}} = \frac{19.3}{\sqrt{10}} = 6.1$$

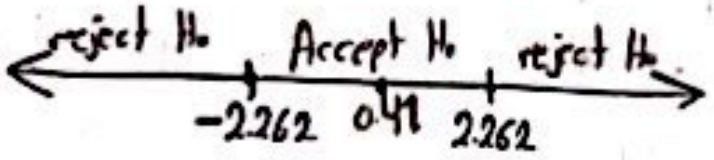
$$= \frac{72.5 - 70}{6.1} = 0.41$$

critical value(s): $t_{\frac{\alpha}{2}}, -t_{\frac{\alpha}{2}}$; $df = n - 1 = 9$.

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$



∴ critical values: $-2.262, 2.262$



Conclusion: Accept H_0. (Teacher's belief is wrong).

*ch 18: Introduction to regression analysis
مقدمة في الانتشار الخطي

.In this chapter, we will analyze simple linear relationships between two variables X and Y.
سندرس العلاقة بين متغيرين

* Independent variable X: predicts or changes the other variable Y. المتغير المستقل: هو المتغير الذي يؤثر على الآخر ونرمز له X.

* Dependent variable Y: predicted or changed by X. المتغير التابع: هو المتغير الذي يتأثر بالمتغير الأول ونرمز له Y.

- Ex: Let:

1) X: # of absence. عدد الغيابات
Y: final grade. العلامة

2) X: Amound of advertising. الإعلانات
Y: Volume of Sales. حجم المبيعات

3) X: Age العمر
Y: Blood pressure ضغط الدم

* Types of relationships: أنواع العلاقات

① Positive relationship: when X increases, Y increases.
العلاقة الطردية (الموجبة): كلما زاد X يزداد Y وبالعكس.

② Negative relationship: when X increases, Y decreases.
العلاقة العكسية: كلما زاد X، يقل Y وبالعكس.

③ No relationship: X and Y are not related.
لا يوجد علاقة بين X و Y.

Scatter diagram: A graph used to determine the type of relation ship between the variables X and Y by plotting the ordered pairs (X_i, Y_i).
المخطط التبعثري: يمثل X مع Y لأي متغيرين.

-Ex: The following represents the # of absences X and final gracles Y of a sample of students.

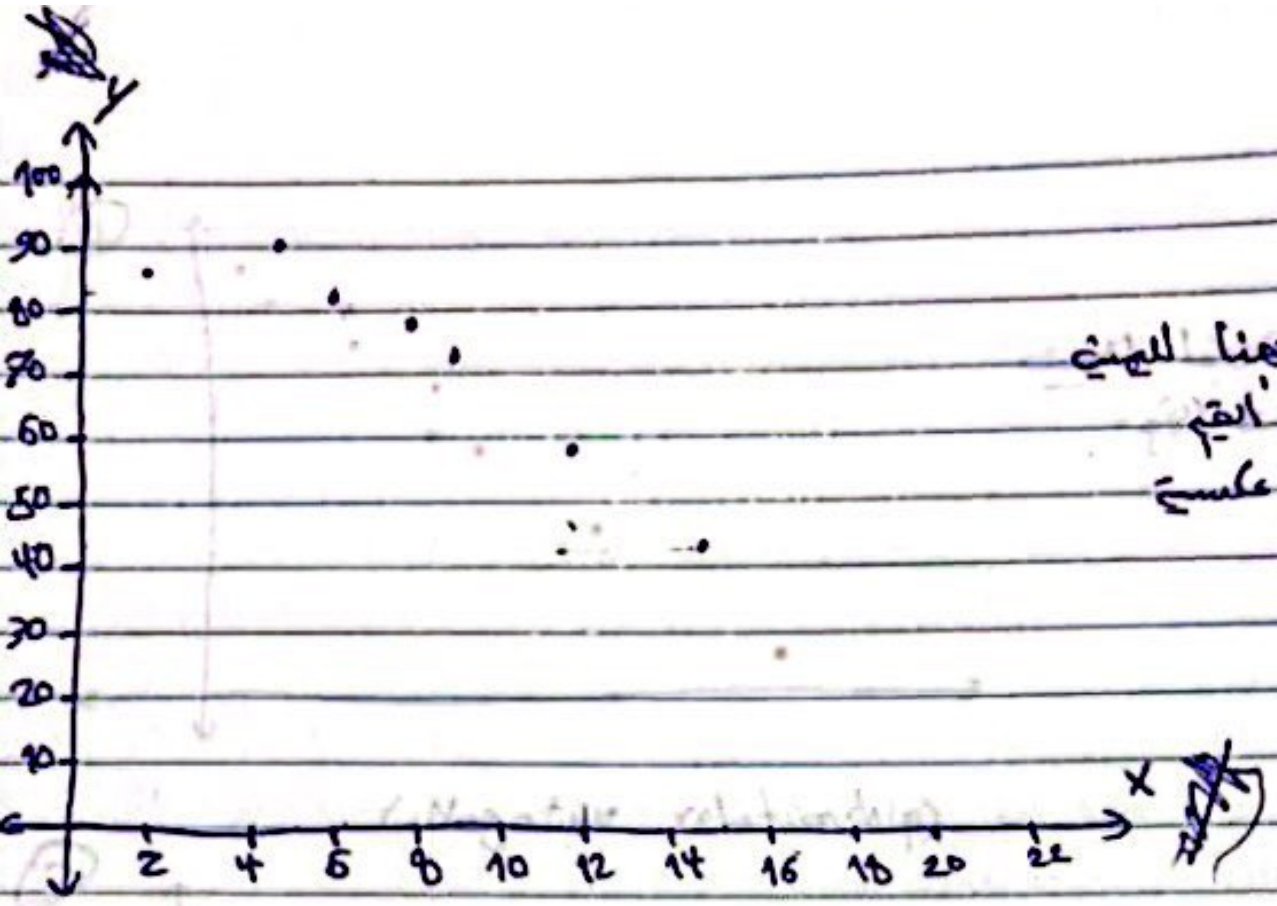
.Draw the scatter diagram.

X: عدد الغيابات
Y: العلامة

X_i	Y_i
6	82
2	86
15	43
9	74
12	58
5	90
8	48

هذه قيم X و Y
حيث أن قيم
X تمثل عدد الغيابات
و Y تمثل علامة الطالب.
" وهذا الجدول يكون معطى "

scatter diagram
~~scatterdiagram~~
المطلوب: رسم

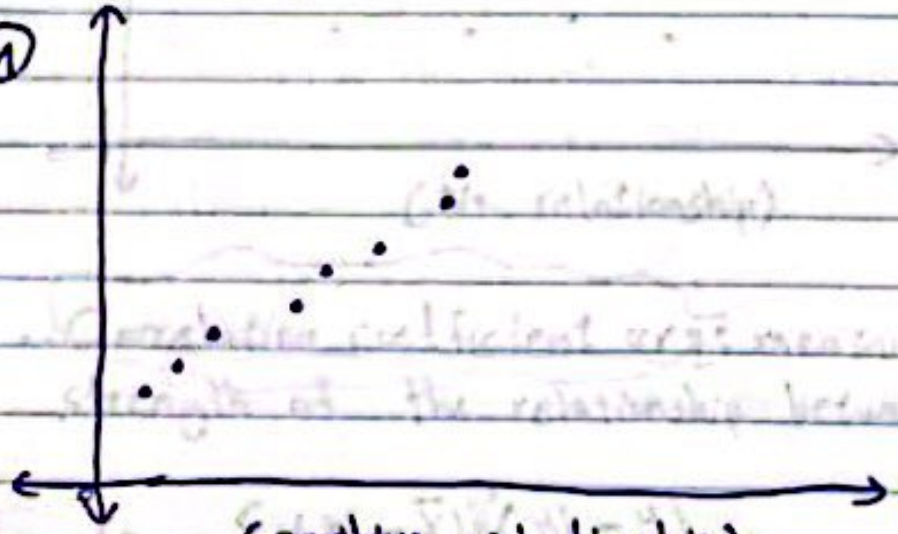


كلما اتجهنا لليمين
تتناقص القيم
← العلاقة عكسية

(Negative relationship)

-Ex! Determine the type of the relationship from the scatter diagram.. حدد نوع العلاقة من خلال الرسم

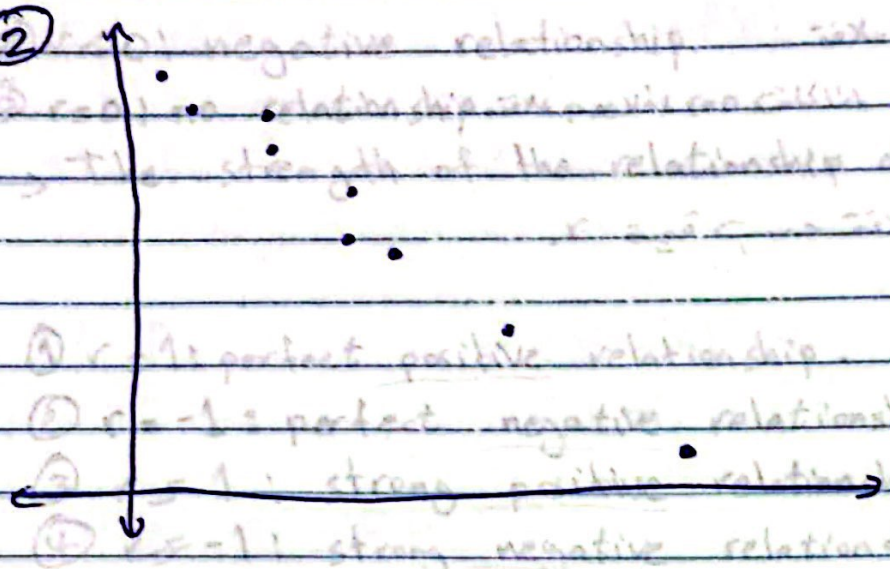
①



كلما اتجهنا لليمين
تزداد القيم.
← العلاقة طردية.

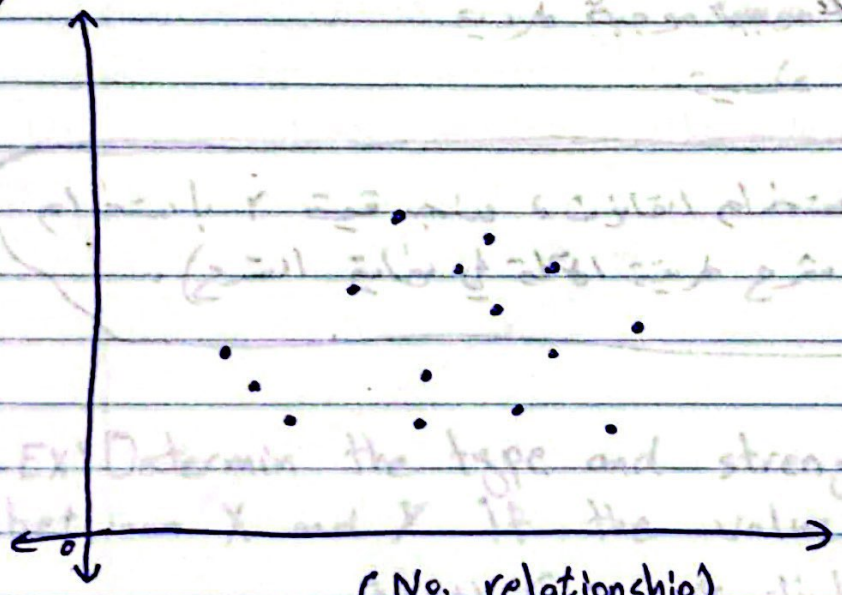
(postive relationship).

② كلما اتجهنا للمين تتناقص القيم.



(Negative relationship)

③ لا يوجد علاقة



(No. relationship)

- Correlation coefficient (r): measures the type and strength of the relationship between X and Y.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

معامل الارتباط: يقيس مدى قوة ونوع العلاقة بين X و Y.

→ The value of r is: $-1 \leq r \leq 1$ قيمة r دائماً محصورة بين -1 و 1.

① $r > 0$: positive relationship. اذا كانت r+ تكون العلاقة طردية.

② $r < 0$: negative relationship. إذا كانت سالبة تكون علاقة عكسية.
③ $r = 0$: no relationship. إذا كانت r=0 تكون العلاقة معدومة
→ The strength of the relationship depends on r:
تحدد r أيضاً قوة العلاقة حسب قيمة r.

① $r = 1$: perfect positive relationship. علاقة طردية تامة
② $r = -1$: perfect negative relationship. علاقة عكسية تامة
③ $r \approx 1$: strong positive relationship. علاقة طردية قوية
④ $r \approx -1$: strong negative relationship. علاقة عكسية قوية
⑤ $r \approx 0$: weak positive/negative relationship.
علاقة ضعيفة إذا موجبة طردية
وإذا سالبة عكسية.

*ملاحظة كاريه الحل باستخدام القانون، سنجد قيمة r باستخدام الآلة الحاسبة. (سيتم شرح طريقة الآلة في نهاية الشرح).

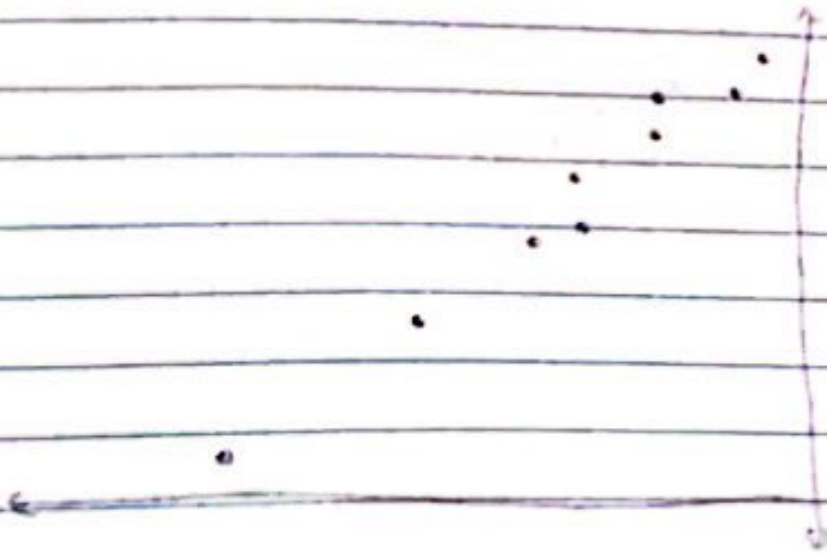
-Ex: Determin the type and strength of the relationship between X and Y if the value of r equals:
حدد نوع العلاقة وقوتها بناء على قيمة r المعطاة.

① $r = 0.99$
→ strong positive relationship.

② $r = -0.96$
→ strong negative relationship.

③ $r = -0.16$
→ weak negative relationship.

④ $r = 0$
→ No relationship.



(positive correlation)

$r > 0$ علاقة طردية

$r < 0$ علاقة عكسية

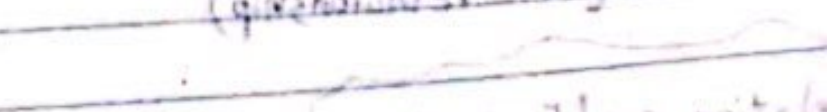
$r = 0$ لا توجد علاقة

إذا كانت قريبة من 100

strong علاقة

إذا كانت بعيدة عن 100 أو قريبة من 0

weak علاقة



(no relationship)

Correlation coefficient is a measure the type and strength of the relationship between X and Y.

$$\frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

* Regression line: $y = A + BX$ معادلة الانتشار

The best line fits the sample points of X and Y.
لإيجاد معادلة أفضل خط يلائم النقاط نحتاج قيمتي A و B:

$$B = \text{slope} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

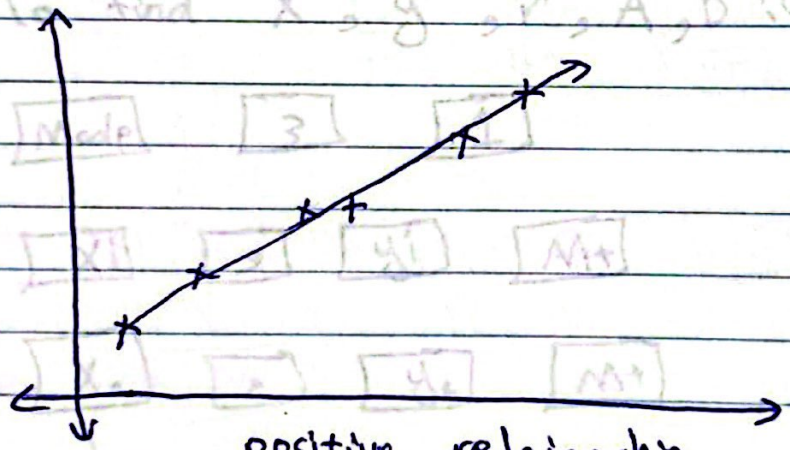
$\bar{y}$: mean of y

$\bar{x}$: mean of X.

$$A = \bar{y} - B\bar{x}$$

لن نستخدم القوانين بسبب باستخدام الآلة الحاسبة.

- If the relationship is positive then $B > 0$.

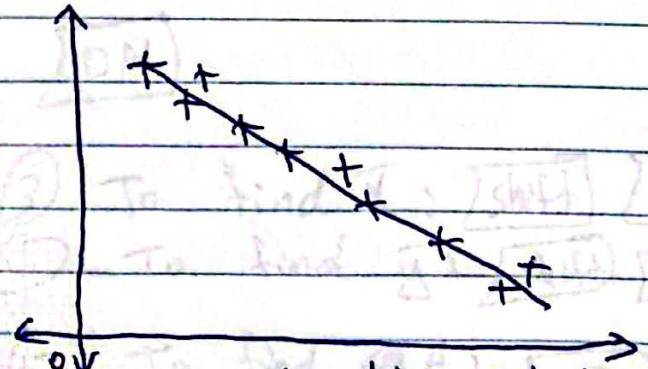


positive relaionship
«B>0»

* نلاحظ أن إشارة r و B تكون نفس الإشارة + أو -

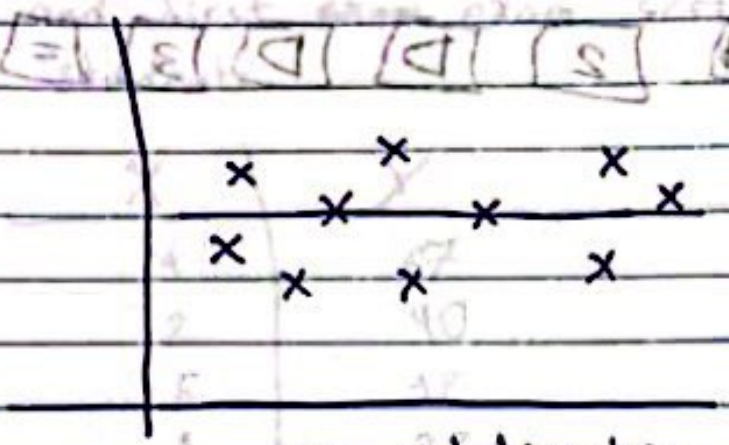
- If the relationship is negative then $B < 0$.

إذا كانت العلاقة عكسية فإن قيمة B -.



Negative relationship
«B<0»

If there is no relationship then B = 0
إذا كانت B=0 لا يوجد علاقة وانكاس صحيح.



no relationship
(B = 0)

* Using Calculator : باستخدام الآلة الحاسبة

To find $\bar{X}$, $\bar{y}$, r , A , B : (لإيجاد قيمة B, A, r نتبع الخطوات.)

① Mode 3 1

② Xi , yi , M+

X_2 , y_2 , M+

X_n , y_n , M+

ON

X	Y
X_1	y_1
X_2	y_2
⋮	⋮
X_n	y_n

نفرض كان لدينا هذه القيم.

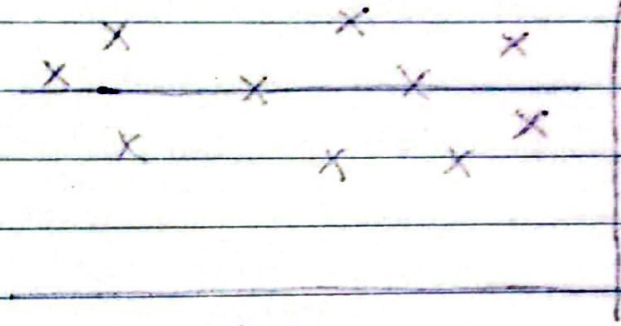
③ To find $\bar{X}$: shift 2 1 =

④ To find $\bar{y}$: shift 2 السهم لليمين D =

⑤ To find A : shift 2 D D 1 =

(6) To find θ = [shift] [2] [D] [D] [2] [=]

(7) To find r: [shift] [2] [D] [D] [3] [=]



Ex: The following represents the hours of study X and first exam scores Y of a sample of students.

find the correlation Coefficient

X	Y
4	67
2	40
6	92
1	32
0	20
8	100

→ Mode 3 1

4 , 1 67 M+

2 1 40 M+

⋮

8 1 100 M+

ON

→ To find r: shift 2 3 =

∴ r = 0.99 shift 2 ▷ ▷ 3 =

-EX: The following represents the population (millions) X and # cafes Y of a sample of cities.

• find the regression line.

X	Y
20	70
15	65
30	75
10	40
50	80

→ To find A, B:

Mode 3 1

20 , 70 M+
15 , 65 M+
30 , 75 M+
10 , 40 M+
50 , 80 M+

ON

A: shift 2 ▷ ▷ 1 = 46.625
B: shift 2 ▷ ▷ 2 = 0.775

→ Regression line: $y = A + Bx$

$= 46.625 + 0.775x$

- EX: The following represents the price (NIS) X and weight (grams) Y of a sample of iphones.

X	Y
800	350
1000	200
500	400
1500	120
900	220

① find the correlation coefficient.

Mode 3 1

800 2 350 M+
1000 2 200 M+
500 2 400 M+
1500 2 120 M+
900 2 220 M+

ON

→ r: shift 2 D D 3 = -0.925

② Determine the type and strength of the relationship between price and wight.

→ strong negative relationship.