

MAST90104: A First Course in Statistical Learning

Week 8 Practical and Workshop

1 Practical questions

1. Consider the filter question in Week 7. Recall that we are interested in comparing the lifespan of 5 different types of filters. Six filters of each type are tested, and the time to failure in hours is given in the dataset (on the website) `filters` (in `csv` format).

Read the data. Then convert the `type` component into a factor. Recall that we fit a one-way classification model using the treatment contrast

```
> model <- lm(y~type, data=filters)
```

- (a) Calculate a 95% confidence interval for the difference in lifespan between filter types 3 and 4.
 - (b) Show that the hypothesis that the filters all have the same lifespan is testable.
 - (c) Test this hypothesis, using matrix theory.
 - (d) Test the same hypothesis using the `linearHypothesis` function from the `car` package.
 - (e) Repeat part d using the sum-to-zero contrast (`contr.sum`)
2. We study the effect of various breeds and diets on the milk yield of cows. A study is conducted on 9 cows and the following data obtained:

Breed	Diet		
	1	2	3
1	18.8	16.7	19.8
	21.2		23.9
2	22.3	15.9	21.8
		19.2	

- (a) Input this data into R. Plot an interaction plot between breed and diet.
- (b) Fit an additive model. What is the estimated amount of milk produced from breed 2 and diet 3 now?
- (c) Test the hypothesis (under the additive model) that the 2nd and 3rd diets are equivalent in terms of milk produced.
- (d) Find a 95% confidence interval, under the additive model, for the amount of milk produced from breed 2 and diet 3. Use both matrix calculations and the `estimable` function from the `gmodels` package.

2 Workshop questions

1. An industrial psychologist is investigating absenteeism among production-line workers, based on different types of work hours: (1) 4-day week with a 10-hour day, (2) 5-day week with a flexible 8-hour day, and (3) 5-day week with a structured 8-hour day. A study is conducted and the following data obtained of the average number of days missed:

	Work plan		
	1	2	3
Mean	9	6.2	10.1
Number	100	85	90

They also find $s^2 = 110.15$.

- (a) Test the hypothesis that the work plan has no effect on the absenteeism.
 - (b) Test the hypothesis that work plans 1 and 3 have the same rate of absenteeism.
2. Prove Theorem 6.2 using the following steps.
 - (a) Show that under the conditions of Theorem 6.1, the column space of XC is the same as the column space of X .
 - (b) Show that if two full-rank linear models have the same column space, the eigenvectors of their hat matrices are the same.
 - (c) Hence show that if the column space for two linear models is the same, the fitted values are the same.
 - (d) Complete the proof of Theorem 6.2.
 3. Consider question 2 in practical class this week, where we study the effect of various breeds and diets on the milk yield of cows. A study is conducted on 9 cows and the following data obtained:

Breed	Diet		
	1	2	3
1	18.8	16.7	19.8
	21.2		23.9
2	22.3	15.9	21.8
		19.2	

- (a) Express this as a two-factor model with no interaction in matrix form.
- (b) Express this as a two-factor model with interaction in matrix form.
- (c) Express the hypothesis that there is no interaction in terms of your parameters. Eliminate any redundancies.