Physics Laboratory
(Vp141/Vp241)

Linear Fit Tutorial & Criteria

rev 1.1

Edited by: Yao Yuan

1 Uncertainty for a sample average

If we perform n independent measurements of a physical quantity X, obtaining a set of results $x_1, x_2, \ldots, x_n$, the average value $\bar{X}$ of X is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

and the standard deviation for X may be estimated by the quantity

$$s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}. \tag{2}$$

The standard deviation for $\bar{X}$ is then

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \sqrt{\frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2}.$$

The sample average $\bar{X}$ follows a *Student T Distribution* with **n-1** degrees of freedom (This will be taught in probabilistic courses). Thus, the Type-A uncertainty with 95% confidence is calculated as

$$\Delta_A = t_{n-1,0.95} \cdot \frac{s_X}{\sqrt{n}} = \frac{t_{n-1,0.95}}{\sqrt{n}} \cdot s_X. \tag{3}$$

With Type-B uncertainty Δ_B given, you should be able to calculate **the total uncertainty** $u = \sqrt{\Delta_A^2 + \Delta_B^2}$.

The values of $t_{n-1,0.95}$ and $t_{n-1,0.95}/\sqrt{n}$ are given in Table 1.

n	3	4	5	6	7	8	9	10	15	20
$t_{n-1,0.95}$	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.14	2.09
$\frac{t_{n-1,0.95}}{\sqrt{n}}$	2.48	1.59	1.204	1.05	0.926	0.834	0.770	0.715	0.553	0.467

Table 1: The values of $t_{n-1,0.95}$ and $t_{n-1,0.95}/\sqrt{n}$.

For other values of n that are not included in the table, please refer to the table at the end of this document for (two-sided) t-values, using two parameters **n-1** and **0.95**.

*__Note__ that this is a brief summary for calculation of confidence interval for *sample average*. For more details, please refer to **"handbook-uncertainty analysis.pdf" page 8-9**. Also, this is aim to make it clear that the confidence interval for sample average is different from the confidence interval for *linear regression*, which will be explained in the next section.

2 Uncertainty for a linear fit

Typically, a linear fit is performed using software such as *Origin, Matlab or Microsoft Excel.* The uncertainty(,which leads to a confidence interval) could be directly obtained from the output. Mostly, we would like to know the uncertainty of the slope.

First, we should know that the slope in a simple linear regression follows a *Student T Distribution* with **n-2** degrees of freedom

The values of $t_{n-2,0.95}$ are given in Table 2.

n	3	4	5	6	7	8	9	10	15	20
$t_{n-2,0.95}$	12.71	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.16	2.10

Table 2: The values of $t_{n-2,0.95}$.

For other values of n that are not included in the table, please refer to the table at the end of this document for (two-sided) t-values, using two parameters **n-2** and **0.95**.

Let us consider a set of sample data for a linear fit, where $n = 12$.

t_i [s]	s_i [m]	t_i [s]	s_i [m]
1.0	110	7.0	139
2.0	115	8.0	144
3.0	120	9.0	149
4.0	125	10.0	154
5.0	129	11.0	159
6.0	134	12.0	164

Table 3: Sample data for s-t linear fit.

Time t_i [s]	$\pm$ 0.1 [s]
Distance s_i [m]	$\pm$ 1 [m]

Table 4: Uncertainty of the measurement (used for error bar drawing).

Then $n - 2 = 12 - 2 = 10$ and the $t_{10,0.95} = 2.228$ is read from the table in appendix.

2.1 OriginLab

First, import the data, make sure you have included all X, Y, X error and Y error, otherwise you won't have error bars in the plot.

	A(X)	B(Y)	C(xEr±)	D(yEr±)
Long Name	t	s	t error	s error
Units	[s]	[m]	[s]	[m]
Comments				
F(x)=				
1	1	110	0.1	1
2	2	115	0.1	1
3	3	120	0.1	1
4	4	125	0.1	1
5	5	129	0.1	1
6	6	134	0.1	1
7	7	139	0.1	1
8	8	144	0.1	1
9	9	149	0.1	1
10	10	154	0.1	1
11	11	159	0.1	1
12	12	164	0.1	1

Figure 1: Import data.

Then, select all the data and select Linear Fit in menu.

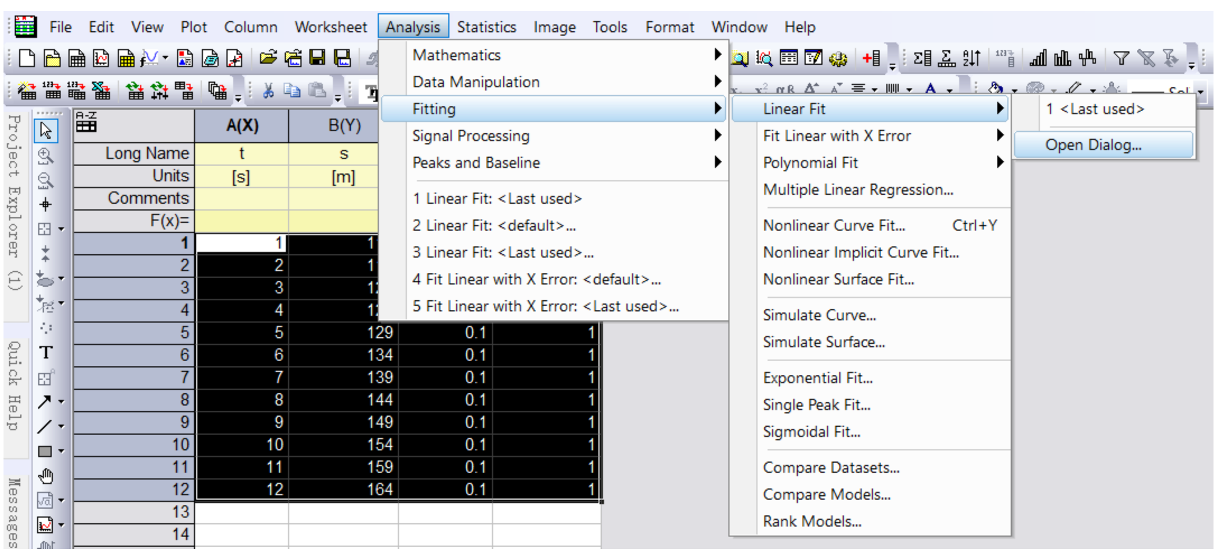


Figure 2: Select linear fit.

In this menu, tick the quantities as is shown in the figure below, including the necessary **Standard Error (Standard Deviation), Lower Confidence Limit (LCL), Upper Confidence Limit(UCL), Pearson's r (Correlation Coefficient) and R-Square.** Some of them will be included in your report if specified.

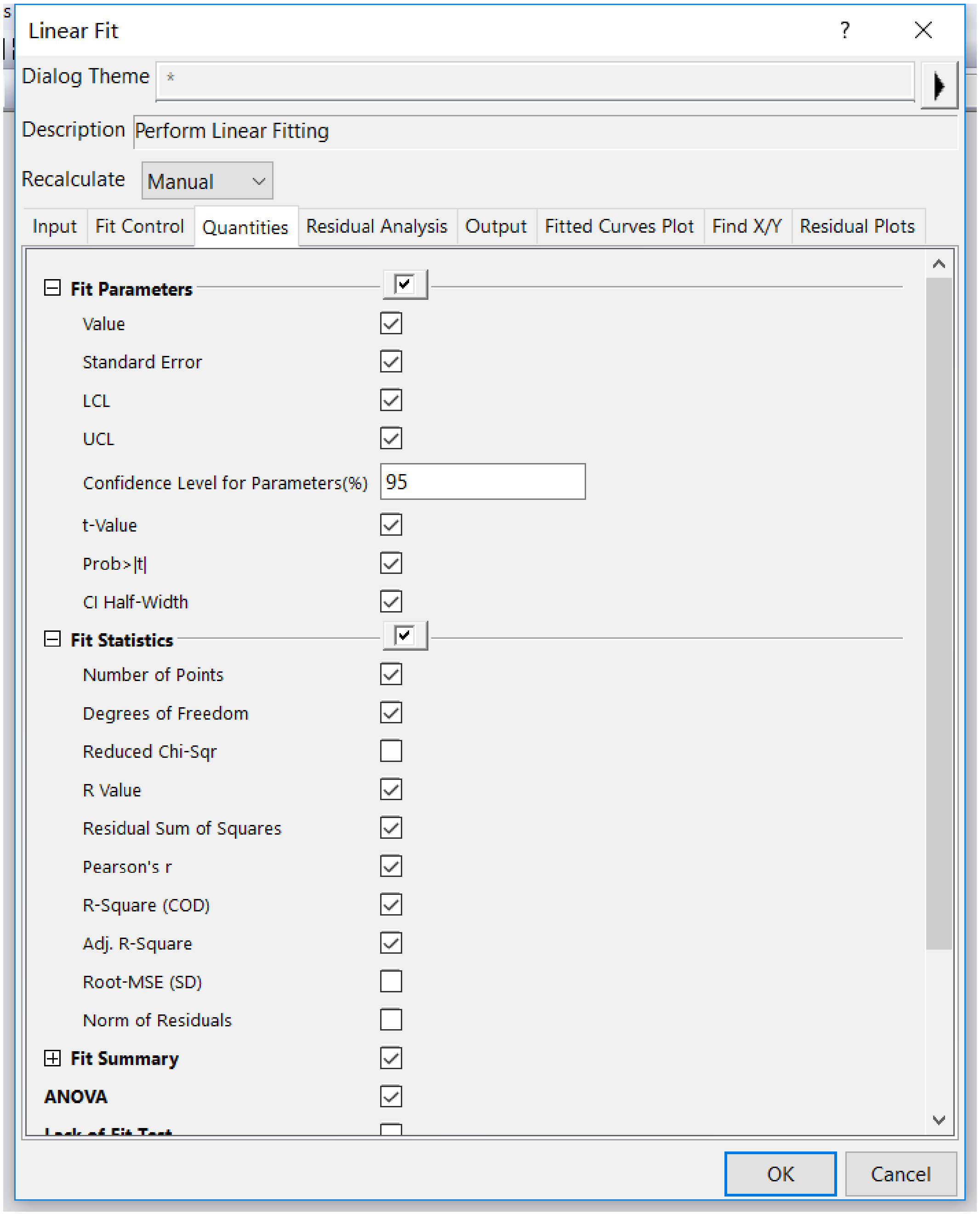
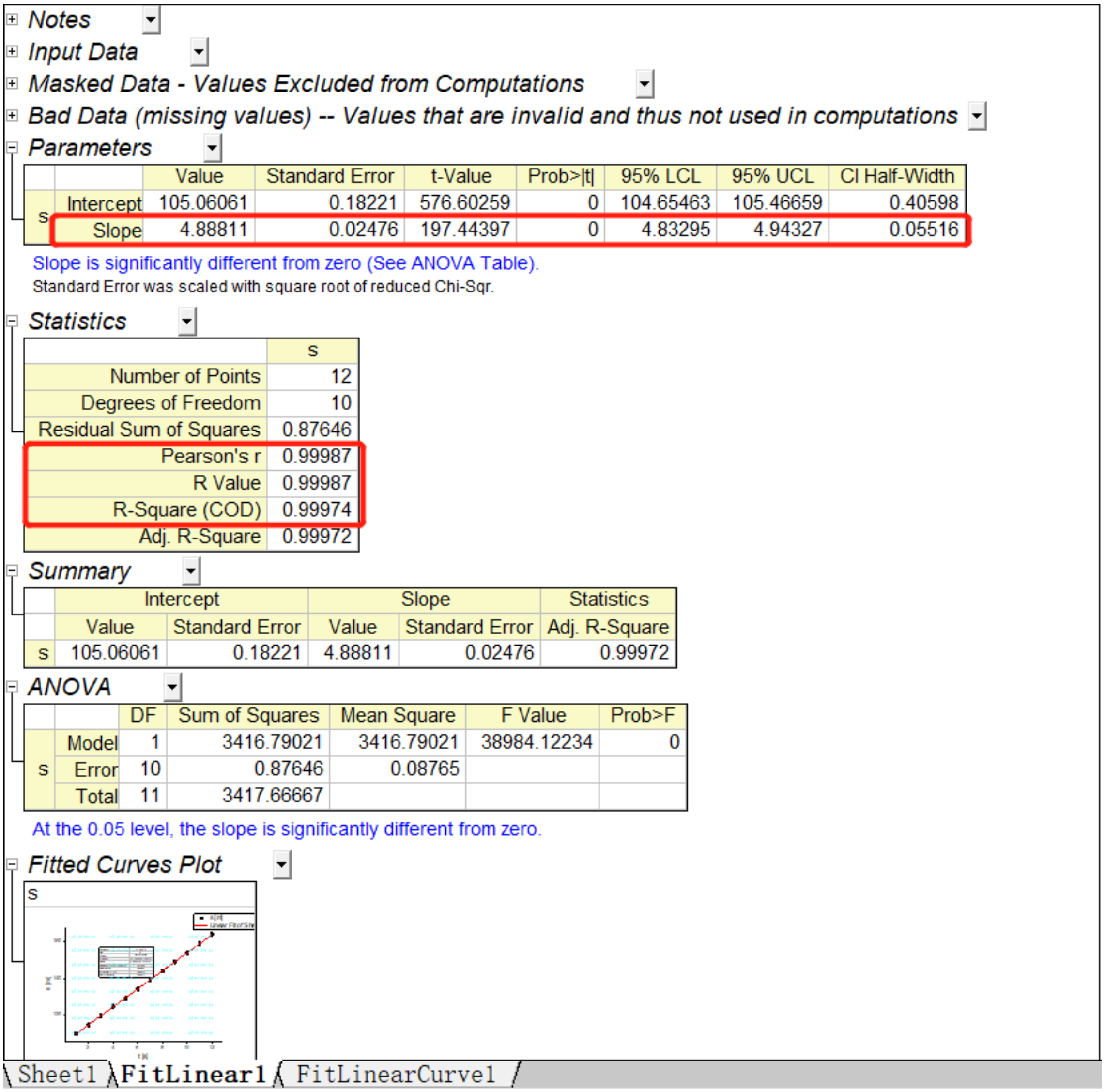


Figure 3: Select quantity display.



Notes

Input Data

Masked Data - Values Excluded from Computations

Bad Data (missing values) -- Values that are invalid and thus not used in computations

Parameters

		Value	Standard Error	t-Value	Prob>|t|	95% LCL	95% UCL	CI Half-Width
s	Intercept	105.06061	0.18221	576.60259	0	104.65463	105.46659	0.40598
	Slope	4.88811	0.02476	197.44397	0	4.83295	4.94327	0.05516

Slope is significantly different from zero (See ANOVA Table).

Standard Error was scaled with square root of reduced Chi-Sqr.

Statistics

	s
Number of Points	12
Degrees of Freedom	10
Residual Sum of Squares	0.87646
Pearson's r	0.99987
R Value	0.99987
R-Square (COD)	0.99974
Adj. R-Square	0.99972

Summary

	Intercept		Slope		Statistics
	Value	Standard Error	Value	Standard Error	Adj. R-Square
s	105.06061	0.18221	4.88811	0.02476	0.99972

ANOVA

		DF	Sum of Squares	Mean Square	F Value	Prob>F
s	Model	1	3416.79021	3416.79021	38984.12234	0
	Error	10	0.87646	0.08765		
	Total	11	3417.66667			

At the 0.05 level, the slope is significantly different from zero.

Fitted Curves Plot

s

Sheet1 FitLinear1 FitLinearCurve1

Figure 4: Summary chart.

Click OK and you will get this table including all the information you need. Here, the **CI**(Confidence Interval) **Half-Width** is the value of **uncertainty**.

You should notice that the uncertainty (0.05516) is the standard error (0.02476) times $t_{10,0.95} = 2.228$. Round the uncertainty to 0.06, and you obtain the uncertainty for the slope.

2.2 Matlab

To do linear fit in Matlab, you need to acquire Curve Fitting in APP. If you don't have this toolbox, click "Acquire more App" on the left and download it(probably need VPN).

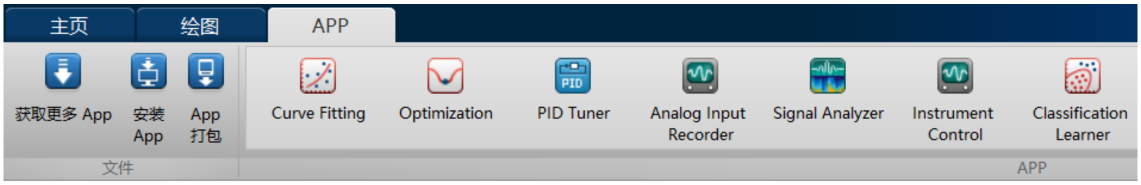
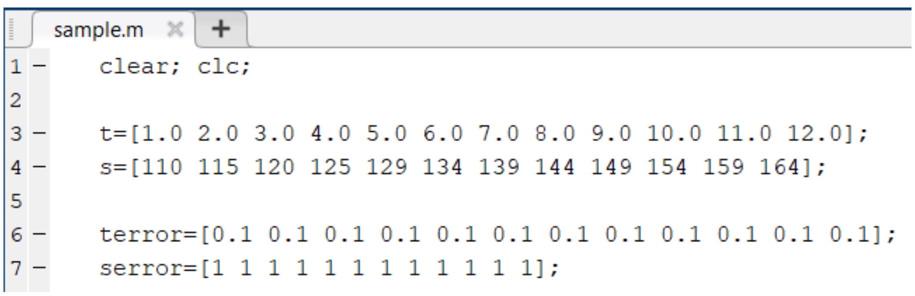


Figure 5: Curve fitting toolbox.

Import your data into two matrix variables.




```
clear; clc;

t=[1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0];
s=[110 115 120 125 129 134 139 144 149 154 159 164];

terror=[0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1];
serror=[1 1 1 1 1 1 1 1 1 1 1 1];
```

Figure 6: Import data.

Select x data and y data and parameters will be shown in the “Results” window.

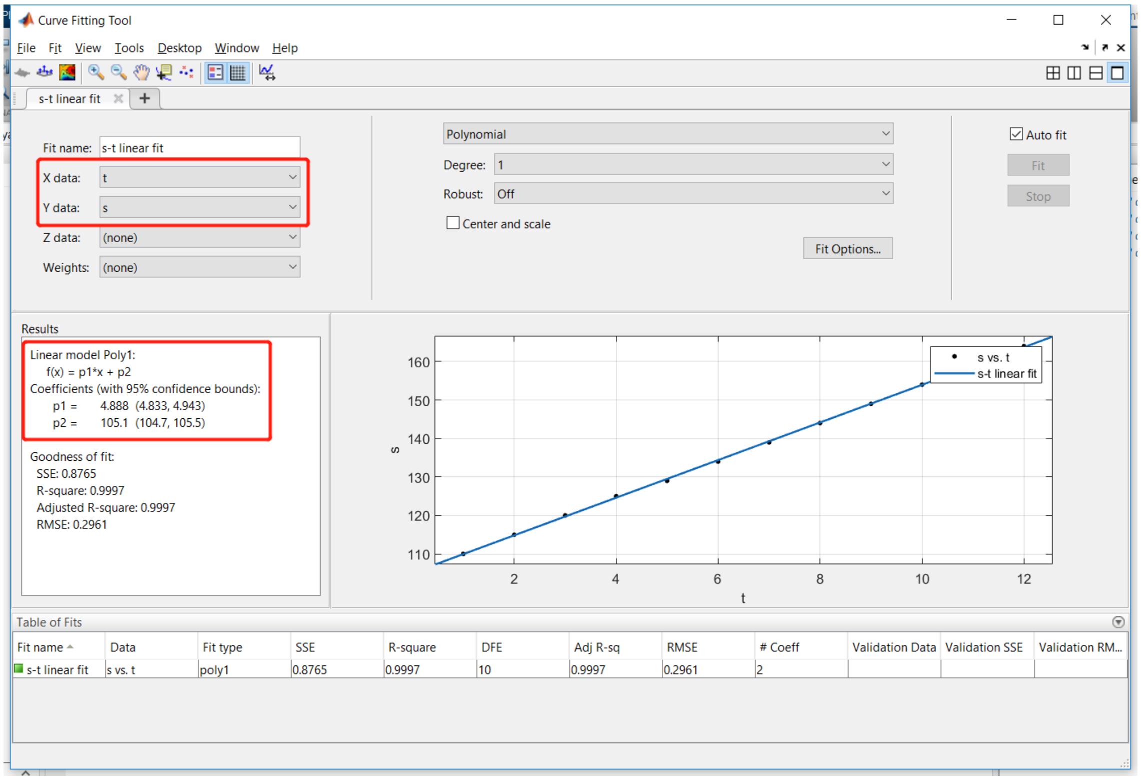


Figure 7: Display settings.

As we can see, “p1” is the slope we need and its 95% confidence is given. Divide the length of confidence interval by 2, we obtain the uncertainty (4.943-4.833)/2=0.055≈0.06.

However, the standard error is not given and we can obtain it by dividing the uncertainty by $t_{10,0.95}$.

Furthermore, error bars can be drawn on the plot if you use properly hold on and function errorbar. The remaining tasks are left to you.

2.3 Excel

First, import all the data you got.

	A	B	C	D
1	t	s	t error	s error
2	1	110	0.1	1
3	2	115	0.1	1
4	3	120	0.1	1
5	4	125	0.1	1
6	5	129	0.1	1
7	6	134	0.1	1
8	7	139	0.1	1
9	8	144	0.1	1
10	9	149	0.1	1
11	10	154	0.1	1
12	11	159	0.1	1
13	12	164	0.1	1

Figure 8: Import data.

The following steps show how to find the Data Analysis toolbox necessary for a linear regression in Excel.

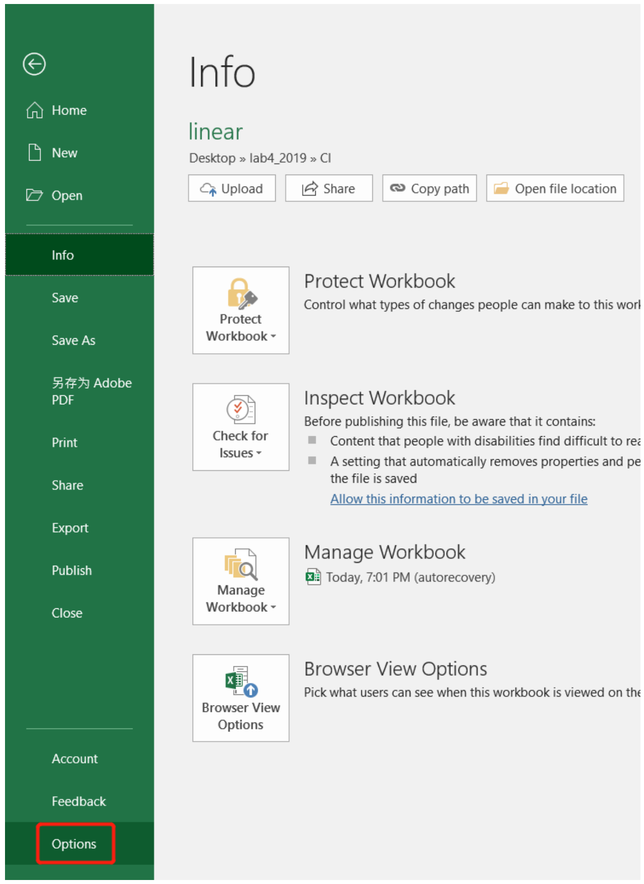


Figure 9: File -> Options.

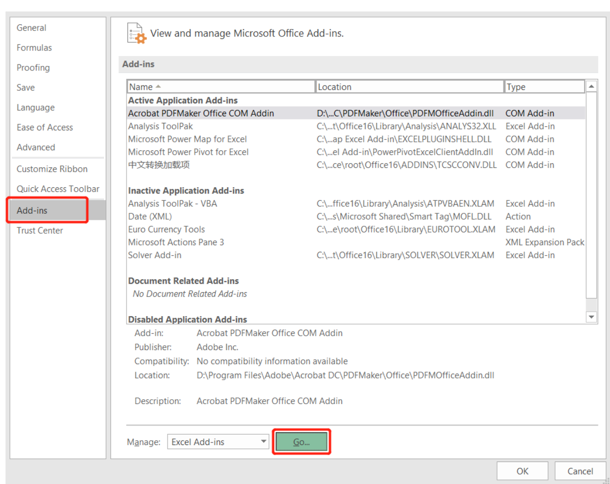


Figure 10: Add-ins -> Go.

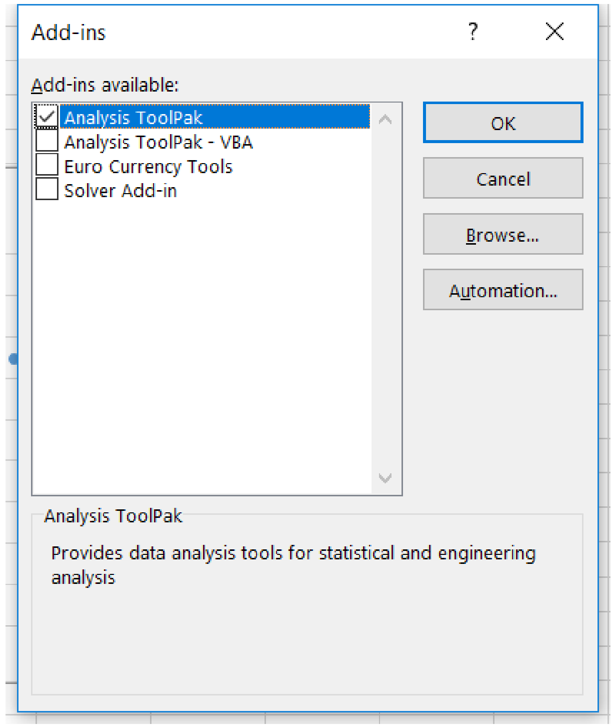


Figure 11: Analysis toolpak.

After this, the toolbox will appear in tab Data.

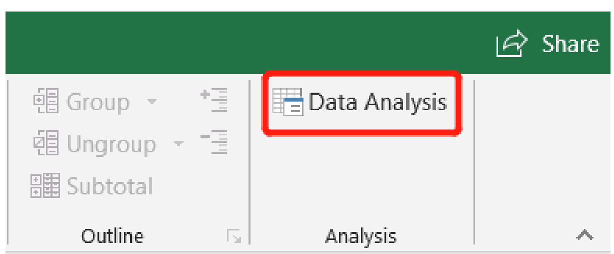


Figure 12: Data analysis.

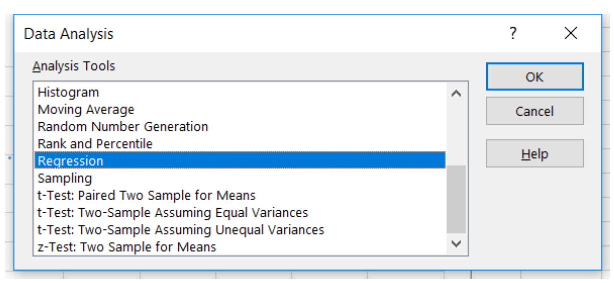
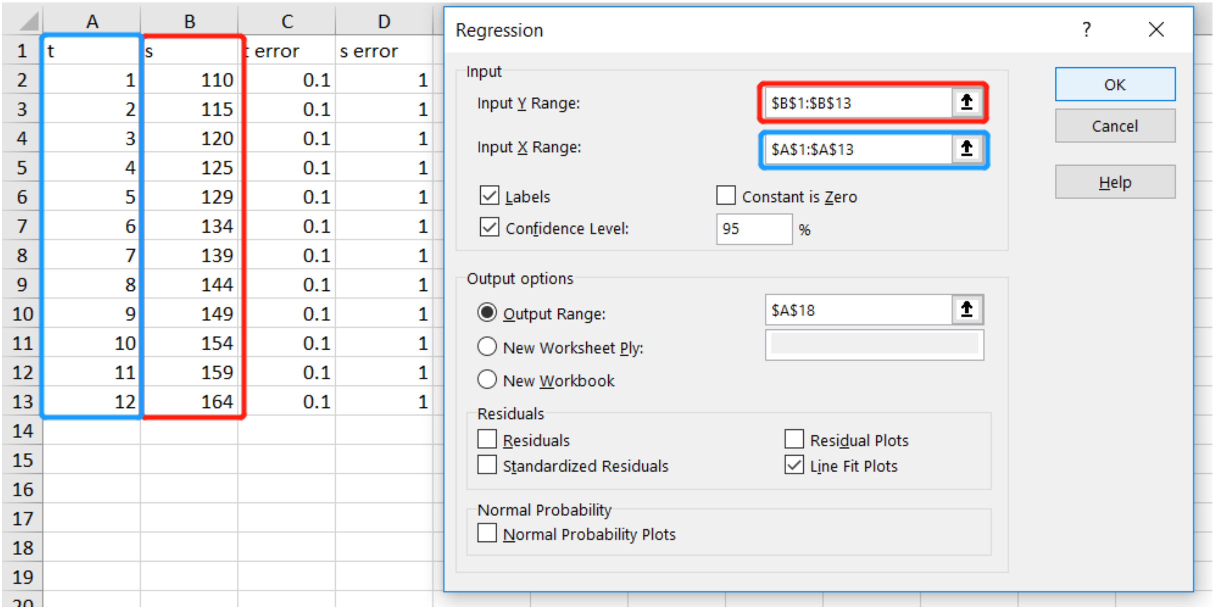


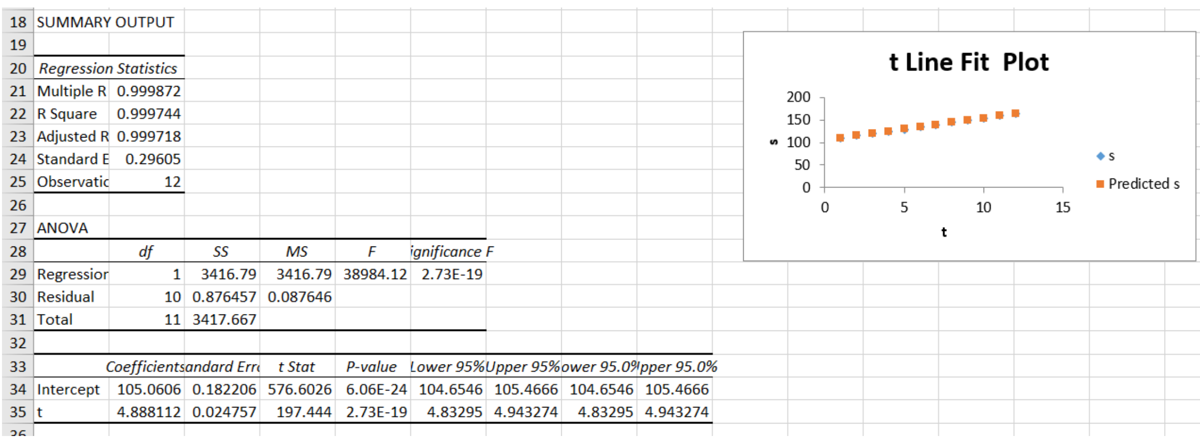
Figure 13: Regression.



	A	B	C	D
1	t	s	t error	s error
2	1	110	0.1	1
3	2	115	0.1	1
4	3	120	0.1	1
5	4	125	0.1	1
6	5	129	0.1	1
7	6	134	0.1	1
8	7	139	0.1	1
9	8	144	0.1	1
10	9	149	0.1	1
11	10	154	0.1	1
12	11	159	0.1	1
13	12	164	0.1	1

Figure 14: Regression settings.

Use this toolbox and you should get all the quantities you need.



18	SUMMARY OUTPUT								
19									
20	*Regression Statistics*								
21	Multiple R	0.999872							
22	R Square	0.999744							
23	Adjusted R	0.999718							
24	Standard E	0.29605							
25	Observatic	12							
26									
27	ANOVA								
28		*df*	*SS*	*MS*	*F*	*ignificance F*			
29	Regressior	1	3416.79	3416.79	38984.12	2.73E-19			
30	Residual	10	0.876457	0.087646					
31	Total	11	3417.667						
32									
33		*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
34	Intercept	105.0606	0.182206	576.6026	6.06E-24	104.6546	105.4666	104.6546	105.4666
35	t	4.888112	0.024757	197.444	2.73E-19	4.83295	4.943274	4.83295	4.943274

Figure 15: Summary output.

Then, the uncertainty can be obtained by

$$u = t_{10,0.95} \cdot \text{“Standard Error”} = 2.228 \times 0.024757 = 0.5516 \approx 0.06.$$

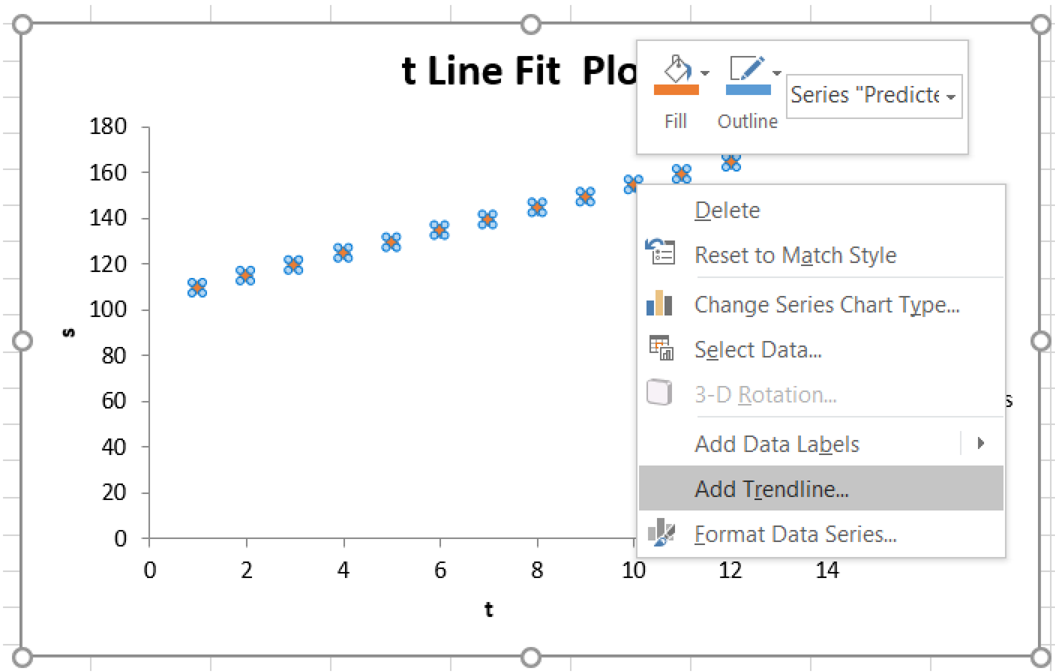


Figure 16: Plot.

Also, you will get a fit line by choosing Add Trendline. The remaining tasks are left to you, including adding error bars, a summary chart and legends in the plot.

2.4 Other Software

For other software, please tell apart standard error and confidence interval(uncertainty). If only one of them is given, calculate another using the same method given above.

3 Sample Plot

The linear fit plot included in your report should look like this. Of course **it should have different looks depending on which software you use**, but please make sure that the elements are included.

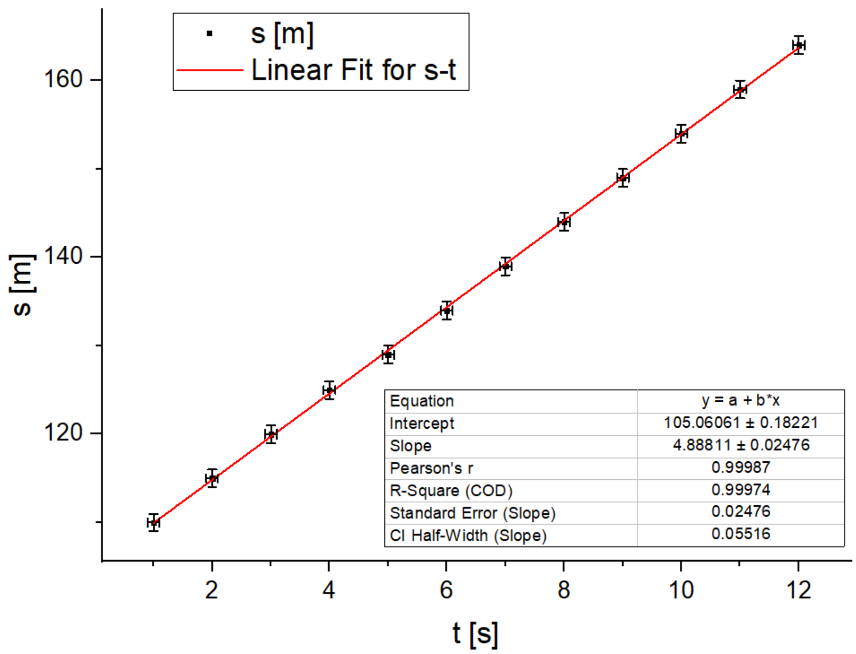


Figure 17: Sample Plot.

Checklist:

1. The **fit line**.
2. **Data points**.
3. **Error bars** in both directions.

(The ones that look like "H". Sometimes one set of data is constant and we only have error bars in the other direction. *Must be visible when printed out!!!*)

4. Proper **legends and labels.**
5. **Summary chart**

   Equation

   Intercept and Slope

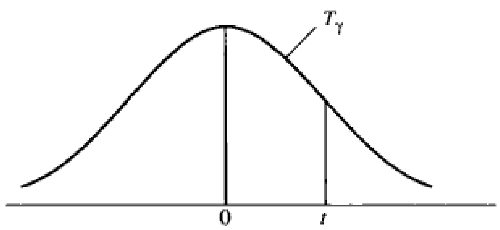
   Standard Error (of the quantity we need)

   Confidence Interval Half-Width (of the quantity we need)

   Pearson's r

   R-Square

4 Appendix



Column heading = cumulative probability

Row heading = degrees of freedom

Row ∞ = standard normal values

	$P[T_\gamma \le t]$								
γ	.6	.75	.9	.95	.975	.99	.995	.999	.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.317	636.607
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.611	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
31	0.256	0.682	1.309	1.696	2.040	2.453	2.744	3.375	3.633
32	0.255	0.682	1.309	1.694	2.037	2.449	2.738	3.365	3.622
33	0.255	0.682	1.308	1.692	2.035	2.445	2.733	3.356	3.611
34	0.255	0.682	1.307	1.691	2.032	2.441	2.728	3.348	3.601
35	0.255	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591
36	0.255	0.681	1.306	1.688	2.028	2.434	2.719	3.333	3.582
37	0.255	0.681	1.305	1.687	2.026	2.431	2.715	3.326	3.574
38	0.255	0.681	1.304	1.686	2.024	2.429	2.712	3.319	3.566
39	0.255	0.681	1.304	1.685	2.023	2.426	2.708	3.313	3.558
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
41	0.255	0.681	1.303	1.683	2.020	2.421	2.701	3.301	3.544
42	0.255	0.680	1.302	1.682	2.018	2.418	2.698	3.296	3.538
43	0.255	0.680	1.302	1.681	2.017	2.416	2.695	3.291	3.532
44	0.255	0.680	1.301	1.680	2.015	2.414	2.692	3.286	3.526

Figure 18: Table for t-values. Two-sided $t_{0.95}$ values are expressed as one-sided $t_{0.975}$ values, which we choose to use.

The table is taken from *Introduction to Probability and Statistics Principles and Application for Engineering and Computer Sciences*.