Stat631HW2

Brandon Keck

$N = 4 \ x \ 10 = 40, \ g = 4 \ and \ n = 10$

```r
# Dataset
drug1 <- c(18.2, 16.4, 10.0, 13.5, 13.5, 6.7, 12.2, 18.2, 13.5, 16.4)
drug2 <- c(5.5, 12.2, 11.0, 6.7, 15, 8.2, 7.4, 12.2, 6.7, 11.0)
drug3 <- c(5.5, 5.5, 8.2, 9.0, 10.0, 6.0, 7.4, 5.5, 11, 8.2)
drug4 <- c(6.0, 7.4, 7.8, 8, 5.0, 7.4, 7.4, 5.5, 6.7, 5.5)

# Combine data into single vector
Response <- c(drug1, drug2, drug3, drug4)

# Create grouping variable
drug <- rep(c("Drug1", "Drug2", "Drug3", "Drug4"), each = 10)

# Create a data frame
df <- data.frame(Response, drug)

# Covert drug to a factor variable
df$drug <- as.factor(df$drug)

# Check the structure
str(df)
```

```
'data.frame':   40 obs. of  2 variables:
 $ Response: num  18.2 16.4 10 13.5 13.5 6.7 12.2 18.2 13.5 16.4 ...
 $ drug    : Factor w/ 4 levels "Drug1","Drug2",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Exercise 1

(a)

i.

Step 1: Hypothesis

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

all treatments means are the same

vs

H_A : at least one treatment mean is different from the others.

ii.

Step 2: Check conditions (assumptions)

Independence: Each guinea pig was randomly selected, and we can reasonably assume that the drug's effects on one guinea pig do not influence its effects on another. Additionally, the drugs themselves are administered independently of each other.

Checking Normality assumptions: The residuals align closely with the QQ plot's reference line, indicating that the assumption of normality is satisfied.

```
# Fit the model
fit <- lm(Response ~ drug, data = df)
anova(fit)
```

```
Analysis of Variance Table

Response: Response
          Df Sum Sq Mean Sq F value    Pr(>F)
drug       3 305.08 101.693   14.45 2.446e-06 ***
Residuals 36 253.35   7.038
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
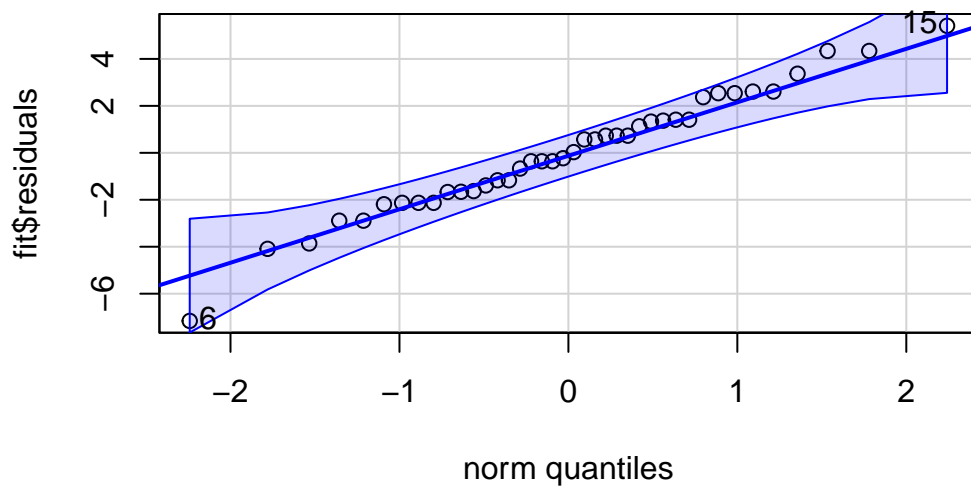
```
names(fit)
```

```
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "contrasts"     "xlevels"       "call"          "terms"
[13] "model"
```

```
# Check for Normality
library(car)
```

Loading required package: carData

```
qqPlot(fit$residuals)
```


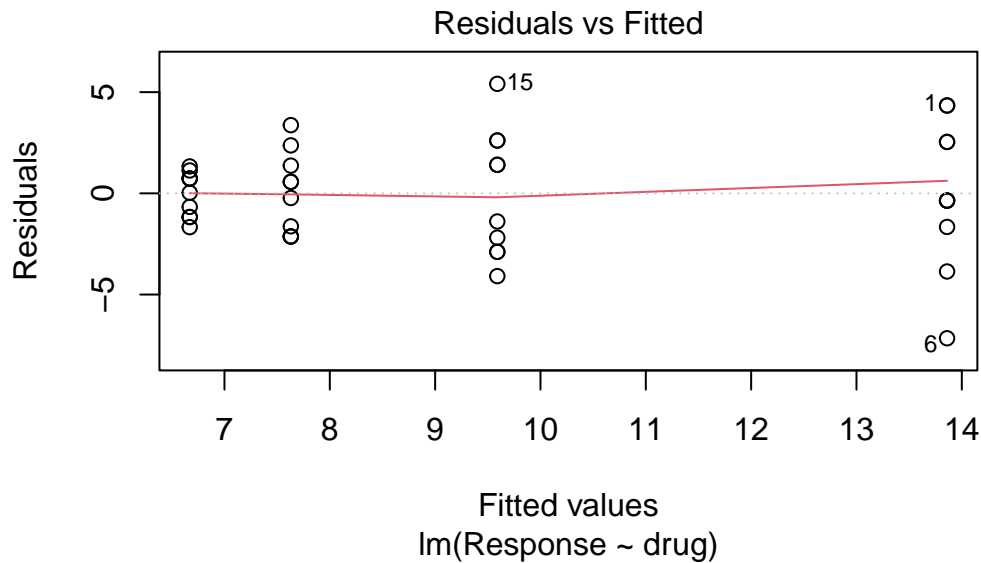[1]   6 15

```
# Testing for Normality
shapiro.test(fit$residuals)
```

    Shapiro-Wilk normality test

data: fit$residuals
W = 0.98456, p-value = 0.8506

Checking Equal Variance Assumptions:

```
# Residuals vs Fitted plot
plot(fit, which = 1)
```

Residuals vs Fitted

lm(Response ~ drug)

```
# Bartlett Test for Normality
bartlett.test(Response ~ drug, data = df)
```

```
    Bartlett test of homogeneity of variances

data:  Response by drug
Bartlett's K-squared = 12.041, df = 3, p-value = 0.007243
```

The Q-Q plot visually confirms that the residuals follow an approximately normal distribution. Additionally, the Shapiro-Wilk test yields a p-value of 0.8506, which is greater than the significance level of $\alpha = 0.05$ This suggests that we do not have enough evidence to reject the normality assumption.

However, the Residuals vs Fitted plot indicates heteroscedasticity. This concern is confirmed by Bartlett's test, which tests for homoscedasticity of variance and results in a p-value of 0.007243. Since this is less than the significance level of $\alpha = 0.05$ we reject the null hypothesis of equal variances, indicating that the assumption of homoscedasticity is violated.

iii.

```
powerTransform(Response ~ drug, data = df)
```

```
Estimated transformation parameter
        Y1
-0.1739748
```

Yes, a transformation is needed because the equal variance assumption is violated, as indicated by Bartlett's test. To determine an appropriate transformation, we use the Box-Cox powerTransform method, which estimates the optimal transformation parameter λ .

From the powerTransform test, the estimated λ value is -0.1739748, which is close to 0. Since a λ value near 0 suggests a log transformation, we apply the natural logarithm to the response variable to stabilize variance.

iv.

```
# Log transformation of Response
df$logResponse <- log(df$Response)

# Refitting the model
mod2 <- lm(logResponse ~ drug, data = df)
anova(mod2)
```
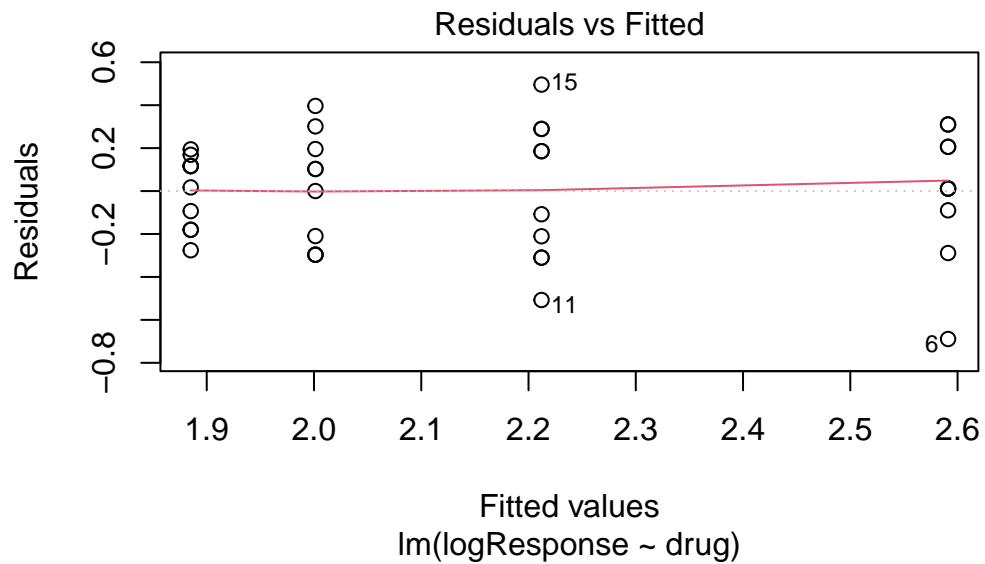
```
Analysis of Variance Table

Response: logResponse
          Df Sum Sq Mean Sq F value    Pr(>F)
drug       3 2.8875 0.96250  12.812 7.534e-06 ***
Residuals 36 2.7044 0.07512
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
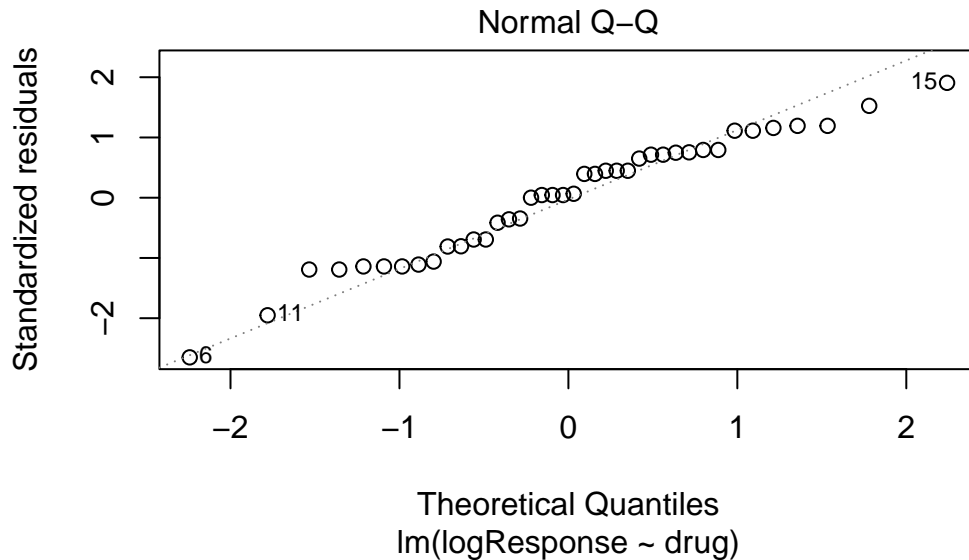
```
plot(mod2, 1) # Residual vs Fitted Values
```

Residuals vs Fitted

Fitted values
lm(logResponse ~ drug)

```r
# Equal variance Check using Levene's Test
leveneTest(mod2)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.5595  0.216
      36
```

```r
plot(mod2, 2) # Q-Q Plot to check normality
```

Normal Q–Q

Theoretical Quantiles
lm(logResponse ~ drug)

```
shapiro.test(mod2$residuals) # Shapiro.Wilk Test
```

```
        Shapiro-Wilk normality test

data:  mod2$residuals
W = 0.96438, p-value = 0.2357
```

After applying a log transformation to stabilize the variance, I conducted an ANOVA test to compare the mean lethal dosages of four different drugs. The ANOVA results indicated a statistically significant difference among at least one of the drugs with a p-value of 7.534e-6 suggesting that some drugs require different dosages to be lethal. Since all ANOVA assumptions were met after transformation, we conclude that the effectiveness of the four drugs is not the same.

(b)

Non-parametric ANOVA

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs H_A : at least one treatment mean is different from the others.

```r
kruskal.test(Response ~ drug, data = df)
```

```
    Kruskal-Wallis rank sum test

data:  Response by drug
Kruskal-Wallis chi-squared = 17.839, df = 3, p-value = 0.0004748
```

```r
anova(mod2)$"Pr(>F)"
```

```
[1] 7.53391e-06          NA
```

Conclusion:

Since the p-value is 0.0004748 which is less than our significance value of 0.05, we reject the null hypothesis and can conclude that at least one drug has significantly different lethal dosage. This finding aligns with the ANOVA test from earlier confirming the effectiveness of the drugs is not the same.