0.1 Influence model

With the groups created in the previous section, we can create a model, which predicts the performance of a given system. For this, we measure the non-functional property we want to predict for each group of features. Table 1 shows an example of such measurements. We only enable the features of one group at a time. The taken measurement then serves as the influence value of the group itself. In Table 1 each round of groupings is separated by a horizontal line. If we look at the first round of groupings, the influence value of G_1 is 3 and the features F_1 and F_2 are part of the group.

If in another measurement, the same configuration is chosen, we expect to see the influence value of the group as the measurement. Since measuring all possible groups (configurations) is not feasible, we want to determine the influence of single features to predict unseen configurations. By definition, only a few features are influential, this lets us assume that the influence measured in a grouping stems from only one or just a few features. In our example, this lets us assume that F_1 and F_2 have the influence of 3. With multiple different groupings, we can correct or confirm this assumption. The average influence value the feature has is our best approximation with the data available. If an influential feature is in a group, the group influence is most likely significantly higher. With enough groupings of different features, we can determine an influential parameter due to the fact, that the average influence of this feature is higher than that of the rest. In our example, we can see that the groups with F_6 assigned to them have, on average, a higher influence than the groups not containing F_6. With the average influence of the features, we can create a model of the system. We can use the formula for multiple linear regression as described by **?**].

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \varepsilon \tag{1}$$

We adjust the formula so that our parameters β_n represents the influence of a feature. The feature influences we determined in Table 1 represent the measured non-functional property of the system and include its baseline performance. This means, the performance of the system if all features would be disabled. We need to compensate for this since it would otherwise make the prediction of the model unusable. We can do this by determining the baseline performance and subtracting it from the influence values. An approximation of the baseline performance would be the average of all measurements taken. The model constructed would be described by following forumlas:

$$f_0 = \frac{1}{n} \sum_{n=1}^{|I|} I_n \tag{2}$$

$$f_n = I_n - f_0 \tag{3}$$

$$y = f_0 + f_1x_1 + f_2x_2 + ... + f_nx_n + \varepsilon \tag{4}$$

Where I_n is the average of all measurements taken which included F_n and x_n is either one, if the feature is selected or 0, if the feature is not selected. We can use this formula to predict the behaviour of the system on unseen configurations, but the accuracy of the prediction is most likely not very good. We can see why if we look at F_3 and F_5 and their respective average measurements I_3 and I_4. They both have high average values, due to the fact, that they share a group with our influential feature F_6. The effect on their influence values due to sharing a group with an influential feature would average down if the number of groupings would be increased, but it is still a problem. We try to compensate for this problem with a stepwise analysis of the influence values in **??**.

Table 1: Feature groupings and their influences

G_1	G_2	G_3	R	F_1	F_2	F_3	F_4	F_5	F_6	I_1	I_2	I_3	I_4	I_5	I_6
1	0	0	3	1	1	0	0	0	0	3	3				
0	1	0	6	0	0	1	1	0	0			6	6		
0	0	1	28	0	0	0	0	1	1					28	28
1	0	0	2	1	0	0	1	0	0	2			2		
0	1	0	7	0	1	0	0	1	0		7			7	
0	0	1	25	0	0	1	0	0	1			25			25
										2,5	5	15,5	4	17,5	**26,5**