

Department of Statistics

COURSE STATS 330

Assignment 1, 2008

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 7 Aug.

The data sets for this assignment are in the files **romania.txt** (for Question 1) and **weights.txt** (for Question 2) which are available on the course web page.

Question 1.

The data in the file **romania.txt** relate to a peasant rebellion which took place in Romania in 1907, and which was viciously suppressed by the Romanian government with considerable loss of life.

In an article in the *American Sociological Review* (Chirot, D and Ragin, C, (1975). The market, tradition and peasant rebellion: the case of Romania in 1907, *American Sociological Review*, 40, 428-444) Daniel Chirot and Charles Ragin examine the causes of this rebellion. They gathered data on 32 counties in Romania where the rebellion had taken place, with a view to performing a statistical analysis of the causes of the rebellion.

The variables measured for each county were

intensity: Intensity of the rebellion in the county, measured by an index made up of the number of deaths and the number of villages affected by the rebellion, scaled to have mean zero. The higher the index, the more intense the rebellion in that county.

commerce: Commercialization of agriculture, measured by the percentage of arable land in the county devoted to wheat (a cash crop)

tradition: Traditionalism, measured by the percentage of the rural population that are illiterate

midpeasant : Strength of middle peasantry, measured by the percentage of land owned in units of 7 to 50 haectares

inequality: Inequality of land tenure, as measured by the Gini index (A Gini index of 0 means perfect equality, with everyone owning the same amount of land, while 1 is

perfect inequality, with one person owing all the land. See http://en.wikipedia.org/wiki/Gini_coefficient for more details.)

The authors were interested in testing two theories. The first, the “transitional” theory, hold that rebellion is more likely when agriculture when there has been a extensive recent commercialization of agriculture in a traditional peasant society. The second theory, the “structural” theory, holds that rebellion is caused by class resentments, which would be greater in counties where the middle peasantry is strong, and where land inequality is great.

Instructions

1. Load the data into R, and make a data frame **Chirot.df** to contain the data. Check for any typographical errors. [5 marks]
2. Are either of these theories supported by the data? What is the relationship between intensity and the other variables? Is the relationship planar? Draw suitable plots to answer this question. Don't try and fit any models. [20 marks, 10 for the plots and 10 for the discussion]
3. Are there any features of the data that might make fitting a regression model difficult? [5 marks]

Question 2

The file `weights.txt` contains data on 200 subjects; these were men and women engaged in regular exercise. There are some missing data. There are 5 variables:

sex: A factor with levels: F, female; M, male.

weight: Measured weight in kg.

height: Measured height in cm.

repwt : Reported weight in kg.

repht: Reported height in cm.

The data were gathered in a study to relate subjects' self-reported height and weight to their actual height and weight.

Instructions

1. Load the data into R, and make a data frame **weights.df** to contain the data. Check for any typographical errors. [3 marks]
2. Is there any evidence in these data of systematic over- or under-reporting of height and weight? Does the sex of the subject make a difference? As in Question 1, don't fit any models, confine yourself to graphical methods. [7 marks, 3 for graph and 4 for discussion]

R hints

1. Use **read.table** to create the data frames.
2. Use regular R graphics, XGobi and trellis graphics to help answer Question 1.
3. Don't forget to label the axes. Use **xlab** and **ylab**. Skim the paper "A Tour of Trellis Graphics" on the web site for helpful hints.
4. A trellis plot is useful for Question 2. You may consider adding a reference line to the plot: The function **panel.abline** may be useful. See "A Tour of Trellis Graphics".



Romania data

intensity	commerce	tradition	midpeasant	inequality
-1.37	13.8	86.2	6.2	0.6
0.67	20.4	86.7	2.9	0.72
1.9	27.6	79.3	16.9	0.66
-0.13	18.6	90.1	3.4	0.74
-0.84	17.2	84.5	9	0.7
0.13	21.5	81.5	5.2	0.6
-0.48	11.6	82.6	5.1	0.52
-0.84	20.4	82.4	6.3	0.64
-0.22	19.5	87.5	4.8	0.68
-0.75	8.9	85.6	9.5	0.58
-0.93	25.8	82.2	10.9	0.68
-1.55	24.1	83.5	8.4	0.74
-0.48	22	88.3	6.2	0.7
-1.55	24.2	84.9	6.1	0.62
-0.48	30.6	76.1	1.3	0.76
-1.1	33.9	85.5	5.8	0.7
-1.19	28.6	84.2	2.9	0.58
-1.19	36.5	78.1	4.3	0.72
-0.84	40.9	84.4	2.3	0.64
-1.36	6.8	76.3	3.6	0.58
2.83	41.9	89.7	6.6	0.66
-1.02	25.4	83.2	2.5	0.68
1.59	30.5	80.2	4.1	0.76
4.33	48.2	91	4.2	0.7
3.8	46	90.5	3.7	0.68
3.8	45.1	85.5	5.1	0.64
-1.72	12.5	83.8	7.2	0.5
0.84	39.3	85.6	4.9	0.6
2.61	47.7	87.6	5.2	0.58
-0.84	15.2	87.3	10.8	0.42
-1.81	11.7	82.3	81.7	0.42
-1.81	25.6	80.1	68.4	0.26

Weight data

sex	weight	height	repwt	repht
M	77	182	77	180
F	58	161	51	159
F	53	161	54	158
M	68	177	70	175
F	59	157	59	155
M	76	170	76	165
M	76	167	77	165
M	69	186	73	180
M	71	178	71	175
M	65	171	64	170
M	70	175	75	174
F	166	57	56	163
F	51	161	52	158
F	64	168	64	165
F	52	163	57	160
F	65	166	66	165
M	92	187	101	185
F	62	168	62	165
M	76	197	75	200
F	61	175	61	171
M	119	180	124	178
F	61	170	61	170
M	65	175	66	173
M	66	173	70	170
F	54	171	59	168
F	50	166	50	165
F	63	169	61	168
F	58	166	60	160
F	39	157	41	153
M	101	183	100	180
F	71	166	71	165
M	75	178	73	175
M	79	173	76	173
F	52	164	52	161
F	68	169	63	170
M	64	176	65	175
F	56	166	54	165
M	69	174	69	171
M	88	178	86	175
M	65	187	67	188

F	54	164	53	160
M	80	178	80	178
F	63	163	59	159
M	78	183	80	180
M	85	179	82	175
F	54	160	55	158
M	73	180	NA	NA
F	49	161	NA	NA
F	54	174	56	173
F	75	162	75	158
M	82	182	85	183
F	56	165	57	163
M	74	169	73	170
M	102	185	107	185
M	64	177	NA	NA
M	65	176	64	172
F	66	170	65	NA
M	73	183	74	180
M	75	172	70	169
M	57	173	58	170
M	68	165	69	165
M	71	177	71	170
M	71	180	76	175
F	78	173	75	169
M	97	189	98	185
F	60	162	59	160
F	64	165	63	163
F	64	164	62	161
F	52	158	51	155
M	80	178	76	175
F	62	175	61	171
M	66	173	66	175
F	55	165	54	163
F	56	163	57	159
F	50	166	50	161
F	50	171	NA	NA
F	50	160	55	150
F	63	160	64	158
M	69	182	70	180
M	69	183	70	183
F	61	165	60	163
M	55	168	56	170
F	53	169	52	175

F	60	167	55	163
F	56	170	56	170
M	59	182	61	183
M	62	178	66	175
F	53	165	53	165
F	57	163	59	160
F	57	162	56	160
M	70	173	68	170
F	56	161	56	161
M	84	184	86	183
M	69	180	71	180
M	88	189	87	185
F	56	165	57	160
M	103	185	101	182
F	50	169	50	165
F	52	159	52	153
F	55	155	NA	154
F	55	164	55	163
M	63	178	63	175
F	47	163	47	160
F	45	163	45	160
F	62	175	63	173
F	53	164	51	160
F	52	152	51	150
F	57	167	55	164
F	64	166	64	165
F	59	166	55	163
M	84	183	90	183
M	79	179	79	171
F	55	174	57	171
M	67	179	67	179
F	76	167	77	165
F	62	168	62	163
M	83	184	83	181
M	96	184	94	183
M	75	169	76	165
M	65	178	66	178
M	78	178	77	175
M	69	167	73	165
F	68	178	68	175
F	55	165	55	163
M	67	179	NA	NA
F	52	169	56	NA

F	47	153	NA	154
F	45	157	45	153
F	68	171	68	169
F	44	157	44	155
F	62	166	61	163
M	87	185	89	185
F	56	160	53	158
F	50	148	47	148
M	83	177	84	175
F	53	162	53	160
F	64	172	62	168
F	62	167	NA	NA
M	90	188	91	185
M	85	191	83	188
M	66	175	68	175
F	52	163	53	160
F	53	165	55	163
F	54	176	55	176
F	64	171	66	171
F	55	160	55	155
F	55	165	55	165
F	59	157	55	158
F	70	173	67	170
M	88	184	86	183
F	57	168	58	165
F	47	162	47	160
F	47	150	45	152
F	55	162	NA	NA
F	48	163	44	160
M	54	169	58	165
M	69	172	68	174
F	59	170	NA	NA
F	58	169	NA	NA
F	57	167	56	165
F	51	163	50	160
F	54	161	54	160
F	53	162	52	158
F	59	172	58	171
M	56	163	58	161
F	59	159	59	155
F	63	170	62	168
F	66	166	66	165
M	96	191	95	188

F	53	158	50	155
M	76	169	75	165
F	54	163	NA	NA
M	61	170	61	170
M	82	176	NA	NA
M	62	168	64	168
M	71	178	68	178
F	60	174	NA	NA
M	66	170	67	165
M	81	178	82	175
M	68	174	68	173
M	80	176	78	175
F	43	154	NA	NA
M	82	181	NA	NA
F	63	165	59	160
M	70	173	70	173
F	56	162	56	160
F	60	172	55	168
F	58	169	54	166
M	76	183	75	180
F	50	158	49	155
M	88	185	93	188
M	89	173	86	173
F	59	164	59	165
F	51	156	51	158
F	62	164	61	161
M	74	175	71	175
M	83	180	80	180
M	81	175	NA	NA
M	90	181	91	178
M	79	177	81	178