Ulas Eraslan 25058

ECON 301 - Problem Set 3

1. 38.4% of the population (across 807 observation) is smoker.
24.65% of the population (across 807 observation) resides in a state with restaurant smoking restrictions.
The difference in average smoking probability between states with restaurant smoking restrictions and states without restrictions 10.3%

2. In regression model; education, cigarette prices, being white , age , income level, resides in a state with restaurant smoking restrictions and age squared is used.

Results are the following:

```
. reg smoker educ cigpric white age income restaurn
```

Source	SS	df	MS		
				Number of obs	= 807
				F(6, 800)	= 5.57
Model	7.65188433	6	1.27531406	Prob > F	= 0.0000
Residual	183.265092	800	.229081365	R-squared	= 0.0401
				Adj R-squared	= 0.0329
Total	190.916976	806	.236869698	Root MSE	= .47862

smoker	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.0249073	.0059394	-4.19	0.000	-.036566	-.0132486
cigpric	-.0008891	.0036019	-0.25	0.805	-.0079595	.0061812
white	-.0125129	.0519488	-0.24	0.810	-.114485	.0894592
age	-.0036367	.0010083	-3.61	0.000	-.005616	-.0016574
income	2.66e-06	1.97e-06	1.35	0.176	-1.20e-06	6.52e-06
restaurn	-.1035212	.0398964	-2.59	0.010	-.1818353	-.0252071
_cons	.8834777	.2372363	3.72	0.000	.4177986	1.349157

With 10% significance level, we may omit cigarette price, being white and income level since they are not statistically significant.

Increase in one year of education, probability that being a smoker decreases by 2.8 %.
Increase of one year in age, increase the probability of being a smoker by 2 %.
Being in a state that have smoking restrictions in restaurants, decreases being a smoker by 10%.

3. Since the agesq is statistically significant, we can say that the effect of age on smoking probability is quadratic. To find the age that impact of age on smoking probability becomes negative, we can take the derivate of the formula by respect to age. So the result is 0.0202 + 2*-0.000264*age = 0 -> after **age 38** impact of age becomes negative.

smoker	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.028552	.0059364	-4.81	0.000	-.0402049	-.0168992
cigpric	-.001053	.0035634	-0.30	0.768	-.0080477	.0059417
white	-.0263784	.0514915	-0.51	0.609	-.127453	.0746962
age	.0202079	.0056356	3.59	0.000	.0091456	.0312703
income	4.64e-07	2.01e-06	0.23	0.818	-3.48e-06	4.41e-06
restaurn	-.1006295	.0394733	-2.55	0.011	-.1781131	-.0231459
agesq	-.000264	.0000614	-4.30	0.000	-.0003846	-.0001435
_cons	.5348554	.248302	2.15	0.032	.0474541	1.022257

4.

Source	SS	df	MS	Number of obs	=	807
				F(7, 799)	=	7.54
Model	11.8361161	7	1.69087372	Prob > F	=	0.0000
Residual	179.08086	799	.22413124	R-squared	=	0.0620
				Adj R-squared	=	0.0538
Total	190.916976	806	.236869698	Root MSE	=	.47343

smoker	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.0288802	.0059008	-4.89	0.000	-.0404632	-.0172973
lcigpric	-.0689254	.2041088	-0.34	0.736	-.4695782	.3317275
white	-.0256801	.0515172	-0.50	0.618	-.1268051	.0754449
age	.0198158	.005666	3.50	0.000	.0086938	.0309378
lincome	.012162	.0257245	0.47	0.636	-.0383335	.0626575
restaurn	-.1007616	.0394431	-2.55	0.011	-.1781859	-.0233374
agesq	-.0002598	.0000617	-4.21	0.000	-.0003809	-.0001387
_cons	.6560602	.854962	0.77	0.443	-1.022177	2.334297

As it can be seen from the results, cigarette prices are effecting the probability of smoking negatively. Increase of 1% in cigarette prices , decreases the probability of smoking by 0.068/100.

5.

```
. gen white_restrict = white*restaurn

. reg smoker educ white age restaurn agesq white_restrict
```

Source	SS	df	MS	Number of obs	=	807
				F(6, 800)	=	8.93
Model	11.977995	6	1.9963325	Prob > F	=	0.0000
Residual	178.938981	800	.223673727	R-squared	=	0.0627
				Adj R-squared	=	0.0557
Total	190.916976	806	.236869698	Root MSE	=	.47294

smoker	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	-.028523	.0057087	-5.00	0.000	-.0397287	-.0173172
white	-.0626063	.063624	-0.98	0.325	-.1874961	.0622834
age	.0208004	.0054581	3.81	0.000	.0100866	.0315142
restaurn	-.1906341	.0992685	-1.92	0.055	-.3854916	.0042234
agesq	-.0002706	.0000594	-4.55	0.000	-.0003873	-.000154
white_restrict	.105592	.108003	0.98	0.329	-.1064108	.3175948
_cons	.5013577	.1344813	3.73	0.000	.2373799	.7653355

By creating an interaction variable, we may control the claim that impact of restaurant smoking restrictions on smoking probability might be different for white and non-white individuals. As it can be seen from the results, this is not statistically significant at 10% significance level because of the p-value 0.329 >0.10.

6. No we can not see this as a casual effect. Because there might be some omitted variables that directly effect this variable. For instance, living in a child abundant area might affect the probability of smoking negatively, so the direction is negative and also it is rela-

ted with the restaurant smoking restrictions positively. Therefore, since restaurn includes the negative coefficient that comes form the child area issue, the restaurn is underestimated.

7. In this context, it might be logical to examine whether the company is productive or not. If the company is less productive therefore they might applied for the grant. So since we do not have any independent variable in the formula that represents the productivity level, we might omit some variables related to grant. In this sense, self-selection problem may lead to endogeneity as well. Another related variable in error term might be company's total capital level. If it is high then company might not apply for grant.

8. The result is the following:

```
. reg lscrap grant
```

Source	SS	df	MS	Number of obs	=	54
				F(1, 52)	=	0.02
Model	.039451758	1	.039451758	Prob > F	=	0.8895
Residual	105.323208	52	2.02544631	R-squared	=	0.0004
				Adj R-squared	=	-0.0188
Total	105.36266	53	1.98797472	Root MSE	=	1.4232

lscrap	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
grant	.0566004	.4055519	0.14	0.890	-.757199	.8703998
_cons	.408526	.2405616	1.70	0.095	-.0741962	.8912482

Since the p-value is quite high and the confidence interval includes 0, the effect of grant of scarp is not statistically significant. So it is significantly lower.

9. Result is the following:

```
. reg lscrap grant lscrap_1
```

Source	SS	df	MS	Number of obs	=	54
				F(2, 51)	=	174.94
Model	91.9584791	2	45.9792396	Prob > F	=	0.0000
Residual	13.4041809	51	.262827077	R-squared	=	0.8728
				Adj R-squared	=	0.8678
Total	105.36266	53	1.98797472	Root MSE	=	.51267

lscrap	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
grant	-.2539697	.1470311	-1.73	0.090	-.5491469	.0412076
lscrap_1	.8311606	.0444444	18.70	0.000	.7419347	.9203865
_cons	.021237	.0890967	0.24	0.813	-.1576321	.2001061

After adding the lscrap_1, the effect of grant on lscarp deceased from 0.56 to -0.25. So this shows that scrap rate of the company in year 1987 has an important positive effect on grant (grant and lscrap_1 is positively correlated). The effect of lscrap_1 on lscrap is significant because the p-value is almost 0 and confidence interval does not include 0 even at 95% confidence level. The change in grant can be explained by omitted variable bias. Since grant is correlated with lscrap_1, including lscrap_1 to the model decreases the bias on the grant.

10. Equation : **log wage = B0 + B1marijuana + B2education + B3female + u**

11. Equation for testing whether marijuana usage has different effects on wages for men and women: **log wage = B0 + B1marijuana*female + B2education + B3female + u**

To test that there are no differences in the effects of marijuana usage on wage for men and women, t-test can be used. So we can apply t-test on wage based on marijuana usage for female/male. If the result is statistically significantly then we can say that the the difference is different than 0.

12. **log wage = B0 + sigma1*marijuana_light + sigma2*marijuana_moderate + sigma3*marijuana_heavy+ B2*education + B3female + u** (Effects are measured in comparison to non-marijuana user (base category))

13. I think it is not much logical to estimate a causal effect from this survey. Because since there are others aspects in the error term and these might be correlated with marijuana usage such as age (if the age is high marijuana usage might be lower and also if the age is high wage might be higher), we over/under estimate the parameter. So due to endogeneity problem it is not easy to come up with casual relation.