SECI2143-07 Probability and Statistical Analysis

PROJECT 2 : INFERENTIAL STATISTICS

By : Spongebob & Jojo

Team Member :

Wong Li Jie A21EC0238

Ng Qian Hui A21EC0212

Toh Kang Lun A21EC0234

Ho Wei Chun A21EC0184

INTRODUCTION

- To investigate the relationship between a country's GDP per capita, brain drain rate and inflation rate
- To study the monthly average income and unemployment rate of different countries
- To learn how various economic environments can affect the performance of a country and its citizens

Graph Visualizations and Calculations :

Hypothesis Testing using one sample

Test statistic :

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

t = test statistic
$\bar{x}$ = sample mean
μ = population mean
s = standard deviation
n = number of observations

Hypothesis Testing using single sample:

Statement: The average monthly income is greater than USD 800.
Significance level, $\alpha = 0.05$
H_0 : μ = USD 800
H_1 : μ > USD 800

Degree of freedom = 10 - 1 = 9

$t_{0.05, 9} = 2.262$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{1292.02 - 800}{1486.06 / \sqrt{10}} = 1.046$$

- t = 1.046
- Smaller than critical value = 2.262
- Not enough evidence to support that average monthly income of 10 different countries is greater than USD 800

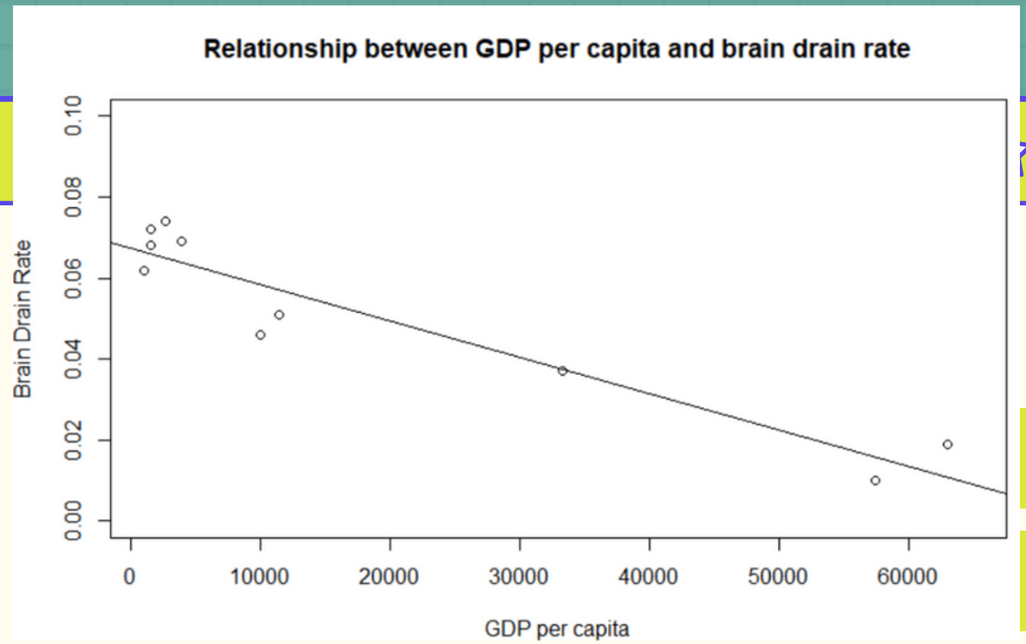
Fail to reject H0 since 1.046 < 2.262

Correlation

Sample correlation coefficient :

$$r = \frac{\Sigma xy - (\Sigma x\, \Sigma y)/n}{\sqrt{[(\Sigma x^2) - (\Sigma x)^2/n]\,[(\Sigma y^2) - (\Sigma y)^2/n]}}$$

r = sample correlation coefficient,
n = 10,
x = GDP per capita,
y = Brain drain rate.



- r = -0.9490472
- Negative relationship between the GDP per capita and brain drain rate
- Strong negative relationship, -1 < r < -0.8

Significance Test for Correlation :

Significance level, $\alpha = 0.05$

$H_0 : \rho = 0$ (no linear correlation)

$H_A : \rho \neq 0$ (linear correlation exists)

$$t_{0.025,\,8} = \pm 2.306$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.9490472}{\sqrt{\frac{1-(-0.9490472)^2}{10-2}}} = -8.51799$$

Reject H0 since -8.518 < -2.306

REGRESSION

The coefficient of determination:

$$R^2 = \frac{SSR}{SST} = \frac{sum\ of\ square\ explained\ by\ regression}{total\ sum\ of\ squares}$$

Estimated Regression Model:

Y = b0 + b1 X

Y = Estimated (or predicted) Y value
b0= Estimate of the regression intercept
b1= Estimate of the regression slope
X= Independent variable

Y = 39373 - 664411 x

R2 = 0.2647. Since, 0 <R2 < 1, shows weaker linear relationship between x and y

Test Statistical of Regression

$H_0 : \beta_1 = 0$ (non linear relationship)

$H_1 : \beta_2 \neq 0$ (linear relationship)

Test statistic, $t = \frac{b_1 - \beta_1}{S_{b_1}}$

$= -1.69699$

Degree of freedom = 10-2 =8

Where:

b_1 = Sample regression slope coefficient

β_1 = Hypothesized slope

S_{b_1} =Estimator of the standard error of the slope

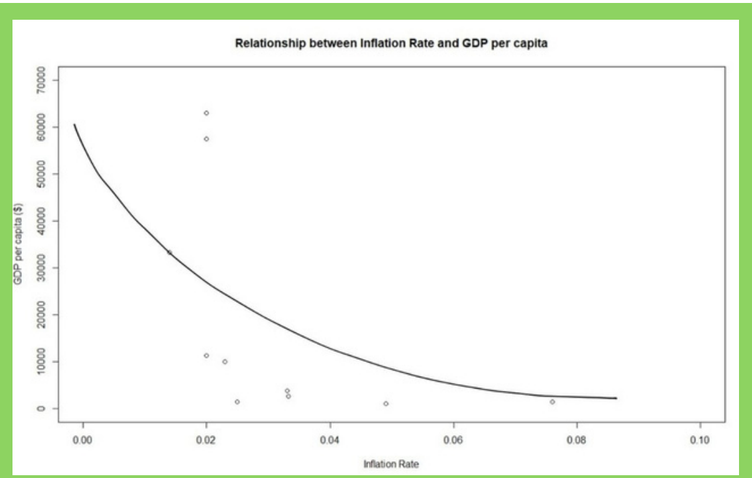
```
> summary(model)

Call:
lm(formula = y ~ x)

Residuals:
   Min     1Q Median     3Q    Max
-21251 -14542  -9666  10270  36912

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    39373      14072   2.798   0.0233 *
x            -664411     391524  -1.697   0.1281
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21830 on 8 degrees of freedom
Multiple R-squared:  0.2647,    Adjusted R-squared:  0.1728
F-statistic:  2.88 on 1 and 8 DF,  p-value: 0.1281
```



***Line is shwoing non-linear relationship

∴ Fail to reject H0. Since P-value is 0.1281 is more than significance level 0.05, there's sufficient evidence that a non-linear relationship exists between inflation rate and GDP per capita.

SAVE

Chi-Square Test of Independence

$$x^2 = \sum \frac{(O-E)^2}{E}$$

Test hypothesis:

H0: Unemployment rate is independent with the countries status

H1: Unemployment rate is dependent with the countries status

By using,

Degree of freedom= 4

Significance level= 0.05

Result:

Test statistic= 4.22222

Critical Value= 16.91898

```
> # output critical value
> print(x.alpha)
[1] 16.91898
> # output the chi-square value
> output$statistic
X-squared
 4.222222
> # output the parameter of degree of freedom
> output$parameter
df
 4
> # output the observed value table
> output$observed

      Developed Developing Least developed
0<x<2         0          0               2
2<x<4         2          1               3
4<x<6         1          1               0
> # output the expected value table
> output$expected

      Developed Developing Least developed
0<x<2       0.6        0.4               1
2<x<4       1.8        1.2               3
4<x<6       0.6        0.4               1
```

Since the test statistic value < critical value, we fail to reject the null hypothesis.

Therefore, there is sufficient evidence that status of countries and unemployment rate are dependent.

Conclusion

Prepare data set
-Fulfill the objective of performing data analysis test

Data pre-processing stage
-Data cleaning and data transforming

Hypothesis testing
-There is not sufficient evidence to prove that the average monthly income of 10 countries is greater than USD 800

Correlation test
-There is a linear relationship exists between GDP per capita and brain drain rate

Regression test
-There exists a non-linear relationship between inflation rate and GDP per capita

Chi-square test of Independence
-Conclude that the status of 10 countries and unemployment rate are dependent

THANK YOU

By Spongebob & Jojo