Software Measurement - 6611

Assignment – 2

Saswati Chowdhury - 40184906

1.1.

a) Averaging : mean, median, standard deviation :

[Calculated in Excel Sheet]

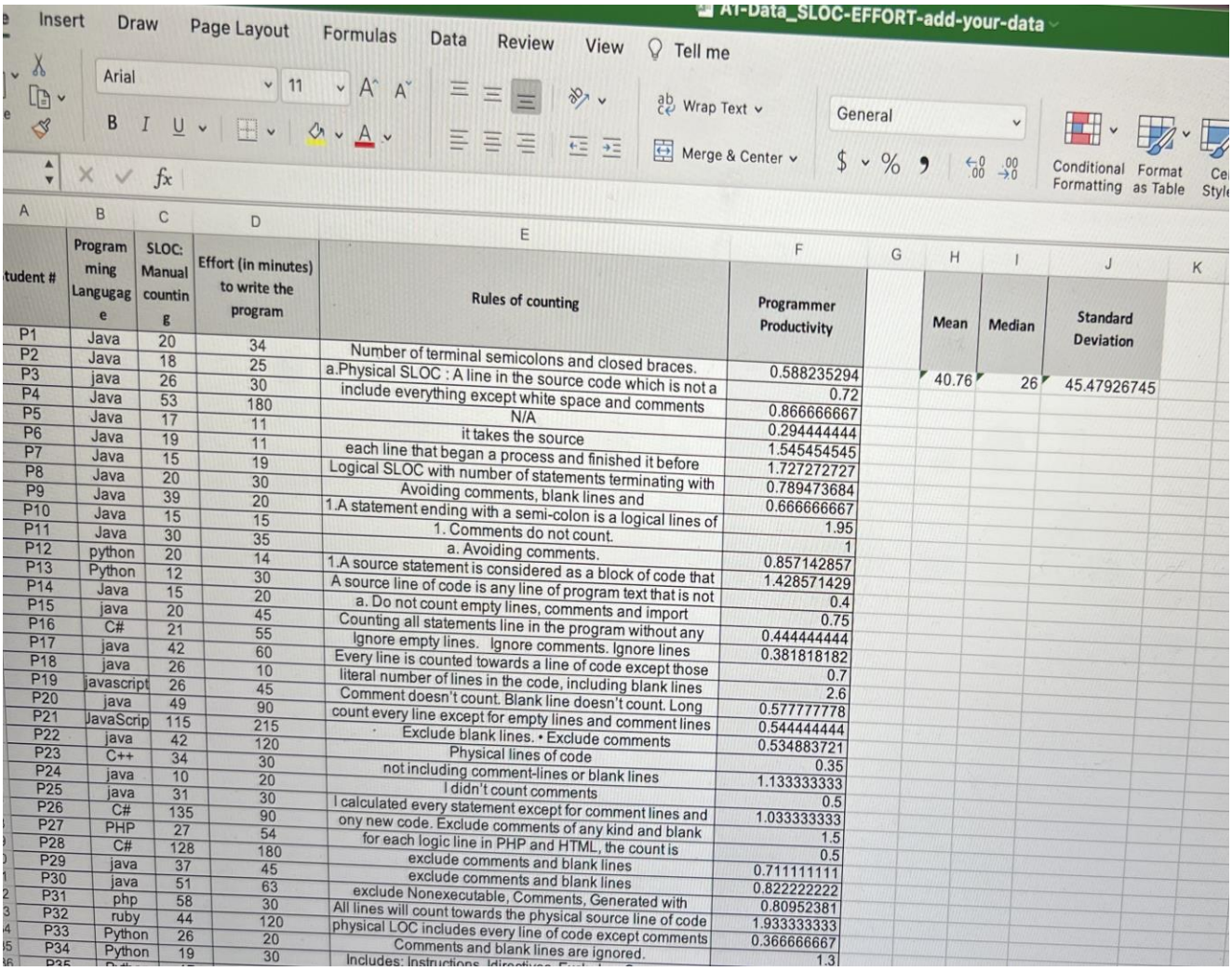
Mean : mean is the numerical average of data set. It is calculated by adding all the SLOC values of the students and then dividing it by the total number of students.

Median : Median is the number that is in the middle of a set of data.

- Arrange the SLOC values in the set in order from least to greatest.
- Then find the number that is in the middle.

Standard Deviation : The standard deviation is the average amount of variability in your dataset. It tells you, on average, how far each value lies from the mean.

Mean	Median	Standard Deviation
40.76	26	45.47926745



tudent #	Programming Langugage	SLOC: Manual counting	Effort (in minutes) to write the program	Rules of counting	Programmer Productivity		Mean	Median	Standard Deviation
P1	Java	20	34	Number of terminal semicolons and closed braces.	0.588235294		40.76	26	45.47926745
P2	Java	18	25	a.Physical SLOC : A line in the source code which is not a	0.72				
P3	java	26	30	include everything except white space and comments	0.866666667				
P4	Java	53	180	N/A	0.294444444				
P5	Java	17	11	it takes the source	1.545454545				
P6	Java	19	11	each line that began a process and finished it before	1.727272727				
P7	Java	15	19	Logical SLOC with number of statements terminating with	0.789473684				
P8	Java	20	30	Avoiding comments, blank lines and	0.666666667				
P9	Java	39	20	1.A statement ending with a semi-colon is a logical lines of	1.95				
P10	Java	15	15	1. Comments do not count.	1				
P11	Java	30	35	a. Avoiding comments.	0.857142857				
P12	python	20	14	1.A source statement is considered as a block of code that	1.428571429				
P13	Python	12	30	A source line of code is any line of program text that is not	0.4				
P14	Java	15	20	a. Do not count empty lines, comments and import	0.75				
P15	java	20	45	Counting all statements line in the program without any	0.444444444				
P16	C#	21	55	Ignore empty lines. Ignore comments. Ignore lines	0.381818182				
P17	java	42	60	Every line is counted towards a line of code except those	0.7				
P18	java	26	10	literal number of lines in the code, including blank lines	2.6				
P19	javascript	26	45	Comment doesn't count. Blank line doesn't count. Long	0.577777778				
P20	java	49	90	count every line except for empty lines and comment lines	0.544444444				
P21	JavaScrip	115	215	Exclude blank lines. • Exclude comments	0.534883721				
P22	java	42	120	Physical lines of code	0.35				
P23	C++	34	30	not including comment-lines or blank lines	1.133333333				
P24	java	10	20	I didn't count comments	0.5				
P25	java	31	30	I calculated every statement except for comment lines and	1.033333333				
P26	C#	135	90	ony new code. Exclude comments of any kind and blank	1.5				
P27	PHP	27	54	for each logic line in PHP and HTML, the count is	0.5				
P28	C#	128	180	exclude comments and blank lines	0.711111111				
P29	java	37	45	exclude comments and blank lines	0.822222222				
P30	java	51	63	exclude Nonexecutable, Comments, Generated with	0.80952381				
P31	php	58	30	All lines will count towards the physical source line of code	1.933333333				
P32	ruby	44	120	physical LOC includes every line of code except comments	0.366666667				
P33	Python	26	20	Comments and blank lines are ignored.	1.3				
P34	Python	19	30	Includes: Instructions, directives, Exec...					

b) Box Plot :

Step 1: Order (increasing) : [set of attribute values (SLOC) obtained from A1_Data_SLOC-Effort-add-your-data]

10, 12, 15, 15, 15, 17, 17, 18, 18, 19, 19, 20, 20, 20, 20, 21, 23, 23, 23, 26, 26, 26, 26, 26, 26, 27, 27, 28, 30, 31, 32, 34, 37, 39, 42, 42, 44, 49, 51, 51, 53, 58, 73, 115, 128, 135, 280

Step 2: Median – 26

Step 3: Lower and Upper Quertiles(Fourths) -

LQ(Lower Forth) = ¼ *(n +1) and round to the nearest integer; the measure with this rank represents the lower quartile.

= round (¼ * 48) -> 12th position

LQ = 20

UQ(Upper Forth) = ¾ *(n +1) and round to the nearest integer; the measure with this rank represents the upper quartile.

= round (¾ * 48) -> 36th position

UQ = 42

Step 4: Box Length - the 'distance' between the lower to upper fourth:

=UQ-LQ

=42-20

=22

Step 5: Upper and Lower Tails:

multiplying the box length by 1.5 (22x1.5=33)

adding and subtracting 9 from the upper and lower forths

Upper Tails : UQ+33

=42+33

=75

Lower Tails: LQ-33

=20-33 -> 0

Here, lower tail is truncated at 0 because a negative number of paths is not meaningful.

Acceptable Range: [Lower forth, Upper forth]

[20..42]

(20, 20, 20, 20, 21, 23, 23, 23, 26, 26, 26, 26, 26, 26, 27, 27, 28, 30, 31, 32, 34, 37, 39, 42, 42)

Quick Review: [lower tail .. lower forth[U]upper forth .. upper tail]

[0..20] U [42..75]

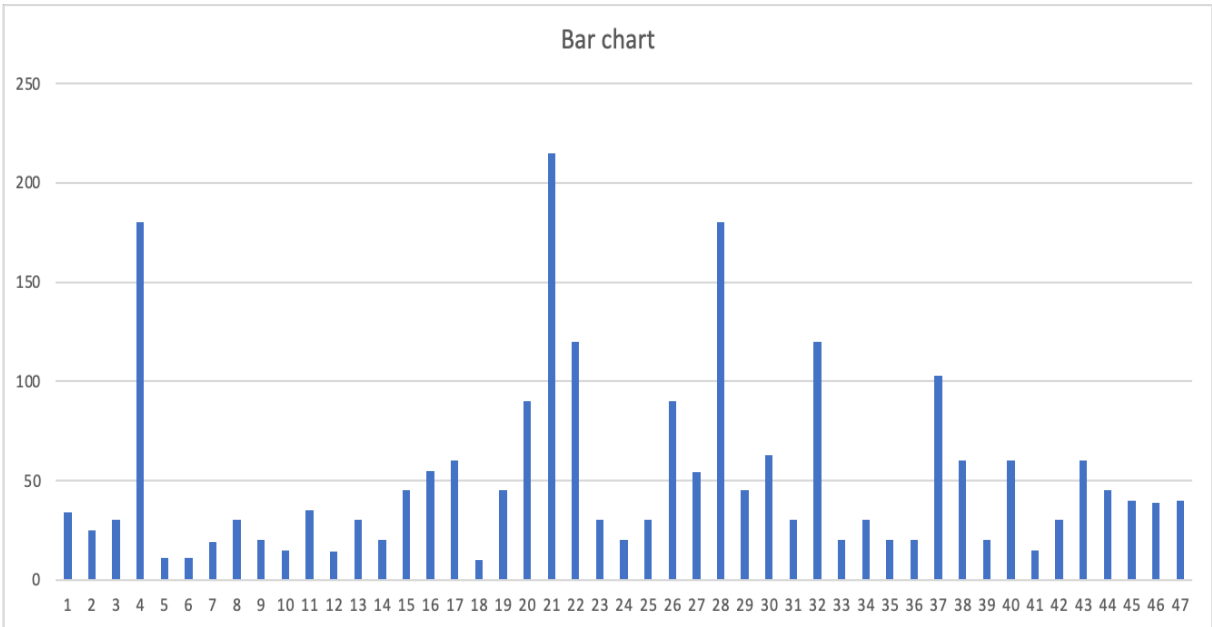
(10, 12, 15, 15, 15, 17, 17, 18, 18, 19, 19, 20, 20, 20, 20, 21, 23, 23, 23, 26, 26, 26, 26, 26, 26, 27, 27, 28, 30, 31, 32, 34, 37, 39, 42, 42, 44, 49, 51, 51, 53, 58, 73)

Range of outliers: Components that are statistical outliers with values greater than the upper tail or less than the lower tail

>75

(115, 128, 135, 280)

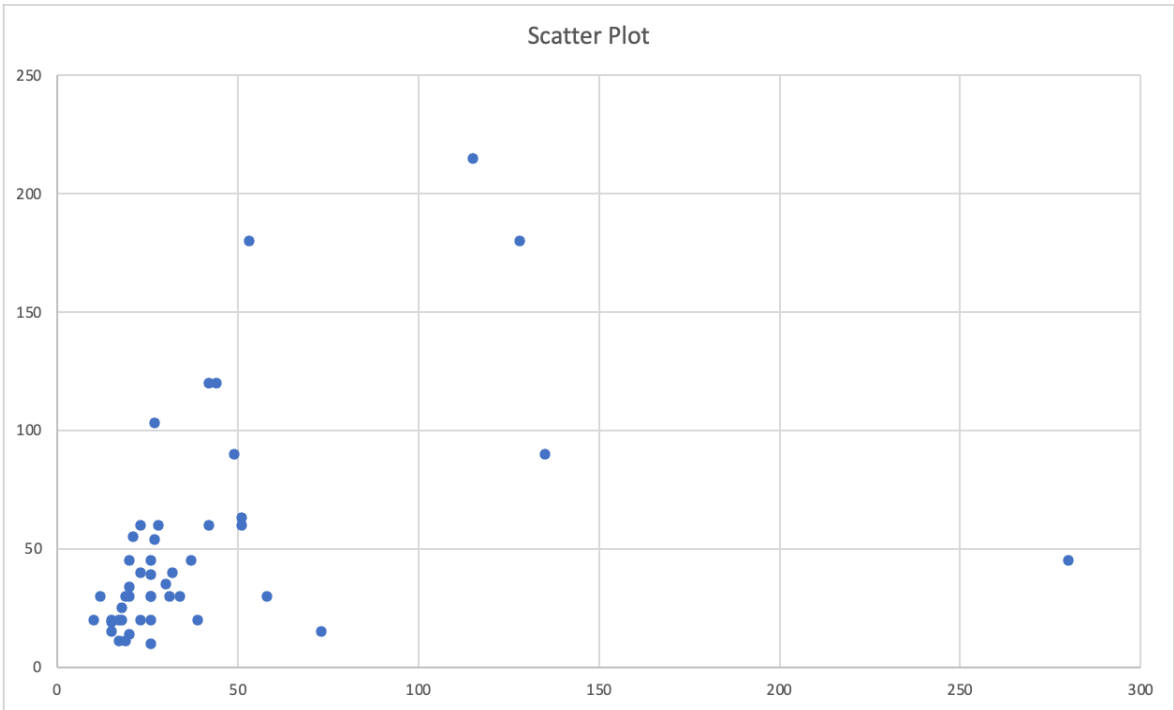
1.2) Bar Chart :



The maximum effort in the excel sheet is shown by the student **P21**. He wrote 115 SLOC in JavaScript in 215 person-minutes. As a result, this student's productivity is 0.53. The minimum effort is by two students (**P5** and **P6**). They both used 11 person-minutes for the programming language Java, and their SLOCs are 17 and 19 respectively. Also there are three students whose effort exceeds 150 person-minutes.

2.1.

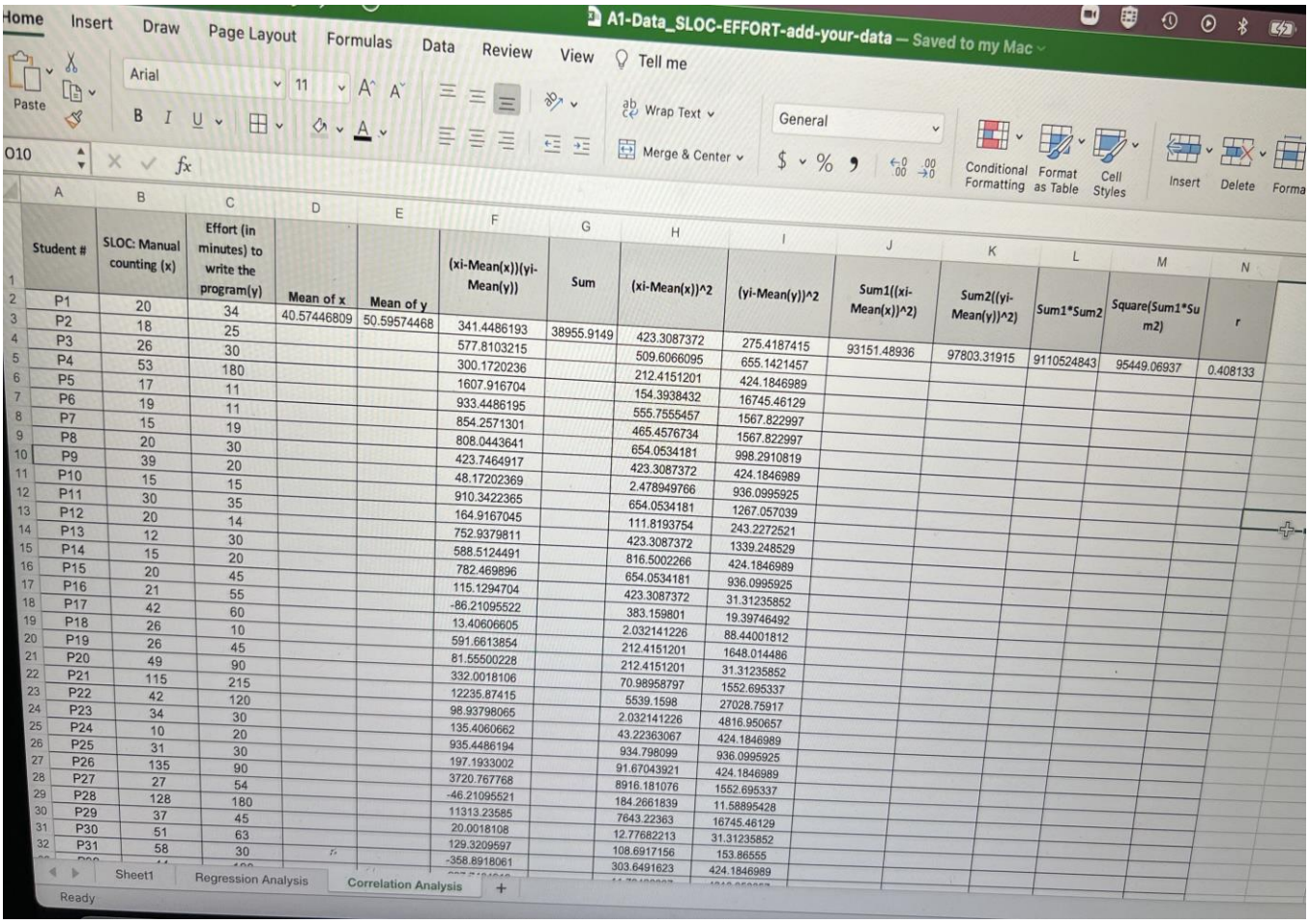
a) Scatter Plot: [attached in the excel sheet]



In this graph I observed 4 data points that are atypical, and those are P21 (115,215), P28 (128,180), P26(135,90), P44 (280,45). These are not organized the same way as the other points.

b) Correlation Analysis:

[calculated in the excel sheet A1-data(Correlation Analysis)]

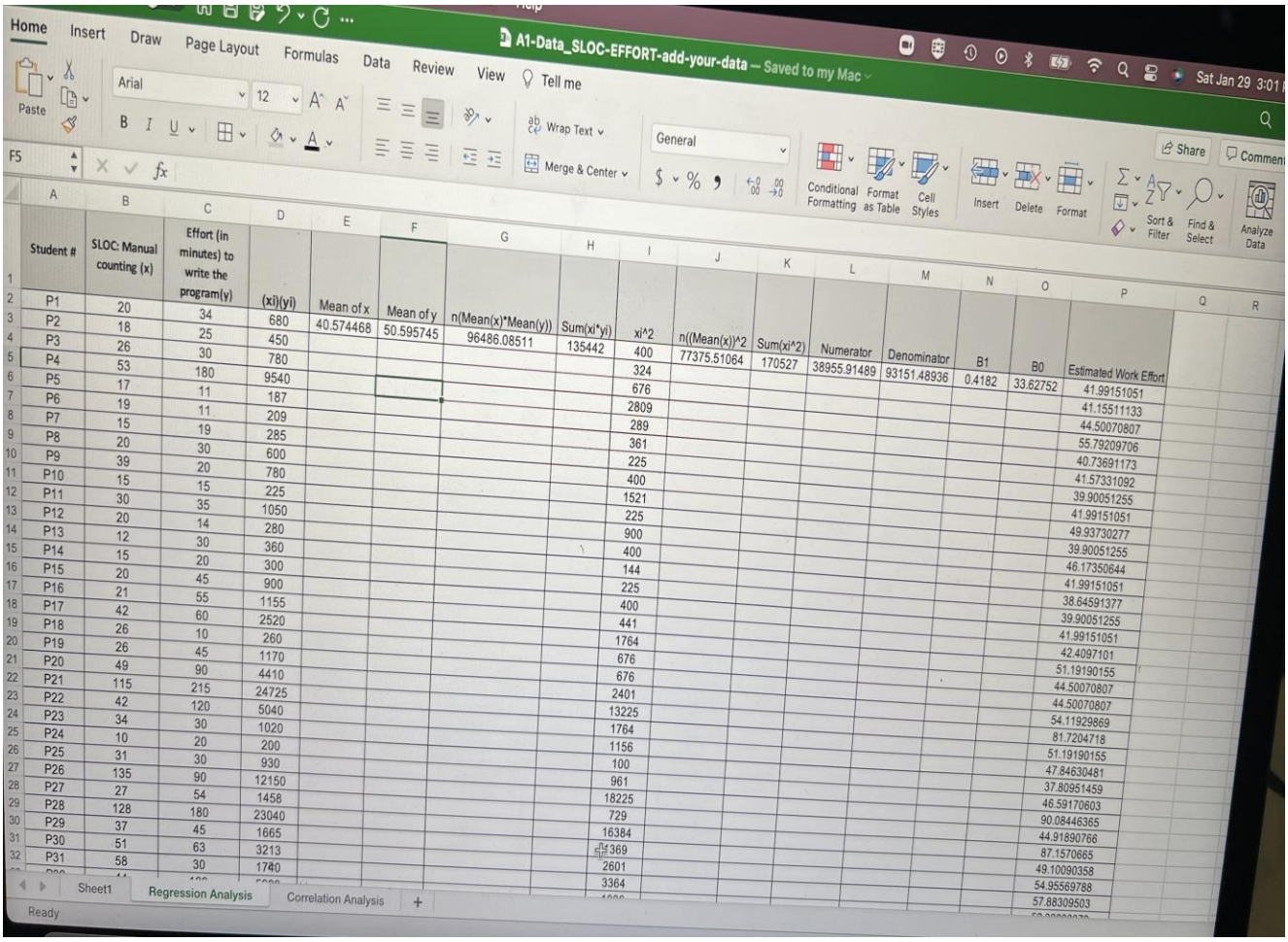


Student #	SLOC: Manual counting (x)	Effort (in minutes) to write the program(y)	Mean of x	Mean of y	(xi-Mean(x))(yi-Mean(y))	Sum	(xi-Mean(x))^2	(yi-Mean(y))^2	Sum1((xi-Mean(x))^2)	Sum2((yi-Mean(y))^2)	Sum1*Sum2	Square(Sum1*Sum2)	r
P1	20	34	40.57446809	50.59574468	341.4486193	38955.9149	423.3087372	275.4187415	93151.48936	97803.31915	9110524843	95449.06937	0.408133
P2	18	25			577.8103215		509.6066095	655.1421457					
P3	26	30			300.1720236		212.4151201	424.1846989					
P4	53	180			1607.916704		154.3938432	16745.46129					
P5	17	11			933.4486195		555.7555457	1567.822997					
P6	19	11			854.2571301		465.4576734	1567.822997					
P7	15	19			808.0443641		654.0534181	998.2910819					
P8	20	30			423.7464917		423.3087372	424.1846989					
P9	39	20			48.17202369		2.478949766	936.0995925					
P10	15	15			910.3422365		654.0534181	1267.057039					
P11	30	35			164.9167045		111.8193754	243.2272521					
P12	20	14			752.9379811		423.3087372	1339.248529					
P13	12	30			588.5124491		816.5002266	424.1846989					
P14	15	20			782.469896		654.0534181	936.0995925					
P15	20	45			115.1294704		423.3087372	31.31235852					
P16	21	55			-86.21095522		383.159801	19.39746492					
P17	42	60			13.40606605		2.032141226	88.44001812					
P18	26	10			591.6613854		212.4151201	1648.014486					
P19	26	45			81.55500228		212.4151201	31.31235852					
P20	49	90			332.0018106		70.98958797	1552.695337					
P21	115	215			12235.87415		5539.1598	27028.75917					
P22	42	120			98.93798065		2.032141226	4816.950657					
P23	34	30			135.4060662		43.22363067	424.1846989					
P24	10	20			935.4486194		934.798099	936.0995925					
P25	31	30			197.1933002		91.67043921	424.1846989					
P26	135	90			3720.767768		8916.181076	1552.695337					
P27	27	54			-46.21095521		184.2661839	11.58895428					
P28	128	180			11313.23585		7643.22363	16745.46129					
P29	37	45			20.0018108		12.77682213	31.31235852					
P30	51	63			129.3209597		108.6917156	153.86555					
P31	58	30			-358.8918061		303.6491623	424.1846989					

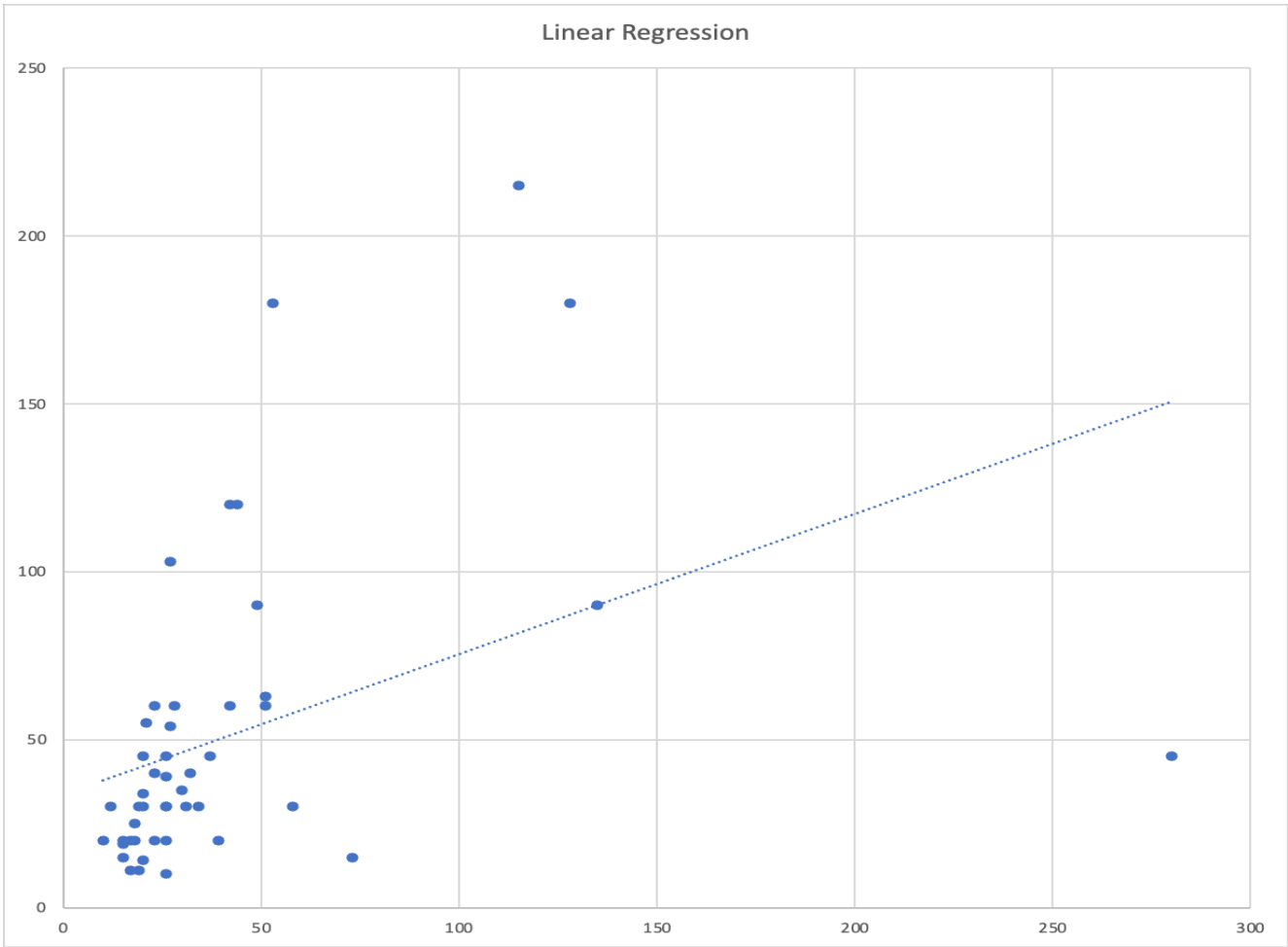
So, correlation coefficient is **0.408133**, which is just greater than 0 but less than 1. Hence, we can say that the correlation between Length and Effort is **not true**.

b) Regression Analysis :

[Calculated in Excel sheet A1-data (Regression Analysis)]



Student #	SLOC: Manual counting (x)	Effort (in minutes) to write the program(y)	(xi)(yi)	Mean of x	Mean of y	n(Mean(x)*Mean(y))	Sum(xi*yi)	xi^2	n((Mean(x))^2	Sum(xi^2)	Numerator	Denominator	B1	B0	Estimated Work Effort
P1	20	34	680	40.574468	50.595745	96486.08511	135442	400	77375.51064	170527	38955.91489	93151.48936	0.4182	33.62752	41.99151051
P2	18	25	450					324							41.15511133
P3	26	30	780					676							44.50070807
P4	53	180	9540					2809							55.79209706
P5	17	11	187					289							40.73691173
P6	19	11	209					361							41.57331092
P7	15	19	285					225							39.90051255
P8	20	30	600					400							41.99151051
P9	39	20	780					1521							49.93730277
P10	15	15	225					225							39.90051255
P11	30	35	1050					900							46.17350644
P12	20	14	280					400							41.99151051
P13	12	30	360					144							38.64591377
P14	15	20	300					225							39.90051255
P15	20	45	900					400							41.99151051
P16	21	55	1155					441							42.4097101
P17	42	60	2520					1764							51.19190155
P18	26	10	260					676							44.50070807
P19	26	45	1170					676							44.50070807
P20	49	90	4410					2401							54.11929869
P21	115	215	24725					13225							81.7204718
P22	42	120	5040					1764							51.19190155
P23	34	30	1020					1156							47.84630481
P24	10	20	200					100							37.80951459
P25	31	30	930					961							46.59170603
P26	135	90	12150					18225							90.08446365
P27	27	54	1458					729							44.91890766
P28	128	180	23040					16384							87.1570665
P29	37	45	1665					1369							49.10090358
P30	51	63	3213					2601							54.95569788
P31	58	30	1740					3364							57.88309503



Regression Equation = $y = \beta 1x + \beta 0$

$\beta 1 = 0.4182$

$\beta 0 = 33.62752$

So, $\hat{y} = 0.4182X + 33.62752$

The relation between the independent variable (SLOC) and the dependent variable (Effort) is **linear.**

2.2) Assumption in effort estimation model :

- Experience of the programmer in the particular programming language
- Clear requirement in order to develop it
- Buffer time added for detecting the bug and resolving it
- Previous experience of the same kind of development
- Size (depends on the requirements. Too many requirements or less?)

3.1) [Calculated in the Excel Sheet – Test A2 data]

Estimated Effort calculation : $= 0.4182X + 33.62752$

$= 0.4182 * SLOC + 33.62752$

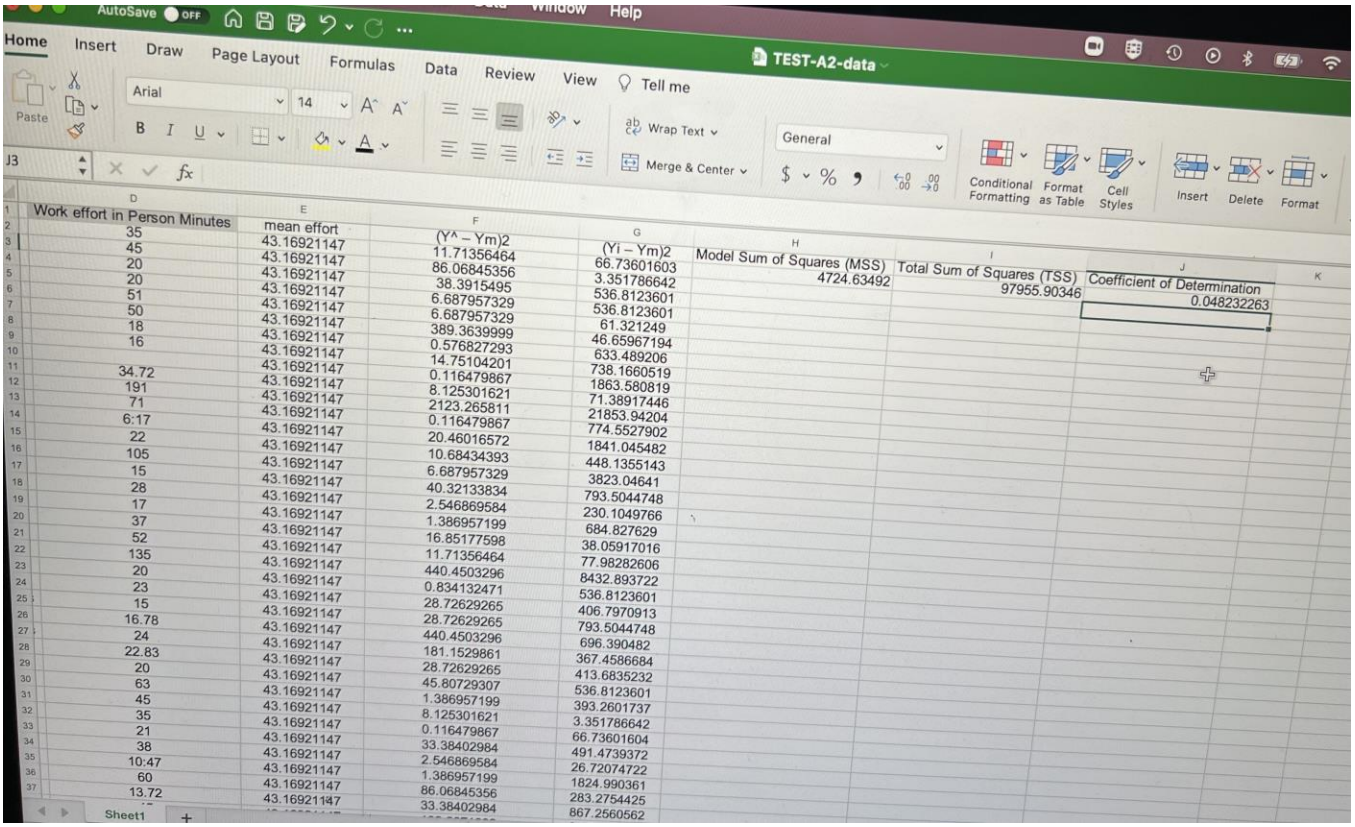
3.2) Coefficient of determination(R-square)

[Calculated in Excel sheet – Test A2]

R-square = MSS/TSS

= 4724.63492/97955.90346

= 0.048232263



Work effort in Person Minutes	mean effort	(Y^ – Ym)2	(Yi – Ym)2	Model Sum of Squares (MSS)	Total Sum of Squares (TSS)	Coefficient of Determination
35	43.16921147	11.71356464	66.73601603	4724.63492	97955.90346	0.048232263
45	43.16921147	86.06845356	3.351786642			
20	43.16921147	38.3915495	536.8123601			
20	43.16921147	6.687957329	536.8123601			
51	43.16921147	6.687957329	61.321249			
50	43.16921147	389.3639999	46.65967194			
18	43.16921147	0.576827293	633.489206			
16	43.16921147	14.75104201	738.1660519			
34.72	43.16921147	0.116479867	1863.580819			
191	43.16921147	8.125301621	71.38917446			
71	43.16921147	2123.265811	21853.94204			
6:17	43.16921147	0.116479867	774.5527902			
22	43.16921147	20.46016572	1841.045482			
105	43.16921147	10.68434393	448.1355143			
15	43.16921147	6.687957329	3823.04641			
28	43.16921147	40.32133834	793.5044748			
17	43.16921147	2.546869584	230.1049766			
37	43.16921147	1.386957199	684.827629			
52	43.16921147	16.85177598	38.05917016			
135	43.16921147	11.71356464	77.98282606			
20	43.16921147	440.4503296	8432.893722			
23	43.16921147	0.834132471	536.8123601			
15	43.16921147	28.72629265	406.7970913			
16.78	43.16921147	28.72629265	793.5044748			
24	43.16921147	440.4503296	696.390482			
22.83	43.16921147	181.1529861	367.4586684			
20	43.16921147	28.72629265	413.6835232			
63	43.16921147	45.80729307	536.8123601			
45	43.16921147	1.386957199	393.2601737			
35	43.16921147	8.125301621	3.351786642			
21	43.16921147	0.116479867	66.73601604			
38	43.16921147	33.38402984	491.4739372			
10:47	43.16921147	2.546869584	26.72074722			
60	43.16921147	1.386957199	1824.990361			
13.72	43.16921147	86.06845356	283.2754425			
	43.16921147	33.38402984	867.2560562			

As R-square value is 0.048232263 which means it is less than 0.5. Hence, this result implies that the relationship is **not reliable** for planning purposes.