Income Model

James Eapen, Myungha Kim, Ivanna Rodriguez

December 18, 2019

The following document outlines our attempt to mimic the GAC Livelihoods regression on income and animal production activities that was done by the World Renew consultant. This was difficult to do because BIHS dataset did not contain neither change in income variable nor animal husbandry variable which were the two main variables used for GAC Livelihoods regression. Because of this challenge, we tried measure income by counting the number of technological assets a household possessed. We used this as our reponse variable. To measure the effects of animal production, we tried to use livestock asset variables in the BIHS dataset. We know that this regression would not perfectly align with what was measured by the consultant, but it was the closest we could get with our dataset.
The following section shows our data preparation process to construct our response variable.

Data Preparation

```
# selecting variables we need for the regression
bihs_original <- bihs_original%>%
  select(
          fcs,
          survey_year,
          asset_qty_poultry,
          asset_qty_cattle,
          asset_qty_otherlivestock,
          asset_qty_sheepgoat,
          memb_total,
          memb_und15,
          memb_15_44,
          hhs_total,
          bio_bio_1,
          bio_bio_12,
          house_owned,
          asset_cartplough,
          asset_telephone,
          asset_tractor,
          asset_motorbike,
          asset_television,
          asset_radio
        )
```

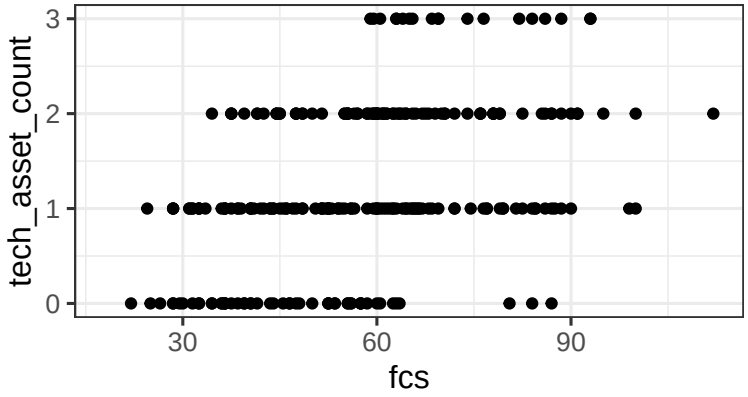
```
# changing yes to 1 and no to 0
bihs_original = bihs_original%>%mutate(asset_cartplough = ifelse(asset_cartplough=="Yes", "1", "0"),
                                       asset_television = ifelse(asset_television=="Yes", 1, 0),
                                       asset_telephone = ifelse(asset_telephone=="Yes", 1, 0),
                                       asset_tractor = ifelse(asset_tractor=="Yes", 1, 0),
                                       asset_radio = ifelse(asset_radio=="Yes", 1, 0),
                                       asset_motorbike = ifelse(asset_motorbike=="Yes", 1, 0),
                                       asset_cartplough = as.numeric(asset_cartplough)
                                       )
```

```r
# create the new variable to use as the response
bihs_original <- bihs_original%>%mutate(tech_asset_count = asset_cartplough + asset_television +
                                          asset_telephone + asset_tractor + asset_radio +
                                          asset_motorbike, tech_asset_count =
                                          as.numeric(tech_asset_count))
```

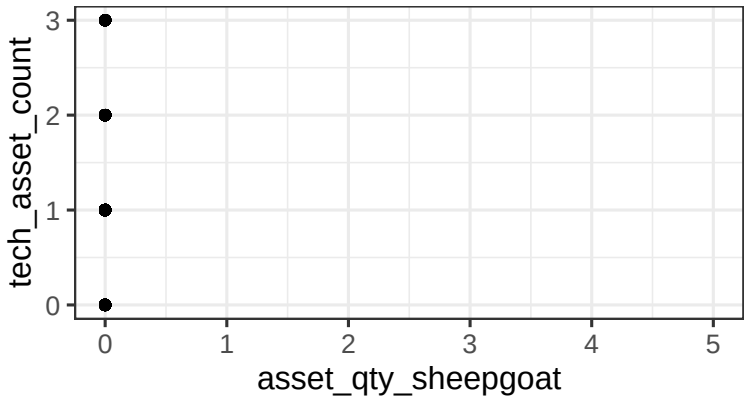
Exploratory plots

We selected a few variables for the BIHS dataset to see if there might be a relationship between the response `tech_asset_count` and the predictors. The boxplots are shown below.

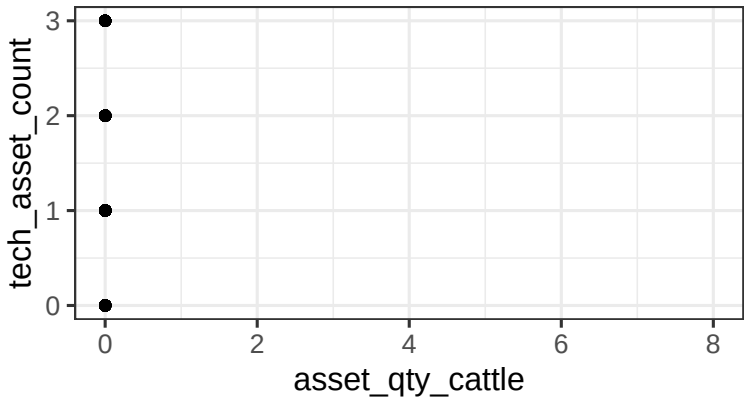
```r
# whether they owned/rented a house
gf_point(data = bihs_original, tech_asset_count ~ fcs )
```



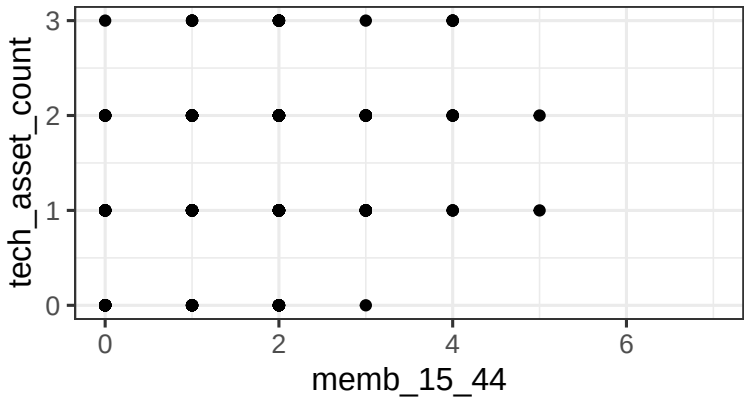

```r
gf_point(data = bihs_original, tech_asset_count ~ asset_qty_sheepgoat)
```



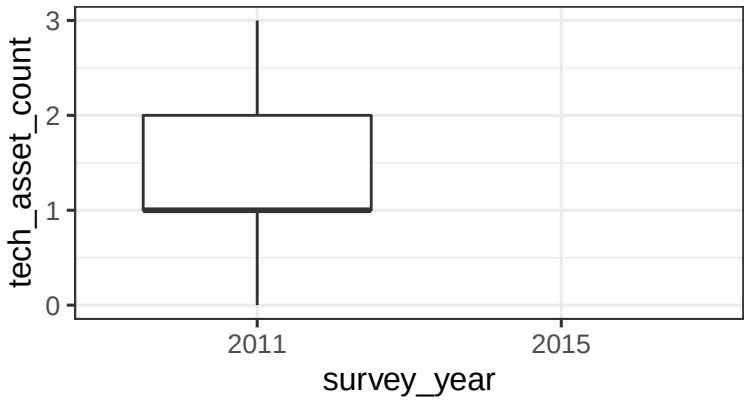

```r
gf_point(data = bihs_original, tech_asset_count ~ asset_qty_cattle)
```




```
gf_point(data = bihs_original, tech_asset_count ~ memb_15_44)
```




```
gf_boxplot(data = bihs_original, tech_asset_count ~ survey_year)
```



As it is shown above, there were too many missing values in our response variable to start with. The plots for each livestock asset do not look vaild. Also, because of the missing value problem, the boxplot does not show any data from 2015, which would make our regression useless.

```
# regression
income_glm <- glm(data = bihs_original, tech_asset_count ~ asset_qty_poultry +
                  asset_qty_cattle + asset_qty_otherlivestock + asset_qty_sheepgoat +
                  memb_total + memb_und15 + memb_15_44 + hhs_total + bio_bio_1 +
                  bio_bio_12 + house_owned, family=poisson(link='log'), na.action = 'na.fail')
```

```
## Error in na.fail.default(structure(list(tech_asset_count = c(3, 1, 1, : missing values in object
```

```r
summary(income_glm)
```

```
## Error in summary(income_glm): object 'income_glm' not found
```

Our regression cannot be run. It cannot be compiled. We will explore other regression methods.