Variance and StDev

By M.Aek Progs AD

https://www.facebook.com/M.Aek.Progs.Angedevil.AD/

https://www.youtube.com/channel/UC6AhJIORlsp56XDwqfNSTsg

https://github.com/Angedevil-AD

Variance:

In probability theory and statistics, variance is the expectation of the squared deviation of a random variable from its population mean or sample mean. Variance is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value

the equation of variance is given by :

$$\operatorname{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

where:
x: observation
u: mean
n: size of observation

exemple:

set of observation A:
(5,3,7,8,2,4)
set of observation B:
(3,18,10,45,100,80)

mean of A:
u=4.8

Variance of A:
$Var(X) =((5-4.8)^2+(3-4.8)^2+(7-4.8)^2+(8-4.8)^2+(2-4.8)^2+(4-4.8)^2 \ ) / 6$
$Var(X)= (0{,}04+3{,}24+4{,}84+10{,}24+7{,}84+0{,}64) / 6$
Var(X) = 4.47

mean of B:

u=42.66

Variance of B:

$$Var(X) = ((3-42.66)^2+(18-42.66)^2+(10-42.66)^2+(45-42.66)^2+(100-42.66)^2+(80-42.66)^2) / 6$$

$$Var(X) = (1572,91+608,11+1066,67+5,47+3287,875+1394,27)/6$$

$$Var(X) = 1322,55$$

the variance of B > variance of A

thats means the set of observation B is dispersed unlike A.

if you see (B):

(3,18,10,45,100,80)

from 3 to 100 ...and 18 <--> 80 ...also 45<-->80

there is a large spacing in the set ...therfore the variance gives a big number

unlike A:

(5,3,7,8,2,4)

from 5 to 3 .or 3 to 7 ..or 2 to 8......is not too much differance ...

so, when the variance is large, the data is widely scattered

Standard deviation (StDev):

In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

Standard deviation may be abbreviated SD, and is most commonly represented in mathematical texts and equations by the lower case Greek letter sigma σ, for the population standard deviation, or the Latin letter s, for the sample standard deviation.

The stansard deviation = square root of the variation

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

lets asume ; you are a teacher, and you want to evaluate he level of your students: number of students 7

score of every student: 10,8,12,14,15,18,16

lets calc the mean:
u = (10+8+12+14+15+18+16) / 7 = 13.28
clac variance:
Var(X) = ($(10-13.28)^2 + (8-13.28)^2 +(12-13.28)^2 +(14-13.28)^2 +(15-13.28)^2$
$+(18-13.28)^2 +(16-13.28)^2$) / 6
Var(X)=10,7584+ (27,8784+ 1,6384+ 0,5184+ 2,9584+ 7,3984) / 6
Var(X) = 5.76
Stdev = square root (5.76)
Stdev = 2.4

we have :
u (mean) = 13.28

u + Sigma = 13.28 + 2.4 = 15,68
u - Sigma = 13.28 - 2.4 = 10,88

we deduce that all students with medium level their scores are between the range :
10.88 and 15.68
all student with score above 15.68.. they are excellent..
and all student with scores under 10,88 they are weak

if you ask me ! what if the mean = 5 and stdev is 2 ..it will be
range = 3 <-- to --> 7
so we cant say that scores above 7 ..excellent
or scores inside the range are medium because .....8 or 9 = bad score!!!!
its wrong!
the method doesnt work like that! its show you the level of student.. and compare them to the rest ..

another ex:

lets supose that A is group of prices of food or cloths or anything ......
low prices means bad quality where hight price means good quality
and medium price is reasonable..

lets supose that the mean of group A u= 80 and the Stdev = 20

we conclude that all price above (u+stdev) = 100 are good quality
and all price under (u-stdev) = 60 are bad quality
where all price between 60 and 100 are resonable quality

if you have 90$, you avoid the bad quality because 90 > 60 ..
but you cant buy somthing >(u+stdev) > 100
only what you can to buy is everything between 60 and 100 ... (resonable quality)

the previous equation is about the entire population.

another equation to calcul variation for a sample is given by

$$s^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

such like the population eq but this time you devide the sum by (N-1).
X: observation
X bar : mean

---

lets suppose that the mean of the student = 11,125 and the StDev 3.95

scores = 9 , 17 ,7, 14 ,5 , 10 ,13 , 12 , 16 ,14 , 16 ,8 ,10 ,3 ,10 ,14

question 1:
whats is the propability of student with score <9.0
question2:
if we select 7 student randomly , what is the propability of score greather than 14.0
in this case central limit theorem CLT & cumulative destibution

Central limit theorem:

$$Z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right).$$

answer 1:
Z = sqrt(1) . (9 - 11.125) / 3.95 = -0.53

P(Z) = P(mean + (Sigma*Z)) = P(11.125+(3.95*-0.53)) = 9,0315
P(Z) = 0.29 (29%)

answer 2:
Z = sqrt(7) . ((14 - 11.125) / 3.95) =0.727*2.64 = 1.92
P(Z) = P(mean + (Sigma*Z)) = P(11.125+(3.95*1.53)) = 17.16
P(Z) = 0.06 (6%)

question3:
if we select 10 student randomly , what is the propability of score greather than 12.0

answer 3:
$Z = sqrt(10) . ((12 - 11.125) / 3.95) =3.162*0,22= 0.700$
$P(Z) = P(mean + (Sigma*Z)) = P(11.125+(3.95*0.700)) =13.89$
$P(Z) = 0.24$ (24%)

suppose, that you want to evaluate 2 class ,
you make foe exemple 3 defferent score, (less than 10) between (10-15) ..and greather than 15
you take 8 samples

class1 :
7,10,14,18,3,13,16,3,14,9
14,8,13,13,5,17,10,5,12,10

class2:
12,14,3,5,9,17,20,13,14,3
4,12,15,8,10,7,10,15,16,12

lets begin with class1:

$u=10.7$
$stdev=4.36$
less than 10.00: -------------------------------------

$Z = sqrt(8) . (10-10.7)/4.36 = -0.454$
$p(<10) = (10.7+ (4.36*-0.454)) = 12.67$
$p(<10) = 0.3248$ (32.48%)

between 10.00 & 15.00 : -------------------------------------

$p(>10) = 1-0.3248 = 0,6752$

$Z = sqrt(8) . (15-10.7)/4.36 = 2.77$
$p(<15) = (10.7+ (4.36*2.77)) = 22.77$
$p(<15) = 0.9972 =$ (99.72%)

$P(10<x<15) = 0.99 - 0.32 = 0.3222$ (32.22%)

greather than 15.00: ------------------------------------

Z = sqrt(8) . (15-10.7)/4.36 = 2.77
p(>15) = (10.7+ (4.36*2.77)) = 22.77

p(>15) = 0.0028 (0.2 %)

class2:

u=10.95
stdev=4.70

less than 10.00: ------------------------------------

Z = sqrt(8) . (10-10.95)/4.70 = -0.57
p(<10) = 0.2840 (28.40%)

between 10.00 & 15.00 : ------------------------------------

p(>10) = 1-0.2840 = 0.715

Z = sqrt(8) . (15-10.95)/4.70 = 2.43
p(<15) = (10.95+ (4.70*2.43)) = 22.37
p(<15) = 0.9924 = (99.24%)

P(10<x<15) = 0.9924 - 0.715 = 0.2765 (27.65%)

greather than 15.00: ------------------------------------

Z = sqrt(8) . (15-10.95)/4.70 = 2.43
p(>15) = (10.95+ (4.70*2.43)) = 22.37
p(>15) = 0.0076 (0.7%)

result class1:
p(<10) = 0.32(32.48%)
P(10<x<15) = 0.3222 (32.22%)
p(>15) = 0.0028 (0.2 %)

result class2:
p(<10) = 0.28 (28.40%)
P(10<x<15)= 0.27 (27.65%)
p(>15) = 0.007 = (0.7%)