ECO 251: Take-Home R Exam

Roll: 231270012.
Name: Debargha Som
M.Sc Economics
No. of pages: 5

Answers

i) The variables 'ID_P' and 'SURVEY' are the unique identifiers (variables) in my dataset. Since the other variables have information about the individuals like standard deviation from mean years of education, etc, those cannot be used as a unique identifier. From 'var_label' function we know 'ID_P' - is the "Unique cross-wave person ID" and 'SURVEY' - is the "Wave of Survey". Thus, these two variables act as identifiers. Together they are a unique identifier.

I confirmed it by using the “get_dupes(mydata, c(ID_P,SURVEY))” command in R.

---

ii) Summary Statistics Table (Table 1):

Summary Statistics Table			
widow	widow_support	hh_assets	std_edu
Min. :0.00000	Min. :1.000	Min. :1.000	Min. :-3.171
1st Qu.:0.00000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.: 1.306
Median :0.00000	Median :2.000	Median :3.000	Median : 4.225
Mean :0.03621	Mean :2.255	Mean :3.062	Mean : 3.979
3rd Qu.:0.00000	3rd Qu.:3.000	3rd Qu.:4.000	3rd Qu.: 6.306
Max. :1.00000	Max. :4.000	Max. :5.000	Max. :12.568

Summary Statistics Table (Continued)			
hh_children	ln_cons	girl	cons_capita
Min. :1.000	Min. : 6.532	Min. :0.0000	Min. : 58.86
1st Qu.:3.000	1st Qu.:10.954	1st Qu.:0.0000	1st Qu.: 816.98
Median :3.000	Median :11.341	Median :0.0000	Median : 3806.45
Mean :3.659	Mean :11.383	Mean :0.4662	Mean : 8584.95
3rd Qu.:4.000	3rd Qu.:11.778	3rd Qu.:1.0000	3rd Qu.: 12472.10
Max. :9.000	Max. :14.703	Max. :1.0000	Max. :240360.53

In the above Summary Statistics table, the columns represent “key variables in our analysis”. We see that the mean of the ‘widow’ is 0.03621, i.e., 3.621% of our sample observations are widows.

Yes, 3.621% is a reasonable rate of widowhood for our given sample. This is because a 2015 WEF (Which country has the largest number of widows? | World Economic Forum (weforum.org)) report suggested that the number of widows in India was estimated to be 46 million. Also, the estimated population of India in 2015 was 1.32 billion (indian population 2015 - Google Search). Thus, roughly we get 3.48% of the population was widows, which is very close to our sample observations.

---

iii) To answer – "Is it true the households headed by widowed mothers are poorer (in the sense that they earn less)?" We use the information in Table 2, given below. We use consumption per capita as a proxy for income. Specification of models,

Model (1): $cons\_capita = \beta_0 + \beta_1 widow + u$

Model (2): $log(cons\_capita) = \beta_0 + \beta_1 widow + u$

Regression Table (Table 2)

	Dependent variable:	
	cons_capita	log(cons_capita)
	(1)	(2)
widow	-2,903.459***	-0.486***
	(345.875)	(0.042)
Constant	8,702.926***	8.134***
	(65.803)	(0.008)
Observations	36,137	36,137
R2	0.002	0.004
Adjusted R2	0.002	0.004
Residual Std. Error (df = 36135)	12,280.530	1.492
F Statistic (df = 1; 36135)	70.468***	133.974***
Note:	*p<0.1; **p<0.05; ***p<0.01	

First, we regressed consumption per capita (cons_capita) on the 'widow' dummy. However, the overall model (inferring from F statistic) and all the parameters were statistically significant at 1% level of significance, the standard errors were too high. There can be multiple reasons for that (like nonlinearity, heteroskedasticity, multicollinearity, etc.), but taking the natural log of consumption per capita (log(cons_capita)) reduces them.

Thus, we use model (2) in Table 2, i.e., log(cons_capita) regressed on the 'widow' dummy. Now, the overall model and all the parameters are statistically significant at 1% level of significance. Thus, the slope coefficient of widow in model (2), is interpreted as - for our given sample, widows earn an estimated 48.6% less than non-widows on average.

---

iv) In studying the association between widowhood and educational attainments, we use standardized years of education rather than completed years of education because:

- It accounts for differences in educational systems and cohorts. As different educational systems may have varying durations and structures. Also, educational attainment may vary across different cohorts due to changes in educational policies, access to education, etc. Thus, Standardizing years of education allows for comparisons across individuals in different cohorts and systems.

- Handling Educational Inequality: Completed years of education may not capture the quality or intensity of education received. Standardizing education levels tells how far an observation is from the mean. This provides a more nuanced measure of variations in educational quality.

---

v) To find the association between widowhood and children's educational attainment I will regress std_edu on widow (dummy), household children, and household assets, and will obtain the following results (in Table 3). Specification of models,

Model (1): $std\_edu = \beta_0 + \beta_1\, widow + u$

Model (2): $std\_edu = \beta_0 + \beta_1\, widow + \beta_2\, hh\_children + \beta_3\, hh\_assets + u$

Regression Table (Table 3)

	Dependent variable:	
	std_edu	
	(1)	(2)
widow	0.682***	0.861***
	(0.097)	(0.091)
hh_children		-0.243***
		(0.011)
hh_assets		0.763***
		(0.013)
Constant	3.951***	2.498***
	(0.018)	(0.068)
Observations	36,130	36,125
R2	0.001	0.120
Adjusted R2	0.001	0.120
Residual Std. Error	3.448 (df = 36128)	3.237 (df = 36121)
F Statistic	49.378*** (df = 1;36128)	1,644.251*** (df = 3;36121)
Note:		*p<0.1; **p<0.05; ***p<0.01

For the given sample, in model (2), we can see widowhood positively affects the standardized years of a child's education. That is, **controlling other factors fixed like the level of household assets (hh_assets) and the number of children (hh_children),** a child is estimated to attain 0.861 units (standard deviation from mean) more education, than a child whose mother is not a widow, on average. The coefficient is statistically highly significant as the p-value is less than 1%.

Eg: Say, when **hh_children=1** (as the min. of this variable is 1, see Table 1, Summary Statistics) and **hh_asset=1** (again the min. of this variable is 1, see Table 1) then a widowed child's standardized years of education is supposed to be 0.861 units more than a non-widowed child, on average.

E[std_edu | hh_asset, hh_children, **widow=1**] - E[std_edu | hh_asset, hh_children, **widow=0**] = 0.861 = $\hat{\beta}_1$ (Note: we are considering model 2)

The coefficient for the number of household children suggests that for each additional household child, the child's standardized years of education are estimated to decrease by 0.243 units, on average, holding other variables constant. This suggests that larger household sizes may be associated with lower educational attainment for children.

The coefficient for the household assets suggests that for each unit increase in household assets, the child's standardized years of education are estimated to increase by 0.763 units, on average, holding other variables constant. This suggests that higher household assets may be associated with higher educational attainment for children.

The intercept is interpreted as – the estimated child's standardized years of education, when all independent variables are zero (i.e., when there is no widow in the household, no household children, and no household assets), holding other variables constant. However, this makes no sense, if there are no children in the household, then whose standardized years of education we are talking about.

---

vi) To analyse if widowhood appears to affect girls differently than it affects boys, we will regress std_edu on widow (dummy), girl (dummy) and widow*girl (interaction dummy). Specification of the model,

Model (1): $std\_edu = \beta_0 + \beta_1\, widow + \beta_2\, girl + \beta_3\, widow{*}girl + u$

```
Regression Table (Table 4)
================================================
                        Dependent variable:
                    ---------------------------
                               std_edu
------------------------------------------------
widow                         0.504***
                               (0.129)

girl                          -0.178***
                               (0.037)

widow:girl                     0.398**
                               (0.196)

Constant                      4.034***
                               (0.025)

------------------------------------------------
Observations                   36,130
R2                              0.002
Adjusted R2                     0.002
Residual Std. Error      3.447 (df = 36126)
F Statistic          24.597*** (df = 3; 36126)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

The coefficient of widow is interpreted as, for children in households where the mother is widow, a child's standardized years of education are estimated to increase by 0.504 units, on average, holding other variables constant. The coefficient is statistically significant at 1% level of significance.

The coefficient of girl is interpreted as, for girls (compared to boys), the standardized years of education are estimated to decrease by 0.178 units, on average, holding other variables constant. The coefficient is statistically significant at 1% level of significance.

The coefficient of the interaction term is interpreted as, the additional effect on a girl child's standardized years of education when the mother is widow. Specifically, it suggests that the effect of widowhood on standardized years of education is 0.398 units higher for girls than boys, holding other variables constant. However, the coefficient is statistically significant at 5% level of significance.

Thus to answer the question "Does widowhood appear to affect girls differently than it affects boys?" We do the following calculations,

- E[std_edu | widow=1, girl=0] = 4.034 + 0.504
  (average standardized years of education of boys whose mothers are widowed)
- E[std_edu | widow=1, girl=1] = 4.034 + 0.504 – 0.178 + 0.398
  (average standardized years of education of boys whose mothers are widowed)

Therefore,
E[std_edu | widow=1, **girl=1**] - E[std_edu | widow=1, **girl=0**]
= – 0.178 + 0.398 = 0.22 = $\hat{\beta}_2 + \hat{\beta}_3$

Thus, we can conclude that, on average, a girl child's standardized years of education are more than that of a boy child's standardized years of education, for widowed women.