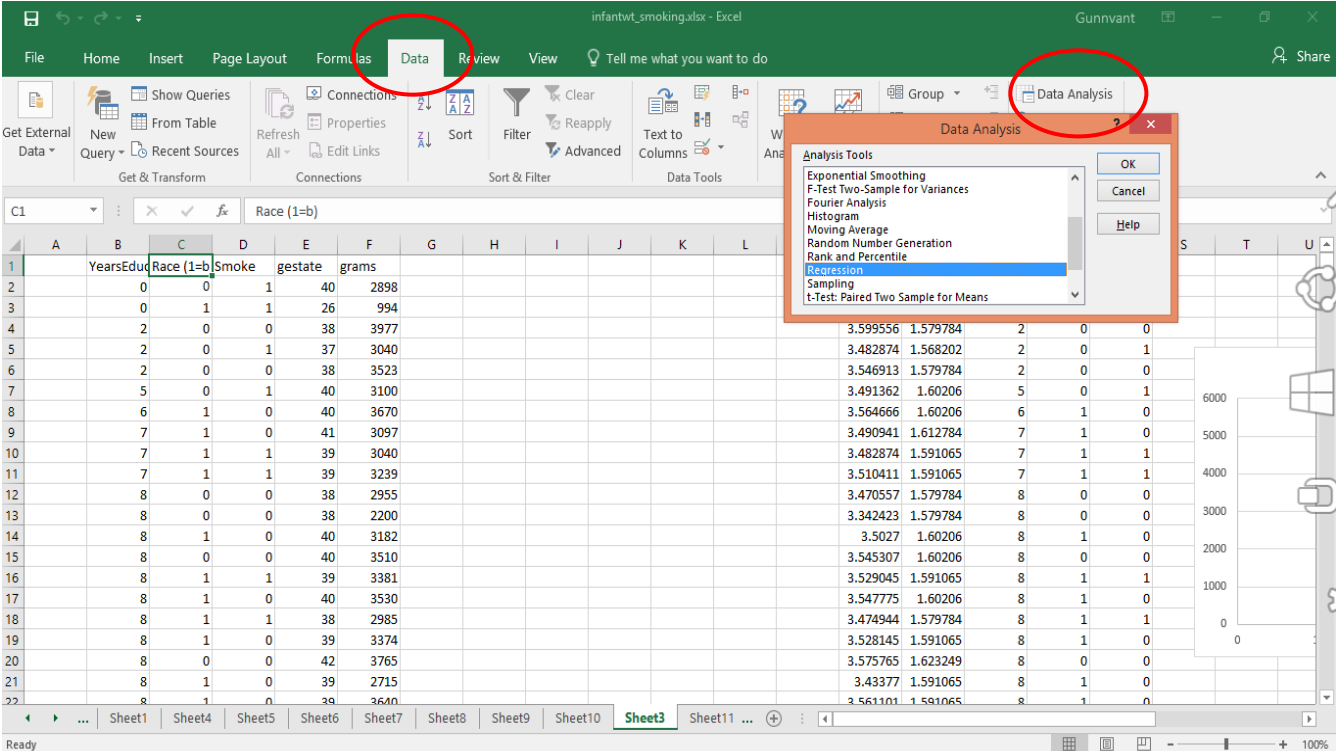


Doing Multiple Linear Regression in Excel

In this tutorial we will see how we can do multiple linear regression in Excel. We will use the same data set that we have been discussing in the preclass videos. We will try to predict **birthweight** using different attributes. Remember while doing a multiple linear regression we use more than one predictor. Let's build a model taking **years of education** and **gestation period** as two predictors for the **birthweight**.

Let's see how we can use excel to build a linear regression model.

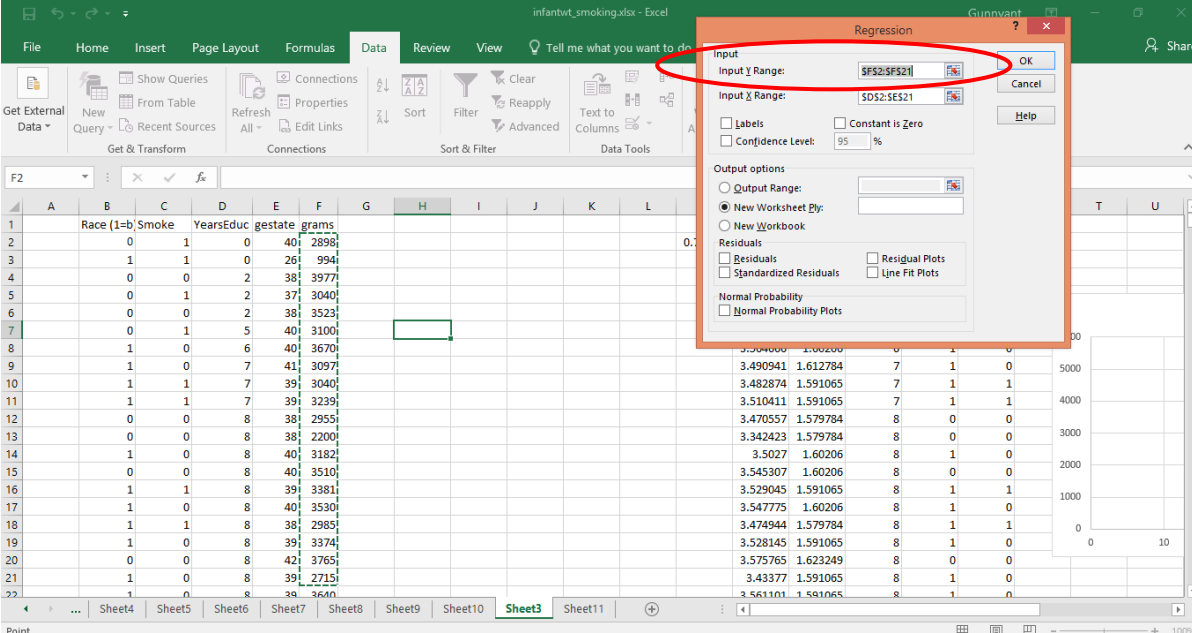
Step 1 - Use the data analysis tool pack to select the regression option (Please zoom in the document)



The screenshot shows the Microsoft Excel interface with the 'Data' tab selected in the ribbon. The 'Data Analysis' button in the 'Data Tools' group is circled. The 'Data Analysis' task pane is open, displaying a list of analysis tools. 'Regression' is highlighted in the list. The background shows a spreadsheet with columns labeled 'YearsEduc', 'Race (1=b)', 'Smoke', 'gestate', and 'grams'.

| | YearsEduc | Race (1=b) | Smoke | gestate | grams |
|----|-----------|------------|-------|---------|-------|
| 1 | | | | | |
| 2 | 0 | 0 | 1 | 40 | 2898 |
| 3 | 0 | 1 | 1 | 26 | 994 |
| 4 | 2 | 0 | 0 | 38 | 3977 |
| 5 | 2 | 0 | 1 | 37 | 3040 |
| 6 | 2 | 0 | 0 | 38 | 3523 |
| 7 | 5 | 0 | 1 | 40 | 3100 |
| 8 | 6 | 1 | 0 | 40 | 3670 |
| 9 | 7 | 1 | 0 | 41 | 3097 |
| 10 | 7 | 1 | 1 | 39 | 3040 |
| 11 | 7 | 1 | 1 | 39 | 3239 |
| 12 | 8 | 0 | 0 | 38 | 2955 |
| 13 | 8 | 0 | 0 | 38 | 2200 |
| 14 | 8 | 1 | 0 | 40 | 3182 |
| 15 | 8 | 0 | 0 | 40 | 3510 |
| 16 | 8 | 1 | 1 | 39 | 3381 |
| 17 | 8 | 1 | 0 | 40 | 3530 |
| 18 | 8 | 1 | 1 | 38 | 2985 |
| 19 | 8 | 1 | 0 | 39 | 3374 |
| 20 | 8 | 0 | 0 | 42 | 3765 |
| 21 | 8 | 1 | 0 | 39 | 2715 |
| 22 | 8 | 1 | 0 | 39 | 3640 |

Step 2 - After selecting the **Regression** option click ok and select the variable ranges (Please zoom in for more details)



The screenshot shows the Excel interface with the 'Data' tab selected. The 'Regression' dialog box is open, showing the 'Input' section. The 'Input Y Range' is set to '\$F\$2:\$F\$21' and the 'Input X Range' is set to '\$D\$2:\$E\$21'. The 'OK' button is highlighted with a red circle. The background shows a spreadsheet with columns labeled 'Race (1=b/Smoke)', 'YearsEduc', 'gestate', and 'grams'.

Notice that the **Input Y Range** cell should point out to the column of data that is to be *predicted*. You can select *multiple columns* when you are building a multiple regression model. As you can see in the **Input X Range** we have selected the two columns: **YearsEduc** and **gestate**.

Step 3 - Click on **OK** and you will be able to see regression output as given below:

| SUMMARY OUTPUT | | | | | | | |
|------------------------------|-----------------------|---------------|----------------|------------------|-----------------------|--------------------|--------------------|
| Regression Statistics | | | | | | | |
| Multiple R | 0.811493 | | | | | | |
| R Square | 0.658521 | | | | | | |
| Adjusted R Square | 0.618347 | | | | | | |
| Standard Error | 393.8539 | | | | | | |
| Observations | 20 | | | | | | |
| ANOVA | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Significance F</i> | | |
| Regression | 2 | 5085403 | 2542701 | 16.39174 | 0.000108 | | |
| Residual | 17 | 2637055 | 155120.9 | | | | |
| Total | 19 | 7722458 | | | | | |
| Coefficients | | | | | | | |
| | <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> |
| Intercept | -3553.68 | 1192.244 | -2.98066 | 0.008394 | -6069.09 | -1038.26 | -6069.09 |
| X Variable 1 | -40.7461 | 36.27129 | -1.12337 | 0.276895 | -117.272 | 35.77966 | -117.272 |
| X Variable 2 | 179.0617 | 33.41932 | 5.35803 | 5.22E-05 | 108.5531 | 249.5703 | 108.5531 |

Remarks:

1. The demo was created using Microsoft Office 2016
2. If you don't see a data analysis tab, you will need to enable it, please use the documentation from Microsoft to enable the data analysis tool pack.