Nhan Le
ECO 231W
HW 4

1)

a. i) crime rate (which might be linked to school quality)
   ii) yes, negative
b. i) population density (which might be linked to unemployment and income per capita)
   ii) yes, negative because urban areas have less criminal activity per capita than rural areas
c. i) student's ability (which might be linked to whether student attends public school)
   ii) yes, positive

2) The two-variable linear regression equation can be derived using the following formulas.

$$B_1 = \frac{\sum temp^2 \sum(price \times sold) - \sum(price \times temp) \sum(temp \times sold)}{\sum price^2 \sum temp^2 - ((\sum(temp \times price))^2}$$

$$B_2 = \frac{\sum price^2 \sum(temp \times sold) - \sum(price \times temp) \sum(price \times sold)}{\sum price^2 \sum temp^2 - ((\sum(temp \times price))^2}$$

$$B_0 = \overline{sold} - B_1(\overline{price}) - B_2(\overline{temp})$$

3)

```
##
## Call:
## lm(formula = attend.df$final ~ attend.df$attend)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.3570  -3.2361  -0.1152   3.1568  12.7639
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      22.72992    0.87691  25.921  < 2e-16 ***
## attend.df$attend  0.12090    0.03283   3.683 0.000249 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.667 on 678 degrees of freedom
## Multiple R-squared:  0.01961,    Adjusted R-squared:  0.01816
## F-statistic: 13.56 on 1 and 678 DF,  p-value: 0.0002493
```

a. The equation received is: $\text{final} = 22.73 + 0.12(\text{attend})$ which indicates that the base score for the final is 22.73 and increases by 0.12 for each class attended. Both values are statistically significant from 0.

Nhan Le
ECO 231W
HW 4

b.

```
##
## Call:
## lm(formula = attend.df$final ~ attend.df$attend + attend.df$skipped)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.3570  -3.2361  -0.1152   3.1568  12.7639
##
## Coefficients: (1 not defined because of singularities)
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       22.72992    0.87691  25.921  < 2e-16 ***
## attend.df$attend   0.12090    0.03283   3.683 0.000249 ***
## attend.df$skipped       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.667 on 678 degrees of freedom
## Multiple R-squared:  0.01961,    Adjusted R-squared:  0.01816
## F-statistic: 13.56 on 1 and 678 DF,  p-value: 0.0002493
```

The equation received is the same as above because skipped and attend are directly linked by $\text{skipped} = 32 - \text{attend}$. No new information can be gained from skipped.

c.

```
##
## Call:
## lm(formula = attend.df$final ~ attend.df$attend + attend.df$hwrte +
##     attend.df$priGPA + attend.df$ACT + attend.df$frosh + attend.df$soph)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.1292  -2.6933  -0.1603   2.8520  10.9175
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       9.45642    1.51824   6.229 8.33e-10 ***
## attend.df$attend  0.03732    0.04354   0.857   0.3917
## attend.df$hwrte   0.01976    0.01090   1.814   0.0702 .
## attend.df$priGPA  2.02491    0.39125   5.176 3.01e-07 ***
## attend.df$ACT     0.40050    0.05345   7.493 2.14e-13 ***
## attend.df$frosh  -0.26759    0.51868  -0.516   0.6061
## attend.df$soph   -0.81788    0.43021  -1.901   0.0577 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.216 on 667 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.211,  Adjusted R-squared:  0.2039
## F-statistic: 29.74 on 6 and 667 DF,  p-value: < 2.2e-16
```

The estimate for B_1 changed from 0.12 to 0.037 but is not statistically significant from 0. The final grade is better explained by the other variables.

Nhan Le
ECO 231W
HW 4

```
##
##  F test to compare two variances
##
## data:  attend.df$attend and attend.df$hwrte
## F = 0.080143, num df = 679, denom df = 673, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.06891621 0.09319385
## sample estimates:
## ratio of variances
##         0.08014297
```

d. The joint significance is statistically significant from 0 at 5% significance level.

```
##
##  F test to compare two variances
##
## data:  attend.df$attend and attend.df$hwrte
## F = 0.080143, num df = 679, denom df = 673, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 99 percent confidence interval:
##  0.06571953 0.09772452
## sample estimates:
## ratio of variances
##         0.08014297
```

The joint significance is statistically significant from 0 at 1% significance level.

4)

a. B_0 indicates that ln(wage) starts at 8.5 and B_1 indicates that ln(wage) increases by 1.4 per year of education.
b. B_0 indicates that ln(wage) starts at 8.5, B_1 indicates that ln(wage) increases by 1.5 per year of education for women, B_2 indicates that ln(wage) decreases by 2.3 for women, and B_3 indicates that ln(wage) increases by 0.97 per year of education regardless of gender.