

**GPA Modelling Exercise**

Document Version: 1

Date: Sep 2019

GPA HK

**DBS Bank Limited**

**Topic:**

Please build a regression model using the dataset attached. The dataset is randomly sampled from elementary schools in a region . Please build a model to explore possible factors affecting the academic performance. Please refer to the appendix for the explanations of each data field.

Main Steps：

1. Select model factors
2. Run multivariate regression model
3. Test the classical problems in the regression model: Multicollinearity, Heteroscedasticity and Auto-correlation
4. Provide remedies to the problems identified in the model

Please select at least one method for each category from the below table. On top of that, feel free to try more methods not listed in the table.

|  |  |  |
| --- | --- | --- |
| **Test Methods** | Multicollinearity | VIF; Auxiliary Regression； Partial correlationcoefficient； |
| Heteroscedasticity | Residual Normality； White Test; Park Test |
| Auto-Correlation | Time Series Plot; D-W Test |
| **Remedies** | Multicollinearity | Factor Analysis; Principal Component Analysis |
| Heteroscedasticity | WLS; White |
| Auto-Correlation | NW; GLS |

**Appendix：**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Variable Label** | **Description** |
| snum | school number | School code |
| dnum | district number | Area code |
| api00 | api2000 | School 2000 Academic Performance |
| api99 | api1999 | School 1999 Academic Performance |
| growth | growth 1999 to 2000 | School performance changed from 1999 to 2000 |
| meals | pct free meal | The school provides a percentage of free lunches for students |
| ell | english language learner | Number of students learning English |
| yr\_rnd | Year around school | School type code |
| mobility | pct 1st year in school | Percentage of students in their first year of school |
| acs\_k3 | avg class size k-3 | Average student size in grades K-3 |
| acs\_46 | avg class size 4-6 | Average number of students in grades 4-6 |
| not\_hsg | parent not hsg | The rate of parents who did not attend high school |
| hsg | parent hsg | The rate of parents attending high school |
| some\_col | parent some college | The percentage of students whose parents attended college |
| avg\_ed | avg parent ed | Average parental education |
| full | pct full credential | Percentage of teachers with full-time teacher certification |
| emer | pct emer credential | Percentage of teachers with a substitute teacher certificate |
| enroll | number of students | Number of students |