Assignment 4 - Applied Econometrics - ECON6645

2022-03-17

Justin Desrosier

Question 1

Table 1: Summary Statistics

Marital Status	Substance Use	Respondent Count	Median Income
Married	No	28168	90089
	Yes	1782	89770
	NA	374	78637
Common-law	No	7356	86409
	Yes	1400	76947
	NA	63	69379
Widowed/Divorced/Separated	No	8232	49192
	Yes	1231	43321
	NA	79	37040
Single	No	11212	49758
	Yes	3070	43645
	NA	331	44975
Did not respond (NA)	No	172	61763
	Yes	16	48953
	NA	6	56677

Table 2: Summary Statistics

Highest Education	Substance Use	Respondent Count	Median Income
Less than secondary school	No	5313	46155
	Yes	1106	35711
	NA	180	38300
Secondary school	No	11667	68987
	Yes	1768	53597
	NA	219	60119
Post-secondary diploma or university degree	No	37606	84552
	Yes	4549	69819
	NA	435	73175
Did not respond	No	554	76807
	Yes	76	69492
	NA	19	85759

Table 3: Summary Statistics

Visible Minority	Immigrant	Substance Use	Respondent Count	Median Income
No	No	No	40004	81965
		Yes	5669	63186
		NA	577	64670
	Yes	No	3320	81449
		Yes	305	72235
		NA	62	70016
	NA	No	341	69512
		Yes	39	56625
		NA	4	58609
Yes	No	No	1098	84489
		Yes	196	69102
		NA	24	65816
	Yes	No	5732	67626
		Yes	206	52942
		NA	83	56988
	NA	No	292	72326
		Yes	18	48722
		NA	5	86647
NA	No	No	3088	64462
		Yes	940	43506
		NA	60	61177
	Yes	No	130	71463
		Yes	8	19017
		NA	5	42347
	NA	No	1135	37773
		Yes	118	35027
		NA	33	37334

Table 4: Summary Statistics

				(Age Quantile %)		
Female	Substance Use	Respondent Count	Median Income	25th	50th	75th
No	No	24209	82611	37	48	57
	Yes	4741	64747	32	40	52
	NA	535	61960	35	49	58
Yes	No	30931	75092	36	47	56
	Yes	2758	53030	30	38	50
	NA	318	64766	34	50	58

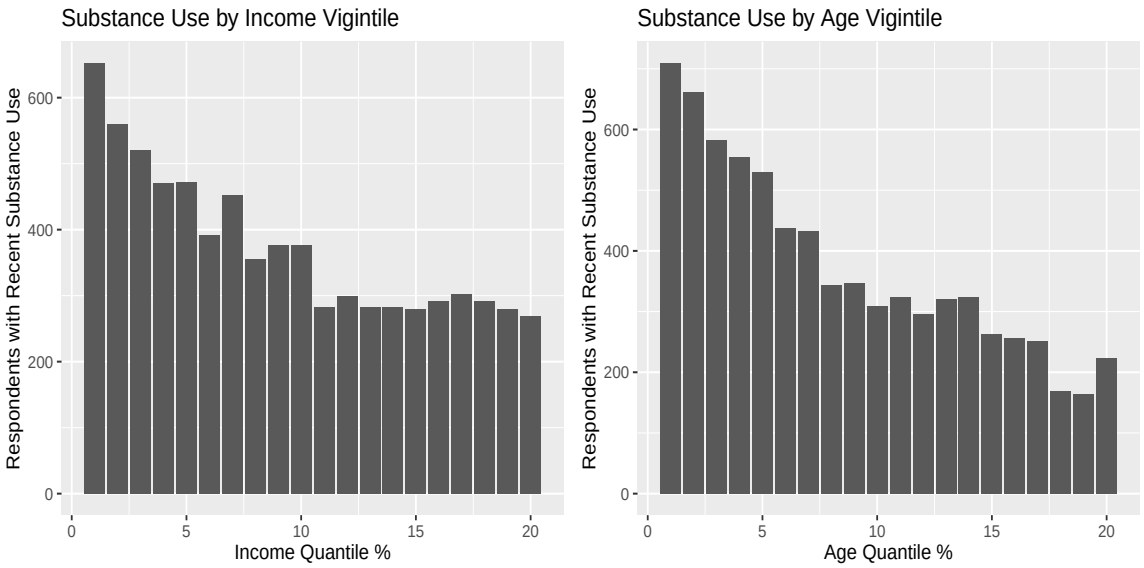

Figure 1: This is a figure.

Tables 1-4 display the summary statistics for our primary variable of interest, illict drug use in the past 12 months. The dependent variable *Substance Use* is grouped against the explanatory variables income, education, marital status, immigrant and visible minority status, sex, and age. For these control variables, there will be a log trasnformation applied to income, as it is a continous variable that tends to increse exponentially. Also, as can be ascertained from Figure 1, there is a diminshing decrease in the number of respondents who have recently used illicit substances as income increases, for this reason it is also necessary to use a quadratic (squared) income control. The same does not apply to age in regard to either concern so age will be left as a continuous variable.

For the remaining control variables, they are all discrete, some are categorical while others are binary. It will be necessary to select reference categories for the non-binary variables, and leave unchanged the binary immigrant, visible minority, and sex variables. For the categorical education variable, post-secondary diploma or greater will be the reference category because a near four-fold majority of respondents are in this category. For the marital status variable, the textitmarried category contains a majority of the respondents except to a lesser extent, for this variable texitsingle will be the reference category as it the best representation of uninhibited human experience.

Table 5 (2) displays the output from a probabilistic regression of the likelihood that a responded has used an illicit drug in last 12 months on our control variables outlined above. Since the result coefficients of this model are not modified to display average marginal effects, the coefficients can be interpreted as the change in the z-score or probit index for a one unit change in the predictor [1]. The categorical factor variables, $HighestEducation$ and $MaritalStatus$ have a slightly different interpretation. For example, having attained less than a secondary school education, versus having attained at least a post-secondary diploma (the reference group), increases teh z-score by 0.028.

[1] https://stats.oarc.ucla.edu/r/dae/probit-regression/

Table 5: Q1 Probabilistic Regression

	Dependent variable:	
	illicit	
	OLS/LPM (1)	*probit (2)*
$log(Income)$	-0.037^{***}	-0.186^{***}
	(0.003)	(0.017)
$Income^2$	0.000^{***}	0.000^{***}
	(0.000)	(0.000)
Highest Education		
$LessthanSecondarySchool$	0.028^{***}	0.124^{***}
	(0.005)	(0.025)
$SecondarySchool$	0.015^{***}	0.072^{***}
	(0.003)	(0.018)
Marital Status		
$Married$	-0.095^{***}	-0.498^{***}
	(0.004)	(0.020)
$Commonlaw$	-0.038^{***}	-0.133^{***}
	(0.005)	(0.023)
$Widowed/Divorced/Separated$	-0.027^{***}	-0.039^{*}
	(0.005)	(0.024)
$Immigrant$	-0.037^{***}	-0.268^{***}
	(0.005)	(0.030)
$VisibleMinority$	-0.051^{***}	-0.346^{***}
	(0.005)	(0.034)
$Female$	-0.078^{***}	-0.448^{***}
	(0.003)	(0.015)
Age	-0.004^{***}	-0.021^{***}
	(0.0001)	(0.001)
Constant	0.698^{***}	1.732^{***}
	(0.035)	(0.189)
Observations	55,822	55,822
R^2	0.073	
Adjusted R^2	0.073	
Log Likelihood		-17,588.420
Akaike Inf. Crit.		35,200.850
Residual Std. Error	0.305 (df = 55810)	
F Statistic	398.054^{***} (df = 11; 55810)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Question 2

Since our dependent variable, perceived mental health, is a multi-categorical ordinal variable, but it is not cardinal (i.e. moving between categories is not necessarily an equivalent change), we must estimate this variable's coefficients using maximum likelihood estimation.

It will also be necessary to scale (standardize) the income and age variables for this regression since they are measured in different orders of magnitude, scaling will make the coefficient results more interpretable. The categorical variables will have the same reference categories as outlined in the previous regression specification.

Table 6(2) shows the coefficients from the ordered probit regression of perceived mental health on the control variables. From the coefficients, mental health is better (from poor to fair, good, very good, excellent) with higher income although there are diminshing returns to income as can be see by the $Income^2$ coefficient. Better mental health is also associated with lower education levels, as can be seen in the categorical education coefficients with respect to the reference category, post-secondary diploma or greater level of educational attainment.

Table 6(2) also indciates that being single is associated with better mental health versus the other marital statuses. Although marriage is assocaited with at least three times less of a decline in mental health versus being single when compared to the other categories, common-law and widowed/divorced/separate. better mental health is assocaited with being an immigrant or a visible minority and not being female. Age is statistically insignificant but notably appears to have no effect.

The average marginal effects of the ordered probit regression from table 6 (2) can be interpreted in the $Mean$ column as a percentage point rise (fall) in level of mental health (coeff. X 100) following a 1 standard deviation increase in the *Statistic*. For instance, a 1 standard deviation increase in income is assocaited with a decrease of 1.2 percentage points on the perceived mental health scale.

Table 6: Ordered Probabilistic Regression

	Dependent variable:	
	Perceived Mental Health	
	OLS/LPM (2)	*ordered probit (1)*
$log(Income)$	0.209^{***}	0.219^{***}
	(0.007)	(0.008)
$Income^2$	-0.078^{***}	-0.079^{***}
	(0.006)	(0.007)
Highest Education		
$LessthanSecondarySchool$	0.125^{***}	0.136^{***}
	(0.016)	(0.018)
$SecondarySchool$	0.187^{***}	0.205^{***}
	(0.014)	(0.016)
Marital Status		
$Married$	-0.042^{***}	-0.052^{***}
	(0.012)	(0.014)
$Commonlaw$	-0.187^{***}	-0.209^{***}
	(0.012)	(0.014)
$Widowed/Divorced/Separated$	-0.142^{***}	-0.162^{***}
	(0.011)	(0.013)
$Immigrant$	0.066^{***}	0.078^{***}
	(0.014)	(0.017)
$VisibleMinority$	0.063^{***}	0.072^{***}
	(0.016)	(0.019)
$Female$	-0.060^{***}	-0.074^{***}
	(0.008)	(0.009)
Age	0.002	0.004
	(0.004)	(0.005)
Constant	2.856^{***}	
	(0.015)	
Observations	55,999	55,999
R^2	0.055	
Adjusted R^2	0.055	
Residual Std. Error	0.936 (df = 55987)	
F Statistic	295.705^{***} (df = 11; 55987)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 7: Ordered Probabilistic Regression - Average Marginal Effects

Statistic	Mean	St. Dev.	Min	Max
$log(Income)$	−0.012	0.008	−0.053	−0.004
$Income^2$	0.000	0.000	0	0
$Immigrant$	−0.003	0.002	−0.013	−0.001
$VisibleMinority$	−0.003	0.002	−0.012	−0.001
$Female$	0.003	0.002	0.001	0.012
Age	−0.00001	0.00001	−0.0001	−0.00000
Highest Education				
$LessthanSecondarySchool$	−0.006	0.003	−0.021	−0.002
$SecondarySchool$	−0.009	0.005	−0.030	−0.003
Marital Status				
$Married$	0.002	0.001	0.001	0.007
$Commonlaw$	0.008	0.004	0.003	0.030
$Widowed/Divorced/Separated$	0.006	0.003	0.002	0.022

Question 3

(a) To derive a model that determines whether tobacco consumption fell as a result of the policy enactment, it is possible to use the status of similar policies in Nova Scotia as the control group in a natural experiment. The extent to which the policy effects tobacco consumption could be modelled as follows:

$$\text{TobaccoConsumption}_{i,t} = \beta_0 + \beta_1 T_t + \beta_3 DiD_{i,t} + X_{i,t}\beta + e_i$$

Where T_t is the difference in time from the point when the policy (T) was enacted, and the time (t) that amount of tobacco consumption is being observed. $DiD_{i,t}$ is the difference in differences coefficient that in this case estimates the difference in tobacco consumption between observations (i) (respondents) in Nova Scotia versus New Brunswick, and between the before and after time periods (t) relative to the policy change. The caveat of the model is that competing policy changes and general macroeconomic conditions could bias the results.

(b) It would be possible to test the assumption of parallel trends by by viewing the difference-in-differences relationship between tobacco consumption and the explanatory variables at a point prior to the policy change. That is, change T to an earlier point in time, and view the $DiD_{i,t}$ for a before and after period within a time frame prior to the actual policy change. If there is no effect, then there is a parallel trend between the NS and NB groups and they are adequate for comparison.

(c) To control for heterogenous effects between indigenous and non-indigenous peoples, we would need to define a dummy variable Z with values corresponding to whether the repondent belongs to each group.

$$\text{TobaccoConsumption}_{i,t} = \beta_0 + \beta_1 T_t + \beta_2 Z_{i,t} + \beta_3 DiD_{i,t} + \beta_4 (Z * T)_{i,t} + \beta_5 INT + X_{i,t}\beta + e_i$$

Where, in the case of tobacco consumption following the policy change, INT would be the time relative to the policy change interacted with whether the respondent was a NB or NS resident, and whether the respondent was indigenous. $INT = (Time) * (NBResident) * (Z)$. The differnce in the coefficient INT when Z=1 and Z=0 would be interpreted as the magnitude of the hetergenous effect on indigenous peoples tobacco consumption following such a policy change.