Exercise session 4

1. Your aim is to estimate how the number of prenatal examinations and several other characteristics influence the birth weight of a baby. Your initial hypothesis is that more responsible pregnant women visit the doctor more often and this leads to healthier and thus also bigger babies.

   (a) In your first specification, you run the following model:

   $$bwght = \beta_0 + \beta_1\, npvis + \beta_2\, npvis^2 + \beta_3\, monpre + \beta_4\, male + \varepsilon\ ,$$

   where $bwght$ is birth weight of the baby (in grams), $npvis$ is the number of prenatal doctor's visits, $monpre$ is the month on pregnancy in which the prenatal care began and $male$ is a dummy, equal to one if the baby is a boy and zero if it is a girl. You obtain the following results from Stata[1]:

   | Source | SS | df | MS | | Number of obs = | 1726 |
   |---|---|---|---|---|---|---|
   | | | | | | F( 4, 1721) = | 9.70 |
   | Model | 12848047.5 | 4 | 3212011.87 | | Prob > F = | 0.0000 |
   | Residual | 570003184 | 1721 | 331204.639 | | R-squared = | 0.0220 |
   | | | | | | Adj R-squared = | 0.0198 |
   | Total | 582851231 | 1725 | 337884.772 | | Root MSE = | 575.5 |

   | bwght | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
   |---|---|---|---|---|---|---|
   | npvis | 53.50974 | 11.41313 | 4.69 | 0.000 | 31.12468 | 75.8948 |
   | npvissq | -1.173175 | .3591552 | -3.27 | 0.001 | -1.877601 | -.4687481 |
   | monpre | 30.47033 | 12.40794 | 2.46 | 0.014 | 6.134091 | 54.80657 |
   | MALE | 76.69243 | 27.76083 | 2.76 | 0.006 | 22.24391 | 131.141 |
   | _cons | 2853.196 | 101.3073 | 28.16 | 0.000 | 2654.498 | 3051.895 |

   i. Is there strong evidence that $npvissq$ (stands for $npvis^2$) should be included in the model?
   ii. How do you interpret the negative coefficient of $npvissq$?
   iii. Holding $npvis$ and $monpre$ fixed, test the hypothesis that newborn boys weight by 100 grams more than newborn girls (at 95% confidence level).

---

[1] Stata is a statistical software, which can be used to for econometric purposes. The Stata output is quite similar to the Gretl output you are familiar with. In particular, *Coef.* denotes the estimated coefficients, *Std.Err.* denotes the standard errors of these coefficients, *t* denotes the *t*-statistic of the test of significance of the coefficients, $P > |t|$ denotes the corresponding *p*-value.

(b) A friend of yours, student of medicine, reminds you of the fact that the age of the parents (especially of the mother) might be a decisive factor for the health and for the weight of the baby. Therefore, in your second specification, you decide to include in your model also the age of the mother ($mage$) and of the father ($fage$). The results of your estimation are now the following:

| Source | SS | df | MS | | Number of obs = | 1720 |
|---|---|---|---|---|---|---|
| | | | | | F( 6, 1713) = | 8.25 |
| Model | 16270165.8 | 6 | 2711694.3 | | Prob > F = | 0.0000 |
| Residual | 563258231 | 1713 | 328813.912 | | R-squared = | 0.0281 |
| | | | | | Adj R-squared = | 0.0247 |
| Total | 579528396 | 1719 | 337131.121 | | Root MSE = | 573.42 |

| bwght | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| npvis | 52.43859 | 11.40558 | 4.60 | 0.000 | 30.06826 | 74.80891 |
| npvissq | -1.138545 | .3585648 | -3.18 | 0.002 | -1.841816 | -.4352743 |
| monpre | 34.35661 | 12.69477 | 2.71 | 0.007 | 9.457725 | 59.2555 |
| MALE | 74.45482 | 27.75247 | 2.68 | 0.007 | 20.02252 | 128.8871 |
| MAGE | .5285275 | 4.218069 | 0.13 | 0.900 | -7.744582 | 8.801637 |
| FAGE | 8.697342 | 3.465973 | 2.51 | 0.012 | 1.899357 | 15.49533 |
| _cons | 2592.813 | 139.6173 | 18.57 | 0.000 | 2318.974 | 2866.651 |

i. Comment on the significance of the coefficients on $mage$ and $fage$ sepa- rately: are they in line with your friend's claim?

ii. Test the hypothesis that $mage$ and $fage$ are jointly significant (at 95% confidence level). Is the result in line with your friend's claim?

iii. How can you reconcile you findings from the two previous questions?

(c) In your third specification, you decide to drop $fage$ and you get the following results:

| Source | SS | df | MS | | Number of obs = | 1726 |
|---|---|---|---|---|---|---|
| | | | | | F( 5, 1720) = | 8.75 |
| Model | 14451685.6 | 5 | 2890337.13 | | Prob > F = | 0.0000 |
| Residual | 568399545 | 1720 | 330464.852 | | R-squared = | 0.0248 |
| | | | | | Adj R-squared = | 0.0220 |
| Total | 582851231 | 1725 | 337884.772 | | Root MSE = | 574.86 |

| bwght | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| npvis | 52.27885 | 11.41406 | 4.58 | 0.000 | 29.89196 | 74.66575 |
| npvissq | -1.142647 | .3590214 | -3.18 | 0.001 | -1.846811 | -.4384821 |
| monpre | 35.25912 | 12.58328 | 2.80 | 0.005 | 10.57898 | 59.93927 |
| MALE | 79.38175 | 27.75667 | 2.86 | 0.004 | 24.94136 | 133.8221 |
| MAGE | -6.91257 | 3.137972 | -2.20 | 0.028 | -13.06721 | -.757928 |
| _cons | 2648.851 | 137.2778 | 19.30 | 0.000 | 2379.602 | 2918.1 |

Comment on the significance of the coefficient on *mage*, compared to the results

2

from part (b). Is your finding in line with your reasoning in part (b)? Does it confirm your friend's claim?

(d) Having regained trust in your friend, you consult your results once more with him. Together, you come up with an interesting question: whether smoking during pregnancy can affect the weight of the baby. Fortunately, you have at your disposition the variable *cigs*, standing for the average number of cigarettes each woman in your sample smokes per day during the pregnancy, and so you can include it in your model. However, your friend warns you that women who smoke during pregnancy are in general less responsible than those who do not smoke, and that these women also tend to visit the doctor less often. (In other words, the more the women smokes, the less prenatal doctor's visits she has). This is an important fact that you have to take into consideration while interpreting your final results, which are:

| Source | SS | df | MS | | Number of obs = | 1622 |
|---|---|---|---|---|---|---|
| | | | | | F( 6, 1615) = | 7.49 |
| Model | 14560828.9 | 6 | 2426804.81 | | Prob > F = | 0.0000 |
| Residual | 523281374 | 1615 | 324013.235 | | R-squared = | 0.0271 |
| | | | | | Adj R-squared = | 0.0235 |
| Total | 537842203 | 1621 | 331796.547 | | Root MSE = | 569.22 |

| bwght | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| npvis | 42.43442 | 11.59582 | 3.66 | 0.000 | 19.68999 | 65.17885 |
| npvissq | -.8948737 | .3624432 | -2.47 | 0.014 | -1.605782 | -.1839653 |
| monpre | 31.77658 | 12.78156 | 2.49 | 0.013 | 6.706395 | 56.84676 |
| male | 82.39438 | 28.34937 | 2.91 | 0.004 | 26.78897 | 137.9998 |
| mage | -6.980738 | 3.227181 | -2.16 | 0.031 | -13.31064 | -.6508356 |
| cigs | -10.209 | 3.398309 | -3.00 | 0.003 | -16.87456 | -3.54344 |
| _cons | 2748.856 | 141.868 | 19.38 | 0.000 | 2470.591 | 3027.12 |

i. Interpret the coefficient on *cigs*.
ii. What evidence do you find that *cigs* really should be included in the model? List at least two arguments.
iii. Compare the coefficient on *npvis* with the one you obtained in part (c). Do you think there was a bias? If yes, explain where it came from and interpret its sign.

3