Notes on regression analysis

## Factors associated with variations in food security and income measures

For the first two tabs in the spreadsheet (“**Country**” and “**Partner**”), I was looking at the factors that are associated with variations in the 4 food security (FS) indicators (Food Frequency Score (FFS), Food Diversity Score (FDS), Seasonal Food Security Score (SFSS) and the Coping Strategy Index (CSI) as well as the self-reported Change in Income perceived to be associated with participation in the program. This is partly in response to the question “How crucial was literacy training to supporting observed outcomes in other sectors?” – this would include FS and income.

1. These results use the entire dataset since all respondents were asked the questions in parts A and B.
2. With the exception of the Change in Income, simple linear regression (Ordinary Least Squares or OLS) was used with each of the FS indicators as the independent variable. For the income measure, since it was not a continuous variable but an ordered composite of the questions B3 (i.e., did your earnings increase? Did they increase significantly?) that took on the value of 1 if the respondent experienced increased earnings and 2 if the increase was by a significant amount, I used an Ordered Probit regression.
3. The first 5 sets of results look at each of the dependent variables of interest. The last 3 incorporate the SFSS and the CSI as explanatory variables for the two main FS indicators (FFS and FDS) and the Change in Income.
4. The two tabs (“Country” and “Partner”) have the same sets of independent variables with the exception of the country and partner – the first controlling for country and the second for partner. You will note that the significance of the other independent variables remains much the same regardless of whether including partner or country.
5. There are some cells where there is no estimated coefficient (such as “age\_50\_plus” and “bangladesh” or one of the other countries or partners). These are omitted from the regression analysis because of collinearity – it is not possible to include all of the dummy variables for either age or country or partner. That being said, this does not present a problem for interpretation of the data. For example, statistically significant coefficients on the other age-related variables are simply interpreted as saying that the result (e.g. the SFSS) for respondents under 30 was less by 0.429 than for those aged 50 plus.
6. **Gender and Age** – for the most part, there were no significant differences in the FS or income results based on age or gender. The exception to this was the SFSS, where men had a lower score than women and younger respondents had a lower SFSS than those over 50 years of age.
7. **Country and Partner** – the country and partner coefficients are, for the most part, highly significant as would be expected. Each country and partner is in a different context and all this says is that the levels of FS differ from one to the other. However, by including them in the regression we are able to account for the country- or partner-specific variations in the FS and income variables so that the variation associated with other variables stands out more clearly.
8. **Program participation** – the results for both sets of models show that participation in all 4 programs was associated with better FS outcomes (higher FFS, higher FDS, lower SFSS and lower CSA). The strong association between increased income and participation in both the VSL and agriculture programs is also positive. While these results are encouraging, they should be interpreted with care as the FS dependent variables are not measuring the change in FS over the life of the project (this would have required a panel data set based on return to the same sample of respondents as in the baseline – or a self-reported measure of whether FS had improved). It could be that participants in the program were already more food secure than non-participants.
9. **Participation in literacy training** – when controlling for the country, most of the FS coefficients are highly significant for participation in the literacy training. However, when we control for the partner instead of the country, the coefficients are small and insignificant. This may have to do with the fact that not all partners participated in the literacy program and this variation is being accounted for in the partner-specific coefficients. While one could run a separate regression for each partner, I doubt this would yield useful results since the sample size for some would be too small to be of use.
10. **Changes in livelihood activities** – there are few definitive conclusions to be reached by the changes (activities that have been added or discontinued during the life of the project). The negative signs on the addition of fishing and livestock sales to the household’s activities seem to imply that those who added them are less food secure (FFS, FDS decrease), but this cannot be said with any certainty. It may be that these people were less food secure to start with and have adopted these activities in an effort to improve. While their FS measures are still lower than average, they may very well have improved. In fact, given the positive and significant coefficient on the change in income, this might be the case. One might reach a similar conclusion for those who added processing as an activity. Those with lower FS added it and it had a positive impact on income. With respect to discontinued activities, those who stopped selling livestock products seemed to have better FS as measured by FFS and FDS but poorer as measured by the CSI. Perhaps what really happened was that they were facing some challenges (as seen by the CSI) and so stopped selling things, thereby preserving their FFS and FDS. That there was no impact on income might suggest that they no longer needed this income as much.
11. **Changes in household assets** – acquisition of some assets over the life of the project seems to be associated with better FS outcomes – but not universally so. For example, both addition of a mobile phone are positively associated with improvements in the FFS but the opposite in terms of the SFSS (a higher SFSS is a poorer outcome).
12. **IGA-related variables** – there is a significant and positive association between these and both the FS and income measures. Given the emphasis of the project on livelihoods, this is a positive outcome. However, as mentioned above, one does need to be careful with the interpretation. It could be that the more food secure people were the ones who got involved and developed a plan and were already more confident because of their status in society. However, the coefficients are also positive and significant on the self-reported degree of change in income.
13. **Income and food security** – increased and more diverse incomes are positively associated with better food security all across the board.
14. **Periods of hunger and use of coping strategies** – the 3 last regressions used the SFSS and CSI as independent variables that have an impact on FFS, FDS and Income. This is probably the more appropriate place to include them (as opposed to being independent variables on their own) since they are not an outcome in the same way as the others. They are significant and their signs are as expected.
15. **Goodness of fit** – the 1st, 2nd, 6th and 7th linear regressions explain almost 50% of the variation in the data. While this is quite good, it does mean that there is a considerable amount that is not captured. Had the household survey included a more complete socioeconomic and demographic profile of the household, this could have been improved considerably. The ordered probit results (the 5th and 8thcolumns) don’t use the same measure of goodness of fit. However, they could benefit from the same improvements as the others. The inclusion of the SFSS and CSI in the regressions in columns 6, 7 and 8 helps to improve their fit.

# Conservation agriculture (CA) in relation to FS and income outcomes

1. The two sets of 4 linear regressions on the 3rd tab compare the FS outcomes associated with 4 different ways of measuring the adoption of Conservation Agriculture (CA) and are repeated controlling for the country or the partner. In other words, columns 1 and 5 have the same independent variables except for the country or partner – and so on. The results are based on those participating in the agriculture program and who are using one or more CA practice (1,494 observations instead of the 3,074 that are in the entire combined dataset).
2. **Gender and age** – there are no significant differences in FS outcomes for those involved in the agriculture program across age or gender.
3. **Program participation** – note the strongly positive and significant association between involvement in the community leadership program and the FS outcome.
4. **CA principles** – the 1st and 5th columns include a set of interaction terms between the 3 CA principles. Looking at columns 2 and 6, it is clear that all 3 principles are important – 2 and 3 in particular – given the relative magnitude of their coefficients. In the 1st and 5th, the second principle is the most important. However, it is also important to note that it is also the one most difficult for farmers to adopt. Mulch is typically difficult to come by yet very important for the success of CA. Whether one measures CA adoption by the application of any 2 of the 3 principles (“CA\_practitioner”) or by the application of all 3 (“CA\_full\_practitioner”), it is clear that there is a positive and significant association between them and FS as measured by the FFS.
5. **Security of tenure** – it is important to note that security of access to agricultural land is positively and significantly associated with better FS outcomes.
6. **Periods of hunger and use of coping strategies** – as expected there is a significant relationship between these 2 variables and the FFS – and the sign on the coefficients is as expected.

# Links between dietary diversity and animal production activities

1. This set of linear regressions (on the 4th tab) considers the question: to what extent is more dietary diversity associated with animal husbandry? The results are based on all those who indicated their participation in the agriculture program (1,789 respondents).
2. **Gender and age** – there is some indication that men have a lower level of dietary diversity than women. On the other hand, the under-30 age group appears to have a higher level of dietary diversity than their elders.
3. **Program participation** – note the strongly positive and significant association between involvement in the community leadership program and this FS outcome (FDS). The negative (and significant) association with participation in the VSL program is somewhat puzzling. However, it may simply be telling is that the poorer, less food secure respondents are also involved in the VSL program, which would be a positive situation given that the VSL program should contribute positively to improved incomes. Repeating the analysis with the Change in Income as the dependent variable will clarify this (see #31 below).
4. **Livestock types raised** – there are few coefficients that are statistically significant, but those that are have the correct sign. Of those that are (raising bees and raising fish in rice fields), they were only adopted by a very few (18 out of 24 and 47 out of 48 respectively) – so it is probably safe to say that the presence of livestock alone had little association with a more diverse diet. This is particularly true since the data only indicate the presence of one or more of a specific type of animal and nothing about the degree to which they might actually contribute to the household’s livelihood.
5. **Animal husbandry practices** – on the other hand, use of improved feed, improved herd management and vaccination of livestock were practices that were strongly associated with a more diverse diet. While this does not imply causation, one might reasonably conclude that those who adopt these improved practices are serious about animal husbandry and are more likely to have sufficient production to add to the diversity of the diet.
6. **Other variables of note** – increased knowledge, secure tenure, and increased income have significant and positive relationships with dietary diversity. On the other hand, simply practicing animal husbandry on its own does not. However, the negative value on the coefficient here is more than counteracted by the strong positive relationship apparent on specific animal husbandry practices.
7. **Periods of hunger and use of coping strategies** – as expected there is a significant relationship between these 2 variables and the FDS – and the sign on the coefficients is as expected. It should be noted that the magnitude of the coefficient on the SFSS (number of hungry months) is so small as to be insignificant, which may be a sign that raising animals to some extent has a positive role to play in insulating from periods of hunger.

# Links between change in income and animal production activities

1. This last set of linear regressions (on the 5th tab) considers the question: to what extent is an increase in household income associated with some form of animal husbandry? The results are based on all those who indicated their participation in the agriculture program (1,789 respondents).
2. **Gender and age** – there are no significant differences in income outcomes for those involved in some form of animal husbandry across age or gender.
3. **Program participation** – note the strongly positive and significant association between involvement in the VSL program and this self-reported change in household income. This confirms the observation made in #24 above.
4. **Livestock types raised** – there are only a few coefficients that are statistically significant, but they are fairly large and have the correct sign. Poultry have the largest impact, followed by fish ponds and cattle. It is interesting that the coefficients on small livestock are not statistically significant.
5. **Animal husbandry practices** – none of the estimated coefficients for animal husbandry practices are statistically significant. From an income point of view, they do not appear to be as important as the presence of livestock themselves. This is in contract to their connection to dietary diversity.
6. **Other variables of note** – the planting of drought and/or disease resistant crops, increased knowledge, access to seed and fertilizer, and greater diversity of income sources all have significant and positive relationships with increased household income. On the other hand, simply practicing animal husbandry on its own does not. However, the negative value on the coefficient here is more than counteracted by the strong positive relationship apparent on the ownership of specific types of livestock.
7. **Periods of hunger and use of coping strategies** – as expected there is a significant relationship between the use of coping strategies and increased income – and the sign on the coefficients is as expected. One would a higher income would be correlated with a reliance on fewer coping strategies. The fact the coefficient on the SFSS (number of hungry months) is not statistically significant on any of the regressions may be a sign that raising animals to some extent has a positive role to play in insulating households from periods of hunger.