

Name AK

ID AK

Midterm 2 –60 Points

You must answer all questions. Please write your name on every page. The exam is closed book and closed notes. You may use calculators, but they must not be graphing calculators. No cell phones. Do not use your own scratch paper.

You must show your work to receive full credit

I have neither given nor received unauthorized aid on this examination, nor have I concealed any similar misconduct by others.

Signature AK

Problem 1

Smoking has severe health consequences, and many policies are meant to curb smoking behavior. To begin our study of smoking behavior, we focus on a sample of expectant mothers to evaluate their smoking habits as a function of their family characteristics. To establish a baseline, we evaluate impact of respective education levels (in years) of the expectant mother and the father, *motheduc* and *fatheduc*, on the number of cigarette packs consumed per day, *packs*.

$$packs = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + u$$

Education levels are in years. The results from this regression are below.

Source	SS	df	MS	
Model	4.2315727	2	2.11578635	Number of obs = 1191
Residual	80.7221848	1188	.067947967	F(2, 1188) = 31.14
				Prob > F = 0.0000
				R-squared = 0.0498
				Adj R-squared = 0.0482
Total	84.9537575	1190	.071389712	Root MSE = .26067

packs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	-.0190001	.0040835	xxxxxx		
fatheduc	-.0067882	.0036011	xxxxxx		
_cons	.4273789	.0436991	xxxxxx		

+3
if
+ test

a.) Do the variables of the model jointly tell us anything about the expectant mother's smoking behavior? Please test this hypothesis at the 95% level, clearly stating your null and alternative hypotheses. (10 Points)

$H_0: \beta_1 = 0, \beta_2 = 0$
 $H_1: H_0 \text{ not true}$

$F_{stat} = 31.14$, $F_{crit} = 3$ for $q = 2$

$F_{stat} > F_{crit} \Rightarrow$ Reject the Null!

Some combination of *motheduc* and *fatheduc* are a significant predictor of expectant mother cigarette consumption

- b.) Using the 98% confidence level, please test whether the coefficient on *fatheduc* is significantly different from zero. State your null and alternative hypotheses, and show your work! (10 Points)

$$H_0: \beta_2 = 0 \quad +2 \quad t_{stat} = \frac{-0.00679 - 0}{0.0036} = -1.89 \quad t_{crit} = 2.225 \quad +1$$

$$H_A: \beta_2 \neq 0 \quad +3$$

$|t_{stat}| < t_{crit} \Rightarrow$ Fail to reject null! +2

At the 98% level, the father's education level has an insignificant effect on the expectant mother's cigarette consumption +2

- c.) Please **derive** an equation that allows me to test whether the effect of the expectant mother's education on her own smoking behavior is the same as the father's education. Along with the derivation, please state the null and alternative hypothesis, and write down any Stata commands required to generate new variables and run the regression. Show your work! (10 Points)

$$H_0: \beta_1 = \beta_2 \Rightarrow \theta = \beta_1 - \beta_2 = 0 \Rightarrow \beta_1 = \theta + \beta_2$$

$$H_A: \theta \neq 0 \quad +2$$

$$\text{Packs} = \beta_0 + \beta_1 \text{motheduc} + \beta_2 \text{fatheduc} + u$$

$$= \beta_0 + (\theta + \beta_2) \text{motheduc} + \beta_2 \text{fatheduc} + u$$

$$= \beta_0 + \theta \text{motheduc} + \beta_2 (\text{motheduc} + \text{fatheduc}) + u$$

$$\text{gen toteduc} = \text{motheduc} + \text{fatheduc} \quad +2$$

$$\text{reg packs motheduc toteduc} \quad +2$$

Problem 2

Adding to the regression in Problem 1a, we include market and policy variables that may be important for influencing smoking behavior. Specifically, we include the pre-tax price of cigarettes (in cents), *cigprice*, and the tax on cigarettes (in cents), *cigtax*, as follows:

$$packs = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 cigprice + \beta_4 cigtax + u$$

The results from estimating this equation are below:

Source	SS	df	MS		
Model	4.39816711	4	1.09954178	Number of obs =	1191
Residual	80.5555904	1186	.067922083	F(4, 1186) =	16.19
				Prob > F =	0.0000
				R-squared =	0.0518
				Adj R-squared =	0.0486
				Root MSE =	.26062

packs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	-.0189956	.0040828	xxxxxx	xxxxxx	xxxxxx
fatheduc	-.0069949	.0036096	xxxxxx	xxxxxx	xxxxxx
cigprice	-.0012073	.0015401	xxxxxx	xxxxxx	xxxxxx
cigtax	.0027044	.0020261	xxxxxx	xxxxxx	xxxxxx
_cons	.5348374	.1710825	xxxxxx	xxxxxx	xxxxxx

a. Is this model preferred to the model in Problem 1? Please test this at the 95% level, stating your null and alternative hypotheses. Show your work!!! (10 Points)

$H_0: \beta_3 = 0, \beta_4 = 0$ +2
 $H_A: H_0 \text{ not true}$

$g = 2$ +5
 $df = 1191 - 4 - 1 = 1186$ +5

$SSR_{ur} = 80.56$ +4
 $SSR_R = 80.72$

$$F_{stat} = \frac{\frac{SSR_{ur} - SSR_R}{g}}{\frac{SSR_{ur}}{df}} = \frac{\frac{80.72 - 80.56}{2}}{\frac{80.56}{1186}} = 1.178$$

+4

$F_{crit} = 3$ +2
 $F_{stat} < F_{crit}$

Fail to reject null! "cigprice" and "cigtax" jointly have an insignificant effect on the cigarette consumption.

You argue that “cents are cents”, and the value that matters for consumption is the pre-tax price plus the tax, not each of them individually. So, you suggest a new variable, *price*, which is $cigprice + cigtax$, and insist on the following new regression.

$$packs = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_5 cigtax + \beta_6 price + u$$

Source	SS	df	MS	Number of obs =	1191
Model	4.39816718	4	1.09954179	F(4, 1186) =	16.19
Residual	80.5555903	1186	.067922083	Prob > F =	0.0000
				R-squared =	0.0518
				Adj R-squared =	0.0486
Total	84.9537575	1190	.071389712	Root MSE =	.26062

packs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
motheduc	-.0189956	.0040828	xxxxxx	xxxxxx	xxxxxx
fatheduc	-.0069949	.0036096	xxxxxx	xxxxxx	xxxxxx
cigtax	.0039117	.0034592	xxxxxx	xxxxxx	xxxxxx
price	-.0012073	.0015401	xxxxxx	xxxxxx	xxxxxx
_cons	.5348375	.1710825	xxxxxx	xxxxxx	xxxxxx

b. Using the above table, please construct a 95% confidence interval for the effect of mother’s education on cigarette consumption. Please interpret this confidence interval. (10 Points)

$$t_{crit} = 1.96 \quad \underline{+2}$$

$$\hat{\beta}_1 - t_{crit} \cdot se(\hat{\beta}_1) < \beta_1 < \hat{\beta}_1 + t_{crit} \cdot se(\hat{\beta}_1)$$

$$-0.019 - 1.96 \cdot 0.00411 < \beta_1 < -0.019 + 1.96 \cdot 0.00411$$

$$\underline{-0.027 < \beta_1 < -0.011}$$

Within 95% confidence, a one year increase in mother's education decreases her smoking consumption between -0.027 and -0.011 packs per day.

+4

- c. Using the regression results in 'b', I claim that the effects of a cigarette tax (*cigtax*) are different than the effects of the pre-tax cigarette price (*cigprice*), thereby rejecting the claim that "cents are cents" when it comes to tax policy. What is the probability that I'm wrong? Along with calculating this probability, please indicate the null and alternative hypotheses to which the statement is referring. (10 Points)

$$H_0: \beta_5 = 0$$
$$H_A: \beta_5 \neq 0$$

$$t_{stat} = \frac{0.00391 - 0}{0.00346} = 1.13$$

+6

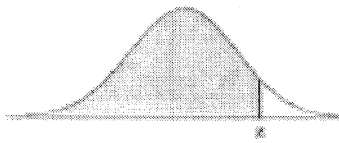
$$P_{value} = \Pr(|T| > t_{stat})$$

$$= \Pr(T < -t_{stat}) + \Pr(T > t_{stat})$$

$$= 2 \cdot (1 - \Pr(T < t_{stat}))$$

$$= 0.2584$$

+11



Normal Distribution from $-\infty$ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

TABLE G.3b

5% Critical Values of the F Distribution

		Numerator Degrees of Freedom:									
		1	2	3	4	5	6	7	8	9	10
∞		3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator $df (\infty)$ is 2.37.

Source: This table was generated using the Stata® function invFtail.