

Department of Statistics

STATS 760: A Survey of Modern Applied Statistics

Assignment 1 2017

Question 1

Suppose we want to fit a linear model with four explanatory variables, two of which are factors (A and B say), and the others (X and Z) are continuous. The model formula is

$$Y \sim A + B*X + B*Z$$

Discuss how R would build up the model matrix in this case, illustrating your answer with some suitable artificial data. How are the coefficients interpreted? (assume the default contrasts are being used).

How would your answer change if instead of the default contrasts (i.e. using the contrast function `contr.treatment`) you used the "sum contrasts" (i.e. executed the code

```
options(contrasts=c("contr.sum", "contr.poly"))?
```

Question 2

Suppose we have 6 observations $y = (y_1, y_2, y_3, y_4, y_5, y_6)$. The first two have mean μ_1 , the next two have mean μ_2 , and the last two have mean μ_3 . Express the model in the form

$$E(y) = X\beta$$

(where X is a matrix with 6 rows and 3 columns and $\beta = (\beta_1, \beta_2, \beta_3)$ is a vector of three elements) in two different ways. Interpret the beta coefficients in terms of the means for each way.

Question 3.

In question 2, calculate the matrices X for each of the cases (i) using sum contrasts, and (ii) using treatment contrasts. Find a matrix T such that multiplication by T turns the "sum" coefficients into "treatment" coefficients.

Email answers to me by Friday March 24.