**What is Linear Regression?**

Linear regression is a way of demonstrating a relationship between a dependent variable (y) and one or more explanatory variables (x).

Demonstration

Connect to sample superstore data.

Create the following calculated fields

**2015 Sales**

{FIXED [Sub-Category]:SUM(

IF YEAR([Order Date])=2015

THEN [Sales]

else 0

END)}

**2018 Sales**

{FIXED [Sub-Category]:SUM(

IF YEAR([Order Date])=2018

THEN [Sales]

else 0

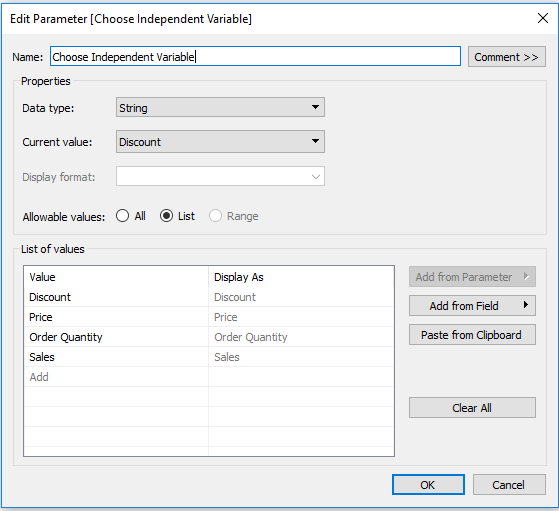
END)}

**Price**

{fixed [Order ID] : avg([Sales])}

**Create a parameter**

Choose Independent Variable



Click on show parameter control.

**Create a calculated field for the parameter control**

**Chosen Variable**

CASE [Choose Independent Variable]

WHEN "Discount" THEN AVG([Discount])

WHEN "Price" THEN AVG ([Price])

WHEN "Order Quantity" THEN sum([Quantity])

WHEN "Sales" THEN sum([Sales])

END

**Min Discount**

window\_min(AVG([Discount]))

**A correlation coefficient is a statistical measure of the degree to which changes to the value of one variable predict change to the value of another. ... A coefficient of -1 indicates a perfect negative correlation: A change in the value of one variable predicts a change in the opposite direction in the second variable.**

**Y=mx+c**

**Correlation Coefficient**

window\_corr(AVG([Discount]),[Profit Ratio])

**Correlation Coefficient Parameterized**

window\_corr([Chosen Variable],[Profit Ratio])

**R square or coeff. of determination shows percentage variation in y which is explained by all the x variables together. Higher the better. It is always between 0 and 1. It can never be negative – since it is a squared value.**

**R-squared**

window\_corr(AVG([Discount]),[Profit Ratio])^2

**Coefficient of Correlation: is the degree of relationship between two variables say x and y. It can go between -1 and 1.  1 indicates that the two variables are moving in unison. They rise and fall together and have perfect correlation. -1 means that the two variables are in perfect opposites. One goes up and other goes down, in perfect negative way. Any two variables in this universe can be argued to have a correlation value. If they are not correlated then the correlation value can still be computed which would be 0. The correlation value always lies between -1 and 1 (going thru 0 – which means no correlation at all – perfectly not related).**

**R-squared Parameterized**

window\_corr([Chosen Variable],[Profit Ratio])^2

**Marker**

IF SUM([2018 Sales])<SUM([2015 Sales])

THEN '▼'

ELSE ''

END

**Even Number**

INDEX()%2 = 0

**Shape**

ZN(SUM([Sales])) - LOOKUP(ZN(SUM([Sales])), -1)<0

**Zn is used to treat null values**

**More Colors**

IF [Shape] THEN 'R'

ELSEIF [Even Number]

THEN '2'

ELSE '1'

END

**Strength text**

if [R-squared] >= .6 then "strong"

elseif [R-squared] >=.4 then "average/medium"

else "weak" end

**Strength text Parameterized**

if [R-squared Parameterized] >= .6 then "strong"

elseif [R-squared Parameterized] >=.4 then "average/medium"

else "weak" end

**Sub Cat + Marker**

[Marker]+' '+ MAX([Sub-Category])

**Title Text**

if sign([Correlation Coefficient]) = -1 then "decreases"

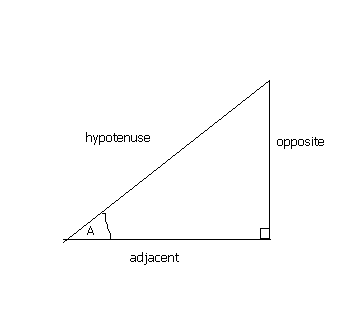
elseif sign([Correlation Coefficient]) = 1 then "increases"

else "doesn't change"

end

**Sin, Cos and Tan**

**A right-angled triangle is a triangle in which one of the angles is a right-angle. The hypotenuse of a right angled triangle is the longest side, which is the one opposite the right angle. The adjacent side is the side which is between the angle in question and the right angle. The opposite side is opposite the angle in question.**

****

**In any right angled triangle, for any angle:**

**The sine of the angle = the length of the opposite side  
                                   the length of the hypotenuse**

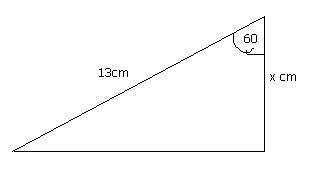
**The cosine of the angle = the length of the adjacent side  
                                      the length of the hypotenuse**

**The tangent of the angle = the length of the opposite side  
                                      the length of the adjacent side**

**So in shorthand notation:  
sin = o/h   cos = a/h   tan = o/a  
Often remembered by: soh cah toa**

***Example***

**Find the length of side x in the diagram below:**

****

**The angle is 60 degrees. We are given the hypotenuse and need to find the adjacent side. This formula which connects these three is:  
cos(angle) = adjacent / hypotenuse  
therefore, cos60 = x / 13  
therefore, x = 13 × cos60 = 6.5  
therefore the length of side x is 6.5cm.**

**Title Text (copy)**

if sign([Correlation Coefficient]) = -1 then "negative"

elseif sign([Correlation Coefficient]) = 1 then "positive"

else "flat"

end

**Title Text Parameterized**

if sign([Correlation Coefficient Parameterized]) = -1 then "decreases"

elseif sign([Correlation Coefficient Parameterized]) = 1

then "increases"

else "doesn't change"

end

**Title Text Direction Parameterized**

if sign([Correlation Coefficient Parameterized]) = -1 then "negative"

elseif sign([Correlation Coefficient Parameterized]) = 1 then "positive"

else "flat"

end

**1.Business Cases**

**Basic Regression What If**

**Drag discount to Columns shelf and change it to AVG(Discount)**

**Drag Profit Ratio to Rows Shelf and change it to Agg (Profit Ratio)**

**Drag Product name to the Details shelf in the marks card.**

**Drag Title text, Title text(copy) and Strength text to the details shelf and right click on each of them to get the drop down and then change it to compute using Product name.**

**Drag sum(profit) to the colors shelf. Change the colors as desired.**

**Drag R-Squared to the details shelf in the marks card.**

**Click on the Analytics tab and then drag the trend line to the worksheet and drop it to linear tab.**

**Click on the AVG(Discount ) on the columns shelf and drag it to the filter shelf and then show filter.**

**Click on the worksheet on the dashboard and the click on worksheet and then show caption and insert the desired details.**

**Change the format of min discount to percentage in the dimensions shelf and drag it to the details self the in the header insert the following text**

***What would be the impact to our profit ratio if we ran a sale offering at least <AGG(Min Discount)> off?***

**2.Regular regression**

**Same as the what if analysis**

**In the title use the following text**

**Comparing Discounts vs. Profit Ratio per Product**

***There is a <AGG(Title Text (copy))> relationship between Discounts and our Profit Margin. As Discounts increase, Profit Margin <AGG(Title Text)>. <AGG(R-squared)> of the variation or changes in Profit Margin can be attributed to changes in Discounts, indicating a <AGG(Strength text)> correlation.***

**3.Comparing the impact on profit margin from other variables**

**Drag Profit ratio to the rows shelf**

**Discount to the filter shelf and show filter**

**Show parameter control for the chosen Independent variable and also drag it to the columns shelf.**

**Sum(profit) to the colors shelf**

**Chosen variable to the columns shelf next to the chosen independent variable parameter (AGG (Chosen variable)).**

**Product name, Min Discount, Title text parameterized, Strength text parameterized, Title text direction parameterized, R-Squared parameterized to the details shelf and then compute using product name.**

**Go to analytics tab and drag trend line to the linear tab.**

**Enable title and caption as required.**

**Use the following text as caption**

**There is a <AGG(Title Text Direction Parameterized)> relationship between Discounts and our Profit Margin. As <Parameters.Choose Independent Variable> increases, Profit Margin <AGG(Title Text Parameterized)>. <AGG(R-squared Parameterized)> of the variation or changes in Profit Margin can be attributed to changes in <Parameters.Choose Independent Variable>, indicating a <AGG(Strength text Parameterized)> correlation.**