1 More examples on Two-factor models

The previous example gave us a brief idea about what two-factor models are. We saw how the slope of the land on which we are looking for the yield of a particular crop and the variety affect the total yield of the land.This two inputs are categorical in nature. The variety had 3 levels and the slope of the land had 2 levels. Problems where we fit such a model that takes two categorical variables as input are generally termed as Two-way Analysis of Variance Model or in the short 2-Way ANOVA model. But for the time being, we just ignore the complexity of the terminology because of the fact that we don't even know what this "2-Way" literally mean . So to keep life simple we just call it the 2-Factor model and the name readily suggests that we are dealing with a situation where we have two factors affecting the situation and we need to fit our model based on these two factors.

The general format of a 2-factor model is simple. We have a black box which is essentially our model (we are calling it a black box because we are not sure about how these factors are interconnected with each other as well aswith the response). It has two input lines (because it is a 2-way model) and one output line. We call the inputs **"factors"** and the output **"response"**. The diagram here explains the

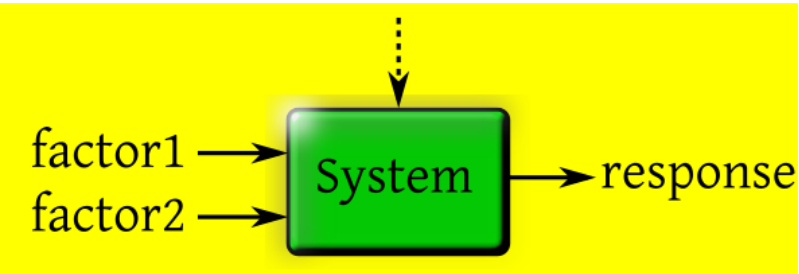

Figure 1: Two factor model

structure of a basic 2-factor model. But we should never expect "nice" things from any real-world structure because it seems God does play dice with the universe. There might be little things that we are ignoring as well as measurements that might not be accurate enough and errors will be there which can not be controlled by the

experimenter. The dotted arrow in the figure denotes the unavoidable random error in the system.

Example. Let us look at another example, similar to the previous set-up where we replace the factor "tilt" of the land with the factor "fertilizer". Our response is still the yield of the land. So what is the difference between the previous model and this one? In the previous model, we had one treatment factor (variety) and one control factor(tilt) as input. But in this case, we have both the factors as treatment factors because it is up to us which land will be cultured with which fertilizer and variety. The two models being different the method of experimenting with both cases will also be different. In the previous case, we had 24 plots and we divided the land into two sections each with 12 plots in it. From there we selected 4 random plots each for every factor. Here we have 3 varieties and two fertilizers. Denote the factors with f_1 and f_2 and the varieties with v_1,v_2 and v_3. Note that after the assignment of a particular variety and fertilizer to a land if we choose randomly then the event that a land is assigned variety v_i and fertilizer f_j should be independent of each other. there are six pairs of possible combinations of the factors (v_1, f_1), (v_1, f_2), (v_2, f_1),(v_2, f_2),(v_3, f_1),(v_3, f_2) . So in this case we randomly divide 24 lands in 6 equal parts and randomly allocate each pair to each set. For the allocation purpose, we may follow the SRSWOR method i.e. first we choose 4 random plots and assign a combination. Then from the remaining, we choose again 4 plots and assign the next pair and so on.

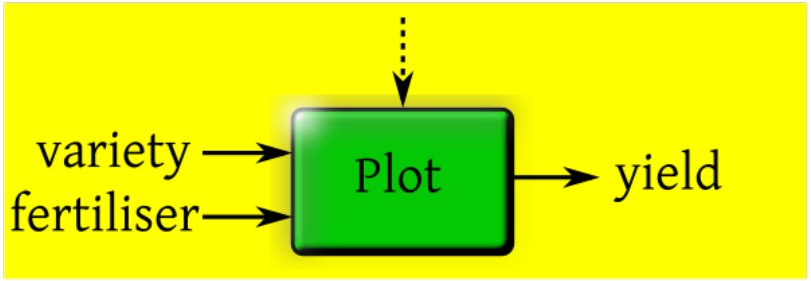

Figure 2: effect of variety and slope of land in crop yield

Example. Now here is yet another example . In a clinical study, we are trying to

find the effect of a particular drug (or different drugs) on a patient. Now the black box a.k.a. the system is the patient and the two factors associated with the patient are "gender" and the "dose of the drug" ("type of the drug"). Note that gender is a control factor. Choosing a random patient and assigning a gender is meaningless. Instead what we can do is we can make the count of each gender equal to make the test more symmetric. To make the example compatible with the previous case suppose we are performing the analysis on 24 patients and hence we will randomly choose 12 men from a large population of male patients and equally 12 females from a large population of female patients. Though we should be cautious about other factors like the patients should be of the same age group or their past medical history should not differ that much. Otherwise, the analysis will be introduced with unwanted errors. Next, we assign the doses (type of drug) randomly. Suppose we have three different doses (three different drugs) and we choose 4 people from males and assign the first type, then choose another 4 from the remaining and assign the second type and so on. Same thing we do for female patients. And after allocation, we perform the experiments where our measurements will be some kind of diagnostic measurements like the amount of sleep or blood pressure. Also there will be random error which is denoted by the dotted arrow in the figure below.

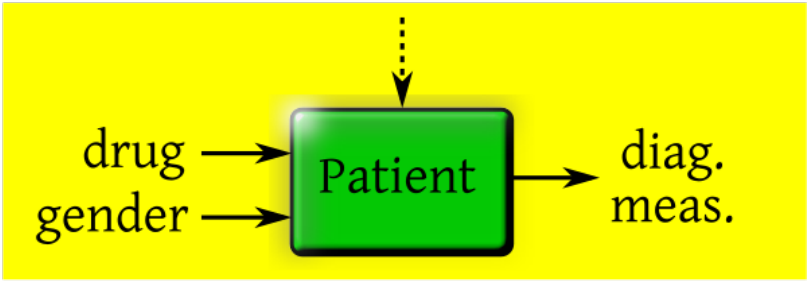

Figure 3: Effect of a certain drug for different doses on different genders

Example. Here is another example. This time our basic system is the student. In India, people get their high school education in different languages. So a social experiment was conducted on whether the vernacular language or the English language affects the performance of the students. The performance of the student is measured

through his/her marks. Another factor playing a role in the system might be the location of the student. Suppose we are trying to answer the question of whether there is a difference between the performance of the students studying in Bengali medium and students studying in English medium. And there are two categories of location let's say, urban and rural. For this experiment note that this system has both the input factors are control factors. Because the experimenter is not going to choose a random student and assign the medium of study and the location. The combination (location, medium) should be set prior to the sampling and then random samples in that particular category should be chosen. Because the students we are sampling from are studying in certain schools and they are already living in certain areas which have been affecting them all this time and we are exactly trying to find that effect. We are just passively observing them. This is also an example of a 2-factor model. there are 4 possible categories and we might choose 10 students and then look at their marks to decide their performance.

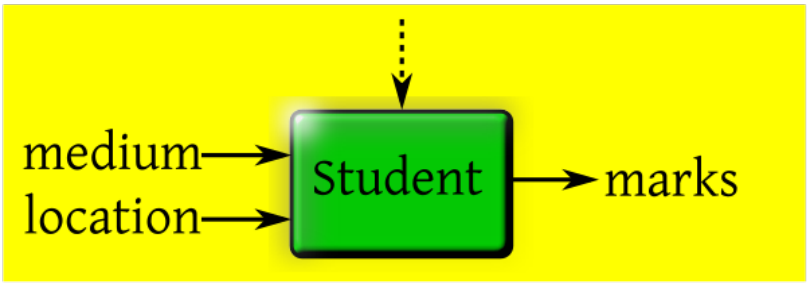

Figure 4: Effect of medium and location of a student on his/her marks

Definition. Treatment factor A factor which can be controlled by the experimenter at will.

Definition. Control factorA factor which can not be be controlled by the experimenter at will.