

Exercise One

Will Barnard

16 01 2026

Format your (reproducible) answers using R Markdown and submit on Moodle as a PDF.

You can obtain the data for this assignment using:

```
library(faraway)
data(truck)
head(truck)
```

```
  B C D E O height
1 - - - - -   7.78
2 + - - + -   8.15
3 - + - + -   7.50
4 + + - - -   7.59
5 - - + + -   7.94
6 + - + - -   7.69
```

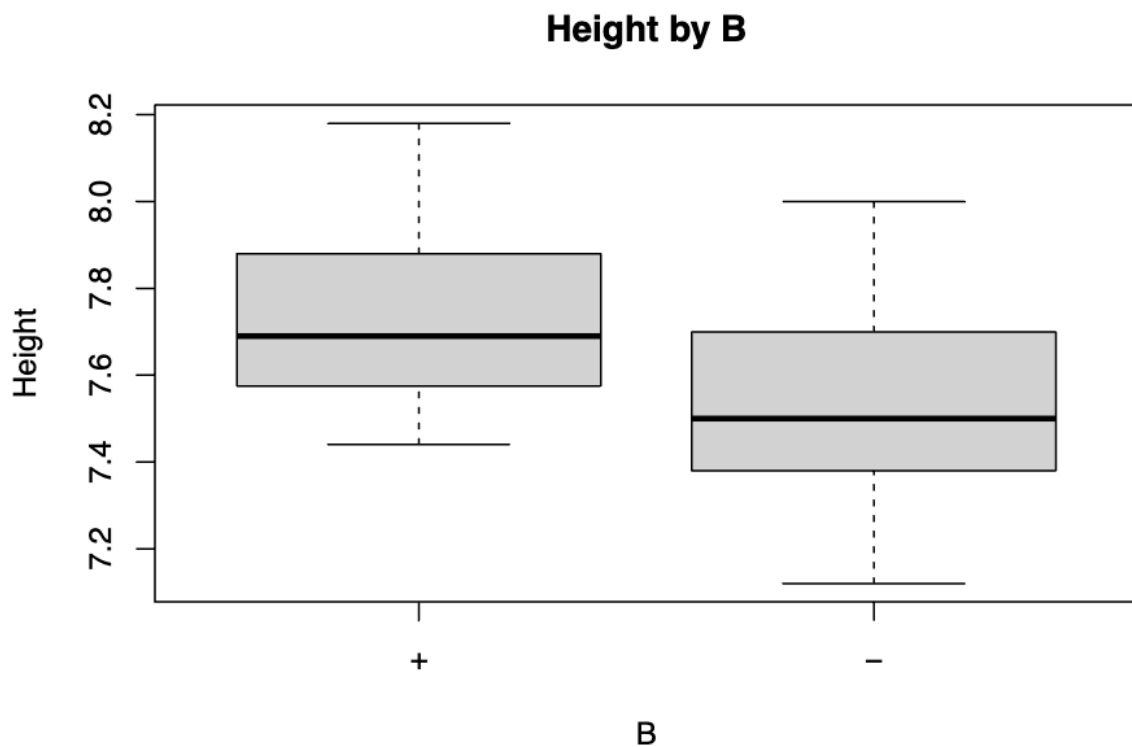
You will need to have installed the `faraway` package for this to work. Read the help page for `truck` for information about the dataset.

Answer the following questions:

Q1

Make a boxplot showing how height varies for the two levels of B.

```
plot(height ~ B, truck,
     main="Height by B",
     xlab="B",
     ylab="Height"
)
```



Q2

Fit a linear model with height as the response and all five factors as predictors. Report the regression coefficients.

```
model = lm(height ~ B + C + D + E + O, truck)
coef(model)
```

```
(Intercept)      B-      C-      D-      E-      O-
  7.5662500  -0.2212500  0.1762500  0.0287500 -0.1037500  0.2595833
```

Q3

Display the first six lines of the design matrix. Identify the coding used for the two levels of each factor.

```
X = model.matrix(model)
head(X)
```

```
(Intercept) B- C- D- E- O-
1           1  1  1  1  1  1
2           1  0  1  1  0  1
3           1  1  0  1  0  1
4           1  0  0  1  1  1
```

```
5      1  1  1  0  0  1
6      1  0  1  0  1  1
```

For each factor, the reference level "-" is coded as 0, and the "+" level is coded as 1.

Q4

Change the coding of all the factors to have reference level -. You will find the `relevel` function useful. Refit the model and compute the regression coefficients. Comment on how they compare to the previously computed coefficients.

```
truck$B = relevel(truck$B, ref = "-")
truck$C = relevel(truck$C, ref = "-")
truck$D = relevel(truck$D, ref = "-")
truck$E = relevel(truck$E, ref = "-")
truck$O = relevel(truck$O, ref = "-")

model_new = lm(height ~ B + C + D + E + O, data = truck)
coef(model_new)
```

```
(Intercept)      B+      C+      D+      E+      O+
  7.7058333   0.2212500 -0.1762500 -0.0287500  0.1037500 -0.2595833
```

These coefficients are identical to the previous ones.

Q5

Display the new design matrix and compare to before.

```
X_new = model.matrix(model_new)
head(X_new)
```

```
(Intercept) B+ C+ D+ E+ O+
1      1  0  0  0  0  0
2      1  1  0  0  1  0
3      1  0  1  0  1  0
4      1  1  1  0  0  0
5      1  0  0  1  1  0
6      1  1  0  1  0  0
```

Again, the design matrix is identical to the previous one.

Q6

Which observed combination of the predictors gives a predicted response close to the target of 8?

```
idx = which.min(abs(predict(model) - 8))
truck[idx, ]
```

```

  B C D E O height
2 + - - + - 8.15

```

```
predict(model)[idx]
```

```

      2
8.030833

```

```

## so the combination of the predictors that gives a predicted response close to the target of 8 is:
# B = +, C = -, D = -, E = +, O = -

```

Q7

There are 3 replicates of each observed combination of the predictors. Compute the SD within each set of three replicates. (This is easier once you notice the repeated pattern in the design.) What does this say about the linear model assumptions?

```

group_id = apply(truck[, c("B","C","D","E","O")], 1, paste, collapse = "_")

sd_values = tapply(truck$height, group_id, sd)
sd_values

```

```

-_-_-_-_- -_-_-_-_+ -_-_+_+_- -_-_+_+_+ -_+_+_+_- -_+_+_+_+ -_+_+_+_-
0.01732051 0.19313208 0.06000000 0.06928203 0.03464102 0.03464102 0.09165151
-_+_+_+_- +_-_-_-_- +_-_-_-_- +_-_-_-_- +_-_-_-_- +_-_-_-_- +_-_-_-_-
0.04041452 0.16522712 0.25403412 0.22278540 0.06506407 0.10214369 0.09609024
+_+_+_+_- ++_+_+_+
0.12503333 0.15947832

```

```

## The relatively low and stable sd's across groups indicates that the linear model is
##in fact a good fit for the model.

```