

Department of Statistics

COURSE STATS 330/762

Assignment 2, 2011

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm August 18th**.

The data sets for this assignment is in the file **savings.txt** which is available on the course web page.

The data consist of economic measurements on several countries collected during a study conducted in the 1970's which investigated the "life-cycle savings hypothesis" which states that the savings rate in a country is determined by the per-capita disposable income, the rate of change in this quantity, and two demographic variable, the percentage of population under 15 years of age and percentage over 75 years of age. The variables in the data set are (all averaged over the period 1960-1970)

sr:	The average aggregate personal savings,
pop15:	The average percentage of the population under 15,
pop75:	The average percentage of the population over 75,
dpi:	The average level of per-capita disposable income (measured in \$US),
ddpi:	The average percentage growth rate of dpi.

Questions and tasks

1. Load the data into R, and make a data frame **savings.df** to contain the data. Check for any typographical errors (the data below may be taken to be the correct data, but the data on the web may have been corrupted). Correct as necessary. Print out the last 10 lines of the data file. [5 marks]
2. Fit a regression model to these data, using **sr** as the response. Do all variables appear to be necessary in the model? Do you think that the model using **ddpi** alone is an adequate submodel? Give reasons.[10 marks]
3. Do you think that the data support the life-cycle savings hypothesis? Give reasons. [5 marks]

4. Are there any data points that unduly influence the results? What aspect of the results are influenced by what data points? Does this make you change your mind about the answer to Q3? Give a full discussion with reasons. [15 marks]

5. Do you think these data have collinearity problems? Give reasons. What is the consequence if they do? [5 marks]

Data follow – you may take these as the **correct** data

	sr	pop15	pop75	dpi	ddpi
Australia	11.43	29.35	2.87	2329.68	2.87
Austria	12.07	23.32	4.41	1507.99	3.93
Belgium	13.17	23.80	4.43	2108.47	3.82
Bolivia	5.75	41.89	1.67	189.13	0.22
Brazil	12.88	42.19	0.83	728.47	4.56
Canada	8.79	31.72	2.85	2982.88	2.43
Chile	0.60	39.74	1.34	662.86	2.67
China	11.90	44.75	0.67	289.52	6.51
Colombia	4.98	46.64	1.06	276.65	3.08
Costa Rica	10.78	47.64	1.14	471.24	2.80
Denmark	16.85	24.42	3.93	2496.53	3.99
Ecuador	3.59	46.31	1.19	287.77	2.19
Finland	11.24	27.84	2.37	1681.25	4.32
France	12.64	25.06	4.70	2213.82	4.52
Germany	12.55	23.31	3.35	2457.12	3.44
Greece	10.67	25.62	3.10	870.85	6.28
Guatamala	3.01	46.05	0.87	289.71	1.48
Honduras	7.70	47.32	0.58	232.44	3.19
Iceland	1.27	34.03	3.08	1900.10	1.12
India	9.00	41.31	0.96	88.94	1.54
Ireland	11.34	31.16	4.19	1139.95	2.99
Italy	14.28	24.52	3.48	1390.00	3.54
Japan	21.10	27.01	1.91	1257.28	8.21
Korea	3.98	41.74	0.91	207.68	5.81
Luxembourg	10.35	21.80	3.73	2449.39	1.57
Malta	15.48	32.54	2.47	601.05	8.12
Norway	10.25	25.95	3.67	2231.03	3.62
Netherlands	14.65	24.71	3.25	1740.70	7.66
New Zealand	10.67	32.61	3.17	1487.52	1.76
Nicaragua	7.30	45.04	1.21	325.54	2.48
Panama	4.44	43.56	1.20	568.56	3.61
Paraguay	2.02	41.18	1.05	220.56	1.03
Peru	12.70	44.19	1.28	400.06	0.67
Philippines	12.78	46.26	1.12	152.01	2.00
Portugal	12.49	28.96	2.85	579.51	7.48
South Africa	11.14	31.94	2.28	651.11	2.19
South Rhodesia	13.30	31.92	1.52	250.96	2.00
Spain	11.77	27.74	2.87	768.79	4.35
Sweden	6.86	21.44	4.54	3299.49	3.01
Switzerland	14.13	23.49	3.73	2630.96	2.70
Turkey	5.13	43.42	1.08	389.66	2.96
Tunisia	2.81	46.12	1.21	249.87	1.13

United Kingdom	7.81	23.27	4.46	1813.93	2.01
United States	7.56	29.81	3.43	4001.89	2.45
Venezuela	9.22	46.40	0.90	813.39	0.53
Zambia	18.56	45.25	0.56	138.33	5.14
Jamaica	7.72	41.12	1.73	380.47	10.23
Uruguay	9.24	28.13	2.72	766.54	1.88
Libya	8.89	43.69	2.07	123.58	16.71
Malaysia	4.71	47.20	0.66	242.69	5.08