Lab 5

AA 501

Kristen Fitzgerald

2024-07-01

Question 1

Data were collected in an effort to determine whether different dose levels of a given drug have an effect on blood pressure for people with one of three types of heart disease. To obtain this data, submit the following code:

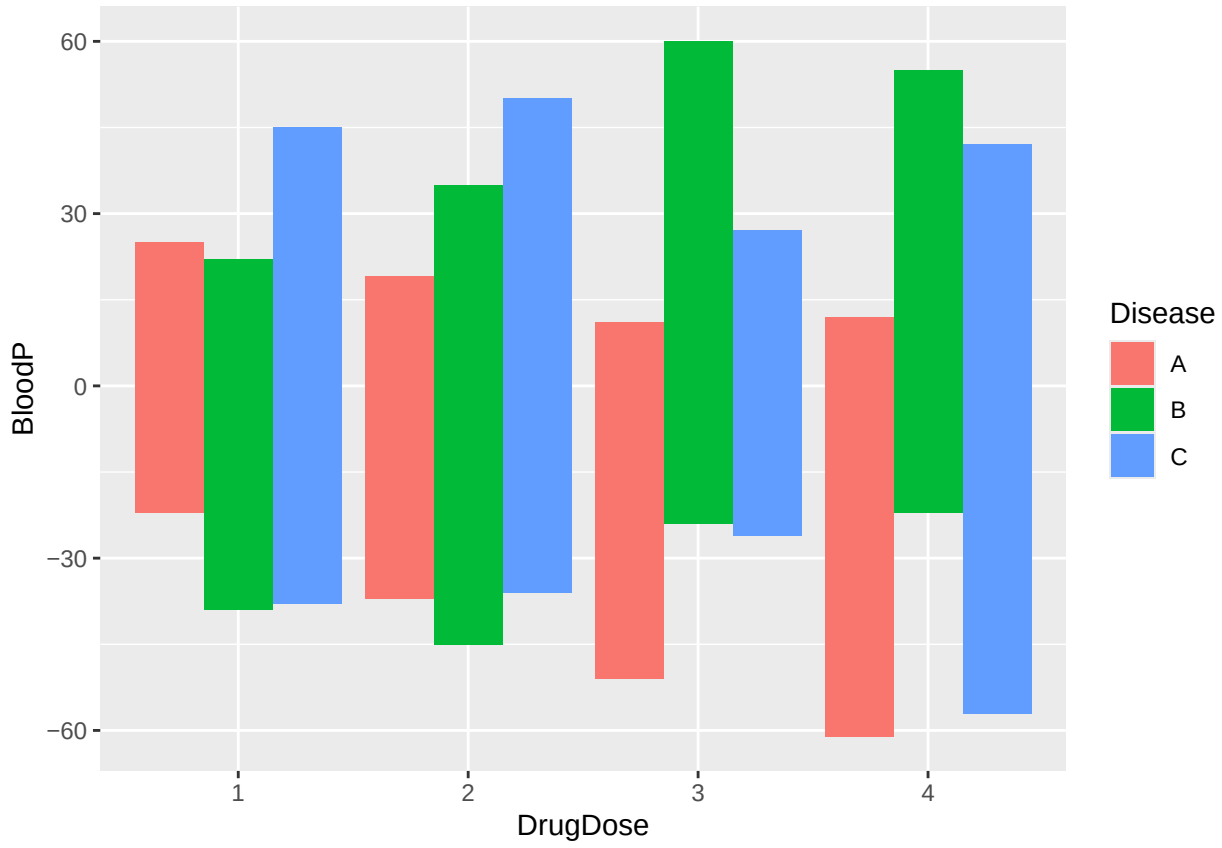
This dataset has the following variables:

- `DrugDose`: dosage level of drug (1, 2, 3, 4), corresponding to (Placebo, 50mg, 100mg, 200mg)
- `Disease`: heart disease category
- `BloodP`: change in diastolic blood pressure after 2 weeks of treatment

a. Examine the data with a side by side bar chart. Put `BloodP` on the Y axis, `DrugDose` on the X axis, and stratify by `Disease`. What information can you obtain from looking at the data?

Solution:

```
ggplot(drugdose, aes(x=DrugDose, y=BloodP, fill = Disease)) +
    geom_bar(stat = "identity", position = position_dodge())
```


60
30
0
−30
−60
BloodP
1
2
3
4
DrugDose
Disease
A
B
C

b. Run a 2-Way ANOVA, making sure to include an interaction term if the graphical analyses that you performed previously indicate that might be advisable (HINT: Make sure `DrugDose` is a factor in your model since it is stored as a number). What conclusions can you reach at this point?

Solution: We see that the interaction between `DrugDose` and `Disease` is statistically significant. Meaning that the impact of `Disease` on `BloodP` has differs across levels of `DrugDose`.

```
drug.anova <- aov(BloodP ~ DrugDose*Disease, data=drugdose)
summary(drug.anova)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## DrugDose            3     54      18   0.042    0.989
## Disease             2  19276    9638  22.275 3.01e-09 ***
## DrugDose:Disease    6  17146    2858   6.604 3.02e-06 ***
## Residuals         158  68366     433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. If an interaction is required in the previous piece, investigate the differences in drug dose by performing a sliced ANOVA across the levels of heart disease.

Solution: Based on the slicing analysis we find that `DrugDose` is statistically significant for patients with Diseases A and B.

```
sliced_aov <- drugdose %>%
  group_by(Disease) %>%
  nest() %>%
  mutate(aov = map(data, ~summary(aov(BloodP ~ DrugDose, data = .x))))

print(sliced_aov$aov)
```

```
## [[1]]
##             Df Sum Sq Mean Sq F value  Pr(>F)
## DrugDose     3  10561    3520   7.549 0.00027 ***
## Residuals   53  24716     466
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## [[2]]
##             Df Sum Sq Mean Sq F value  Pr(>F)
## DrugDose     3   6320  2106.7   6.137 0.00112 **
## Residuals   55  18881   343.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## [[3]]
##             Df Sum Sq Mean Sq F value Pr(>F)
## DrugDose     3    468   156.0   0.315  0.814
```

```
## Residuals   50  24770   495.4
```

Question 2

A computer service center has four technicians who specialize in repairing three brands of computer disk drives for desktop computers. The service center wants to study the effects of the technician and brand of the disk drive on the service time.

This dataset has the following variables:

- `Technician`: name of the technician (Angela, Bob, Justin, or Karen)
- `Brand`: brand of disk drive (1, 2, or 3)
- `Time`: time for repair (in minutes)

a. Generate a 2-Way ANOVA with `Time` as the dependent variable and `Technician` and `Brand` as the independent variables (HINT: Make sure `Brand` is a factor in your model since it is stored as a number). Include the interaction between the independent variables in your model. Assume a level of significance of 0.05. Is the overall F test significant in your model? Is there a significant interaction?

Solution:

```
disk.aov <- aov(Time ~ Technician*Brand, data = disks)
summary(disk.aov)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Technician        3  14798    4933  29.151 5.15e-13 ***
## Brand             2    343     172   1.014    0.367
## Technician:Brand  6   7907    1318   7.789 1.12e-06 ***
## Residuals        84  14213     169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b. Is it appropriate to examine the tests for the main effects shown in the output?

Solution:

c. Determine whether there are differences between the technicians for each brand of disk drive using sliced ANOVA across the brands of disk drives. Also, examine the Tukey HSD among the values of `Technician` within each slice of `Brand`. What are your conclusions?

Solution:

```r
sliced_aov_disk <- disks %>%
  group_by(Technician) %>%
  nest() %>%
  mutate(aov = map(data, ~summary(aov(Time ~ Brand, data = .x))))

print(sliced_aov_disk$aov)
```

```
## [[1]]
##             Df Sum Sq Mean Sq F value Pr(>F)
## Brand        2   1397   698.3   5.128 0.0154 *
## Residuals   21   2859   136.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## [[2]]
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Brand        2   3367  1683.5    10.5 0.00069 ***
## Residuals   21   3367   160.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## [[3]]
##             Df Sum Sq Mean Sq F value Pr(>F)
## Brand        2     28    14.0   0.067  0.935
## Residuals   21   4382   208.7
##
## [[4]]
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Brand        2   3459  1729.5   10.07 0.000857 ***
## Residuals   21   3606   171.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```