

Calculation of Diabetes Risk on Excel

1. Find correlations for all 7 attributes using CORREL function

2. Calculate the weight of each correlation using this formula:

Correlation of "x" variable

Sum(all 7 correlations)

By doing this we can see how the strengths of each health attribute compare with one another. For example, the correlation of glucose level vs Diabetic outcome was **0.309** while Blood Pressure vs outcome was **0.084**. If we see a high glucose level it has more of an effect on the risk level compared to high blood pressure.

3. Normalize the patient health values using this formula:

$$x_N = \frac{x - \text{minimum(all values in column)}}{\text{maximum} - \text{minimum}}$$

This is a basic formula that essentially calculates the percentile between the lowest to highest recorded value.

4. Multiply the normalization (x_N) of each patient health value to the corresponding attribute weighted correlation to calculate risk score

Ex: The 1st patient in the cleaned dataset is 22 yrs old, with a BMI of 2, glucose at 15.....

$$= (0.0167 \cdot 0.166) + (x_{N,BMI} \cdot C_{BMI}) + \dots = 0.0797$$

x_N of
age

weighted
correlation of
age

5. Categorize each risk score as "low", "moderate", or "high".

In this case a score of **>0.4** is high, **0.4 > x > 0.2** is moderate and **<0.2** is low