EE325

ASSIGNMENT - 1

- → KALPIT BORKAR 200070029
- → NAVNEET 200070048
- → SANJHI PRIYA 20D070070
- → YUVRAJ SINGH 200070093

1 b) (i)

METHOD 1 :

Case- 1 : K=10

Calculated average (guess) = 117.902
Calculated standard deviation (guess)
= 12.37

Case-2: K= 20

Calculated Average = 119.031
Calculated Standard deviation = 18.43

Case -3 : K= 50

Calculated Average = 119.959
Calculated standard deviation = 19.798

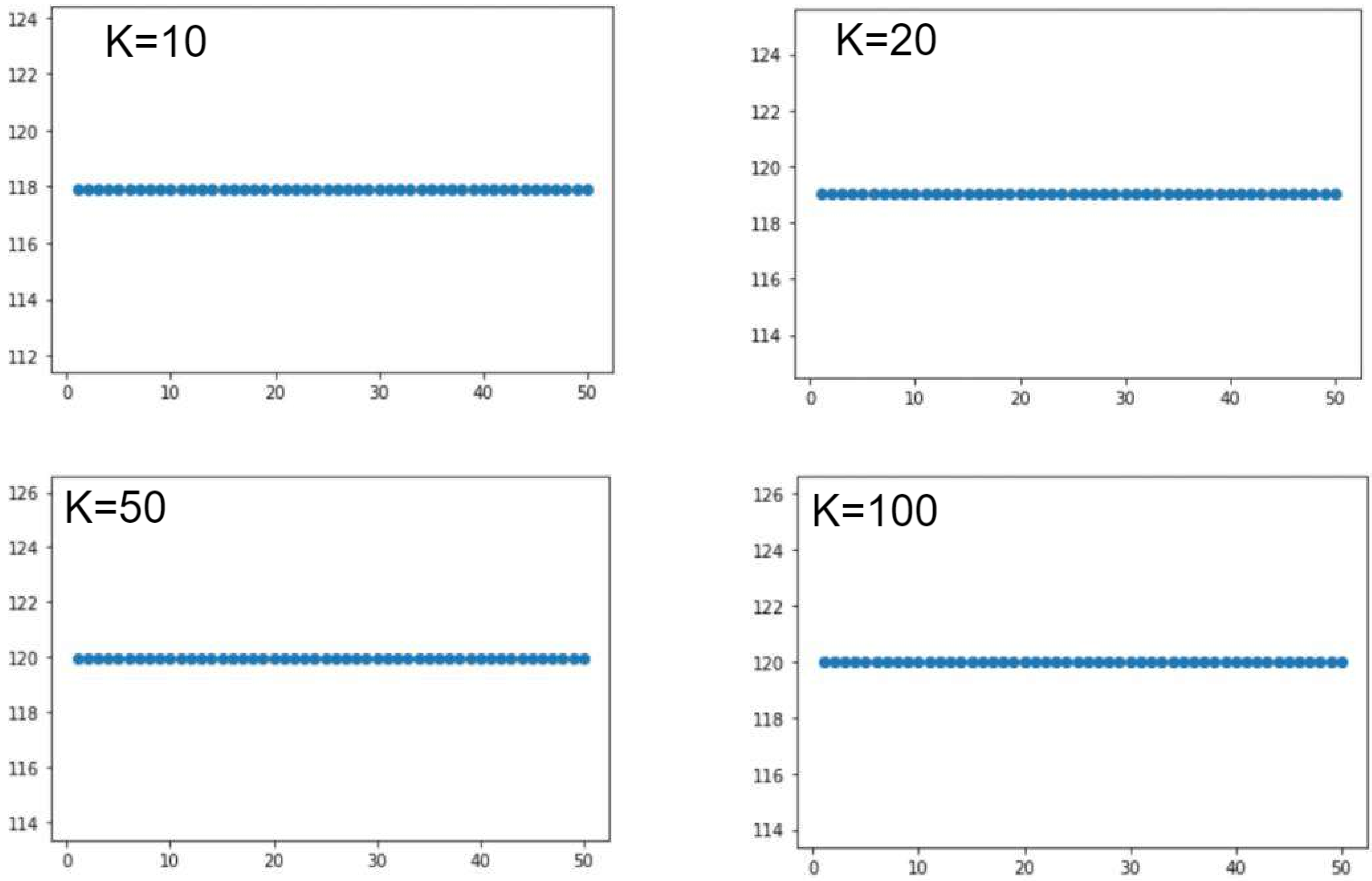
Case- 4 : K= 100

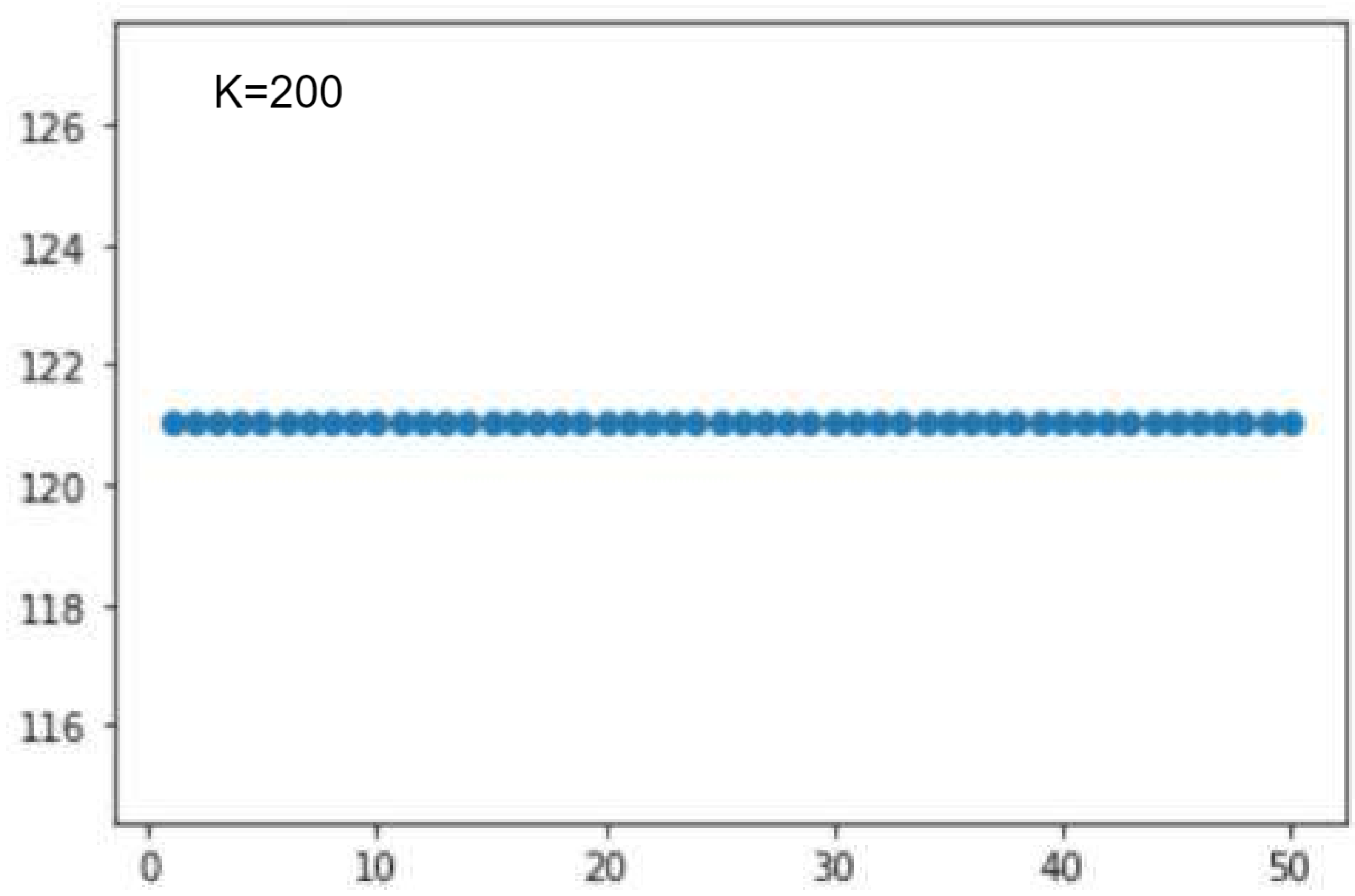
Calculated Average = 119.981
Calculated Standard deviation = 19.537

Case- 5 : $K = 200$

Calculated Average = 121.040
Calculated standard deviation = 18.893

Method 1





K=200
126
124
122
120
118
116
0
10
20
30
40
50

METHOD : 2

Case-1 : K= 10

~~A~~ Calculated Average = 121.675
Calculated Standard deviation = 17.158

Case-2 : K= 20

Calculated Average = 118.251
Calculated Standard deviation = 18.855

Case-3 : K=50

Calculated Average = 120.555
Calculated Standard deviation = 19.514

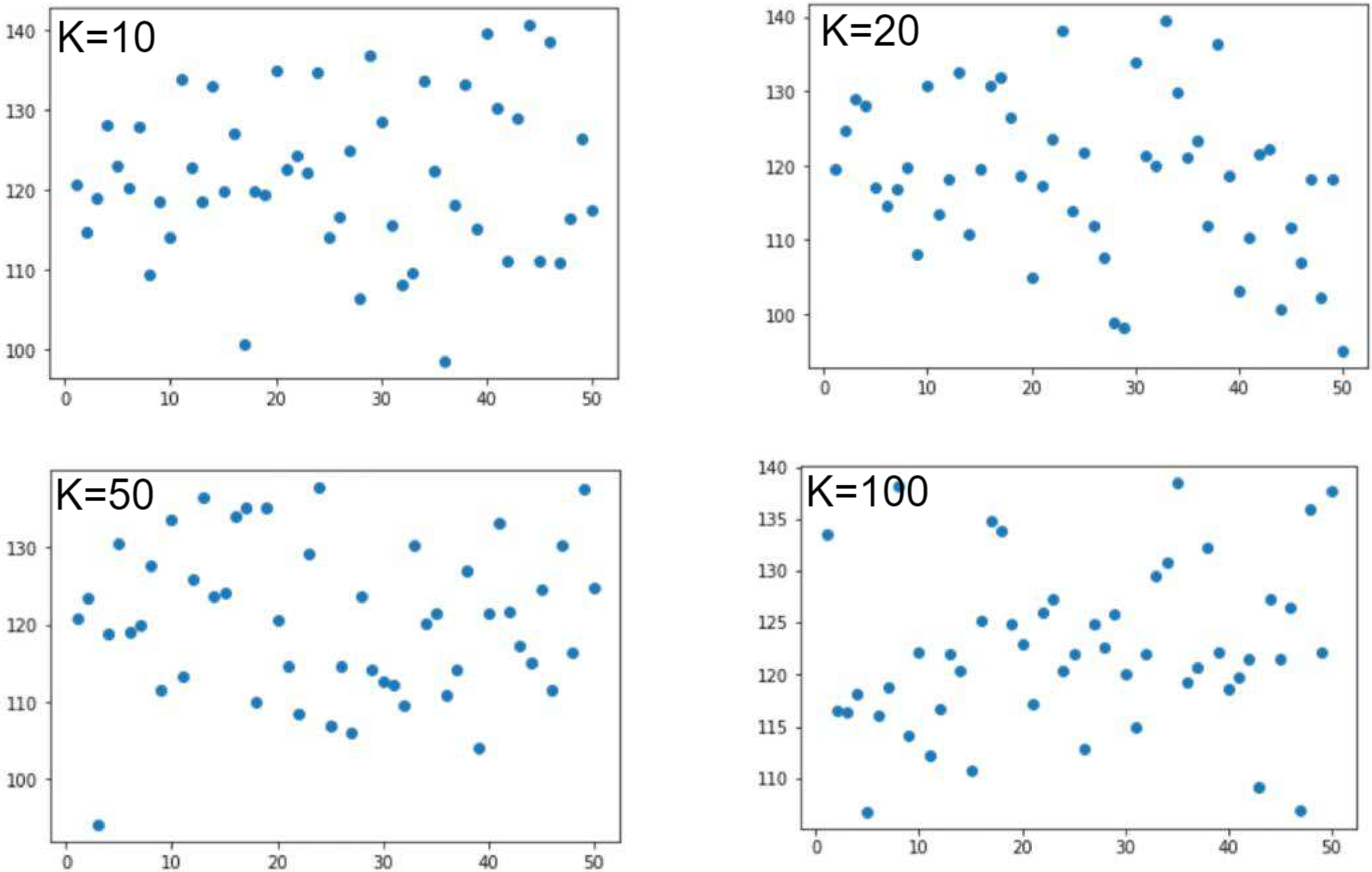
Case-4 : K= 100

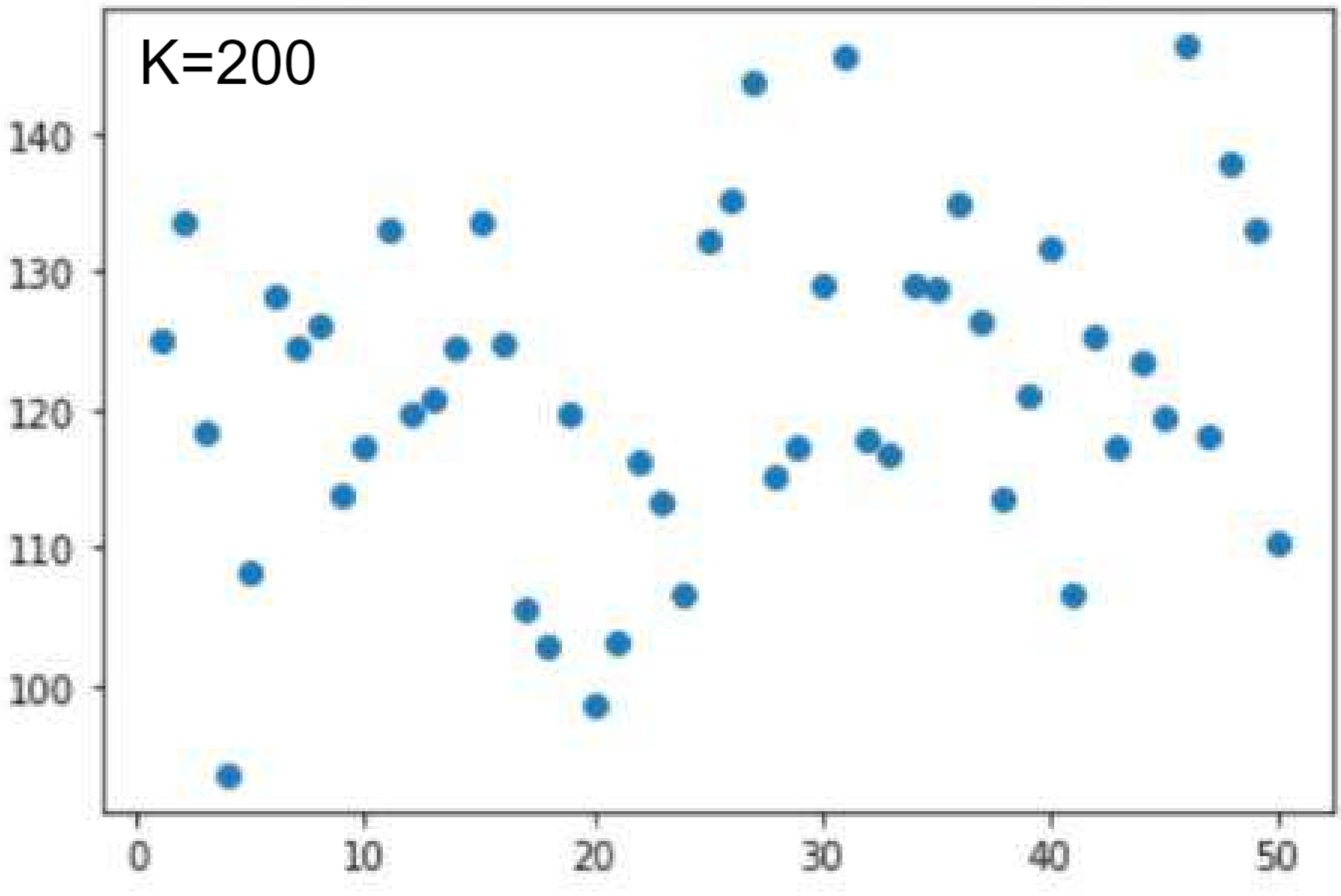
Calculated Average = 122.422
Calculated Standard deviation = 19.638

Case-5: $K = 200$

Calculated Average - = 121.732
Calculated standard deviation = 19.728

Method 2




K=200
140
130
120
110
100
0
10
20
30
40
50

METHOD - 3

Case-1 : K=10

Calculated Average = 120.727
Calculated standard deviation = 18.236

Case-2 : K=20

Calculated Average = 121.547
Calculated standard deviation = 19.512

Case-3 : K=50

Calculated Average = 121.103
Calculated standard deviation = 19.976

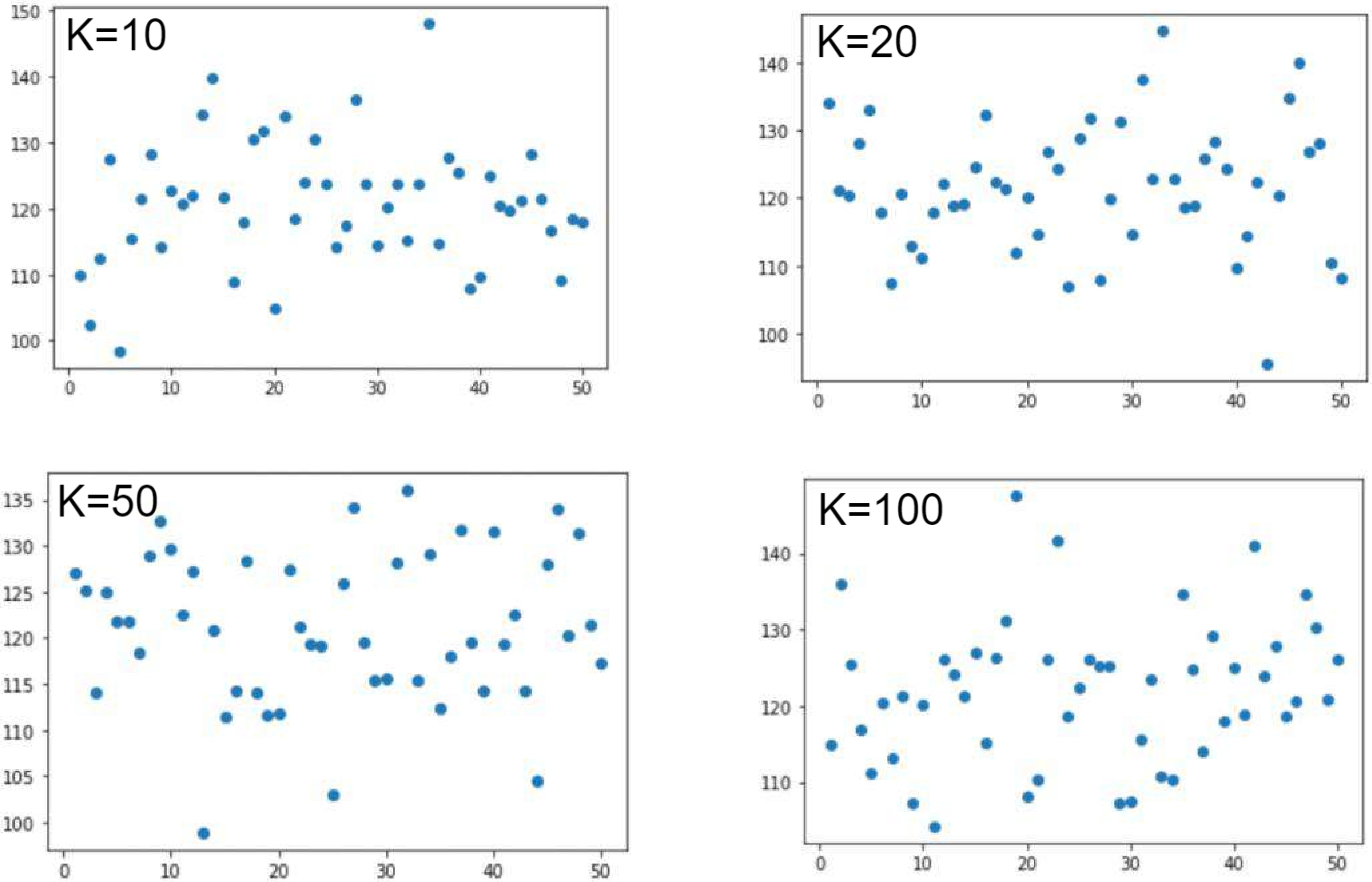
Case-4 : K=100

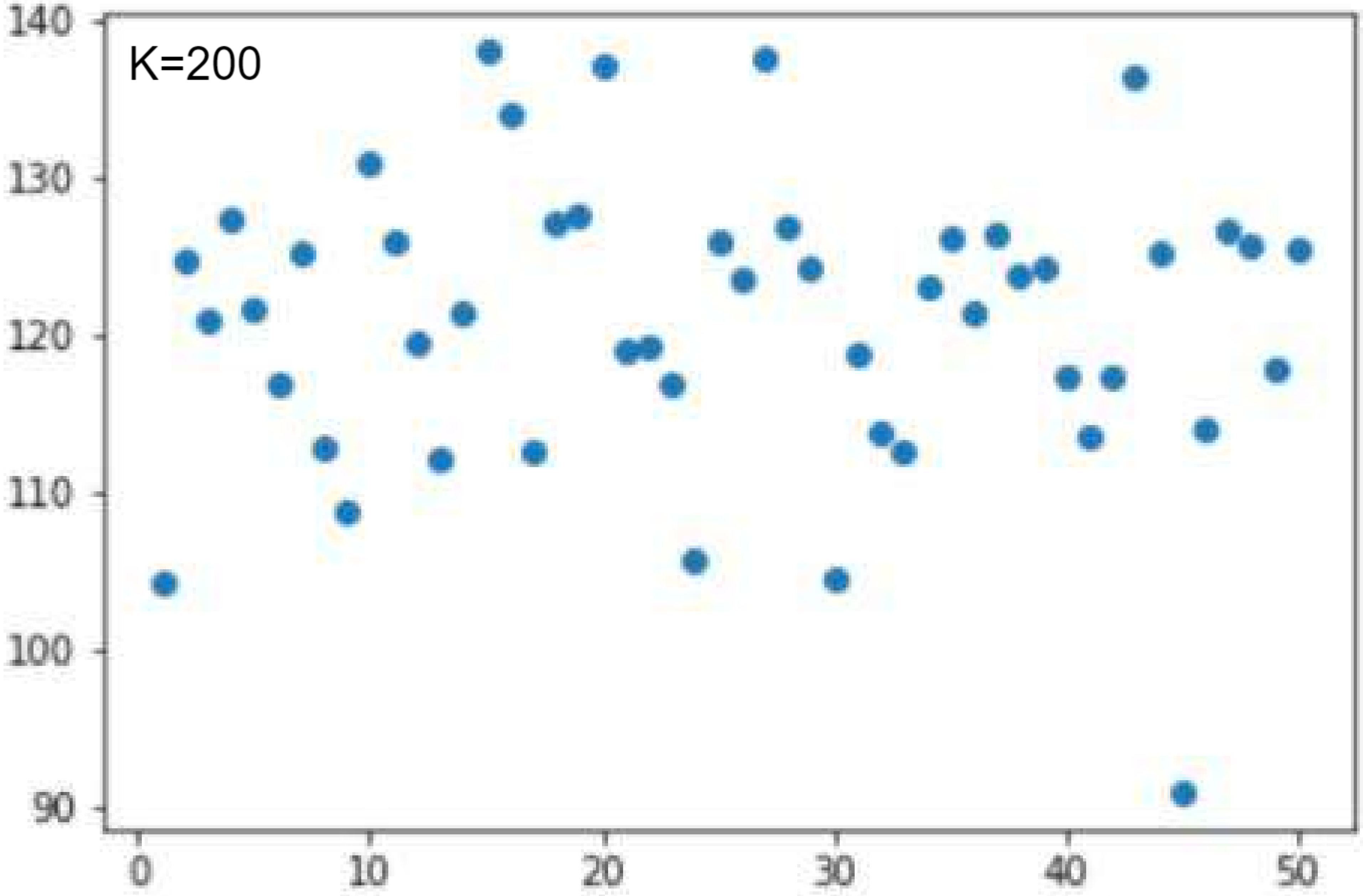
Calculated Average = 121.934
Calculated standard deviation = 19.747

Case-5: K=200

Calculated Average = 120.968
Calculated Standard deviation = 19.783

Method 3


K=10
100
110
120
130
140
150
0
10
20
30
40
50
K=20
100
110
120
130
140
0
10
20
30
40
50
K=50
100
105
110
115
120
125
130
135
0
10
20
30
40
50
K=100
110
120
130
140
0
10
20
30
40
50


K=200
140
130
120
110
100
90
0
10
20
30
40
50

* Actual Average : 120.133

* ~~Act~~ Actual Standard deviation = 19.975

6) ii)

A.) • According to the calculated values (guess), method-3 seems to give the most accurated answer as compared to the actual values.

• Method 3 is the "most random" way of sampling, and hence the calculated values tend to be close to the actual values.

Camlin Page
Date / /

B) • Based on the calculated data, it can be observed that the calculated value is more accurate for higher values of K.

• For a quantative measure of "sureness of our estimate", we can define something called **Confidence interval:**

Let's suppose

$\mu =$ actual average
$\hat{\mu} =$ Calculated average

we can say that for $\epsilon > 0, \delta > 0$

$$\mathbb{P}(|\mu - \hat{\mu}| > \epsilon) \leq \delta$$

where $\delta = 2\exp(-n\epsilon^2)$

So, this inequality says that

$\hat{\mu} \notin (\mu - \epsilon, \mu + \epsilon)$ has a probability less than δ.

$$2\exp(-n\epsilon^2) = \delta \Rightarrow n = \frac{1}{\epsilon^2}\ln\left(\frac{2}{\delta}\right)$$

Here we can see that if we decrease the value of ϵ or δ, we need to increase n.
This relation of ϵ, n, δ can be used as a ~~qu~~ quantitative measure.

Question – 2

define : P_i = probability of events : {H} calculated after the i'th toss.

→ for the ideal case, i.e. when the coin is fair, this probability P_i should converge to 0·5.

→ When P_i is plotted v/s i, the amount of deviation from 0·5 gives measure of 'bias' of the coin.

→ We can introduce a variable 'degree of sureness' which describes the quantitative measure of our 'sureness' that the coin is biased.

→ We should also make sure that the initial P_i are less significant as compared to later P_i while calculating deviation, and hence we should calculate weighted average of error.

→ Weighted average of error, i.e.

$$\left[P_i - 0.5 = \sum_{n=1} \frac{2 \times i \times (P_i - 0.5)}{n(n+1)} \right]$$

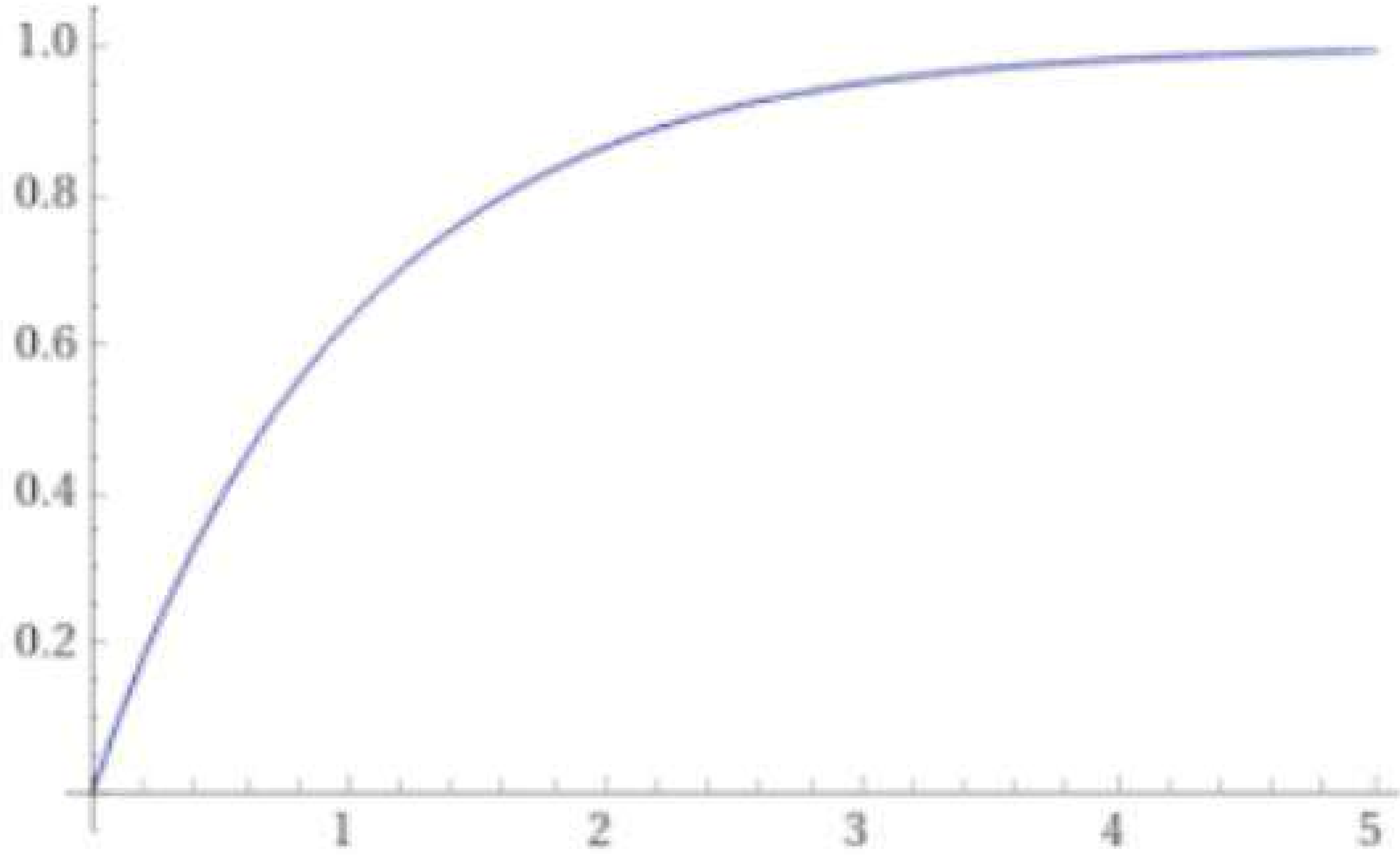
→ This weighted average takes value in (−0·5, 0·5) and more the value is away from zero, the more we are sure that the coin is 'biased'.

→ To make it easy to interpret, we took | weighted average |, so that it takes value from in (0, 0·5).

→ We should also make sure that the same value of weighted average for higher value of N should have more weightage on 'degree of sureness'.

→ So for calculating the 'degree of ~~e~~ sureness' we multiply a factor of: $\left[1 - e^{\frac{-N}{100}}\right]$

Plot



→ define : D_i = degree of sureness that the coin is biased after the i'th toss.

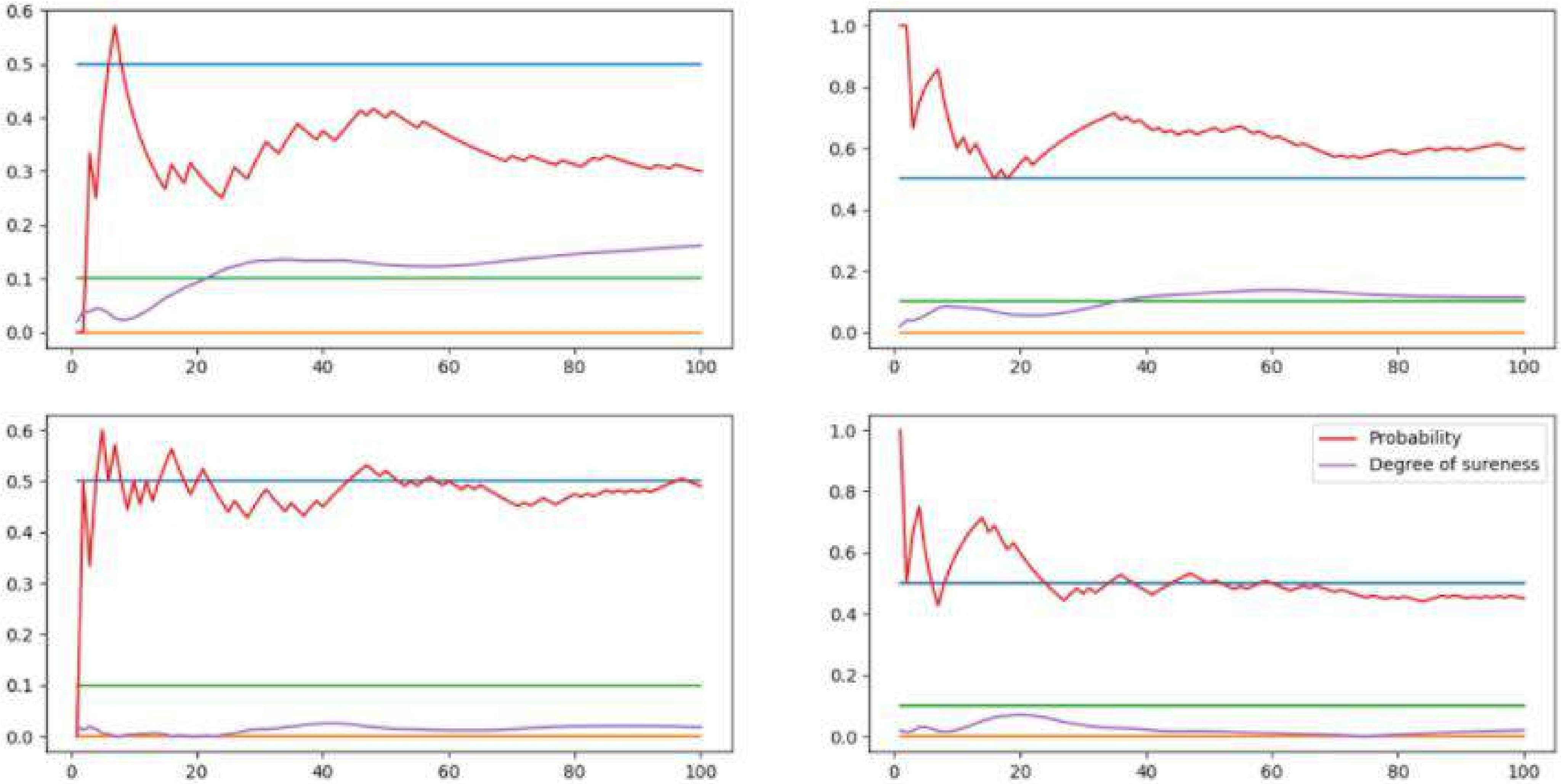
→ degree of sureness takes values in (0,1)

→ if D_i is less than a particular value, then we won't doubt.

→ A value close to 1 implies a biased coin.

→ A value close to 0 implies an unbiased coin.

Plots :



b. → In this situation, one can never be sure about his initial assumptions being wrong with this much amount of data, one can only say about the chances / probability of its assumption being wrong.

→ For any conclusion we would have to look for large number of results, and then determine the convergence of the probability value.

Question - 3

→ After having done straight line fitting of the form $y = ax + b$ on the data given in hw1C1, we get,

$$a = 0.001$$
$$b = 51.78$$

→ let y be the actual weight
let $\hat{y}$ be the prediction using our model.

→ the difference between the actual weight y & the prediction $\hat{y}$ could be assumed as the "error"

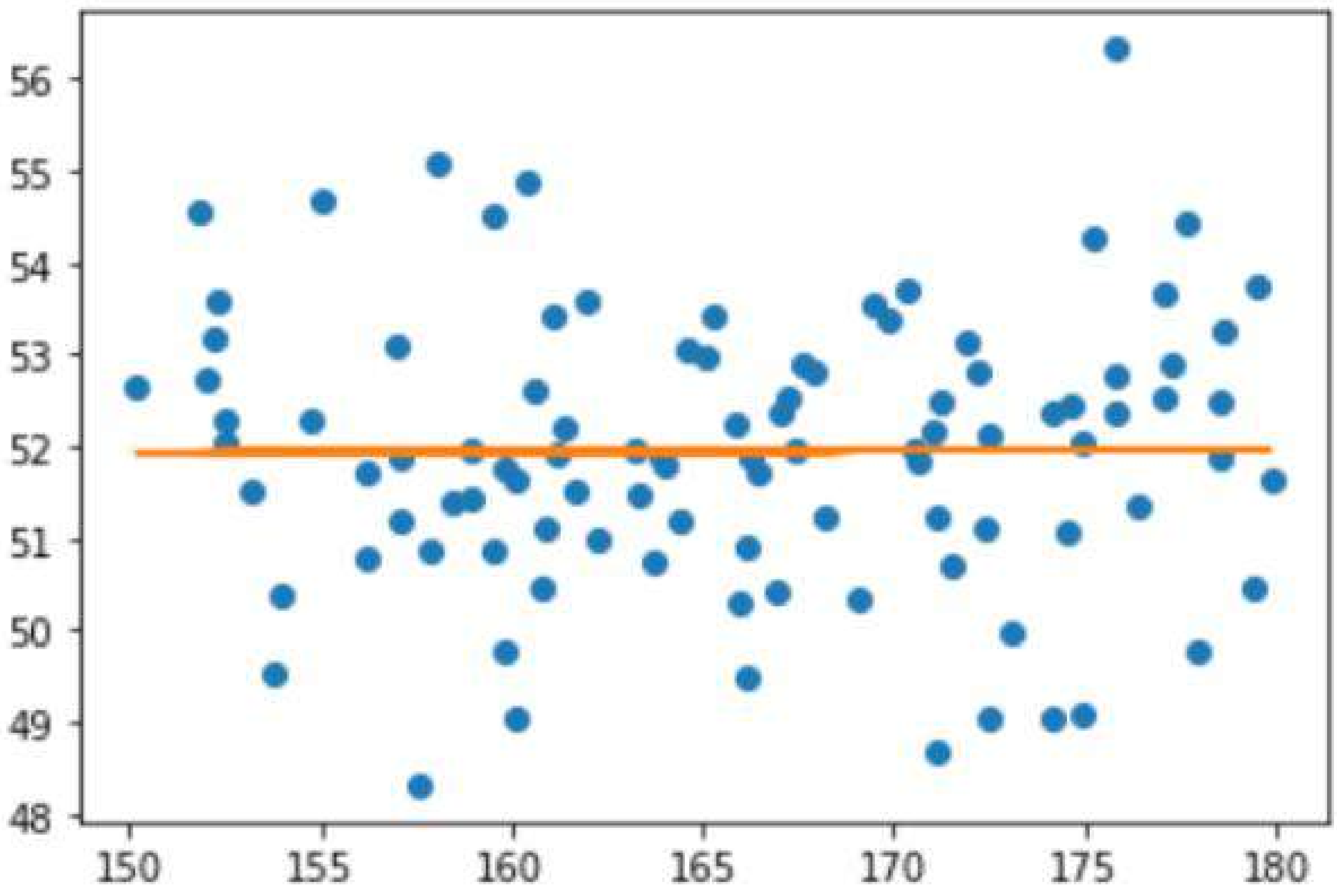
i.e. $\text{error} = y - \hat{y}$.

→ greater the difference, more is the error, i.e. the predictions are more inaccurate.

→ if $y - \hat{y} > 0$:
error is positive.
actual weight > predicted weight

→ if $y - \hat{y} < 0$:
error is negative
actual weight < predicted weight.

→ if $y - \hat{y} = 0$:
error is zero.
actual weight = predicted weight


56
55
54
53
52
51
50
49
48
150
155
160
165
170
175
180

Q3. b.

for height 155:

→ Average of error = -2.71

→ Standard deviation of error = 1.19

for height 160:

→ Average of error = -1.10

→ Standard deviation of error = 1.63

for height 170:

→ Average of error = 1.49

→ Standard deviation of error = 1.33

→ if average error < 0 : implies that people are underweight as compared to original data set.

→ if average error > 0 : implies that people are over weight as compared to original data set.

→ Standard deviation of error signifies how much the weight of people with same height vary among themselves.

→ Greater the value of standard deviation, greater will be the variation among them.

→ coefficient of variance (CV) = $\frac{\text{standard deviation}}{\text{mean}}$

→ if coefficient of variance < 1:

- → low variance data
- → our model is more "accurate".

→ higher the variance, lesser our model is accurate.