Data 301 Data Analytics
Spreadsheets: Microsoft Excel
Part 2 of 3

Dr. Irene Vrbik

University of British Columbia Okanagan
irene.vrbik@ubc.ca

Term 1, 2019

Charts

A *chart* is a graphical representation of spreadsheet data.

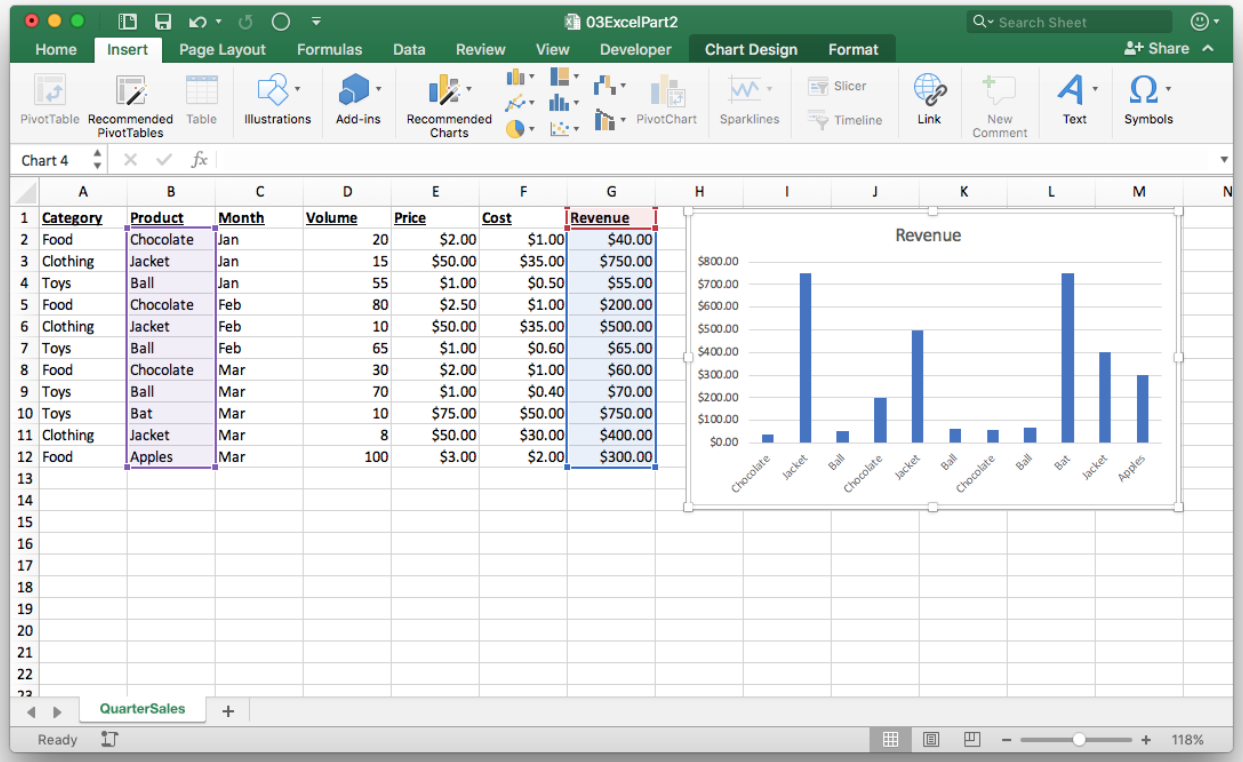
Charts can help you visualize relationships between data and create maximum impact on your audience.

When you create a chart, you can select from many chart types (eg, a stacked column chart or a 3-D exploded pie chart).

To create a line/bar/pic/etc. chart, the user to supply the data that will be displayed.

Charts

Once your data is selected, navigate to the **Insert** tab, then click the desired **Chart** Icon. You can either choose one from the **Recommended Chart** button or select from the variety of chart types.



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$2.00	$1.00	$40.00
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$750.00
4	Toys	Ball	Jan	55	$1.00	$0.50	$55.00
5	Food	Chocolate	Feb	80	$2.50	$1.00	$200.00
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$500.00
7	Toys	Ball	Feb	65	$1.00	$0.60	$65.00
8	Food	Chocolate	Mar	30	$2.00	$1.00	$60.00
9	Toys	Ball	Mar	70	$1.00	$0.40	$70.00
10	Toys	Bat	Mar	10	$75.00	$50.00	$750.00
11	Clothing	Jacket	Mar	8	$50.00	$30.00	$400.00
12	Food	Apples	Mar	100	$3.00	$2.00	$300.00

Sometimes the default graphic is not what we want,. . .

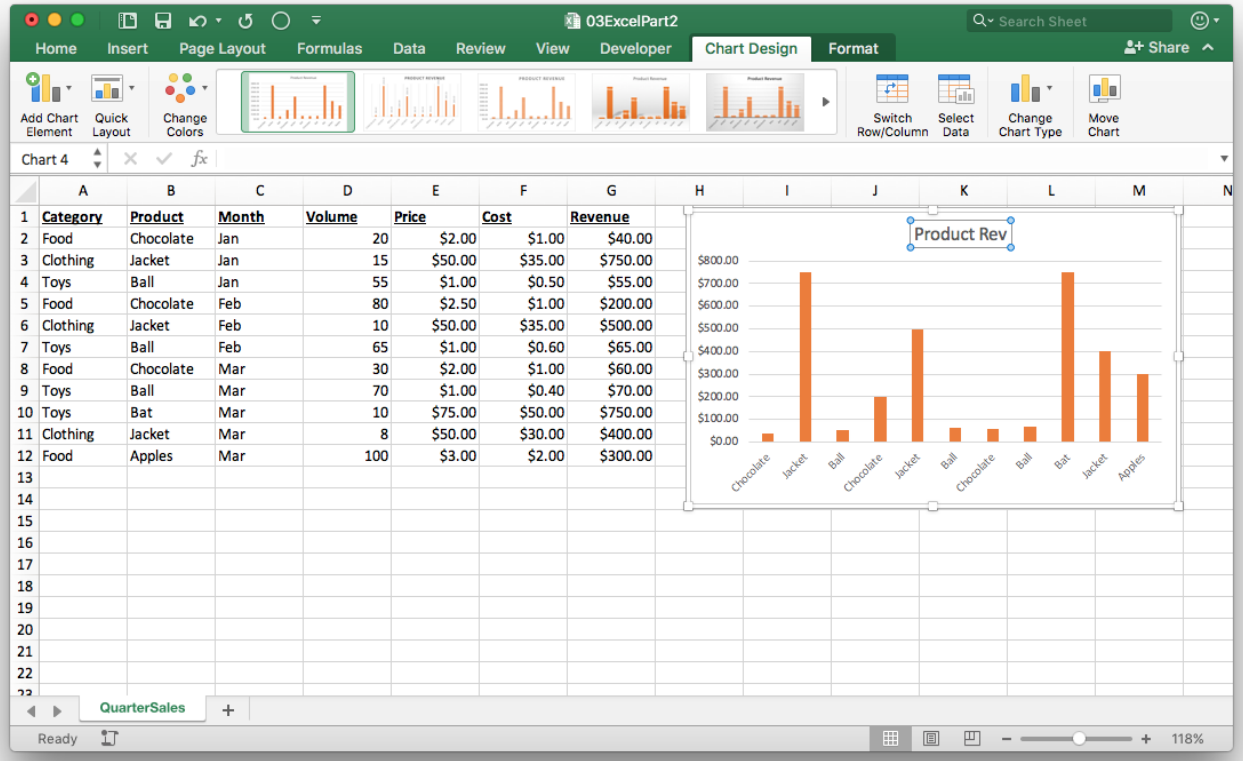
For instance the recommended chart on the previous page shows an individual bar for each instance of the product type.

We could instead create the total revenue for each product by selecting:

- The histogram icon 
- Double clicking on the bars
- Selecting the Series Options icon: 
- Selecting *By Category* in the **Bins** drop down menu.

Chart Options

After you create a chart, you can customize it by applying chart quick layouts or styles. The **Chart Design** tab also provides a number of features for modifying things like chart color/type/position/etc.

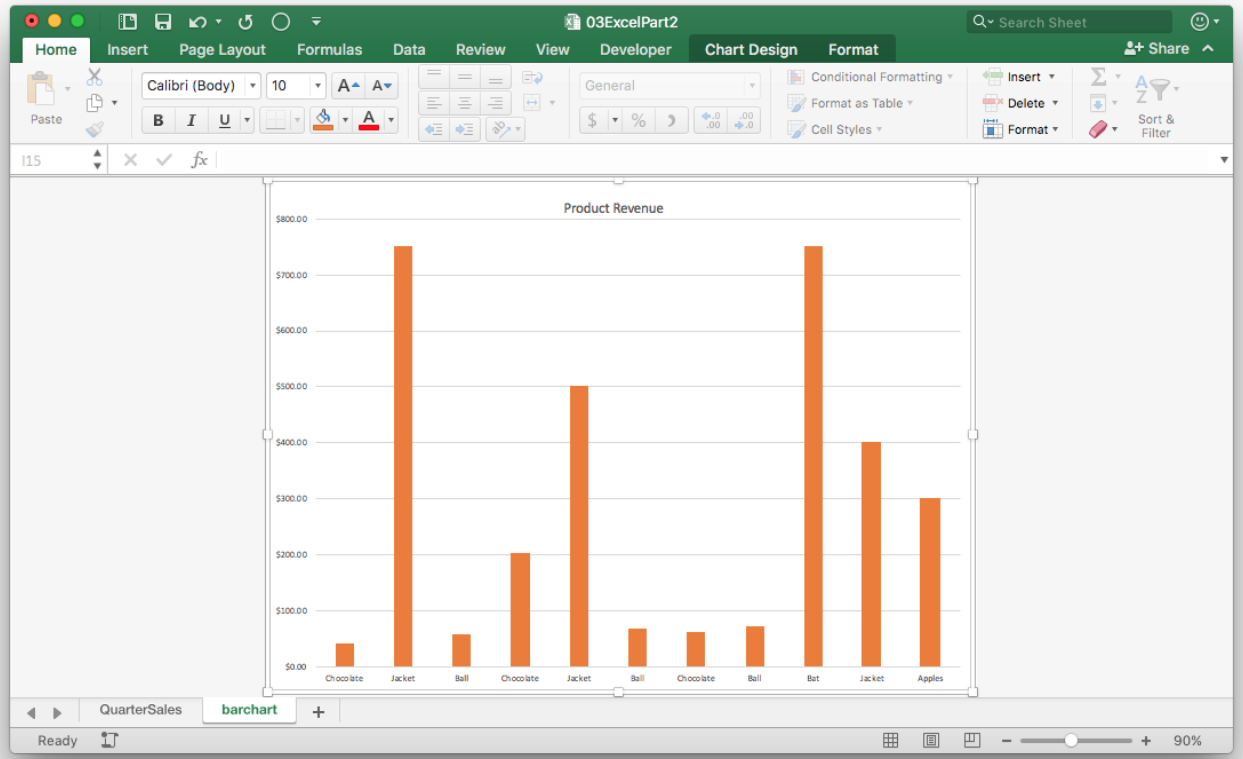


	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$2.00	$1.00	$40.00
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$750.00
4	Toys	Ball	Jan	55	$1.00	$0.50	$55.00
5	Food	Chocolate	Feb	80	$2.50	$1.00	$200.00
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$500.00
7	Toys	Ball	Feb	65	$1.00	$0.60	$65.00
8	Food	Chocolate	Mar	30	$2.00	$1.00	$60.00
9	Toys	Ball	Mar	70	$1.00	$0.40	$70.00
10	Toys	Bat	Mar	10	$75.00	$50.00	$750.00
11	Clothing	Jacket	Mar	8	$50.00	$30.00	$400.00
12	Food	Apples	Mar	100	$3.00	$2.00	$300.00

Chart Options

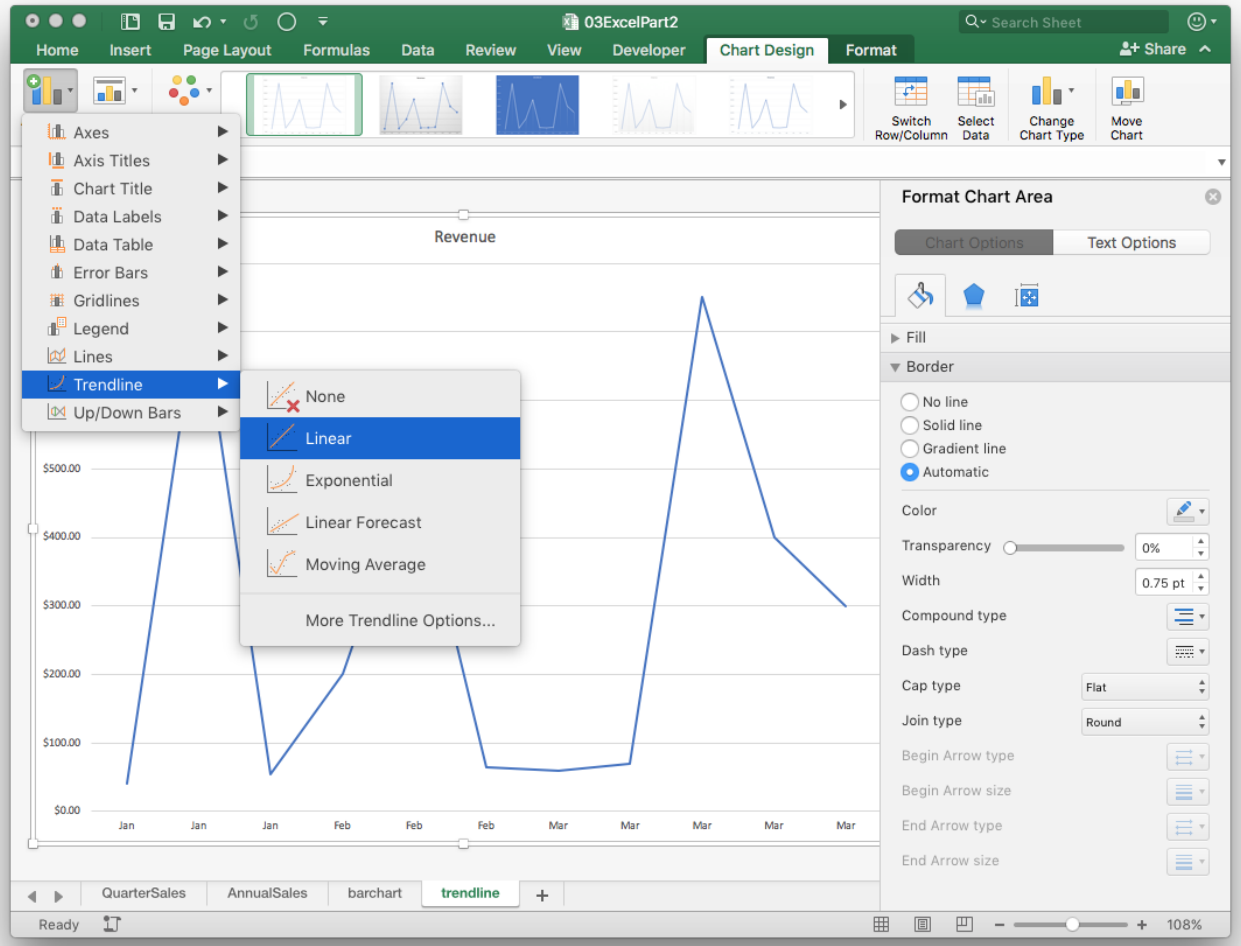
Example 1

Navigate to the **Chart Design** tab and relocate your chart to a new worksheet called *barchart*.



Trendlines

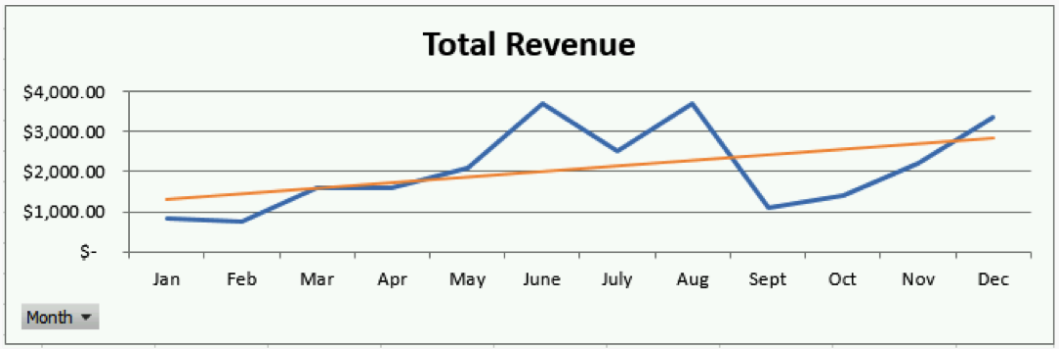
- Trendlines can be easily added to a number of chart types.
- Click on chart, "Add Chart Element"...."Trendline"

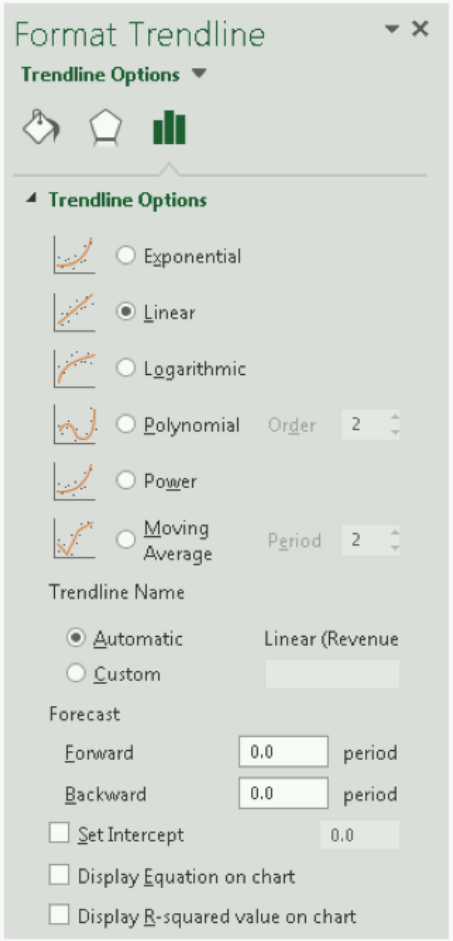


Trendlines

Trendlines can be easily added to any chart.

- Linear treadline for monthly revenue. Good choice?





Probably not, as data appears to be seasonal

Try it! Trendline

Of course it wouldn't really make sense to add a linear trend to the previous line chart since the data do not appear linear.

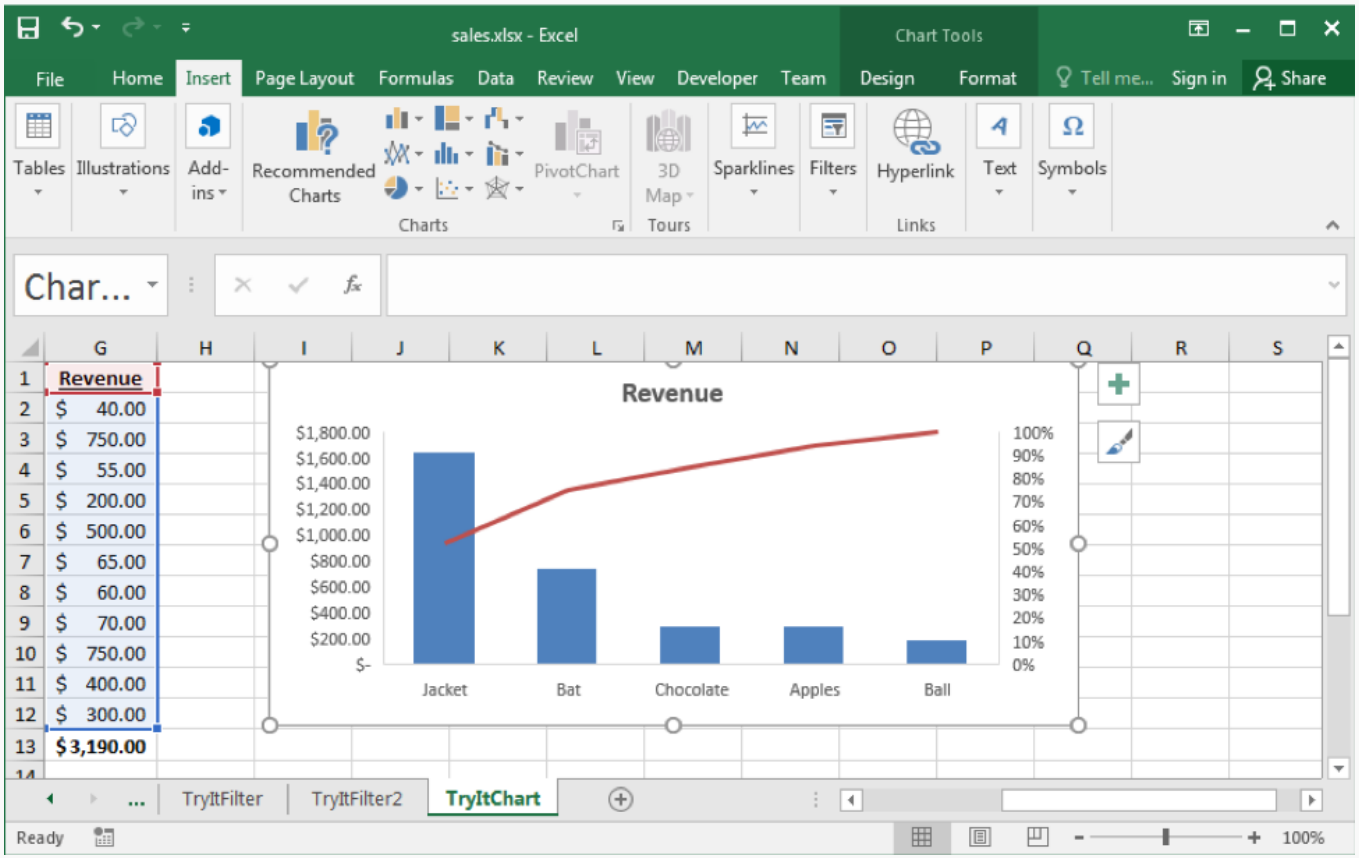
Apples

1. Create a line chart that plots the revenue per month for only **Apples** using the **Annual Sales data**.
2. Add a linear trendline.
3. In **Chart Design** tab, select a design that displays the months on a 45 degree angle.

Try It: Chart

Question:

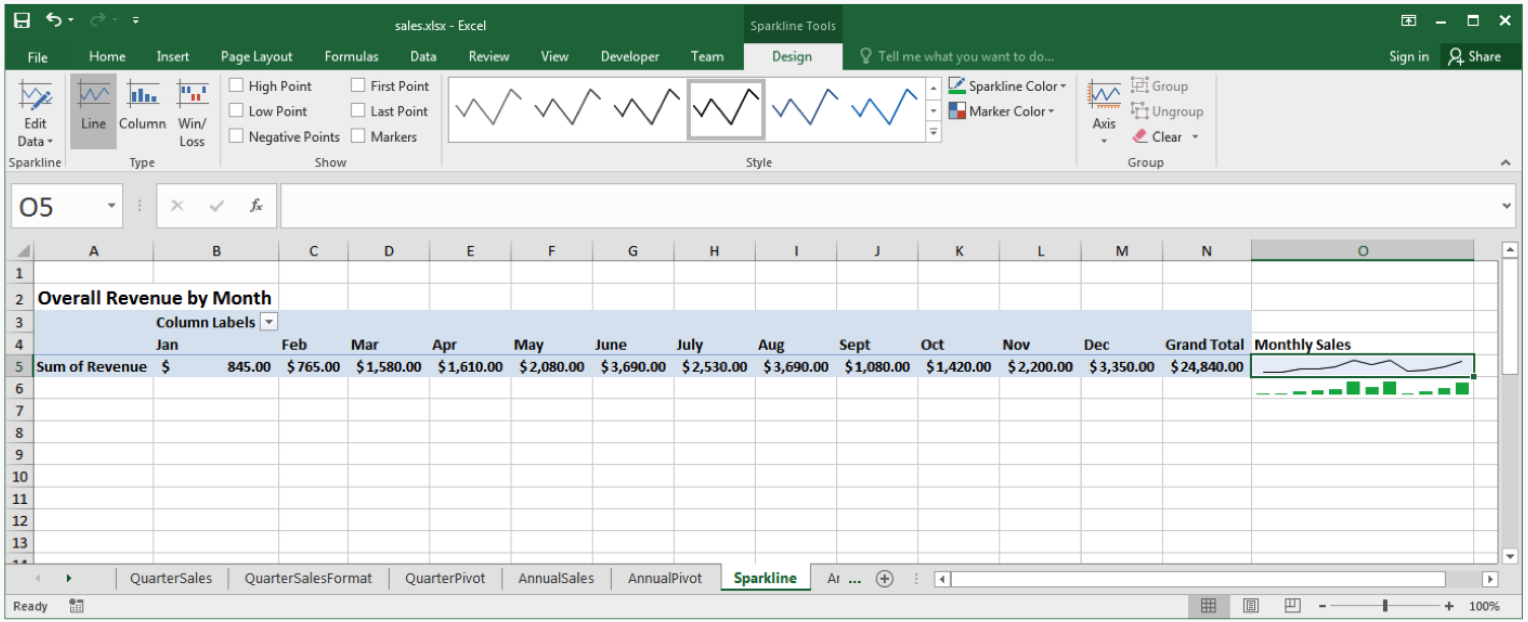
Create a Pareto chart (option under the histogram icon) to make it easy to see the best selling product.



Sparklines

A *sparkline* is a tiny chart in a worksheet cell for a quick data overview.

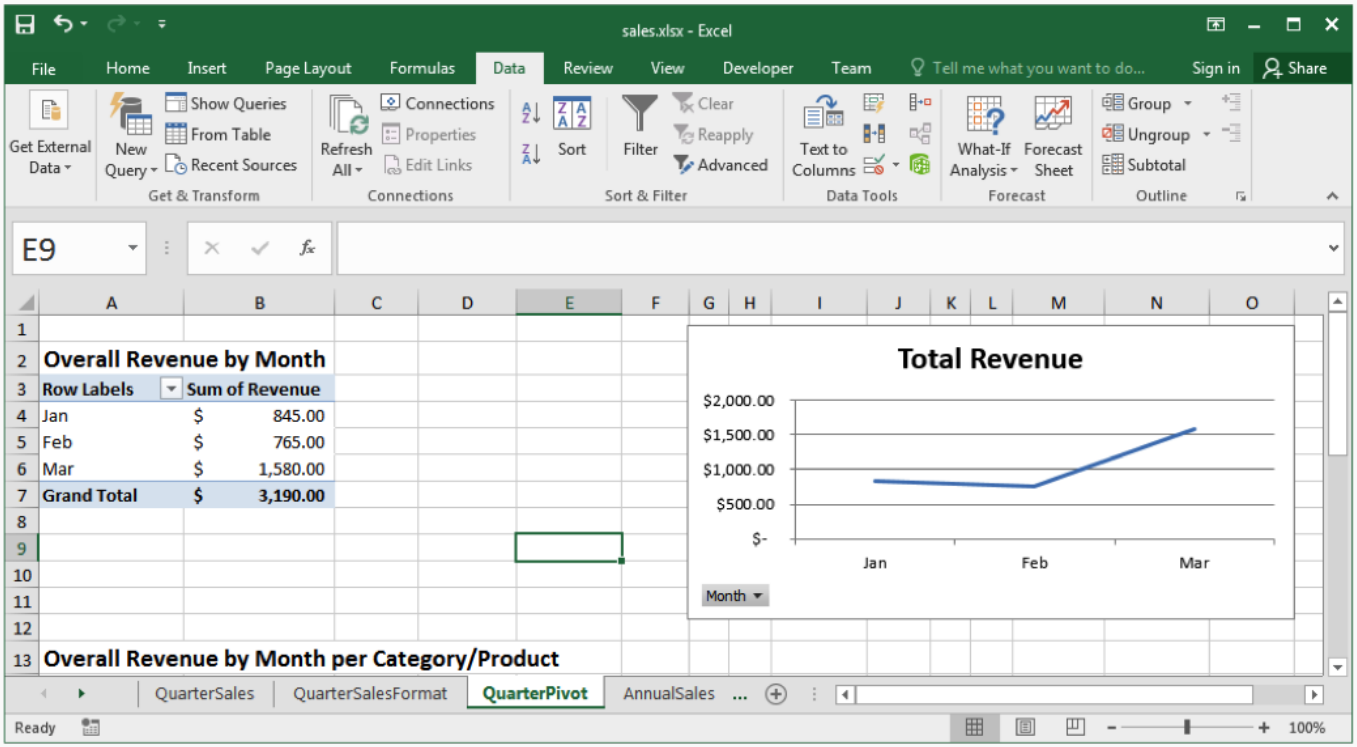
- Click **Insert** then select **Sparklines**...
- Displayed below is an example of a *line* and *column* sparkline.
- You can change the style in the **Sparkline** tab in the ribbon.



Pivot Tabels

Pivot tables allow for easily aggregating and exploring large data sets.

- For example, our data set can be summarized by revenue by month.



Creating a Pivot Table

To create a pivot table

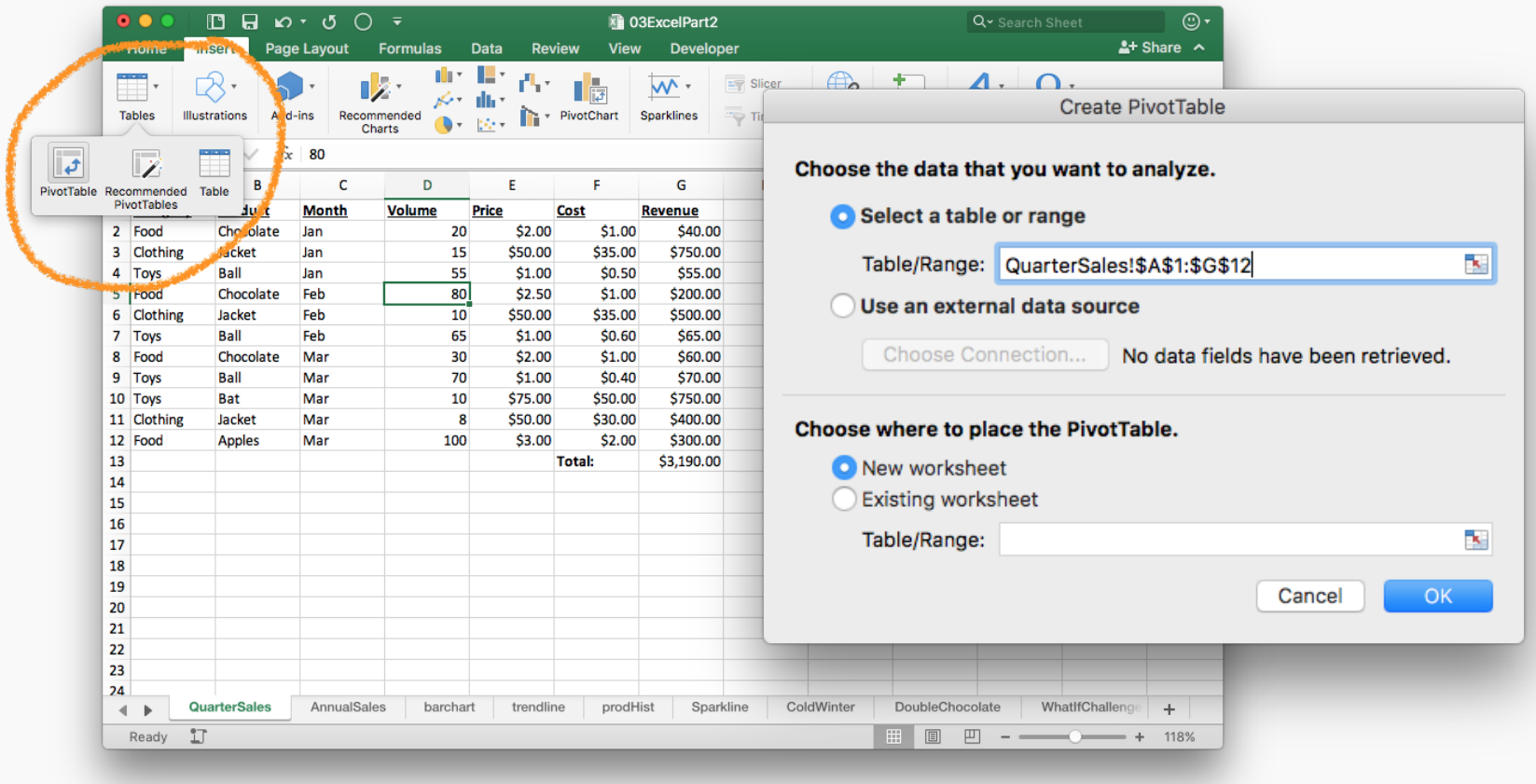
1. Select the data (or click any single cell inside the data set).
2. Navigate to the **Insert** tab, select the drop down **Tables** menu, and click the PivotTable icon.
3. Click OK
4. Add fields in the PivotTable Fields pane

Drag fields To get the total revenue for each month, drag the following fields to the different areas.

1. `Month` to the Rows area.
2. `Revenue` to the Values area.

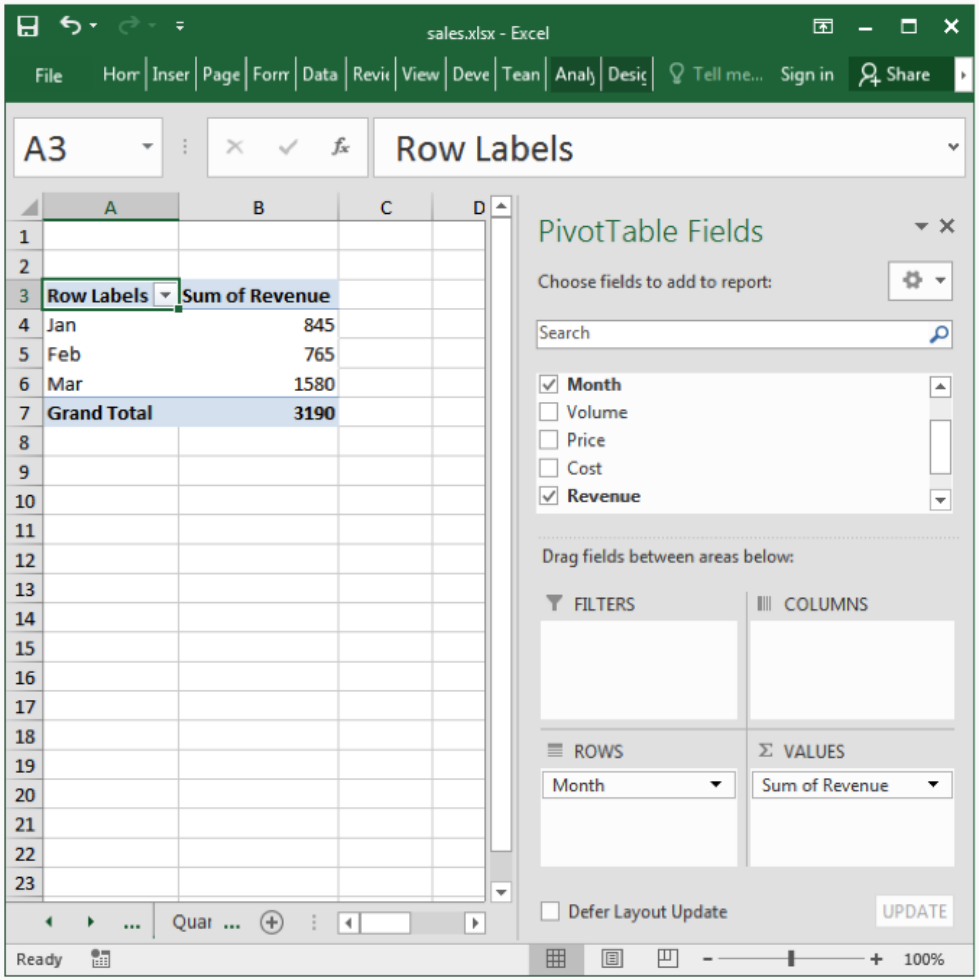
Creating a Pivot Table

The following dialog box appears. Excel automatically selects the data for you. The default location for a new pivot table is New Worksheet.



Creating a Pivot Table

Pivot1 worksheet



Add fields to pivot table.

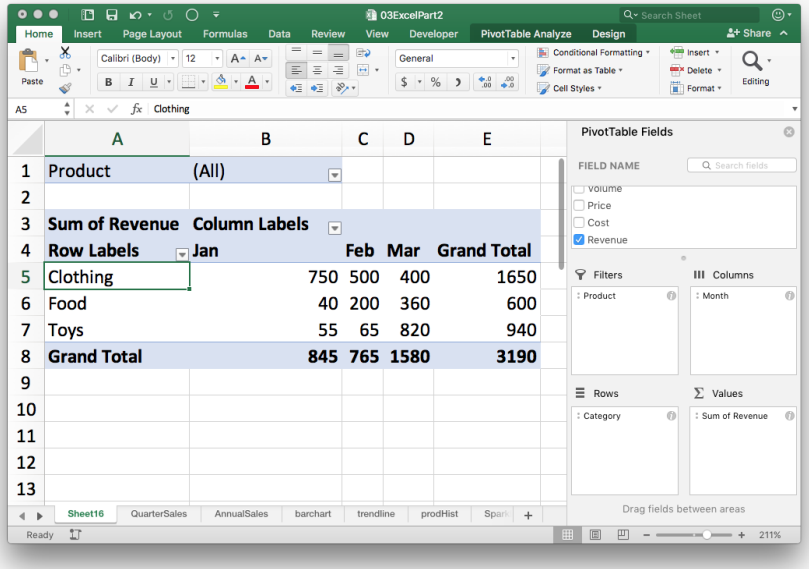
Field may either be:

- ROWS value
- COLUMNS value
- Cell VALUES (aggregated)
- Used in a FILTERS (we could also filter using the drop down filter option; select the dropdown arrow located in cell A3, for example)

N.B Right click on a cell in the pivot table and select Show/Hide Fields List to show/hide the pop-up window on the right.

Pivot Table Example

Pivot2 worksheet

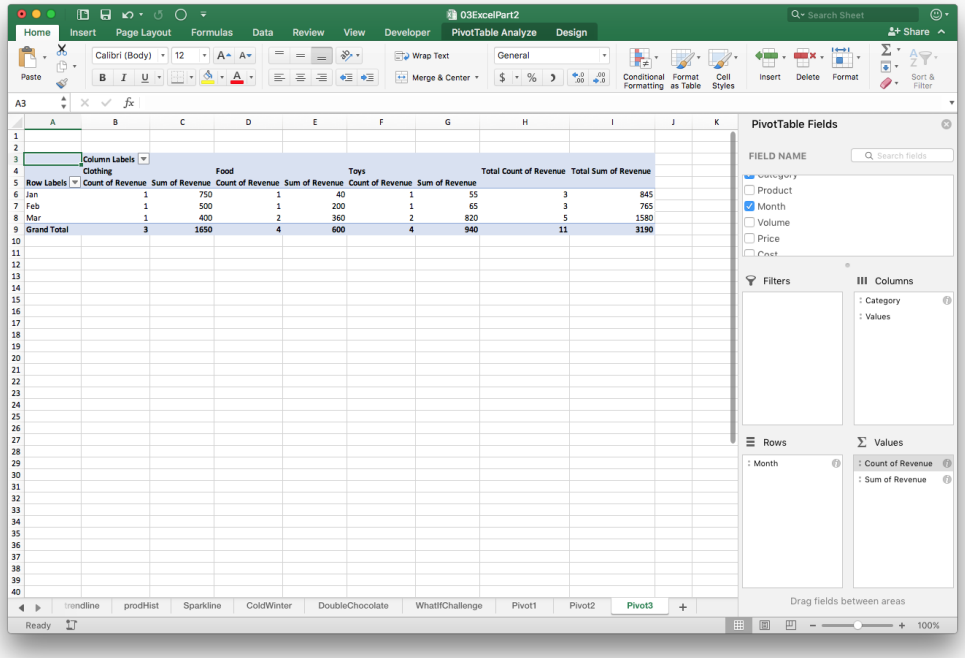


	A	B	C	D	E
1	Product	(All)			
2					
3	**Sum of Revenue**	**Column Labels**			
4	**Row Labels**	**Jan**	**Feb**	**Mar**	**Grand Total**
5	Clothing	750	500	400	1650
6	Food	40	200	360	600
7	Toys	55	65	820	940
8	**Grand Total**	**845**	**765**	**1580**	**3190**
9					
10					
11					
12					
13					

- Categories are rows.
- Months are columns.
- Each cell is a sum of revenue per category for that month.
- We can filter on product (eg. filter out apples).

Pivot Table Example

Pivot3 worksheet



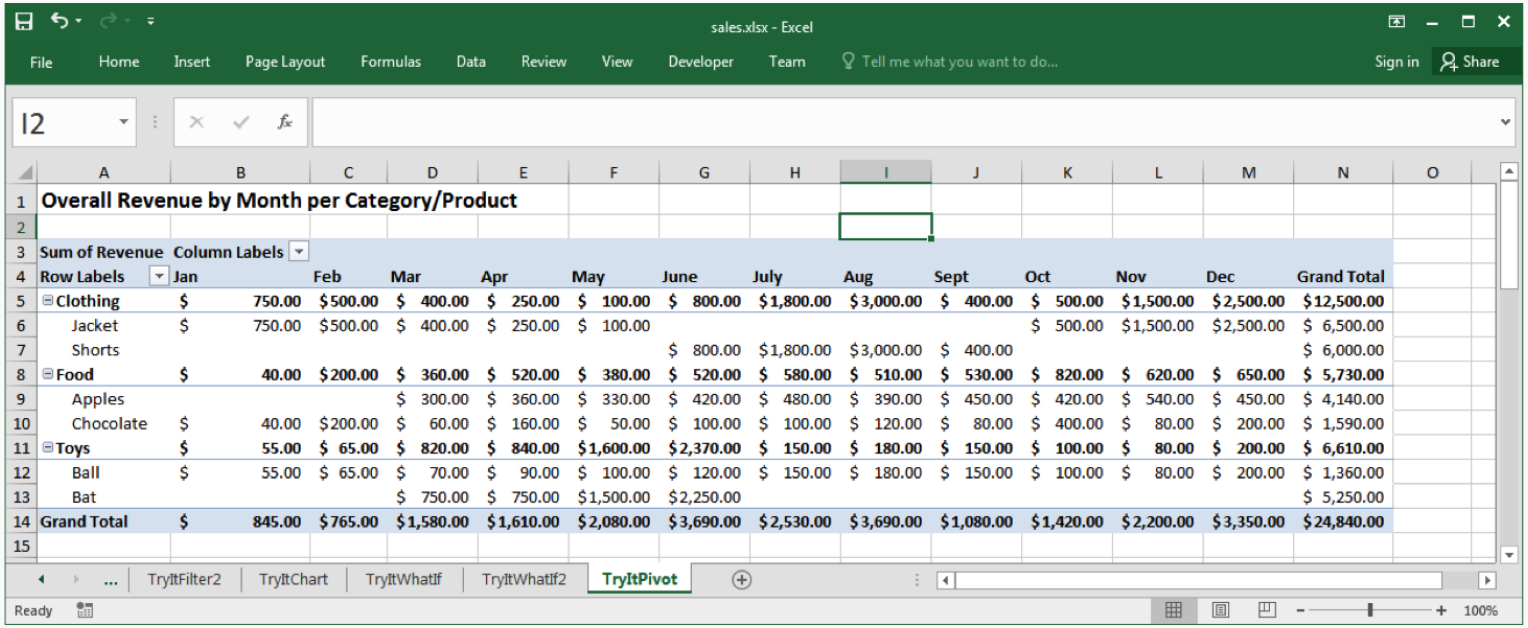
Row Labels	Clothing Count of Revenue	Clothing Sum of Revenue	Food Count of Revenue	Food Sum of Revenue	Toys Count of Revenue	Toys Sum of Revenue	Total Count of Revenue	Total Sum of Revenue
Jan	1	750	1	40	1	55	3	845
Feb	1	500	1	200	1	65	3	765
Mar	1	400	2	360	2	820	5	1580
Grand Total	3	1650	4	600	4	940	11	3190

- Notice that the same field can be used in VALUES multiple times
- In that way we can see multiple aggregate summaries (eg. total and count)
- To change the value aggregation function, click on the  icon
- Notice that a field can NOT be used in ROWS/COLUMNS/FILTER at the same time.

Try It: Pivot Table

Question:

Create a pivot table using the annual sales data that shows the total revenue per month by category/product.

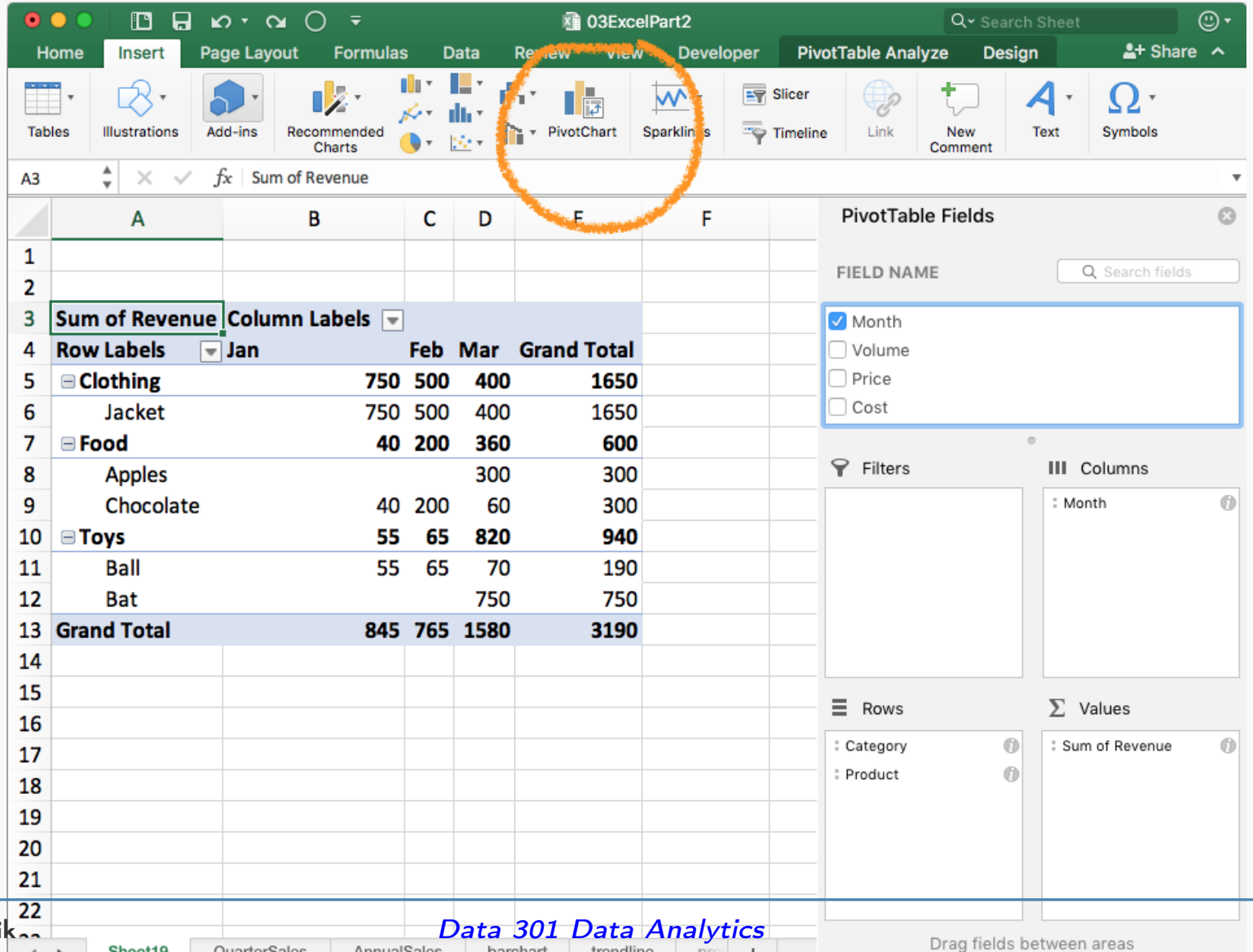


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	**Overall Revenue by Month per Category/Product**														
2															
3	Sum of Revenue	Column Labels													
4	Row Labels	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Grand Total	
5	**Clothing**	$ 750.00	$500.00	$ 400.00	$ 250.00	$ 100.00	$ 800.00	$1,800.00	$3,000.00	$ 400.00	$ 500.00	$1,500.00	$2,500.00	$12,500.00	
6	Jacket	$ 750.00	$500.00	$ 400.00	$ 250.00	$ 100.00					$ 500.00	$1,500.00	$2,500.00	$ 6,500.00	
7	Shorts						$ 800.00	$1,800.00	$3,000.00	$ 400.00				$ 6,000.00	
8	**Food**	$ 40.00	$200.00	$ 360.00	$ 520.00	$ 380.00	$ 520.00	$ 580.00	$ 510.00	$ 530.00	$ 820.00	$ 620.00	$ 650.00	$ 5,730.00	
9	Apples			$ 300.00	$ 360.00	$ 330.00	$ 420.00	$ 480.00	$ 390.00	$ 450.00	$ 420.00	$ 540.00	$ 450.00	$ 4,140.00	
10	Chocolate	$ 40.00	$200.00	$ 60.00	$ 160.00	$ 50.00	$ 100.00	$ 100.00	$ 120.00	$ 80.00	$ 400.00	$ 80.00	$ 200.00	$ 1,590.00	
11	**Toys**	$ 55.00	$ 65.00	$ 820.00	$ 840.00	$1,600.00	$2,370.00	$ 150.00	$ 180.00	$ 150.00	$ 100.00	$ 80.00	$ 200.00	$ 6,610.00	
12	Ball	$ 55.00	$ 65.00	$ 70.00	$ 90.00	$ 100.00	$ 120.00	$ 150.00	$ 180.00	$ 150.00	$ 100.00	$ 80.00	$ 200.00	$ 1,360.00	
13	Bat			$ 750.00	$ 750.00	$1,500.00	$2,250.00							$ 5,250.00	
14	**Grand Total**	$ 845.00	$765.00	$1,580.00	$1,610.00	$2,080.00	$3,690.00	$2,530.00	$3,690.00	$1,080.00	$1,420.00	$2,200.00	$3,350.00	$24,840.00	
15															

N.B. The order in which your fields appear in ROWS will make a difference. Also, we can reformat the cells to display in **currency** format.

Pivot Charts

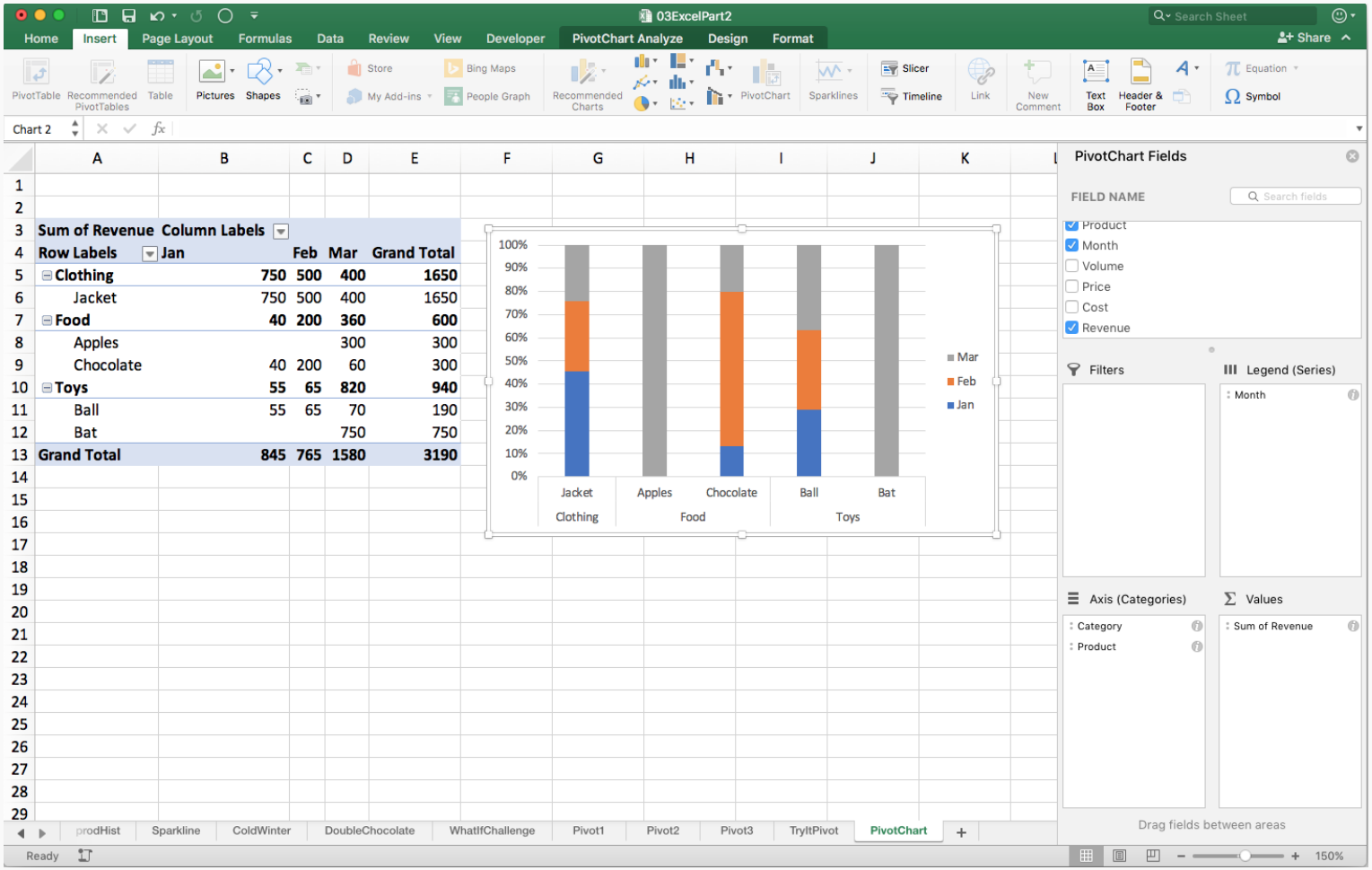
A pivot chart is a chart attached to a pivot table. Create it under **Insert** then **Pivot Chart**.



Sum of Revenue	Column Labels			
Row Labels	Jan	Feb	Mar	Grand Total
Clothing	**750**	**500**	**400**	**1650**
Jacket	750	500	400	1650
Food	**40**	**200**	**360**	**600**
Apples			300	300
Chocolate	40	200	60	300
Toys	**55**	**65**	**820**	**940**
Ball	55	65	70	190
Bat			750	750
Grand Total	**845**	**765**	**1580**	**3190**

Pivot Charts

Below we have chosen a 100% stacked bar chart type.



Pivot Tables Question

Example 2

How many of the following statements are TRUE?

1. A pivot table field can be used in ROWS and COLUMNS at the same time.
2. A pivot table field can be used in VALUES more than once.
3. In our sales spreadsheet example, if Product and Category are both used in ROWS then the order they are list does not matter.

A) 0 **B)** 1 **C)** 2 **D)** 3

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. A pivot table field can be used in ROWS and COLUMNS at the same time.
2. A pivot table field can be used in VALUES more than once.
3. In our sales spreadsheet example, if Product and Category are both used in ROWS then the order they are list does not matter.

A) 0 B) 1 C) 2 D) 3

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. A pivot table field can be used in ROWS and COLUMNS at the same time.
2. A pivot table field can be used in VALUES more than once.
3. In our sales spreadsheet example, if Product and Category are both used in ROWS then the order they are list does not matter.

A) 0 B) 1 C) 2 D) 3

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. A pivot table field can be used in ROWS and COLUMNS at the same time.
2. A pivot table field can be used in VALUES more than once.
3. In our sales spreadsheet example, if Product and Category are both used in ROWS then the order they are list does not matter.

A) 0 B) 1 C) 2 D) 3

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. A pivot table field can be used in ROWS and COLUMNS at the same time.
2. A pivot table field can be used in VALUES more than once.
3. In our sales spreadsheet example, if Product and Category are both used in ROWS then the order they are list does not matter.

A) 0 B) *1* C) 2 D) 3

Conditions and Decisions

A *condition* is an expression that is either TRUE or FALSE.

Conditions are used to make decisions and perform different actions depending on the condition value.

Excel condition and decision functions:

`FALSE()` returns `FALSE`

`TRUE()` returns `TRUE`

`AND(cond1, cond2)` returns `TRUE` if both `cond1` and `cond2` are true

`OR(cond1, cond2)` returns `TRUE` if either or both of `cond1` and `cond2` are true

`NOT(cond)` returns `TRUE` if `cond` is `FALSE`

Decisions using `IF()`

The `IF()` function is used to make a decision based on a condition

- `IF(condition, value_if_true, value_if_false)`

Example: If cell **A2** is less than 5, return 10 otherwise return 20.

- `= IF(A2 < 5, 10, 20)`

If the third argument is not specified, it's default is `FALSE()`

- `= IF(A2 < 5, 10)` returns `FALSE()` if **A2** is ≥ 5

The following statements are equivalent

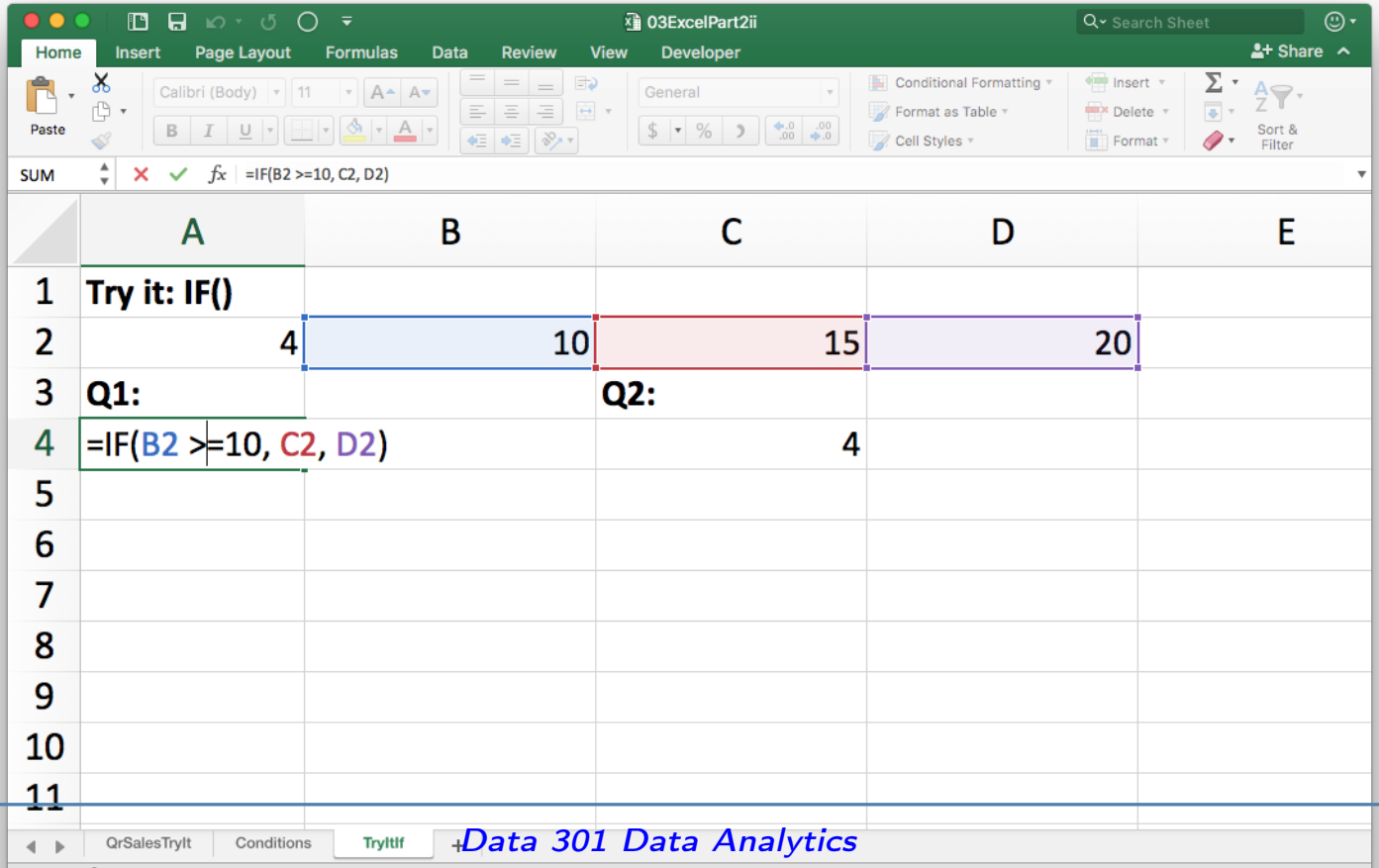
- `= IF(A2 < 5, TRUE(), FALSE())` same as `=A2 < 5`

Hence if we just want to check if a condition is met, we don't need the `IF()` function.

Try it: IF()

Question: Create two conditions:

1. If cell B2 $\leq$ 10, then show C2, otherwise D2.
2. If cell B2$<$ 15 and C2$>$ 20, return B2*C2, otherwise if D2 $<$ 10, return 1, else 4.



Decision Using IF() Question

Example 3

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. `=AND(FALSE(), TRUE())`
2. `=OR(FALSE(), NOT(TRUE()))`
3. `=IF(A1=40, 5, 10)` returns 10.
4. `=IF(OR(A1=40,A2>10),1, 2)` returns 2.
5. `=IF(A2=10,IF(A1=40,FALSE()),TRUE())`

A) 0 B) 1 C) 2 D) 3 E) 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) 0 B) 1 C) 2 D) 3 E) 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) 0 B) 1 C) 2 D) 3 E) 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) 0 B) 1 C) 2 D) 3 E) 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) 0 B) 1 C) 2 D) 3 E) 4

Decision Using IF() Question

Answer:

How many of these statements are TRUE with **A1**=40, **A2**=10?

1. =AND(FALSE(), TRUE())
2. =OR(FALSE(), NOT(TRUE()))
3. =IF(A1=40, 5, 10) returns 10.
4. =IF(OR(A1=40,A2>10),1, 2) returns 2.
5. =IF(A2=10,IF(A1=40,FALSE()),TRUE())

A) *0* B) 1 C) 2 D) 3 E) 4

What-If

What-If scenarios help understand different possibilities.

There are three kinds of What-If Analysis tools:

- Scenarios
- Goal Seek
- Data Tables

A what-if *scenario* is created under **Data** tab then **What-If Analysis** then **Scenario Manager**.

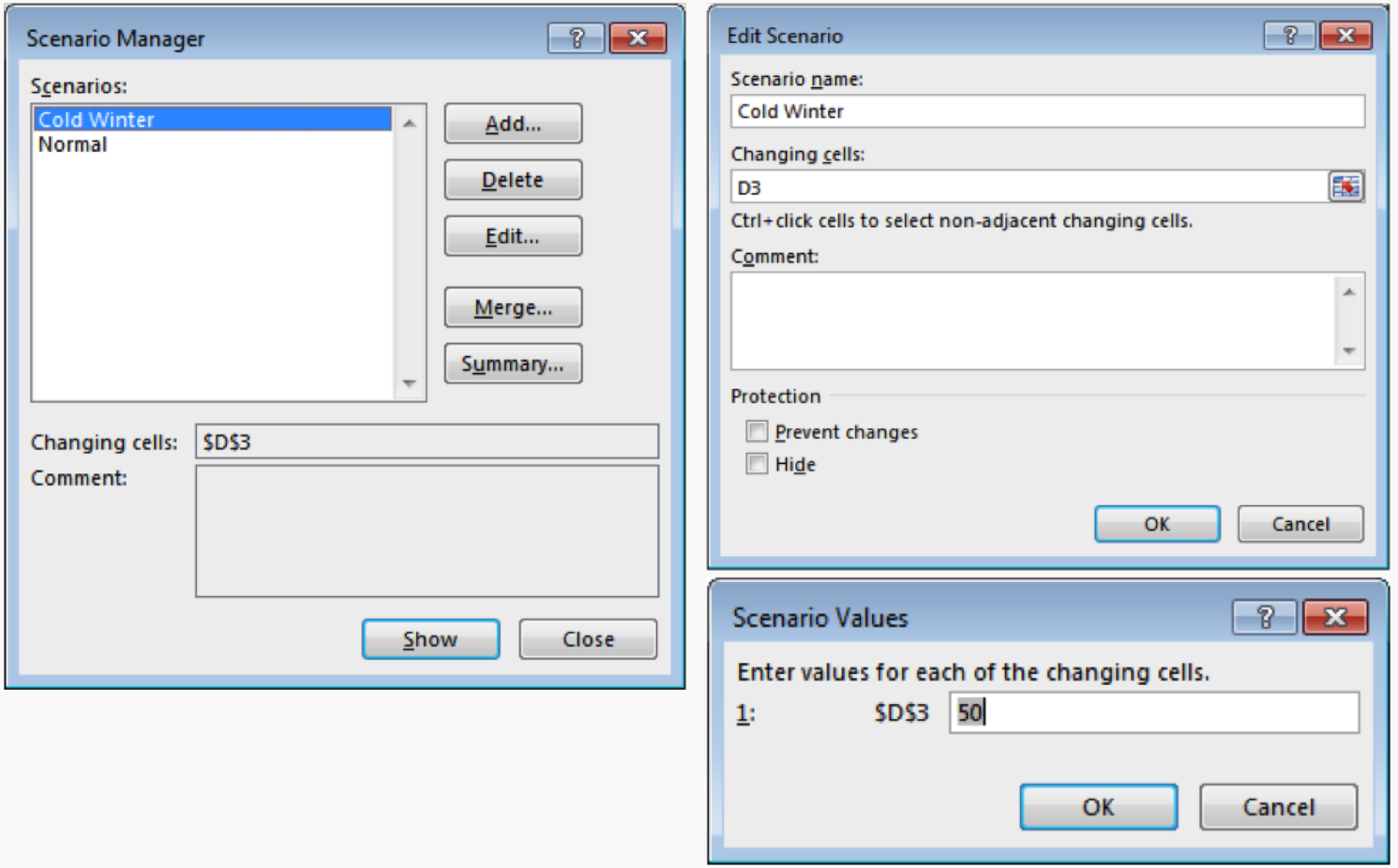
To define a scenario, give it a name and list the cells that will change with this scenario.

Tip

It is usually a good idea to create a "Normal" scenario that you can go back to.

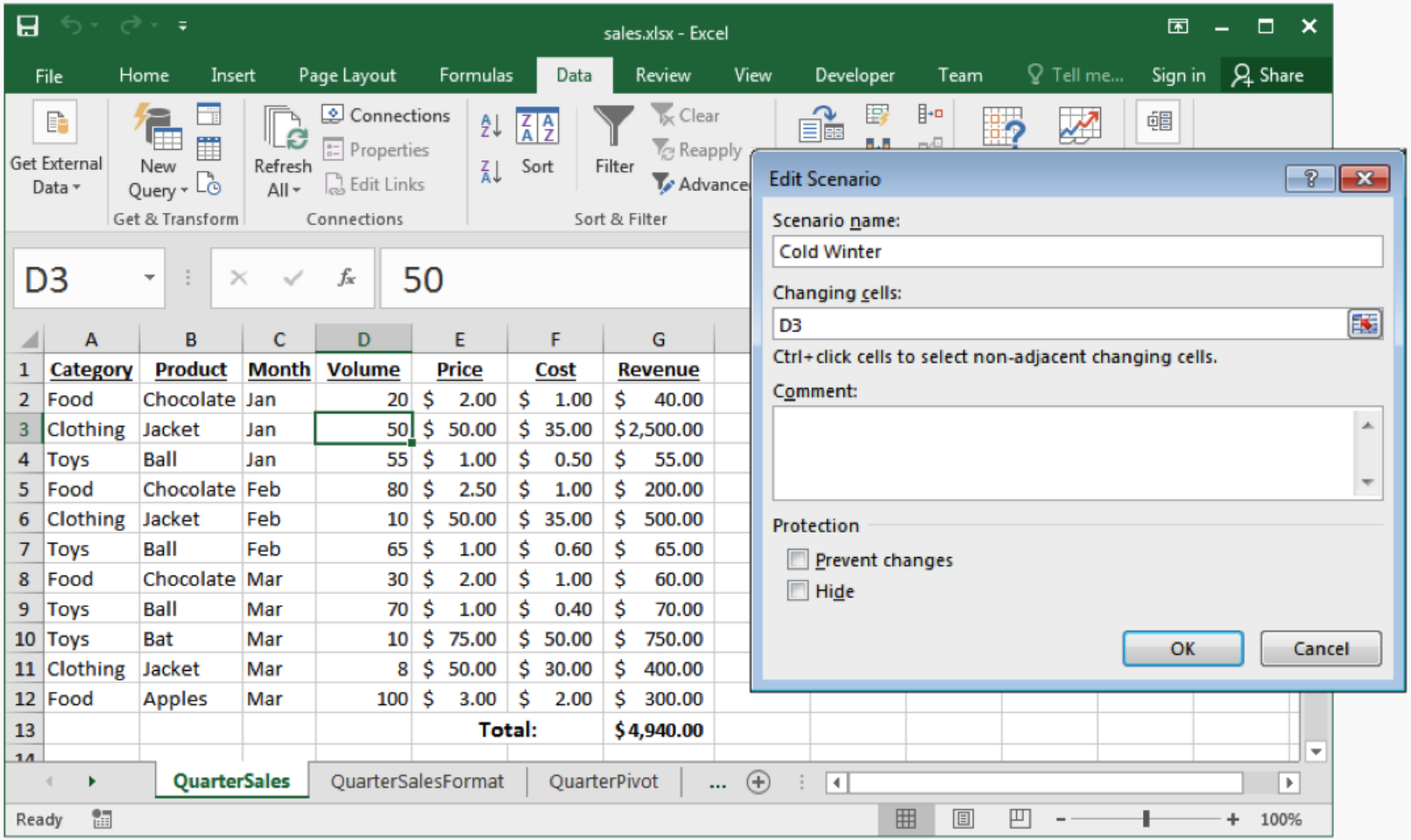
What-If

Consider what happens with a cold winter and we predict to sell 50 jackets instead of the normal 15.



What-If

User can easily select scenario and see the result.



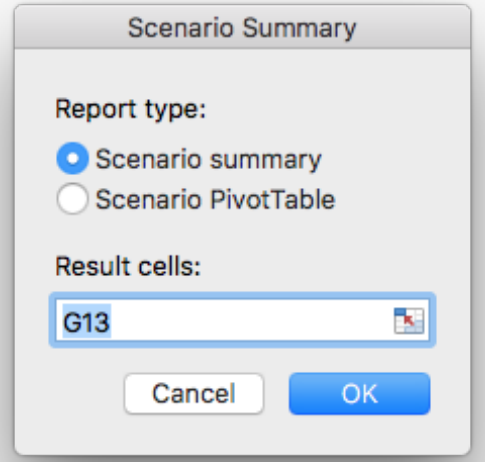
	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	50	$ 50.00	$ 35.00	$ 2,500.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 4,940.00

Scenario Summary

- After creating all of your scenarios, you can create a Scenario Summary Report.
- This report summaries how these scenarios impact a specified **Result Cell**
- For instance, to see how the *Cold Winter* scenario affects total revenue when compared with the *Normal*
  1. go to the **Data** tab then **What-If Analysis** then **Scenario Manager**.
  2. Select **Summary** in the pop up window.
  3. Select the **Scenario summary** report type and specify the **Result cell** (the value of interest these scenarios will impact).

Scenario Summary

A scenario report will display a summary in table form:



Scenario Summary

	Current Values:	Normal	Cold Winter
Changing Cells:			
D3	15	15	50
Result Cells:			
G13	$3,190.00	$3,190.00	$4,940.00

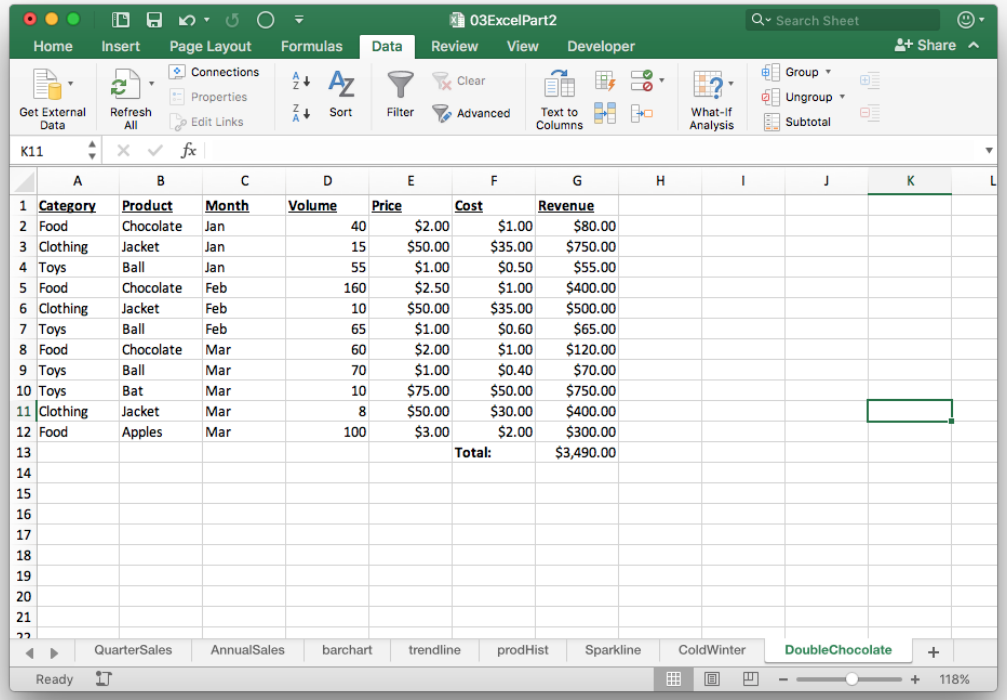
Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

Figure: Left: Step 3 from the previous slide. **Right:** the resulting Scenario Summary

Try it: What-If

Question:

Create a what-if scenario that wherever chocolates are sold, the volume is double than normal. Examine how it affects our total revenue.

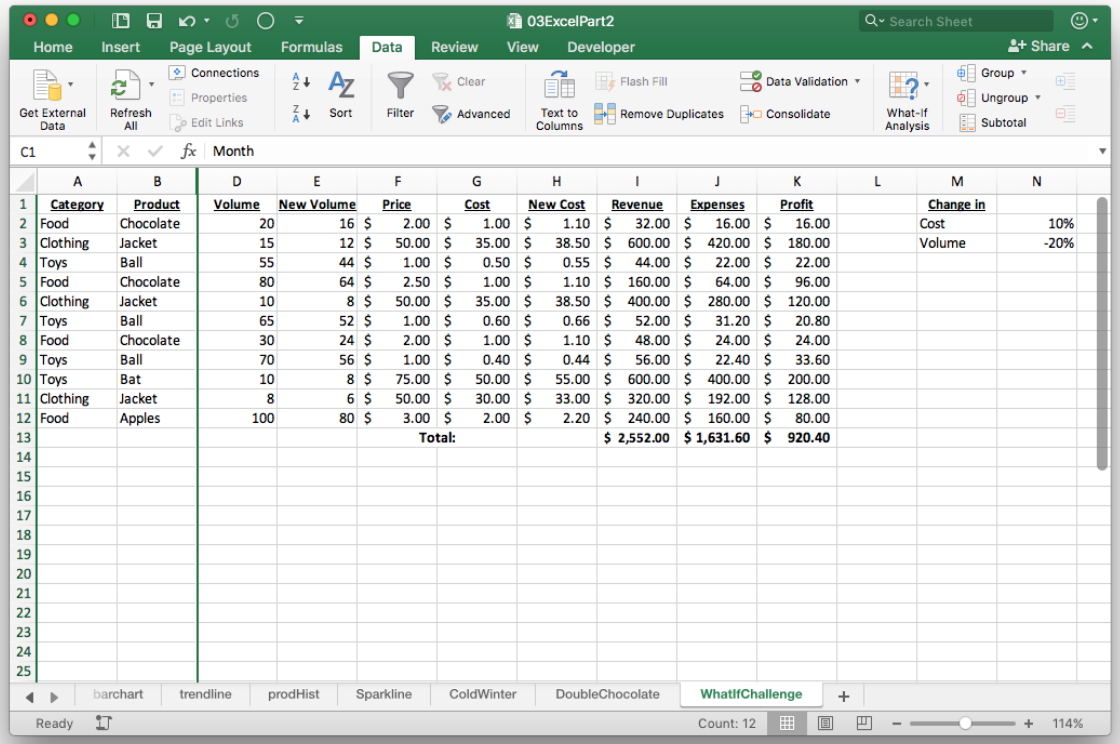


	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	40	$2.00	$1.00	$80.00
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$750.00
4	Toys	Ball	Jan	55	$1.00	$0.50	$55.00
5	Food	Chocolate	Feb	160	$2.50	$1.00	$400.00
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$500.00
7	Toys	Ball	Feb	65	$1.00	$0.60	$65.00
8	Food	Chocolate	Mar	60	$2.00	$1.00	$120.00
9	Toys	Ball	Mar	70	$1.00	$0.40	$70.00
10	Toys	Bat	Mar	10	$75.00	$50.00	$750.00
11	Clothing	Jacket	Mar	8	$50.00	$30.00	$400.00
12	Food	Apples	Mar	100	$3.00	$2.00	$300.00
13						Total:	$3,490.00

Try it: What-If Scenario Challenge

Question:

Create a what-if scenario in which all costs go up by 10% and volume down by 20%.



	A	B	D	E	F	G	H	I	J	K	L	M	N
1	Category	Product	Volume	New Volume	Price	Cost	New Cost	Revenue	Expenses	Profit		Change in	
2	Food	Chocolate	20	16	$ 2.00	$ 1.00	$ 1.10	$ 32.00	$ 16.00	$ 16.00		Cost	10%
3	Clothing	Jacket	15	12	$ 50.00	$ 35.00	$ 38.50	$ 600.00	$ 420.00	$ 180.00		Volume	-20%
4	Toys	Ball	55	44	$ 1.00	$ 0.50	$ 0.55	$ 44.00	$ 22.00	$ 22.00			
5	Food	Chocolate	80	64	$ 2.50	$ 1.00	$ 1.10	$ 160.00	$ 64.00	$ 96.00			
6	Clothing	Jacket	10	8	$ 50.00	$ 35.00	$ 38.50	$ 400.00	$ 280.00	$ 120.00			
7	Toys	Ball	65	52	$ 1.00	$ 0.60	$ 0.66	$ 52.00	$ 31.20	$ 20.80			
8	Food	Chocolate	30	24	$ 2.00	$ 1.00	$ 1.10	$ 48.00	$ 24.00	$ 24.00			
9	Toys	Ball	70	56	$ 1.00	$ 0.40	$ 0.44	$ 56.00	$ 22.40	$ 33.60			
10	Toys	Bat	10	8	$ 75.00	$ 50.00	$ 55.00	$ 600.00	$ 400.00	$ 200.00			
11	Clothing	Jacket	8	6	$ 50.00	$ 30.00	$ 33.00	$ 320.00	$ 192.00	$ 128.00			
12	Food	Apples	100	80	$ 3.00	$ 2.00	$ 2.20	$ 240.00	$ 160.00	$ 80.00			
13					Total:			$ 2,552.00	$ 1,631.60	$ 920.40			
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													

Goal Seek and Solver

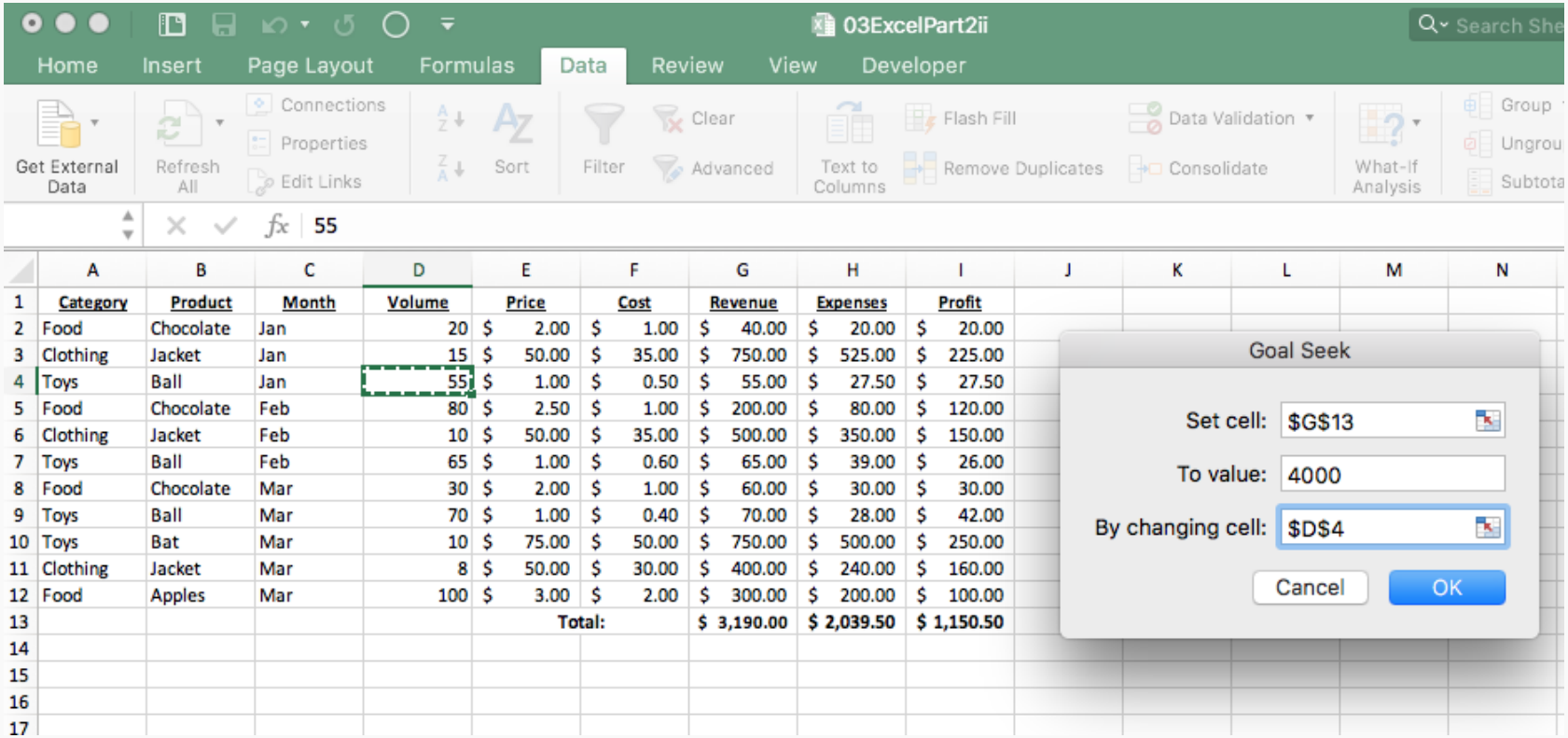
- Another way to ask Excel "what if" is by using the Goal Seek and Solver tools.
- *Goal seek* is used to determine what **value** needs to be in an input cell to achieve a desired result in a formula cell.
- *Solver* determines what **values** need to be in multiple input cells to achieve a desired result in a formula cell.
  - The Solver add-in is similar to Goal Seek, but it can accommodate more variables.
- These methods work to achieve a certain goal in the form of a formula *output*, while what-if scenarios looked at changing formula *inputs*.

Goal Seek

Example: How many balls would we have to sell in January to have total revenue for first 3 months of $4000?

1. Navigate to the **Data** tab, click **What-If Analysis**, and then click **Goal Seek**.
2. In the **Set cell** box, enter the reference for the cell that contains the formula that you want to resolve, ie. **G13**.
3. In the **To value** box, type the formula result that you want. eg. 4000
4. In the **By changing** cell box, enter the reference for the cell that contains the value that you want to adjust, ie, **D4**

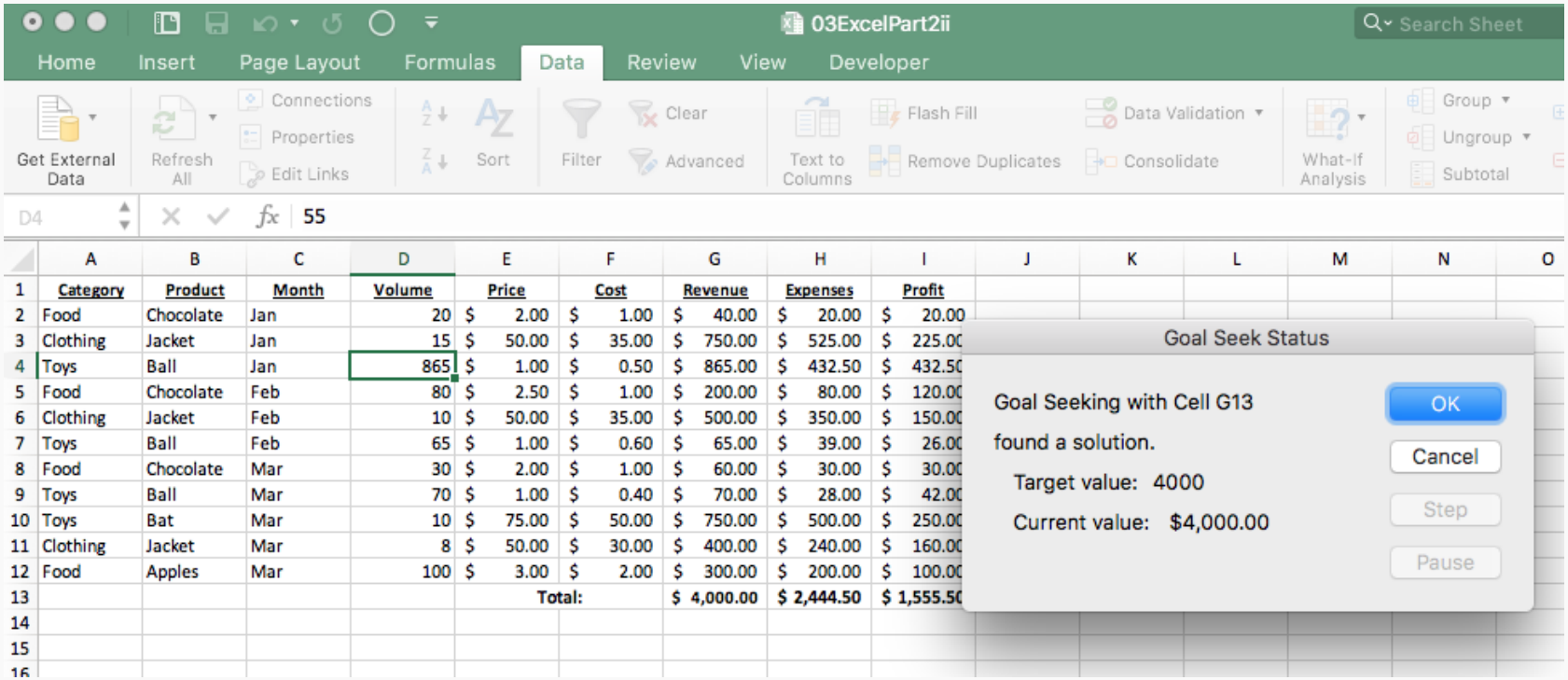
Goal Seek



	A	B	C	D	E	F	G	H	I
1	Category	Product	Month	Volume	Price	Cost	Revenue	Expenses	Profit
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00	$ 20.00	$ 20.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00	$ 525.00	$ 225.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00	$ 27.50	$ 27.50
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00	$ 80.00	$ 120.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00	$ 350.00	$ 150.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00	$ 39.00	$ 26.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00	$ 30.00	$ 30.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00	$ 28.00	$ 42.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00	$ 500.00	$ 250.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00	$ 240.00	$ 160.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00	$ 200.00	$ 100.00
13						Total:	$ 3,190.00	$ 2,039.50	$ 1,150.50
14									
15									
16									
17									

Goal Seek

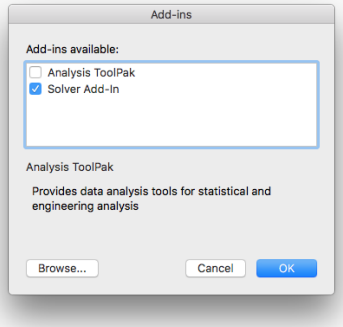
Answer: How many balls would we have to sell in January to have total revenue for first 3 months of $4000? **Answer: 865**



	A	B	C	D	E	F	G	H	I
1	Category	Product	Month	Volume	Price	Cost	Revenue	Expenses	Profit
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00	$ 20.00	$ 20.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00	$ 525.00	$ 225.00
4	Toys	Ball	Jan	865	$ 1.00	$ 0.50	$ 865.00	$ 432.50	$ 432.50
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00	$ 80.00	$ 120.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00	$ 350.00	$ 150.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00	$ 39.00	$ 26.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00	$ 30.00	$ 30.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00	$ 28.00	$ 42.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00	$ 500.00	$ 250.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00	$ 240.00	$ 160.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00	$ 200.00	$ 100.00
13					Total:		$ 4,000.00	$ 2,444.50	$ 1,555.50
14									
15									
16									

Add ons

- Goal Seek only allows a single input value, that is, we can only specify a single cell under the *By changing cell* field.
- In order to specify more than one input value we should instead use the Solver.
- To use the Solver Add-in, however, you first need to load it in Excel (click here for instructions on Windows)
  1. On the **Tools** menu, select **Excel Add-Ins**.
  2. Select the **Solver Add-in** check box from the Add-Ins available box.



- After you load the Solver add-in, the Solver button is available on the Data tab.

Solver

- Like Goal seek, Solver can be used to determine what values need to be changed to achieve a desired result. Unlike Goal seek, we can accommodate *multiple* cell values changing.
- In addition, rather than setting our desired result (called the objective cell) to a specific number, we could use Solver to find an optimal (maximum or minimum) value for a formula in one cell subject to certain constraints.
- Solver works with a group of cells, called decision variables.
  - These cells are necessarily inputs to the objective cell.
  - Solver adjusts these decision variable in order to obtain the desired result in your objective cell.

Solver Example

- For example, lets assume that our final grades are calculated using the following breakdown:

Assignments	10%
Midterm 1	25%
Midterm 2	25%
Exam	40%

- Let's assume we've already received 90% on the assignment component, 65% on midterm 1, and only midterm 2 and the exam remain.
- Let's use Solver to determine a scenario in which our resulting final grade would be 75%.

Solver Example

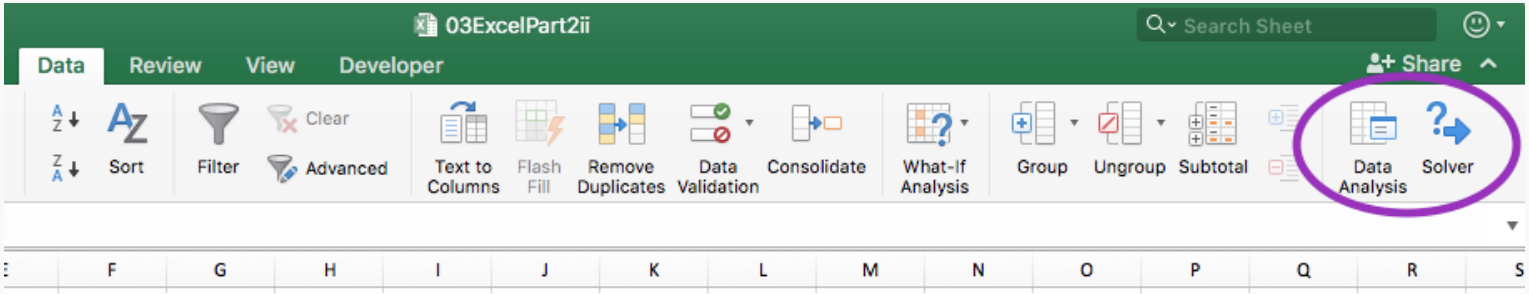
Consider the *Solver* starter worksheet on **03ExcelPart2ii.xlsx**

	A	B	C	D
1		**Score**	**% of total grade**	
2	**Assignment**	90	0.1	=C2*B2
3	**Midterm 1**	65	0.25	=C3*B3
4	**Midterm 2**	0	0.25	=C4*B4
5	**Final**	0	0.4	=C5*B5
6			Final Grade	=SUM(D2:D5)

- Our final grade stored in **D6** is our objective cell.
- **B3** and **B5** will be our decision variables.
- Notice how the calculation of **D6** depends on **B4** and **B5**.

Solver Example

- Navigate to the **Data** tab and click on the Solver button.



- Specify the objective cell, decision variables and any constraints in the the pop-up window.

- For this example our only constraints is that **B3** and **B5** must be ≤ 100 and ≥ 0.

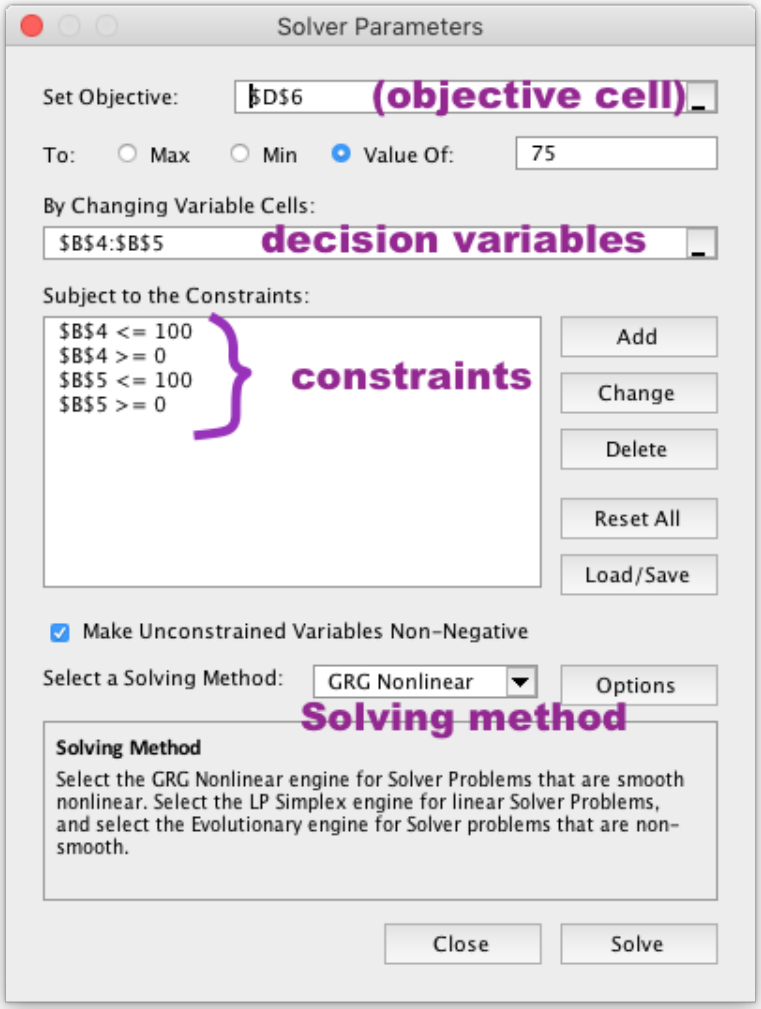
Sidenote other constraints include:

int Restricting cells to be an integer,

bin Restricting cells to be a binary digit (either 0 or 1)

dif Restrict cells to contain different (i.e. non-identical) value

Solver Example



	A	B	C	D
1		**Score**	**% of total grade**	
2	**Assignment**	90	0.1	9
3	**Midterm 1**	65	0.25	16.25
4	**Midterm 2**	55.899	0.25	13.9747192
5	**Final**	89.438	0.4	35.7752818
6			Final Grade	75.000001

Solver

You may have noticed that there are three different solving methods using solver (as chosen in the **Select a Solving Method** drop-down menu:

1. GRG Nonlinear
2. Simplex LP
3. Evolutionary.

A description of these methods are provided directly below the drop-down menu with further options available by clicking the **Options** button located to the right of this drop-down menu.

If you are unsure which one to use, it is usually best to select the GRG Nonlinear engine.

What-If Clicker Question

Example 4

How many of the following statements are TRUE?

1. We can only change a single cell in using Scenario Manager.
2. We can only change a single cell in using Goal Seek.
3. We can only change a single cell in using Solver. on the fitted line.

A) 0 B) 1 C) 2 D) 3

What-If Clicker Question

Answer:

How many of the following statements are TRUE?

1. We can only change a single cell in using Scenario Manager.
2. We can only change a single cell in using Goal Seek.
3. We can only change a single cell in using Solver.

A) 0 B) 1 C) 2 D) 3

What-If Clicker Question

Answer:

How many of the following statements are TRUE?

1. We can only change a single cell in using Scenario Manager.
2. We can only change a single cell in using Goal Seek.
3. We can only change a single cell in using Solver.

A) 0 B) 1 C) 2 D) 3

What-If Clicker Question

Answer:

How many of the following statements are TRUE?

1. We can only change a single cell in using Scenario Manager.
2. We can only change a single cell in using Goal Seek.
3. We can only change a single cell in using Solver.

A) 0 B) 1 C) 2 D) 3

What-If Clicker Question

Answer:

How many of the following statements are TRUE?

1. We can only change a single cell in using Scenario Manager.
2. We can only change a single cell in using Goal Seek.
3. We can only change a single cell in using Solver.

A) 0 B) *1* C) 2 D) 3

The Analysis ToolPak

The Analysis ToolPak is an another Excel *add-in* that has a set of statistical and data analysis tools such as ANOVA, covariance, regression, and t-test.

On a Windows: Analysis ToolPak is not installed by default.

- To install: **File** $\rightarrow$ **Options** $\rightarrow$ **Add-Ins**
- Select **Excel Add-ins** in the **Manage:** box and select **Go** ...
- Choose **AnalysisToolPak** and select OK

On a mac: **Tools** –> **Excel Add-ins** select *AnalysisToolpak*

Regression

Linear regression models the relationship between a dependent variable Y and explanatory/independent variable(s) X.

Simple linear regression has one explanatory variable, X, and can be modelled using the equation

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$

- You can think of β_0 as the y-intercept (b) and β_1 as the slope (m) is an equation for a line $y = m * x + b$ from high school.
- ϵ denotes the random error about the line which we cannot account for in our model

N.B. *Trend lines* are often calculated using linear regression.

Regression[1]

- The technique provides a way to determine linear patterns in the data set.
- The so-called fitted model can be used to:
  1. determine the strength of the relationship between variables Y and X and/or
  2. predict the outcome variable Y for a new variable X.
- To address either of these points, we need the fitted model comprised of the estimates of the regression equation coefficients.
  - $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates for β_0 and β_1 respectively.
  - the fitted line is given by $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

[1] N.B. To learn more on linear regression, I encourage you to take STAT 121/230 or read Faraway's Linear Models with R

Regression in Excel- R^2 value

- Regarding point 1. Excel regression function that will calculate R^2, the coefficient of determination which is a statistical measure of how close the data are to the fitted regression line.
- Mathematically speaking, this is the percentage of the variation in Y that is explained by a linear model.
- This range of this value is from 0–1.
  - A value of 0 indicates that the model explains *none* of the variability of the response data around its mean and
  - a value of 1 indicates that the model explains *all* the variability of the response data around its mean.

Regression in Excel- R^2 value

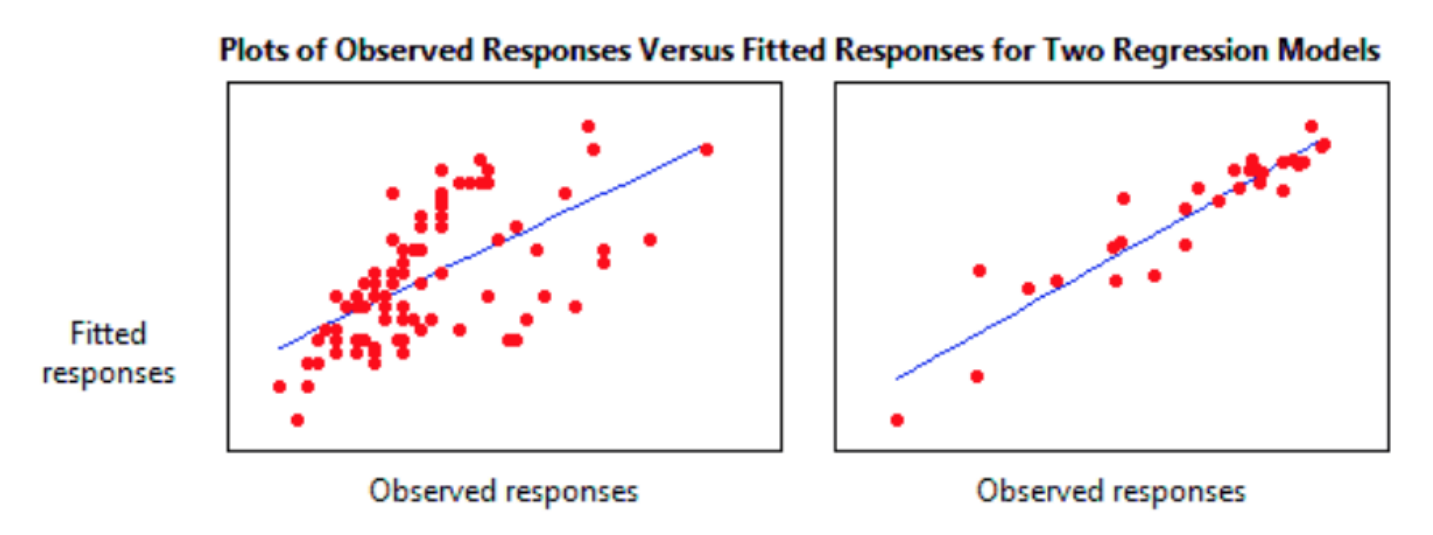


Figure: The regression model on the left accounts for 38.0% of the variance (ie. $R^2 = 0.38$) while the one on the right accounts for 87.4% (i.e has an $R^2 = 0.874$). Source of image here.

Linear Regression assumptions

The legitimacy of this model depends on the following assumptions:

1. **Linearity:** The means of the values of X fall approximately on the straight line $\beta_0 + \beta_1 X$ for simple linear regression, fall approximately on the plane $\beta_0 + \beta_1 X_1, \beta_2 X_2$ for multiple linear regression, . . . (harder to visual in larger dimensions).
2. The error terms $\epsilon \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. That is:
   - **Normality:** For each value of the explanatory variable (X) there is a subpopulation of response variables (Y) that is normally distributed.
   - **Constant variance (homoscedasticity):** All subpopulations have the same standard deviation σ^2.
   - **Independence:** All observations are independent.

N.B. non-constant variance = heteroscedasticity

Regression in Excel- Testing Assumptions

- To check the **linearity** assumption for simple linear regression, it is always a good idea to plot your data in a *scatterplot*.
- To check the **normality** assumption, we can look at the *Normal Probability Plot* of residuals.
  - If the data are normal the points should form an approximate straight line.
- The **constant variance** assumption can be assessed using a *residual plot* (where a residual is the difference between our observed and predicted y, i.e. $y_i - \hat{y}_i$).
  - A 'good' residual plot should show no pattern and look like random noise about 0.
- A pattern in the residual plot may suggests a violation of the **independence** assumption. Reason for this may include:
  - Time series data, related individuals (same family, same household).

Testing Assumptions Using Residual Plots

Figure: Image taken from Faraway's Linear Models with R (2005, p. 59).

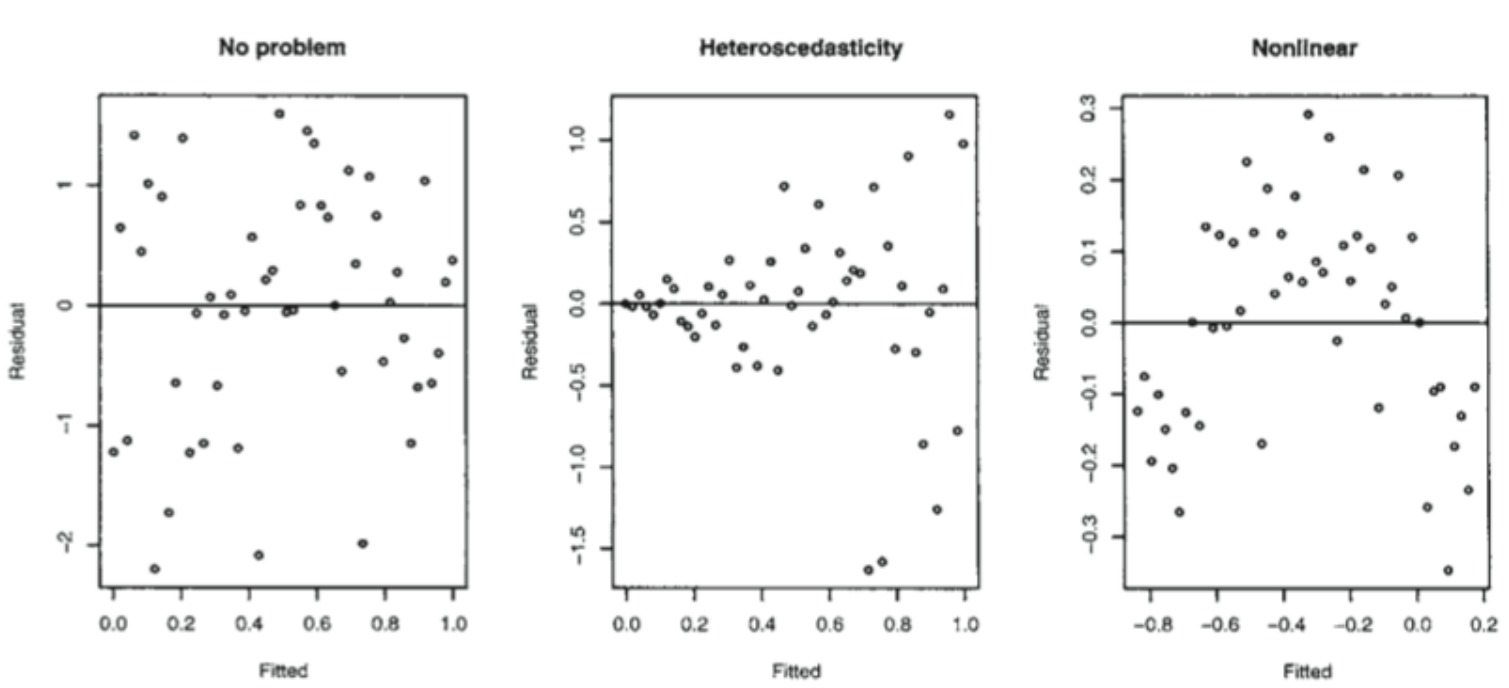
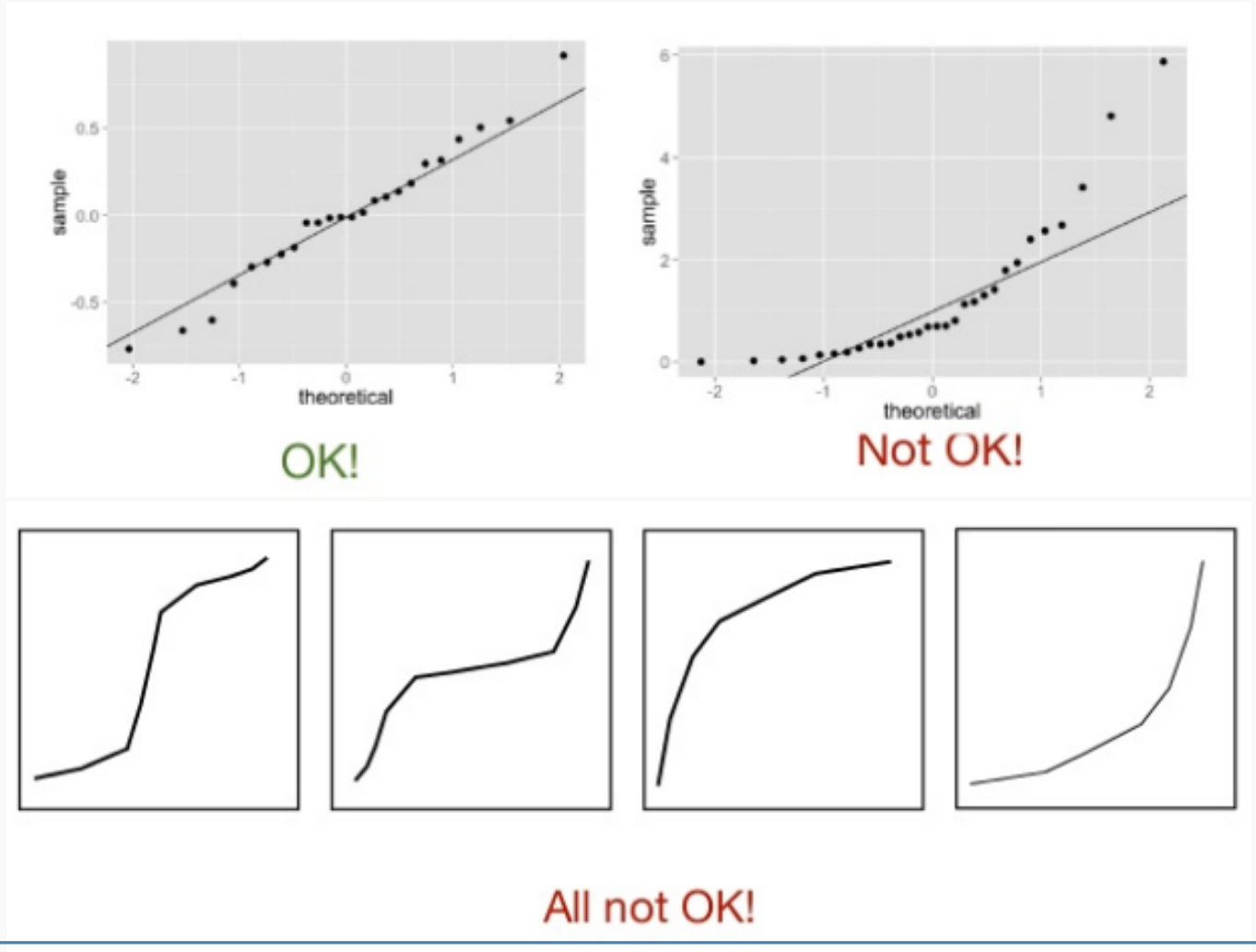


Figure 4.1 *Residuals vs. fitted plots—the first suggests no change to the current model while the second shows nonconstant variance and the third indicates some nonlinearity, which should prompt some change in the structural form of the model.*

Testing Assumptions Using Normal Probability Plots

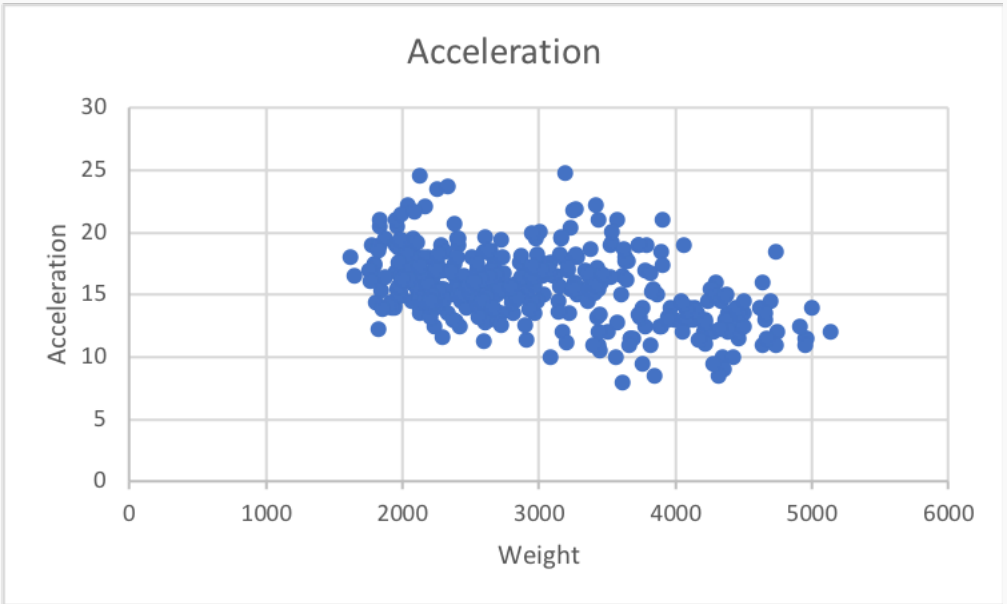
Figure: The points on a "good" normal probability plots should fall approximately on a straight line. Images taken from SlideShare



Regression in Excel

Example:

Given a data set of car weight and acceleration, determine if there is any relationship between them.



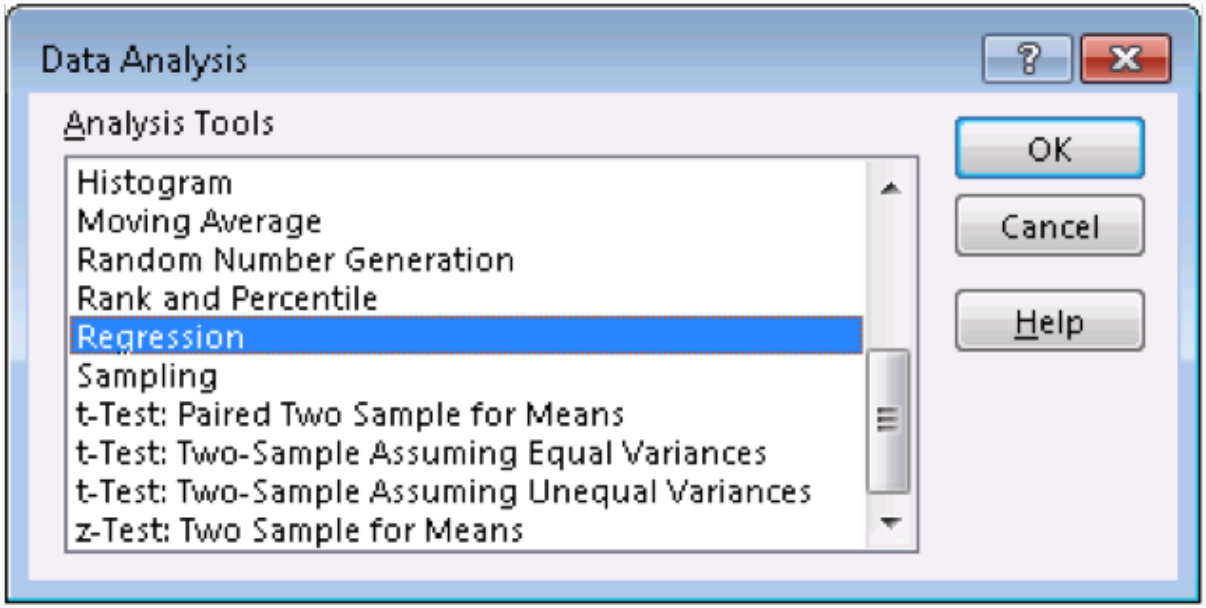
Scatterplot shows weak relationship with no strong patterns, and we would expect to see this shown in the regression analysis.

Regression in Excel

Regression computes constants m and b in formula:

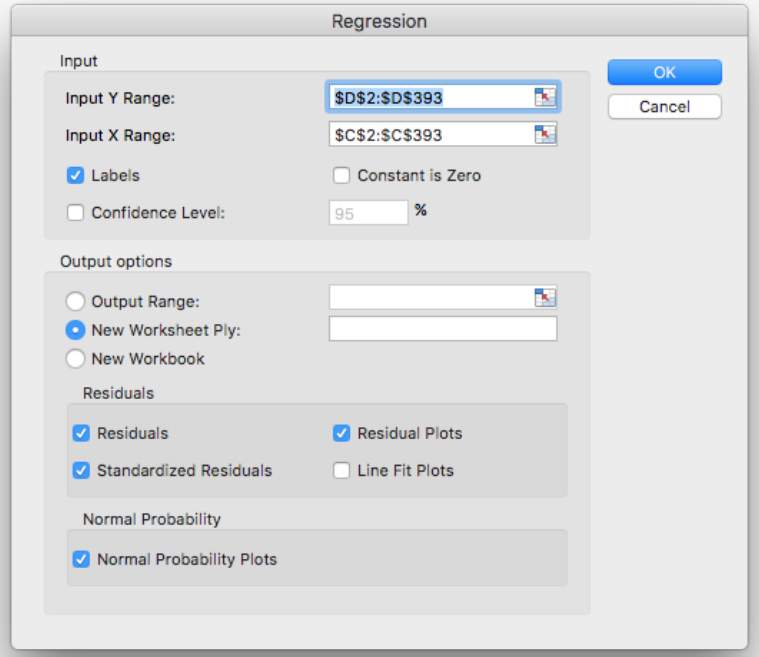
$$\text{acceleration} = m*\text{weight} + b = \beta_1*\text{weight} + \beta_0$$

Acceleration is the dependent variable and weight is the independent variable



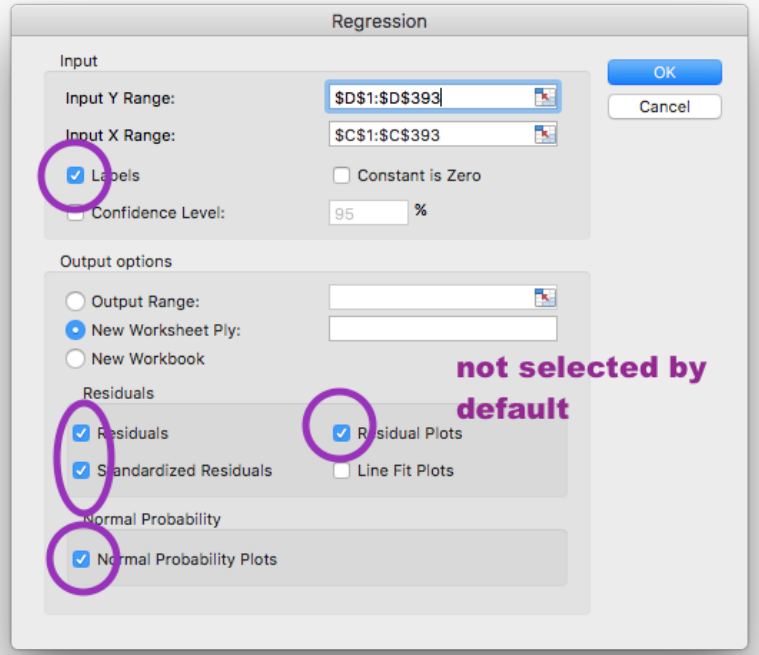
To start select, **Data Analysis** from the **Data** tab and then select **Regression** and **OK** .

Regression Example Settings



Settings:

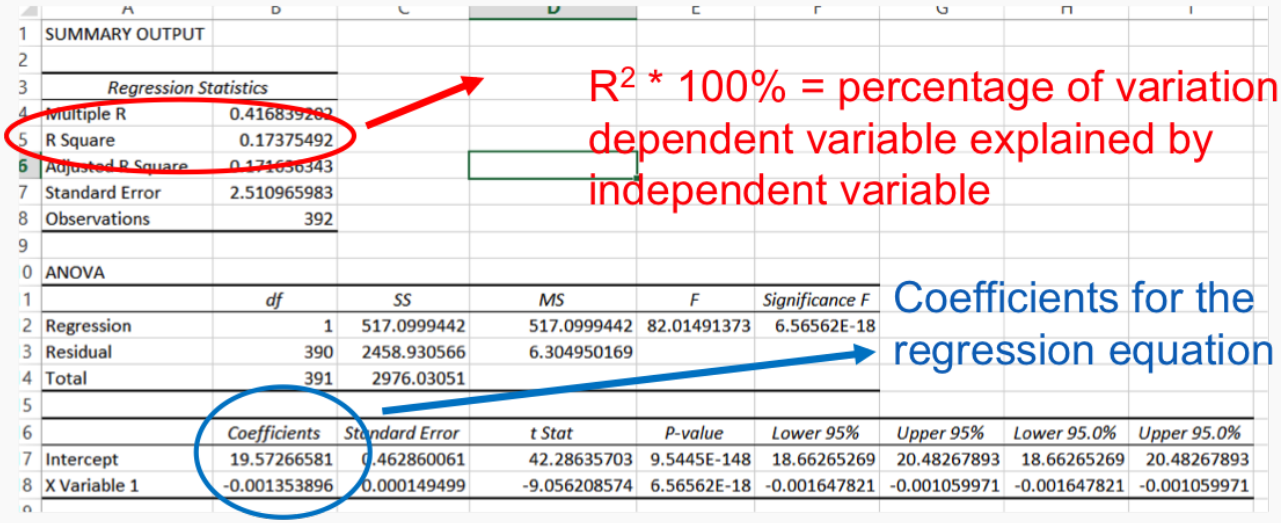
- Response (dependent) data for the Input Y Range
- Columns for the explanatory (independent) data (X Range).
- For residual information select, Residuals, Standardized Residuals, and Residual Plots from the Residuals section.



If you select the column headers, be sure to select the **Labels** options to ensure they are treated as meta-data.

Notice that many of the features regarding the residuals are not selected by default.

Regression Example Results



	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.416839282							
5	R Square	0.17375492							
6	Adjusted R Square	0.171636343							
7	Standard Error	2.510965983							
8	Observations	392							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	517.0999442	517.0999442	82.01491373	6.56562E-18			
13	Residual	390	2458.930566	6.304950169					
14	Total	391	2976.03051						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	19.57266581	0.462860061	42.28635703	9.5445E-148	18.66265269	20.48267893	18.66265269	20.48267893
18	X Variable 1	-0.001353896	0.000149499	-9.056208574	6.56562E-18	-0.001647821	-0.001059971	-0.001647821	-0.001059971

All of the output is put into a new sheet. Read the values off of the table and form the regression equation:

$$\text{acceleration} = -0.001353896*\text{weight} + 19.57266581$$

As one might expect, a weak relation is found as reflected by the R^2 value

- Only 17.4% of the variation in weight is explained by acceleration.

Regression Example Results (cont.)

Below the previous tables are the predicted y values (from the regression equation) as well as the residuals and standardized residuals (residuals divided by their standard deviation).

Observation	Predicted Y	Residuals	Standard Residuals
1	14.82861427	-2.828614269	-1.127947728
2	14.57272793	-3.072727927	-1.22529131
3	14.9206792	-3.920679196	-1.563423207
4	14.92474088	-2.924740884	-1.1662795
5	14.90307855	-4.403078548	-1.755786393
6	13.69540333	-3.695403327	-1.473591445
7	13.67780268	-4.67780268	-1.865336311
8	13.73466631	-5.234666311	-2.087393123
9	13.58167606	-3.581676065	-1.428241179
10	14.36016626	-5.860166257	-2.336819583
11	14.74873441	-4.748734406	-1.893621284

Regression Example Results (cont.)

Observation	Predicted Y	Residuals	Standard Residuals
1	14.82861427	-2.828614269	-1.127947728
2	14.57272793	-3.072727927	-1.22529131
3	14.9206792	-3.920679196	-1.563423207
4	14.92474088	-2.924740884	-1.1662795
5	14.90307855	-4.403078548	-1.755786393
6	13.69540333	-3.695403327	-1.473591445
7	13.67780268	-4.67780268	-1.865336311
8	13.73466631	-5.234666311	-2.087393123
9	13.58167606	-3.581676065	-1.428241179
10	14.36016626	-5.860166257	-2.336819583
11	14.74873441	-4.748734406	-1.893621284

eg. $X_1 = 3504$, hence $\hat{Y}_1 = -0.001353896 * \mathit{weight} + 19.57266581$

$$- 0.001353896 * 3504 + 19.57266581$$

$$= 14.82861423 \text{ (with some rounding error)}$$

Regression Example Results (cont.)

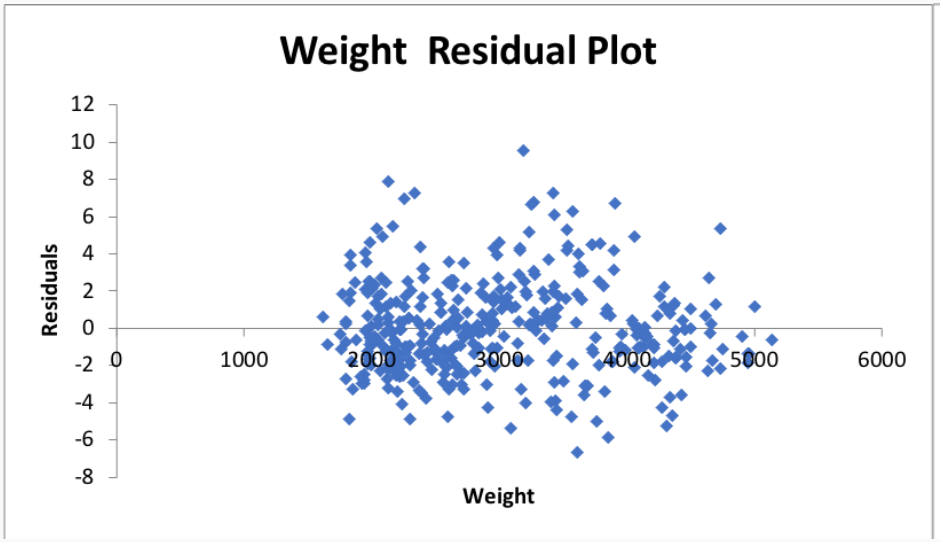
Observation	Predicted Y	Residuals	Standard Residuals
1	14.82861427	-2.828614269	-1.127947728
2	14.57272793	-3.072727927	-1.22529131
3	14.9206792	-3.920679196	-1.563423207
4	14.92474088	-2.924740884	-1.1662795
5	14.90307855	-4.403078548	-1.755786393
6	13.69540333	-3.695403327	-1.473591445
7	13.67780268	-4.67780268	-1.865336311
8	13.73466631	-5.234666311	-2.087393123
9	13.58167606	-3.581676065	-1.428241179
10	14.36016626	-5.860166257	-2.336819583
11	14.74873441	-4.748734406	-1.893621284

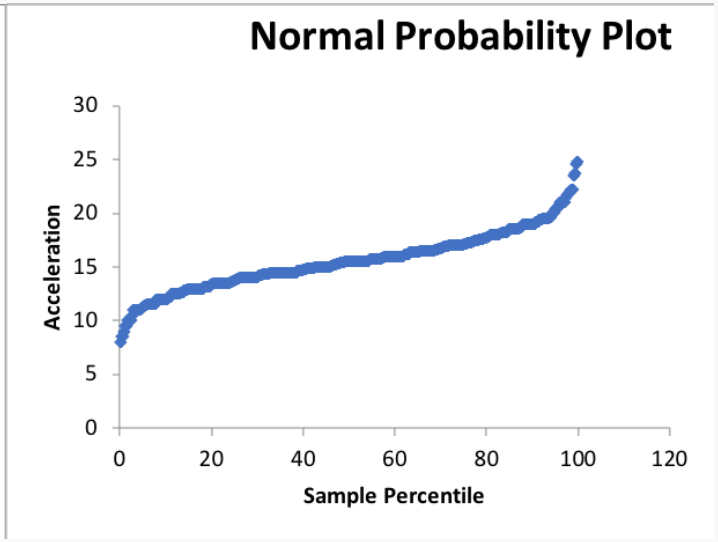
The observed acceleration for observation 1 was $Y_1 = 12$. To calculated the residual, we subtract the true value Y_1 from the predicted value $\hat{Y}_1$

$$\begin{aligned} e_1(\text{the residual of obs. } 1) &= Y_1 - \hat{Y}_1 \\ &= 12 - 14.82861423 \\ &= -2.828614226 \text{ (rounding error)} \end{aligned}$$

Regression Example Results (cont.)

All plots are placed to the right of the charts.





- This is a "good" residual plot since it shows no obvious pattern and points are symmetrically distributed around a horizontal line with a roughly constant variance.
- The Normal Probability Plot shows some curvature in the tails indicating that the residuals are showing some departure from normality.

03ExcelPart2ii.xlsx

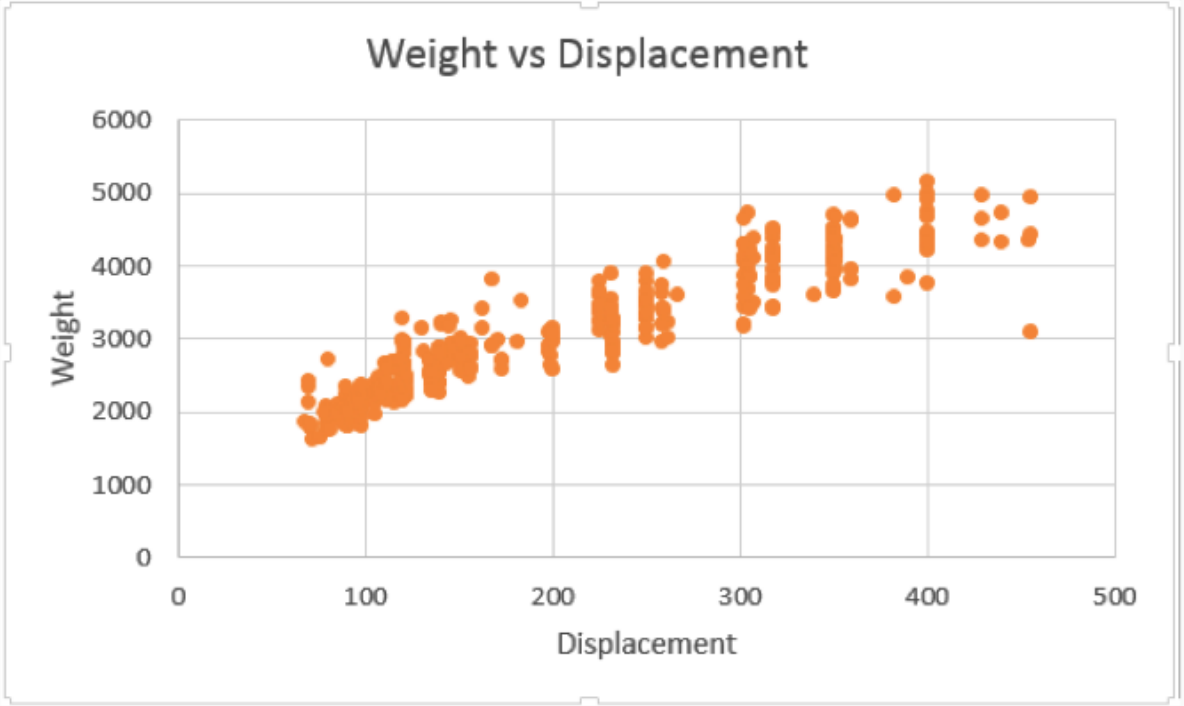
Regression Example Results (cont.)

As you can see, these are also the results obtained when fitting a trendline to the original scatterplot

Try It: Regression

Question:

Perform a regression analysis between weight (dependent) and displacement (independent) variable.



Example 5

How many of the following statements are TRUE?

1. Regression can only be used when we have one dependent variable Y and one explanatory variable X.
2. No assumptions are needed for the distribution of X in order to fit a linear regression model.
3. Residual plots may be used to test linear model assumptions.
4. A R^2 value of 1 means that all points in our data set lie exactly on the fitted line.

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. Regression can only be used when we have one dependent variable Y and one explanatory variable X.
2. No assumptions are needed for the distribution of X in order to fit a linear regression model.
3. Residual plots may be used to test linear model assumptions.
4. A R^2 value of 1 means that all points in our data set lie exactly on the fitted line.

A) 0 B) 1 C) 2 D) 3 E) 4

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. Regression can only be used when we have one dependent variable Y and one explanatory variable X.
2. No assumptions are needed for the distribution of X in order to fit a linear regression model.
3. Residual plots may be used to test linear model assumptions.
4. A R^2 value of 1 means that all points in our data set lie exactly on the fitted line.

A) 0 B) 1 C) 2 D) 3 E) 4

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. Regression can only be used when we have one dependent variable Y and one explanatory variable X.
2. No assumptions are needed for the distribution of X in order to fit a linear regression model.
3. Residual plots may be used to test linear model assumptions.
4. A R^2 value of 1 means that all points in our data set lie exactly on the fitted line.

A) 0 B) 1 C) 2 D) 3 E) 4

What-If and Pivot Tables Question

Answer:

How many of the following statements are TRUE?

1. Regression can only be used when we have one dependent variable Y and one explanatory variable X.
2. No assumptions are needed for the distribution of X in order to fit a linear regression model.
3. Residual plots may be used to test linear model assumptions.
4. A R^2 value of 1 means that all points in our data set lie exactly on the fitted line.

A) 0 B) 1 C) 2 D) *3* E) 4

Conclusion

Spreadsheets are general purpose tools for data analysis that consist of a table of cells which contain data and formulas.

Formulas contain data values, cell references, and functions.

- Aggregate functions summarize multiple data values into a single value.
- Functions exist for statistics, string manipulation, lookup/indexing, and decisions.

Spreadsheets provide tools for data sorting, filtering, visualization using charts, and summarization (pivot tables).

Also contain tools for what-if scenarios, goal seek, linear solvers, and statistical analysis tools.

Objectives

- Explain what a spreadsheet is.
- Explain how cells are addressed in a spreadsheet.
- List some of the ways to select cells in a spreadsheet.
- Define and explain: formula, function, argument, concatenation
- Use these functions: concatenate, lookup, index
- Explain the difference between an absolute and relative address.
- Explain how an aggregate function works. List some examples.
- Explain how to use conditional formatting.
- Use sorting and filtering.
- Be able to create and edit charts and use chart features: trendlines, sparklines
- Explain the usefulness of: what-if scenarios, goal seek, solver
- Use and create pivot tables and charts.
- Evaluate and create conditions. Use IF() to make decisions.