Report 4 (Final) Jensen Miller

Ria Bawiskar

December 2023

1 Introduction

The main contribution of the paper

This paper summarizes the engineered experiment of economists to attempt to prove the real-world evidence of Giffen behavior. To set up their experiment, they define Giffen goods as goods that increase in demand even after increasing in price, a contradiction to the Law of Demand. The posited reason they offered for this, as accepted and disputed among many economists, is that impoverished people in a subsistence zone focus on redistributing their money around to purchase more of the pricier staple good, which is still relatively cheap compared to other goods they may buy.

After defining Giffen behavior, and the theory they set out to prove, they described their experiment to prove the existence of Giffen behavior. To do this, they studied two provinces in China which, among the impoverished, had staple goods of rice and wheat respectively. For a five-month period, they gave randomly selected poor households subsidy coupons for their staple good and studied their purchase behavior for that good and other goods they bought for subsistence, before, during, and after the coupon period.

Even without reading the results portion of this paper, one can tell that the authors found significant evidence to prove Giffen behavior and the existence of Giffen goods within this study, a previously unsuccessful endeavor.

Explanation of Part 1

In part one of this report, I was tasked with replicating Table 1 which presented the means and standard deviations of key variables across the control group, and the groups offered the 0.1, 0.2, and 0.3 yuan/jin subsidy respectively in both the Hunan and Gansu provinces.

Explanation of Part 2

In part two of this report, I used the data to estimate the parameters of the following simple linear regression model: $\%\Delta staple_{i,t} = \beta_0 + \beta_1 \%\Delta p_{i,t} + \epsilon_{i,t}$ for each both the Hunan and Gansu regions where $\%\Delta staple_{i,t}$ is the percent change in household i's consumption of the staple good and $\%\Delta p_{i,t}$ is the percent change in the price of the staple due to the subsidy.

2 Summarizing Statistics

Replication of Table 1

See the code below for my replication of Table 1

```
install.packages('haven')
library(haven)
df <-read_dta("Giffen.dta")
df_hh <- subset(df,(person_id == min_p) & (round == 1))
#Hunan province
#control group
df_A_control<- subset(df_hh,(province == 'Hunan') & (subsidy_group == 0))
fam_size_meanc <- mean(df_A_control$family_size)
fam_size_stdc <- sd(df_A_control$family_size)
fem_head_meanc <- 1 - mean(df_A_control$male)
fem_head_stdc <- sd(df_A_control$male)
inc_per_meanc <- mean(df_A_control$income_per_capita)
inc_per_stdc <- sd(df_A_control$income_per_capita)
exp_per_meanc <- mean(df_A_control$expend_per_capita)
exp_per_stdc <- sd(df_A_control$expend_per_capita)
cal_per_meanc <- mean(df_A_control$hh_cals_percap)
cal_per_stdc <- sd(df_A_control$hh_cals_percap)
rice_per_meanc <- mean(df_A_control$hh_rice_percap)
rice_per_stdc <- sd(df_A_control$hh_rice_percap)
meat_per_meanc <- mean(df_A_control$hh_meat_percap)
meat_per_stdc <- sd(df_A_control$hh_meat_percap)
rice_cal_meanc <- mean(df_A_control$hh_rice_calorie_share)
rice_cal_stdc <- sd(df_A_control$hh_rice_calorie_share)
obsvc <- nrow(df_A_control)
#0.1 yuan subsidy
df_A_one<- subset(df_hh,(province == 'Hunan') & (subsidy_group == 10))
fam_size_meano <- mean(df_A_one$family_size)
fam_size_stdo <- sd(df_A_one$family_size)
fem_head_meano <- 1 - mean(df_A_one$male)
fem_head_stdo <- sd(df_A_one$male)
inc_per_meano <- mean(df_A_one$income_per_capita)
inc_per_stdo <- sd(df_A_one$income_per_capita)
exp_per_meano <- mean(df_A_one$expend_per_capita)
exp_per_stdo <- sd(df_A_one$expend_per_capita)
cal_per_meano <- mean(df_A_one$hh_cals_percap)
cal_per_stdo <- sd(df_A_one$hh_cals_percap)
rice_per_meano <- mean(df_A_one$hh_rice_percap)
rice_per_stdo <- sd(df_A_one$hh_rice_percap)
meat_per_meano <- mean(df_A_one$hh_meat_percap)
```

```
meat_per_stdo <- sd(df_A_one$hh_meat_percap)
rice_cal_meano <- mean(df_A_one$hh_rice_calorie_share)
rice_cal_stdo <- sd(df_A_one$hh_rice_calorie_share)
obsvo <- nrow(df_A_one)
#0.2 yuan subsidy
df_A_two <- subset(df_hh,(province == 'Hunan') & (subsidy_group == 20))
fam_size_meantw <- mean(df_A_two$family_size)
fam_size_stdtw <- sd(df_A_two$family_size)
fem_head_meantw <- 1 - mean(df_A_two$male)
fem_head_stdtw <- sd(df_A_two$male)
inc_per_meantw <- mean(df_A_two$income_per_capita)
inc_per_stdtw <- sd(df_A_two$income_per_capita)
exp_per_meantw <- mean(df_A_two$expend_per_capita)
exp_per_stdtw <- sd(df_A_two$expend_per_capita)
cal_per_meantw <- mean(df_A_two$hh_cals_percap)
cal_per_stdtw <- sd(df_A_two$hh_cals_percap)
rice_per_meantw <- mean(df_A_two$hh_rice_percap)
rice_per_stdtw <- sd(df_A_two$hh_rice_percap)
meat_per_meantw <- mean(df_A_two$hh_meat_percap)
meat_per_stdtw <- sd(df_A_two$hh_meat_percap)
rice_cal_meantw <- mean(df_A_two$hh_rice_calorie_share)
rice_cal_stdtw <- sd(df_A_two$hh_rice_calorie_share)
obsvtw <- nrow(df_A_two)
#0.3 yuan subsidy
df_A_three <- subset(df_hh,(province == 'Hunan') & (subsidy_group == 30))
fam_size_meanth <- mean(df_A_three$family_size)
fam_size_stdth <- sd(df_A_three$family_size)
fem_head_meanth <- 1 - mean(df_A_three$male)
fem_head_stdth <- sd(df_A_three$male)
inc_per_meanth <- mean(df_A_three$income_per_capita)
inc_per_stdth <- sd(df_A_three$income_per_capita)
exp_per_meanth <- mean(df_A_three$expend_per_capita)
exp_per_stdth <- sd(df_A_three$expend_per_capita)
cal_per_meanth <- mean(df_A_three$hh_cals_percap)
cal_per_stdth <- sd(df_A_three$hh_cals_percap)
rice_per_meanth <- mean(df_A_three$hh_rice_percap)
rice_per_stdth <- sd(df_A_three$hh_rice_percap)
meat_per_meanth <- mean(df_A_three$hh_meat_percap)
meat_per_stdth <- sd(df_A_three$hh_meat_percap)
```

```r
rice_cal_meanth <- mean(df_A_three$hh_rice_calorie_share)
rice_cal_stdth <- sd(df_A_three$hh_rice_calorie_share)
obsvth <- nrow(df_A_three)
#Gansu province
#control group
df_A_controlG<- subset(df_hh,(province == 'Gansu') & (subsidy_group == 0))
fam_size_meancG <- mean(df_A_controlG$family_size)
fam_size_stdcG <- sd(df_A_controlG$family_size)
fem_head_meancG <- 1 - mean(df_A_controlG$male)
fem_head_stdcG <- sd(df_A_controlG$male)
inc_per_meancG <- mean(df_A_controlG$income_per_capita)
inc_per_stdcG <- sd(df_A_controlG$income_per_capita)
exp_per_meancG <- mean(df_A_controlG$expend_per_capita)
exp_per_stdcG <- sd(df_A_controlG$expend_per_capita)
cal_per_meancG <- mean(df_A_controlG$hh_cals_percap)
cal_per_stdcG <- sd(df_A_controlG$hh_cals_percap)
rice_per_meancG <- mean(df_A_controlG$hh_all_wheat_and_noodles_percap)
rice_per_stdcG <- sd(df_A_controlG$hh_all_wheat_and_noodles_percap)
meat_per_meancG <- mean(df_A_controlG$hh_meat_percap)
meat_per_stdcG <- sd(df_A_controlG$hh_meat_percap)
rice_cal_meancG <- mean(df_A_controlG$hh_wheat_noodles_calorie_share)
rice_cal_stdcG <- sd(df_A_controlG$hh_wheat_noodles_calorie_share)
obsvcG <- nrow(df_A_controlG)
#0.1 yuan subsidy
df_A_oneG<- subset(df_hh,(province == 'Gansu') & (subsidy_group == 10))
fam_size_meanoG <- mean(df_A_oneG$family_size)
fam_size_stdoG <- sd(df_A_oneG$family_size)
fem_head_meanoG <- 1 - mean(df_A_oneG$male)
fem_head_stdoG <- sd(df_A_oneG$male)
inc_per_meanoG <- mean(df_A_oneG$income_per_capita)
inc_per_stdoG <- sd(df_A_oneG$income_per_capita)
exp_per_meanoG <- mean(df_A_oneG$expend_per_capita)
exp_per_stdoG <- sd(df_A_oneG$expend_per_capita)
cal_per_meanoG <- mean(df_A_oneG$hh_cals_percap)
cal_per_stdoG <- sd(df_A_oneG$hh_cals_percap)
rice_per_meanoG <- mean(df_A_oneG$hh_all_wheat_and_noodles_percap)
rice_per_stdoG <- sd(df_A_oneG$hh_all_wheat_and_noodles_percap)
meat_per_meanoG <- mean(df_A_oneG$hh_meat_percap)
meat_per_stdoG <- sd(df_A_oneG$hh_meat_percap)
```

```r
rice_cal_meanoG <- mean(df_A_oneG$hh_wheat_noodles_calorie_share)
rice_cal_stdoG <- sd(df_A_oneG$hh_wheat_noodles_calorie_share)
obsvoG <- nrow(df_A_oneG)
#0.2 yuan subsidy
df_A_twoG <- subset(df_hh,(province == 'Gansu') & (subsidy_group == 20))
fam_size_meantwG <- mean(df_A_twoG$family_size)
fam_size_stdtwG <- sd(df_A_twoG$family_size)
fem_head_meantwG <- 1 - mean(df_A_twoG$male)
fem_head_stdtwG <- sd(df_A_twoG$male)
inc_per_meantwG <- mean(df_A_twoG$income_per_capita)
inc_per_stdtwG <- sd(df_A_twoG$income_per_capita)
exp_per_meantwG <- mean(df_A_twoG$expend_per_capita)
exp_per_stdtwG <- sd(df_A_twoG$expend_per_capita)
cal_per_meantwG <- mean(df_A_twoG$hh_cals_percap)
cal_per_stdtwG <- sd(df_A_twoG$hh_cals_percap)
rice_per_meantwG <- mean(df_A_twoG$hh_all_wheat_and_noodles_percap)
rice_per_stdtwG <- sd(df_A_twoG$hh_all_wheat_and_noodles_percap)
meat_per_meantwG <- mean(df_A_twoG$hh_meat_percap)
meat_per_stdtwG <- sd(df_A_twoG$hh_meat_percap)
rice_cal_meantwG <- mean(df_A_twoG$hh_wheat_noodles_calorie_share)
rice_cal_stdtwG <- sd(df_A_twoG$hh_wheat_noodles_calorie_share)
obsvtwG <- nrow(df_A_twoG)
#0.3 yuan subsidy
df_A_threeG <- subset(df_hh,(province == 'Gansu') & (subsidy_group == 30))
fam_size_meanthG <- mean(df_A_threeG$family_size)
fam_size_stdthG <- sd(df_A_threeG$family_size)
fem_head_meanthG <- 1 - mean(df_A_threeG$male)
fem_head_stdthG <- sd(df_A_threeG$male)
inc_per_meanthG <- mean(df_A_threeG$income_per_capita)
inc_per_stdthG <- sd(df_A_threeG$income_per_capita)
exp_per_meanthG <- mean(df_A_threeG$expend_per_capita)
exp_per_stdthG <- sd(df_A_threeG$expend_per_capita)
cal_per_meanthG <- mean(df_A_threeG$hh_cals_percap)
cal_per_stdthG <- sd(df_A_threeG$hh_cals_percap)
rice_per_meanthG <- mean(df_A_threeG$hh_all_wheat_and_noodles_percap)
rice_per_stdthG <- sd(df_A_threeG$hh_all_wheat_and_noodles_percap)
meat_per_meanthG <- mean(df_A_threeG$hh_meat_percap)
meat_per_stdthG <- sd(df_A_threeG$hh_meat_percap)
rice_cal_meanthG <- mean(df_A_threeG$hh_wheat_noodles_calorie_share)
```

```r
rice_cal_stdthG <- sd(df_A_threeG$hh_wheat_noodles_calorie_share)
obsvthG <- nrow(df_A_threeG)
#loading values into vectors
rownamess <- c("H: Family Size mean","H: Family Size std","H: Female Head mean",
               "H: Female Head std","H: Income per capita mean",
               "H: Income per capita std","H: Expenditure per capita mean",
               "H: Expenditure per capita std", "H: Calories per capita mean",
               "H: Calories per capita std","H: Rice per capita mean",
               "H: Rice per capita std","H: Meat per Capita mean",
               "H: Meat per Capita std", "H: Rice calorie share mean",
               "H: Rice calorie share std", "H: Observations",
               "G: Family Size mean","G: Family Size std","G: Female Head mean",
               "G: Female Head std","G: Income per capita mean",
               "G: Income per capita std","G: Expenditure per capita mean",
               "G: Expenditure per capita std", "G: Calories per capita mean",
               "G: Calories per capita std","G: Wheat per capita mean",
               "G: Wheat per capita std","G: Meat per Capita mean",
               "G: Meat per Capita std", "G: Wheat calorie share mean",
               "G: Wheat calorie share std", "G: Observations")
#cols
control <- c(fam_size_meanc,fam_size_stdc,fem_head_meanc,fem_head_stdc,
             inc_per_meanc,inc_per_stdc,exp_per_meanc,exp_per_stdc,
             cal_per_meanc,cal_per_stdc,rice_per_meanc,rice_per_stdc,
             meat_per_meanc,meat_per_stdc,rice_cal_meanc,rice_cal_stdc,
             obsvc,fam_size_meancG,fam_size_stdcG,fem_head_meancG,fem_head_stdcG,
             inc_per_meancG,inc_per_stdcG,exp_per_meancG,exp_per_stdcG,
             cal_per_meancG,cal_per_stdcG,rice_per_meancG,rice_per_stdcG,
             meat_per_meancG,meat_per_stdcG,rice_cal_meancG,rice_cal_stdcG,
             obsvcG)
control <- round(control, digits = 4)
one <- c(fam_size_meano,fam_size_stdo,fem_head_meano,fem_head_stdo,
             inc_per_meano,inc_per_stdo,exp_per_meano,exp_per_stdo,
             cal_per_meano,cal_per_stdo,rice_per_meano,rice_per_stdo,
             meat_per_meano,meat_per_stdo,rice_cal_meano,rice_cal_stdo,
             obsvo,fam_size_meanoG,fam_size_stdoG,fem_head_meanoG,fem_head_stdoG,
             inc_per_meanoG,inc_per_stdoG,exp_per_meanoG,exp_per_stdoG,
             cal_per_meanoG,cal_per_stdoG,rice_per_meanoG,rice_per_stdoG,
             meat_per_meanoG,meat_per_stdoG,rice_cal_meanoG,rice_cal_stdoG,
             obsvoG)
```

```r
one <- round(one, digits = 4)
two <- c(fam_size_meantw,fam_size_stdtw,fem_head_meantw,fem_head_stdtw,
         inc_per_meantw,inc_per_stdtw,exp_per_meantw,exp_per_stdtw,
         cal_per_meantw,cal_per_stdtw,rice_per_meantw,rice_per_stdtw,
         meat_per_meantw,meat_per_stdtw,rice_cal_meantw,rice_cal_stdtw,
         obsvtw,fam_size_meantwG,fam_size_stdtwG,fem_head_meantwG,fem_head_stdtwG,
         inc_per_meantwG,inc_per_stdtwG,exp_per_meantwG,exp_per_stdtwG,
         cal_per_meantwG,cal_per_stdtwG,rice_per_meantwG,rice_per_stdtwG,
         meat_per_meantwG,meat_per_stdtwG,rice_cal_meantwG,rice_cal_stdtwG,
         obsvtwG)
two <- round(two, digits = 4)
three <- c(fam_size_meanth,fam_size_stdth,fem_head_meanth,fem_head_stdth,
         inc_per_meanth,inc_per_stdth,exp_per_meanth,exp_per_stdth,
         cal_per_meanth,cal_per_stdth,rice_per_meanth,rice_per_stdth,
         meat_per_meanth,meat_per_stdth,rice_cal_meanth,rice_cal_stdth,
         obsvth,fam_size_meanthG,fam_size_stdthG,fem_head_meanthG,fem_head_stdthG,
         inc_per_meanthG,inc_per_stdthG,exp_per_meanthG,exp_per_stdthG,
         cal_per_meanthG,cal_per_stdthG,rice_per_meanthG,rice_per_stdthG,
         meat_per_meanthG,meat_per_stdthG,rice_cal_meanthG,rice_cal_stdthG,
         obsvthG)
three <- round(three, digits = 4)
rez_df <- data.frame(control, one, two, three)
row.names(rez_df) <- rownamess
```

In short, I created a separate df for just the head of households and created sub-dataframes of it specifying the province, and subsidy amount. I calculate means and standard deviations with the mean() and sd() functions for each value present in the data table. And finally, I put my data into a dataframe, with specific row and column titles, where 'one', 'two', and 'three' represent the 0.1, 0.2, and 0.3 yuan/jin subsidy. I decided to label rows with mean, standard deviation, and G or H for Gansu or Hunan provinces respectively, depending on which province the data represents. I chose to round each value to 4 decimal places, as that's the highest level of specificity recorded in the original table, and I wanted to maintain uniformity. The dataframe looks like this:

	control	one	two	three
H: Family Size mean	2.8323	2.8889	2.9815	2.7107
H: Family Size std	1.3474	1.2805	1.3763	1.1270
H: Female Head mean	0.3354	0.3704	0.3704	0.4088
H: Female Head std	0.4736	0.4844	0.4844	0.4932
H: Income per capita mean	603.6368	556.9779	703.3933	751.4579
H: Income per capita std	1227.4451	796.5283	958.7367	2451.2167
H: Expenditure per capita mean	316.2324	329.5142	299.4996	361.1711
H: Expenditure per capita std	251.8911	316.2363	290.4720	483.2839
H: Calories per capita mean	1767.1914	1783.4783	1817.0374	1851.4563
H: Calories per capita std	628.3830	588.3827	548.6841	600.8500
H: Rice per capita mean	316.9125	324.9798	339.8578	337.8145
H: Rice per capita std	122.0863	128.9773	127.8291	119.6453
H: Meat per Capita mean	50.3825	42.3740	40.7150	52.7636
H: Meat per Capita std	81.6062	61.0201	59.2248	70.2572
H: Rice calorie share mean	0.6387	0.6361	0.6451	0.6415
H: Rice calorie share std	0.1877	0.1861	0.1579	0.1523
H: Observations	161.0000	162.0000	162.0000	159.0000
G: Family Size mean	2.9264	2.6914	2.7346	2.7407
G: Family Size std	1.1031	1.0937	0.9509	1.0606
G: Female Head mean	0.4356	0.4198	0.4506	0.4568
G: Female Head std	0.4974	0.4950	0.4991	0.4997
G: Income per capita mean	693.6210	694.0639	724.3053	726.4058
G: Income per capita std	662.8177	651.8003	800.3756	696.5191
G: Expenditure per capita mean	202.4276	227.7177	198.3299	215.8125
G: Expenditure per capita std	246.6131	213.6109	231.3298	201.0257
G: Calories per capita mean	1736.6886	1732.2584	1716.0521	1654.6273
G: Calories per capita std	496.4061	552.6021	499.6943	519.6920
G: Wheat per capita mean	352.6238	353.4498	340.6775	328.6584
G: Wheat per capita std	132.3133	147.1284	136.3773	119.8066
G: Meat per Capita mean	13.9340	9.6811	13.4558	13.5571
G: Meat per Capita std	30.9121	23.8432	33.6640	31.0554
G: Wheat calorie share mean	0.6905	0.6914	0.6781	0.6804
G: Wheat calorie share std	0.1764	0.1724	0.1810	0.1647
G: Observations	163.0000	162.0000	162.0000	162.0000

Explanation of Findings

One variable that displays significant difference across treatment levels is the Wheat Calorie share in the Gansu province. Mean calorie share was 0.691, 0.691, 0.678, 0.680 for the control group, 0.1, 0.2, and 0.3 subsidies respectively. This displays how, generally, the calorie share of wheat, a staple good in Gansu, decreases as the price decreases. This shows a contradiction in the Law of Demand, and fits all the categories in proving the existence of Giffen goods. This pattern is also reflected in the variable Wheat per capita for the

Gansu province where the the mean goes from 353,353, 341, 329 respective to the control group, 0.1, 0.2, and 0.3 subsidies.

This pattern of displaying Giffen behavior isn't so clear in the variables of the Hunan province, as the staple good's use, as shown in 'Rice per capita' and 'Rice calorie share' actually goes up as the subsidy increases. However, I believe that the breach of pattern can somewhat be attributed to the fluctuating income levels in the Hunan province, given by 604, 557, 703, and 751 respective to control, 0.1, 0.2, and 0.3 yuan subsidy levels. This is because, to exhibit Giffen behavior, people must be in a very specific level of poverty, as described by the paper, and the unpredictability of these income levels makes it harder to attribute staple good consumption to Giffen behavior.

3 Regression

```
install.packages('haven')
library(haven)
df <-read_dta("Giffen.dta")
df_hunan <- subset(df,(province == 'Hunan') & (person_id == min_p))
df_gansu <- subset(df,(province == 'Gansu') & (person_id == min_p))

#hunan indep var: percent change in rice price per household
perc_change_rice_price <- as.vector(df_hunan$pct_ch_sub_rice_arc)
perc_change_rice_price <- na.omit(perc_change_rice_price)

#hunan dep var: percent change in rice consumption per household
perc_change_rice_cons <- as.vector(df_hunan$pct_ch_hh_rice)
perc_change_rice_cons <- na.omit(perc_change_rice_cons)

#gansu indep var: percent change in wheat price per household
perc_change_wheat_price <- as.vector(df_gansu$pct_ch_sub_wheat_arc)
perc_change_wheat_price <- na.omit(perc_change_wheat_price)

#gansu dep var: percent change in wheat consumption per household
perc_change_wheat_cons <- as.vector(df_gansu$pct_ch_hh_wheat)
perc_change_wheat_cons <- na.omit(perc_change_wheat_cons)

#pad shortest vectors with NA's to have same length as longest vectors
length(perc_change_rice_price) <- length(perc_change_wheat_price)
length(perc_change_rice_cons) <- length(perc_change_wheat_cons)

calculation_df <- data.frame(perc_change_rice_price,perc_change_rice_cons,
                             perc_change_wheat_price, perc_change_wheat_cons)
colnames(calculation_df) <- c('p_change_rice_price','p_change_rice_consumption',
                              'p_change_wheat_price','p_change_wheat_consumption')

hunan_slr <- lm(p_change_rice_consumption ~ p_change_rice_price, data=calculation_df)
gansu_slr <- lm(p_change_wheat_consumption ~ p_change_wheat_price, data=calculation_df)
```

```
> summary(hunan_slr)

Call:
lm(formula = p_change_rice_consumption ~ p_change_rice_price,
    data = calculation_df)

Residuals:
     Min       1Q   Median       3Q      Max
-202.123  -22.817    0.459   24.459  203.300

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.45935    1.36794  -0.336    0.737
p_change_rice_price -0.09469    0.08085  -1.171    0.242

Residual standard error: 48.52 on 1256 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.001091,  Adjusted R-squared:  0.0002957
F-statistic: 1.372 on 1 and 1256 DF,  p-value: 0.2417
```

```
> summary(gansu_slr)

Call:
lm(formula = p_change_wheat_consumption ~ p_change_wheat_price,
    data = calculation_df)

Residuals:
    Min      1Q  Median      3Q     Max
-185.74  -49.66   15.76   49.74  244.33

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -28.6038     2.7688 -10.331   <2e-16 ***
p_change_wheat_price   0.4089     0.1339   3.054   0.0023 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 98.62 on 1267 degrees of freedom
Multiple R-squared:  0.007309,  Adjusted R-squared:  0.006526
F-statistic: 9.329 on 1 and 1267 DF,  p-value: 0.002303
```

Above is my code which helps me find the parameters of simple linear regression models for both the Hunan and Gansu provinces, for observations representing a single household. Describing the code, I first create subset datasets for both the Hunan and Gansu provinces and use those to extract vectors of the variables I will use in my regression later on, also subsetting to one household. I then collated them into a dataframe and used the columns to calculate simple linear regression for both provinces.

Printed out is the summary for both regressions and from that, we can obtain the following:

- Hunan: $\%\delta\text{rice} = -0.4594 - 0.0947\beta_1$

- Gansu: $\%\delta\text{wheat} = -28.6038 + 0.4089\beta_1$

The Hunan β_0 coefficient is significant at the 1 p-value level, and the Hunan β_1 coefficient is significant at the 1 p-value level. The Gansu β_0 coefficient is significant at the 0 p-value level, and the Gansu β_1 coefficient is significant at the 0.001 p-value level. The R-squared value for the Hunan province is 0.0011 and the R-squared value for the Gansu province is 0.0073 which indicates that only a small percentage of the variability in the percentage of staple good price can be explained by the linear relationship between the percentage change of the staple's price and the percentage change of the staple's household consumption. This value indicates that the equations are a bad fit for both provinces.

Going back to what the equations mean, the Hunan simple linear regression indicates a weak negative relationship between the percent change in household rice consumption and the percent change in rice price, due to the β_1 being negative, but still relatively small. The Gansu simple linear regression indicates a weak positive relationship between the percent change in household wheat consumption and the percent change in wheat price, due to the β_1 being positive, but still relatively small

The report indicates that for ”all regressions, we present standard errors clustered at the household level” so there is an expected error term, with a mean $\neq$ zero for both regression equations.

4 Conclusion

Altogether, by recreating Table 1, and using the data to develop linear regression models for each province, I learned to understand some of the steps taken by the authors of this study in proving the existence of Giffen goods.

Through this process, I truly understood how impoverished people in a subsistence zone focus on redistributing their money around to purchase more of the pricier staple good, which is still relatively cheap compared to other goods they may buy.

Through the use of RStudio, I was able to become more familiar with large-scale manipulations of data to find important indicators like mean and standard deviation and find meaning within them. I was also able to create simple linear regressions, and use their summaries, including coefficients and R-squared values to measure how weak/strong, and positively/negatively correlated variables like percent change in price of a staple good had to consumption. And through this step specifically, I was able to see a small correlation effect, especially in the Gansu province, which followed Giffen behavior.

Overall, by following and replicating parts of this study, I've become more familiar with steps of economic research and the important implications its results can have on policy which deeply affects many.