Comparing two collections.

(1) Z-score. ← Can be skipped.

A z-score describes the position of a raw score in terms of its distance from the mean.

The z-score is positive if the value lies above the mean, and is negative if it lies below the mean.

The z-score corresponding to an observation x from a sample with mean $\bar{x}$ and Std. S is given by $\boxed{Z = \dfrac{x - \bar{x}}{S}}$.

Ex. Consider the following information about the grades in two sections.

	Section I	Section II
Mean	60	70
Std.	3	5
x	65	68

The z-score of grade $x = 65$ in Sec. I is

$$\frac{65-60}{3} = 1.67 > 0 \text{ as 65 above the average.}$$

The z-score of grade $x = 68$ in Sec. II is

$$\frac{68-70}{5} = -0.4 < 0 \text{ as 68 below the average.}$$

(2) Coefficient of variation. ← Can be skipped.

The coefficient of variation (C.V.) is a measure of relative variability, and it is given by the ratio of the standard deviation to the average. That is, $C.V. = \frac{Std.}{\bar{x}} * 100\%$.

Ex: Consider the following information about the grades in two sections.

	Section I	Section II
Mean	60	70
Std.	4.5	5
C.V.	7.5 %	7.14%

$\frac{4.5}{60} * 100\%$

$\frac{5}{70} * 100\%$

Note that the variability of sec. I is more than the variability of sec. II even though the Std. of Sec. I is less than that of Sec. II.

Applications.

Note: Percentage = Ratio * 100%
Ratio = Proportion which is out of 1.

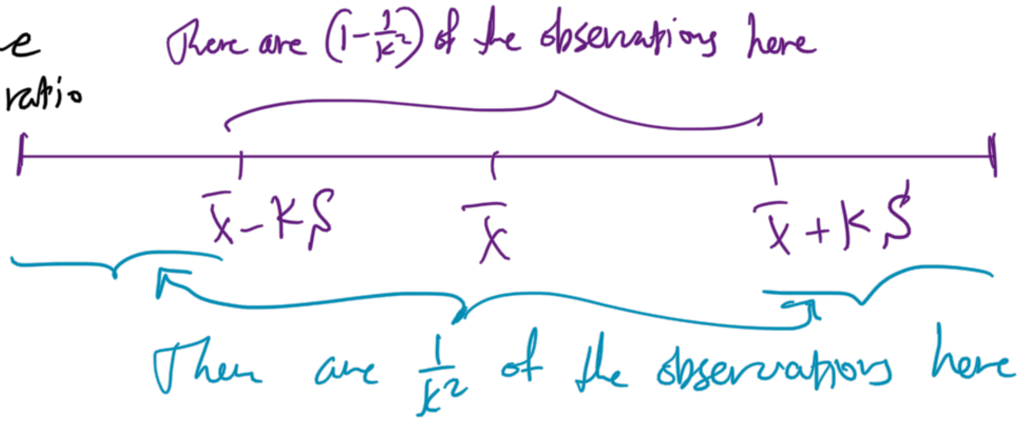
(1) Chebyshev's rule.

This is a ratio (out of 1)

At Least $\left(1 - \frac{1}{K^2}\right)$ of the observations are between $\bar{x} - KS$ and $\bar{x} + KS$ for any $K > 1$.

Or at most $\left(\frac{1}{k^2}\right)$ of the observations are less than $\bar{x} - kS$ or greater than $\bar{x} + kS$.

This is the complement ratio (out of 1)



Ex. Consider a collection of 500 observations with mean 60 and standard deviation 2. Find an interval whose center is 60 which contains at least 450 observations.

Soln. Using Chebeyshev's inequality, we need to solve $1 - \frac{1}{k^2} = \frac{450}{500}$ for k.

$1 - \frac{1}{k^2} = \frac{9}{10}$, then $\frac{1}{k^2} = \frac{1}{10}$, then $k^2 = 10$, then $k = 3.16$.

So, the required interval is

$$[\bar{x} - kS, \bar{x} + kS] = [60 - 3.16(2), 60 + 3.16(2)]$$
$$= [53.68, 66.32].$$

Ex. The mean $\bar{x}$ of grades of 1000 students is $\bar{x} = 55$, and the standard deviation is $S = 16$.

(1) Find the number of students who got grades between 23 and 87.

Soln. $23 = \bar{x} - kS = 55 - k(16)$. So $k = 2$.

Then by Chebeyshev's rule, at least $1 - \frac{1}{2^2} = \frac{3}{4}$ of the students got grades between 23 and 87.

So, the answer is $\frac{3}{4} \times 1000 = 750$ students.

(2) Find the number of students who got grades less than 31 or greater than 79.

Soln. $31 = \bar{x} - kS = 55 - k(16)$. So $k = 3/2$.
Then by Chebeyshev's rule, at most $\frac{1}{(3/2)^2} = \frac{4}{9}$ of the students got grades less than 31 or greater than 79. So, the answer is

$$\frac{4}{9} \times 1000 = 445 \text{ students.}$$

---

(2) Empirical rule

For data bell-shaped frequency curve;

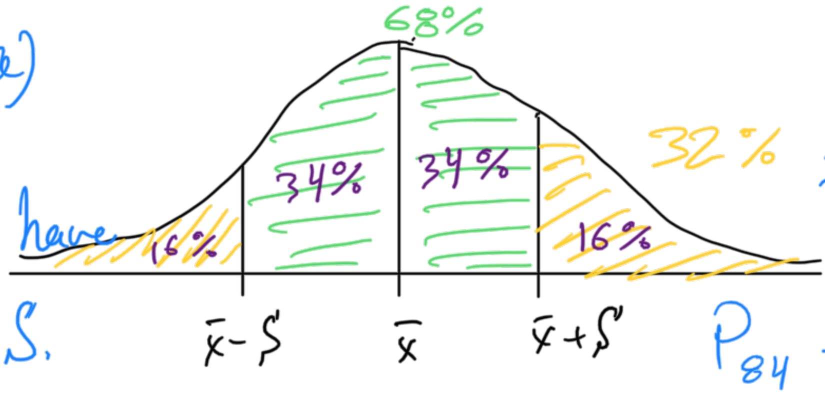
a) The interval $(\bar{x} - S, \bar{x} + S)$ contains about 68% of the observations. $(k=1)$

b) The interval $(\bar{x} - 2S, \bar{x} + 2S)$ contains about 95% of the observations. $(k=2)$

c) The interval $(\bar{x} - 3S, \bar{x} + 3S)$ contains about 99% of the observations. $(k=3)$

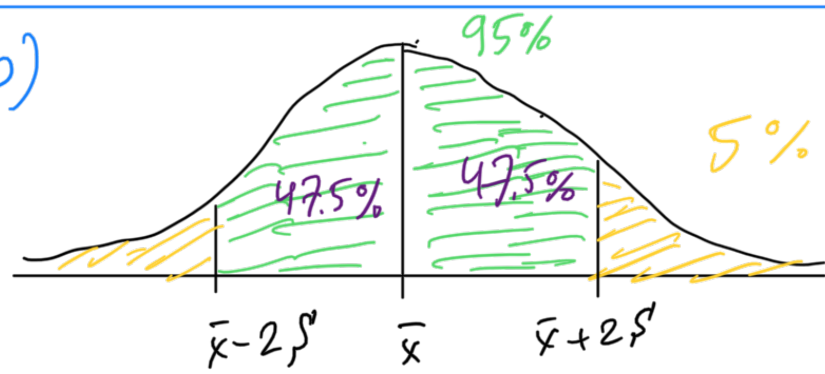
a)

From here we have $P_{16} = \bar{x} - S$.

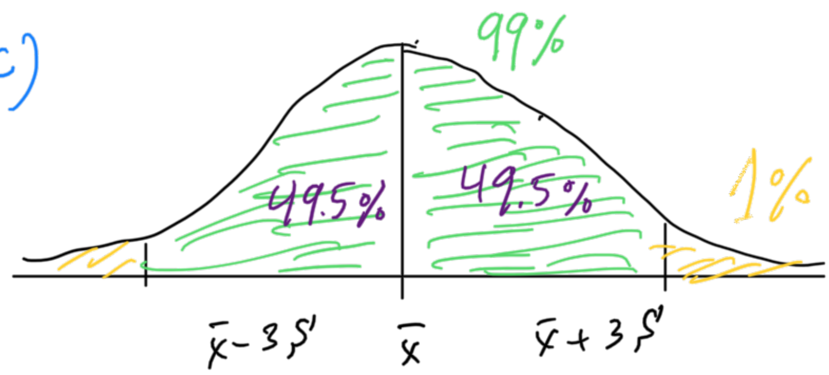


From here we have $P_{84} = \bar{x} + S$.

b)



c)



Ex. Consider a collection of 500 observations with mean 60 and standard deviation 2. Find the number of observations in the interval [56, 62] assuming bell-shaped distribution.

Soln. We solve

$[\bar{x} - K_1 S, \bar{x} + K_2 S] = [56, 62]$ for K_1 and K_2.

Now, $60 - K_1(2) = 56$ then $K_1 = 4/2 = 2$, ⇝ 95%/2 of the observations.

and $60 - K_2(2) = 62$ then $K_2 = 2/2 = 1$. ⇝ 68%/2 of the observations.

We divided by 2 because each border is one of the 2 endpoints.

From the empirical rule ((a) and (b)), the percentage of observations in the given interval is

$\frac{1}{2}(95\%) + \frac{1}{2}\,68\%$, which is

$$\frac{0.95}{2} + \frac{0.68}{2} = \frac{1.63}{2} = 0.82 \text{ or } 82\%$$

So, the number of observations is

$$(500)\left(\frac{82}{100}\right) = 410 \text{ observations.}$$

Ex. The grades of 1000 students are bell-shaped, with mean $\bar{x} = 50$ and std. $S = 13$.

(1) How many students got grades greater than 63?

Soln. $63 = 50 + 13 = \bar{x} + S$

From the empirical rule (a), there are 16%.

Number of students is $\frac{16}{100} \times 1000 = 160$ students.

(2) How many students got grades between 37 and 76.

Soln. $37 = 50 - 13 = \bar{x} - S$.

$76 = 50 + 2(13) = \bar{x} + 2S$.

From the empirical rule (a), there are 34%.

From the empirical rule (b), there are 47.5%.

Total number of students has the percentage

$$34\% + 47.5\% = 81.5\%$$

So, the number is $\frac{81.5}{\cancel{100}} \times 10\cancel{00} = 815$ students

(3) What is P_{84} of the grades?

Soln. $P_{84} = \bar{x} + S = 50 + 13 = 63.$

(4) What is P_{16} of the grades?

Soln. $P_{16} = \bar{x} - S = 50 - 13 = 37.$

Searching keywords:

- Comparing two collections.
- Z-score, coefficient of variation.
- Chebyshev's rule, empirical rule.
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx

For any comments or concerns, please use my email to contact me.

د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com