# Tutorial Problems

## Question 1

A possible balanced incomplete block design with three treatments in three blocks of size two is

| | | |
|---|---|---|
| A | A | B |
| B | C | C |

and a possible randomized version is

| | | |
|---|---|---|
| B | C | A |
| C | A | B |

This question is about balanced incomplete block designs with four treatments in four blocks of size three:

(a) Similarly to above, construct a balanced incomplete block design with four treatments in four blocks of size three, that is $t = 4$, $b = 4$ and $k = 3$, using the four treatment labels A, B, C and D. That is complete the following BIBD:

| | | | |
|---|---|---|---|
| A | A | A | B |
| B | B | | |
| | | | |

(b) Explain how to randomize the design and illustrate this by showing a randomized version of your design in (a).

(c) Write out a model for the data from such a design stating the assumptions implicit in the model.

(d) Write down an explicit equation for the difference of the treatment means of A and B, that is $\bar{Y}_{A\bullet} - \bar{Y}_{B\bullet}$ and an explicit equation for the variance of $\bar{Y}_{A\bullet} - \bar{Y}_{B\bullet}$ using the non-randomised BIBD in (a).

(e) An R analysis of a data frame `dat` for data using your randomised design, constructed with data in the first column called `y` and two factors `block` and `treat` in the next two columns gave the following output:

```
anova(lm(y ~ block + treat, data=dat))
```

```
##             Df  Sum Sq Mean Sq F value  Pr(>F)
## block        * 10.4271  ******  4.3481 0.07369
## treat        * 22.4663  ******  ****** *******
## Residuals    *  3.9968  ******
```

```r
anova(lm(y ~ treat + block, data=dat))
```

```
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## treat        * 30.1893  ******  ****** ********
## block        *  2.7040  ******  1.1276 0.421589
## Residuals    *  3.9968  ******
```

(i) Why do the Treatment SS differ in the summary outputs of model `M1` and `M2`? Which of the models in R is more appropriate to use in this context?

(ii) What is the efficiency factor of the above BIBD?

(iii) For the model of your choice in i. fill in the missing values in the ANOVA table and test for the hypothesis that all four treatments have the same effect.

(iv) Provide a numerical value for the standard error of the difference of treatments A and B, that is $\hat{\alpha}_A - \hat{\alpha}_B$. You can use that $\text{Var}(\hat{\alpha}_A - \hat{\alpha}_B) = \sigma^2 \left(\frac{2k}{t\lambda}\right)$, where $\lambda$ is the number of times the pair AB occurs in the same block.

## Question 2

In a study on the use of drugs in the treatment of leprosy in the Philippines six sites where leprosy bacilli tend to congregate were selected. Drugs A and B were used in the study along with a placebo C and *ten* patients were selected for each treatment. To measure the treatment effect, a score representing the abundance of leprosy bacilli at these sites based on laboratory tests is measured before the experiment (initial score as $X$) and the score is measured again after several months of treatment (after treatment score as $Y$). The ANCOVA model is fitted to compare the treatment effects on the score after allowing for the inital score. The ANCOVA table is

| Source of Variation | d.f. | S.S. |
|---|---|---|
| Regression | | 802.9 |
| Drugs (adjusted for $X$) | | 68.6 |
| Residual | | 417.2 |
| Total | | 1288.7 |

(a) Complete the above table and test for drug effects.

(b) The $X$ and $Y$ totals were

| Drug | A | B | C | Total |
|---|---|---|---|---|
| $X$ | 93 | 100 | 129 | 322 |
| $Y$ | 53 | 61 | 123 | 237 |

Calculate the 1-way ANOVA table with $Y$ as the outcome variable ignoring the initial score $X$ values.

(c) Calculate the 1-way ANOVA table for testing initial scores $X$ differences between the drug groups if the initial score $X$ total sum of squares is 665.9. Are there significant differences in the initial scores between the three drug groups?

(d) Calculate a 95% confidence interval for the difference in the effects of drugs $A$ and $B$ using the analysis of covariance model, given that the regression coefficient is estimated to be $\hat{\beta} = 0.987$.

(e) (Extra) Write down the anova table for the model `Score` $\sim$ `Drugs + InitialScores` which swap the order of `InitialScores` and `Drugs`.

# Computer Problems

## Question 1

% New Question This problem concerns a study on the absorption over time of rubidium and bromide ions in potato slices. Three variables are measured on each slice:

| Variable | Description |
|---|---|
| `Absorption` | amount of ion absorbed in the tissue |
| `Duration` | time of immersion (in hours) in ion solution |
| `Ions` | Rubidium, R, or Bromide, B |

The data is from *A Handbook of Small Data Sets* by Hand et al. (1993) and is stored in `potato.txt`.

(a) Fit the following linear models:

- M1: A simple linear regression of ‘Absorption‘ (the response) on ‘Duration‘.
- M2: Parallel linear regression of ‘Absorption‘ (the response) on ‘Duration‘ for each type of ion.
- M3: Separate linear regressions of ‘Absorption‘ (the response) on ‘Duration‘ for each type of ion.

(b) Compare the three models in (a) using suitable F tests (implemented using the `anova()` command in R. You should find that these tests indicate clearly that model M2 (parallel regressions) is adequate.

(c) Check model diagnostics for M2 and comment briefly

(d) Produce a scatterplot of the data using different colours and plotting symbols for the two ions:

```
plot(Absorption ~ Duration, data=dat, col=as.numeric(Ions))
```

Add the fitted regression lines from $M2$ to this plot. Use colours for these lines to match the plotted data points. (Hint: use the `abline()` command twice. You will need to work out the appropriate intercept and slope for each model as input for the `abline()` command.

(e) Fit model M3 by hand, i.e. create a new design matrix $\mathbf{X}$ and programme the least squares parameter estimators with the well known formula

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

## Question 2

A chemical engineer wished to evaluate the effectiveness of nine alternative formulations of a dishwashing detergent in terms of the extent to which each would maintain foam or suds while in use. Three sinks were available, and three people were instructed to use the sinks to wash plates at a constant rate. Each block consisted of three experimental units, where the experimental unit was a sink with a fixed amount of clean water and a fixed amount of soil added. Three detergent formulations were randomly assigned to the three sinks in each block. The response Y was foam duration, which was measured by the number of plates washed before the suds disappeared. A BIBD (number 18 from Table 28.1 in Kutner 5e) was utilized for this experiment. Data for the randomized BIBD follow:

 

| | Treatments | | | Responses | | |
|---|---|---|---|---|---|---|
| Block | Sink 1 | Sink 2 | Sink 3 | Sink 1 | Sink 2 | Sink 3 |
| 1 | 3 | 8 | 4 | 13 | 20 | 7 |
| 2 | 4 | 9 | 2 | 6 | 29 | 17 |
| 3 | 3 | 6 | 9 | 15 | 23 | 31 |
| 4 | 9 | 5 | 1 | 31 | 26 | 20 |
| 5 | 2 | 7 | 6 | 16 | 21 | 23 |
| 6 | 6 | 5 | 4 | 23 | 26 | 6 |
| 7 | 9 | 8 | 7 | 28 | 19 | 21 |
| 8 | 7 | 1 | 4 | 20 | 20 | 7 |
| 9 | 6 | 8 | 1 | 24 | 19 | 20 |
| 10 | 5 | 8 | 2 | 26 | 19 | 17 |
| 11 | 5 | 3 | 7 | 24 | 14 | 19 |
| 12 | 3 | 2 | 1 | 11 | 17 | 19 |

To read this data in R, complete the following R-code:

```
Y <- c(13, 6, 15, 31, 16, 23, 28, 20, 24, 26, 24, 11,
20, 29, 23, 26, 21, 26, 19, 20, 19, 19, 14, 17,
7, 17, 31, 20, 23, 6, 21, 7, 20, ??, ??, ??)
tmt <- factor(c(3, 4, 3, 9, 2, 6, 9, 7, 6, 5, 5, 3,
?, ?, ?, ?, 7, 5, 8, 1, 8, 8, 3, 2,
4, 2, 9, 1, 6, 4, 7, 4, 1, 2, 7, 1))
block <- factor(c(rep(?:??, ?)))
```

(a) Run a main effects ANOVA with blocking variable Block and treatment variable Detergent.

(b) Plot residuals of the model in (a) versus the fitted values and check the normality assumption of the errors with a boxplot and a Q-Q plot of the residuals.

(c) Give a 95% confidence interval for the fifth treatment effect.

(d) Analyze the nature of the treatment effects by making all pairwise comparisons among the treatment differences (= differences of the estimated parameters). Use Tukey's, Scheff'e's and Bonferroni's method for multiple testing. (Remark: this question requires some skilled Rprogramming, e.g. use `outer(b,b,"-")` to calculate all pairwise differences of the components of vector b. Further, a matrix with standard errors for the (pairwise) contrasts has to be generated.).

 