

Monday, November 21, 11:59PM, BbLearn groups (99 points scaled to 100 percent)

You may discuss this assignment with whomever you wish, but please prepare and submit work in groups of **ONE to THREE** students, no more and no fewer. **Each group** will submit a copy of their group's completed assignment, via BbLearn, including the **names and student ID numbers of all group members** who participated on the assignment. If you discover a mistake, you may submit another version before the deadline. The last submitted (on-time) version will be graded, with all team members receiving the same score, which will be recorded in BbLearn.

GROUP MEMBERS WHO DO NOT CONTRIBUTE SUBSTANTIALLY TO AN ASSIGNMENT MAY BE REQUIRED TO WORK IN THEIR OWN GROUP OF ONE FOR THE REMAINDER OF THE SEMESTER.

While you are permitted to discuss the assignment with other groups, please prepare your own group's code/output and written answers. GROUPS WHOSE CODE AND SOLUTIONS APPEAR SUBSTANTIALLY SIMILAR MAY BE SUBJECT TO A 10% PENALTY.

Please prepare solutions in a **neat, organized and concise fashion!** I prefer typeset presentations (e.g., cut and paste code/output into MS Word with added exposition when appropriate; knitr via EMACS and ESS; knitr or R Markdown via RStudio, the latter being the method preferred by students in recent years). At the very least, you need to ensure code and output are presented with a fixed-width font. Neatly handwritten presentations may also be appropriate for some problems. Sloppily prepared or disorganized solutions will not receive full credit.

To complete the items below, I expect you to find and use material in our lecture notes, including code/output, possibly after some modification. Some questions may be answered with code and output alone, but some exposition may be required beyond code and output for other questions. It's up to you to communicate concisely!

The data for this homework were obtained from a randomized experiment to estimate the effect of a diet (Diet) on the concentration of a substance (Conc) in the blood of participants (no units given). (This is similar to the diet example in our notes.) The data are contained in the file `hw6.rds` in BbLearn and are shown in the table, below. We will conduct analyses of these data using the cell means model and the factor effects models with reference treatment coding and sum-to-zero coding. Use `readRDS(file='hw6.rds')` to read the data frame into R.

| Conc | Diet |
|------|---------|
| 66 | D |
| 90 | D |
| 87 | D |
| 76 | A |
| 106 | A |
| 73 | A |
| 69 | B |
| 71 | B |
| 36 | B |
| 74 | C |
| 66 | C |
| 55 | C |
| 61 | E |
| 66 | E |
| 71 | E |
| 70 | Control |
| 77 | Control |
| 63 | Control |
| 98 | Control |
| 118 | Control |
| 107 | Control |
| 106 | Control |
| 115 | Control |
| 129 | Control |
| 102 | Control |
| 101 | Control |
| 73 | Control |
| 105 | Control |
| 100 | Control |
| 113 | Control |

1. What is the factor in this experiment? (1 point)
2. Give at least one other name for a factor. (1 point)
3. What are the levels of the factor? (1 point)
4. What are the treatments? (1 point)
5. Make sure the Diet variable is seen as a factor variable by **R**. (2 points)
6. It seems reasonable to have the **Control** level be the reference level, at least for the factor effects model with reference treatment coding. Use the **relevel** function in **R** to make the **Control** level the reference level. Make this change to the **Diet** factor,

- keeping remaining levels in their same relative order. (We will use treatment coding, shortly.) Be sure to use the so releveled `Diet` factor throughout the remainder of this homework; replacing the existing `Diet` factor in the data frame will help to ensure this. Show your code and output to convince me that you've reordered the levels as requested. (3 points)
7. Create a scatter plot (dot plot) of concentration level (vertical) by factor level. We did this in our notes with a convenient function in R. Show your code and plot. (5 points)
 8. (a) Compute a one-way ANOVA (i.e., one-factor linear model) using the cell means model and test for ("overall") equality of factor level means ($\alpha = 0.05$). Be sure to state null and alternative hypotheses, report a test statistic, p-value and state your conclusions. In particular, what are the null and alternative hypotheses in terms of $\mathbf{C}\beta$? Etc. You may use `stats::anova` or `gmodels::glh.test`, as in our notes. (Remember, by default, R reports results that are not likely of interest when using the cell means model. You'll have to fix this as we did in our notes.) Also, show your code and output. (10 points)
 - (b) Report the factor level (i.e., treatment) estimated mean concentration, $\hat{\mu}_i$, $i = 1, \dots, a$ (in $\hat{\beta}$). (5 points)
 - (c) It seems plausible that our fellow researchers may have wanted to compare the control group to the remaining groups. Using the cell means model, compare the mean of the control group with the average of the means in the remaining groups. In particular, construct a 95% confidence interval for the difference between the mean concentration of the control group and the average of the mean concentrations of the remaining groups. Report your code/output and summarize your interval very briefly. (10 points)
 9. (a) Change to the factor effects model with sum-to-zero coding and report code/output to convince me that you have done this. (5 points)
 - (b) Repeat 8a, above, now using the factor effects model with sum-to-zero coding. Be sure to report your (new) \mathbf{C} matrix. (Note: R's default behavior should be what we want now that we do not omit the overall constant ("intercept") from our model. That is, its overall F test and R^2 are more likely to be of interest. Still, report a \mathbf{C} matrix and use `stats::anova` or `gmodels::glh.test` as in 8a, above.) (10 points)
 - (c) Repeat 8b, above, now using the factor effects model with sum-to-zero coding. In particular, I want $\hat{\beta}$ for the factor effects model with sum-to-zero coding, of course! (Be sure to compute $\hat{\alpha}_a$, too!) *ALSO*, compute $\widehat{\mu + \alpha_i}$, $i = 1, \dots, a$; compare these a values with the estimated means computed using the cell means model in 8b, above. (5 points)
 - (d) Repeat 8c, above, now using the factor effects model with sum-to-zero coding. (10 points)

10. (a) Change to the factor effects model with (reference) treatment coding, with the control group as the reference level. Report code/output to convince me that you have done this correctly. (5 points)
- (b) Repeat 8a, above, now using the factor effects model with treatment coding. Be sure to report your (new) \mathbf{C} matrix. (Note: R's default behavior should be what we want now that we do not omit the overall constant from our model. Still, report a \mathbf{C} matrix and use `stats::anova` or `gmodels::glh.test` as in 8a, above.) (10 points)
- (c) Repeat 8b, above, now using the factor effects model with treatment coding. In particular, I want $\hat{\beta}$ for the factor effects model with treatment coding, of course! (Be sure to give $\hat{\alpha}_1$, too!) *ALSO*, compute $\widehat{\mu + \alpha_i}$, $i = 1, \dots, a$; compare these a values with the estimated means computed using the cell means model in 8b, above. (5 points)
- (d) Repeat 8c, above, now using the factor effects model with treatment coding. (10 points)