

Associative Statistics

How can we see **relationships** between things in our data?

Correlation

Value range: $-1 < r < +1$	
+	positive correlation: $\uparrow x \rightarrow y \uparrow$
-	negative correlation: $\uparrow x \rightarrow y \downarrow$

The closer is the correlation value to either 1 or -1, the stronger is the relationship, positive or negative.

+	temperature and sales	If the temperature goes up, the sales increase, and otherwise.
-	price and sales	If the price goes up, the sales drop, and otherwise.

Correltaion Matrix

	Lemon	Orange	Temperature	Leaflets	Price	Sales	Revenue
Lemon	1						
Orange	0,996714	1					
Temperature	0,477345	0,453116	1				
Leaflets	0,821304	0,790474	0,245969625	1			
				-			
Price	-0,27053	-0,31808	0,033574567	0,054939	1		
Sales	0,999309	0,999036	0,466616419	0,807828	-0,29257	1	
Revenue	0,469276	0,426955	0,339446501	0,588396	0,702259	0,450239	1

What is interesting with correlations is that they're **not causation**. A good example of this is that if we take a look at the relationship between temperatures and leaflets. It's a strong positive correlation, so if correlation meant causation, it would say - well, \uparrow temperature \rightarrow \uparrow leaflets, or if I put more leaflets out, then the temperature must be increasing. Those are what we call **spurious correlations**. There is a relationship, but odds are good that there's a third factor that's driving that relationship to be so strong. So, strong correlation does not necessarily mean that there is a causal relationship between those two variables. It could be a spurious relationship that is caused by a third, yet-to-be-identified factor that's in our data.

Excel does not tell you the significance of these correlations. So, we have some big correlations in here, but they may or may not be significant.