Homework 10

Kevin Linares and Jamila Sani

2024-12-02

Use rats data in the faraway package.

- Description: An experiment was conducted as part of an investigation to combat the effects of certain toxic agents.

- Format: A data frame with 48 observations on the following 3 variables:

- time: survival time in tens of hours

- poison: the poison type - a factor with levels I, II, and III randomly assigned to subjects

- treat: the treatment - a factor with levels A, B, C, and D randomly assigned to subjects

```
data(rats)
summary(rats)
```

```
      time          poison   treat
 Min.   :0.1800   I  :16    A:12
 1st Qu.:0.3000   II :16    B:12
 Median :0.4000   III:16    C:12
 Mean   :0.4794             D:12
 3rd Qu.:0.6225
 Max.   :1.2400
```

```
glimpse(rats)
```

```
Rows: 48
Columns: 3
$ time   <dbl> 0.31, 0.82, 0.43, 0.45, 0.45, 1.10, 0.45, 0.71, 0.46, 0.88, 0.6~
$ poison <fct> I, I, I, I, I, I, I, I, I, I, I, I, I, I, I, I, II, II, II, II,~
$ treat  <fct> A, B, C, D, A, B, C, D, A, B, C, D, A, B, C, D, A, B, C, D, A, ~
```

```
rats %>%
  tabyl(poison,treat) # janitor::tabyl()
```

```
 poison A B C D
      I 4 4 4 4
     II 4 4 4 4
    III 4 4 4 4
```
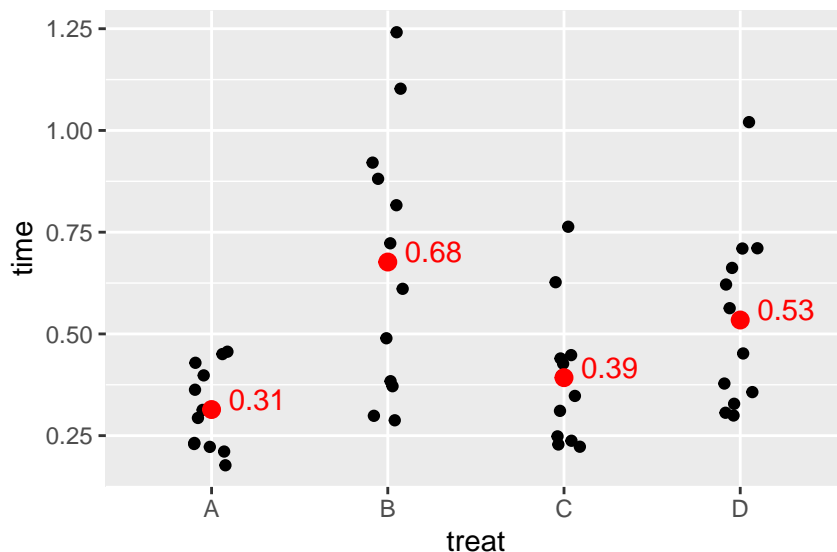
1. Focus on time as the response and treat as the predictor. Plot time as a function of treat using a jittered scatter plot. Add a mean of time for each level of treat to the plot. What pattern do you see?

- Survival time is longest for treatment B followed by treatment D. Treatment A has the shortest survival time and the smallest variability.

```
(means <- aggregate(time~treat,rats, mean))
```

```
  treat      time
1     A 0.3141667
2     B 0.6766667
3     C 0.3925000
4     D 0.5341667
```

```
ggplot(rats,aes(x=treat, y=time))+
  geom_point(position=position_jitter(width=0.1)) +
  stat_summary(fun="mean", colour = "red")+
  geom_text(data = means,
            aes(label = round(time, 2)),
            vjust = 0, hjust = -0.3, size = 4, color="red")
```

2. Perform a one-way ANOVA with time as the response and treat as the predictor. Use aov() as well as lm(). For each, describe what you observe about what can be concluded.

- They have the same F-statistic (6.48) and p-value (0.000992) indicating that there is significant effect of treatment on time.
- aov() neither included the intercept nor the levels of treatments. Its p-value (0.000992) supports rejecting the null that there is no difference among the group means - in favor of the alternative hypothesis (at least one group mean is different).
- lm() included the intercept (reference group) and all the other three levels (B,C,D) of treatment. The p-values of the treatment coefficients suggests that the means for treatments B and D are different from the reference group (treatment A) while treatment C is not significantly different from treatment A.

```
summary(aov(time~treat,rats))
```

```
           Df Sum Sq Mean Sq F value   Pr(>F)
treat       3 0.9212 0.30707   6.484 0.000992 ***
Residuals  44 2.0839 0.04736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(lm(time~treat,rats))
```

```
Analysis of Variance Table

Response: time
          Df  Sum Sq  Mean Sq F value    Pr(>F)
treat      3 0.92121 0.307069  6.4836 0.0009921 ***
Residuals 44 2.08387 0.047361
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm(time~treat,rats))
```

```
Call:
lm(formula = time ~ treat, data = rats)

Residuals:
     Min       1Q   Median       3Q      Max
-0.38667 -0.15292 -0.01417  0.12833  0.56333

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.31417    0.06282   5.001 9.62e-06 ***
treatB       0.36250    0.08885   4.080 0.000186 ***
treatC       0.07833    0.08885   0.882 0.382739
treatD       0.22000    0.08885   2.476 0.017196 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2176 on 44 degrees of freedom
Multiple R-squared:  0.3065,    Adjusted R-squared:  0.2593
F-statistic: 6.484 on 3 and 44 DF,  p-value: 0.0009921
```

3. Perform pairwise comparisons using Tukey's difference test. Which pairs of treatment are different? How does this compare to your results from #2?

- The following pairs are different: B-A and C-B as they have p-values < 0.05
- The p-values in the Tukey's difference test were larger than in #2 as Tukey's test accounts for the number of comparisons making it more robust while controlling for Type I error.

```
TukeyHSD(aov(time~treat,rats))
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = time ~ treat, data = rats)

$treat
          diff         lwr         upr     p adj
B-A  0.36250000  0.12528287  0.59971713 0.0010358
C-A  0.07833333 -0.15888380  0.31555047 0.8143113
D-A  0.22000000 -0.01721713  0.45721713 0.0778376
C-B -0.28416667 -0.52138380 -0.04694953 0.0131752
D-B -0.14250000 -0.37971713  0.09471713 0.3869986
D-C  0.14166667 -0.09555047  0.37888380 0.3921830
```

4. Go back to your plot from #1 and add lines showing time as at the interaction of poison and treat. Does the pattern from #1 hold across poison types?

- The pattern from #1 is consistent with the pattern observed here - treatment B has the longest survival time, next is treatment D, then treatment C, while treatment A has the shortest survival time.

- In the first plot we see that treatment B and poison type I and II have higher survival time, followed by the same poison types and treatment D. Flipping the graphic around, we can easily see how treatment B declines across poison types with the largest surviving time for poison type I, followed by II, and lastly III. We see the same trend for treatment types A and C. However, for treatment D, the highest survival time is in combination with poison II, followed by I, and lastly III.

```
ggplot(rats,aes(x=treat,y=time))+
  geom_point(position=position_jitter(width=0.2))+
  stat_summary(fun=mean, geom="point", color="red",size=2)+
  stat_summary(fun=mean, geom="line",
               aes(group=poison, linetype=poison))
```
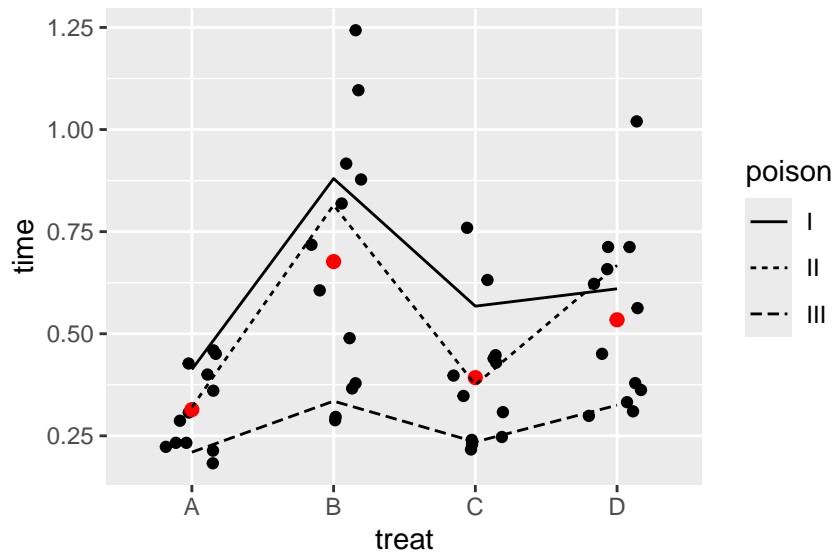
```
ggplot(rats,aes(x=poison,y=time))+
  geom_point(position=position_jitter(width=0.2))+
  stat_summary(fun=mean, geom="point", color="blue",size=2)+
  stat_summary(fun=mean, geom="line",
               aes(group=treat, linetype=treat, color=treat))
```
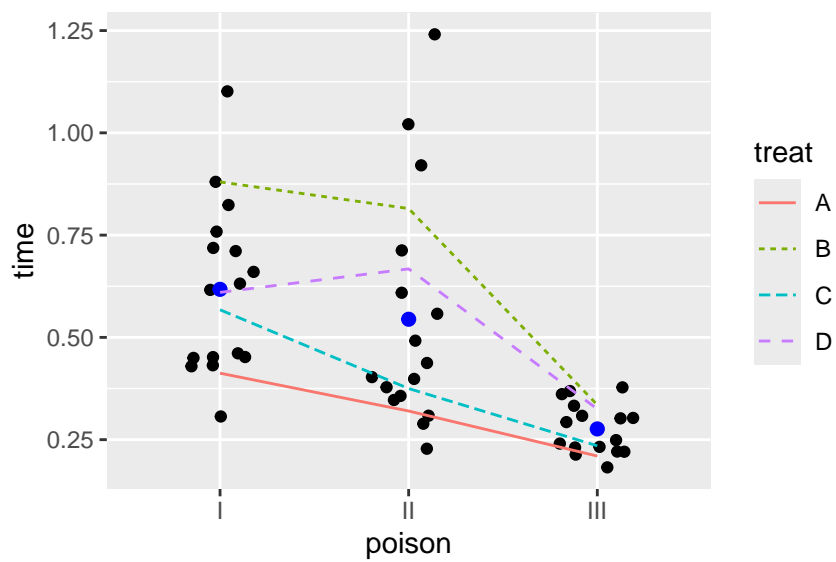
5. Model time as a function of treat, poison, and their interaction in a linear regression. Check whether OLS assumptions hold, and if not, what (if any) transformation of the outcome might present a solution.

- The OLS assumptions do not appear to be met. Data on the qq-plot are positively skewed and the constant variance appears to be violated.
- Since lambda is closer to -0.5, an inverse square root transformation is a potential solution. Other possible recommendations based on choice of lambda are:

  - lambda $= 0 =>$ log transformation
  - lamda $= -1 =>$ reciprocal transformation

```
summary(rat_ols <- lm(time~treat*poison,rats))
```

```
Call:
lm(formula = time ~ treat * poison, data = rats)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32500 -0.04875  0.00500  0.04312  0.42500

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.41250    0.07457   5.532 2.94e-06 ***
treatB            0.46750    0.10546   4.433 8.37e-05 ***
treatC            0.15500    0.10546   1.470   0.1503
treatD            0.19750    0.10546   1.873   0.0692 .
poisonII         -0.09250    0.10546  -0.877   0.3862
poisonIII        -0.20250    0.10546  -1.920   0.0628 .
treatB:poisonII   0.02750    0.14914   0.184   0.8547
treatC:poisonII  -0.10000    0.14914  -0.671   0.5068
treatD:poisonII   0.15000    0.14914   1.006   0.3212
treatB:poisonIII -0.34250    0.14914  -2.297   0.0276 *
treatC:poisonIII -0.13000    0.14914  -0.872   0.3892
treatD:poisonIII -0.08250    0.14914  -0.553   0.5836
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1491 on 36 degrees of freedom
Multiple R-squared:  0.7335,    Adjusted R-squared:  0.6521
F-statistic:  9.01 on 11 and 36 DF,  p-value: 1.986e-07
```
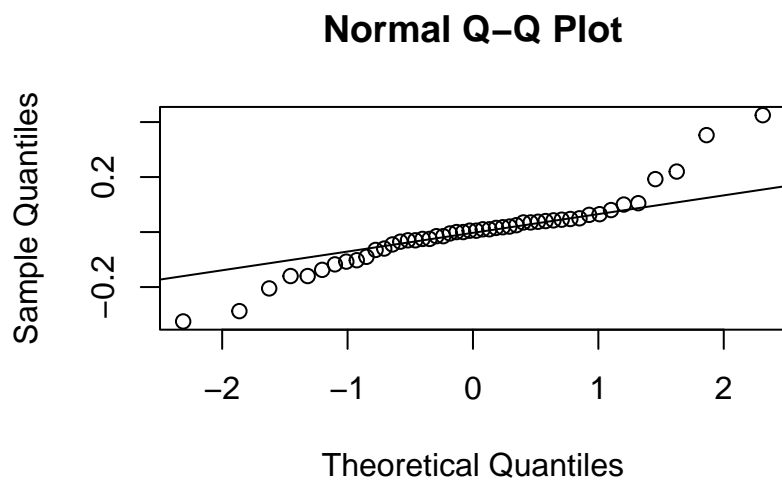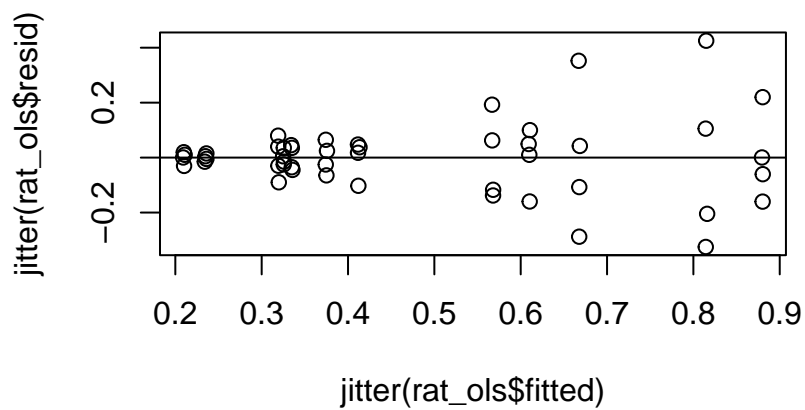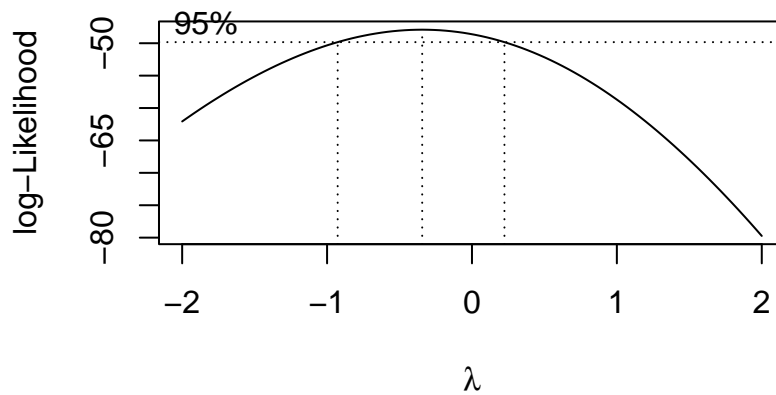
7

```
qqnorm(rat_ols$resid);qqline(rat_ols$resid)
```

Normal Q–Q Plot


```
plot(jitter(rat_ols$fitted),jitter(rat_ols$resid));abline(h=0)
```



```
boxcox(lm(time~treat,rats))
```

8

6. Now fit a model of time as a function of treat, but with treat as a random effect. Interpret the effect of treat on time.

- The fixed intercept of .48, survival time was significant (p<.05) with a 95% confidence interval between .32 and .64 and does not include zero => reject null hypothesis: variance from treatment does contribute to the overall variance therefore, there may be meaningful differences between treatments. Furthermore, we calculate the intraclass correlation (ICC) to be .31 which is the proportion of variance in time values explained by the grouping structure of treatment.

```
# (1|treat) means we are fitting as random effect
summary(randmod<-lmer(time~(1|treat), rats))
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: time ~ (1 | treat)
   Data: rats

REML criterion at convergence: -0.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6369 -0.7218 -0.1514  0.5186  2.7284

Random effects:
```

```
 Groups    Name          Variance Std.Dev.
 treat    (Intercept) 0.02164  0.1471
 Residual              0.04736  0.2176
Number of obs: 48, groups:  treat, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.47937    0.07998   5.993
```

confint(randmod)

```
                2.5 %     97.5 %
.sig01       0.04474072 0.3338817
.sigma       0.17900826 0.2724161
(Intercept) 0.30345222 0.6552976
```

performance::icc(randmod)

```
# Intraclass Correlation Coefficient

    Adjusted ICC: 0.314
  Unadjusted ICC: 0.314
```