Python Program Output (Terminal)

```
(base) divitshetty@Divits-MBP-2 A3 % python3 A3.py
Question 1

                            OLS Regression Results
==============================================================================
Dep. Variable:            birthweight   R-squared:                       0.029
Model:                            OLS   Adj. R-squared:                  0.028
Method:                 Least Squares   F-statistic:                     88.28
Date:                Fri, 27 Oct 2023   Prob (F-statistic):           1.09e-20
Time:                        20:07:26   Log-Likelihood:                -23364.
No. Observations:                3000   AIC:                         4.673e+04
Df Residuals:                    2998   BIC:                         4.674e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   3432.0600     11.871    289.115      0.000    3408.784    3455.336
smoker      -253.2284     26.951     -9.396      0.000    -306.074    -200.383
==============================================================================
Omnibus:                      473.891   Durbin-Watson:                   1.973
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1247.472
Skew:                          -0.858   Prob(JB):                    1.30e-271
Kurtosis:                       5.652   Cond. No.                         2.64
==============================================================================
```

```
What is the estimated effect of smoking on birthweight?
The effect of smoking on birthweight is that infants born to mothers who smoke have a birthweight that is lower by 253.23 grams.
```

```
Question 2

                            OLS Regression Results
==============================================================================
Dep. Variable:            birthweight   R-squared:                       0.073
Model:                            OLS   Adj. R-squared:                  0.072
Method:                 Least Squares   F-statistic:                     78.47
Date:                Fri, 27 Oct 2023   Prob (F-statistic):           7.31e-49
Time:                        20:07:26   Log-Likelihood:                -23294.
No. Observations:                3000   AIC:                         4.660e+04
Df Residuals:                    2996   BIC:                         4.662e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   3051.2486     34.016     89.701      0.000    2984.552    3117.946
smoker      -217.5801     26.680     -8.155      0.000    -269.892    -165.268
alcohol      -30.4913     76.234     -0.400      0.689    -179.968     118.985
nprevist      34.0699      2.855     11.933      0.000      28.472      39.668
==============================================================================
Omnibus:                      374.095   Durbin-Watson:                   1.974
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              869.220
Skew:                          -0.729   Prob(JB):                    1.78e-189
Kurtosis:                       5.197   Cond. No.                         85.2
==============================================================================
```

```
Question 2(c)
The predicted birthweight of Jane's child is 3106 grams

Question 2(d)
R-squared: 0.073
Adjusted R-squared: 0.072

Question 2(e)
Coefficient (slope) of nprevist: 34.07
```

```
Question 3

                            OLS Regression Results
==============================================================================
Dep. Variable:            birthweight   R-squared:                       0.077
Model:                            OLS   Adj. R-squared:                  0.075
Method:                 Least Squares   F-statistic:                     41.43
Date:                Fri, 27 Oct 2023   Prob (F-statistic):           9.39e-49
Time:                        20:07:26   Log-Likelihood:                -23288.
No. Observations:                3000   AIC:                          4.659e+04
Df Residuals:                    2993   BIC:                          4.663e+04
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   3051.5460     42.579     71.668      0.000    2968.059    3135.033
smoker      -218.8952     26.754     -8.182      0.000    -271.354    -166.437
alcohol      -15.1126     76.315     -0.198      0.843    -164.747     134.522
nprevist      33.8179      3.417      9.897      0.000      27.118      40.518
tripre0     -299.2764    112.636     -2.657      0.008    -520.128     -78.425
tripre2       10.4501     31.244      0.334      0.738     -50.812      71.712
tripre3      115.8851     63.962      1.812      0.070      -9.530     241.300
==============================================================================
Omnibus:                      373.949   Durbin-Watson:                   1.974
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              876.725
Skew:                          -0.726   Prob(JB):                    4.18e-191
Kurtosis:                       5.215   Cond. No.                          129.
==============================================================================
```

Python Program Text

Question 1

OLS Regression Results

==========

Dep. Variable: birthweight R-squared: 0.029

Model: OLS Adj. R-squared: 0.028

Method: Least Squares F-statistic: 88.28

Date: Fri, 27 Oct 2023 Prob (F-statistic): 1.09e-20

Time: 18:14:58 Log-Likelihood: -23364.

No. Observations: 3000 AIC: 4.673e+04

Df Residuals: 2998 BIC: 4.674e+04

Df Model: 1

Covariance Type: nonrobust

==========

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3432.0600	11.871	289.115	0.000	3408.784	3455.336

smoker -253.2284 26.951 -9.396 0.000 -306.074 -200.383

==========

Omnibus: 473.891 Durbin-Watson: 1.973
Prob(Omnibus): 0.000 Jarque-Bera (JB): 1247.472
Skew: -0.858 Prob(JB): 1.30e-271
Kurtosis: 5.652 Cond. No. 2.64

==========

The effect of smoking on birthweight is that infants born to mothers who smoke have a birth weight that is lower by 253.23 grams.

Question 2

OLS Regression Results

==========

Dep. Variable: birthweight R-squared: 0.073
Model: OLS Adj. R-squared: 0.072
Method: Least Squares F-statistic: 78.47
Date: Fri, 27 Oct 2023 Prob (F-statistic): 7.31e-49
Time: 18:14:58 Log-Likelihood: -23294.
No. Observations: 3000 AIC: 4.660e+04
Df Residuals: 2996 BIC: 4.662e+04
Df Model: 3
Covariance Type: nonrobust

==========

coef std err t P>|t| [0.025 0.975]

Intercept 3051.2486 34.016 89.701 0.000 2984.552 3117.946
smoker -217.5801 26.680 -8.155 0.000 -269.892 -165.268
alcohol -30.4913 76.234 -0.400 0.689 -179.968 118.985
nprevist 34.0699 2.855 11.933 0.000 28.472 39.668

==========

Omnibus: 374.095 Durbin-Watson: 1.974
Prob(Omnibus): 0.000 Jarque-Bera (JB): 869.220

Skew:	-0.729	Prob(JB):	1.78e-189
Kurtosis:	5.197	Cond. No.	85.2

==========

- Question 2(a)
    - **'alcohol' and 'nprevist' may have direct effects on 'birthweight.' If these variables aren't included in the model, the regression will incorrectly attribute these effects to the 'smoker' variable, leading to a biased estimate of the effect of smoking on birthweight.**

- Question 2(b)
    - **The estimated effect of smoking on birthweight in both regressions are negative, but the inclusion of 'alcohol' and 'nprevist' led to a more negative estimate of the effect of smoking on birth weight.**
        - **Thus, the regression in Q1 does seem to suffer from omitted variable bias**
- Question 2(c)
    - **The predicted birth weight of Jane's child is 3106 grams**
- Question 2(d)
    - **R-squared: 0.073**
    - **Adjusted R-squared: 0.072**
    - **The adjusted R-squared value is similar to the R-squared value because the number of predictors is not very high relative to the sample size.**
- Question 2(e)
    - The coefficient on Nprevist is 34.07.
        - **This means that for every additional prenatal care visit, birth-weight is estimated to increase by 34.07 grams, assuming other factors remain constant.**
            - **However, correlation doesn't imply causation. The relationship between 'nprevist' and 'birthweight' might be influenced by other unmeasured factors (error u).**
                - **The coefficient on 'nprevist' doesn't measure a causal effect of prenatal visits on birth weight, it instead quantifies the association between 'nprevist' and 'birthweight'**

Question 3

OLS Regression Results

```
==============================================================================
=========
Dep. Variable:          birthweight   R-squared:                       0.077
Model:                          OLS   Adj. R-squared:                  0.075
Method:               Least Squares   F-statistic:                     41.43
Date:              Fri, 27 Oct 2023   Prob (F-statistic):           9.39e-49
Time:                      19:26:59   Log-Likelihood:                -23288.
No. Observations:              3000   AIC:                         4.659e+04
Df Residuals:                  2993   BIC:                         4.663e+04
Df Model:                         6
Covariance Type:          nonrobust
==============================================================================
=========
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   3051.5460     42.579     71.668      0.000    2968.059    3135.033
smoker      -218.8952     26.754     -8.182      0.000    -271.354    -166.437
alcohol      -15.1126     76.315     -0.198      0.843    -164.747     134.522
nprevist      33.8179      3.417      9.897      0.000      27.118      40.518
tripre0     -299.2764    112.636     -2.657      0.008    -520.128     -78.425
tripre2       10.4501     31.244      0.334      0.738     -50.812      71.712
tripre3      115.8851     63.962      1.812      0.070      -9.530     241.300
==============================================================================
=========
Omnibus:                      373.949   Durbin-Watson:                   1.974
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              876.725
Skew:                          -0.726   Prob(JB):                    4.18e-191
Kurtosis:                       5.215   Cond. No.                        129.
==============================================================================
=========
```

- Question 3(a)
    - **If 'tripre1' were included, it would represent the fourth possibility: prenatal visits that started in the 1st trimester.**
        - **Including 'tripre1' would lead to perfect collinearity because 'tripre0' and 'tripre1' would be perfectly negatively correlated**
            - **Thus, 'tripre1' is excluded from the regression to avoid perfect collinearity.**
- Question 3(b)

- **Yes, the regression in Q3 explains a slightly larger fraction of the variance in birth weight than the regression in Q2 .**
  - **The R-squared value in Q3 is slightly higher than the R-squared value in Q2, which means that the regression model in Q3 is a slightly better explanation of the variation in birth weight compared to the regression model in Q2.**