ECO220Y: Homework, Lecture 20 – SOLUTIONS

(1)

(a) We have observational data and hence violation of Assumption # 5. We know there must be factors we haven't included in our model, but that are correlated with both education and spending on lotteries. From the estimated regression, we can conclude that education and spending on lotteries are negatively linearly related and that on average spending on lotteries as a share of income is smaller by 0.7% in the households where the head of the household has on average one more year of education. This is simply a descriptive statement (we can also compute the sample means for groups of individuals who have different education levels and compare). However, we cannot claim that if an individual gets one more year of education, his/her spending on lotteries will decline, i.e. we should not make any casual statements about the relationship between lottery spending and education level.

(b) We can say that the intercept in this regression has no interpretation: in our sample we do not have respondents who have 0 years of education (refer to the Stata summary). That's why we cannot say that if individual has zero years of education, he/she spends 14% of income on lotteries. This is simply because we do not have reliable basis to predict what would happen if individual has no education at all.

(c) First, write down the estimated model:

$$Lottery\_hat = 14.33 - 0.699 * Education$$

Now, check the means of Education and Lottery in Stata summary tables: E[Education]=12.78 and E[Lottery]=5.39

Now, plug these values into the estimated model – left-hand side should be equal to the right-hand side:

5.39=14.33-0.699*12.78

(d)

Prediction interval

$\hat{y}$	$t_{\alpha/2}$	s	$(x_g-\bar{x})^2$	$(n-1)s_x^2$	$\sqrt{1+\frac{1}{n}+\frac{(x_g-\bar{x})^2}{(n-1)s_x^2}}$	LP	UP	LC	UC
3.845	1.98	2.99	4.93	1114.74	1.00714	-2.117	9.807		
3.845	1.66	2.99	4.93	1114.74	1.00714	-1.154	8.844		
					$\sqrt{\frac{1}{n}+\frac{(x_g-\bar{x})^2}{(n-1)s_x^2}}$				
3.845	1.98	2.99	4.93	1114.74	0.11975			3.136	4.554
3.845	1.66	2.99	4.93	1114.74	0.11975			3.251	4.440

Confidence interval