Kitu Komya
UID: 404-491-375

13.1a.

```
> #13.1
> parts <- read.csv(file = "wk9.csv")
> fit <- aov(Capability~factor(PartNo)+factor(Operator)+factor(PartNo)*factor(Opera
tor), data = parts)
> summary(fit)
                                Df Sum Sq Mean Sq F value   Pr(>F)    
factor(PartNo)                   9  99.02  11.002   7.335 3.22e-06 ***
factor(Operator)                 1   0.42   0.417   0.278    0.601    
factor(PartNo):factor(Operator)  9   5.42   0.602   0.401    0.927    
Residuals                       40  60.00   1.500                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

At a significance level of 5%, the p-values are significant for only part number. Thus, we conclude that while the effect of parts is large, operators has no significant effect. There is also no significant interaction between part and operator.

13.1b We may use the following formulas to estimate the variance components:

$$\hat{\sigma}^2 = MS_E$$
$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_E}{n}$$
$$\hat{\sigma}^2_{\beta} = \frac{MS_B - MS_{AB}}{an}$$
$$\hat{\sigma}^2_{\tau} = \frac{MS_A - MS_{AB}}{bn}$$

$\sigma_B^2 = (11.02 – 0.602)/(3*2) = 1.736$

$\sigma_T^2 = (0.417 – 0.602)/(10*3) = -0.006166$

$\sigma_{TB}^2 = (0.602 – 1.500)/(3) = -0.2993$

$\sigma^2 = 1.500$

However, since both σ_T^2 and σ_{TB}^2 are negative, which doesn't make sense since by definition variances are nonnegative, we can instead assume that both σ_T^2 and σ_{TB}^2 are equal to 0 instead. Moreover, since the p-value of the interactional effect between part number and operator is high, we can assume that there is no interactional effect and that σ_{TB}^2 truly is 0. Thus, we can remove the interaction effect from the model for a reduced model.

```
> fit2 <- aov(Capability~factor(PartNo)+factor(Operator), data = parts)
> summary(fit2)
                 Df Sum Sq Mean Sq F value   Pr(>F)    
factor(PartNo)    9  99.02  11.002   8.241 2.46e-07 ***
factor(Operator)  1   0.42   0.417   0.312    0.579    
Residuals        49  65.42   1.335                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Then, the new estimates for variances are as follows, since both main effects are tested against the error term instead:

$\sigma_B^2 = (11.02 - 1.335)/(3*2) = 1.614$

$\sigma_T^2 = (0.417 - 1.335)/(10*3) = -0.0306$

$\sigma^2 = 1.335$

$\sigma_{measurement\_system}^2 = \sigma^2 + \sigma_B^2 = 1.335 + 1.614 = 2.949$. The variability in the measurement system is quite large relative to the variability in the product. This is unfortunately not desirable since the measurement system is thus implied to not have capability in distinguishing among different grades of product.

13.2a

```
> fit3 <- aov(Capab~factor(PartNum) + factor(Inspector) + factor(PartNum)*factor(In
spector), data = inspect)
> summary(fit3)
> summary(fit3)
                                  Df Sum Sq Mean Sq F value   Pr(>F)    
factor(PartNum)                    9   3910   434.5 814.685  < 2e-16 ***
factor(Inspector)                  2     42    21.0  39.396 1.18e-11 ***
factor(PartNum):factor(Inspector) 18     47     2.6   4.859 1.74e-06 ***
Residuals                         60     32     0.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At a significance level of 5%, the p-values are significant for all factors. Thus, we conclude that the effects of part number, inspector, as well as the interaction between part number and inspector are all large and significant.

13.2b We may use the following formulas to estimate the variance components:

$$\hat{\sigma}^2 = MS_E$$
$$\hat{\sigma}^2_{\tau\beta} = \frac{MS_{AB} - MS_E}{n}$$
$$\hat{\sigma}^2_{\beta} = \frac{MS_B - MS_{AB}}{an}$$
$$\hat{\sigma}^2_{\tau} = \frac{MS_A - MS_{AB}}{bn}$$

$\sigma_B^2 = (434.5 - 2.6)/(3*3) = 47.98$

$\sigma_T^2 = (21.0 - 2.6)/(10*3) = 0.6133$

$\sigma_{TB}^2 = (2.6 - 0.5)/(3) = 7$

$\sigma^2 = 0.5$

$\sigma_{induction\_motor\_starter} = \sigma^2 + \sigma_B^2 = 0.5 + 47.98 = 48.48$. The variability in the induction motor starter is quite large relative to the variability in the product. This is unfortunately not desirable since the induction motor starter is thus implied to not have capability in distinguishing among different grades of product.

14.1

```
> fit4 <- lm(Rate~factor(Process)/factor(Batch), data = batch)
> anova(fit4)
Analysis of Variance Table

Response: Rate
                              Df  Sum Sq Mean Sq F value    Pr(>F)    
factor(Process)                2  676.06  338.03  17.869 1.768e-05 ***
factor(Process):factor(Batch)  9 2077.58  230.84  12.203 5.477e-07 ***
Residuals                     24  454.00   18.92                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To find the correct F-values and p-values:

<u>Process</u>

F-value = 338.03/230.84 = 1.464348

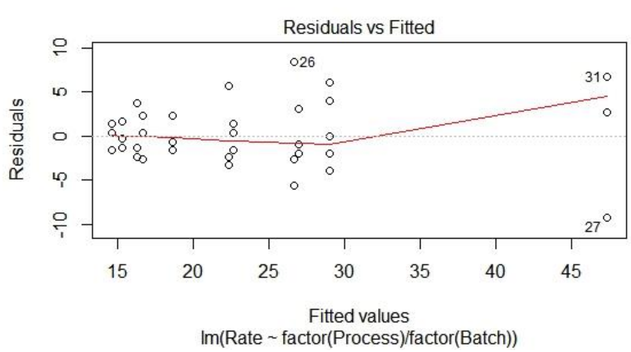
p-value = 1 – pf(1.464348, 2, 9) = 0.281464 #used that R code

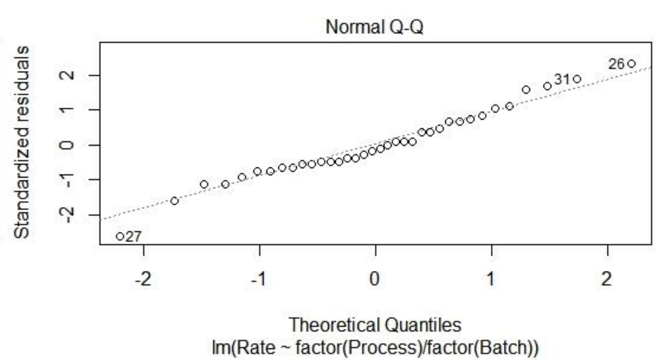
<u>Batch</u>

F-value = 230.84/18.92 = 12.20085

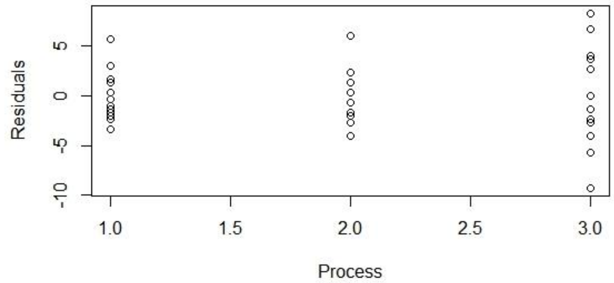
p-value = 1 – pf(12.20085, 9, 24) = 5.486321 * 10^{-7} #used that R code

From examining the p-values at a 5% significance level, we conclude that there is no significant effect on rate due to process, but the purity of batches of raw materials from the same process does differ significantly, hence having a significant effect on rate.





From the residual plot, we can see that since the residuals are randomly scattered with no obvious pattern, our assumptions of linearity and constant variance are met. Moreover, in the normal qq plot, we see that since the majority of the residuals follow the straight line, the errors are normally distributed, hence making our model a valid one since the normality assumption is met.



Looking at the residuals vs processes plot, we can affirm that the variability within batches is approximately the same for all processes. Although process three has a lightly larger variability, in general, we can conclude that batch to batch variability is about the same in the three processes since the spreads of their residuals are approximately similar for all three processes.

14.2

```
> fit5 <- lm(Finish~factor(Machine)/factor(Operators), data = machine)
> anova(fit5)
Analysis of Variance Table

Response: Finish
                                  Df Sum Sq Mean Sq F value   Pr(>F)    
factor(Machine)                    3 3617.7 1205.89 14.2709 0.000291 ***
factor(Machine):factor(Operators)  8 2817.7  352.21  4.1681 0.013408 *  
Residuals                         12 1014.0   84.50                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

To find the correct F-values and p-values:

<u>Machine:</u>

F-value = 1205.89/352.21 = 3.423781

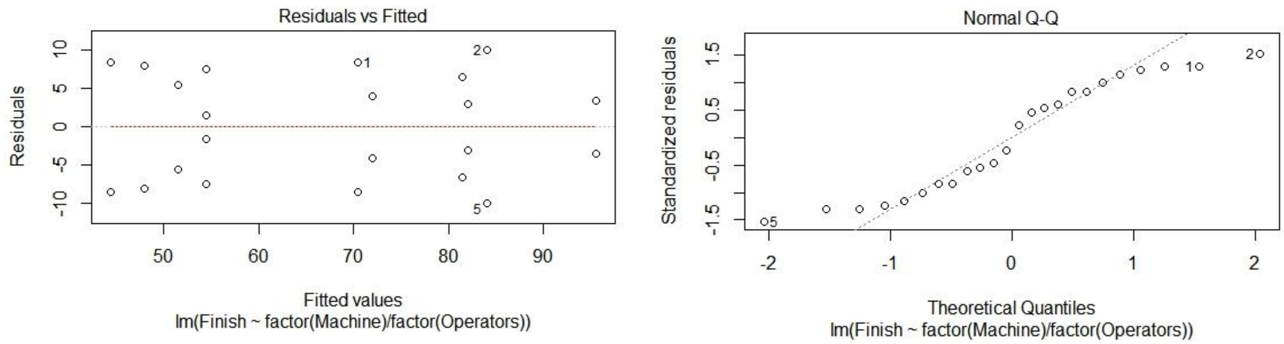
p-value = 1 – pf(3.423781, 3, 8) = 0.07279735 #used that R code

<u>Operators</u>:

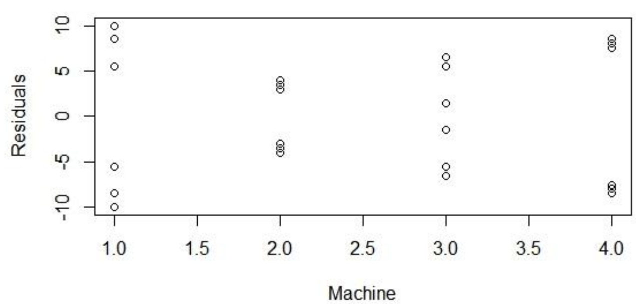
F-value = 352.21/84.50 = 4.168166

p-value = 1 – pf(4.168166, 8, 12) = 0.0134081 #used that R code

From examining the p-values at a 5% significance level, we conclude that there is no significant effect on rate due to machine, but the operators from the same machine do differ significantly, hence having a significant effect on surface finish.



From the residual plot, we can see that since the residuals are randomly scattered with no obvious pattern, our assumptions of linearity and constant variance are met. Moreover, in the normal qq plot, we see that since the majority of the residuals follow the straight line, the errors are normally distributed, hence making our model a valid one since the normality assumption is met. Although there is some deviation at the extrema, for the most part it is straight enough, but we could try to investigate this issue further.



Looking at the residuals vs machines plot, we can affirm that the variability within operators is approximately the same for all processes. Although there are some deviation between the machines, in general, we can conclude that operator to operator variability is about the same in the three machines since the spreads of their residuals are approximately similar for all three machines.

14.20

```
> fit6 <- lm(Reflectance~factor(Day)*factor(Mix)*factor(Application), data = mix)
> anova(fit6)
Analysis of Variance Table

Response: Reflectance
                                          Df  Sum Sq Mean Sq F value Pr(>F)
factor(Day)                                2   2.042   1.021
factor(Mix)                                3 307.479 102.493
factor(Application)                        2 222.095 111.047
factor(Day):factor(Mix)                    6   4.529   0.755
factor(Day):factor(Application)            4   1.963   0.491
factor(Mix):factor(Application)            6  10.036   1.673
factor(Day):factor(Mix):factor(Application) 12  8.786   0.732
Residuals                                  0   0.000
```

Here, the block is "Day." To find the correct F-values and p-values:

<u>Mix</u>

F-value = 102.492/0.755 = 135.751

p-value = 1 – pf(135.751, 3, 6) = 6.658491 * 10^{-6} #used that R code

<u>Application</u>

F-value = 111.047/0.491 = 226.165

p-value = 1 – pf(226.165, 2, 4) = 7.68355 * 10^{-5} #used that R code

<u>Interaction</u>

F-value = 1.673/0.732 = 2.285519

p-value = 1 – pf(2.285519, 6, 12) = 0.1050986 #used that R code

Looking at the p-values at the 5% significance level, we conclude that both the mix and the application have a significant effect on reflectance, while their interactional effects is not significant.