INFSCI 2725: Data Analytics

Assignment 2: Fundamental concepts from statistics

Yulu Meng

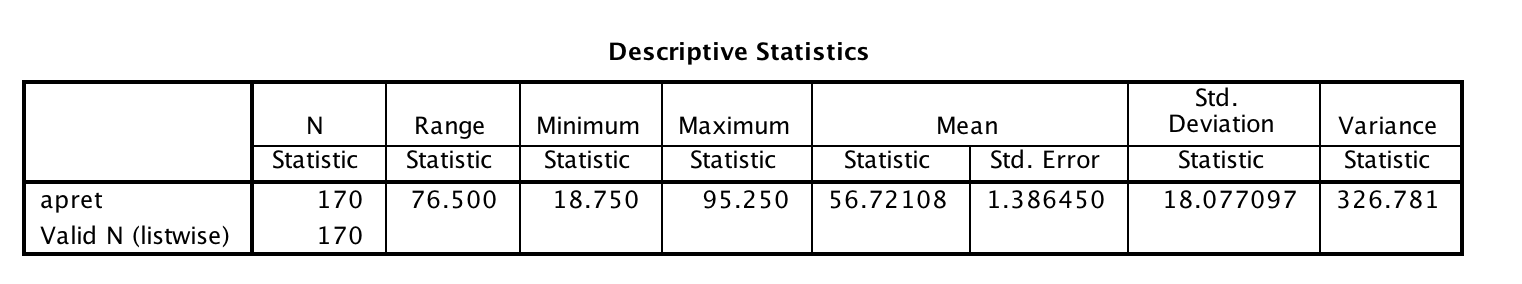
Zhaoxuan Ren

Tao Li

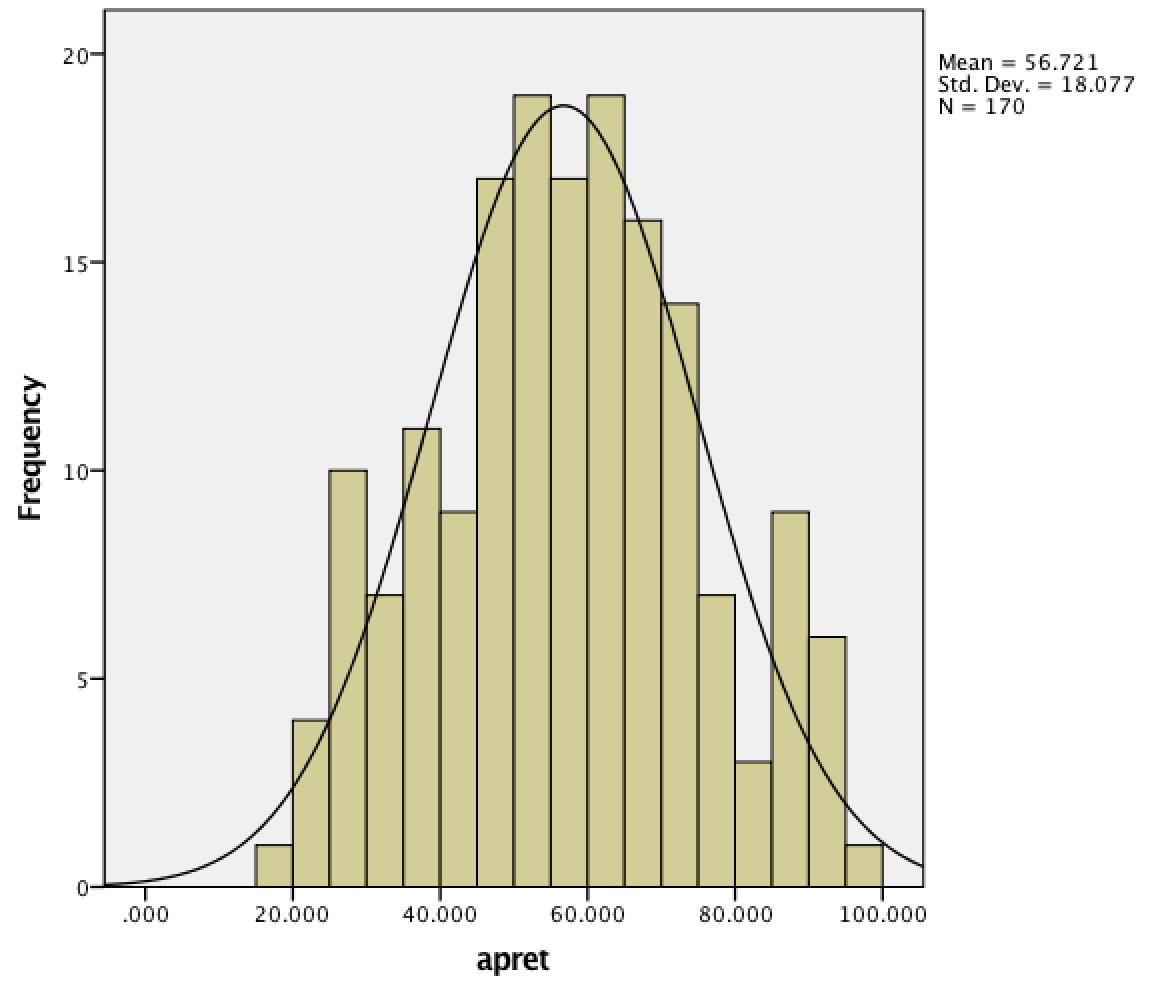
In this assignment 2, we use SPSS to analyze the given data.

1. Generate descriptive statistics and histograms for apret, tstsc, and salar.

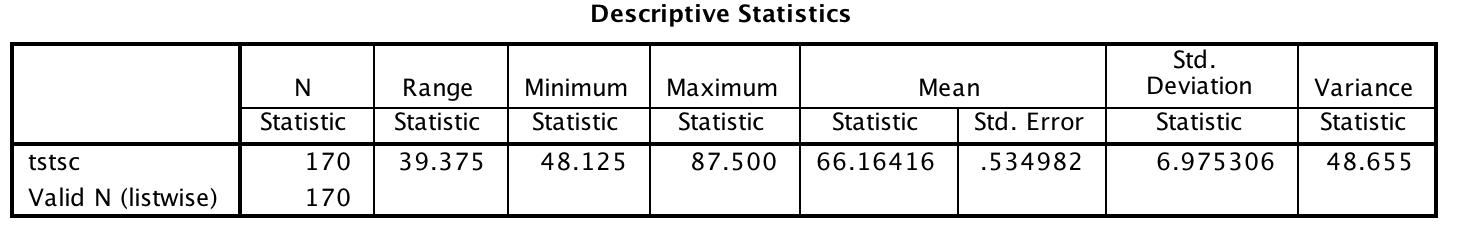
apret:



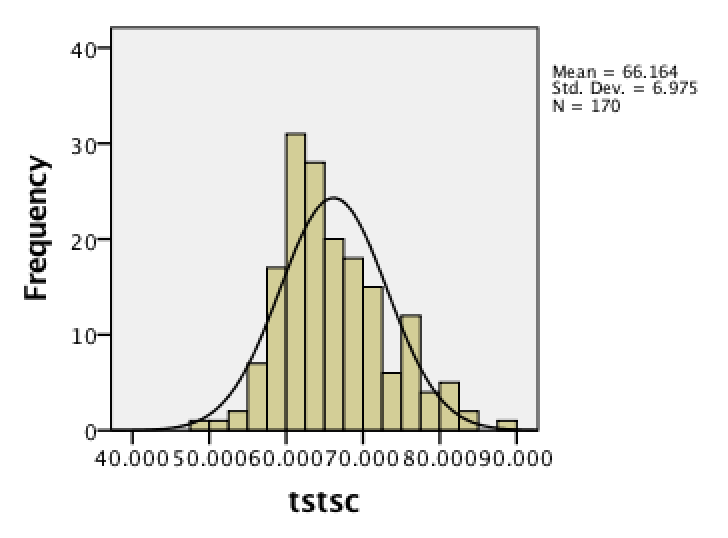
Histogram for apret



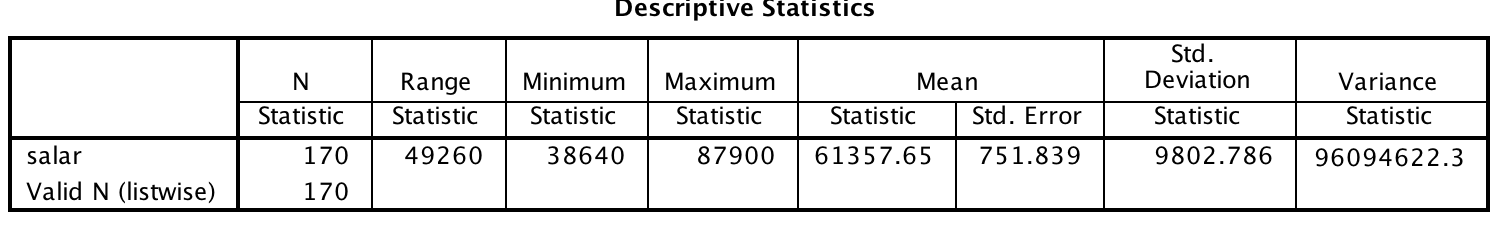
tstsc:



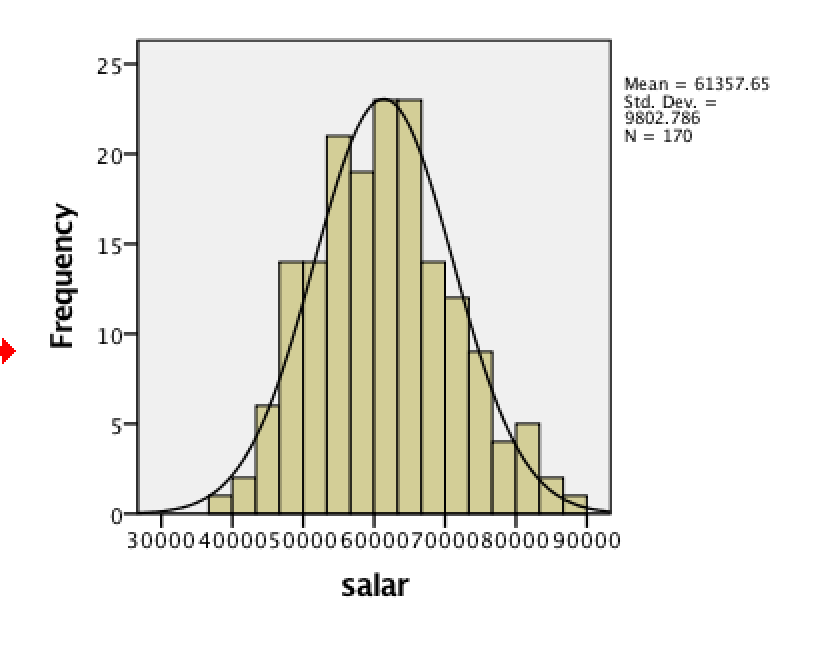
Histogram for tstsc



salar:



Histogram for tstsc

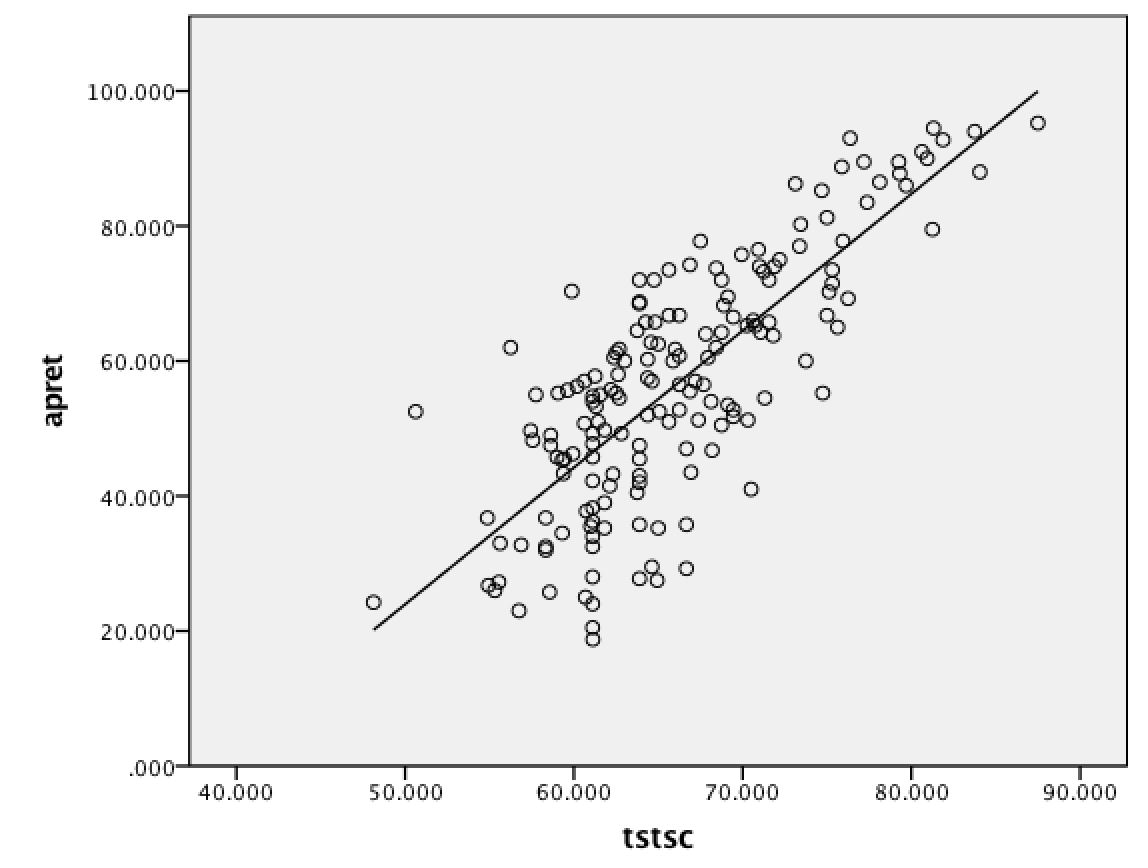


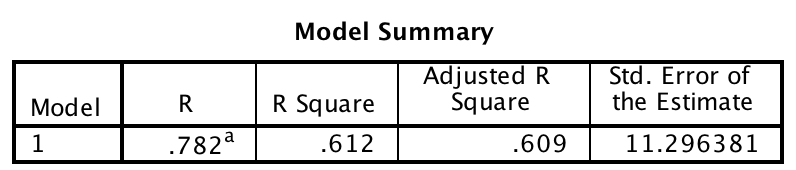
**（2）perform linear regression**

1. **linear regression of apret on tstsc**

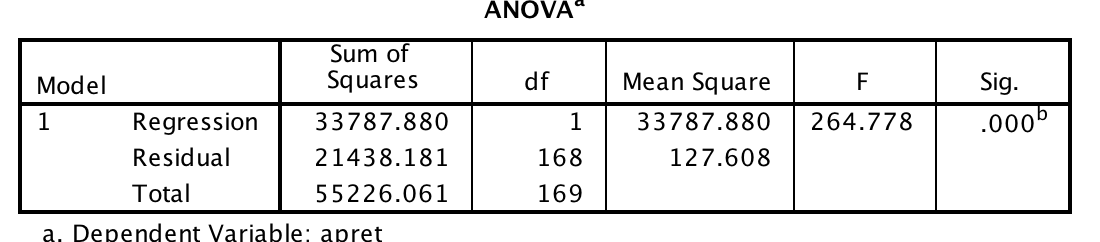
|  |
| --- |
| Dependent Variable: apret |
| Variables Entered: tstsc  Linear model:Y=AX+B |

95% confidence intervals

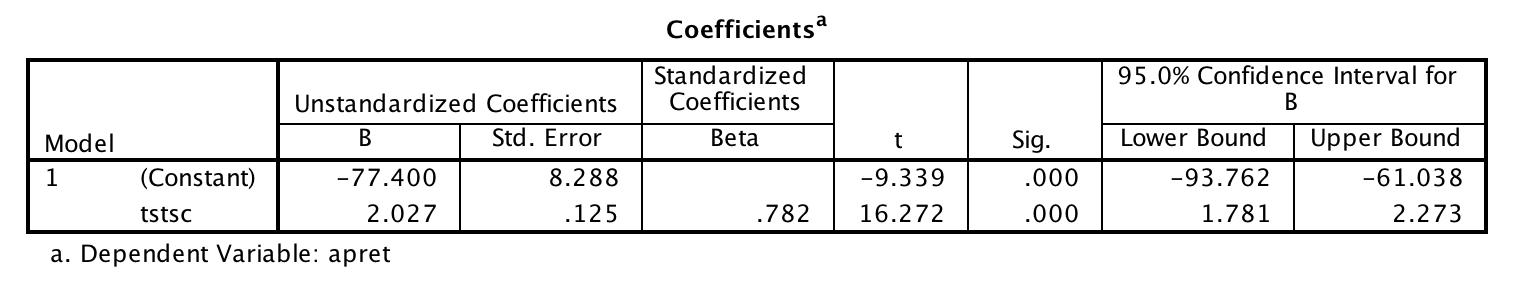




R2=0.612 and adjusted R2=0.609, so the fitting degree of model and data is good



The sig=0.000<0.05, so the regression equation is valid



Constant=-77.400

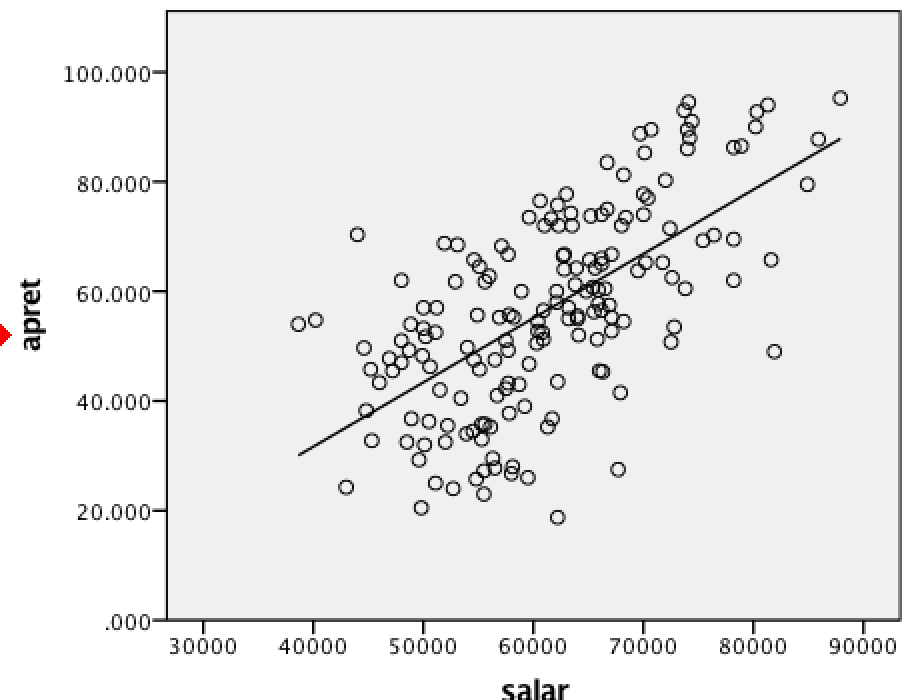
tstsc(B)=2.027

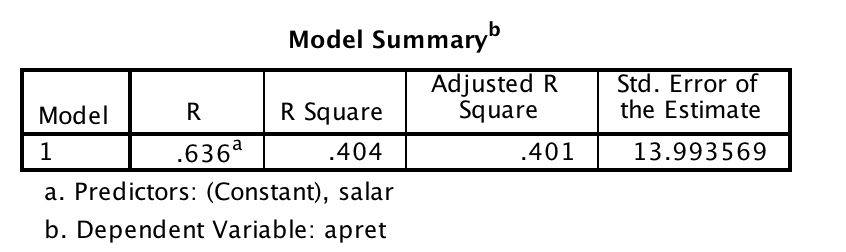
apret=2.027tstsc-77.4000

**（b）linear regression of apret on salar**

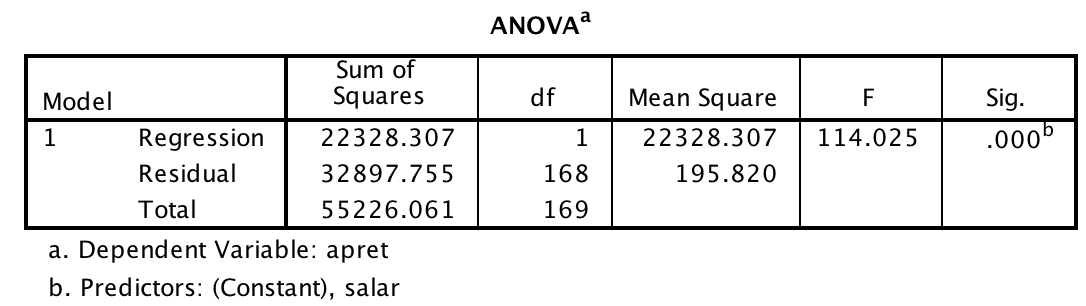
|  |
| --- |
| Dependent Variable: apret |
| Variables Entered: salar  Linear model:Y=AX+B |

95% confidence intervals

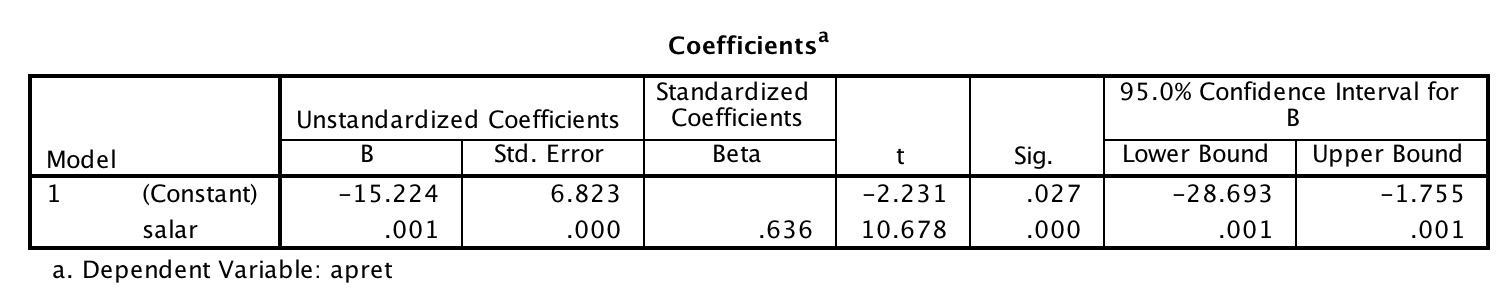




R2=0.404 and adjusted R2=0.404, so the fitting degree of model and data is good



Sig=0.000<0.05 , so the regression equation is valid



Constant=-15.224

salar(B)=0.001

apret=0.001salar-15.224

**(c) linear regression of apret on tstsc and salar**

Dependent Variable: apret

Variables Entered: salar， tstsc

R2=0.624

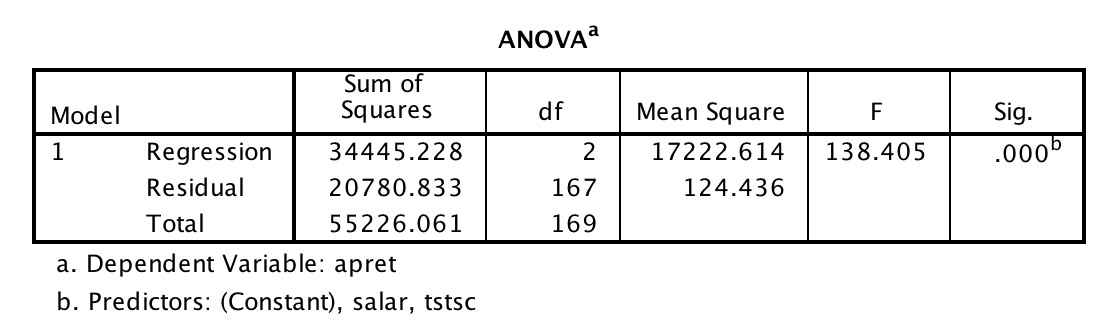
Adjusted R2=0.619

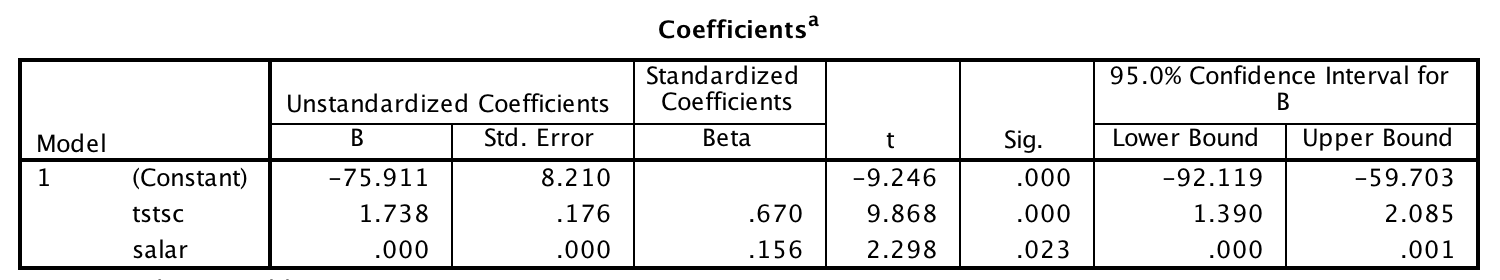
Sig=0.000

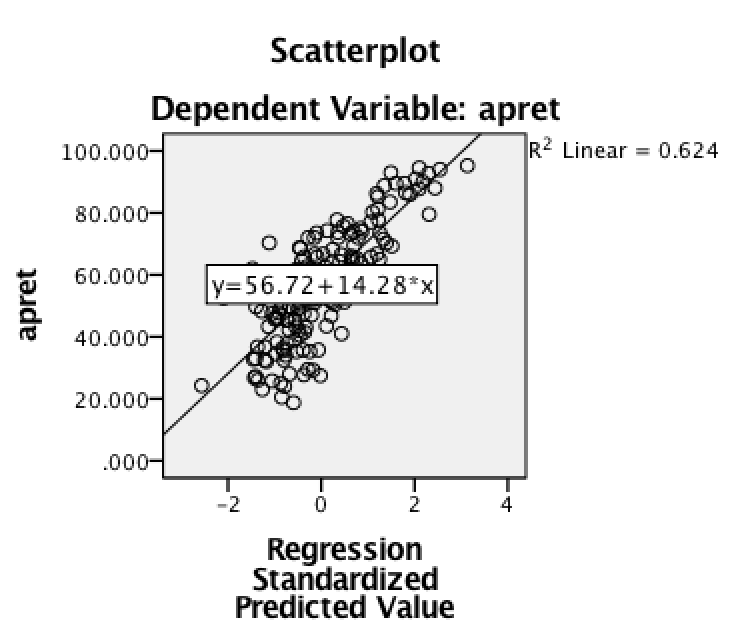
Constant=-75.911

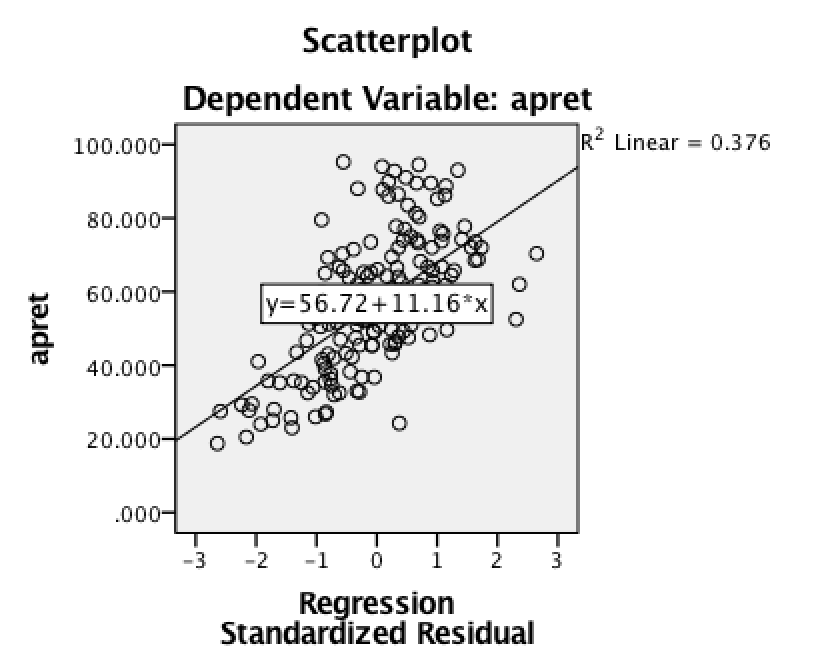
tstsc(B)=1.738

salar(B)=0.000



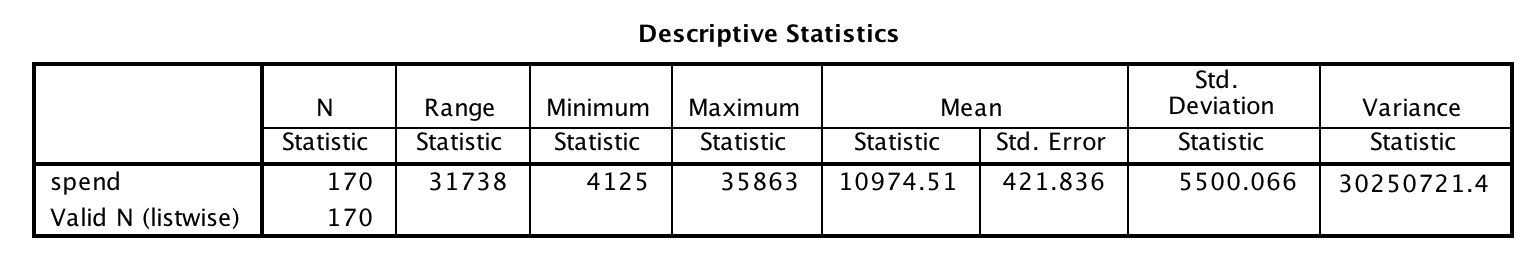


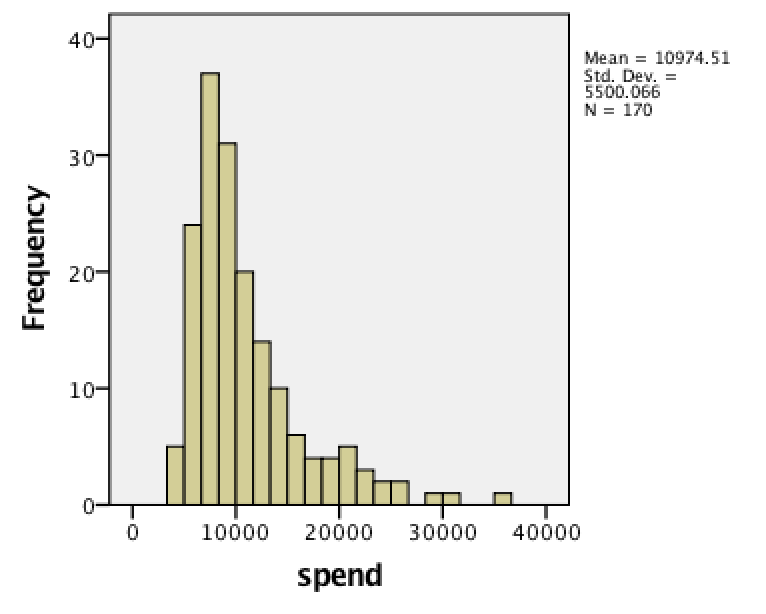


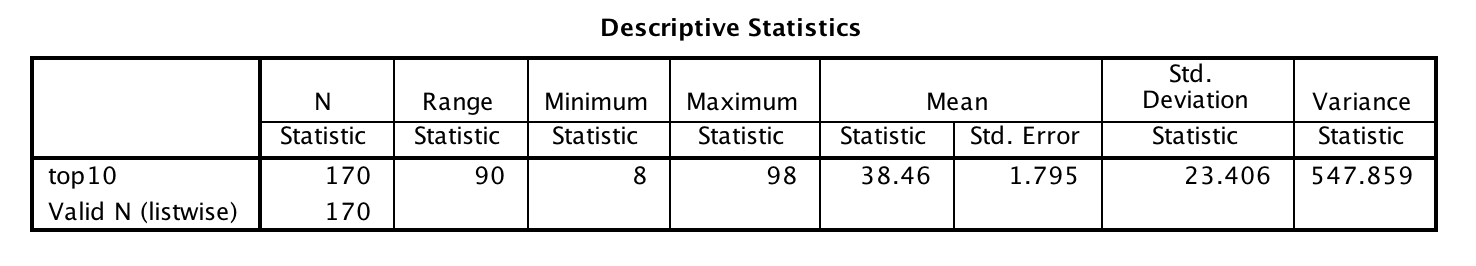


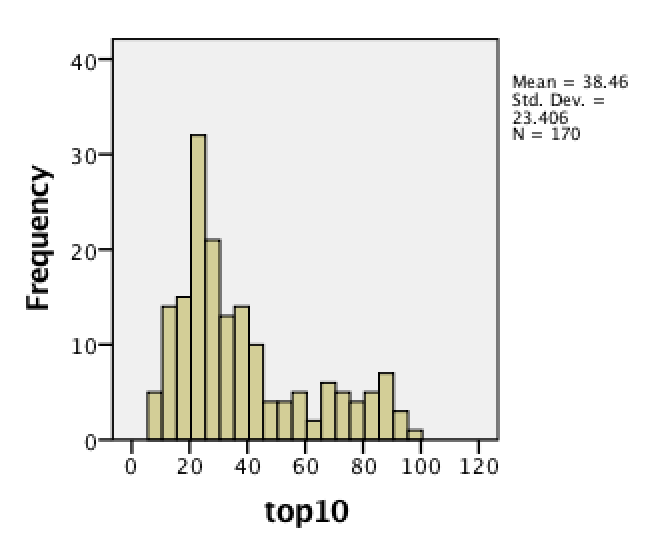
**（3）Generate descriptive statistics and histograms for spend，top10，rejr，pacc，strat**

(a)spend

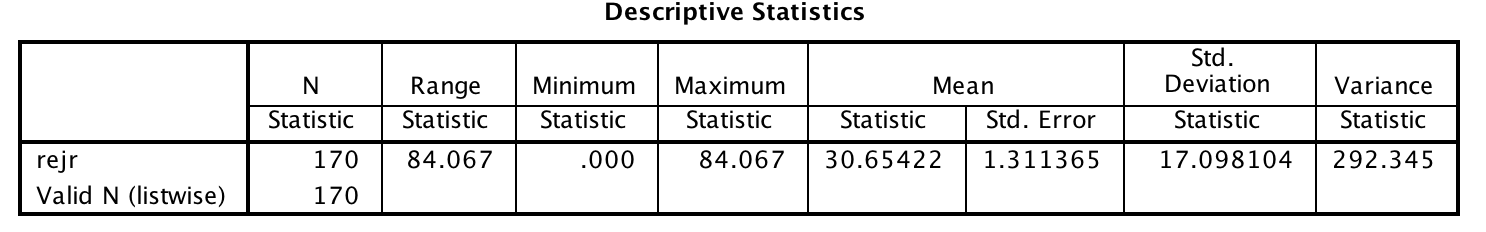


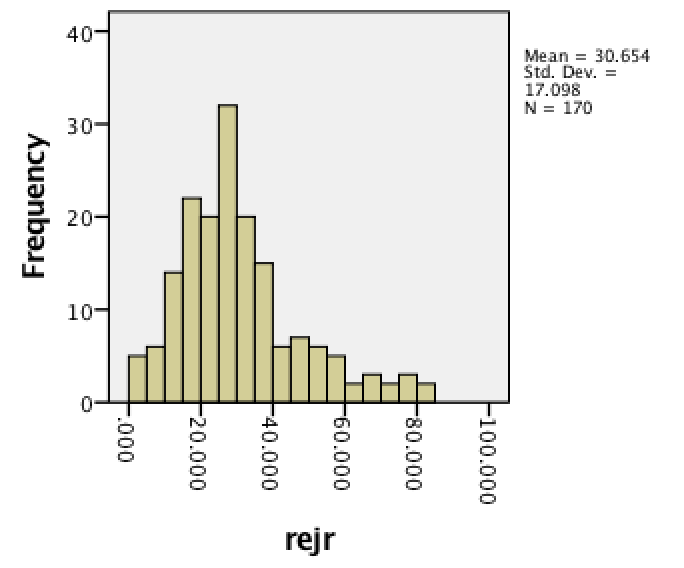


(b) top 10

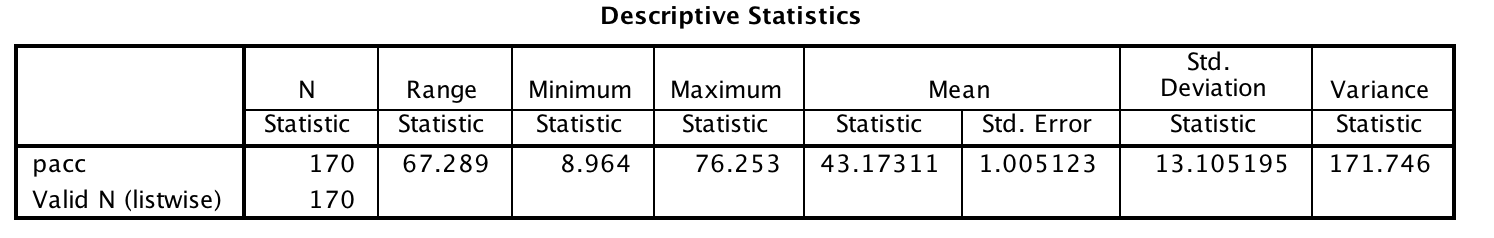


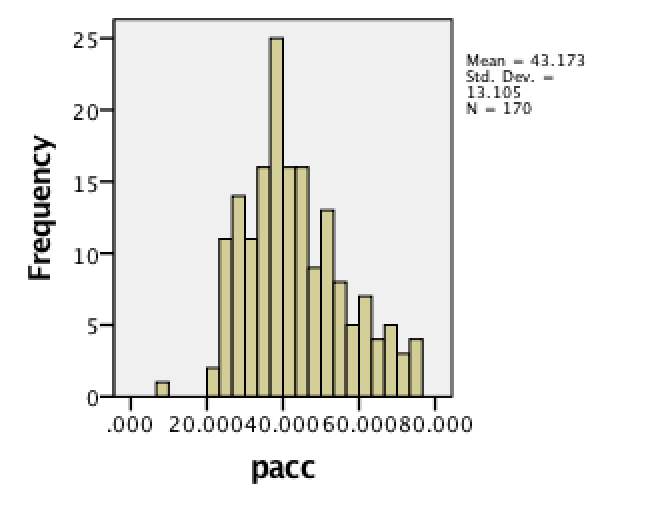
1. rejr



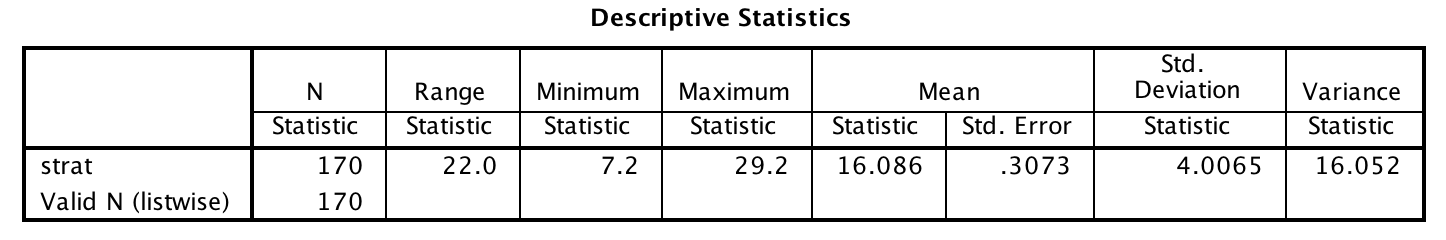


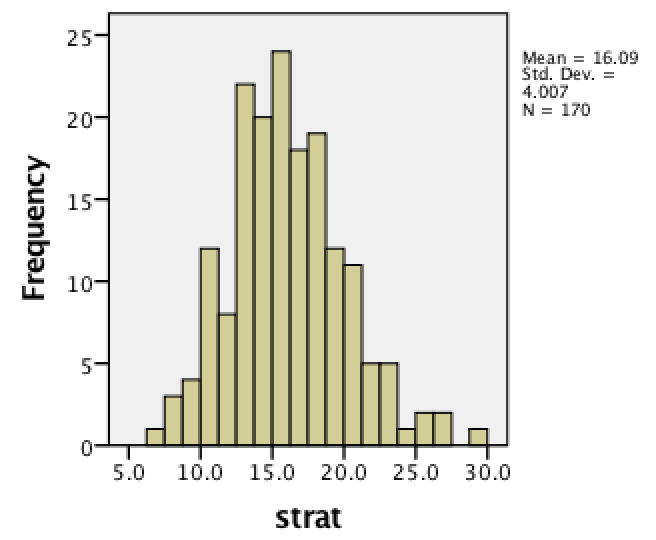
1. pacc





1. strat

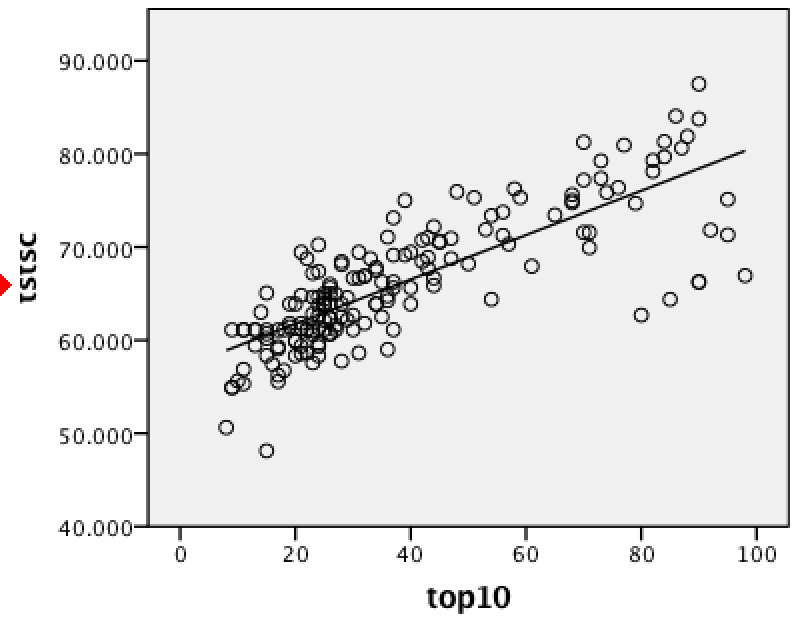


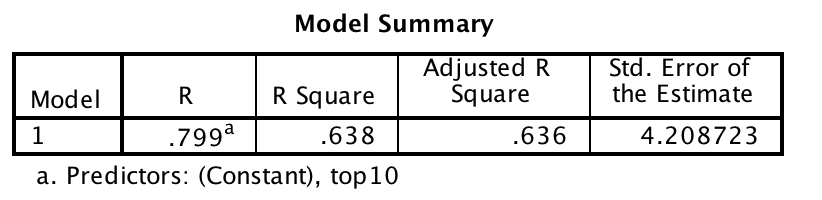


**(4) Analyze the tstsc and top10**

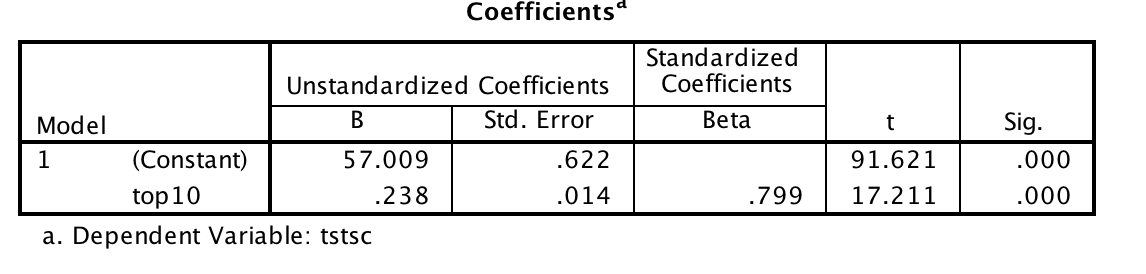
Dependent Variable: tstsc

Variables Entered: top10





R2=0.638 and Adjusted R2=0.636 so the fitting degree of model and data is good



Constant=57.009

Top10(B)=0.238

So tstsc=0.238top10+57.009