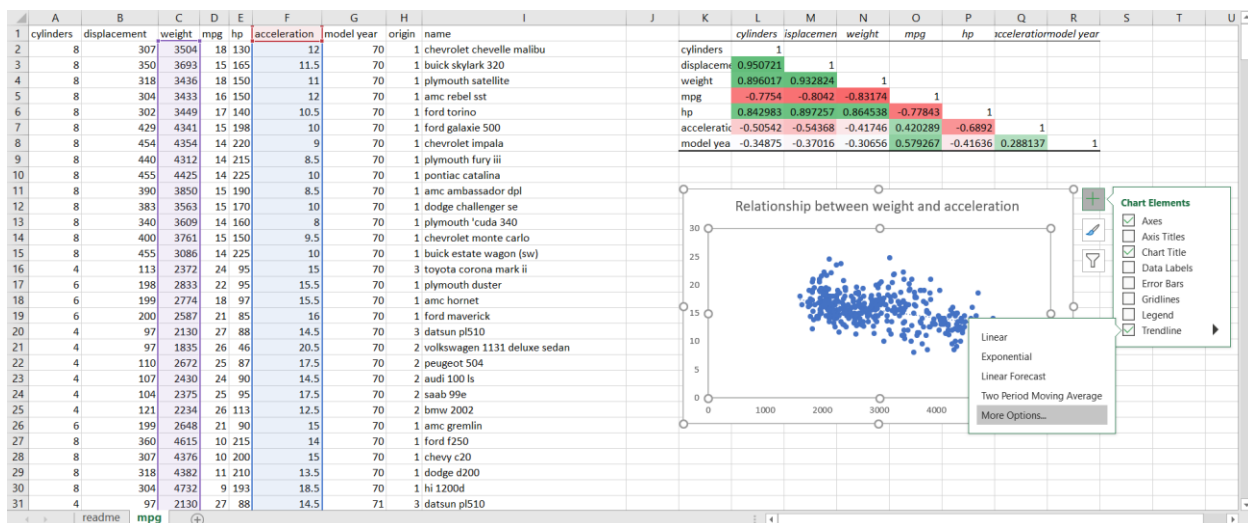


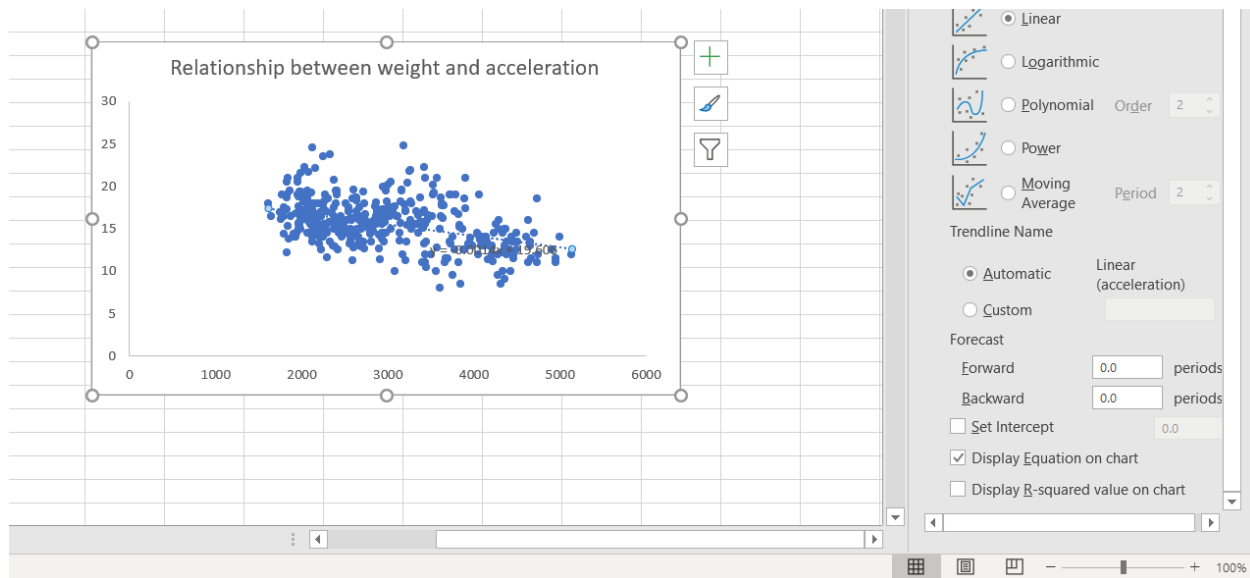
## UP AND RUNNING WITH LINEAR REGRESSION: DEMO NOTES

File: mpg-regression.xlsx

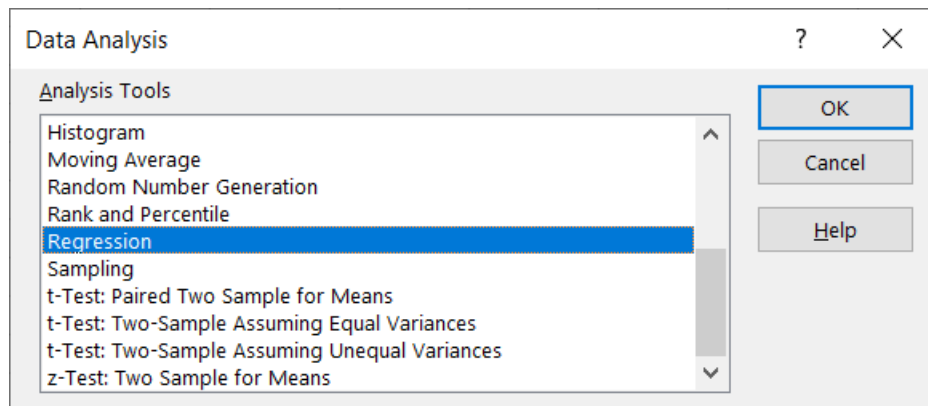
1. The easiest way to find the regression coefficients in Excel is to add a fitted line to a scatter plot:
  - a.



- b. If we click on More Options, we can even insert the regression equation on the chart. Pretty cool, but we can get additional important information by running a full regression model via the ToolPak.

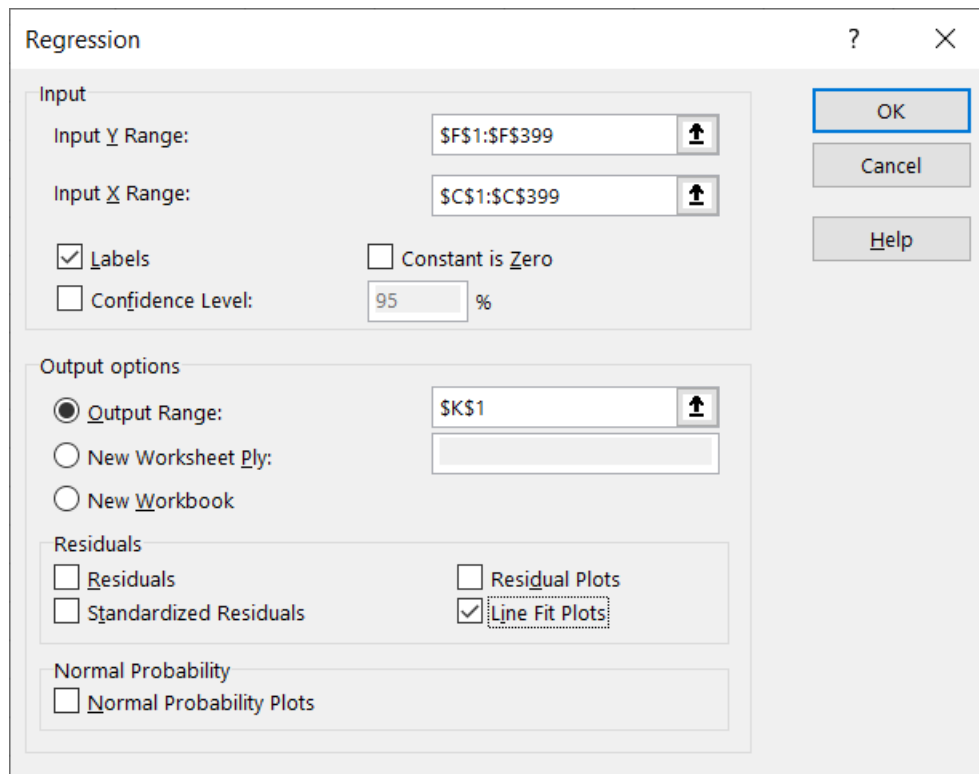


2. This won't give you the statistical significance along with some other key metrics. For that we will select the ToolPak.



- a. Fill it out. Turn line fit on.





The image shows the 'Regression' dialog box in Microsoft Excel. The 'Input' section has 'Input Y Range' set to '\$F\$1:\$F\$399' and 'Input X Range' set to '\$C\$1:\$C\$399'. The 'Labels' checkbox is checked, and 'Confidence Level' is set to '95 %'. The 'Constant is Zero' checkbox is unchecked. The 'Output options' section has 'Output Range' selected with '\$K\$1' entered. 'New Worksheet Ply' and 'New Workbook' are unselected. The 'Residuals' section has 'Residuals' and 'Standardized Residuals' unchecked, while 'Residual Plots' is unchecked and 'Line Fit Plots' is checked. The 'Normal Probability' section has 'Normal Probability Plots' unchecked. On the right, there are 'OK', 'Cancel', and 'Help' buttons.

Regression

Input

Input Y Range: \$F\$1:\$F\$399

Input X Range: \$C\$1:\$C\$399

☒ Labels ☐ Constant is Zero

☐ Confidence Level: 95 %

Output options

☒ Output Range: \$K\$1

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☐ Residual Plots

☐ Standardized Residuals ☒ Line Fit Plots

Normal Probability

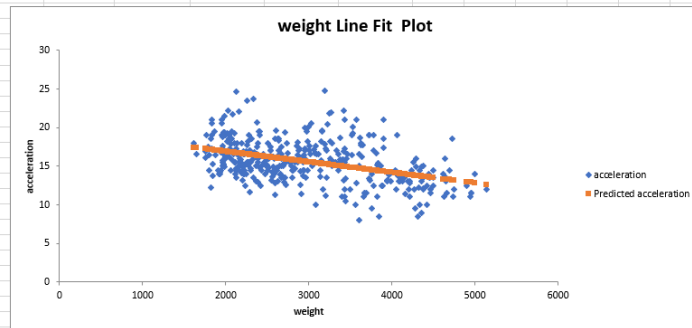
☐ Normal Probability Plots

OK Cancel Help

- b. From here we can see that the slope of the fitted line is significantly different from zero. By turning on the line fit plot, we can also see our predicted values against the actual values.

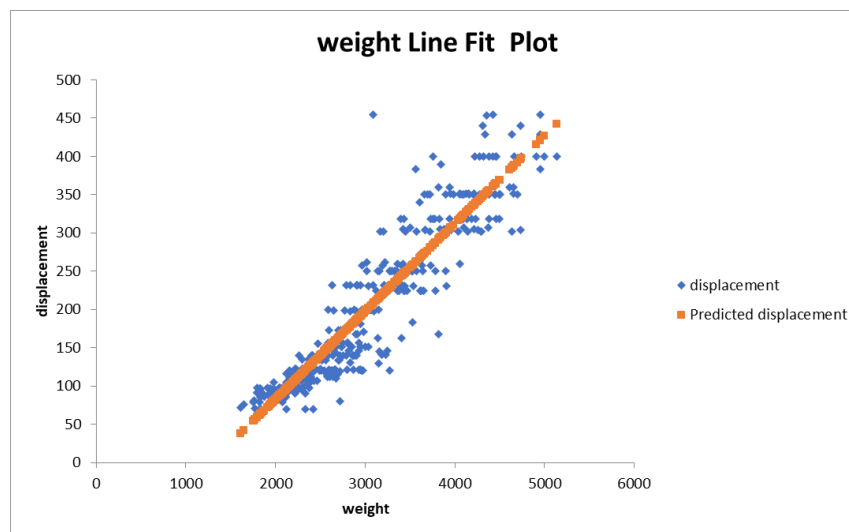


	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1		SUMMARY OUTPUT																
2																		
3		Regression Statistics																
4		Multiple R	0.41745732															
5		R Square	0.174270614															
6		Adjusted R Square	0.172185439															
7		Standard Error	2.509064388															
8		Observations	398															
9																		
10		ANOVA																
11			df	SS	MS	F	Significance F											
12		Regression	1	526.1447236	526.1447236	83.5760048	3.24379E-18											
13		Residual	396	2492.980025	6.295404104													
14		Total	397	3019.124749														
15																		
16			Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%								
17		Intercept	19.60615837	0.45926175	42.69059718	2.949E-150	18.70326235	20.50905439	18.70326235	20.50905439								
18		weight	-0.001359424	0.000148701	-9.14199129	3.2438E-18	-0.00165177	-0.00106708	-0.00165177	-0.00106708								
19																		
20																		
21																		
22		RESIDUAL OUTPUT																
23																		
24			Observation	Predicted acceleration	Residuals													
25			1	14.84273503	-2.842735028													
26			2	14.5858038	-3.085803804													
27			3	14.93517589	-3.935175892													
28			4	14.93925417	-2.939254166													
29			5	14.91750337	-4.417503374													
30			6	13.70489675	-3.704896747													
31			7	13.68722423	-4.687224229													
32			8	13.74432006	-5.244320057													
33			9	13.59070509	-3.590705092													
34			10	14.37237416	-5.872374162													
35			11	14.76252898	-4.762528985													
36			12	14.69999546	-6.699995459													
37			13	14.49336294	-4.99336294													
38			14	15.41097446	-5.410974457													
39			15	16.38160353	-1.381603528													
40			16	15.75490885	-0.254908848													
41			17	15.83511489	-0.335114891													
42			18	16.08932727	-0.089327267													



File: mpg-regression-drill-solutions.xlsx

1. Keep in mind that linear regression assumes no influential cases. From visual inspection, there appears to be at least one influential case in this analysis.
  - a. Spotting and adjusting for influential cases will be addressed in the next course in this sequence.



File: mpg-regression-diagnostics.xlsx

To increase our understanding and use of the model, let's do two more things with our model.

1. R-square: This tells us what percent of our variability in Y is explained by X. We can get the r-square from the results of the ToolPak.
  - a. In this case, 17% of variability in acceleration is explained by weight.

	J	K	L	M	N	O	P	Q
1		SUMMARY OUTPUT						
2								
3		<i>Regression Statistics</i>						
4		Multiple R	0.41745732					
5		R Square	0.174270614	<- 17% of the variability in acceleration is explained by weight				
6		Adjusted R Square	0.172185439					
7		Standard Error	2.509064388					
8		Observations	398					
9								

2. Making point predictions: Using the intercept and slope of our regression line, we can predict a value for Y given a value for X.
  - a. This is done by adding our intercept to the product of the slope and our value for X.

	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1		SUMMARY OUTPUT												
2														
3		<i>Regression Statistics</i>												
4		Multiple R	0.41745732											
5		R Square	0.174270614	<- 17% of the variability in acceleration is explained by weight										
6		Adjusted R Square	0.172185439											
7		Standard Error	2.509064388											
8		Observations	398											
9														
10		<i>ANOVA</i>												
11			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>		<i>X</i>	<i>Y</i>				
12		Regression	1	526.1447236	526.1447236	83.5760048	3.24379E-18		3000	=L17+(R12*L18)				
13		Residual	396	2492.980025	6.295404104									
14		Total	397	3019.124749										
15														
16		<i>Coefficients</i>		<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>				
17		Intercept	19.60615837	0.45926175	42.69059718	2.9489E-150	18.70326235	20.50905439	18.70326235	20.50905439				
18		weight	-0.001359424	0.000148701	-9.141991293	3.24379E-18	-0.001651767	-0.001067082	-0.001651767	-0.001067082				
19														
20														

- a.
  - b. In this example, we predict that a car weighing 3,000 pounds will accelerate from 0-60 in 15.5 seconds.



