Hair – Eye Color Data: Gaining insights by ACE

September 18, 2017

Table 1 shows data on the relation between hair color and eye color among 592 subjects (students in a statistics course) as presented in Snee, R. D. (1974): "Graphical display of two-way contingency tables"; **The American Statistician**, vol. 28, pp. 9-12.

We first treat the variables as nominal (unordered categories). In order to gain some insight, we use the `ace` function from `R`. ACE stands for Alternating Conditional Expectations. It is an algorithm for computing the Renyi maximal correlation. For any two random variables X and Y, this is defined as

$$R^2(X,Y) \ = \ \sup\left\{Corr(\phi(X),\psi(Y))\right\},$$

where ϕ and ψ are arbitrary (measurable) transformations of X and Y. Here, we must have that both transformed random variables have positive variance (i.e., the transformations are not constant). The optimal transformations $\phi(X)$ and $\psi(Y)$ are often interpretable and can lead to insight into the relationship between the variabls, which we now illustrate with the data Hair-Eye data set.

In order to use the ace function, it is necessary to "unpack" the table, i.e. to create variables X and Y which are matched up according to the bivariate observations summarized in the table. For the current table stored as

```
> haireye
      BLACK BROWN RED BLOND
brown    68   119  26     7
blue     20    84  17    94
hazel    15    54  14    10
green     5    29  14    16
```

we wrote a script to create two vectors of length 592:

```
# script file to create variables Eye and Hair
Eye = rep(NA,592); Hair = Eye
ind = 0
for(i in 1:4){
for(j in 1:4){
n = haireye[i,j]
```

```
Eye[(ind+1):(ind+n)] = i
Hair[(ind+1):(ind+n)] = j
ind = ind+n
}
}
```

Next, we run the ACE function:

```
> acehe = ace(Eye,Hair,cat=c(1,0))
> acehe$rsq
[1] 0.2087724
```

The "`cat=c(1,0)`" option specifies that the first column of Eye is to be treated as categorical (there is only one column) and that Hair (the "0") is to be treated as categorical. The numerical value is the estimated maximal correlation, which isn't very large, but probably significantly different from 0 (this is treated later).

What is of interest here is the plots of the estimated optimal transformations, shown below. Since numerical "scores" are assigned to the nominal values, there is a natural order. Further, the larger scores for the different variables tend to go together. So for hair color, we see BLACK has lowest score, followed in order by BROWN, RED (almost the same as BROWN), and BLOND. For eye color, we see that the ordering is brown ¡ hazel ¡ green ¡ blue. Further we now blue eyes and blond hair tend to go together, as well as brown eyes and black hair. The last numeric

		Hair Color				
Eye Color	Value	Black	Brown	Red	Blond	Row Sums
Brown	Obs.	68	119	26	7	220
Blue	Obs.	20	84	17	94	215
Hazel	Obs.	15	54	14	10	93
Green	Obs.	5	29	14	16	64
Column Sums		108	286	71	127	592

Table 1: The Hair/Eye Color cross tabulation.

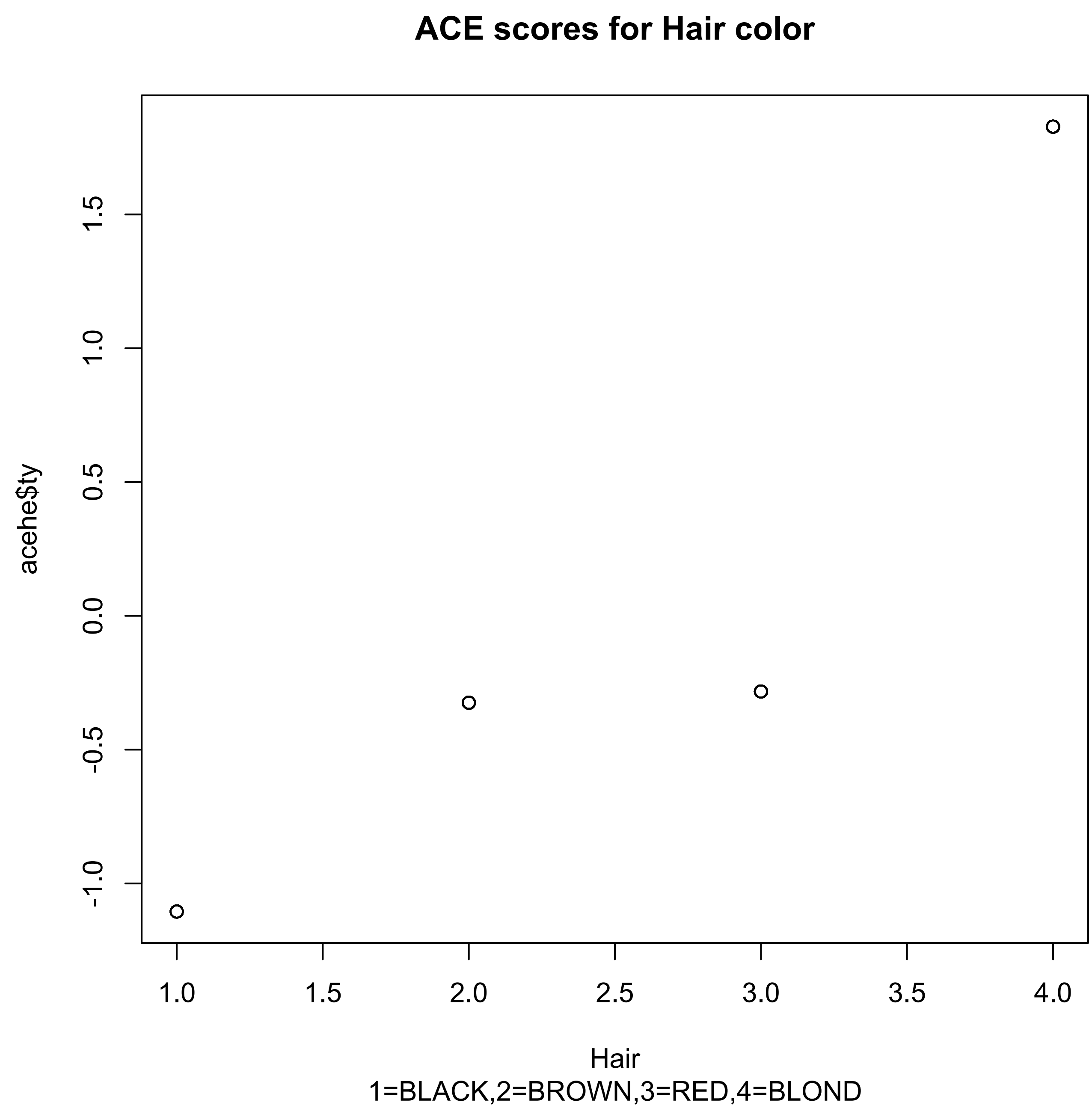

Figure 1: Plot of ACE transformation for Hair color from the Hair-Eye data.

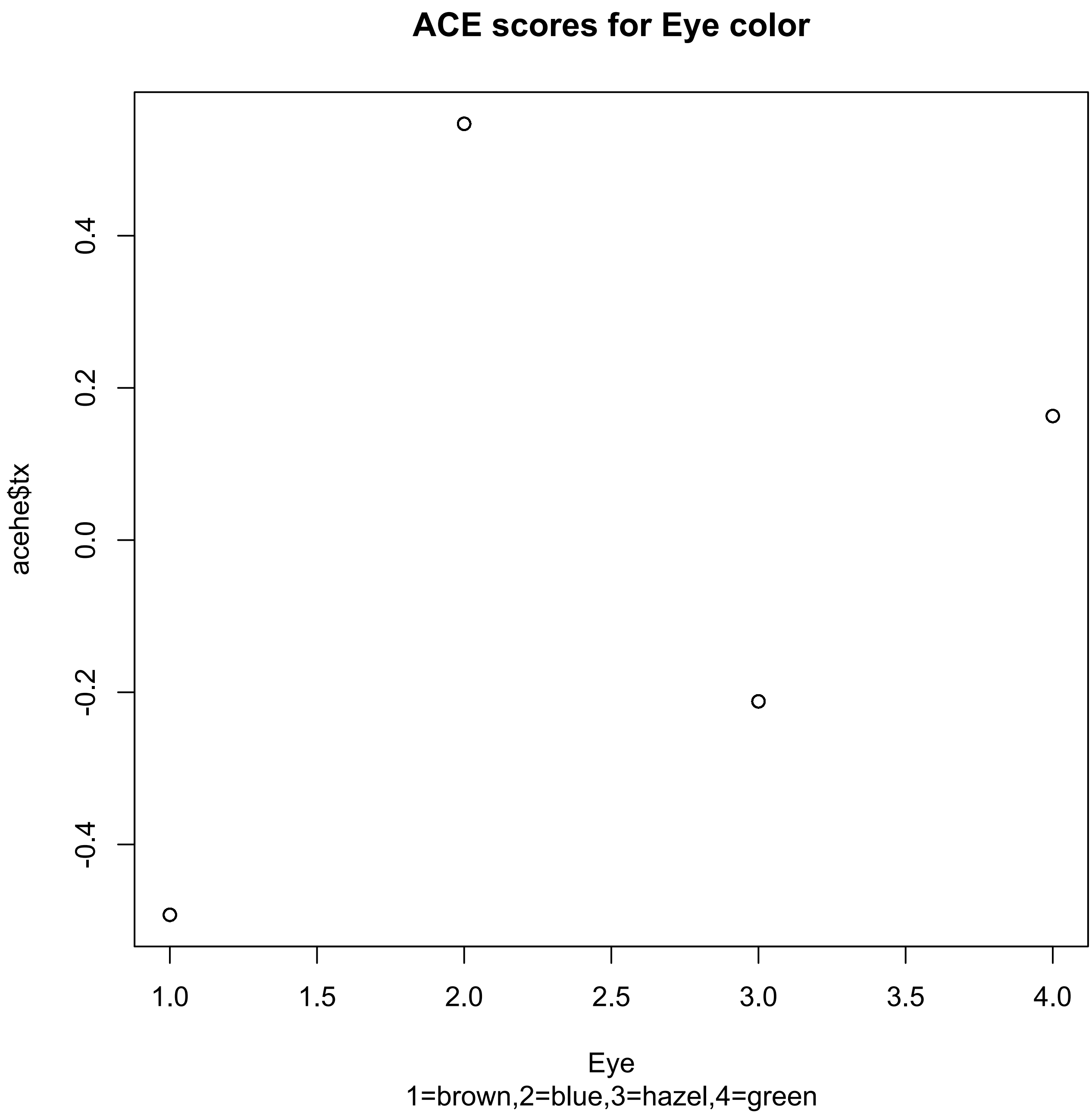

Figure 2: Plot of ACE transformation for Eye color from the Hair-Eye data.