4/17/2021

Last week: linear functions and interpreting the slope and intercept

Linear regression:
Least-squares line/equation
Line-of-best-fit
Trendline

$$y = 0.0241x - 413.28$$

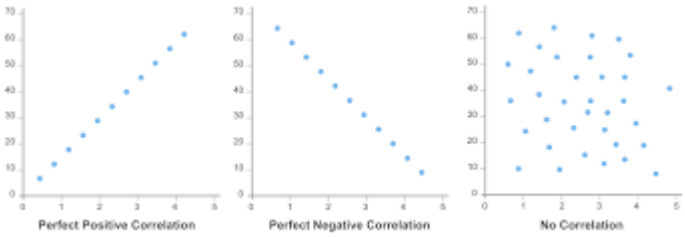
Our equation from Excel. X is cost, and y is units

For each unit increase in cost (in dollars), the number of units produced increased by 0.0241
For each $1000 increase in cost, the number of units produced increases by 24.1

Intercept? Not interpretable because it means that if you spend no money, you produce negative 413 units. Implication is that there are certain costs that have be spent before you can produce any units

Generally, use regression equations within the range of the available data. Our minimum cost is around $25,000 and so don't predict values far below that value.

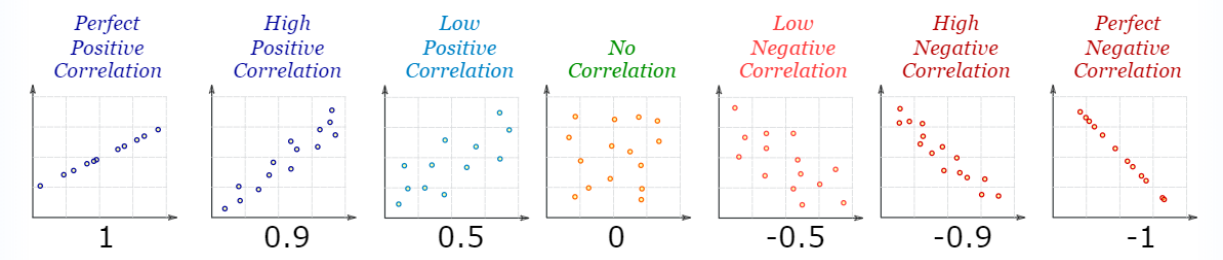
Correlation:



r if its from a sample, ρ is the correlation of the population
Takes on a value between -1 and 1. Positive one is the perfect correlation with a positive slope. Negative one is the perfect correlation with a negative slope. No correlation has a zero value.

When the correlation is zero, you generally: either the trendline is horizontal, or the data is curved. Correlation is always about the strength of the linear relationship.

Closer to 1 (or -1) the better the fit of the data to the trendline, and the closer to

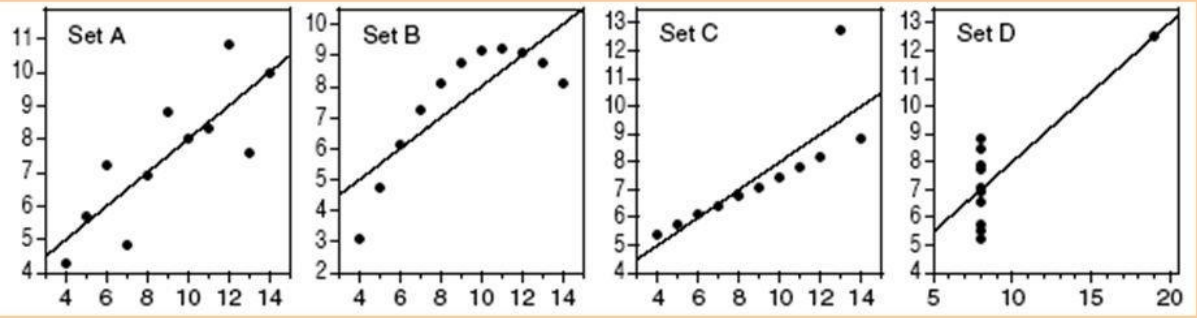


Strong correlation is generally considered to be 0.7 (or 0.8) or higher. (stress: this is an absolute value)
A moderate correlation would be roughly between 0.4 and 0.7.
A weak correlation would be less than 0.4

0.7 could classify as moderately strong.

All the same R value

However, making the scatterplots shows us that the correlation/ regression analysis is not appropriate for all data sets.



Moderate linear association; regression OK.	Obvious nonlinear relationship; regression inappropriate.	One point deviates from the (highly linear) pattern of the other points; it requires examination before a regression can be done	Just one very influential point and a series of other points all with the same *x* value; a redesign is due here…

Coefficient of determination r^2 or R^2 (lower case is specifically linear, capital is in general)

Formally, it is the percent of the variability in the y-value that can be explained by the relationship with the x-value.

For an R^2 of 0.7359, this means that of the overall variability in units (from first scatterplot), we have reduced the variability by 74%.

That means that our predictions are more accurate if we have the cost information.

The values of R^2 are between 0 and 1.