TOPIC: RELATIONSHIP BETWEEN EDUCATION AND ADOLESCENT FERTILITY RATE

TAI WEN JUN

A19EC0207

LECTURER: DR CHAN WENG HOWE

PROBABILITY AND STATISTICAL DATA ANALYSIS (SECI2143-06)

Introduction

Develop countries such as America is facing a problem which is high adolescent fertility rate which is among age 15-19 which is different from other developed country such as Japan and German. Although that the America have the top university in the globe however the enrolment rate of secondary education is relative low compare to other developed country due to the society culture of America that support the adolescent drop out from the school to achieve their dream compare to develop country like Japan where most parents will guide their child to complete their university or at least secondary education. Adolescent fertility rate is an issue that should be concerned around the globe. Child bearing in early age have dozen of disadvantages for the teenage girl no to mention the early birth of adolescent will result in the mother to take the responsibility of taking care of the child instead of enjoying her early hood, she might able to pursue her study and open doors in greater career but have to take care her child while she can have it when she is success or stable in her career. Furthermore, the mother will also face chance of dying in childbearing twice as much as the women that birth after her 20. Singapore is consider as a developed country that have advanced education system and also a country that people are safe and protected by the law. In this study, we will be using the data in Singapore to interpret the relationship between secondary education level and the adolescent fertility rate. This study will be using the data of Singapore secondary education enrolment rate from the official Singapore open data and the fertility rate among adolescent between 15-19 from world bank open data. We will also study what is the effect of the education level of the citizen and the expenditure of the government in the education.

Focus on Topic

We will using the aid of rstudio to visualize the data in graph for presentation and calculate the test-statistic and so on to help on the test.

1. 1 sample test

The purpose of this analysis is to test whether the mean of the adolescent fertility rate in Singapore is lower than the average of the adolescent fertility rate in the globe because Singapore is rated as developed country with high gross domestic product (GPD) and advanced law and famous in low corruption and so is its education system that famous in the globe. This test will proof whether this developed country is the country that pull down the average of adolescent fertility rate in the globe.

2. Chi-square test on dependency

The purpose of this analysis is to test whether is there any dependency between the female secondary education enrolment rate and male secondary education enrolment rate. If there is any reason such as the male will persuade the female to avoid enroll in secondary education or the male will have priority in enrolling in education due to the culture, the data of both female and male secondary education enrolment has to be used in this study.

3. Correlation

The purpose of this analysis is to get the relationship between the adolescent fertility rate and female secondary education enrolment rate in the expression of correlation coefficient, r. This analysis will also test the significance of correlation to proof the existence of linear correlation.

4. Regression

The purpose of this analysis is to get the coefficient of determination and the linear equation of the relationship between the adolescent fertility rate and female secondary education enrolment rate in Singapore. This analysis will also test the significance of the regression to proof the existence of linear relationship.

Support on Topic

1. 1 sample test on mean

To determine is the seriousness of the adolescent fertility rate in Singapore, we will compare the mean of adolescent fertility rate of Singapore with the mean of adolescent fertility rate of the globe according to WHO. According to WHO, the average global fertility rate among adolescent which is age 15-19 from 2000 to 2008 is 47 per 1000 (Adolescent fertility, n.d.). We going to test is the mean of adolescent fertility of Singapore is lower than the claimed value of average adolescent fertility rate by WHO at 0.05 accuracy. The data we will be using in this 1 sample test is the data of Singapore fertility rate among adolescent per 1000 from year 2000-2008 (Gender Statistics, 2020).

Null hypothesis, H_0: $\mu = 47$

Alternate hypothesis, H_1: $\mu < 47$

By using rstudio we get the mean of the adolescent fertility rate of Singapore per 1000 is 6.786511 and the variance of the adolescent fertility rate of Singapore per 1000 is 0.5415614.

$$Z = \frac{\varkappa - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{6.786511 - 47}{\frac{0.5415614}{\sqrt{9}}} = -222.8 \quad (1.1)$$

Equation 1.1 is the formula for test statistic of population mean

From the t-table, the critical value $t_{0.05,8}$ = -1.860, the test statistic that we had calculated from the equation 1.1 is -222.8 which is lower than -1.860 so we reject the null hypothesis (How to Use the t-Table to Solve Statistics Problems, 2019). We can conclude that the mean of adolescent fertility rate of Singapore per 1000 from 2000 to 2008 is smaller than the average of global adolescent fertility rate. From the test shown that Singapore succeed in maintaining low adolescent fertility rate as a developed country.

2. Independency test using chi-square test

Since the secondary education level enrolment rate include both the data of female and male, this test will test is there dependency between the secondary education enrolment rate of female and secondary education enrolment rate of male at an accuracy of 0.05. The data will be used in this test is the data of the enrolment rate of secondary education in Singapore from 1990 to 2017 (Education, 2019).

Null hypothesis, H_0: variables are independent

Alternate hypothesis, H_1: variable are dependent

From the rstudio, chisq.test() function is used as chi-square test on the variables male secondary education enrolment rate and female secondary education enrolment rate.

```
> chisq.test(EducationFemaleMale)

        Pearson's Chi-squared test

data:  EducationFemaleMale
X-squared = 0.42366, df = 27, p-value = 1
```

Figure 2.1 chisq.test() from rstudio of secondary education enrolment of female and male

From the test shown, the p-value is greater than 0.05, we fail to reject the null-hypothesis. The female secondary education enrolment rate and male secondary education enrolment rate are independent. So we are going to use the data of female secondary education enrolment rate instead of using both female and male secondary education enrolment rate due to the enrolment rate of female is not effected or related to the male secondary education enrolment rate. Moreover, the child bearing is happen on the female instead of male, the male secondary education enrolment rate will not be used as it don't affect the female secondary education enrolment rate.

3. Correlation test using Pearson's method

In this test will going to test the correlation between the adolescent fertlity rate per 1000 and the female secondary education enrolment rate from year 1990 to 2017. This test will use the function *ggscatter* from rstudio that using the packages *ggplot2, ggpmisc and ggpubr* to calculate the correlation, r of the variable. *Ggscatter* is a function from rstudio that invented for the plotting of graph and analysis the coefficient of correlation value of the variables.

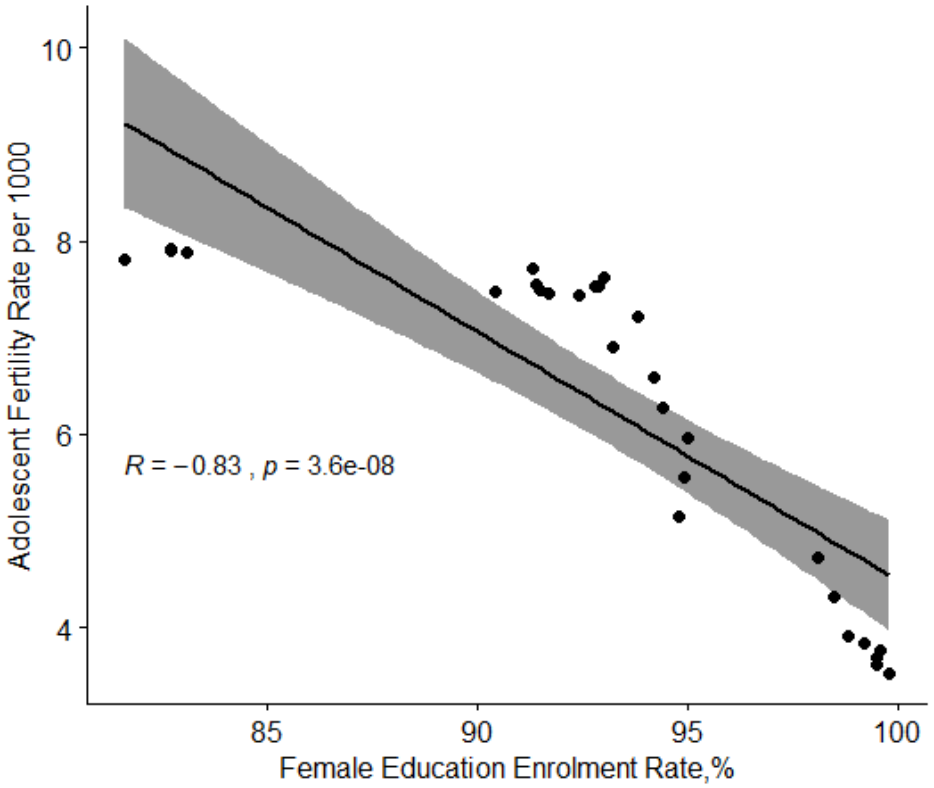

Figure 3.1 Comparison between the female secondary education enrolment rate and adolescent fertility rate per 1000

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\frac{\sum x^2 - (\sum x)^2}{n}\right]\left[\frac{\sum y^2 - (\sum y)^2}{n}\right]}} (3.1)$$

Equation 3.1 Pearson's Product-Moment Correlation Coefficient

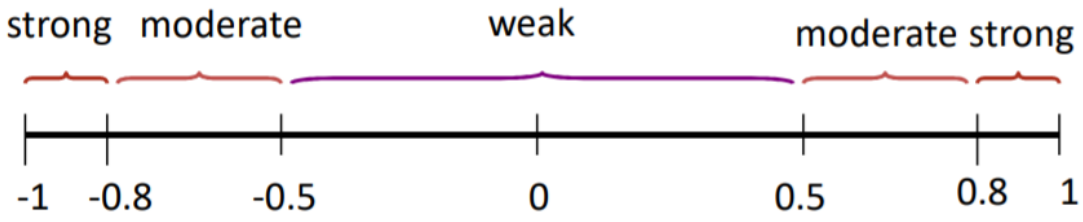

Figure 3.2 explanation of coefficient of correlation, r in term of strength

From the plotting, we can conclude that the coefficient of r=-0.83, the coefficient of r is negative which mean that when the female secondary education enrolment rate increase, the fertility rate among adolescent per 1000 decrease. From figure 1.3, the r value,-0.83 which is greater than -0.8 also shown a strong relationship between the variables. For further confirmation, we will do significance test for correlation with accuracy 0.05.

Null hypothesis, H_0: $\rho = 0$ (There is no linear correlation)

Alternate hypothesis, H_1: $\rho \neq 0$ (There is linear correlation)

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.83}{\sqrt{\frac{1-(-0.83)^2}{28-2}}} = -7.59 (3.2)$$

Equation 3.2 t-test statistic for significance of correlation

This test is using two-tailed test, the number of data used, n is 28, so the degree of freedom is n-2 = 28-2 = 26. For the critical value of this test is $t_{a/2, n-2}$ = 2.0555 and $-t_{a/2, n-2}$ = -2.0555. The test statistic calculated from equation 3.2 is -7.59 is lower than the critical value $-t_{a/2, n-2}$, -2.0555. This test success to reject the null hypotheses. There is enough evidence to proof there is linear correlation between the variable female secondary education enrolment rate and adolescent fertility rate.

4. Regression test

In this analysis, the independent variable is female secondary education enrolment rate and the dependent variable is fertility rate among adolescent. This analysis will test is there a linear relationship between female secondary education enrolment rate and fertility rate among adolescent at an accuracy of 0.05.

Null hypothesis, H_0: $\beta_1 = 0$ (There is no linear relationship)

Alternate hypothesis, H_1: $\beta_1 \neq 0$ (There is linear relationship)

```
Coefficients:
                                (Intercept)  DataCorrelation$secondaryEducationFemale
                                    30.1786                                  -0.2569
```

Figure 4.1 output from rstudio lm() function

In this test, the function lm() from rstudio is used on variable adolescent fertility rate and female secondary education enrolment rate to get the linear equation of the relationship of the variable, we get the linear equation of the relationship of female secondary education enrollment and the fertility rate among adolescent is:

$$y = -0.2569x+30.1786$$

Coefficient of determination is used to explain the variability can cause on a variable by another variable. Coefficient of determination can also indicated the percentage of perfect linear relationship between the variables such as $R^2 = 1$ indicate the variation of y is 100% explained by the variation of x. To calculate the coefficient of determination of our variable, we going to use equation 4.1 and get the result $R^2 = 0.6952848$.

$$R^2 = \frac{SSR(sum\ of\ squares\ regression)}{SST(total\ sum\ of\ squares)} \quad (4.1)$$

Equation 4.1 is the formula for calculating the coefficient of determination

$$s_\varepsilon = \sqrt{\frac{SSE(sum\ of\ squares\ error)}{n-k-1}} \textit{(4.2)}$$

Equation 4.2 is the formula for standard deviation of the variation of observation around the regression line

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} \textit{(4.3)}$$

Equation 4.3 is the formula for standard error of the regression slope coefficient (b_1).

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-0.83-0}{0.00094} = -883 \textit{(4.4)}$$

Equation 4.4 is the formula test statistic for the inference about the slope

To get the test statistic value, the value of standard deviation of the variation of observation around the regression line value, s_ε and the value of standard deviation of regression slope, s_{b_1} is first calculated using equation 4.2 and equation 4.3. The value of s_{b_1} after calculate is 0.00094. By referring to the t-statistic table, the critical value, $t_{a/2, 26}$ = 2.0555 and - $t_{a/2, 26}$ = -2.0555 (How to Use the t-Table to Solve Statistics Problems, 2019). The test statistic which calculated from the equation 4.4 is -883 which is lower than the critical value, -883 < -2.0555 so this analysis reject the null hypothesis. We can conclude that there is a linear relationship between the female secondary education enrolment rate and the adolescent fertility rate per 1000 in Singapore.

Addition

There is a suggestion that the secondary education complete rate is influenced by the amount of effort of the government put into it. We visualize the effort of government into numeric data of the total government expenditure on education. We going to test is there any correlation relationship between the total expenditure of government on education with the female secondary education enrolment rate. In this test will going to test the correlation between the expenditure of government on education and the female secondary education enrolment rate from year 1990 to 2017. This test also use the function *ggscatter* from rstudio that using the packages *ggplot2, ggpmisc and ggpubr* to calculate the correlation, r of the variables.

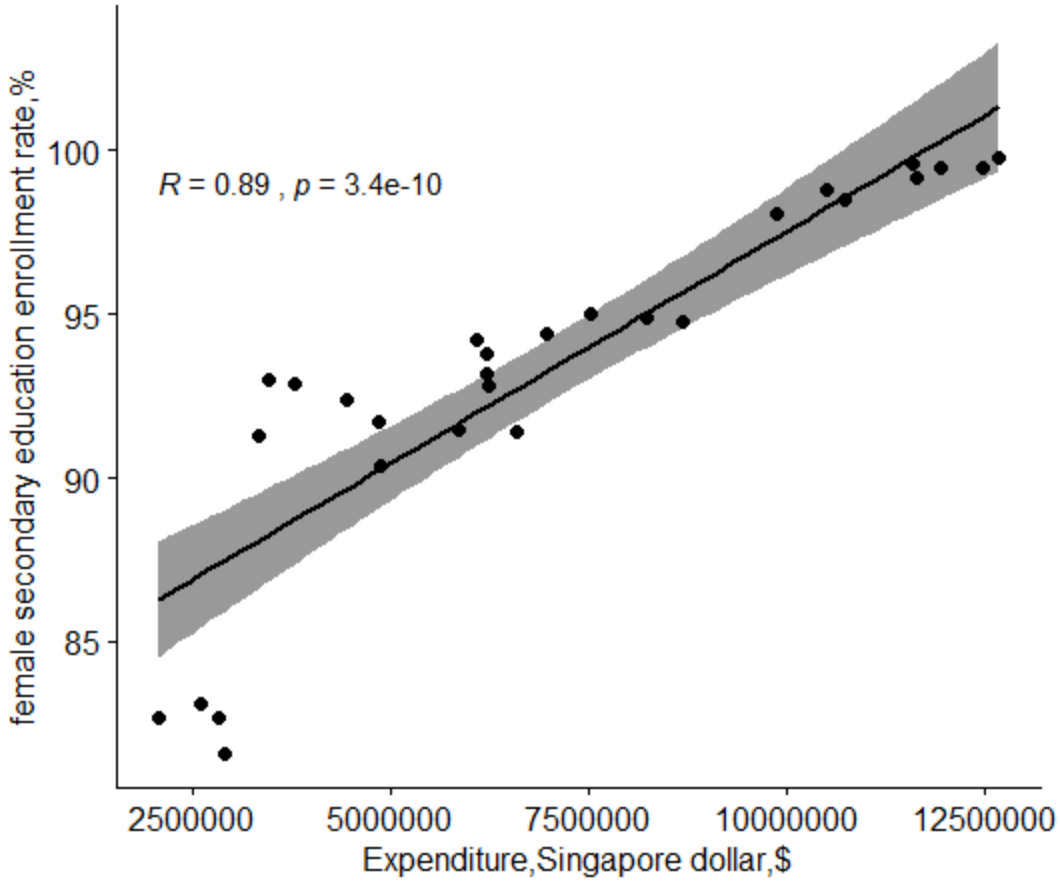

Figure 5.1 Comparison between the expenditure of government on education and female secondary education enrollment rate

From the plotting, we can conclude that the coefficient of r=0.89, the coefficient correlation, r is positive which mean that when the expenditure of government on education increase, the female secondary education enrolment rate also increase. From figure 3.2, the r value, 0.89 which greater than 0.8 also shown a strong relationship between the variables. For further confirmation, we will do significance test for correlation with accuracy 0.05.

Null hypothesis, H_0: $\rho = 0$ (There is no linear correlation)

Alternate hypothesis, H_1: $\rho \neq 0$ (There is linear correlation)

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.89}{\sqrt{\frac{1-(0.89)^2}{28-2}}} = 9.95$$

This test is two-tailed test, the number of data used, n is 28, so the degree of freedom is n-2 = 28-2 = 26. For the critical value of this test is $t_{a/2, n-2} = 2.0555$ and $-t_{a/2, n-2} = -2.0555$. The test statistic calculated from equation 3.2 is 9.95 higher than the critical value $t_{a/2, n-2}$, 2.0555. This test success to reject the null hypotheses. There is enough evidence to proof there is linear correlation between the expenditure of government on education and female secondary education enrolment rate.

The next analysis is the regression test. In this test, the independent variable is expenditure of government on education and the dependent variable is female secondary education enrolment rate. This analysis will test is there a linear relationship between expenditure of government on education and female secondary education enrolment rate at an accuracy of 0.05.

Null hypothesis, H_0: $\beta_1 = 0$ (There is no linear relationship)

Alternate hypothesis, H_1: $\beta_1 \neq 0$ (There is linear relationship)

```
Coefficients:
                                          (Intercept)
                                            8.338e+01
DataCorrelation$total_expenditure_on_education
                                            1.416e-06
```

Figure 4.1 output from rstudio lm() function

In this test, the function lm() from rstudio is used on variable government expenditure on education and female secondary education enrolment rate to get the linear equation of the relationship of the variable, we get the linear equation of the relationship of government expenditure on education and female secondary education enrollment is:

$$y = 1.416*10^{-6}x + 83.38$$

To calculate the coefficient of determination of our variable, we going to use equation 4.1 and get the result $R^2 = 0.7859923$. The variation of variation of female secondary education enrolment rate is 78% explained by the variation of government expenditure on education.

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{1.416 * 10^{-6} - 0}{1.033006 * 10^{-14}} = 1.4 * 10^8$$

To get the test statistic value, the value of standard deviation of the variation of observation around the regression line value, s_ε and the value of standard deviation of reg ression slope, s_{b_1} is first calculated using equation 4.2 and equation 4.3. The value of s_{b_1} after calculate is 1.033006e-14. By referring to the t-statistic table, the critical value, $t_{a/2, 26} = 2.0555$ and $- t_{a/2, 26} = -2.0555$ (How to Use the t-Table to Solve Statistics Problems, 2 019). The test statistic that calculated form equation 4.4 is $1.4*10^8$ is higher than the critical value, $1.4*10^8 > -2.0555$, so in this test reject the null hypothesis. We can conclude that there is a linear relationship between the secondary education complete rate and the adolescent fertility rate per 1000 in Singapore.

Conclusion

As a summary, Singapore is a country that succeed to maintain low adolescent fertility rate with lower mean of adolescent fertility per 1000 than the average from the globe. From the correlation and regression test, the analysis also conclude that there is linear correlation and linear relationship between the female secondary education enrolment rate and adolescent fertility rate. The adolescent fertility rate may affected by the female education enrolment rate because from a few aspect. One of the aspect is the female adolescent may acquire more information on the sexual knowledge in school including the disadvantages for them to child bearing in early age. More over there is an adult in school that watch over their activity during adolescent age instead of letting them unwatched which definitely increase the chance for the adolescent to explore their curiosity in sexuality. To increase the female education enrollment rate, the testing on the relationship of the expenditure of government on education does proof a positive impact on the fertility rate among adolescent. This is might due to a better advertisement of disadvantages of dropping from the school and better secondary school environment with cleaner and more comfortable classroom that allow the student to concentrate in their study encourage female adolescent to enroll in secondary education. To decrease the fertility rate among adolescent in the country, government from the globe should start to increase their expenditure and attention on education.

References

Adolescent fertility. (n.d.). Retrieved from World Health Organization: https://www.who.int/gho/mdg/maternal_health/adolescent_fertility_text/en/#:~:text=The%20adolescent%20fertility%20rate%20(births,over%20the%20period%202000%E2%80%932008.

Education, M. o. (2019, October 7). *Net Enrolment Ratio For Primary And Secondary Education*. Retrieved from Data.gov.sg: https://data.gov.sg/dataset/net-enrolment-ratio-for-primary-and-secondary-education?view_id=ab69441f-951a-4d99-9b31-735fa8530121&resource_id=7b184af5-b718-4c93-b217-c3bb3ab304f4

Gender Statistics. (2020, May 19). Retrieved from The World Bank: https://datacatalog.worldbank.org/dataset/gender-statistics

How to Use the t-Table to Solve Statistics Problems. (2019). Retrieved from Dummies: https://www.dummies.com/education/math/statistics/how-to-use-the-t-table-to-solve-statistics-problems/