

Project 5 Individual

Oliver and Minh

2024-12-10

```
#all possible models
large_model <-lm(log(time)~kidney_fn+alcohol+kidney_fn*alcohol,data=kidney_data)
medium_model<-lm(log(time)~kidney_fn+alcohol,data=kidney_data)
alcohol_model<-lm(log(time)~alcohol,data=kidney_data)
kidney_model<-lm(log(time)~kidney_fn,data=kidney_data)
intercept_model<-lm(log(time)~1,data=kidney_data)
```

```
#hypothesis test to see if all variables are necessary
anova(large_model,intercept_model)
```

```
## Analysis of Variance Table
##
## Model 1: log(time) ~ kidney_fn + alcohol + kidney_fn * alcohol
## Model 2: log(time) ~ 1
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     132 11.543
## 2     137 29.091 -5    -17.547 40.132 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(large_model,kidney_model)
```

```
## Analysis of Variance Table
##
## Model 1: log(time) ~ kidney_fn + alcohol + kidney_fn * alcohol
## Model 2: log(time) ~ kidney_fn
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     132 11.543
## 2     136 12.993 -4    -1.4499 4.1449 0.003397 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(large_model,alcohol_model)
```

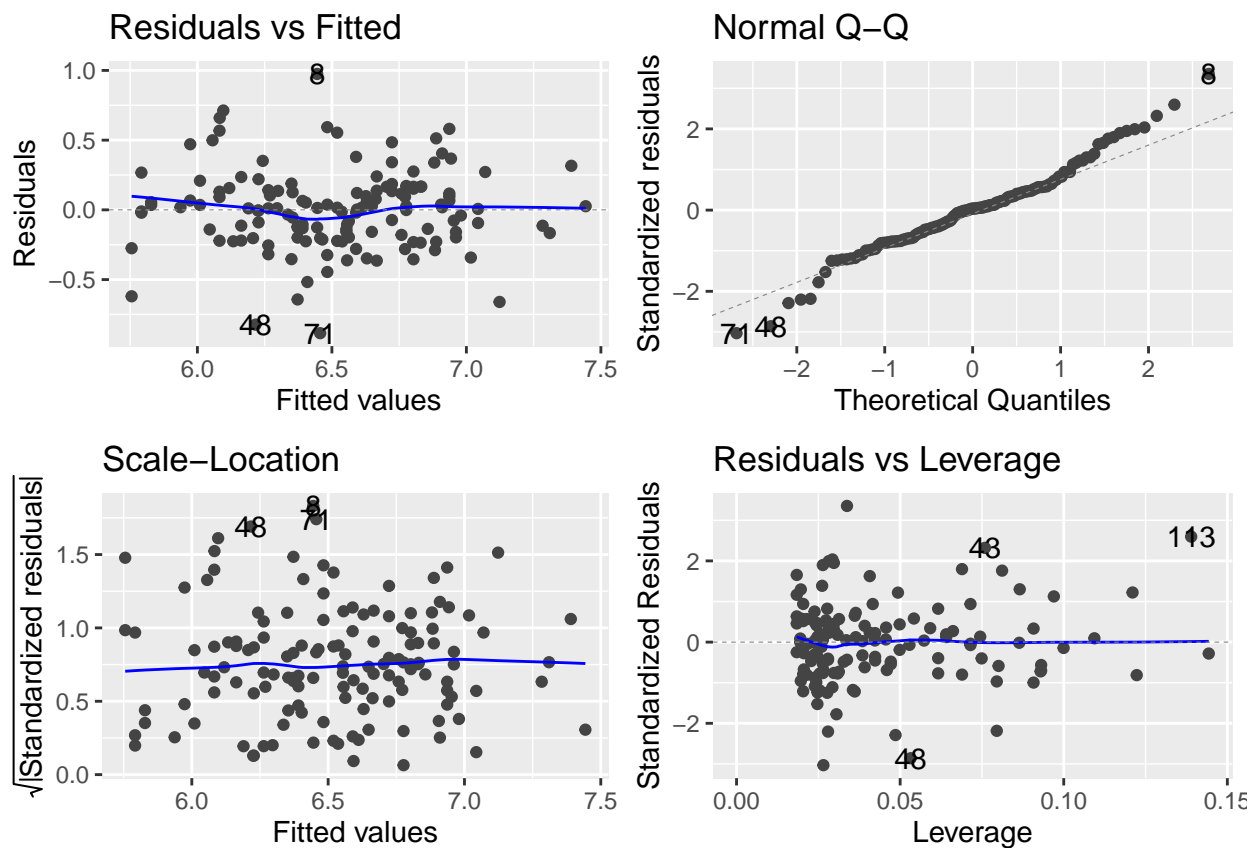
```
## Analysis of Variance Table
##
## Model 1: log(time) ~ kidney_fn + alcohol + kidney_fn * alcohol
## Model 2: log(time) ~ alcohol
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1    132 11.543
## 2    135 24.127 -3    -12.584 47.968 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(large_model,medium_model)
```

```
## Analysis of Variance Table
##
## Model 1: log(time) ~ kidney_fn + alcohol + kidney_fn * alcohol
## Model 2: log(time) ~ kidney_fn + alcohol
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     132 11.543
## 2     134 12.237 -2    -0.69395 3.9678 0.02121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#assumptions check and coefficients
autoplot(large_model)
```



```
summary(large_model)
```

```
##
## Call:
```

```
## lm(formula = log(time) ~ kidney_fn + alcohol + kidney_fn * alcohol,
##     data = kidney_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88433 -0.19244  0.01088  0.14053  0.97474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      5.71936    0.08782   65.126 < 2e-16 ***
## kidney_fn         0.36294    0.05226    6.945 1.56e-10 ***
## alcoholModerate   0.23031    0.12974    1.775  0.07818 .
## alcoholSevere     0.30356    0.15204    1.996  0.04794 *
## kidney_fn:alcoholModerate -0.09620    0.06099   -1.577  0.11715
## kidney_fn:alcoholSevere  -0.17895    0.06449   -2.775  0.00633 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2957 on 132 degrees of freedom
## Multiple R-squared:  0.6032, Adjusted R-squared:  0.5882
## F-statistic: 40.13 on 5 and 132 DF, p-value: < 2.2e-16
```

```
#code for plot of kidney function and survival time by alcohol consumption
library(ggplot2)

ggplot(data = kidney_data, aes(x = kidney_fn, y = log(time), color = alcohol)) +
  geom_point() +
  xlab("Kidney Function") +
  ylab("Log(Survival Time)") +
  ggtitle("") +
  scale_color_discrete(name = "Alcohol Consumption") +
  geom_smooth(method = "lm", se = F)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

