

Homework 3

2024-09-27

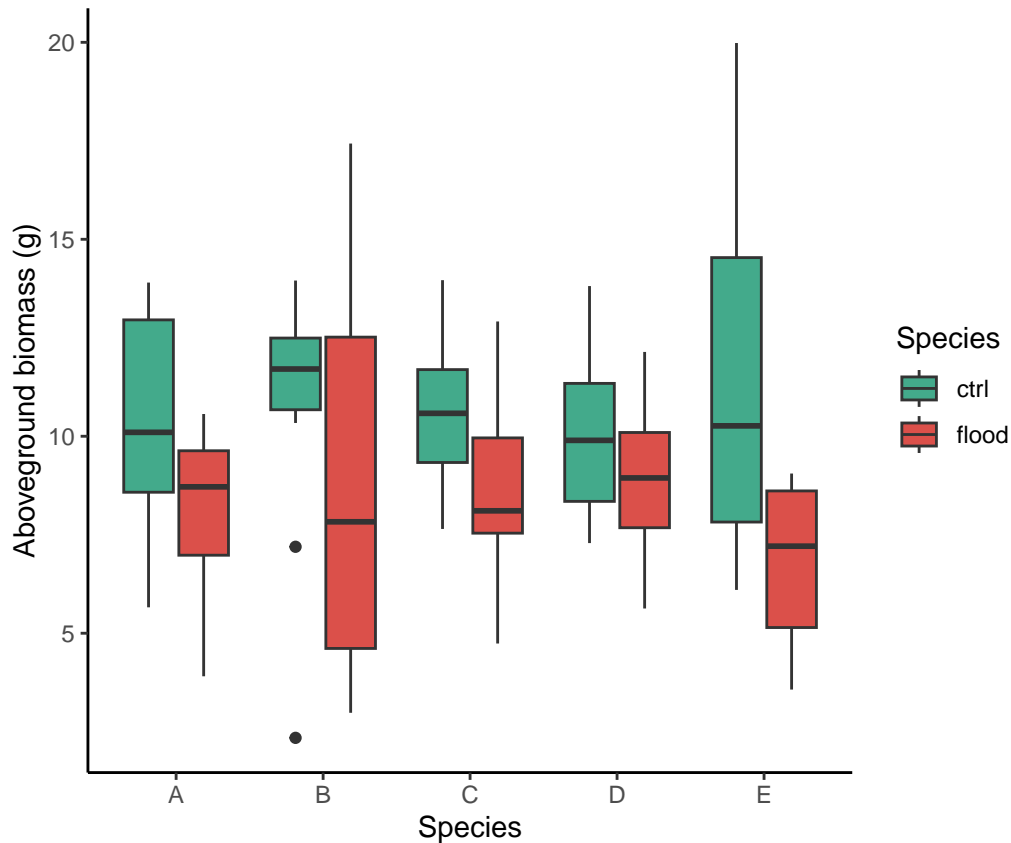
Read the data in the chunk below. The data come from of a study that investigated the flood tolerance of three different clover species.

The data can be interpreted as arising from an experiment with a 5×2 factorial treatment structure in a completely randomized design with 24 repetitions. The experimental units were pots in the greenhouse. The researchers cultivated the pots under two growing conditions: one was normally irrigated (control treatment), and one was left 12 days under flooded conditions. After the flooding, the researchers measured the aboveground biomass (in grams). The objective of the study is to find out if the flooding treatment affects the plant biomass, and if this effect is different for different species.

```
library(tidyverse)
```

```
url <- "https://raw.githubusercontent.com/jlacasa/stat705_fall2024/main/classes/data/lotus_hw3.csv"
dd <- read.csv(url)
```

```
dd %>%
  ggplot(aes(species, agb_g))+
  geom_boxplot(aes(fill = trt))+
  scale_fill_manual(values = c("#43AA8B", "#DB504A"))+
  labs(x = "Species",
       y = "Aboveground biomass (g)",
       fill = "Species")+
  theme_classic()+
  theme(aspect.ratio = 1)
```



a. Propose a statistical model (using mathematical notation) to describe the relationship between aboveground biomass, legume species and treatment. Fit that model to the data.

We can describe the data using the model

$$y_{ijk} = \mu + \tau_i + \rho_j + (\tau\rho)_{ij} + \varepsilon_{ijk},$$

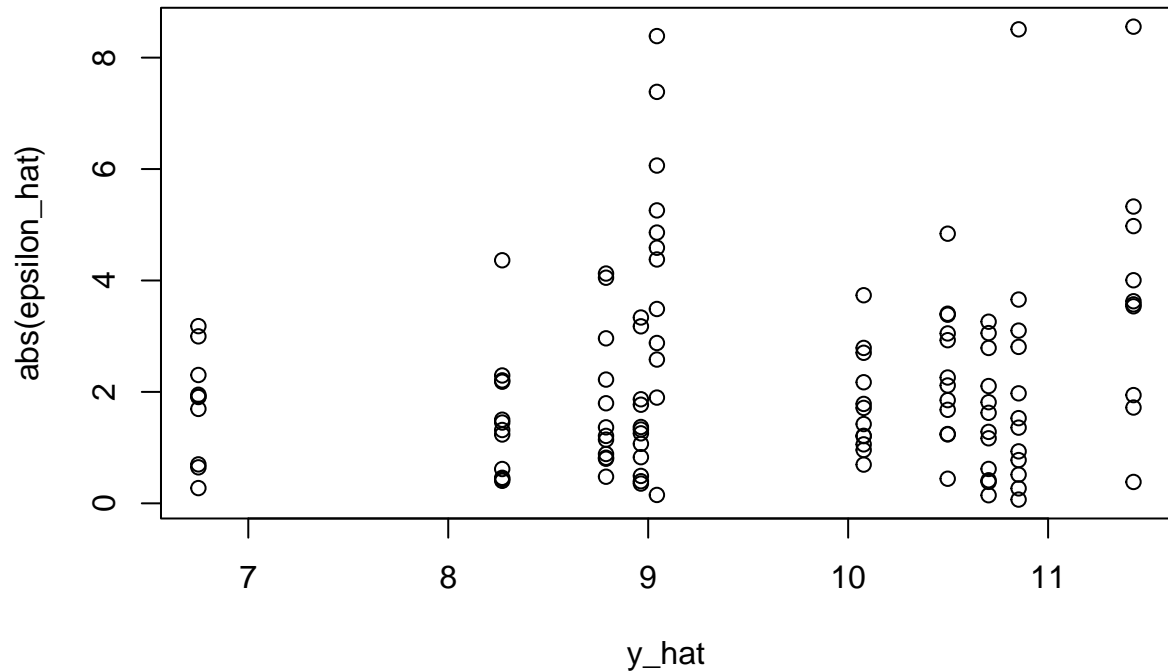
where y_{ijk} is the observation of biomass for the i th treatment ($i = 1, 2$), the j th species ($j = 1, 2, \dots, 5$) and k th repetition, μ is the overall mean, τ_i is the effect of the i th treatment, ρ_j is the effect of the j th species, $(\tau\rho)_{ij}$ is the interaction between the i th treatment and j th species, and ε_{ijk} is the error for the i th treatment ($i = 1, 2$), the j th species ($j = 1, 2, \dots, 5$) and k th repetition. We assume that ε_{ijk} is i.i.d. $\sim N(0, \sigma^2)$.

```
m1 <- lm(agb_g ~ trt*species, data = dd)
```

b. Mention the assumptions in your model and check them. If you believe one of the assumptions is badly wrong and should be changed, name a possible alternative. (Note: you only have to state if you think they are approximately appropriate or not, and name a possible alternative. You do not need to do it!)

The deterministic part of the model

```
y_hat <- m1$fitted.values
epsilon_hat <- m1$residuals
plot(y_hat, abs(epsilon_hat))
```



Constant variance

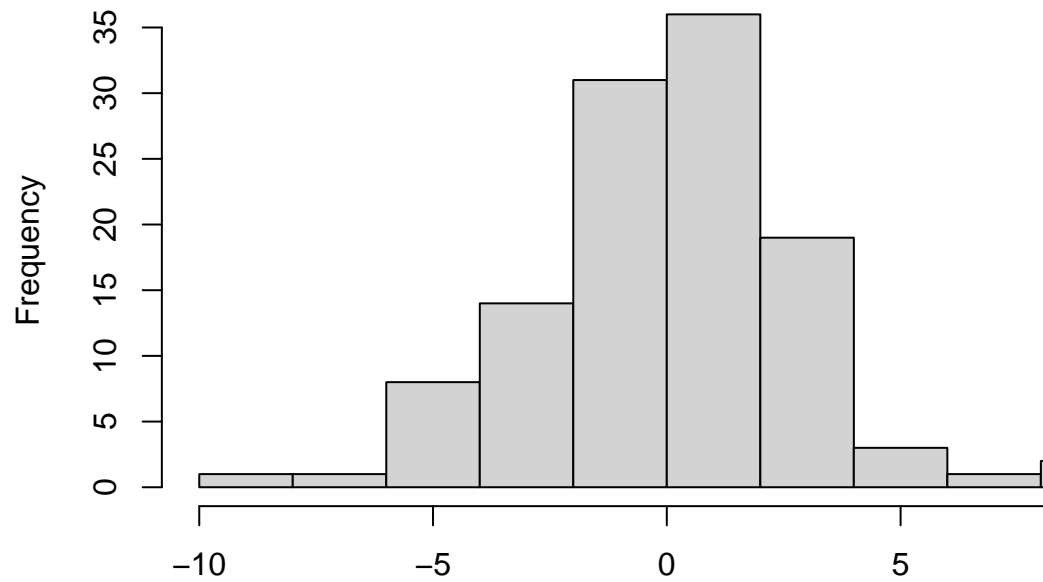
```
summary(lm(abs(epsilon_hat) ~ y_hat))
```

```
##
## Call:
## lm(formula = abs(epsilon_hat) ~ y_hat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4061 -1.1775 -0.4944  0.8293  6.2516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.4417     1.1974   0.369   0.713
## y_hat         0.1872     0.1242   1.508   0.134
##
## Residual standard error: 1.747 on 114 degrees of freedom
## Multiple R-squared:  0.01955,    Adjusted R-squared:  0.01095
## F-statistic: 2.273 on 1 and 114 DF,  p-value: 0.1344
```

Independent residuals

```
hist(epsilon_hat)
```

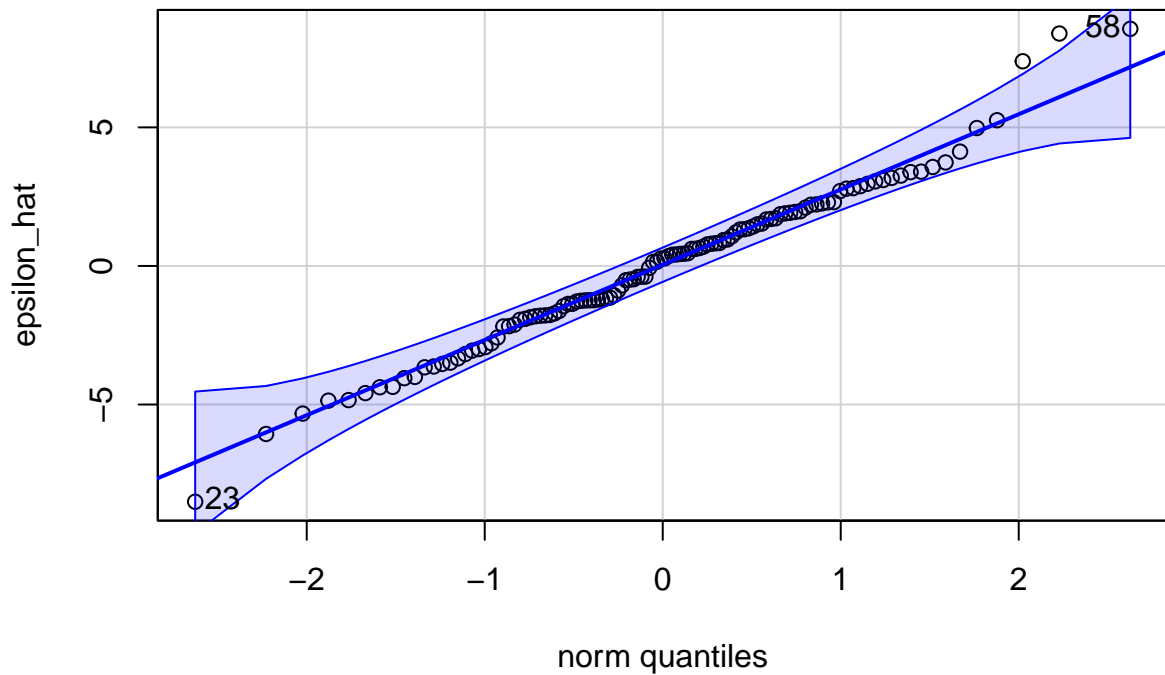
Histogram of epsilon_hat



Normally sitributed residuals

epsilon_hat

```
car::qqPlot(epsilon_hat)
```



```
## [1] 58 23
```

```
shapiro.test(epsilon_hat)
```

```
##  
## Shapiro-Wilk normality test  
##
```

```
## data: epsilon_hat
## W = 0.98512, p-value = 0.2298
```

c. Provide point estimates and 95% confidence intervals of the different species under both treated and control conditions.

```
newdata <- expand.grid(trt = c("ctrl", "flood"),
                      species = c("A", "B", "C", "D", "E"))
```

```
newdata %>%
  bind_cols(as.data.frame(predict(m1, newdata = newdata, interval = "confidence"))) %>%
  mutate(across(fit:upr, ~round(., 2)))
```

```
##      trt species   fit lwr  upr
## 1  ctrl      A 10.50 8.80 12.19
## 2 flood      A   8.27 6.57  9.97
## 3  ctrl      B 10.85 9.16 12.55
## 4 flood      B   9.04 7.35 10.74
## 5  ctrl      C 10.70 9.01 12.40
## 6 flood      C   8.79 7.09 10.49
## 7  ctrl      D 10.08 8.38 11.77
## 8 flood      D   8.96 7.27 10.66
## 9  ctrl      E 11.43 9.57 13.29
## 10 flood     E   6.75 4.89  8.61
```