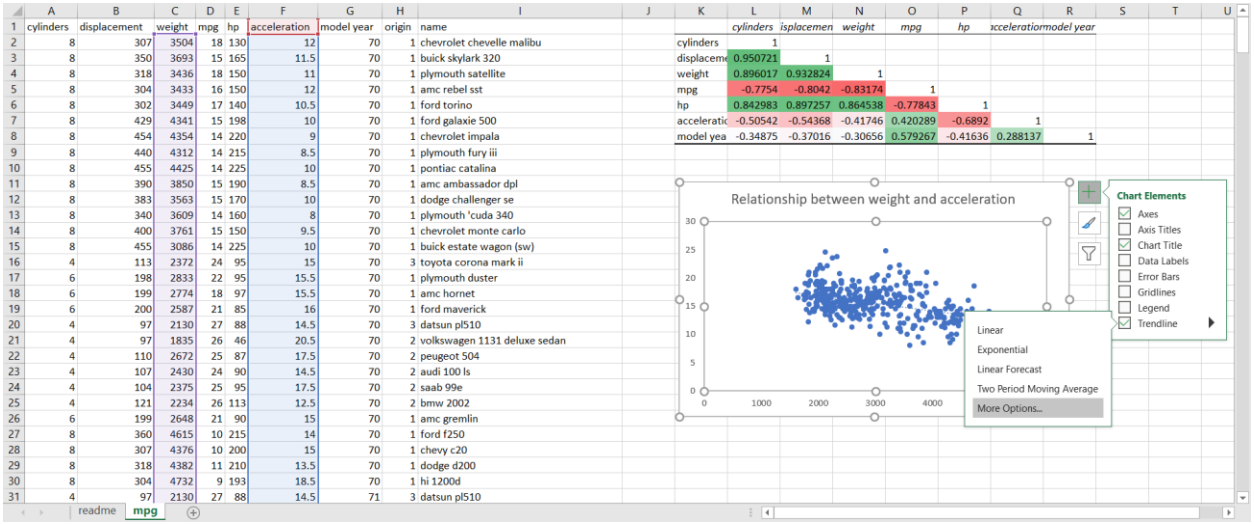
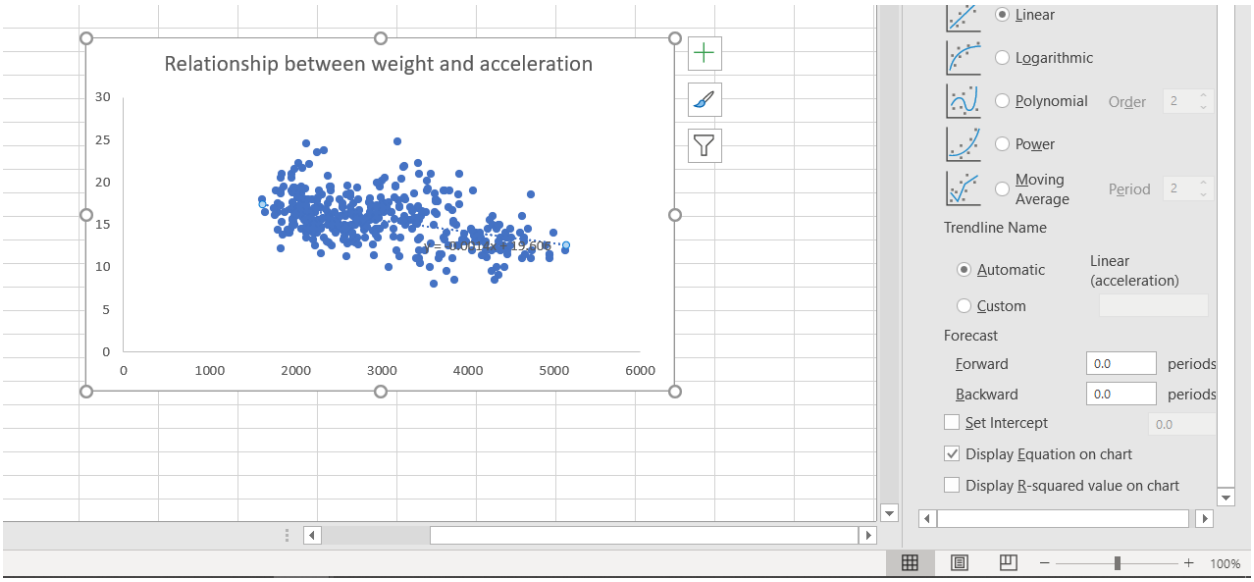
# Up and running with linear regression: demo notes

File: `mpg-regression.xlsx`

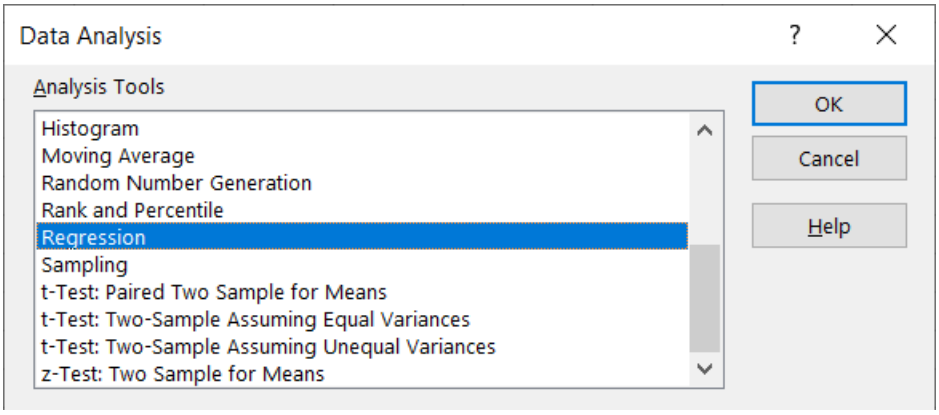
1. The easiest way to find the regression coefficients in Excel is to add a fitted line to a scatter plot:
   a.



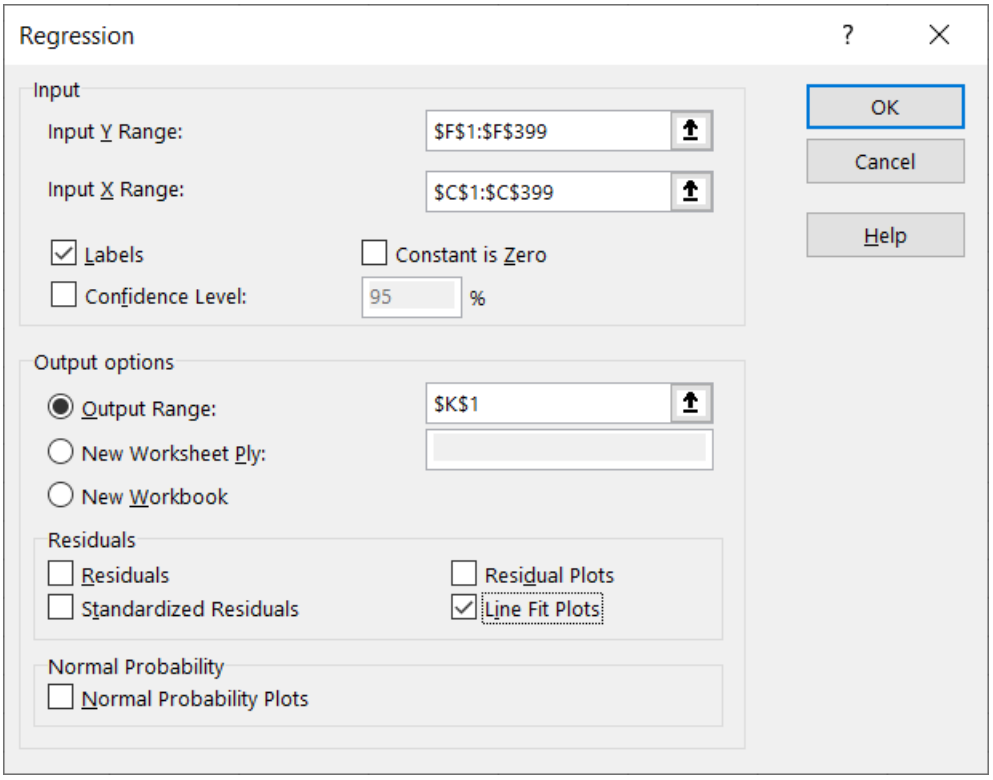

   b. If we click on More Options, we can even insert the regression equation on the chart. Pretty cool, but we can get additional important information by running a full regression model via the ToolPak.




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.




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.

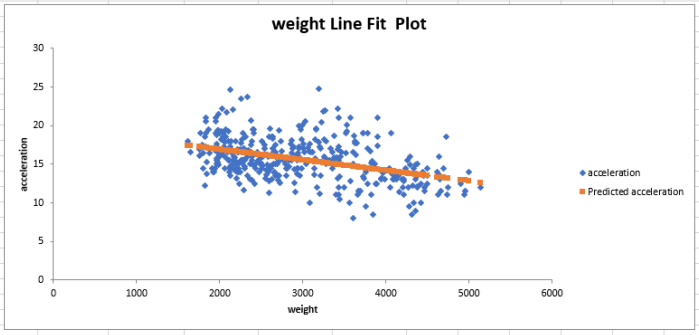
SUMMARY OUTPUT

| Regression Statistics | |
|---|---|
| Multiple R | 0.41745732 |
| R Square | 0.174270614 |
| Adjusted R Square | 0.172185439 |
| Standard Error | 2.509064388 |
| Observations | 398 |

ANOVA

| | df | SS | MS | F | Significance F |
|---|---|---|---|---|---|
| Regression | 1 | 526.1447236 | 526.1447236 | 83.5760048 | 3.24379E-18 |
| Residual | 396 | 2492.980025 | 6.295404104 | | |
| Total | 397 | 3019.124749 | | | |

| | Coefficients | Standard Error | t Stat | P-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|---|---|---|---|---|---|---|---|
| Intercept | 19.60615837 | 0.45926175 | 42.69059718 | 2.949E-150 | 18.70326235 | 20.50905439 | 18.70326235 | 20.50905439 |
| weight | -0.001359424 | 0.000148701 | -9.14199129 | 3.2438E-18 | -0.00165177 | -0.00106708 | -0.00165177 | -0.00106708 |




RESIDUAL OUTPUT

| Observation | Predicted acceleration | Residuals |
|---|---|---|
| 1 | 14.84273503 | -2.842735028 |
| 2 | 14.5858038 | -3.085803804 |
| 3 | 14.93517589 | -3.935175892 |
| 4 | 14.93925417 | -2.939254166 |
| 5 | 14.91750337 | -4.417503374 |
| 6 | 13.70489675 | -3.704896747 |
| 7 | 13.68722423 | -4.687224229 |
| 8 | 13.74432006 | -5.244320057 |
| 9 | 13.59070509 | -3.590705092 |
| 10 | 14.37237416 | -5.872374162 |
| 11 | 14.76252898 | -4.762528985 |
| 12 | 14.69999546 | -6.699995459 |
| 13 | 14.49336294 | -4.99336294 |
| 14 | 15.41097446 | -5.410974457 |
| 15 | 16.38160353 | -1.381603528 |
| 16 | 15.75490885 | -0.254908848 |
| 17 | 15.83511489 | -0.335114891 |
| 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 | | *Regression Statistics* | | | | | | |
| 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 | | *Regression Statistics* | | | | | | | | | | | | |
| 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 | | ANOVA | | | | | | | | | | | | |
| 11 | | | *df* | *SS* | *MS* | *F* | *Significance F* | | *X* | *Y* | | | | |
| 12 | | Regression | 1 | 526.1447236 | 526.1447236 | 83.5760048 | 3.24379E-18 | | 3000 | =L17+(R12*L18) | | *L18) | | |
| 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 | 2.9489E-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 | | | | | | | | | | | | | | |

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