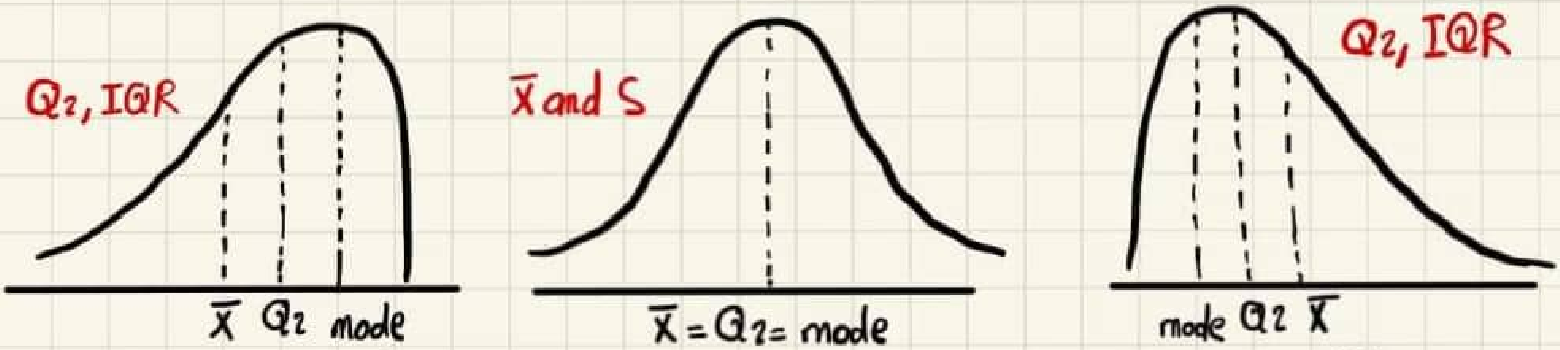
Statistics, lecture 8:

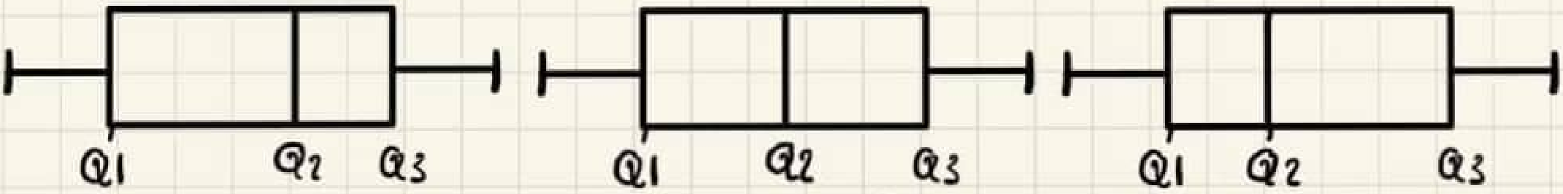
*Skewness:



$\bar{x} < Q_2 < \text{mode}$ — skewed to the left

Symmetric

$\text{mode} < Q_2 < \bar{x}$ — skewed to the right



$Q_2 - Q_1 > Q_3 - Q_2$

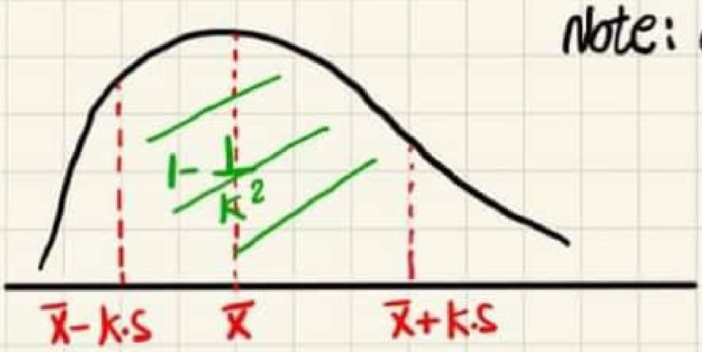
$Q_2 - Q_1 = Q_3 - Q_2$

$Q_2 - Q_1 < Q_3 - Q_2$

* Chebyshev's Inequality:

For $k > 1$ and for any collection of data there are at least $1 - \frac{1}{k^2}$ of the observations lie in $(\bar{x} - k \cdot S, \bar{x} + k \cdot S)$

Note: at most $\frac{1}{k^2}$ lie outside $(\bar{x} - k \cdot S, \bar{x} + k \cdot S)$



e.g) for a collection of 500 observations, we have:
$\bar{X} = 50$, $S = 5$, Find:
1) the interval whose center is 50 and contains at least 450 of the observations

b) The interval whose center is 50 and at most 125 of the observations lie outside it

c) Find the no. of observations that at least in (35, 65)

d) the no. of observations that are at most outside (30, 70)

sol) $\bar{X} = 50$, $S = 5$, $n = 500$

a) $\frac{450}{500} = 0.90$

$1 - \frac{1}{k^2} = 0.90$

$\bar{X} - ks$ ———— $\bar{X} + k.s$

$1 - \frac{1}{k^2} = 0.90 \Rightarrow k^2 = 10 \Rightarrow k = \sqrt{10}$

$k \approx 3.16$

$\bar{X} - ks = 50 - 3.16 \times 5 = 34.2$
$\bar{X} + ks = 50 + 3.16 \times 5 = 65.8$
at least 450 out of 500 observations lie in (34.2, 65.8)

b) $\frac{125}{500} = 0.25 = \frac{1}{4}$

$\frac{1}{k^2} = \frac{1}{4} \Rightarrow k^2 = 4 \Rightarrow k = 2$

$\bar{X} - ks = 50 - 2 \times 5 = 40$
$\bar{X} + ks = 50 + 2 \times 5 = 60$
at most 125 out of 500 observations lie outside (40, 60)

c) $\bar{X} + ks = 65$ or $\bar{X} - ks = 35$
$50 + 5k = 65 \rightarrow 5k = 3$
$k = 3$
at least $1 - \frac{1}{k^2}$ in (35, 65)

$1 - \frac{1}{9} = \frac{8}{9} = 0.889$

$500 \times 0.889 = 444.\bar{4}$
at least 445 lie in (35, 65)

c) $(30, 70) = (\bar{X} - ks, \bar{X} + ks)$

$\bar{X} + ks = 70$
$50 + 5k = 70$
$5k = 20 \therefore k = 4$
at most $\frac{1}{k^2} = \frac{1}{4^2} = \frac{1}{16} = 0.0625$

lie out side (30, 70)
$500 \times \frac{1}{16} = 31.25$

at most 31 observations lie outside (30, 70)

* The Emperical Rule :-

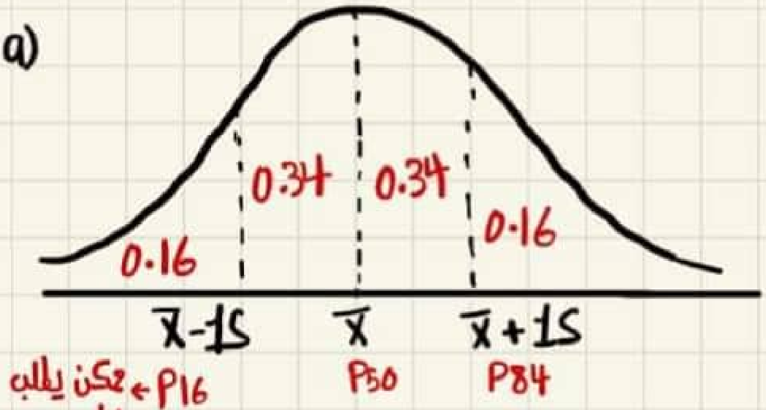
For a bell-shaped frequency graph we have :

a) The percentage of data that lie within 1 standard deviation about the mean is about 68%.

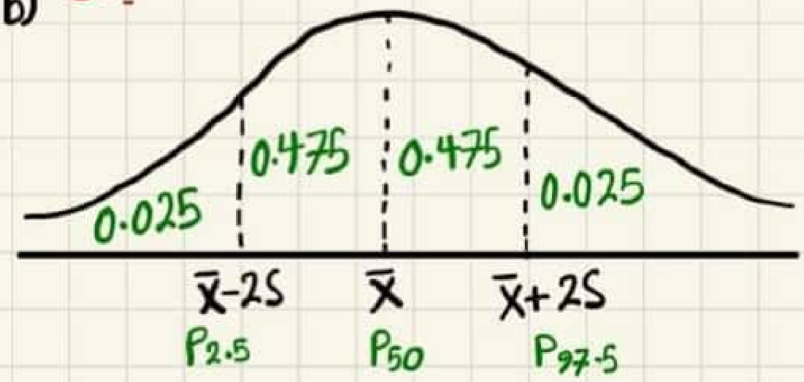
b) The percentage of data that lie within 2 standard deviations of the mean is about 95%.

c) The percentage of data that lie within 3 standard deviations about the mean is about 99%.

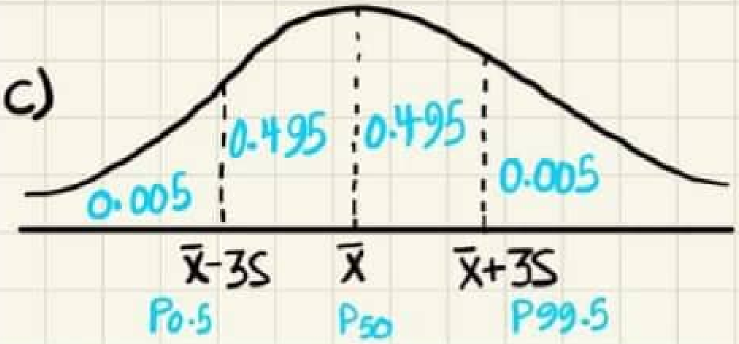
a)



b)



c)



Ex) for a bell-shaped distribution with $\bar{x}=50$ and $S=5$. find:

a) The percentage of data that lie in the interval (40, 60)

b) The percentage of data that lie in the interval (45, 65)

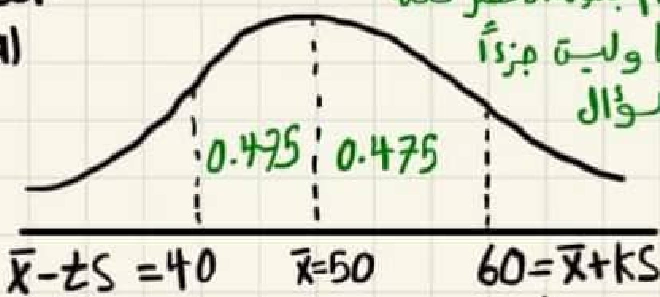
c) The percentage of data that lie in the interval (55, 65)

D) The percentage of data that lie in the interval (35, 45)

Sol:

الارقام باللون الاخضر نحن نكتبها وليست جزءاً من السؤال

a)



ليس شرطاً ان تكون K

$50-5t=40$

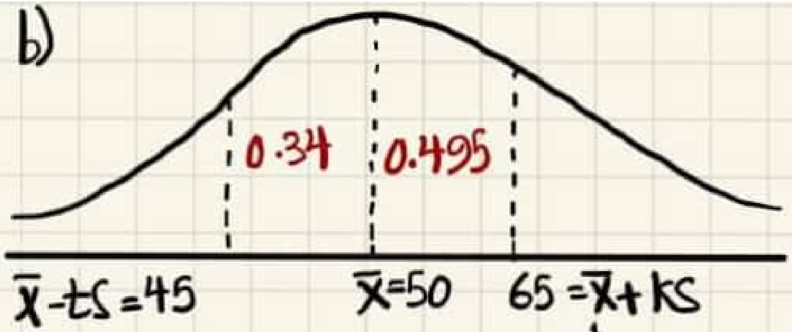
$t=2$

$50+5k=60$

$k=2$

95% lie in the interval (40, 60)

b)



$50-5t=45$

$t=1$

$50+5k=65$

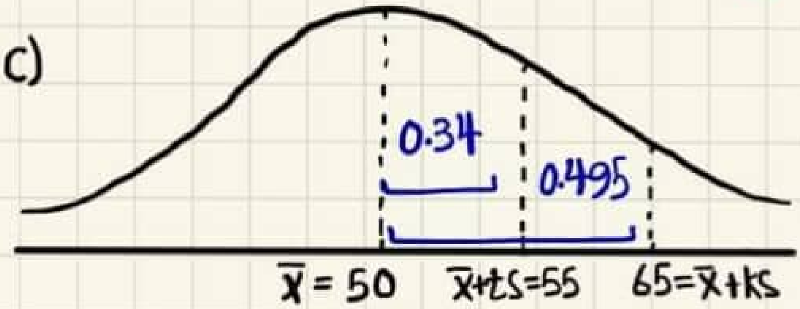
$k=3$

$0.34+0.495=0.835$

$=83.5\%$ in$(45, 46)$

ملاحظة: الارقام باللون الاحمر نحن نكتبها وليست جزءاً من السؤال

c)



$50+5k=65$

$k=3$

$50+5t=55$

$t=1$

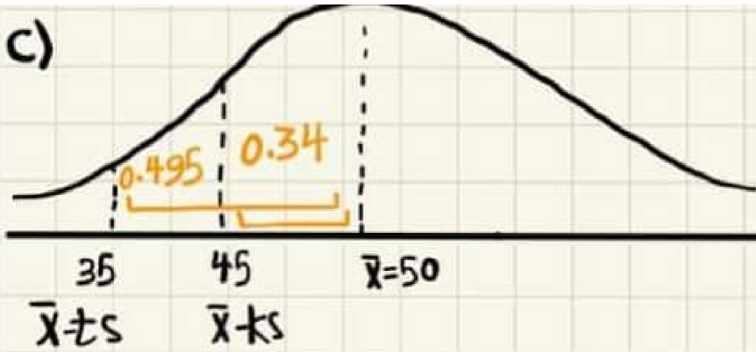
$0.495-0.34=0.155$

15.5% lies in $(55, 65)$

ملاحظة: الارقام باللون الازرق نحن نكتبها و ليست جزءاً من السؤال.

6

c)



$*\ 50-5t=35 \Rightarrow t=3$

$*\ 50-5k=45 \Rightarrow k=1$

$0.495-0.34=0.155 \Rightarrow 15.5\%$ lie in$(35,45)$

ملاحظة: الارقام باللون البرتقالي نحن نكتبها وليست جزءاً من السؤال

7