

STAT 222 Spring 2022 HW9

Matthew Zhao

```
pr8.2 = read.table("http://www.stat.uchicago.edu/~yibi/s222/pr8_2.txt", h=T)
pr8.2$Slash = factor(pr8.2$slash, labels=c("0%", "25%", "50%"))
pr8.2$Target = factor(pr8.2$target, labels=c(42, 48))
pr8.2$Resin = factor(pr8.2$resin, labels=c("6%", "9%", "12%"))
```

Q1 — 2 points

We cannot because there is only 1 replicate per group. As a result, the error term has no degrees of freedom and so we cannot estimate σ^2 and calculate the MSE for hypothesis testing. This means that to perform testing, we need to pool higher order interactions into the error which means we cannot test all interactions.

Q2 — 4 points

```
lm1 = lm(y~Slash*Target*Resin, data = pr8.2)
anova(lm1)
## Analysis of Variance Table
##
## Response: y
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## Slash	2	9.74	4.87		
## Target	1	105.61	105.61		
## Resin	2	52.12	26.06		
## Slash:Target	2	0.88	0.44		
## Slash:Resin	4	2.67	0.67		
## Target:Resin	2	0.59	0.30		
## Slash:Target:Resin	4	1.85	0.46		
## Residuals	0	0.00			

$$\text{MSE} = 0.46 + 0.3 + 0.67 + 0.44 = 1.87$$

$$\text{df}_{\text{MSE}} = 4 + 2 + 4 + 2 = 12$$

```
mse <- 1.87
df_mse <- 12
# Slash
F.stat <- 4.87/mse
sprintf("Slash p-value: %f", pf(F.stat, 2, df_mse, lower.tail = F))
## [1] "Slash p-value: 0.114980"
```

```
# Target
F.stat <- 105.61/mse
sprintf("Target p-value: %f",pf(F.stat,1,df_mse,lower.tail = F))
## [1] "Target p-value: 0.000007"

# Resin
F.stat <- 26.06/mse
sprintf("Resin p-value: %f",pf(F.stat,2,df_mse,lower.tail = F))
## [1] "Resin p-value: 0.000743"
```

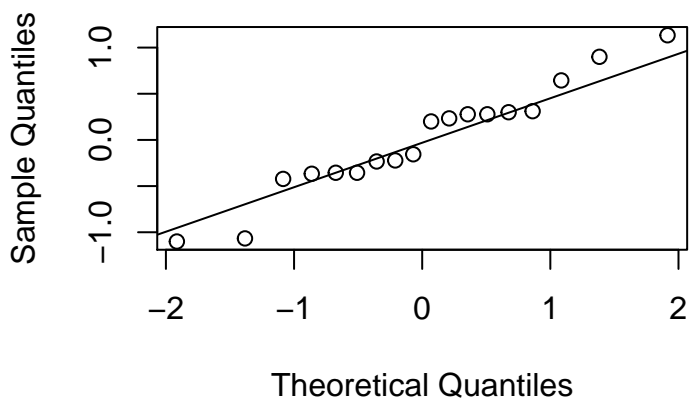
We can conclude that Target and Resin have an effect on the density of board produced.

Q3 — 3 points

```
lm1 = lm(y~Slash+Target+Resin, data = pr8.2)
```

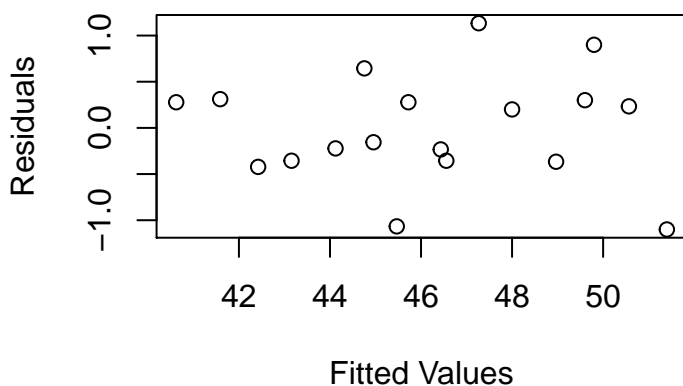
```
qqnorm(lm1$residuals)
qqline(lm1$residuals)
```

Normal Q–Q Plot

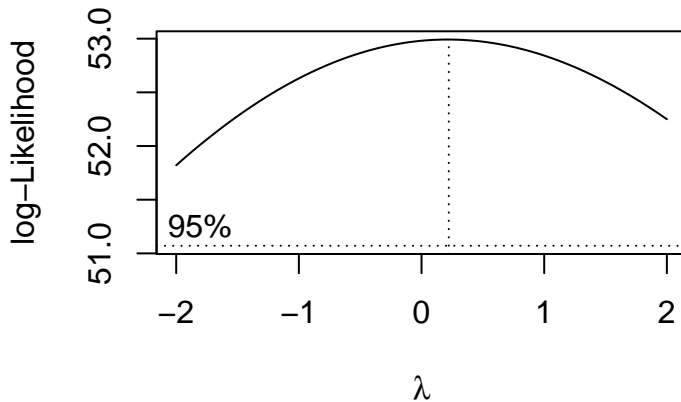


```
plot(lm1$fitted.values, lm1$residuals, xlab = "Fitted Values",
     ylab = "Residuals", main = "lm1")
```

lm1



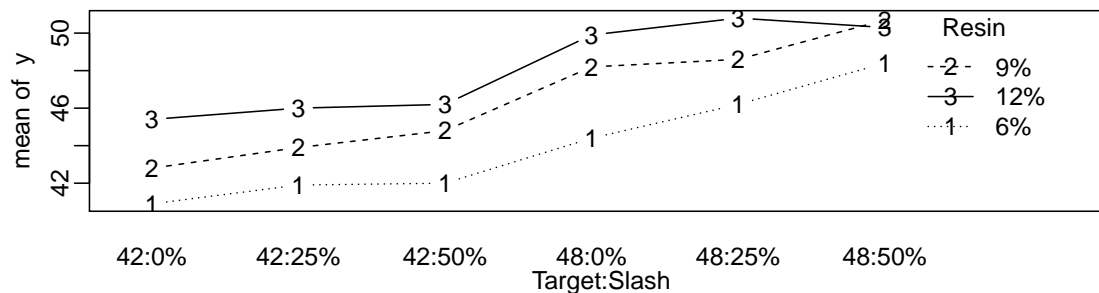
```
boxcox(lm1)
```



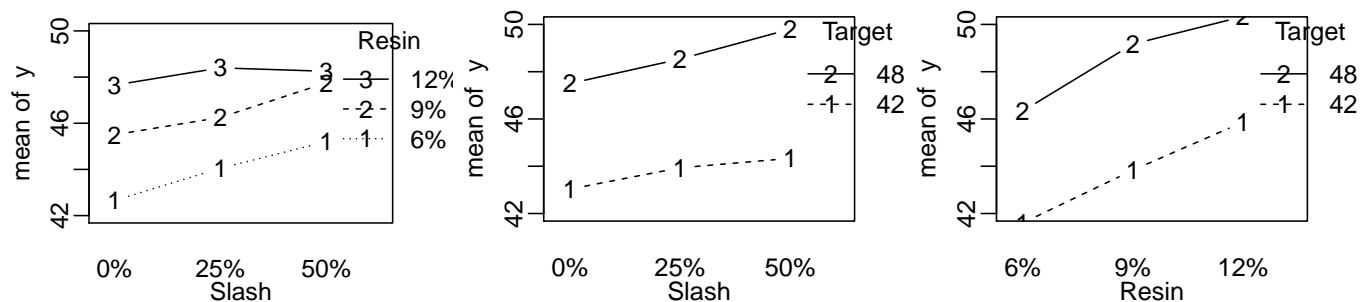
The points mostly fall on the line, meaning that there's no indication that the normality assumption is violated. Additionally, the fitted values vs residuals plot does not reveal significant nonconstant variance, which is supported by the boxcox plot whose 95% CI includes 1, indicating that there is no need for power transformation of the response.

Q4 — 3 points

```
par(mai=c(.56,.56,.1,.1),mgp=c(1.8,.5,0))
with(pr8.2, interaction.plot(Target:Slash, Resin,y,type="b"))
```



```
par(mai=c(.56,.56,.1,.4),mgp=c(1.8,.5,0))
with(pr8.2, interaction.plot(Slash, Resin,y,type="b", ylim=c(42,50)))
with(pr8.2, interaction.plot(Slash, Target,y,type="b", ylim=c(42,50)))
with(pr8.2, interaction.plot(Resin, Target,y,type="b", ylim=c(42,50)))
```



For three-way interactions, the lines at 42 target density are parallel, while for 48 target they are no longer parallel. This is evidence of a three-way interaction since the interaction between Slash and Resin appears to change from 42 target density to 48 target density. For two-way interactions, it appears that there are interactions between Slash and Resin since the lines are not parallel. However,

the interactions of Target-Slash and Target-Resin are absent since the lines are parallel.

Q5 — 4 points

```
library(mosaic)
mean(y ~ Resin, data=pr8.2)
##      6%      9%     12%
## 43.9667 46.5000 48.1000
mean(~y, data = pr8.2)
## [1] 46.1889
```

$$\beta_j = \bar{y}_{.j} - \bar{y}_{...}$$

$$j = 6\%: 43.9667 - 46.1889 = -2.2222$$

$$j = 9\%: 46.5000 - 46.1889 = 0.3111$$

$$j = 12\%: 48.1000 - 46.1889 = 1.9111$$

$$SS_j = acn \sum_j (\hat{\beta}_j)^2 = 2 * 3 * 1 * ((-2.2222)^2 + (0.3111)^2 + (1.9111)^2) = 52.1236$$

Q6 — 9 points

```
lm1emmean <- emmeans(lm1, ~Slash)
summary(contrast(lm1emmean, method="pairwise", adjust="tukey"),
infer=c(T,F), level=0.95)
## contrast estimate SE df lower.CL upper.CL
## 0% - 25% -0.967 0.408 12 -2.05 0.121
## 0% - 50% -1.800 0.408 12 -2.89 -0.712
## 25% - 50% -0.833 0.408 12 -1.92 0.255
##
## Results are averaged over the levels of: Target, Resin
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

```
lm1emmean <- emmeans(lm1, ~Target)
summary(contrast(lm1emmean, method="pairwise", adjust="tukey"),
infer=c(T,F), level=0.95)
## contrast estimate SE df lower.CL upper.CL
## 42 - 48 -4.84 0.333 12 -5.57 -4.12
##
## Results are averaged over the levels of: Slash, Resin
## Confidence level used: 0.95
```

```
lm1emmean <- emmeans(lm1, ~Resin)
summary(contrast(lm1emmean, method="pairwise", adjust="tukey"),
infer=c(T,F), level=0.95)
## contrast estimate SE df lower.CL upper.CL
## 6% - 9% -2.53 0.408 12 -3.62 -1.445
## 6% - 12% -4.13 0.408 12 -5.22 -3.045
```

```
## 9% - 12%    -1.60 0.408 12    -2.69    -0.512
##
## Results are averaged over the levels of: Slash, Target
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
```

We use Tukey's method: $\hat{C} \pm \frac{q_{g,dfE,\alpha}}{\sqrt{2}} \times \sqrt{\text{MSE}(\frac{1}{r} + \frac{1}{r})}$. We can conclude that target density is associated with an increase in actual density, increasing the amount of resin is also associated with an increase in actual density, and that increasing the fraction of slash has inconclusive effects on the actual density since 0 is in the CI for both contrasts.