

Department of Statistics

COURSE STATS 330

Assignment 4, 2003

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 9th October.

Your answers to this assignment are **not to be given in the form of a report**. Just answer the questions below. Set your answer out in 6 sections corresponding to the 6 questions.

The data for this assignment were collected by the famous English scientist Sir Francis Galton over a hundred years ago as part of his study of natural inheritance. The study was concerned with the genetic transmission of eye colour and data was collected on the eye colour of children, as well as that of their parents and grandparents. Each line of the data file reproduced below (and available on the web site as file `galton.txt`) contains data on a single family. Information is given about

- The number of children in the family,
 - The number having light eyes,
 - The eye colour distribution of the 2 parents,
 - The eye colour distribution of the 4 grandparents.
1. Set up a 6-level factor, P , whose levels correspond to the 6 distinguishable combinations of the parents' eye colour. Note that the eye colours of fathers and mothers were not separately recorded. Suggested R-code for doing this is given at the end of the assignment.
 2. Fit a logistic regression to the proportions of children in each family that have light-coloured eyes, using P as the explanatory variable. Are there any outliers? Delete any clearly discrepant points. (You have an outlier budget of 2.)
 3. Refit the model to the remaining data. Compute the fitted probabilities for all 6 eye-colour combinations. Now make a 3×3 table of all the 9 possible eye-colour combinations of the 2 parents, distinguishing mother from father. Assume that:
 - a. the probability of a child having light eyes is the same for the combination (mother Light, father Hazel) as for the combination (mother Hazel, father Light)
 - b. the probability of a child having light eyes is the same for the combination (mother Hazel, father Dark) as for the combination (mother Dark, father Hazel)
 - c. the probability of a child having light eyes is the same for the combination (mother Light, father Dark) as for the combination (mother Dark, father Light)

4. Using these assumptions, transfer your fitted probabilities to the 3×3 table. Display the table and comment on any marked trends and patterns. What you think these data say about the inheritance of eye colour?
5. Now set up the factor G (as you did for P in Question 1) for all the distinguishable eye colour combinations of the grandparents. How many levels does this factor potentially have? How many of these levels are actually in the data set? Include your R code and print out the whole data frame with variables P, G, r and n.
6. Add this factor to your model. What evidence is there for a grandparent effect over and above the parental effect?

Sample R code for setting up the factor in Question 1:

```
galton.df<-read.table(file.choose(),header=T)

P<-character(78)

for(i in 1:78){
v<-galton.df[i,1:3]
P[i]<-if(all(v==c(2,0,0))) "P1" else
  if(all(v==c(0,2,0))) "P2" else
  if(all(v==c(0,0,2))) "P3" else
  if(all(v==c(1,1,0))) "P4" else
  if(all(v==c(1,0,1))) "P5" else
  if(all(v==c(0,1,1))) "P6" else NA
}

P<-factor(P)
```

Data for the assignment: in data file **galton.txt** on the web site. Also shown below:

<i>Number of parents in family having eye colour</i>			<i>Number of grandparents having eye colour</i>			<i>Total children in family</i>	<i>Number having light eyes</i>
<i>Light</i>	<i>Hazel</i>	<i>Dark</i>	<i>Light</i>	<i>Hazel</i>	<i>Dark</i>	<i>n</i>	<i>r</i>
2	0	0	4	0	0	6	6
2	0	0	4	0	0	6	6
2	0	0	4	0	0	6	6
2	0	0	4	0	0	6	5
2	0	0	4	0	0	7	7
2	0	0	4	0	0	7	7
2	0	0	4	0	0	7	7
2	0	0	4	0	0	7	7
2	0	0	4	0	0	7	7
2	0	0	4	0	0	8	8
2	0	0	4	0	0	8	8
2	0	0	4	0	0	8	8
2	0	0	4	0	0	8	8
2	0	0	4	0	0	8	7
2	0	0	4	0	0	8	7
2	0	0	4	0	0	12	12
2	0	0	3	1	0	7	7
2	0	0	3	1	0	10	4
2	0	0	3	1	0	12	12
2	0	0	3	0	1	7	6
2	0	0	3	0	1	8	8
2	0	0	3	0	1	9	9
2	0	0	3	0	1	9	9
2	0	0	3	0	1	9	7
2	0	0	3	0	1	10	10
2	0	0	2	2	0	7	7
2	0	0	2	2	0	10	9
2	0	0	2	1	1	6	6
2	0	0	2	1	1	10	10
0	2	0	2	1	1	7	4
0	2	0	2	0	2	8	5
0	0	2	3	0	1	6	2
0	0	2	2	0	2	9	1
0	0	2	1	0	3	6	1
0	0	2	1	0	3	11	3
0	0	2	1	1	2	6	0
0	0	2	1	1	2	7	4
1	1	0	3	1	0	6	6
1	1	0	3	1	0	7	6

1	1	0	3	1	0	8	6
1	1	0	3	1	0	9	7
1	1	0	3	1	0	11	10
1	1	0	3	0	1	9	6
1	1	0	3	0	1	11	7
1	1	0	2	2	0	7	6
1	1	0	2	2	0	9	9
1	1	0	2	2	0	11	1
1	1	0	2	0	2	6	6
1	1	0	2	0	2	6	4
1	1	0	2	0	2	8	5
1	1	0	2	0	2	9	7
1	1	0	2	1	1	6	6
1	1	0	2	1	1	10	9
1	1	0	1	3	0	9	4
1	1	0	1	1	2	8	5
1	0	1	4	0	0	7	3
1	0	1	3	0	1	6	4
1	0	1	3	0	1	7	3
1	0	1	3	0	1	8	6
1	0	1	3	0	1	8	5
1	0	1	3	0	1	8	4
1	0	1	3	0	1	9	6
1	0	1	3	0	1	9	5
1	0	1	2	0	2	6	5
1	0	1	2	0	2	6	3
1	0	1	2	0	2	8	4
1	0	1	2	0	2	10	2
1	0	1	2	0	2	14	9
1	0	1	2	1	1	7	5
1	0	1	1	2	1	7	3
1	0	1	1	1	2	7	4
1	0	1	1	0	3	8	4
1	0	1	1	0	3	8	3
1	0	1	0	1	3	6	3
0	1	1	2	0	2	6	3
0	1	1	2	1	1	9	4
0	1	1	1	0	3	13	8
0	1	1	0	4	0	7	2