We wish to understand the variables which affect the incidence of malnutrition. The possibilities are many; the type of work undertaken by the mother, the caste or religion of the family, the education status of the mother and many such more. For now, we will begin by looking at whether the work profile of the mother has any effect on the nutrition status of both, the mother as well as the children.

For the mothers, BMI is used as a yardstick to determine the nutrition status. Whereas for children; their height-for-age is considered for the same purpose.

We are performing tests to see whether the proportion of malnourished mothers/children in one work category, differs statistically from the proportion in another category.

Before looking at the output of the statistical test and their interpretations, let us understand the variables we are going to use as well as the statistical theory behind the tests:

* Although the nutrition status is primarily split into four categories, for the purpose of this analysis we are going to use only two statuses; ‘Adequately nourished’ and ‘Undernourished.’
* The nutrition status variable is named ‘Nutri\_status\_bmi’ for mothers and ‘Nutri\_status\_childheightforage’ for the children.
* As we will be testing out various ways of categorizing the work, variables named ‘Work\_category\_1’, ‘Work\_category\_2’ etc. may appear. A description of each type of categorization will be provided at the beginning of each set of analysis
* The analysis is performed using R software
* We are performing a test named ‘Fisher’s exact test for Count data’ with a continuity correction. [This](http://en.wikipedia.org/wiki/Fisher%27s_exact_test) is what Wiki has to say about the test. But simply speaking, this test is doing exactly what we require! It tests whether the proportion of samples with characteristic A (say) rather than B, is different across two categories of samples. In our case, we test to see whether the proportion of undernourished mothers/children is statistically different across various work categories.
* Now the most important part. How do we decide whether the difference in proportions is ‘statistically significant’ or not? This decision is made using something called the ‘p-value’ of the test. Most of us might be aware of the idea of null hypothesis. A statistical test is performed by first assuming a status-quo situation. In the test for proportions, the null hypothesis is; ‘the proportions across two groups in question, are equal.’ After performing certain stipulated calculations, we declare whether we can reject the null hypothesis with a certain degree of confidence, or not.

There is something more to it. Simply saying that the proportions are ‘not equal’ is one of the options. In testing to see whether we can reject the null hypothesis (that of equality), our alternate hypothesis can also be that proportion of A in group 1 is greater than that in group 2, or is lesser than in group 2 or simply that they are note equal. In case of our tests, I have considered each of these alternate hypotheses to see which one fetches the lowest p-value.

So coming back to the p-value! A p-value tells us the percentage of times we may be wrong, if we reject the hypothesis. A p-value of 0.37 would imply that 37% percent of the times I may be wrong in saying that, the proportion of undernourished children with mothers working as casual labour is different than the proportion of undernourished children to mothers with no income source (this is just an example). Notice that our null hypothesis is that the proportion of undernourished children in both these categories is equal. And therefore rejecting it, is saying that those proportions are equal.

But just knowing the interpretation of p-value is not enough. So which p-values may be good? Which may be harmful? Well, that differs from context to context. In a medical science experiment to see whether the proportion of a certain chemical in a cough syrup is equal across two brands, the maximum tolerated p-value may be 0.001 (which means that we do not want to be wrong more than 0.1% of the times; as we are dealing with people’s lives and well-being in general). But in most social science studies, slightly higher p-values are acceptable for inference, given that we are studying human behavior, social structure etc. and it is very difficult for it to be stable. Even in social science, there is no set limit. Generally, a p-value of upto 0.1 (10% chance for error) is considered alright. But in our context, you would be the best judges. You understand the ground realities, you have an intuitive understanding of what intervention may work and which may not and foremost, you know how much we are ready to experiment knowing the number of times there is a ‘possibility’ of it not working. For afterall, no amount of number churning can paint an exact picture of how people would behave and there will always be unobservable factors playing a part in the whole process.

With this; given below are the results of some of the tests I have performed till now. I am reporting the tests with the lowest p-value, within each section.

1. **Classification one:** No Livestock and Livestock+
2. Nutrition status of mothers:

**p-value = 0.4789**

alternative hypothesis: greater

(Proportion undernourishedNo livestock>Proportion undernourished Livestock+)

(**Interpretation: approximately 48% of the times, we will be wrong in saying that the above relationship holds)**

1. Nutrition status of children: \*

**p-value = 0.1577**

alternative hypothesis: less

(Proportion undernourishedNo livestock<Proportion undernourished Livestock+)

(**Interpretation: approximately 16% of the times, we will be wrong in saying that the above relationship holds, which in my opinion, is not that bad)**

1. **Classification two:** Casual labour only and Livestock+
2. Nutrition status of mothers:

**p-value = 0.4093**

alternative hypothesis: less

(Proportion undernourishedCasual labour only< Proportion undernourished Livestock+)

(**Interpretation: approximately 41% of the times, we will be wrong in saying that the above relationship holds)**

1. Nutrition status of children: \*\*\*

**p-value = 0.05856**

alternative hypothesis: greater

(Proportion undernourishedCasual labour only> Proportion undernourished Livestock+)

(**Interpretation: approximately 6% of the times, we will be wrong in saying that the above relationship holds, which is excellent!)**

1. **Classification three:** Livestock+ and No Livestock (but there exists some other source of income, as indicated by the variable ‘Do not have a source of income.’)
2. Nutrition status of mothers:

**p-value = 0.4989**

alternative hypothesis: less

(Proportion undernourishedLivestock+ <Proportion undernourishedNo livestock, some income)

(**Interpretation: approximately 50% of the times, we will be wrong in saying that the above relationship holds)**

1. Nutrition status of children:

**p-value = 0.3074**

alternative hypothesis: greater

(Proportion undernourishedLivestock+> Proportion undernourishedNo Livestock, some income)

(**Interpretation: approximately 31% of the times, we will be wrong in saying that the above relationship holds)**