**Linear Regression / Multiple Linear Regression**

Simple Linear Regression = y = b0 + b1 \* x1

Multiple Linear Regression = y = b0 + b1 \* x1 + b2 \* x2 + … + bn \* xn

X = independent variables

Y = independent variables.

**Assumptions of a linear regression:**

1. **Data is linear** – test for this by plotting a scatter plot
2. **Homoscedasity**– homoscedasticity literally means“ having the same scatter.” Test for this by plotting on a scatter plot. If cone shape the data is likely hetroscedastic. A rule of thumb is to calculate the sample variance of each group. If the ratio between the largest to smallest is <= 1.5 then they meet the homoscedasicity requirement. Also homo in Greek means same scedastic means to scatter. Use Bartlett’s test if your data follow a normal, bell-shaped distribution. If your samples are small, or your data are not normal (or you don’t know whether they’re normal), use Levene’s test.
3. **Outliers in the data** – plot the data and check how much the outliers affect the result
4. **Independence of errors** – If data points follow a clear pattern – one point to the next might be affecting the next point
5. **Lack of multicollinearity** – if multi collinearity is high between independent variables it can be difficult to disentangle in the regression model which one is contributing to the model. To

For example: if tv and radio advertising spend are correlated then using regression it would be difficult to disentangle tv’s effect from radios effect. We might have a high R square. Our model might explain 90% of the variance in the companies sales. We might find that using tv or radio on its own might not be statistically significant. However together they are quite good at explaining variance in companies’ sales.

Bivariate correlations

VIF – Variance inflation factors

It should be noted that with multicollinearity that it doesn’t affect the fit of the model thus doesn’t deter its ability to predict. However, between independent variables – it is difficult to distinguish each of their contribution to the model due to high collinearity. <https://www.youtube.com/watch?v=Cba9LJ9lS8s>

