# Regression Analysis on School Absenteeism

## a. Regression Setup and Output in JASP

A multiple linear regression analysis was conducted in JASP to examine how student characteristics predict school absenteeism. The dependent variable was the number of days absent from school. The explanatory variables (entered as factors) were:

• eth: Ethnicity (0 = Aboriginal, 1 = Not Aboriginal)

• sex: Gender (0 = Female, 1 = Male)

• lnr: Learning rate (0 = Average learner, 1 = Slow learner)

JASP Settings:  
- Menu: Regression → Classical → Linear Regression  
- Dependent Variable: days  
- Factors: eth, sex, lnr

Model Summary:

|  |  |
| --- | --- |
| Metric | Value |
| R | 0.299 |
| R² | 0.089 |
| Adjusted R² | 0.070 |
| RMSE | 15.673 |
| F(3, 142) | 4.643 |
| p-value (ANOVA) | 0.004 |

## b. Regression Model Equation

The regression model based on unstandardized coefficients is:  
Ŷ = 18.932 + 2.154(lnr) - 9.112(eth) + 3.104(sex)  
  
Where:  
- Ŷ: Predicted number of days absent  
- lnr: 1 for slow learner, 0 for average learner  
- eth: 1 for not aboriginal, 0 for aboriginal  
- sex: 1 for male, 0 for female

## c. Interpretation of Slopes and Significance

Interpretation and statistical significance of each slope:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient (B) | Interpretation | p-value | Significant at α = 0.05? |
| lnr | 2.154 | Slow learners are predicted to be absent ~2.15 more days than average learners | 0.418 | No |
| eth | -9.112 | Not aboriginal students are predicted to be absent ~9.11 fewer days than aboriginal students | <0.001 | Yes |
| sex | 3.104 | Male students are predicted to be absent ~3.10 more days than female students | 0.241 | No |

## d. Residual for the Second Observation

To calculate the residual for the second observation, we use the regression equation:  
Ŷ = 18.932 + 2.154(lnr) - 9.112(eth) + 3.104(sex)  
For the second observation:  
- days (observed) = 6  
- lnr = 1 (Slow Learner)  
- eth = 0 (Aboriginal)  
- sex = 1 (Male)  
  
Predicted Value (Ŷ) = 18.932 + 2.154(1) - 9.112(0) + 3.104(1) = 24.19  
Residual = Observed - Predicted = 6 - 24.19 = -18.19  
  
The residual for the second observation is -18.19, indicating that the model overpredicted absenteeism for this student by 18.19 days.

## e. Adjusted R-Squared Value and Interpretation

The adjusted R-squared value from the regression output is 0.070.  
  
This means that approximately 7.0% of the variance in the number of absenteeism days is explained by the predictors in the model (ethnicity, sex, and learning rate), after accounting for the number of predictors.  
  
This relatively low adjusted R-squared suggests that the model has limited explanatory power and that other variables not included in the model may significantly influence absenteeism.