

# Assignment 6

## Reading data

Reading the data from the csv file, and do some cleaning, since we only need two variables for later analysis, we will only keep the following two variables: x1stuedexpct and x1paredexpct.

After cleaning, there are 16429 rows and 2 columns in the dataset.

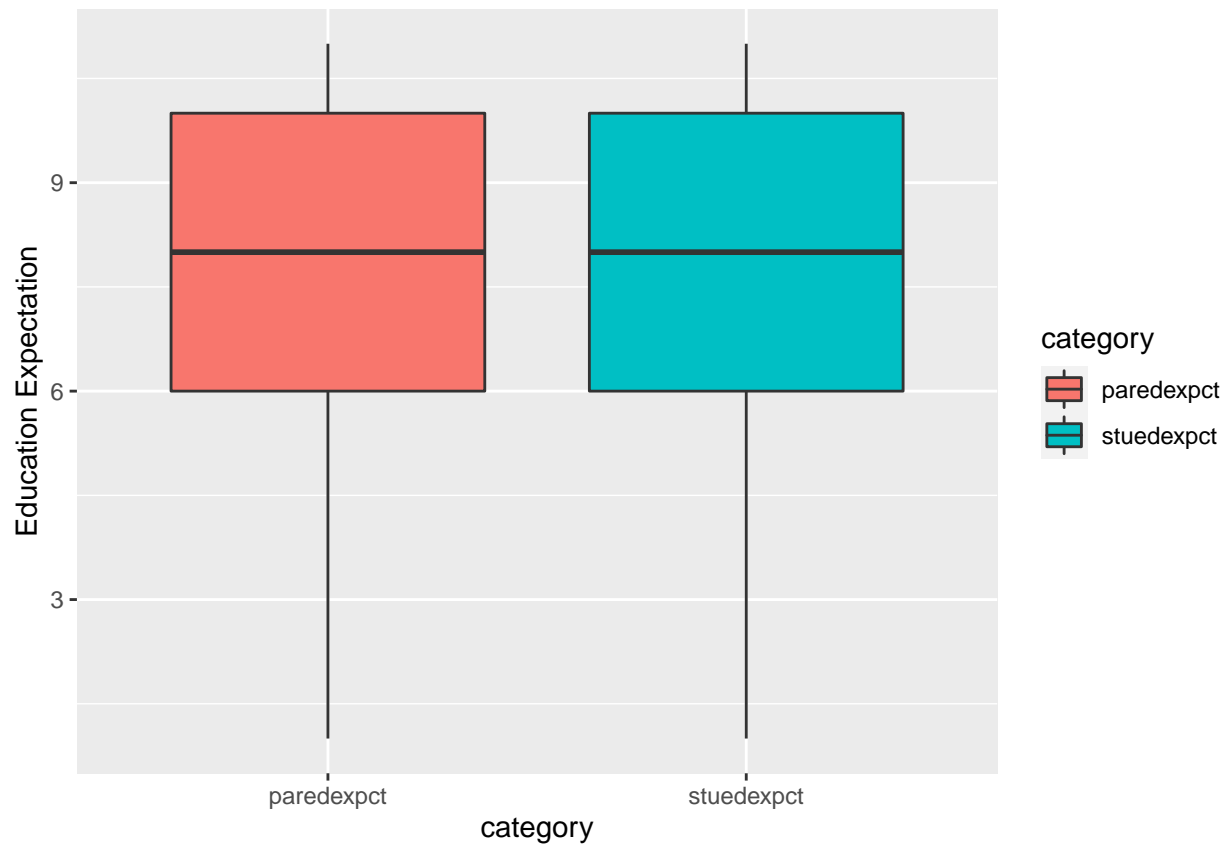
## Summary statistic

Table 1: Summary Statistic

category	n	mean	median	sd
paredexpct	16429	7.33	8	2.63
stuedexpct	16429	7.78	8	2.91

Calculate the summary statistic of the education expectation for parents and students, the mean education expectation of the parents is 7.33, the mean education expectation of the students is 7.78, the standard deviation of the education expectation for parents is 2.63, the standard deviation of the education expectation for students is 2.91.

## Visualization



The boxplot suggests that there is no difference in the education expectation between parents and students, the median values are quite similar.

## Hypothesis

State the hypothesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

Welch Two Sample t-test

```
data: df$stuedexpct and df$paredexpct
t = 14.801, df = 32516, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.3928336 0.5127602
sample estimates:
mean of x mean of y
 7.784588  7.331791
```

Since the p-value is less than 0.05, we have sufficient evidence to reject the null hypothesis.

## **Conclusion**

The hypothesis suggests that there is a significant difference in the average education expectation for parents and students.