MOOC Econometrics

Test Exercise 2

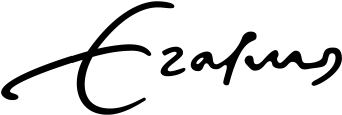
**Notes:**

* See website for how to submit your answers and how feedback is organized.
* This exercise uses the datafile TestExer2 and requires a computer.
* The dataset TestExer2 is available on the website.

**Goals and skills being used:**

* Experience the process of practical application of multiple regression.
* Get hands-on experience with performing multiple regression.
* Give correct interpretation of regression outcomes.

**Questions**

This test exercise is of an applied nature and uses data that are available in the data file TestExer2. The exercise is based on Exercise 3.14 of ‘Econometric Methods with Applications in Business and Economics’. The question of interest is whether the study results of students in Economics can be predicted from the scores on entrance tests taken before they start their studies. More precisely, you are asked to investigate whether verbal and mathematical entrance tests predict freshman grades of students in Economics. Data are available for 609 students on the following variables:

* FGPA: Freshman grade point average (scale 0-4)
* SATV: Score on SAT Verbal test (scale 0-10)
* SATM: Score on SAT Mathematics test (scale 0-10)
* FEM: Gender dummy (1 for females, 0 for males)

1. (i) Regress FGPA on a constant and SATV. Report the coefficient of SATV and its standard error and p-value (give your answers with 3 decimals).
   1. Determine a 95% confidence interval (with 3 decimals) for the effect on FGPA of an increase by 1 point in SATV.
2. Answer questions (a-i) and (a-ii) also for the regression of FGPA on a constant, SATV, SATM, and FEM.
3. Determine the (4 × 4) correlation matrix of FGPA, SATV, SATM, and FEM. Use these correlations to explain the differences between the outcomes in parts (a) and (b).
4. (i) Perform an *F*-test on the significance (at the 5% level) of the effect of SATV on FGPA, based on the regression in part (b) and another regression.

Note: Use the *F*-test in terms of SSR or *R*2 and use 6 decimals in your computations. The relevant critical value is 3.9.

* 1. Check numerically that *F* = *t*2.

============================== Answer =====================================================  
**(a)**  
(i) Linear Regression: FGPA = b0 + b1\*SATV + eps

The intercept b0 = 2.442  
 The coeff(SATV) b1 = 0.063  
 The std error(SATV) se1 = 0.028  
 The p-value(SATV) p1 = 0.023  
  
(ii) The 95% confidence interval for the effect on FGPA of an increase by 1 point in SATV:  
 [0.009, 0.117]

**(b)**  
(i) Linear Regression: FGPA = b0 + b1\*SATV + b2\*SATM + b3\*FEM + eps  
 The intercept b0 = 1.557  
 The coeff(SATV) b1 = 0.014  
 The std error(SATV) se1 = 0.028  
 The p-value(SATV) p1 = 0.612  
  
(ii) The 95% confidence interval for the effect on FGPA of an increase by 1 point in SATV:  
 [-0.041, 0.069]  
  
 OLS Regression Results  
=======================================================  
Dep. Variable: y R-squared: 0.083  
Model: OLS Adj. R-squared: 0.078  
Method: Least Squares F-statistic: 18.24  
Date: Sat, 11 Nov 2017 Prob (F-statistic): 2.41e-11  
Time: 00:30:39 Log-Likelihood: -364.67  
No. Observations: 609 AIC: 737.3  
Df Residuals: 605 BIC: 755.0  
Df Model: 3   
Covariance Type: nonrobust  
=======================================================  
 coef std err t P>|t| [0.025 0.975]  
-----------------------------------------------------------------------------------------  
const 1.5570 0.216 7.205 0.000 1.133 1.981  
**x1 0.0142 0.028 0.507 0.612** -0.041 0.069  
x2 0.1727 0.032 5.410 0.000 0.110 0.235  
x3 0.2003 0.037 5.358 0.000 0.127 0.274  
=======================================================

**(c)**The correlation matrix(4x4) is as follows.  
  
 FGPA SATV SATM FEM  
FGPA 1.000000 0.092167 0.195040 0.176491  
SATV 0.092167 1.000000 0.287801 0.033577  
SATM 0.195040 0.287801 1.000000 -0.162680  
FEM 0.176491 0.033577 -0.162680 1.000000  
  
The regression analysis in (a) shows that SATV has significant impact on FGPA (p-value = 0.028).

However, in part (b), when all three variables (SATV, SATM, FEM) are included in OLS regression, the impact of SATM is significant (p-value is near 0), but SATV is less significant (p-value of 0.612 way too high).  
  
As shown in the correlation matrix(4x4), the correlation between SATV and SATM is positive 0.29. The impact of SATV on FGPA in regression model (a) may be due to the impact of SATM and the positive correlation between SATV and STAM.

**(d)**

(i)

Test H0: βSATV = 0 against. H1: βSATV ≠ 0

g = 1  
n = 609  
k = 4  
n-k = 605

(restricted) = 0.082575  
 (unrestricted) = 0.082965

F = = 0.257155

At 5% level, the critical value of F(g, n-k) = F(1, 605) is 3.9  
  
As F = 0.257155 < 3.9, H0 is not rejected (at 5% level).

(ii)

t-stat of SATV in (b): t = 0.507   
  
t2 = 0.257155

Thus, F = t2.  
  
  
  
=====================TestExer2-GPA-round2.py===============  
  
**import** pandas **as** pd  
**import** numpy **as** np  
**import** statsmodels.api **as** sm  
**import** seaborn **as** sns  
**import** matplotlib.pyplot **as** plt  
**import** scipy.stats **as** ss  
  
df = pd.read\_excel(**'TestExer2-GPA-round2.xls'**)  
  
**print**(**"\n(a):"**)  
**print**(**'(i)'**)  
  
y = np.array(df[**'FGPA'**])  
x = np.array(df[**'SATV'**])  
  
*# sns.set(color\_codes=True)  
# ax = sns.regplot(data=df, x='SATV', y='FGPA', marker='+')  
# plt.show()*X = sm.add\_constant(x)  
model = sm.OLS(y, X)  
r = model.fit()  
**print**(**'Linear Regression: FGPA = b0 + b1\*SATV + eps'**)  
**print**(**'The intercept = %.3f'** % r.params[0])  
**print**(**'The coeff(SATV) = %.3f'** % r.params[1])  
**print**(**'The std error(SATV) = %.3f'** % r.bse[1])  
**print**(**'The p-value(SATV) = %.3f'** % r.pvalues[1])  
  
**print**(**'(ii)'**)  
b1 = r.params[1]  
sb1 = r.bse[1]  
**print**(**'The 95%% confidence interval for the effect on FGPA of an increase by 1 point in SATV: [%.3f, %.3f]'** % (b1-1.96\*sb1, b1+1.96\*sb1))  
**print**()  
**print**(r.summary())  
  
**print**(**'\n(b):'**)  
**print**(**'(i)'**)  
y = np.array(df[**'FGPA'**])  
x3 = np.array(df[[**'SATV'**, **'SATM'**, **'FEM'**]])  
  
X3 = sm.add\_constant(x3)  
model3 = sm.OLS(y, X3)  
r3 = model3.fit()  
**print**(**'Linear Regression: FGPA = b0 + b1\*SATV + b2\*SATM + b3\*FEM + eps'**)  
**print**(**'The intercept = %.3f'** % r3.params[0])  
**print**(**'The coeff(SATV) = %.3f'** % r3.params[1])  
**print**(**'The std error(SATV) = %.3f'** % r3.bse[1])  
**print**(**'The p-value(SATV) = %.3f'** % r3.pvalues[1])  
  
**print**(**'(ii)'**)  
b1 = r3.params[1]  
sb1 = r3.bse[1]  
**print**(**'The 95%% confidence interval for the effect on FGPA of an increase by 1 point in SATV: [%.3f, %.3f]'** % (b1-1.96\*sb1, b1+1.96\*sb1))  
**print**()  
**print**(r3.summary())  
  
**print**(**'\n(c):'**)  
df4 = df[[**'FGPA'**, **'SATV'**, **'SATM'**, **'FEM'**]]  
  
**print**(df4.corr())  
  
**print**(**'\n(d):'**)  
**print**(**'(i)'**)  
  
g = 1  
n = 609  
k = 4  
  
**print**(**'H0: beta(SATV) = 0 ....... that is, SATV has no effect.'**)  
  
y = np.array(df[**'FGPA'**])  
x1 = np.array(df[[**'SATM'**, **'FEM'**]])  
  
X1 = sm.add\_constant(x1)  
model1 = sm.OLS(y, X1)  
r1 = model1.fit()  
  
R1\_squared = r3.rsquared *# unrestricted model*R0\_squared = r1.rsquared *# restricted model*F = ((R1\_squared-R0\_squared)/g)/((1-R1\_squared)/(n-k))  
**print**(**'g=%d, n=%d, k=%d, n-k=%d'** % (g,n,k, n-k))  
**print**(**'R0\_squared(restricted) = %.6f, R1\_squared(unrestricted) = %.6f'** % (R0\_squared, R1\_squared))  
**print**(**'F = ((R1^2 - R0^2)/g) / ((1-R1^2)/(n-k)) = %.6f'** % F)  
  
**print**(**'At 5%% level, the critical value of F(g,n-k) = F(%d, %d) is %.1f'** % (g, n-k, ss.f.ppf(q=1-0.05, dfn=g, dfd=n-k))) *# 3.9***print**(**'\n(ii)'**)  
t = r3.tvalues  
**print**(**'t^2 = %.6f'** % (t[1]\*\*2))  
**print**(**'F = %.6f'** % F)