AAEC 6311

LAB #2

Objectives:

1) Estimate treatment effects using SAS

We will analyze the effect of education on female fertility. The data is for women of childbearing age in Botswana. The following variables are used in the model:
*variables used

children = number of living children (measure of fertility)
educ7 = 1 if attaining at least 7 years of education, =0 otherwise (this is our 'treatment")
age = age
evermarr = 1 if the person has been ever married, =0 otherwise
urban = 1 if lives in urban area, =0 otherwise
electric = 1 if lives in a house with electricity, =0 otherwise
tv = 1 if the household owns a television, = otherwise
frsthalf = 1 if woman was born during the first half of the year, =0 otherwise (this is our outside instrument).

Part 1. Basic Operations Using SAS

1.1. Import and manipulate the data

Proc import

Data

1.2. Calculate basic summary statistics:

2. The SAS System

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
children	**children**	4361	2.2678285	2.2220319	0	13.0000000
w		4361	0.5556065	0.4969553	0	1.0000000
age	**age**	4361	27.4051823	8.6852327	15.0000000	49.0000000
evermarr	**evermarr**	4361	0.4767255	0.4995153	0	1.0000000
urban	**urban**	4361	0.5166246	0.4997808	0	1.0000000
electric	**electric**	4358	0.1402019	0.3472363	0	1.0000000
tv	**tv**	4359	0.0929112	0.2903413	0	1.0000000
frsthalf	**frsthalf**	4361	0.5404724	0.4984164	0	1.0000000

Part 2. Estimate ATE and ATT under the assumption that the treatment is independent of fertility.

2.1. Calculate and report mean values for the fertility measure for women with and without at least 7 years of education:

The SAS System

The MEANS Procedure

w=0

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
children	**children**	1938	3.2512900	2.4697051	0	13.0000000
w		1938	0	0	0	0
age	**age**	1938	31.1682147	9.0200592	15.0000000	49.0000000
evermarr	**evermarr**	1938	0.6186791	0.4858365	0	1.0000000
urban	**urban**	1938	0.4112487	0.4921872	0	1.0000000
electric	**electric**	1937	0.0686629	0.2529453	0	1.0000000
tv	**tv**	1937	0.0268456	0.1616739	0	1.0000000
frsthalf	**frsthalf**	1938	0.5995872	0.4901085	0	1.0000000

w=1

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
children	**children**	2423	1.4812216	1.6177810	0	9.0000000
w		2423	1.0000000	0	1.0000000	1.0000000
age	**age**	2423	24.3953776	7.0943630	15.0000000	49.0000000
evermarr	**evermarr**	2423	0.3631861	0.4810171	0	1.0000000
urban	**urban**	2423	0.6009080	0.4898128	0	1.0000000
electric	**electric**	2421	0.1974391	0.3981487	0	1.0000000
tv	**tv**	2422	0.1457473	0.3529256	0	1.0000000
frsthalf	**frsthalf**	2423	0.4931903	0.5000568	0	1.0000000

2.2. Test if the average value of children for women with and without at least 7 years of education is statistically different than zero (α=0.05).

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: children children

Number of Observations Read 4361

Number of Observations Used 4361

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3373.65898	3373.65898	810.08	<.0001
Error	4359	18154	4.16461		
Corrected Total	4360	21527			

Root MSE	2.04074	**R-Square**	0.1567
Dependent Mean	2.26783	**Adj R-Sq**	0.1565
Coeff Var	89.98637		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	3.25129	0.04636	70.14	<.0001
w		**1**	-1.77007	0.06219	-28.46	<.0001

Since $\alpha = -1.77$, and the p-value is < 0.0001, we find evidence to reject the null hypothesis and conclude that women with and without at least 7 years of education have more than zero children.

2.3. Can we really give this difference in mean values a causality interpretation? In other words, is education independent of fertility? Why yes or no?

No, we cannot give this difference in mean values a causality interpretation. Although education may or may not cause changes in fertility, we cannot conclude from this analysis since there are other variables interacting with education. Additionally, there are unobservables from our error term that may also affect fertility.

Part 3. Estimate ATE and ATT under the assumption that the treatment is independent of fertility, conditional on a set of demographic factors (age, evermarr, urban, electric, tv).

Part 3.A. Assuming homogenous treatment effects

3.1. Express mathematically the regression model to be estimated

$$\hat{y} = \gamma + \alpha w + x'\beta_0 + \hat{\epsilon}$$

$$\hat{y} = \hat{b}_0 + \hat{b}_1 age + \hat{b}_2 evermarr + \hat{b}_3 urban + \hat{b}_4 electric + \hat{b}_5 tv + \hat{\epsilon}$$

3.2. Estimate and report the regression model based on 3.1.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: children children

Number of Observations Read	4361
Number of Observations Used	4358
Number of Observations with Missing Values	3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	12510	2085.04780	1008.03	<.0001
Error	4351	8999.74531	2.06843		
Corrected Total	4357	21510			

Root MSE	1.43820	**R-Square**	0.5816
Dependent Mean	2.26755	**Adj R-Sq**	0.5810
Coeff Var	63.42536		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-1.96811	0.09478	-20.77	<.0001
w		**1**	-0.38665	0.04981	-7.76	<.0001
age	age	**1**	0.15604	0.00309	50.45	<.0001
evermarr	evermarr	**1**	0.77829	0.05125	15.19	<.0001

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
urban	urban	**1**	-0.22901	0.04622	-4.95	<.0001
electric	electric	**1**	-0.33188	0.07585	-4.38	<.0001
tv	tv	**1**	-0.33947	0.09023	-3.76	0.0002

3.3. Interpret the coefficient corresponding to education and estimated in 3.2.

On average, if a woman has at least 7 years of education, she would have 0.38665 fewer children, ceteris paribus.

Part 3.B. Assuming heterogeneous treatment effects

3.4. Express mathematically the regression model to be estimated

$$\hat{y} = \gamma + \alpha w + x'\hat{\beta}_0 + w(x - \bar{x})'\delta + \hat{\epsilon}$$

$$\hat{y} = \widehat{b_0} + \widehat{b_1}age + \widehat{b_2}evermarr + \widehat{b_3}urban + \widehat{b_4}agedm + \widehat{b_5}evermarrdm + \widehat{b_6}urbandm + \widehat{b_7}electric + \widehat{b_8}tv + \hat{\epsilon}$$

3.5. Estimate and report the regression model based on 3.4. Use interactions of education and age, education and evermarr, and education and urban.

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: children children

Number of Observations Read	4361
Number of Observations Used	4358
Number of Observations with Missing Values	3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	12565	1396.07331	678.58	<.0001

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	4348	8945.37236	2.05735		
Corrected Total	4357	21510			

Root MSE	1.43435	**R-Square**	0.5841
Dependent Mean	2.26755	**Adj R-Sq**	0.5833
Coeff Var	63.25529		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-2.06384	0.12208	-16.91	<.0001
w		**1**	-0.37918	0.04980	-7.61	<.0001
age	age	**1**	0.15677	0.00395	39.74	<.0001
evermarr	evermarr	**1**	0.98662	0.07320	13.48	<.0001
urban	urban	**1**	-0.36926	0.06711	-5.50	<.0001
w_agedm		**1**	-0.00245	0.00632	-0.39	0.6987
w_evermarrdm		**1**	-0.38408	0.10329	-3.72	0.0002
w_urbandm		**1**	0.25962	0.08997	2.89	0.0039
electric	electric	**1**	-0.31086	0.07584	-4.10	<.0001
tv	tv	**1**	-0.32484	0.09099	-3.57	0.0004

3.6. Interpret the coefficient corresponding to education and estimated in 3.5. Interpret also the coefficients corresponding to interactions of education and the sociodemographic characteristics.

- If a woman has at least 7 years of education, she would have 0.37918 fewer children, ceteris paribus. In other words, our average treatment effect for a woman having at least 7 years of education would be 0.37918 fewer children, ceteris paribus. Additionally, there are interaction effects between having at least 7 years of education with age, marriage, and living in an urban environment.

- For each additional year of age, a woman with at least 7 years of education would have an additional -0.00245 affect on the average treatment effect, ceteris paribus.
- If a woman has been married and has at least 7 years of education, she would have 0.76326 fewer children, ceteris paribus.
- If a woman is living in an urban environment and has at least 7 years of education, she would have 0.11956 fewer children, ceteris paribus.
- This is also not to say the age, marriage, or the urban environment *causes* fertility rates or causes changes in education. There is simply an interaction effect between those variables, not a causal one.

3.7. Interpret the coefficient corresponding to tv and estimated in 3.5.

If a woman owns a TV, with or without at 7 years of education (the general population), she would have -0.32484 fewer children, on average, ceteris paribus.

3.8. Compare the ATE (and standard errors) estimated in 3.2 and 3.5.

The estimated ATE coefficient assuming homogeneous treatment effect was -0.38665 and the standard error was 0.04981. The estimated ATE coefficient assuming heterogeneous treatment effect was -0.37918 and the standard error was 0.04980. For the most part, the coefficient of our estimates and the corresponding standard errors are very similar.

Part 4. Estimate ATE and ATT under the assumption that the treatment is NOT independent of fertility, conditional on a set of demographic factors (age, evermarr, urban, electric, tv).

Part 4.A. Assuming homogenous treatment effects

4.1. Express mathematically the regression model to be estimated

$$\hat{y} = \gamma + \alpha w + x'\beta_0 + \hat{\epsilon}$$

$$\hat{y} = \hat{b}_0 + \widehat{\mathrm{b}_1}\mathrm{frsthalf} + \hat{b}_2 age + \hat{b}_3 evermarr + \hat{b}_4 urban + \hat{b}_5 electric + \hat{b}_6 tv + \hat{\epsilon}$$

4.2. Why do you think education is endogenous? Is this due to simultaneity? Is it due to unobserved or omitted variables?

Education may be endogenous because it affects the number of children a woman may have and the number of children a woman has may affect the education she has. This is because a woman with less education may have fewer work opportunities. It may also be due to omitted and unobserved variables. If we were to ask every women from the study the factors that led to their decision(s) of having more or fewer children, they would certainly give different and more varied reasons and variables.

4.3. Provide an "argument "for the validity (i.e., exogeneity and relevance of instrument) "frsthalf" as an instrument for education (see class notes).

4.4. Estimate and report the regression model based on 4.1. Use 2SL procedures with "frsthalf" as an instrument for treatment effect.

-Evaluate the potential weakness of frsthalf as an instrument for treatment effect.

Since we are only using one instrument, then the F-value is $t^2=-1.78^2=3.1684<10$, so frsthalf is a weak instrument.

-Report the 2SLS regression results

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: children children

Number of Observations Read	4361
Number of Observations Used	4358
Number of Observations with Missing Values	3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	12392	2065.38101	985.60	<.0001
Error	4351	9117.74606	2.09555		
Corrected Total	4357	21510			

Root MSE	1.44760	**R-Square**	0.5761
Dependent Mean	2.26755	**Adj R-Sq**	0.5755
Coeff Var	63.83981		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-1.24202	0.63774	-1.95	0.0515
childrenhat	Predicted Value of w	**1**	-1.08639	0.60974	-1.78	0.0749
age	age	**1**	0.14247	0.01219	11.69	<.0001
evermarr	evermarr	**1**	0.70769	0.08012	8.83	<.0001
urban	urban	**1**	-0.15578	0.07879	-1.98	0.0481
electric	electric	**1**	-0.23451	0.11393	-2.06	0.0396
tv	tv	**1**	-0.16072	0.17985	-0.89	0.3715

4.5. Interpret the coefficient corresponding to education and estimated in 4.4.

When a woman is born within the first half of the year, then we expect, on average, her to have 1.08639 fewer children, ceteris paribus.

4.6. Use a simple control function approach with "frsthalf" as an instrument for treatment effect.

- Express mathematically the regression model to be estimated

$$\hat{y} = x_1\beta + x_2\lambda + \rho\mu + w$$

$$\hat{y} = \widehat{b_0} + \widehat{b_1}frsthalf + \widehat{b_2}age + \widehat{b_3}evermarr + \widehat{b_4}urban + \widehat{b_5}electric + \widehat{b_6}tv + \widehat{b_7}ehat$$

-Report the control function estimation results

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: children children

Number of Observations Read	4361
Number of Observations Used	4358
Number of Observations with Missing Values	3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	12513	1787.58077	864.29	<.0001
Error	4350	8996.96675	2.06827		
Corrected Total	4357	21510			

Root MSE	1.43815	**R-Square**	0.5817
Dependent Mean	2.26755	**Adj R-Sq**	0.5811
Coeff Var	63.42286		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	**1**	-1.24202	0.63357	-1.96	0.0500
w		**1**	-1.08639	0.60576	-1.79	0.0730
age	age	**1**	0.14247	0.01211	11.77	<.0001
evermarr	evermarr	**1**	0.70769	0.07960	8.89	<.0001
urban	urban	**1**	-0.15578	0.07828	-1.99	0.0466
electric	electric	**1**	-0.23451	0.11318	-2.07	0.0383
tv	tv	**1**	-0.16072	0.17867	-0.90	0.3684
ehat1	Residual	**1**	0.70450	0.60782	1.16	0.2465

-Report the results of a test for endogeneity of education.

H_0: rho = 0

H_A: rho ≠ 0

Since the t-value for rho (in this case, ehat1) is 1.16 and the p-value is 0.2465, this is not significant at a α=0.05 level. Therefore, we fail to reject the null hypothesis and find that there is no evidence of endogeneity

4.7. Compare the coefficients and standard errors corresponding to education estimated using OLS (3.2.), 2SLS (4.4) and the simple control function approach in 4.6.

Part 5.B. Assuming heterogeneous treatment effects

4.10. Express mathematically the regression model to be estimated

$$\hat{y} = \gamma + \alpha w + x'\hat{\beta}_0 + w(x - \bar{x})'\delta + \hat{\epsilon}$$

$$\hat{y} = \widehat{b_0} + \widehat{b_1} frsthalf + \widehat{b_2} age + \widehat{b_3} evermarr + \widehat{b_4} urban + \widehat{b_5} agedm + \widehat{b_6} evermarrdm + \widehat{b_7} urbandm + \widehat{b_8} electric + \widehat{b_9} tv + \hat{\epsilon}$$

4.11. Estimate and report the regression model based on 4.10 and estimated using 2SLS. Use interactions of education and age, education and evermarr, and education and urban.

The SAS System

The MODEL Procedure

Nonlinear 2SLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Label
children	10	4348	10806.8	2.4855	1.5765	0.4976	0.4966	children

Nonlinear 2SLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
bo	-3.52059	1.2596	-2.80	0.0052
b1	-1.00888	0.7584	-1.33	0.1835
b2	0.193428	0.0345	5.61	<.0001
b3	1.354224	0.6706	2.02	0.0435
b4	0.367731	0.6489	0.57	0.5709
b5	-0.11358	0.1362	-0.83	0.4044
b6	0.185616	0.2829	0.66	0.5118
b7	-0.12015	0.0801	-1.50	0.1339
b8	-0.90494	1.2573	-0.72	0.4717
b9	-0.97946	1.1176	-0.88	0.3809

Number of Observations		Statistics for System	
Used	4358	**Objective**	1.254E-25
Missing	3	**Objective*N**	5.466E-22

4.12. Interpret the coefficient corresponding to education and estimated in 4.11. Interpret also the coefficients corresponding to interactions of education and the sociodemographic characteristics.

- On average, a woman born within the first half of the year will have 1.00888 fewer children, ceteris paribus.
- On average, a woman born within the first half of the year for every additional year of age, she will have 0.11358 fewer children, ceteris paribus.
- On average, a woman born within the first half of the year and has been married, will have 1.008888+0.185616 more children, ceteris paribus.
- On average, a woman born within the first half of the year and lives in an urban environment, she will have 1.008888-0.12015.
- Again, this is not to say age, marriage, and living in an urban environment causes differences in fertility rates.

4.13. Compare the coefficients and standard errors corresponding to education estimated using 2SLS (4.4) (homogenous treatment effects) and the 2SLS with heterogeneous treatment effects in 4.11.

For women born in the first half of the year, in the homogeneous treatment, the average treated effect was -1.08639 and the standard error was 0.60974. For women born in the first half of the year, in the heterogenous treatment, the average treated effect was -1.00888 and the standard error was 0.7584. We find the coefficient for education to be similar between the homogeneous and heterogeneous treatment effects, however, the standard error for the heterogeneous effect was larger.