**Data Science II**

**Predicting School Enrollment: R Final Project**

**REGRESSION OUTPUTS**

**Overall Regression model with all predictors (Both best and overall model):**

> overallmodel1 <- lm(county\_enroll\_DM ~ population\_DM + prop\_age1524\_DM + prop\_nonwhite\_DM + prop\_male\_DM +  
+ total\_crime\_rate\_old\_DM + new\_tot\_crime\_rate\_DM +  
+ violent\_rate\_old\_DM + new\_violent\_rate\_DM +  
+ percap\_income\_DM + employment\_rate\_DM + year, data = school\_dm\_final)  
> summary(overallmodel1)  
  
Call:  
lm(formula = county\_enroll\_DM ~ population\_DM + prop\_age1524\_DM +   
 prop\_nonwhite\_DM + prop\_male\_DM + total\_crime\_rate\_old\_DM +   
 new\_tot\_crime\_rate\_DM + violent\_rate\_old\_DM + new\_violent\_rate\_DM +   
 percap\_income\_DM + employment\_rate\_DM + year, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-127792 -327 69 459 66023   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -6.419e+04 2.761e+04 -2.325 0.020068 \*   
population\_DM 2.348e-01 1.325e-03 177.198 < 2e-16 \*\*\*  
prop\_age1524\_DM -4.148e+04 3.112e+03 -13.327 < 2e-16 \*\*\*  
prop\_nonwhite\_DM 5.872e+04 2.454e+03 23.924 < 2e-16 \*\*\*  
prop\_male\_DM -3.299e+04 3.849e+03 -8.572 < 2e-16 \*\*\*  
total\_crime\_rate\_old\_DM -4.126e-02 1.137e-02 -3.627 0.000287 \*\*\*  
new\_tot\_crime\_rate\_DM 1.459e-01 4.718e-02 3.092 0.001993 \*\*   
violent\_rate\_old\_DM -1.643e-01 1.393e-01 -1.179 0.238291   
new\_violent\_rate\_DM 7.980e-01 3.135e-01 2.545 0.010919 \*   
percap\_income\_DM -1.772e-02 1.624e-02 -1.092 0.275053   
employment\_rate\_DM -5.947e-02 7.455e-01 -0.080 0.936418   
year 3.217e+01 1.383e+01 2.325 0.020078 \*   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 3244 on 30752 degrees of freedom  
 (6880 observations deleted due to missingness)  
Multiple R-squared: 0.5699, Adjusted R-squared: 0.5698   
F-statistic: 3704 on 11 and 30752 DF, p-value: < 2.2e-16

**Explanation -**

This is the overall model which has all the predictor variables.

The R-squared value of the regression model is 0.5699 which indicates that 56.99% variability can be explained by the predictor variables. The F-value is 3704 with a corresponding value of < 0.0001 which is less than the significance level of 0.001. This means that the model is statistically significant.

Predictor variables: population\_DM, prop\_age1524\_DM, prop\_nonwhite\_DM, prop\_male\_DM, total\_crime\_rate\_old\_DM, new\_tot\_crime\_rate\_DM, new\_violent\_rate\_DM, year are statistically significant as indicated by high t-values (more than 2) and corresponding p-value less than the significance level of 0.05.

Whereas, violent\_rate\_old\_DM, percap\_income\_DM, employment\_DM are not statistically significant as indicated by the lower t-values and corresponding p-values.

Population\_DM (County population), prop\_nonwhite\_DM (proportion of nonwhite people), new\_tot\_crime\_rate (new total crime rate), new\_violent\_rate\_DM (new violent crime rates), have a positive impact on the county\_enroll\_DM (school enrollment) which means they lead to increase in school enrollment.

On the other hand, prop\_age1524\_DM (proportion of people aged 15-24 years), prop\_male\_DM (proportion of males), total\_crime\_rate\_old\_DM (old total crime rate) have negative impact on the county\_enroll\_DM (school enrollment) which means they lead to decrease in school enrollment.

population\_DM: Positive association. For each unit increase, county\_enroll\_DM increases by 0.2348.

prop\_age1524\_DM: Negative association. For each unit increase, county\_enroll\_DM decreases by 41,480.

prop\_nonwhite\_DM: Positive association. Suggests higher enrollment in counties with a greater proportion of nonwhite individuals.

prop\_male\_DM: Negative association. Suggests lower enrollment in counties with a higher proportion of males.

total\_crime\_rate\_old\_DM and new\_tot\_crime\_rate\_DM: Crime rates have small but significant effects.

new\_violent\_rate\_DM: A higher new violent crime rate is associated with increased enrollment, though the effect size is small.

year: Enrollment appears to increase over time.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

> overallmodel2 <- lm(county\_enroll\_DM ~ population\_DM + prop\_age1524\_DM + prop\_nonwhite\_DM + prop\_male\_DM +  
+ total\_crime\_rate\_old\_DM + new\_tot\_crime\_rate\_DM + new\_violent\_rate\_DM +  
+ year, data = school\_dm\_final)  
> summary(overallmodel2)  
  
Call:  
lm(formula = county\_enroll\_DM ~ population\_DM + prop\_age1524\_DM +   
 prop\_nonwhite\_DM + prop\_male\_DM + total\_crime\_rate\_old\_DM +   
 new\_tot\_crime\_rate\_DM + new\_violent\_rate\_DM + year, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-127599 -341 74 481 65301   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -3.743e+04 1.247e+04 -3.001 0.002690 \*\*   
population\_DM 2.364e-01 1.326e-03 178.311 < 2e-16 \*\*\*  
prop\_age1524\_DM -4.632e+04 3.100e+03 -14.940 < 2e-16 \*\*\*  
prop\_nonwhite\_DM 6.212e+04 2.472e+03 25.131 < 2e-16 \*\*\*  
prop\_male\_DM -3.342e+04 3.831e+03 -8.725 < 2e-16 \*\*\*  
total\_crime\_rate\_old\_DM -4.986e-02 8.911e-03 -5.596 2.21e-08 \*\*\*  
new\_tot\_crime\_rate\_DM 1.488e-01 4.468e-02 3.331 0.000866 \*\*\*  
new\_violent\_rate\_DM 6.383e-01 2.435e-01 2.621 0.008770 \*\*   
year 1.876e+01 6.250e+00 3.001 0.002694 \*\*   
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 3336 on 31346 degrees of freedom  
 (6289 observations deleted due to missingness)  
Multiple R-squared: 0.5594, Adjusted R-squared: 0.5593   
F-statistic: 4975 on 8 and 31346 DF, p-value: < 2.2e-16

**Explanation -**

This is the best model which has only the significant predictor variables from previous model.

The R-squared value of the regression model is 0.5594 which indicates that 55.94% variability can be explained by the predictor variables. The F-value is 4975 with a corresponding value of < 0.0001 which is less than the significance level of 0.001. This means that the model is statistically significant.

All predictor variables have high t-values with corresponding p-values less than the significance level of 0.01.

Population\_DM (County population), prop\_nonwhite\_DM (proportion of nonwhite people), new\_tot\_crime\_rate (new total crime rate), new\_violent\_rate\_DM (new violent crime rates), year have a positive impact on the county\_enroll\_DM (school enrollment) which means they lead to increase in school enrollment.

On the other hand, prop\_age1524\_DM (proportion of people aged 15-24 years), prop\_male\_DM (proportion of males), total\_crime\_rate\_old\_DM (old total crime rate) have negative impact on the county\_enroll\_DM (school enrollment) which means they lead to decrease in school enrollment.

population\_DM: Strong positive relationship. For every unit increase in population, enrollment increases by 0.2364.

prop\_age1524\_DM: Strong negative relationship. Suggests fewer enrollments in areas with a higher proportion of the population aged 15-24.

prop\_nonwhite\_DM: Strong positive relationship, suggesting a higher proportion of nonwhite individuals correlates with increased enrollment.

prop\_male\_DM: Negative relationship, indicating lower enrollment in counties with a higher proportion of males.

total\_crime\_rate\_old\_DM: Small but significant negative effect.

new\_tot\_crime\_rate\_DM: Small but significant positive effect.

new\_violent\_rate\_DM: Small positive relationship.

year: Enrollment increases slightly over time, with an estimated effect of approximately 18.76 per year.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

> bestmodel <- lm(county\_enroll\_DM ~ prop\_age1524\_DM + prop\_nonwhite\_DM + prop\_male\_DM +  
+ total\_crime\_rate\_old\_DM + new\_violent\_rate\_DM +  
+ percap\_income\_DM + employment\_rate\_DM + year, data = school\_dm\_final)  
> summary(bestmodel)  
  
Call:  
lm(formula = county\_enroll\_DM ~ prop\_age1524\_DM + prop\_nonwhite\_DM +   
 prop\_male\_DM + total\_crime\_rate\_old\_DM + new\_violent\_rate\_DM +   
 percap\_income\_DM + employment\_rate\_DM + year, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-154276 -735 164 903 169698   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 4.572e+05 3.898e+04 11.728 < 2e-16 \*\*\*  
prop\_age1524\_DM -7.848e+04 4.412e+03 -17.789 < 2e-16 \*\*\*  
prop\_nonwhite\_DM 1.292e+05 3.444e+03 37.520 < 2e-16 \*\*\*  
prop\_male\_DM -4.784e+04 5.472e+03 -8.743 < 2e-16 \*\*\*  
total\_crime\_rate\_old\_DM -7.972e-02 1.185e-02 -6.726 1.77e-11 \*\*\*  
new\_violent\_rate\_DM 1.505e+00 2.736e-01 5.502 3.79e-08 \*\*\*  
percap\_income\_DM 5.990e-01 2.251e-02 26.617 < 2e-16 \*\*\*  
employment\_rate\_DM -6.943e+00 1.058e+00 -6.560 5.47e-11 \*\*\*  
year -2.291e+02 1.954e+01 -11.728 < 2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4614 on 30755 degrees of freedom  
 (6880 observations deleted due to missingness)  
Multiple R-squared: 0.13, Adjusted R-squared: 0.1298   
F-statistic: 574.7 on 8 and 30755 DF, p-value: < 2.2e-16

**Explanation -**

This model only has the best predictors (significant predictors) from the overall model.

The R-squared value of the regression model is 0.13 which indicates that 13% variability can be explained by the predictor variables. The F-value is 574.7 with a corresponding value of < 0.0001 which is less than the significance level of 0.01. This means that the model is statistically significant.

All the predictor variables (first-differenced) are statistically significant with a significance level of 0.001. Moreover, the t-value for all the predictor variables is also more than 2 which also indicates that the results are statistically significant.

Predictor variables: prop\_nonwhite\_DM (proportion of nonwhite people), new\_violent\_rate\_DM (new violent crime rates), percap\_income\_DM (percap income of the people) have a positive impact on the county\_enroll\_DM (school enrollment). On the other hand, prop\_age1524\_DM (proportion of people aged 15-24 years), prop\_male\_DM (proportion of males), total\_crime\_rate\_old\_DM (old total crime rate), employment\_rate\_DM (employment rates), year have negative impact on the county\_enroll\_DM (school enrollment).

prop\_age1524\_DM: Strong negative effect (*−78,480-*). Counties with a higher proportion of individuals aged 15-24 tend to have lower enrollment.

prop\_nonwhite\_DM: Strong positive effect (*129,200)*. Counties with a higher proportion of nonwhite individuals are associated with higher enrollment.

prop\_male\_DM: Negative effect (*−47,840*). Higher male population proportions correspond to lower enrollment.

total\_crime\_rate\_old\_DM: Negative but small effect (*−0.0797*).

new\_violent\_rate\_DM: Positive but small effect (*1.505*).

percap\_income\_DM: Positive effect (*0.599*). Wealthier counties tend to have slightly higher enrollment.

employment\_rate\_DM: Negative effect (*−6.943*), suggesting that higher employment rates may be associated with lower enrollment.

year: Negative effect (*−229.1*), counterintuitive given earlier models with a positive trend.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Regression with individual predictors:**

1. **Year**

> model\_year <- lm(county\_enroll\_DM ~ year, data = school\_dm\_final)  
> summary(model\_year)  
  
Call:  
lm(formula = county\_enroll\_DM ~ year, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-469531 -640 125 755 1432561   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -493619.39 49243.13 -10.02 <2e-16 \*\*\*  
year 247.37 24.68 10.02 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 16530 on 37642 degrees of freedom  
Multiple R-squared: 0.002662, Adjusted R-squared: 0.002636   
F-statistic: 100.5 on 1 and 37642 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (247.37): Indicates a small positive trend in county\_enroll\_DM over time, with an average increase of 247.37 units per year.

R² (0.00266): Year alone explains a negligible portion of the variation in county\_enroll\_DM.

Significance: Highly significant (p < 2e-16), but the practical relevance is minimal due to the very low R².

1. **Population**

> model\_population <- lm(county\_enroll\_DM ~ population\_DM, data = school\_dm\_final)  
> summary(model\_population)  
  
Call:  
lm(formula = county\_enroll\_DM ~ population\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-129982 -151 35 292 63572   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -3.427e-14 1.719e+01 0.0 1   
population\_DM 2.419e-01 1.222e-03 197.9 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 3332 on 37568 degrees of freedom  
 (74 observations deleted due to missingness)  
Multiple R-squared: 0.5104, Adjusted R-squared: 0.5104   
F-statistic: 3.916e+04 on 1 and 37568 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (0.2419): For every unit increase in population\_DM, county\_enroll\_DM increases by approximately 0.24 units.

R² (0.5104): Population is a strong predictor, explaining ~51% of the variability in county\_enroll\_DM.

Significance: Highly significant, suggesting a meaningful linear relationship.

1. **Prop\_Age1524**

> model\_prop\_age1524 <- lm(county\_enroll\_DM ~ prop\_age1524\_DM, data = school\_dm\_final)  
> summary(model\_prop\_age1524)  
  
Call:  
lm(formula = county\_enroll\_DM ~ prop\_age1524\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-156656 -360 37 488 172485   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -9.079e-14 2.446e+01 0.00 1   
prop\_age1524\_DM -7.442e+04 3.988e+03 -18.66 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4740 on 37568 degrees of freedom  
 (74 observations deleted due to missingness)  
Multiple R-squared: 0.009183, Adjusted R-squared: 0.009157   
F-statistic: 348.2 on 1 and 37568 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (-74420): A negative association indicates that higher proportions of this age group are linked to lower county\_enroll\_DM.

R² (0.00918): Explains only ~0.9% of the variability, meaning other factors are more influential.

Significance: Statistically significant, but the low R² suggests limited practical relevance.

1. **Prop\_nonwhite**

> model\_prop\_nonwhite <- lm(county\_enroll\_DM ~ prop\_nonwhite\_DM, data = school\_dm\_final)  
> summary(model\_prop\_nonwhite)  
  
Call:  
lm(formula = county\_enroll\_DM ~ prop\_nonwhite\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-156642 -427 139 555 172263   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -1.203e-13 2.376e+01 0.00 1   
prop\_nonwhite\_DM 1.415e+05 2.773e+03 51.03 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4605 on 37568 degrees of freedom  
 (74 observations deleted due to missingness)  
Multiple R-squared: 0.06483, Adjusted R-squared: 0.06481   
F-statistic: 2604 on 1 and 37568 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (141500): A higher proportion of nonwhite individuals is associated with higher county\_enroll\_DM.

R² (0.06483): Explains ~6.5% of the variability in county\_enroll\_DM.

Significance: Significant, indicating a measurable but moderate effect.

1. **Prop\_male**

> model\_prop\_male <- lm(county\_enroll\_DM ~ prop\_male\_DM, data = school\_dm\_final)  
> summary(model\_prop\_male)  
  
Call:  
lm(formula = county\_enroll\_DM ~ prop\_male\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-157959 -175 46 325 172753   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -1.035e-13 2.456e+01 0.00 1   
prop\_male\_DM 2.206e+04 4.635e+03 4.76 1.94e-06 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4761 on 37568 degrees of freedom  
 (74 observations deleted due to missingness)  
Multiple R-squared: 0.0006027, Adjusted R-squared: 0.0005761   
F-statistic: 22.66 on 1 and 37568 DF, p-value: 1.944e-06

**Explanation -**

Coefficient (22060): Indicates a positive relationship between the proportion of males and county\_enroll\_DM.

R² (0.0006): The variability explained is extremely small (0.06%).

Significance: Statistically significant (p < 0.001), but with very little explanatory power.

1. **Total\_crime\_rate\_old**

> model\_total\_crime\_rate\_old <- lm(county\_enroll\_DM ~ total\_crime\_rate\_old\_DM, data = school\_dm\_final)  
> summary(model\_total\_crime\_rate\_old)  
  
Call:  
lm(formula = county\_enroll\_DM ~ total\_crime\_rate\_old\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-157653 -204 58 399 172375   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -14.34014 28.52211 -0.503 0.615   
total\_crime\_rate\_old\_DM -0.08802 0.01281 -6.870 6.55e-12 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 5052 on 31367 degrees of freedom  
 (6275 observations deleted due to missingness)  
Multiple R-squared: 0.001502, Adjusted R-squared: 0.001471   
F-statistic: 47.19 on 1 and 31367 DF, p-value: 6.547e-12

**Explanation -**

Coefficient (-0.08802): Suggests a slight negative relationship between crime rate and county\_enroll\_DM.

R² (0.0015): Explains only 0.15% of the variation, indicating a weak effect.

Significance: Statistically significant, but not practically impactful.

1. **New\_tot\_crime\_rate**

> model\_new\_tot\_crime\_rate <- lm(county\_enroll\_DM ~ new\_tot\_crime\_rate\_DM, data = school\_dm\_final)  
> summary(model\_new\_tot\_crime\_rate)  
  
Call:  
lm(formula = county\_enroll\_DM ~ new\_tot\_crime\_rate\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-157886 -198 61 389 172726   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -14.34654 28.37905 -0.506 0.61319   
new\_tot\_crime\_rate\_DM -0.16072 0.05074 -3.168 0.00154 \*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 5025 on 31353 degrees of freedom  
 (6289 observations deleted due to missingness)  
Multiple R-squared: 0.0003199, Adjusted R-squared: 0.000288   
F-statistic: 10.03 on 1 and 31353 DF, p-value: 0.001539

**Explanation -**

Coefficient (-0.16072): Similar negative relationship as the old crime rate, with slightly stronger magnitude.

R² (0.00032): Explains even less variability than the old crime rate (0.03%).

Significance: Statistically significant but minimal effect size.

1. **Violent\_rate\_old**

> model\_violent\_rate\_old <- lm(county\_enroll\_DM ~ violent\_rate\_old\_DM, data = school\_dm\_final)  
> summary(model\_violent\_rate\_old)  
  
Call:  
lm(formula = county\_enroll\_DM ~ violent\_rate\_old\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-157998 -227 52 419 172137   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -14.3401 28.5185 -0.503 0.615   
violent\_rate\_old\_DM -1.0102 0.1361 -7.420 1.2e-13 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 5051 on 31367 degrees of freedom  
 (6275 observations deleted due to missingness)  
Multiple R-squared: 0.001752, Adjusted R-squared: 0.001721   
F-statistic: 55.06 on 1 and 31367 DF, p-value: 1.197e-13

**Explanation -**

Coefficient (-1.0102): Indicates a stronger negative relationship compared to total crime rates.

R² (0.00175): Explains 0.18% of the variability, still weak.

Significance: Statistically significant but limited explanatory power.

1. **New\_violent\_rate**

> model\_new\_violent\_rate <- lm(county\_enroll\_DM ~ new\_violent\_rate\_DM, data = school\_dm\_final)  
> summary(model\_new\_violent\_rate)  
  
Call:  
lm(formula = county\_enroll\_DM ~ new\_violent\_rate\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-157888 -185 54 380 172727   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
(Intercept) -14.3465 28.3828 -0.505 0.613  
new\_violent\_rate\_DM -0.3805 0.2874 -1.324 0.186  
  
Residual standard error: 5026 on 31353 degrees of freedom  
 (6289 observations deleted due to missingness)  
Multiple R-squared: 5.59e-05, Adjusted R-squared: 2.401e-05   
F-statistic: 1.753 on 1 and 31353 DF, p-value: 0.1855

**Explanation -**

Coefficient (-0.3805): Weak negative relationship, not statistically significant (p = 0.186).

R² (0.000056): Near-zero explanatory power.

1. **Percap\_income**

> model\_percap\_income <- lm(county\_enroll\_DM ~ percap\_income\_DM, data = school\_dm\_final)  
> summary(model\_percap\_income)  
  
Call:  
lm(formula = county\_enroll\_DM ~ percap\_income\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-156687 -753 181 877 170527   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.006272 23.579677 0.00 1   
percap\_income\_DM 0.398653 0.007853 50.76 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 4533 on 36959 degrees of freedom  
 (683 observations deleted due to missingness)  
Multiple R-squared: 0.06518, Adjusted R-squared: 0.06515   
F-statistic: 2577 on 1 and 36959 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (0.3987): A small positive association between income and county\_enroll\_DM.

R² (0.06518): Explains ~6.5% of variability, moderate compared to other predictors.

Significance: Highly significant, indicating per capita income is moderately relevant.

1. **Employment\_rate**

> model\_employment\_rate <- lm(county\_enroll\_DM ~ employment\_DM, data = school\_dm\_final)  
> summary(model\_employment\_rate)  
  
Call:  
lm(formula = county\_enroll\_DM ~ employment\_DM, data = school\_dm\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-196065 -164 36 290 90065   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.006272 18.638723 0.0 1   
employment\_DM 0.239558 0.001477 162.2 <2e-16 \*\*\*  
---  
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
Residual standard error: 3583 on 36959 degrees of freedom  
 (683 observations deleted due to missingness)  
Multiple R-squared: 0.4159, Adjusted R-squared: 0.4159   
F-statistic: 2.632e+04 on 1 and 36959 DF, p-value: < 2.2e-16

**Explanation -**

Coefficient (0.2396): Positive association between employment and county\_enroll\_DM.

R² (0.4159): Explains ~41.6% of variability, second only to population.

Significance: Highly significant, making employment one of the most impactful predictors.