

Vietnamese student's learning habit due to Covid-19 pandemic

Learn more at ndcuong869.github.io/learning_habit

This report represent how to analyst learning habit of Vietnamese students during in the pandemic and determine which factors affect to their behaviors.

My contribute only is data analysis from a statistics perspective in existed dataset.

1

Data collecting

The dataset was collected from 460 students in Hanoi from 7th to 28th February by snowball sampling method. Read full dataset description in bit.ly/vietnamese_dataset

2

Data cleaning

The dataset has 460 responses, however, only 420 responses is valid by the authors. The cleaned dataset was used for next stage without any modify steps.

3

Data mining

Represent the result after analysis stage to provide an insight from student learning habit by different approaches such as hypothesis testing, proportion estimation, linear regression, etc.

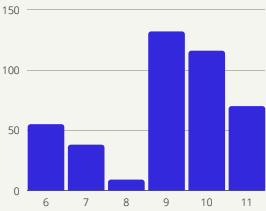
Table of contents

- 1 Dataset overview
- 2 Learning habit overview
- 3 Learning habit trend
- 4 Learning habit by genders
- 5 Linear regression analysis

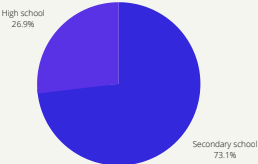
Dataset overview

The dataset has 420 responses with 40 features include 6 nominal data, 25 ordinal data and 9 numerical data.

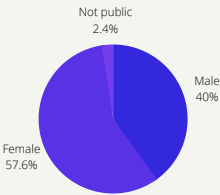
Class



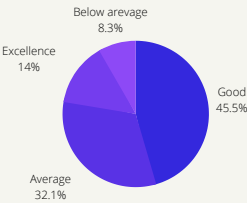
Grade



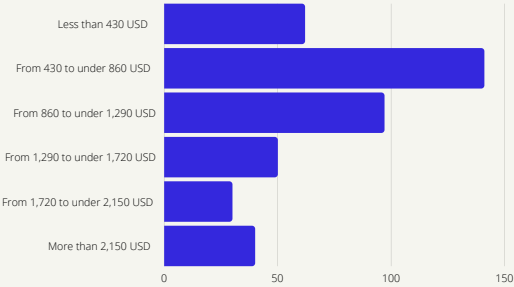
Gender



English



Income



Learning habit overview

On average, each student has spent



2.43 - 3.04 hours
for online learning



2.67 - 3.2 hours
for offline learning

How-to

To estimate population's learning hours, a confidence interval value should be used with significant level. This result is calculated by using bootstrap approach.

Comments

Although the students couldn't go to school in during the pandemic, the learning hours in offline program is higher than online. Obviously, this is explained that the students not familiar with new study method as online learning when it started only 2 weeks.

Learning habit trend

Did the students change self-learning habit when the pandemic began?

	Before	During
Under 4h	74.29%	66.03% ↓
From 4 to 7h	22.14%	30.03% ↑
Over 7h	3.57%	3.85% ↑

Table 3.1 Different learning hours between before and during the Covid-19 pandemic

33.8 - 43.1 %

students changed their learning habit

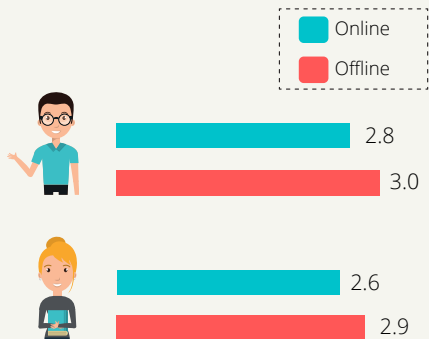
How-to

Using 2 categorical variables which present learning hours before and during the pandemic to calculate different value and its confident interval by bootstrap again.

Comments

While the tables 3.1 didn't provide a clear view about student learning habit changing, the percentage illustrate with over 30% students changed their habit.

Learning habit by genders



- 1 There is no difference between male and female learning hours.
- 2 While learning with instruction, the male student learn more female. However, without the lecture, the female student spend time to learn than other.

How-to

Using hypothesis testing to assess the different between male and female learning habit during the pandemic. Because the p-value larger than significant level, cannot reject H_0 .

Comments

This section showed that no difference between genders when they learn in your home instead of school in general. However, with or without instruction, they can spend more time to learn.

Linear regression analysis

Determine which factors affect to learning habit of Vietnamese student during the pandemic.

Model	R-Square	Adjust R-square	F statistics
Total online ~ School type + Income + English	0.435	0.431	1.82e-57
Total online ~ School type + Income + English + Gender	0.439	0.434	4.75e-51
Total online ~ School type + Income + English + Total offline	0.539	0.534	1.26e-67
Total online ~ School type + Income + English + Total offline ^ 2	0.55	0.546	7.72e-71

Table 5.1 Total online modeling by different predictor variables.

	coef	p-value
school_type	0.328	0.049
income	0.146	0.162
English	0.129	0.321
Total_offline	0.47	0.000

Table 5.2 Modeling with school type, income, English and total offline variables

	coef	p-value
school_type	0.4564	0.049
income	0.1375	0.162
English	0.3141	0.321
Total_offline_2	0.0371	0.000

Table 5.3 Modeling with school type, income, English and total offline power 2 variables.

Comments

Different linear regression models showed that coefficient of predictor variables. In 4th model, the total offline learning hours variable didn't affect to response variable and this is cannot be true. The 3rd model seem likely fit with assumption in above sections, however, some variables include income and English have a huge p-value, it's removed from model.