Summary Statistics

We have presented the summary statistics of our study in West Bengal including all the key variables relevant for our analysis.

URBAN2011	variable	N	mean	SD	min	max
rural 0	SURVEY	1290	2	0	2	2
	COTOTAL	1290	61250.3	40369.24	6728	515862
	COPC	1290	14773.83	8970.961	3824	124965
	INCOME	1290	73044.61	323073.9	-204532	1.12E+07
	INCOMEPC	1290	18008.75	81978.98	-22725.8	2792455
	NPERSONS	1290	4.400775	1.9357	1	19
	NADULTM_n	1290	1.091473	0.724826	0	5
	NADULTF_n	1290	1.118605	0.604389	0	6
	NCHILDM	1290	0.648062	0.816918	0	4
	NCHILDF	1290	0.627907	0.839943	0	4
	NTEENF	1290	0.273643	0.524352	0	4
	NELDERM	1290	0.192248	0.396183	0	2
	NELDERF	1290	0.174419	0.381654	0	2
	NTEENM	1290	0.273643	0.528772	0	3
	CO1X	1290	886.4801	539.7312	0	6300
	CO2X	1290	80.18298	102.8287	0	800
	CO3X	1290	60.21605	75.7352	0	1800
	CO5X	1290	1.249612	15.678	0	360
	CO6X	1290	102.7403	80.47303	0	1200
	CO7X	1290	375.3136	321.6716	0	5000
	CO8X	1290	40.39341	64.70487	0	900
	CO9X	1290	237.1994	124.5735	0	1380
	CO10X	1290	64.74442	46.8328	0	400
	CO11X	1290	115.0467	183.1568	0	1500
	CO12X	1102	24.66152	75.4844	0	1000
	CO13X	1269	88.974	97.58786	0	1200
	CO14X	1289	450.6796	254.7989	1	2800
	CO15	1290	135.0062	108.8829	0	1000
	CO16	1290	44.41318	48.83473	0	450
	CO17	1290	46.09612	48.67635	0	500
	CO19	1290	45.20465	71.17043	0	1000
	CO20	1290	40.62791	113.5026	0	1500
	CO21	1290	64.14884	141.2105	0	1000
urban 1	SURVEY	1145	2	0	2	2
	COTOTAL	1145	117583.9	106952.5	4020	1408460
	COPC	1145	30054.73	28454.98	4020	395829.1
	INCOME	1145	164830	178854.3	-35000	1670000
	INCOMEPC	1145	42510.94	46818.55	-7000	417500
	NPERSONS	1145	4.274236	1.835901	1	17

	NADULTM_n	1145	1.282096	0.834088	0	5
	NADULTF_n	1145	1.21048	0.649283	0	5
	NCHILDM	1145	0.448908	0.672422	0	4
	NCHILDF	1145	0.406114	0.688414	0	5
	NTEENF	1145	0.20262	0.458974	0	3
	NELDERM	1145	0.241921	0.444457	0	3
	NELDERF	1145	0.267249	0.446648	0	2
	NTEENM	1145	0.213974	0.479088	0	2
	CO1X	1145	762.5053	463.5749	0	3900
	CO2X	1145	150.7517	141.8578	0	1200
	CO3X	1145	84.36279	81.33858	0	1750
	CO5X	1145	2.803493	19.21462	0	400
	CO6X	1145	140.0423	102.7769	0	900
	CO7X	1145	677.8009	525.9877	0	7000
	CO8X	1145	105.0716	159.9932	0	3000
	CO9X	1145	287.8384	125.1921	0	1350
	CO10X	1145	88.3821	62.67773	0	880
	CO11X	1145	205.6932	236.2297	0	2500
	CO12X	1063	96.48071	138.0603	0	2000
	CO13X	1139	123.2511	94.06221	0	1000
	CO14X	1145	629.3319	398.9068	0	5000
	CO15	1145	186.5555	138.0182	0	1850
	CO16	1145	79.59563	63.67016	0	500
	CO17	1145	86.98603	80.54831	0	1200
	CO19	1145	141.1886	179.6949	0	2800
	CO20	1145	139.4603	235.8639	0	2500
	CO21	1145	129.5092	196.8773	0	2000
Total	SURVEY	2435	2	0	2	2
	COTOTAL	2435	87739.81	83846.71	4020	1.41E+06
	COPC	2435	21959.3	21940.19	3824	395829.1
	INCOME	2435	116204.5	269092	-204532	1.12E+07
	INCOMEPC	2435	29530.31	68839.98	-22725.8	2792455
	NPERSONS	2435	4.341273	1.890099	1	19
	NADULTM_n	2435	1.181109	0.783753	0	5
	NADULTF_n	2435	1.161807	0.627449	0	6
	NCHILDM	2435	0.554415	0.758827	0	4
	NCHILDF	2435	0.523614	0.780144	0	5
	NTEENF	2435	0.240246	0.495856	0	4
	NELDERM	2435	0.215606	0.42022	0	3
	NELDERF	2435	0.21807	0.415993	0	2
	NTEENM	2435	0.245585	0.50679	0	3
	CO1X	2435	828.1839	509.0266	0	6300
	CO2X	2435	113.3662	127.6675	0	1200
	CO3X	2435	71.57047	79.32495	0	1800
	CO5X	2435	1.980287	17.44421	0	400

	CO6X	2435	120.2806	93.49343	0	1200
	CO7X	2435	517.5509	455.6687	0	7000
	CO8X	2435	70.80678	123.6565	0	3000
	CO9X	2435	261.0112	127.3729	0	1380
	CO10X	2435	75.85947	56.09999	0	880
	CO11X	2435	157.671	214.5737	0	2500
	CO12X	2165	59.92425	116.3736	0	2000
	CO13X	2408	105.1873	97.43188	0	1200
	CO14X	2434	534.721	342.2648	0	5000
	CO15	2435	159.246	126.071	0	1850
	CO16	2435	60.95688	58.96442	0	500
	CO17	2435	65.32361	68.70877	0	1200
	CO19	2435	90.33881	141.9688	0	2800
	CO20	2435	87.10144	188.16	0	2500
	CO21	2435	94.88296	172.7499	0	2000

If we look at the data carefully, we observe that there are a few missing values in only three of our consumption items – 270 missing values in CO12X which denotes consumption of milk products (highlighted in yellow),27 missing values in CO13X which denotes cereal items and only 1 missing value in CO14X which denotes consumption of vegetables. For better econometric analysis, we have replaced them with 0 as both these variables are not qualitative.

As we have no missing values for the COTOTAL and income items, we do not need to bother about them in our analysis.

Another problem in our data is the negative income values (highlighted in red) which in unrealistic and also some outliers. So, we have to address these two issues before starting our analysis.

Let us have a look at the histogram and kernel density of income.
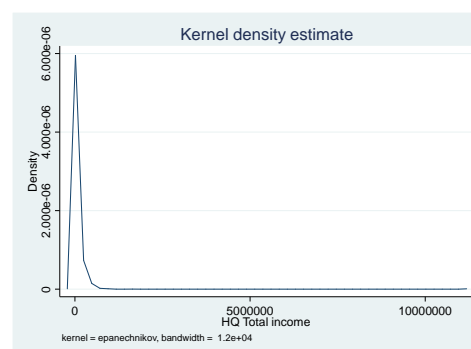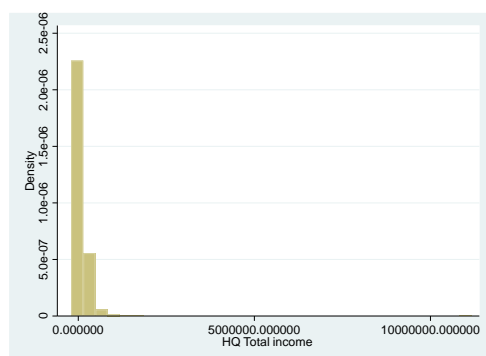
Both of the graphs above confirm that income have negative and zero values in our data.

Running the summarise command, we find that there total 17 such observations in our data.

Now, we take a log transformation in both monthly per capita consumption expenditure and annual income per capita and again check the density functions. (see next page)
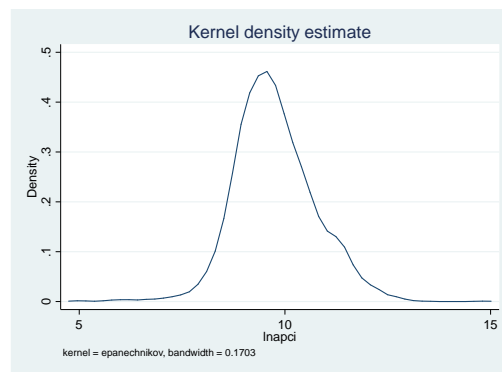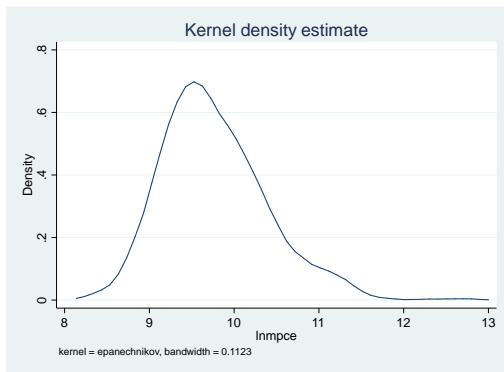
Now, we see that the transformations make our data better as both non-positive values and outliers are solved.

Now, in our analysis we are regressing food share in expenditure items on log (annual per capita income). But we could have also taken log (monthly per capita consumption expenditure).

So, we try to find the correlation between them to see if our results would vary significantly or not.
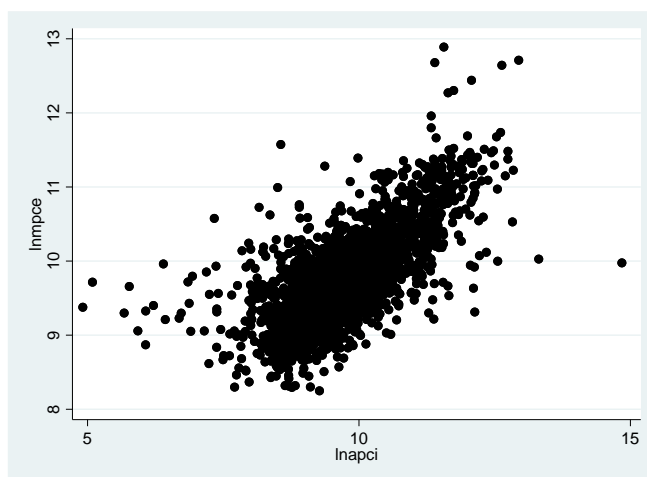
Our scatter plot tells us that both are strongly positively correlated but not perfectly correlated. Hence, we can proceed with our annual per capita income.

Now, we generate food share in expenditure variable and try to see how it behaves with percentage increase in annual income per capita.
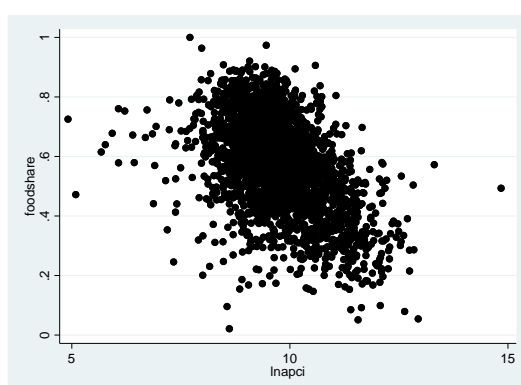
Now, we again check the summary statistics with sample size N- uniform across all categories as we have dropped the missing values (i.e lnmpce~=. & lnapci~=.)

stats	SURVEY	COTOTAL	COPC	lnmpce	INCOME	INCOMEP(	lnapci	totalf~d	foodsh~e	CO1X	CO2X	CO3X	CO5X	CO6X
N	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418
mean	2	87855.44	21976.66	9.767752	117161.5	29763.67	9.754994	42581.32	0.571464	828.7849	113.7449	71.69897	1.99421	120.4367
sd	0	83925.88	21996.55	0.621571	269764.5	69023.53	0.977328	22085.22	0.156472	508.7739	127.8913	79.52465	17.50465	93.6571
min	2	4020	3824	8.249052	500	136	4.912655	153.6	0.021199	0	0	0	0	0
max	2	1408460	395829.1	12.88874	1.12E+07	2792455	14.84243	239520	1	6300	1200	1800	400	1200

stats	CO7X	CO8X	CO9X	CO10X	CO11X	CO12X	CO13X	CO14X	CO15	CO16	CO17	CO18	CO19	CO20
N	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418
mean	518.8985	71.08333	261.2823	75.96187	158.1087	53.53019	103.7655	535.3313	159.531	61.05045	65.42928	108.0476	90.80852	87.04797
sd	456.3745	124.0004	127.441	56.10795	214.9831	111.6455	95.0071	342.0639	126.1967	59.07244	68.66619	188.994	142.3045	188.0787
min	0	0	0	0	0	0	0	0	0	0	0	0	0	0
max	7000	3000	1380	880	2500	2000	1200	5000	1850	500	1200	5000	2800	2500

stats	NPERSON!	lnhhsz	NADU~M_	NADU~F_r	NCHILDM	NCHILDF	NTEENF	NELDERM	NELDERF	NTEENM				
N	2418	2418	2418	2418	2418	2418	2418	2418	2418	2418				
mean	4.346567	1.375246	1.183623	1.162945	0.556245	0.523573	0.241108	0.215881	0.217535	0.24483				
sd	1.888227	0.448613	0.784751	0.628189	0.759671	0.778049	0.496756	0.420468	0.415652	0.505272				
min	1	0	0	0	0	0	0	0	0	0				
max	19	2.944439	5	6	4	5	4	3	2	3				

Finally, we can start our econometric assignment.