BS2280 – Econometrics I
Homework 2: Introduction to Regression Analysis

1

The data below show alcohol expenditure and income (both in £s per week) for sample of 17 families.

Family	Alcohol Expenditure	Income
1	26.17	487
2	19.49	574
3	17.87	439
4	16.90	367
5	4.21	299
6	32.08	743
7	30.19	433
8	22.62	547
9	9.86	303
10	13.32	370
11	9.24	299
12	47.35	531
13	26.80	506
14	33.44	613
15	21.41	472
16	16.06	253
17	24.98	374

a. Draw and XY plot of the data and comment.

b. From the chart, would you expect the line of best fit to slope up or down? In theory, which way should it slope?

c. What would you expect the correlation coefficient to be, approximately?

d. Calculate the Covariance between alcohol expenditure and income.

e. Calculate the correlation coefficient between alcohol spending and income.

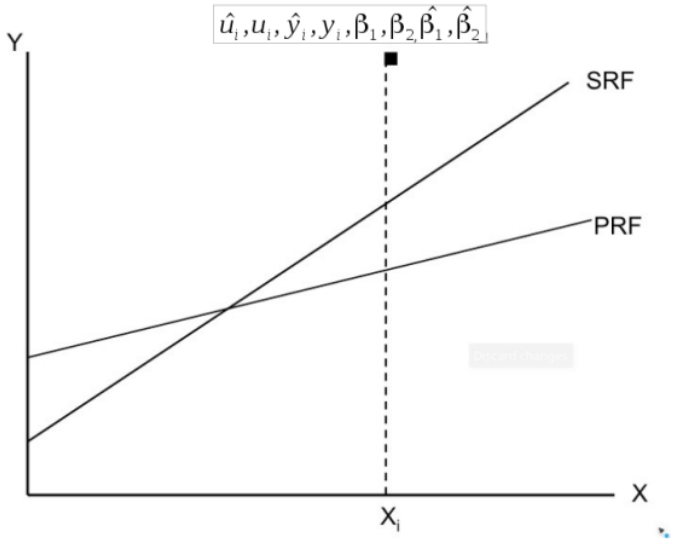
f. Answer questions a to e now using R. To load the data into R, copy the following command:
data <- data.frame(alc = c(26.17, 19.49, 17.87, 16.90, 4.21, 32.08, 30.19, 22.62, 9.86, 13.32, 9.24, 47.35, 26.80, 33.44, 21.41, 16.06, 24.98), inc = c(487, 574, 439, 367, 299, 743, 433, 547, 303, 370, 299, 531, 506, 613, 472, 253, 374))

2

What is the role of the stochastic error term u_i in regression analysis? What is the difference between the stochastic error term and the residual, $\hat{u}_i$?

3

The figure below shows the regression line based on variable X ans Y. SRF represents the sample regression function and PRF the population regression function. Label the diagram with the following:



4

The table below shows the average annual percentage rates of growth of employment, e, and real GDP, g, for 31 OECD countries for the period 2002–2007. The regression output shows the result of regressing e on g. Provide an interpretation of the coefficients.

Average annual percentage rates of growth of employment and real GDP, 2002–2007

	Employment	*GDP*		*Employment*	*GDP*
Australia	2.57	3.52	Korea	1.11	4.48
Austria	1.64	2.66	Luxembourg	1.34	4.55
Belgium	1.06	2.27	Mexico	1.88	3.36
Canada	1.90	2.57	Netherlands	0.51	2.37
Czech Republic	0.79	5.62	New Zealand	2.67	3.41
Denmark	0.58	2.02	Norway	1.36	2.49
Estonia	2.28	8.10	Poland	2.05	5.16
Finland	0.98	3.75	Portugal	0.13	1.04
France	0.69	2.00	Slovak Republic	2.08	7.04
Germany	0.84	1.67	Slovenia	1.60	4.82
Greece	1.55	4.32	Sweden	0.83	3.47
Hungary	0.28	3.31	Switzerland	0.90	2.54
Iceland	2.49	5.62	Turkey	1.30	6.90
Israel	3.29	4.79	United Kingdom	0.92	3.31
Italy	0.89	1.29	United States	1.36	2.88
Japan	0.31	1.85			

Regression output:

```
Call:
lm(formula = e ~ g, data = oecd_exercises)

Residuals:
     Min       1Q   Median       3Q      Max
-1.03915 -0.42605 -0.08701  0.30295  1.65834

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.49195    0.28325   1.737  0.09303 .
g            0.23794    0.07025   3.387  0.00205 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6692 on 29 degrees of freedom
Multiple R-squared:  0.2834,    Adjusted R-squared:  0.2587
F-statistic: 11.47 on 1 and 29 DF,  p-value: 0.002049
```

5 *

In the lecture we introduced the first order conditions of the RSS minimisation problem:

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 2n\hat{\beta}_1 - 2\sum_{i=1}^{n} Y_i + 2\hat{\beta}_2 \sum_{i=1}^{n} X_i = 0$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 2\hat{\beta}_2 \sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i Y_i + 2\hat{\beta}_1 \sum_{i=1}^{n} X_i = 0$$

Derive $\hat{\beta}_1$ and $\hat{\beta}_2$ mathematically using the first order conditions. (Hint: $\sum_{i=1}^{n} X_i = n\bar{X}$)