(Lab assignment done in introduction to econometrics course at Stanislaus State)

Econ 3205 Lab 7

Ethan Machado

1. First regression:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 3432.06 | 11.8709 | 289.1154 | 0 | 3408.784 | 3455.336 | 3408.784 | 3455.336 |
| smoker | -253.228 | 26.95149 | -9.39571 | 1.09E-20 | -306.074 | -200.383 | -306.074 | -200.383 |

Birthweight (grams) = 3432.06 – 253.23(Smoker)

1. Second regression:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 3051.249 | 34.01596 | 89.7005 | 0 | 2984.552 | 3117.946 | 2984.552 | 3117.946 |
| nprevist | 34.06991 | 2.854994 | 11.93344 | 4.18E-32 | 28.47197 | 39.66786 | 28.47197 | 39.66786 |
| alcohol | -30.4913 | 76.23405 | -0.39997 | 0.689208 | -179.968 | 118.9851 | -179.968 | 118.9851 |
| smoker | -217.58 | 26.6796 | -8.1553 | 5.07E-16 | -269.892 | -165.268 | -269.892 | -165.268 |

Birthweight (g) = 3051.25 – 217.58(Smoker) – 30.49(Alcohol) + 34.07(Nprevisit)

1. Third regression:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 3134.4 | 35.656 | 87.90667 | 0 | 3064.487 | 3204.313 | 3064.487 | 3204.313 |
| nprevist | 29.60254 | 2.898385 | 10.21346 | 4.27E-24 | 23.91951 | 35.28557 | 23.91951 | 35.28557 |
| alcohol | -21.0835 | 75.60748 | -0.27885 | 0.780376 | -169.331 | 127.1644 | -169.331 | 127.1644 |
| smoker | -175.377 | 27.09873 | -6.47178 | 1.13E-10 | -228.511 | -122.243 | -228.511 | -122.243 |
| unmarried | -187.133 | 26.00749 | -7.19536 | 7.84E-13 | -238.128 | -136.139 | -238.128 | -136.139 |

Birthweight (g) = 3134.4 – 175.38(Smoker) – 21.08(Alcohol) + 29.60(Nprevisit) – 187.13(Unmarried)

1. First regression: The estimated effect of smoking on birth weight is -253.23 grams in birthweight if the mother is indeed a smoker.

Second regression: The estimated effect of smoking on birth weight is -217.58 grams in birthweight if the mother is a smoker.

Third regression: The estimated effect of smoking on birth weight is -175.38 grams in birthweight if the mother is a smoker.

1. First regression: 95% confidence interval: -306.07 <= b1 (smoker coeff.) <= -200.38

Second regression: 95% confidence interval: -269.89 <= b1 (smoker coeff.) <= -165.27

Third regression: 95% confidence interval: -228.51 <= b1 (smoker coeff.) <= -122.24

1. Yes, the smoker coefficient does suffer from omitted variable bias in the first regression. The coefficient decreased from -253.23 to -217.58 from regression 1 to regression 2, a change of 14.08%, compared to the coefficient’s standard error which decreased by 1%. The coefficient’s magnitude of change and small improvement in its standard error shows that it does suffer from OVB. Adding Nprevisit and Alcohol into regression 2 reduced the overestimated effect of Smoker on birthweight. Additionally, the R2 increased (2.83% to 7.19%) and standard error of regression decreased (583.73 to 570.47), so the independent variables explained more of the variation in birthweight and the average mistake of the regression was reduced.
2. Yes, the smoker coefficient does suffer from omitted variable bias in the second regression. The coefficient decreased from -217.58 to -175.38 from regression 2 to regression 3, a change of 19.39%, compared to the coefficient’s standard error which increased by 1.57%. The coefficient’s magnitude of change and small worsening in its standard error shows that it does suffer from OVB. Adding Unmarried into regression 3 reduced the overestimated effect of Smoker on birthweight. Additionally, the R2 increased (7.19% to 8.74%) and standard error of regression decreased (570.47 to 565.70), so the independent variables explained more of the variation in birthweight and the average mistake of the regression was reduced.
3. Third regression (Unmarried):
4. Unmarried 95% confidence interval: -238.13 <= b4 <= -136.14
5. Yes, Unmarried’s coefficient is statistically significant. Its test statistic is -7.19 (absolute value = 7.19) which is much greater than 1.96 (5% significance), and its p-value is 7.84E-13 which is much less than 0.05. The critical value test and p-value test both signify that the coefficient is statistically significant, and its confidence interval does not include zero.
6. Yes, Unmarried’s coefficient is large (-187.13). It is larger than every other coefficient in regression three. This means that it has the largest impact on a baby’s birthweight. (Nprevisit could become larger in total magnitude, but a mother would need to have more than 6 prenatal visits.)
7. No, I do not agree that marriage should be the focus for public policy. The Unmarried variable is a control variable, so its coefficient should not be interpretated. While in regression three, one would think that being married leads to healthier babies, the Unmarried variable is controlling for factors that are correlated with being married but not the effect of marriage on birthweight itself. Unmarried may possibly control for other factors such as income, age, psychological factors, etc. that can have effects on birthweight that are difficult to record data for. Married families usually have higher incomes than single mothers. Young mothers are usually not married as much as older mothers. Single mothers may suffer from increased psychological effects that are not as harsh when having support from a spouse.
8. My model: Birthweight (g) = b0 + b1(Smoker) + b2(Unmarried) + b3(Nprevisit)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regressor | (1) | (2) | (3) | My model |
| Smoker | -253.23  (26.95)  [-306.07,-200.38] | -217.58  (26.78)  [-269.89,-165.27] | -175.38  27.09  [-228.51,-122.243] | -176.21  (26.93)  [-229.01,-123.41] |
| Alcohol |  | -30.49  (76.23)  [-179.97,118.95] | -21.08  (75.61)  [-169.33,127.16] |  |
| Nprevisit |  | 34.07  (2.85)  [28.47,39.67] | 29.60  (2.90)  [23.92,35.28] | 29.62  (2.90)  [23.94,35.30] |
| Unmarried |  |  | -187.13  (26.01)  [-238.13,-136.14] | -187.26  (26)  [-238.24,-136.28] |
| Intercept | 3432.06  (11.87)  [3408.78,3455.34] | 3051.25  (34.01)  [2984.55,3117.95] | 3134.4  (35.66)  [3064.49,3204.31] | 3133.96  (35.61)  [3064.12,3203.79] |
| Adj. R2 | 0.028 | 0.072 | 0.087 | 0.087 |
| SER | 583.730 | 570.471 | 565.697 | 565.61 |
| n | 3000 | 3000 | 3000 | 3000 |

I chose to include the variables Nprevisit and Unmarried in my model because it provided the best results in a compact manner. Nprevisit is important to include in regression model because it helps to reduce OVB and reduce the bias on my variable of interest, Smoker. Unmarried is a great control variable to use in this regression because it can control for parameters such as income, psychological support, etc. that can have a noticeable effect on birthweight. I believe that they make my variable of interest (Smoker) less biased than if the model only had one regressor (Smoker). My model provides nearly the same coefficients, intercepts, and goodness of fit indicators as regression 3 with the exclusion of alcohol. Alcohol has little impact on Smoker when Unmarried and Nprevisit are included in the regression. Thus, my model provides the same variation in birthweight with little change to coefficients compared to regression 3 in a more compact manner. Sensitivity tests show almost no change in the magnitude of the coefficients between my model and regression 3. A reasonable 95% confidence interval for the effect of smoking on birthweight is -229.01 <= b1 (smoker coeff.) <= -123.41.