

# Department of Statistics

## COURSE STATS 330/762

### Assignment 3, 2012

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm September 13<sup>th</sup>**.

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

### Instructions

It is 1988 in the UK, and you are an employee of STATSINC, a statistical consulting firm. You have been approached by a client in the Department of Sociology of a nearby university, who is interested in the problem of teenage gambling. In particular, she is interested in whether gender is a risk factor in teenage gambling. She tells you that other possible risk factors are income, social status, and intelligence.

She has surveyed 46 teenagers and collected data on these variables, resulting in the data file **gamble.txt** referred to above. The variables are

- sex :** 0=male, 1=female
- status :** Socioeconomic status score based on parents' occupation
- income :** in pounds per week
- verbal :** verbal score in words out of 12 correctly defined ( a proxy for intelligence)
- gamble :** expenditure on gambling in pounds per year

**Task 1.** Your task is to write a report of no more than 6 pages (including appendices) describing the effect of gender on gambling behaviour. The client may wish to have your report peer reviewed by another statistician so you should support your conclusion by a suitable analysis, described in sufficient detail so that a reviewer can follow your reasoning. [25 marks]

**Task 2.** As well as examining the effect of gender, a secondary purpose of the research is to construct a scale which will allow the researcher to predict how much an individual is likely to gamble. Construct a prediction equation from the data supplied and estimate

the prediction error. Test your prediction by predicting the gambling expenditure for a teenage boy whose parents have a socioeconomic status score of 71, who has an income of £2.50 per week and scores 9 out of 12 words correct on the verbal score.

[15 marks]

**Note:** there will be a prize for the best prediction. In the event of a tie, a winner will be selected by a stochastic mechanism.

**Total for assignment: 40 marks**

### Points to note

1. Your report should have the usual structure: Introduction, main section, conclusions, appendix. See the Powerpoint presentation on the website for some hints on report writing.
2. The variable sex is binary, not continuous. Interpreting regression coefficients for binary variables will be discussed in Lecture 17.
3. Put technical details (including R code) in the appendix of your report.
4. For Task 2, make sure you include all relevant R code and describe fully how you arrived at your prediction equation.
5. For the purposes of the prize, make sure your prediction of the teenage boy's gambling expenditure is easy for the graders to locate in your answers.

### Data set

gamble.txt

	sex	status	income	verbal	gamble
1	1	51	2.00	8	0.00
2	1	28	2.50	8	0.00
3	1	37	2.00	6	0.00
4	1	28	7.00	4	7.30
5	1	65	2.00	8	19.60
6	1	61	3.47	6	0.10
7	1	28	5.50	7	1.45
8	1	27	6.42	5	6.60
9	1	43	2.00	6	1.70
10	1	18	6.00	7	0.10
11	1	18	3.00	6	0.10
12	1	43	4.75	6	5.40
13	1	30	2.20	4	1.20

14	1	28	2.00	6	3.60
15	1	38	3.00	6	2.40
16	1	38	1.50	8	3.40
17	1	28	9.50	8	0.10
18	1	18	10.00	5	8.40
19	1	43	4.00	8	12.00
20	0	51	3.50	9	0.00
21	0	62	3.00	8	1.00
22	0	47	2.50	9	1.20
23	0	43	3.50	5	0.10
24	0	27	10.00	4	156.00
25	0	71	6.50	7	38.50
26	0	38	1.50	7	2.10
27	0	51	5.44	4	14.50
28	0	38	1.00	6	3.00
29	0	51	0.60	7	0.60
30	0	62	5.50	8	9.60
31	0	18	12.00	2	88.00
32	0	30	7.00	7	53.20
33	0	38	15.00	7	90.00
34	0	71	2.00	10	3.00
35	0	28	1.50	1	14.10
36	0	61	4.50	8	70.00
37	0	71	2.50	7	38.50
38	0	28	8.00	6	57.20
39	0	51	10.00	6	6.00
40	0	65	1.60	6	25.00
41	0	48	2.00	9	6.90
42	0	61	15.00	9	69.70
43	0	75	3.00	8	13.30
44	0	66	3.25	9	0.60
45	0	62	4.94	6	38.00
46	0	71	1.50	7	14.40