ECOM20001: Econometrics 1

Tutorial 8: Suggested Solutions

Dummy Variable Trap

1. As the summary statistics for constant show, it is equal one for every observation. This means it is identical to the constant regressor. The reason why this is the case is because of the definitions of tripre0, tripre1, tripre2, tripre3:
   - tripre1: dummy for baby first had prenatal care in the 1st trimester (tripre1==1)
   - tripre2: dummy for baby first had prenatal care in the 2nd trimester (tripre2==1)
   - tripre3: dummy for baby first had prenatal care in the 3rd trimester (tripre3==1)
   - tripre0: dummy for baby never had prenatal care in any trimesters (tripre0==1)

So either a baby never had prenatal care, or they had prenatal care, and they had it for the first time in either 1st, 2nd, or 3rd trimester. Having prenatal care or not having prenatal care (for the first time in one of the first three trimesters) are mutually exclusive events, and hence either tripre0 equals one or one and only one of tripre1, tripre2, tripre3 equals one for every observation in the sample.

2. The first and third regressions are identical because the constant regressor is identical to constant. The second regression differs because it does not contain a constant at all.

3. The regression is subject to the dummy variable trap because constant and the constant regressor are perfectly collinear (as they are identical). The statistical program R drops constant from the regression, and keeps the constant regressor (or just the constant) to avoid the dummy variable trap.

4. The regression is subject to the dummy variable trap because tripre0, tripre1, tripre2, tripre3 are together perfectly collinear with the constant regressor. In the R code provided, R drops tripre3 to avoid the dummy variable trap. Given this, we interpret the regression coefficient estimates on tripre0, tripre1, tripre2 and their statistical significance as follows:
   - tripre0=-569.321 means RELATIVE to tripre3==1, babies with no prenatal care weigh 569 grams LESS than babies that had their first prenatal care in the 3rd trimester, and this difference is statistically significant at the 1% level

- tripre1=180.603 means RELATIVE to tripre3==1, babies with their first prenatal care in the 1st trimester weigh 181 grams MORE than babies that had their first prenatal care in the 3rd trimester, and this difference is statistically significant at the 1% level
- tripre2=55.707 means RELATIVE to tripre3==1, babies with their first prenatal care in the 2nd trimester weigh 56 grams MORE than babies that had their first prenatal care in the 3rd trimester, and this difference is NOT statistically significant at the 1% or 5% level

5. The base group in the regression is tripe0. Given this, we interpret the regression coefficient estimates on tripre1, tripre2, tripre3 and their statistical significance as follows:
   - tripre1=749.923 means RELATIVE to tripre0==1, babies with their first prenatal care in the 1st trimester weigh 749 grams MORE than babies that had no prenatal care in any trimester, and this difference is statistically significant at the 1% level
   - tripre2=625.028 means RELATIVE to tripre0==1, babies with their first prenatal care in the 2nd trimester weigh 625 grams MORE than babies that had no prenatal care in any trimester, and this difference is statistically significant at the 1% level
   - tripre3=569.321 means RELATIVE to tripre0==1, babies with their first prenatal care in the 3rd trimester weigh 569 grams MORE than babies that had no prenatal care in any trimester, and this difference is statistically significant at the 1% level

6. There is a clearer/easier interpretation of results in question 5 with a base group of babies where tripre0==1 relative to question 4 than when our base group was babies where tripre3==1. It makes clear that having prenatal care in any trimester has a large positive impact on birthweight relative to having no prenatal care in any trimester at all.

7. The regression coefficients and standard errors on alcohol in the regressions from questions 4. and 5. are identical.

Multicollinearity

8. Answering each of the questions in turn as they appear in the tutorial questions.
    - 30 of 3000 observations have tripre0 equal one.
    - Among these variables, 26 of 30 observations have gambles equal one as well. The high degree of correlation between the two variables raises a concern of imperfect multicollinearity between tripre0 and gambles in a regression where both are included as independent variables.
    - Given that tripre0=1-tripre1-tripre2-tripre3, this directly implies an imperfect multicollinearity concerns between gambles and tripre1, tripre2, tripre3 together in a regression where all are included as independent variables.
9. Regression results are presented in the table on the next page. Answering each of the questions in turn as they appear in the tutorial questions:
    - In column (1), we find a statistically significant coefficient on gambles that implies babies with mothers with problem gambling are 559 grams lower in weight relative to babies with mothers without problem gambling.
    - In column (2) the regression coefficient on gambles becomes statistically insignificant as it rises to -276 from -559 once nprevisit is included in the regression. The coefficient on nprevisit is significant, and is 27.88 which implies each additional prenatal visit is associated with a 28 gram higher baby weight.
        - **Omitted variable bias**, the large increase in the gambles coefficient when nprevisit is controlled for could arise if nprevisit is positively related to birthweight AND if nprevisit is negatively related to gambles. This would create a downward bias in the gambles coefficient in column (1) when nprevisit is **not** controlled for.
    - Only nprevisit and tripre3 are statistically significant in column (3). The 32.09 coefficient implies a 32 gram increase in birthweight associated with each additional prenatal visit. The tripre3 coefficient implies babies that have their first visit in the 3rd trimester are 385 grams heavier relative to babies who had no prenatal visits.
        - The coefficients on tripre1 and tripre2 are similarly large at 209 and 269 gram increases in birthweight, but are insignificant.
        - The coefficients on nprevisit, tripre1, tripre2, and tripre3 are similar in column (4) to those in column (3). However, there is more than a doubling of the standard errors on tripre1, tripre2, and tripre3 in column

(4), once gambles is also included as a regressor. This is due precisely to the multicollinearity between gambles and tripre1, tripre2, and tripre3 together discussed above. The collinearity makes it hard to disentangle the influence of gambles and tripre1, tripre2, and tripre3 together on birthweight, resulting in large standard errors on all of the regression coefficients.

- The smoker coefficient is statistically significant and very stable across columns (1)-(4). It ranges between -176.89 and -183.85 grams. The collinearity issues related to gambles and tripre1, tripre2, and tripre3 have no impact on the smoker regression coefficient estimate and standard errors.
- Column (3) is the preferred set of results to present to the Prime Minister. The coefficients on tripre1, tripre2, and tripre3 are important for policy and are much more precisely estimated in column (3) than in column (4) as the column (3) results do not suffer from multicollinearity due to the inclusion of gambles as a regressor. Importantly, the columns (3) and (4) coefficients on tripre1, tripre2, and tripre3 are very similar, so there is no major omitted variable bias to be concerned about in column (3) by not including gambles as a regressor. That is, gambles only influences the standard errors on tripre1, tripre2, and tripre3 not their regression coefficients, meaning it only creates a noisier regression through its inclusion as a regressor.
  - As a general rule: when including additional regressors in regressions, it is critical to assess omitted variable bias on other important regressors in a regression.
  - If the inclusion of a regressor like gambles in a regression has little impact on key regressors in a regression like tripre1, tripre2, and tripre3, but only increases their standard errors, then the best decision is to not include a regressor like gambles in a regression. It does not matter in terms of omitted variable bias, and only serves to create imprecise regression coefficients on important regressors because of multicollinearity.

	Dependent variable:			
	Baby Birthweight in Grams			
	(1)	(2)	(3)	(4)
Smoker	−183.85***	−176.89***	−178.21***	−178.16***
	(28.07)	(27.34)	(27.21)	(27.22)
Drinks Alcohol During Pregnancy	18.55	3.13	3.94	3.73
	(92.44)	(90.13)	(90.75)	(90.70)
Drinks per Week During Pregnancy	−8.88	−1.31	−3.03	−3.02
	(17.42)	(14.87)	(16.43)	(16.43)
Gambles	−559.40***	−276.52*		19.36
	(151.69)	(158.37)		(365.26)
Prenatal Visits		27.88***	32.09***	32.09***
		(3.70)	(4.25)	(4.25)
Prenatal Care in 1st Trimester			209.53	226.31
			(148.87)	(336.56)
Prenatal Care in 2nd Trimester			268.82*	285.58
			(146.65)	(335.29)
Prenatal Care in 3rd Trimester			385.35**	402.10
			(155.44)	(339.15)
Unmarried	−229.99***	−195.19***	−206.86***	−206.85***
	(31.42)	(30.96)	(31.30)	(31.30)
Years of Education	6.89	0.50	1.83	1.83
	(5.60)	(5.53)	(5.54)	(5.54)
Age	−1.97	−2.33	−2.14	−2.15
	(2.48)	(2.45)	(2.46)	(2.46)
Constant	3,439.84***	3,213.76***	2,923.52***	2,906.85***
	(85.74)	(91.58)	(159.57)	(341.32)
Observations	3,000	3,000	3,000	3,000
R2	0.07	0.09	0.09	0.09
Adjusted R2	0.06	0.09	0.09	0.09
Residual Std. Error	573.25 (df = 2992)	565.42 (df = 2991)	564.65 (df = 2989)	564.74 (df = 2988)
F Statistic	29.74*** (df = 7; 2992)	37.30*** (df = 8; 2991)	30.94*** (df = 10; 2989)	28.12*** (df = 11; 2988)
Note:				*p<0.1; **p<0.05; ***p<0.01