EE 325 Assignment 1

The screenshots of the ~~code and~~ scatter plots for every combination of K and the method of selecting the K are attached below.

(i) By inspecting the obtained scatter plots, the hand-wavy guess for the actual average, that we've arrived at is
120 call-min/day/student.

The code we wrote, was designed in such a way so as to calculate average standard deviation (i.e mean of the 50 standard deviations for a given combination of K & method of selecting K). Thus, the guess for the actual standard deviation was arrived at, with the aid of aforementioned provision of the code and it is 20.

(ii) If each of the 50 repetitions should be seen as a separate survey,

(A) Then the preferred scheme of choosing K students would be **Scheme III**. i.e 'Randomly' select 10,000 people in the colony. Rationally justifying, this scheme is more representative of the original data as compared to other schemes.

(B) The preferred value of K is 200. The sureness of the estimate of the value of average, from a single survey of K samples can be quantitatively described using the concept of Confidence Intervals.

$$\bar{x} \pm t_{n-1,\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

($\bar{x}$ is sample mean
s is sample standard deviation
n is sample size)

Where, $t_{n-1,\alpha/2}$ (known as t-score) is some factor dependent on the size of our sample $(n-1)$ and the ~~degree of~~ confidence % we require $(\alpha/2)$ (say 95% or 99%.)

The smaller the confidence interval, for a given confidence % say 99%, the more is sureness of our sample mean with respect to actual mean. Thus,

Width of confidence interval $= 2(t_{n-1,\alpha/2})\left(\frac{s}{\sqrt{n}}\right)$

[Also, $t \propto \frac{1}{n}$]

So, width dereases with increasing value of n, leading us to choose preferred $K = 200$. Thus, all our guesses are justified appropriately.

Also, from the inspection of scatter plots, it can be observed that the distribution of points are more scattered as K increases, further supporting our choice of K.

It can be observed that scatter plots are more precise for scheme III plots, for a given K. Thus, it validates our choice of scheme to select students.

Using the actual data, we find that,

actual average = 120.13317711180902, which is very much close to our guess.

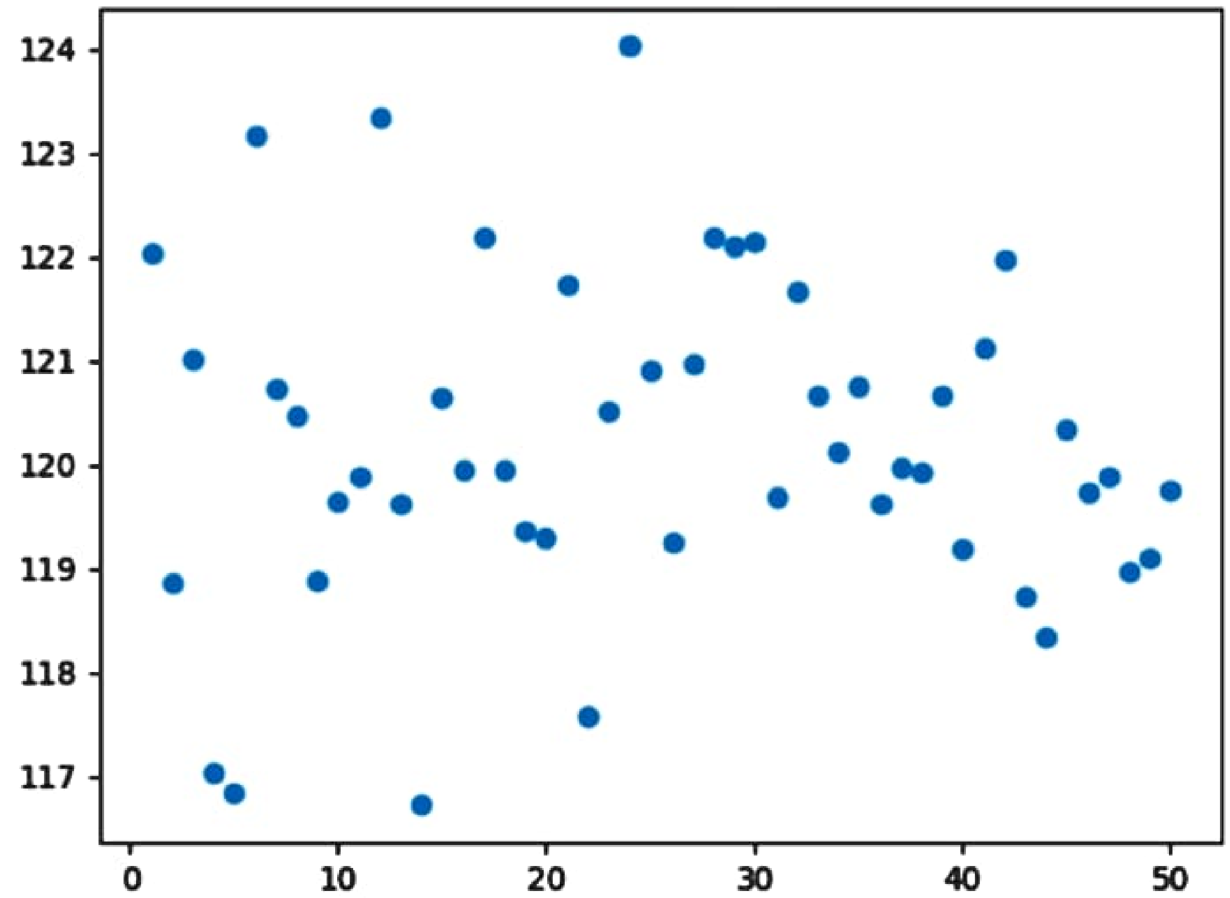
actual standard deviation = 19.975538382855526

Scheme 3 (Random K):

K=200

calculated mean= 120.22922187517852

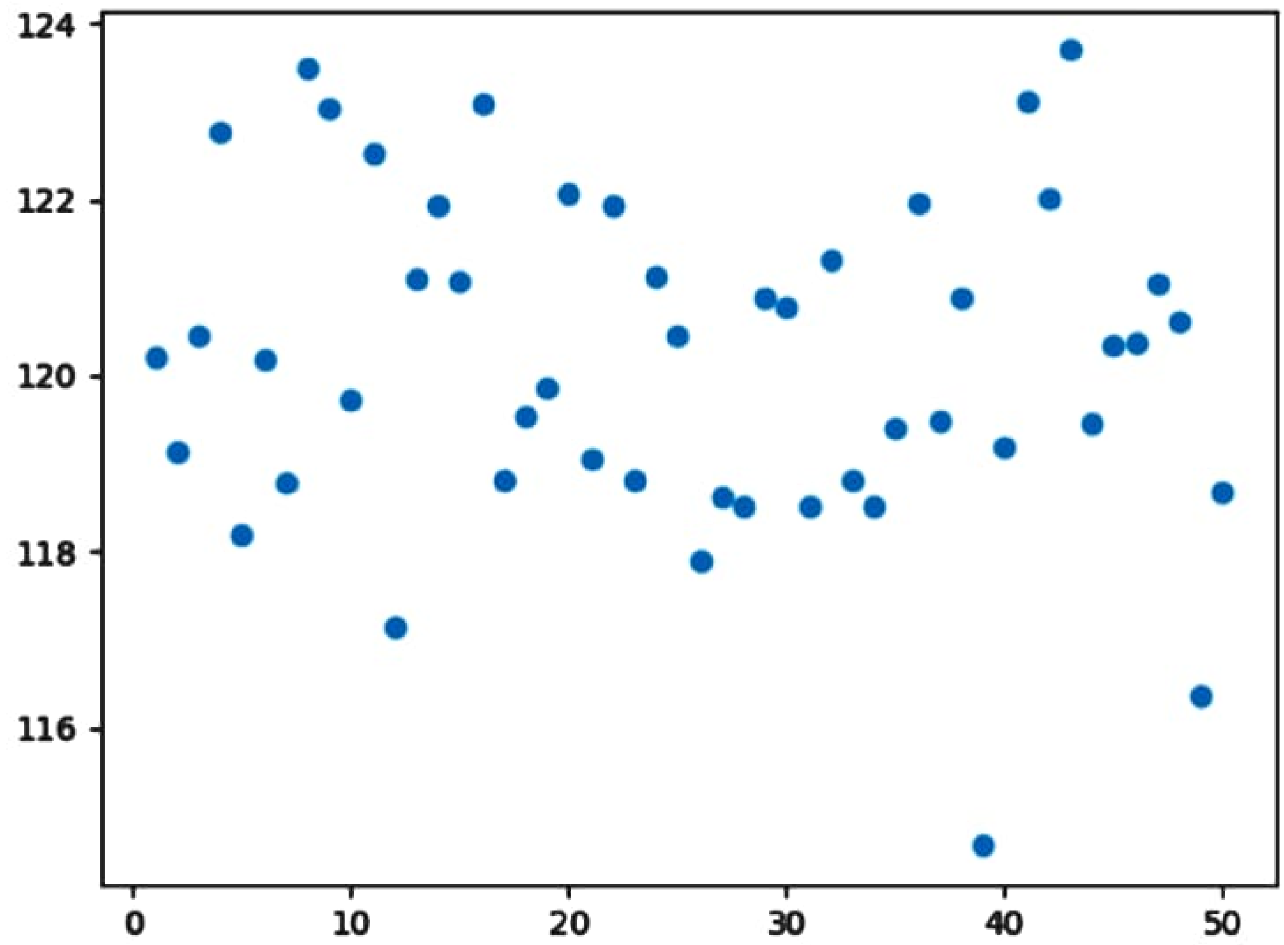
calculated average standard deviation= 20.12935843341547



K=100

calculated mean= 120.19421920555851

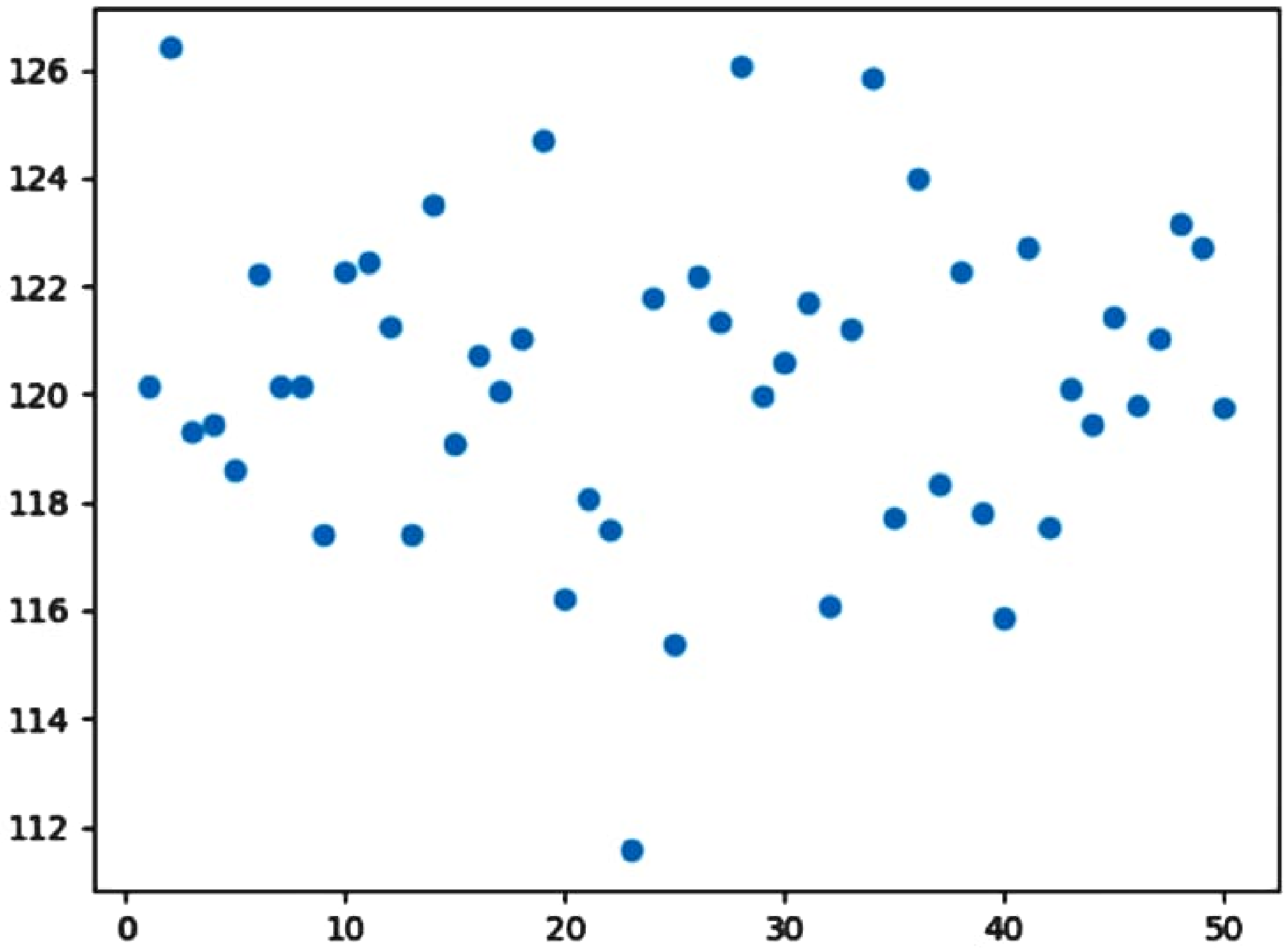
calculated average standard deviation= 19.8157135606269



K=50

calculated mean= 120.31059235747081

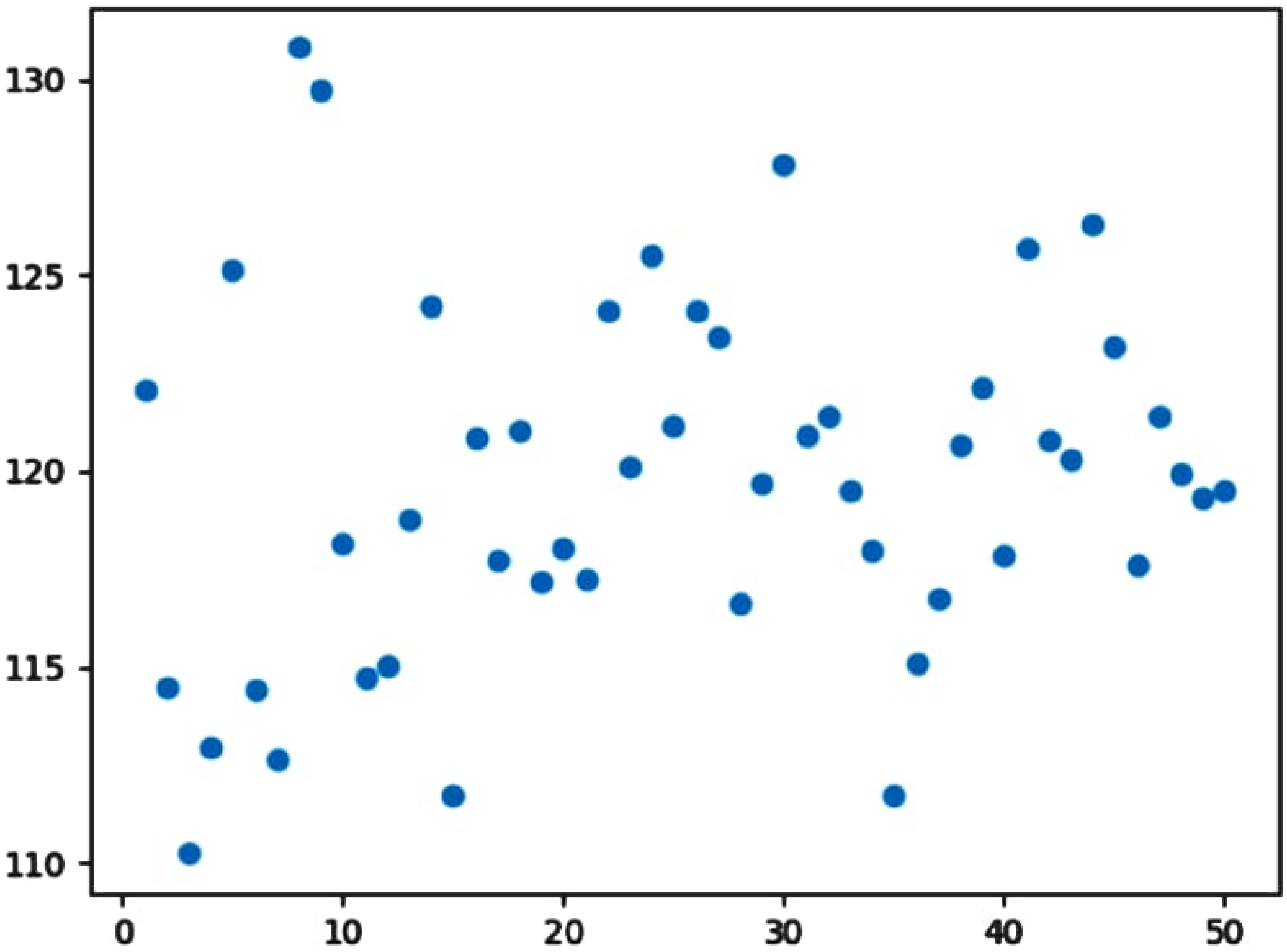
calculated average standard deviation= 19.56177926403608



K=20

calculated mean= 119.76101071571584

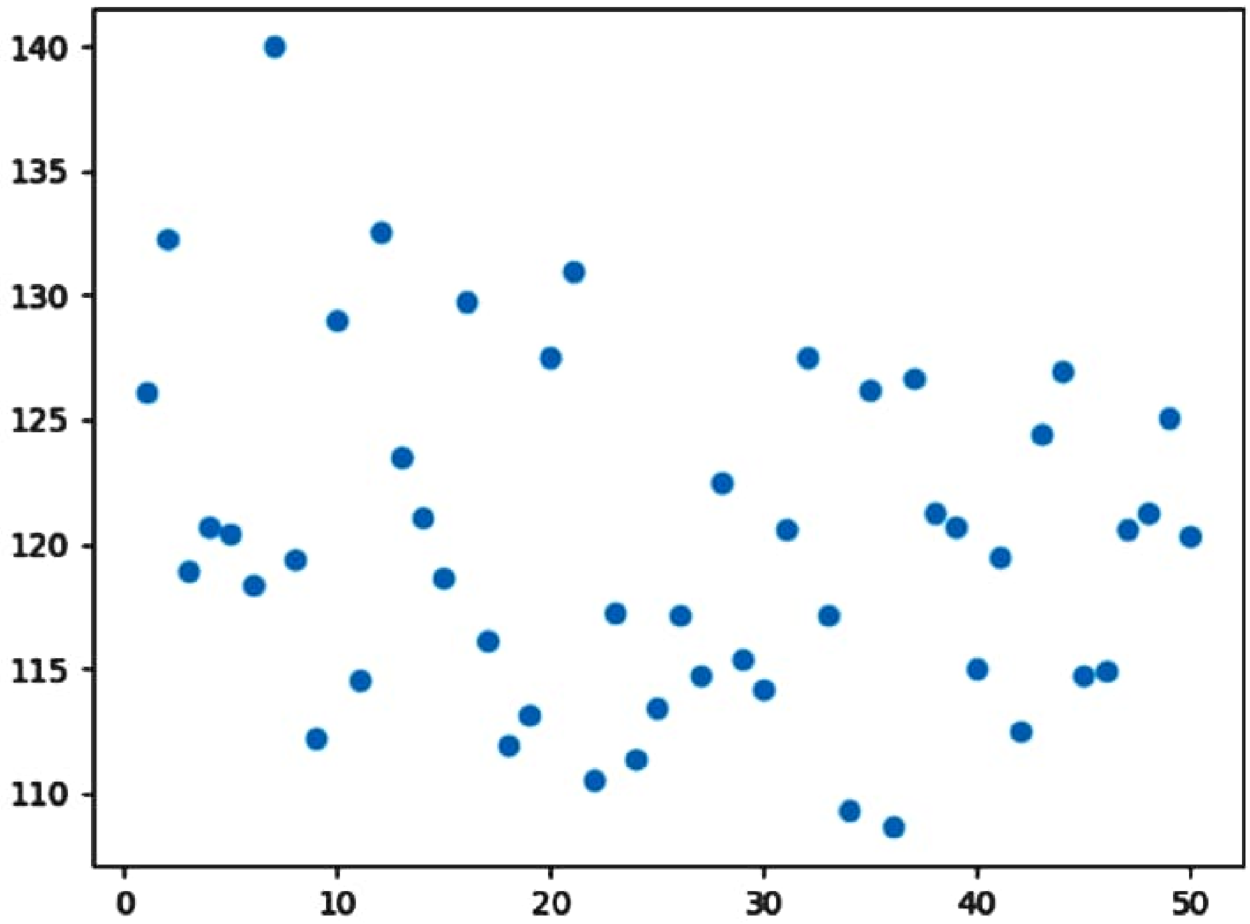
calculated average standard deviation= 18.821670932243602



K=10

calculated mean= 120.14390805720409

calculated average standard deviation= 17.945107775353748

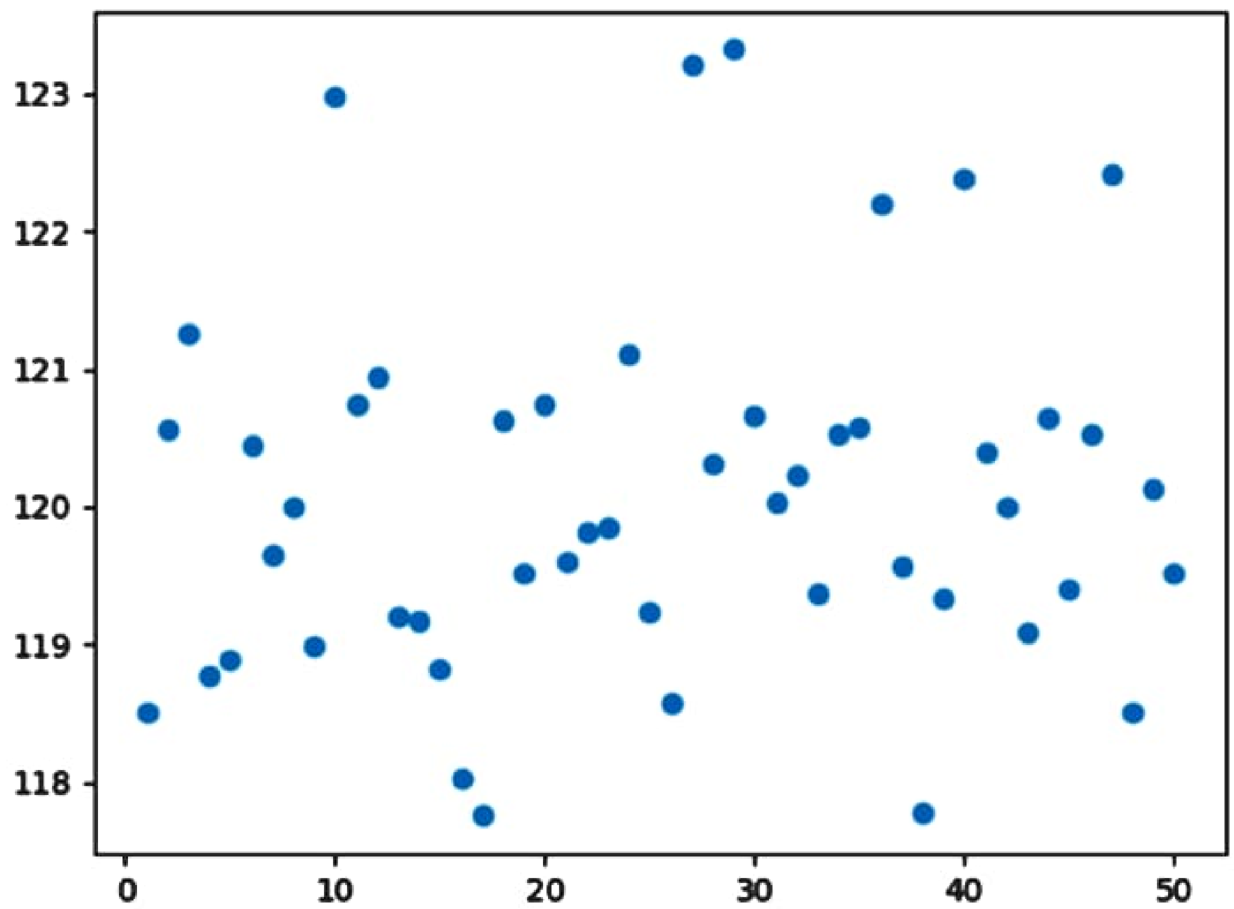


Scheme 2 (Selecting K from an arbitrary point):

K=200

calculated mean= 120.07865788013885

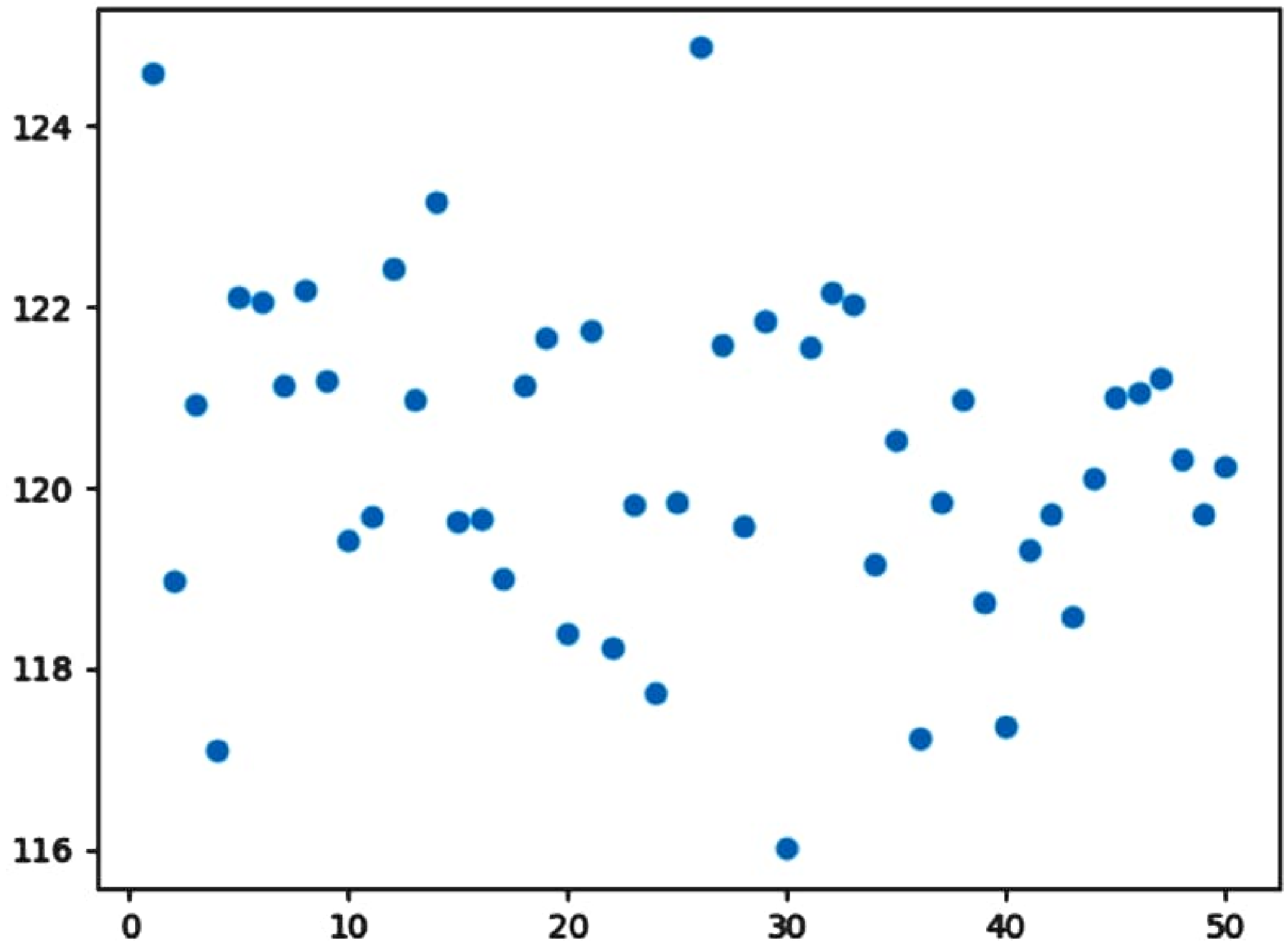
calculated average standard deviation= 19.832993970039976



K=100

calculated mean= 120.34963766310668

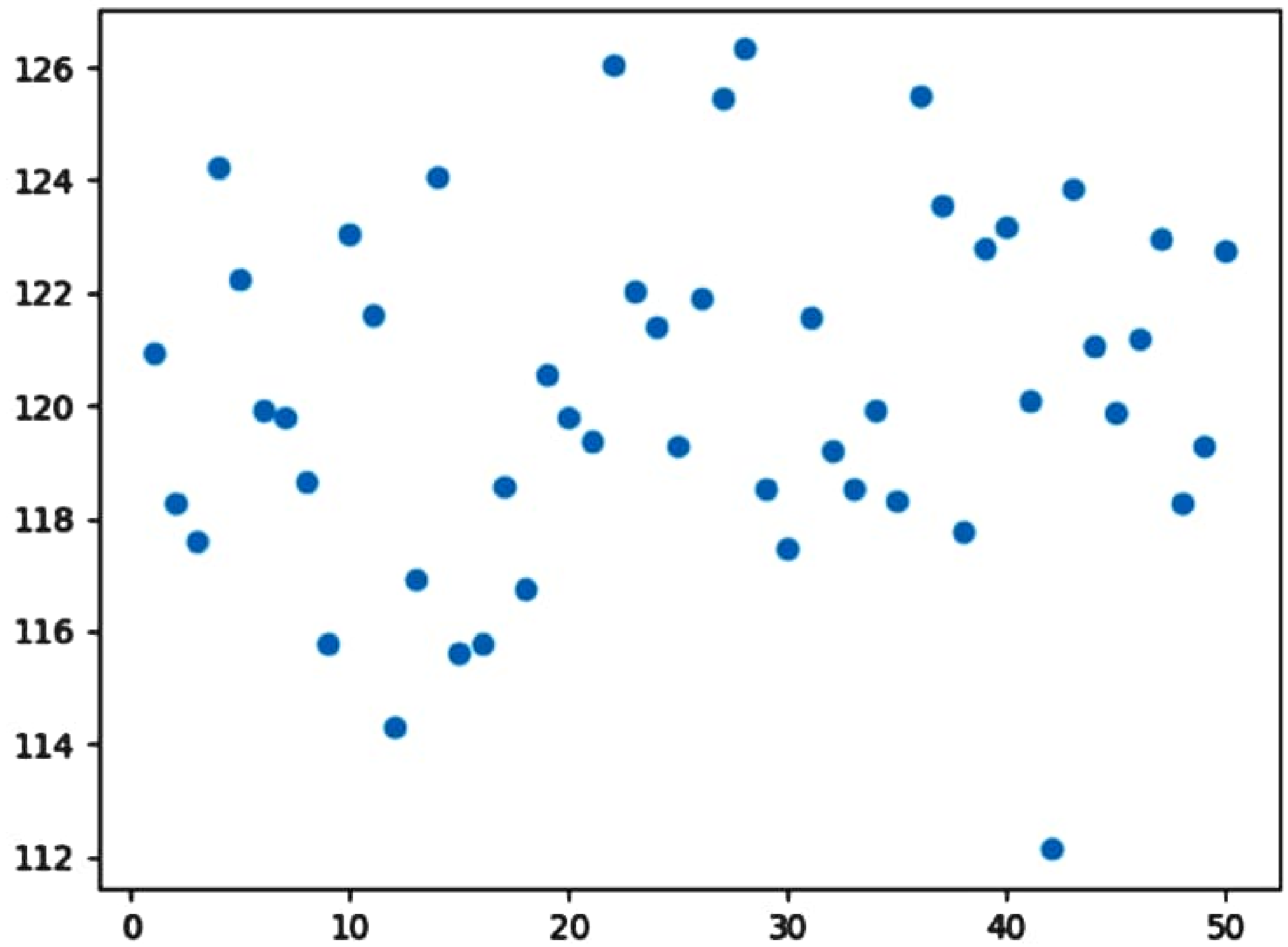
calculated average standard deviation= 19.600241891507256



K=50

calculated mean= 120.28306701108357

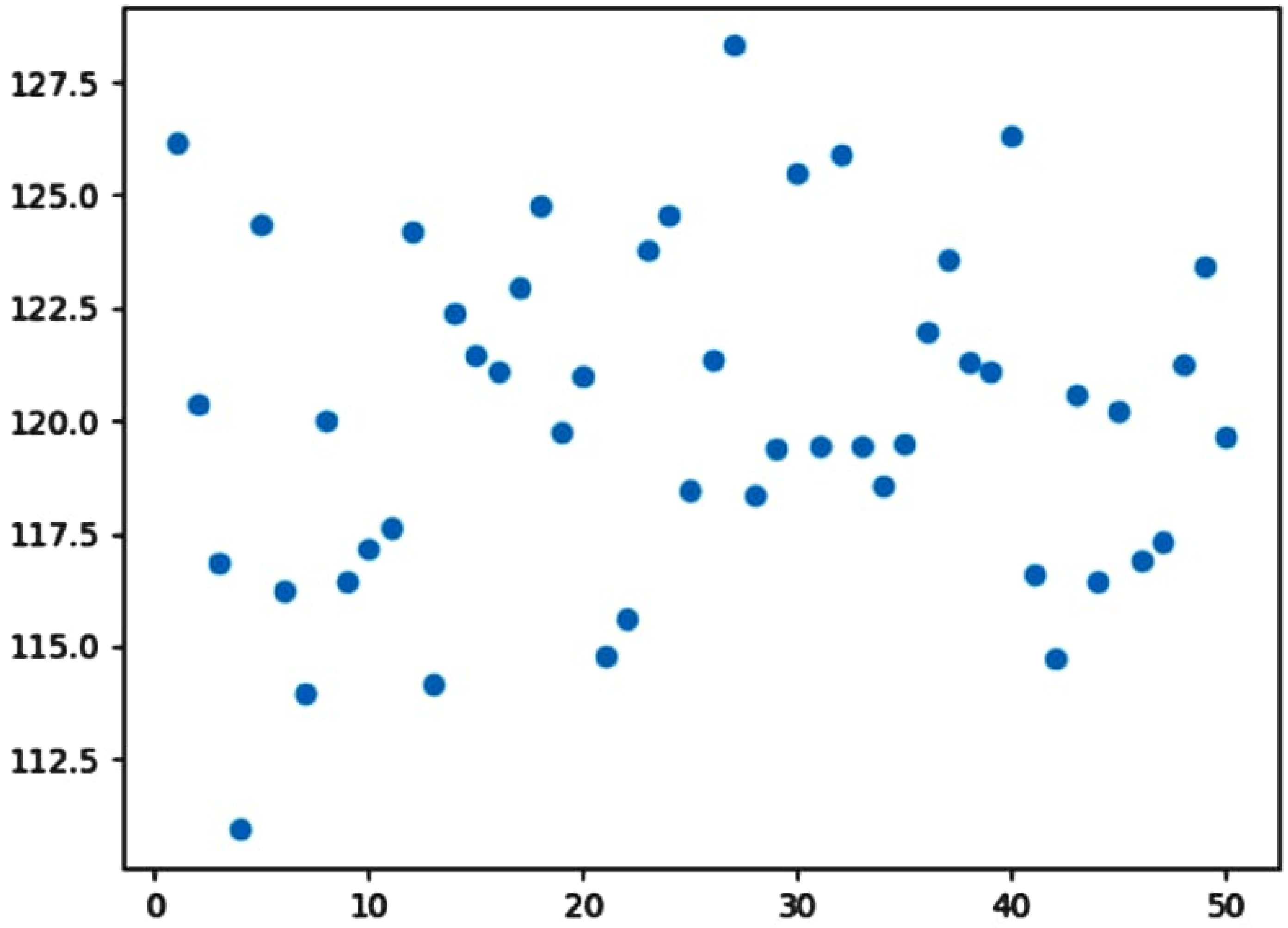
calculated average standard deviation= 19.17156564414576



K=20

calculated mean= 120.12274857002876

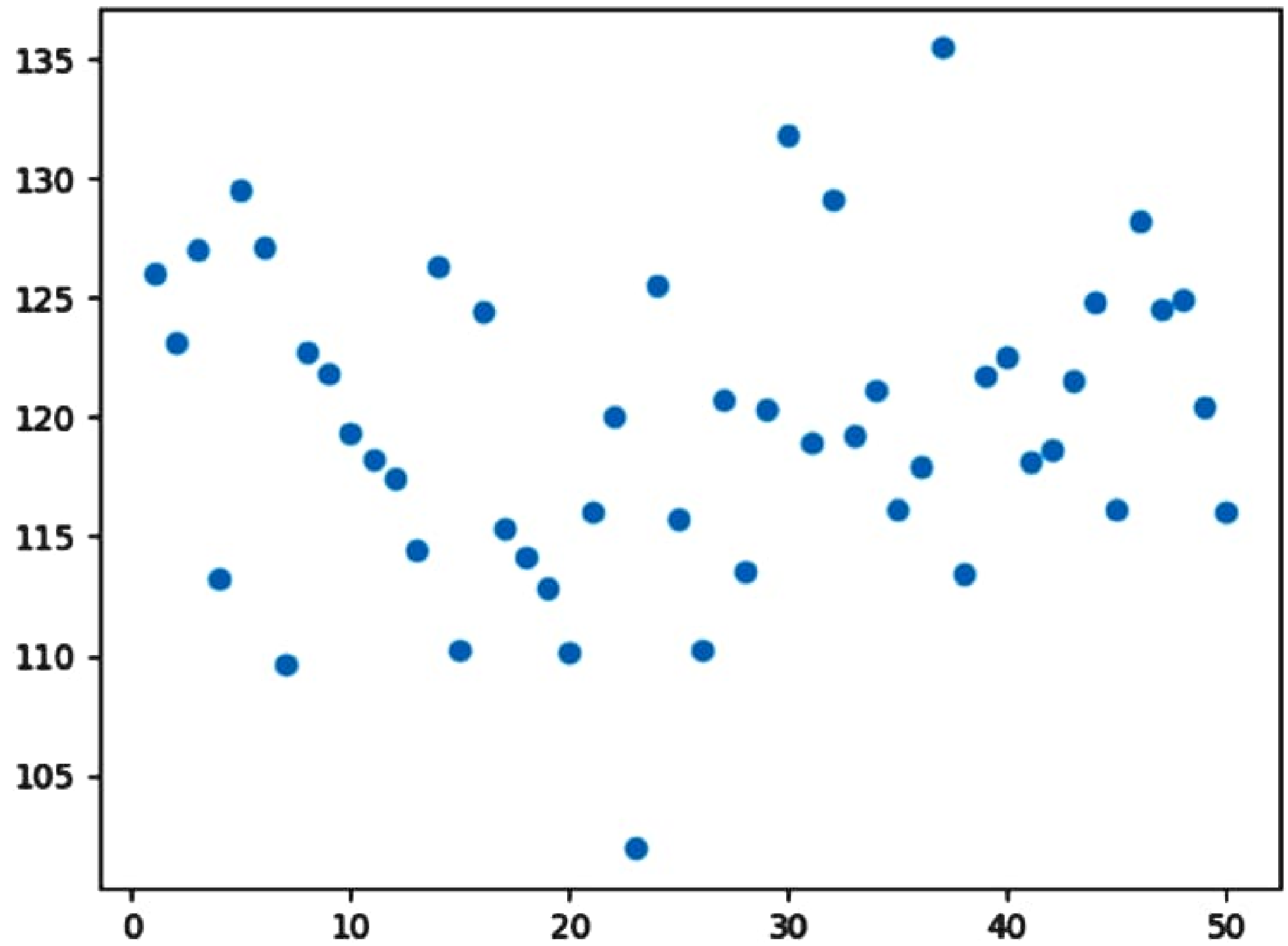
calculated average standard deviation= 19.44943891525027



K=10

calculated mean= 119.77671339775699

calculated average standard deviation= 19.262621008492395

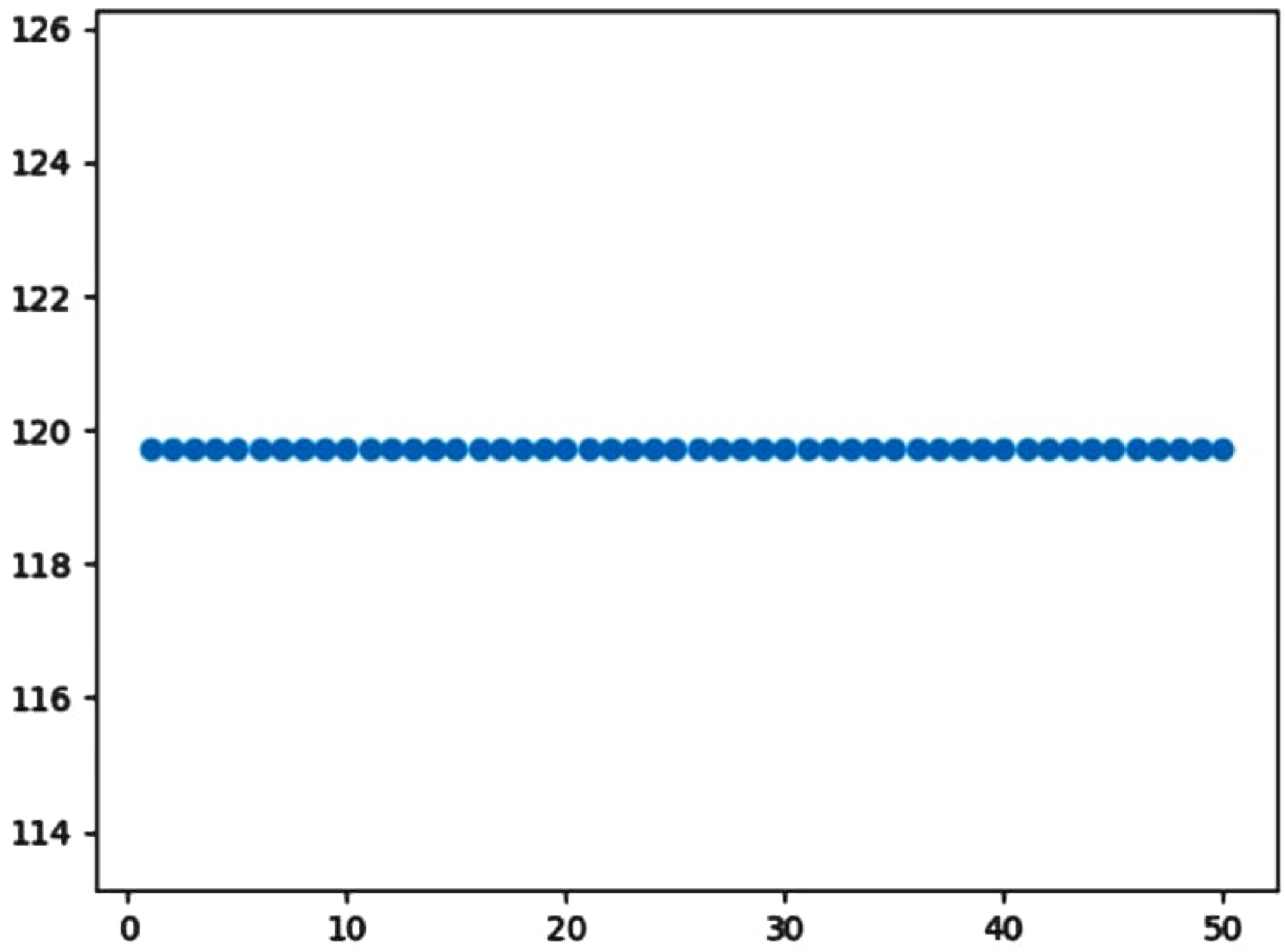


Scheme 1 (Choosing first K):

K=200

calculated mean= 119.7060312390456

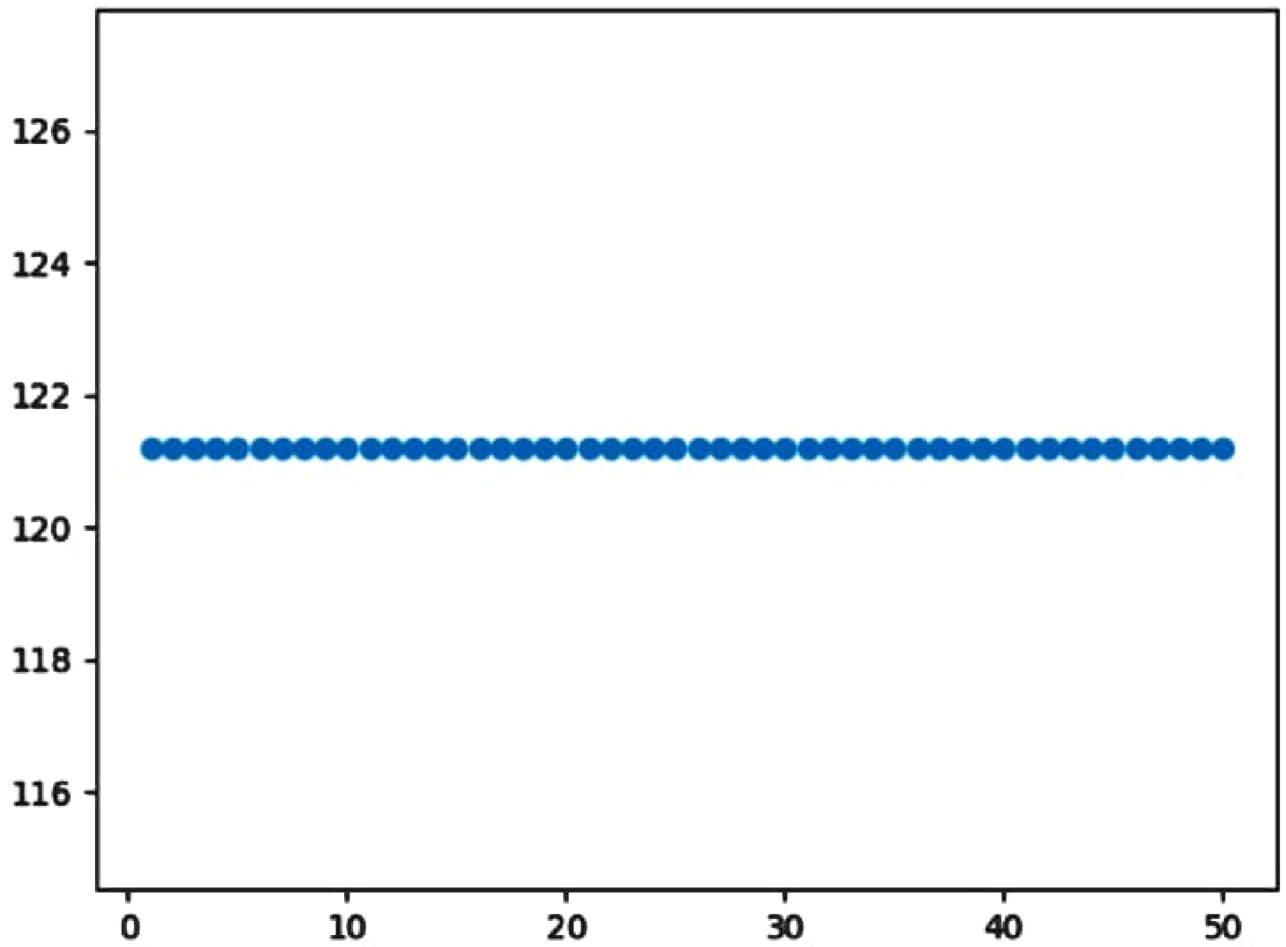
calculated average standard deviation= 20.46215677520817



K=100

calculated mean= 121.18210695177557

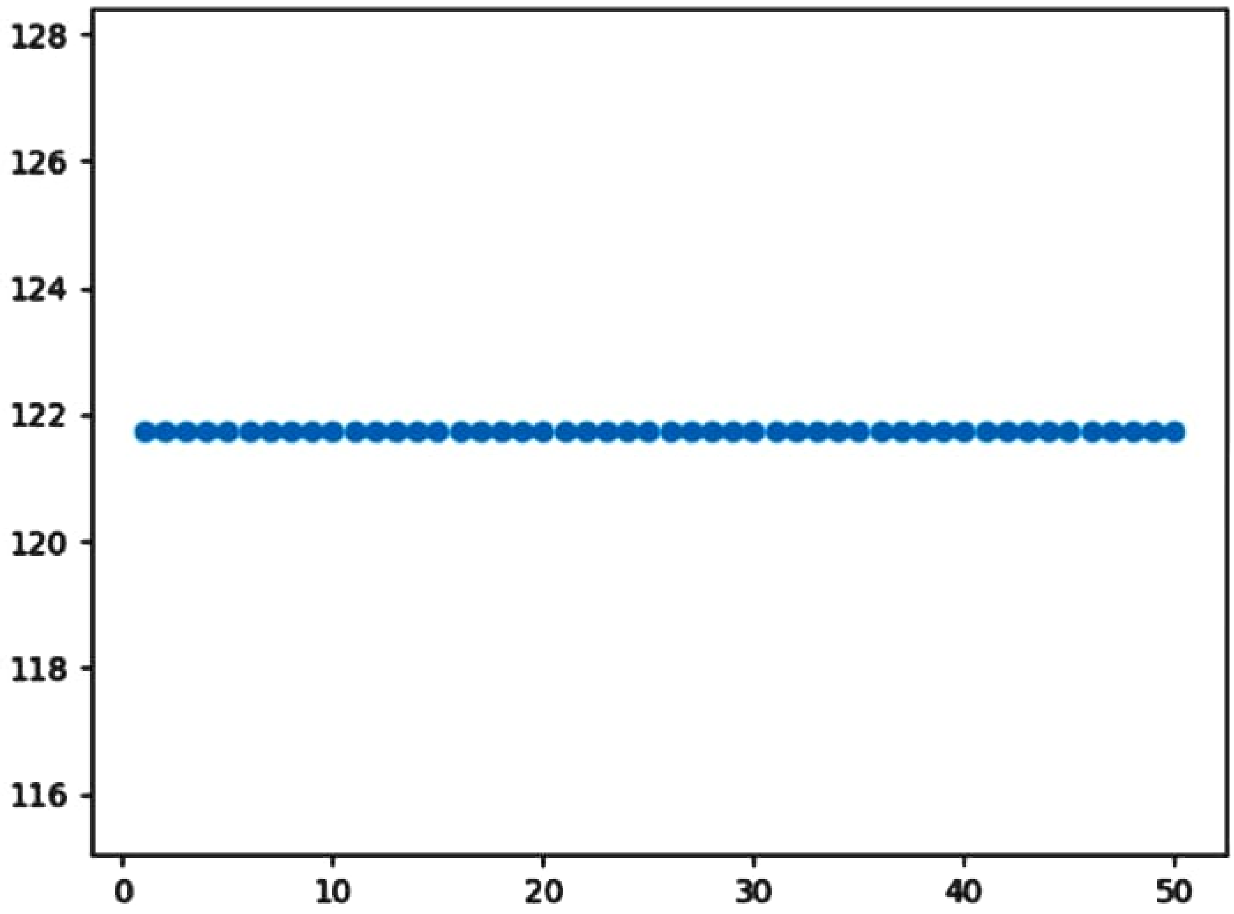
calculated average standard deviation= 19.53385560455709



K=50

calculated mean= 121.72830003447028

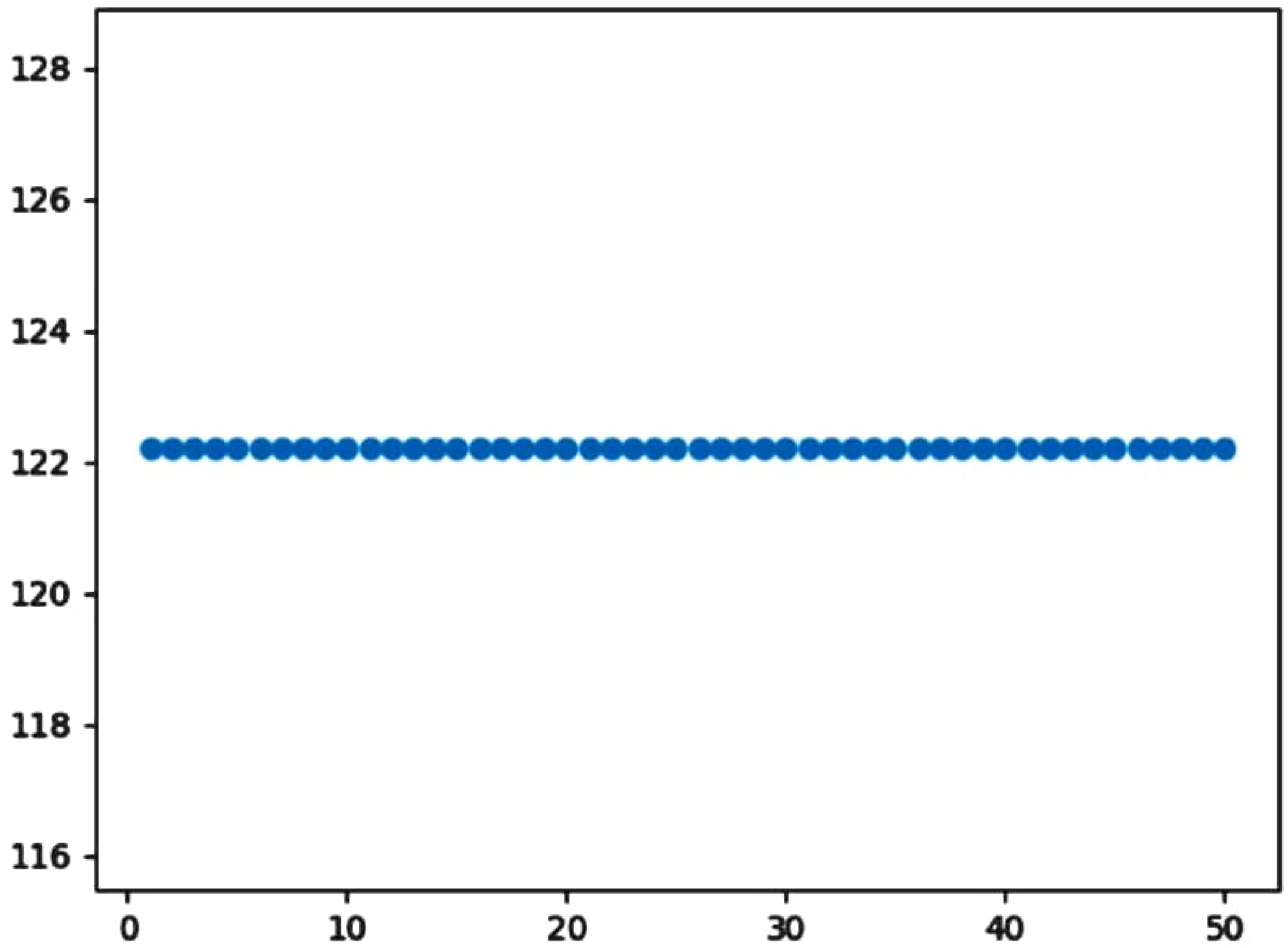
calculated average standard deviation= 20.561181410744524



K=20

calculated mean= 122.1861914235542

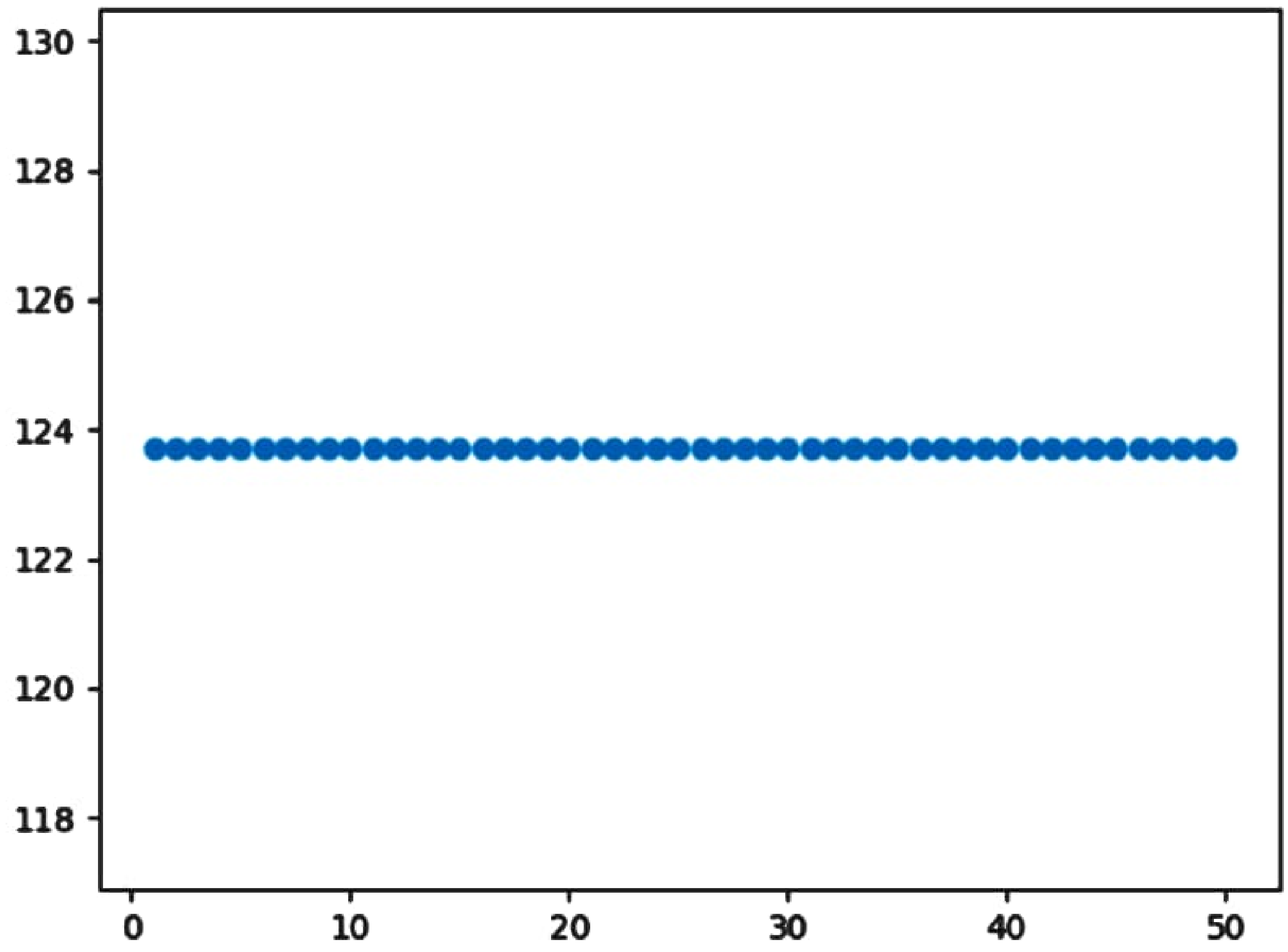
calculated average standard deviation= 18.437072782809306



K=10

calculated mean= 123.70251492505696

calculated average standard deviation= 15.427623599915016



2) As kohli is always picking heads, it means he has in his mind that $p(H) > p(T)$ implying that he thinks $P(H) > 1/2$ and as we have assumed that the coins used is a fair coin.

* Now, let's calculate the problem of below defined events Problem that head comes in n tosses is $\frac{1 \leq n}{100} / \frac{100}{100}$.

So let's divide this into 10 even sets.

From hw, b, txt

$P(1 \to 10) = 0.40$

$P(1 \to 20) = 0.30$

$P(1 \to 30) = 0.33$

$P(1 \to 40) = 0.0275$

$P(1 \to 50) = 0.40$

$P(1 \to 60) = 0.31$

$P(1 \to 70) = 0.32$

$P(1 \to 80) = 0.31$

$P(1 \to 90) = 0.31$

$P(1 \to 100) = 0.3$.

* If we write individual probabilities of tosses, then we can see that after 19th toss, probability is going down from 0.3 and it goes lowest to 0.25. So he might doubt his initial decision after seeing 25th toss.

b) The probability of 25th toss is the least (0.25) but kohli may have in his mind that there are 75 tosses remaining and he maynot lose hope. The probability at 69th toss is 0.318541 which is the least probability in the surroundings. So, it is highly probable that he is make sure by then.

(c)

1) For coin 2 (in h_1b_2.txt), we can see that the probability of each toss is always >0.5

eg:- $P(1) = 1$

$P(10) = 0.6$

$P(30) = 0.67$

$P(60) = 0.633$.

So, there is no need for kohli to doubt his decision and their is no point of being sure.

2) For coin 3 the values are always < 0.5.

He doubt his ~~de~~ decision at the ~~[illegible]~~ 28th toss when the probability is a low at 0.42.

He gets sure at the 37th toss when probabity is 0.43

3) The probabilities were greater than 0.5 but after toss 24 the values fall dow.

He doubts his assumption at the 27th toss ($p = 0.44$).

He gets sure at the 75th toss (~~where~~ $p = 0.45$)

(d). Degree of Sureness

Degree of Sureness = $\dfrac{\text{points in which probability decreases.}}{\text{total no: of points.}}$

* Assume the number of points = 5.

(a) hw_1b_1 degree of sureness = $\frac{5}{5} = 1$

(b) hw_2b_2 degree of sureness = 1
(as probability is always > 0.5)

(c) hw3b3 degree of sureness = $\frac{4}{5}$ = 0.8.

(d) hw4b4 :- degree of sureness = $\frac{4}{5}$ = 0.8.

a). By minimizing the square of distances of each point particular line, say, $\boxed{ax + b = y}$

we can thus find a and b, the data available in hw1c1

The result obtained was

$a = 0.00101$
$b = 51.7805$

$\therefore$ Thus the line is

$$y = 0.00101x + 51.7805$$

b) The result obtained from part a) was used to determine the standard deviation and average value of the distortion produced by using the data from file hw1c2.txt.

The results obtained were :-

$$\text{avg}(y - \hat{y}) = 0.7615$$

$$\text{standard deviation} = 2.2163$$

Both average and standard deviation are important and have significance in their relative space.

Average $(y-\hat{y})$ predicts the expected difference from the previous data.

standard deviation tells how the data is spread out.

INDEX	X	Y(obv)	Y(fitted)	Y(obv)-Y(fitted)
01.	155	48.222	51.9377	-3.71563
02.	155	47.9602	51.9377	-3.97751
03.	155	49.8665	51.9377	-2.07119
04.	155	51.1825	51.9377	-0.75522
05.	155	50.5532	51.9377	-1.38444
06.	155	50.5702	51.9377	-1.36744
07.	155	50.0143	51.9377	-1.92335
08.	155	49.8199	51.9377	-2.11778
09.	155	50.125	51.9377	-1.81269
10.	155	47.9052	51.9377	-4.03243
11.	155	50.2498	51.9377	-1.68787
12.	155	47.2022	51.9377	-4.7355
13.	155	49.9282	51.9377	-2.00951
14.	155	47.7514	51.9377	-4.18629
15.	155	47.5443	51.9377	-4.39341
16.	155	48.3197	51.9377	-3.61794
17.	155	47.0459	51.9377	-4.89173
18.	155	49.0240	51.9377	-2.91371
19.	155	50.6660	51.9377	-1.27168
20.	155	48.6259	51.9377	-3.31173
21.	155	49.7436	51.9377	-2.19412
22.	155	49.2431	51.9377	-2.69461
23.	155	50.8177	51.9377	-1.11994
24.	155	48.7625	51.9377	-3.17514
25.	155	49.4395	51.9377	-2.49821
26.	160	54.2016	51.9427	2.25883
27.	160	52.2194	51.9427	0.276611
28.	160	49.0992	51.9427	-2.84351
29.	160	52.1797	51.9427	0.236948
30.	160	49.1137	51.9427	-2.82908
31.	160	50.5149	51.9427	-1.42788
32.	160	50.9340	51.9427	-1.00869
33.	160	51.8050	51.9427	-0.137735
34.	160	49.9041	51.9427	-2.03863
35.	160	51.3068	51.9427	-0.635953
36.	160	53.1546	51.9427	1.21186
37.	160	47.6418	51.9427	-4.30097
38.	160	48.8849	51.9427	-3.05786
39.	160	52.2924	51.9427	0.349671
40.	160	50.0591	51.9427	-1.88366
41.	160	50.4427	51.9427	-1.50001
42.	160	51.8859	51.9427	-0.056838
43.	160	53.8346	51.9427	1.89185
44.	160	48.5472	51.9427	-3.39558
45.	160	49.1365	51.9427	-2.80623
46.	160	50.3837	51.9427	-1.55899
47.	160	50.5008	51.9427	-1.44197
48.	160	50.6603	51.9427	-1.28246
49.	160	50.0313	51.9427	-1.91144
50.	160	52.3295	51.9427	0.38673
51.	170	52.1357	51.9529	0.182778
52.	170	53.8060	51.9529	1.85309
53.	170	54.9846	51.9529	3.03174
54.	170	54.0628	51.9529	2.10992
55.	170	52.5668	51.9529	0.613931
56.	170	53.6808	51.9529	1.72795
57.	170	53.0239	51.9529	1.07103
58.	170	52.6005	51.9529	0.647594
59.	170	55.5297	51.9529	3.57685
60.	170	53.2035	51.9529	1.25062
61.	170	52.6330	51.9529	0.680117
62.	170	52.3660	51.9529	0.413166
63.	170	53.6818	51.9529	1.72892
64.	170	51.2562	51.9529	-0.696642
65.	170	51.0308	51.9529	-0.922118
66.	170	53.7893	51.9529	1.83647
67.	170	51.5653	51.9529	-0.387534
68.	170	52.9002	51.9529	0.947347
69.	170	54.6631	51.9529	2.71023
70.	170	54.8832	51.9529	2.93035
71.	170	55.8168	51.9529	3.86392
72.	170	51.9857	51.9529	0.032777
73.	170	54.7851	51.9529	2.83218
74.	170	53.9879	51.9529	2.03505
75.	170	55.3787	51.9529	3.42578

Average => 0.7615
Standard deviation => 2.216