

7/22

## STAT 425 Ass03

Daniel Girvitz

02/03/2022

5/4

5.4.

5.4.a. The F-statistic in the "Model" row of the ANOVA represents

1/3

5.5.a.

1.  $SSA = DFA * MSA = 50$ .
2.  $SSError = SST - SSA - SSB - SSAB = 172 - 50 - 80 - 30 = 12$ .
3.  $MSError = SSError / DFErro = 12 / 12 = 1$ .
4.  $F\_A = MSA / MSError = 50 / 1 = 50$ .
5.  $F\_B = MSB / MSError = 40 / 1 = 40$ .
6.  $F\_AB = MSAB / MSError = 15 / 1 = 15$ .

5.5.b.

Interpretation?

5.5.c.

5.7.a.

```
yield <- c(90.4,90.7,90.2,
           90.2,90.6,90.4,
           90.1,90.5,89.9,
           90.3,90.6,90.1,
           90.5,90.8,90.4,
           90.7,90.9,90.1)
temp <- rep(c("150","160","170"),each=6)
pressure <- rep(c("200","215","230"),6)
chemprocess <- data.frame(yield,temp,pressure)

chemprocess.aov <- aov(yield~factor(temp)+factor(pressure),data=chemprocess)
summary(chemprocess.aov)
```

```
##                Df Sum Sq Mean Sq F value    Pr(>F)
## factor(temp)    2  0.3011   0.1506    8.551 0.00426 **
## factor(pressure) 2  0.7678   0.3839   21.803 7.03e-05 ***
## Residuals      13  0.2289   0.0176
```

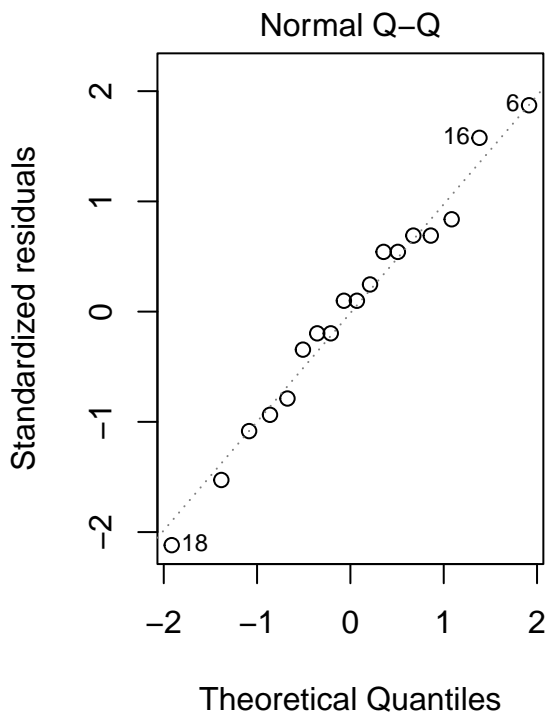
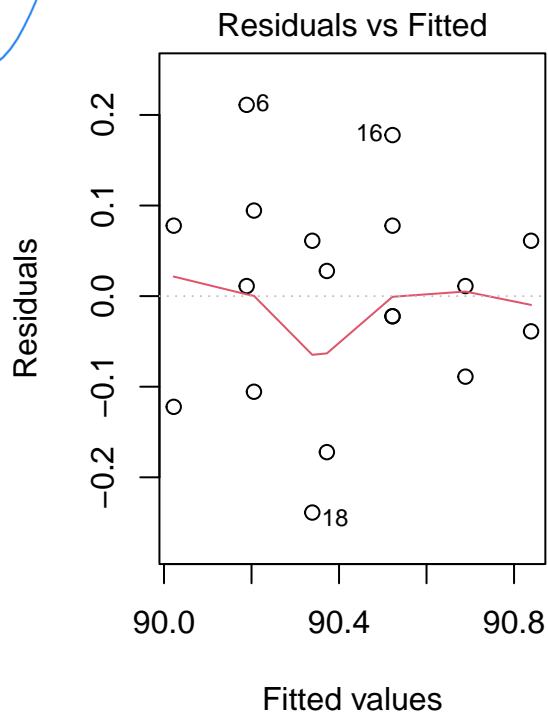
Interaction?

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At a significance level of  $\alpha=0.05$ , the null hypothesis of equality of row & column treatments, corresponding to temperature and pressure, is REJECTED, with p-values of 0.00426 & 7.03e-05.

### 5.7.b.

```
par(mfrow=c(1,2))
plot(chemprocess.aov,1)
plot(chemprocess.aov,2)
```



1. Residuals vs Fitted: Points appear to be, more or less, randomly spread, with increase or decrease in density in one direction or the other, so homogeneity of variance is preserved (homoscedasticity).
2. Normal Q-Q: Points fit the diagonal very well, so normality requirement is preserved.

```
# Numerical tests
```

### 5.7.c.

Constant conditions?

5.15.

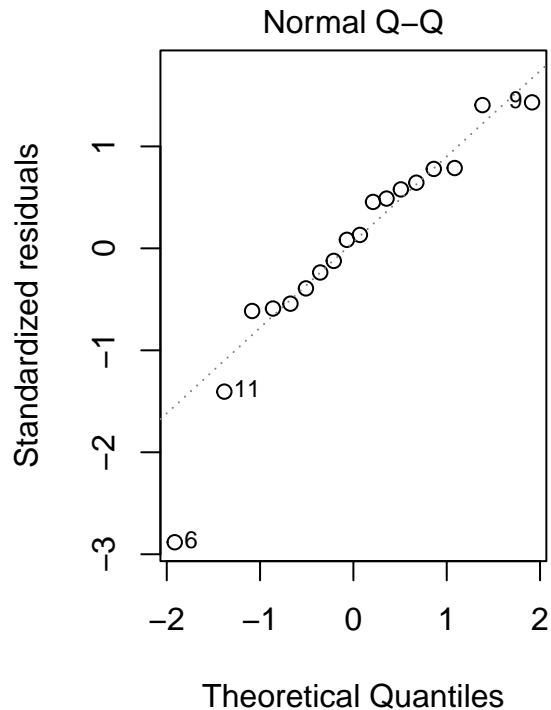
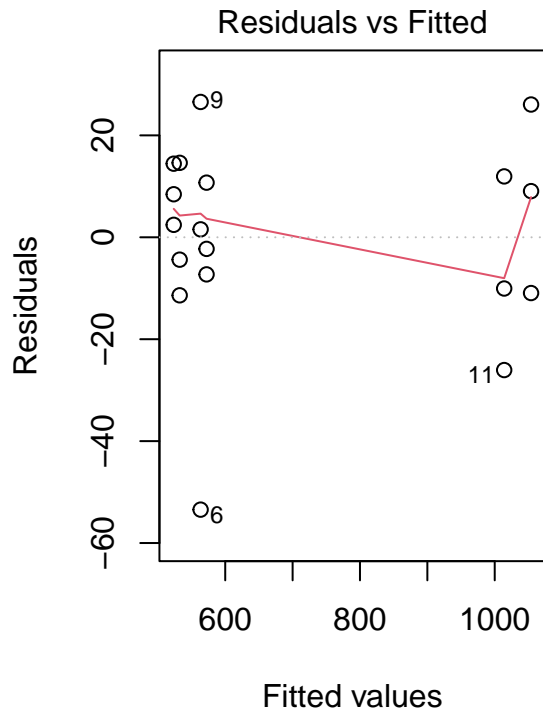
```
density <- c(570, 1063, 565,
            565, 1080, 510,
            583, 1043, 590,
            528, 988, 526,
            547, 1026, 538,
            521, 1004, 532)
position <- rep(c("1","2"),each=9) #row
temp <- rep(c("800","825","850"),6) #column
experiment <- data.frame(density,position,temp)

experiment.aov <- aov(density~factor(position)+factor(temp),data=experiment)
summary(experiment.aov)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(position)  1    7160      7160    16.2 0.00125 **
## factor(temp)      2 945342  472671   1069.3 4.92e-16 ***
## Residuals        14    6189      442
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At a significance level of  $\alpha=0.05$ , we REJECT the null hypothesis of equality of row & column treatment effects.

```
par(mfrow=c(1,2))
plot(experiment.aov,1)
plot(experiment.aov,2)
```



1. Residuals vs Fitted: Points appear to be, more or less, randomly spread, with increase or decrease in density in one direction or the other, so homogeneity of variance is preserved (homoscedasticity).
2. Normal Q-Q: Points fit the diagonal very well, so normality requirement is preserved. There is a bit of an outlier at the lower-left edge, but that is to fine. Numerical testing will deliver “better” analysis.

*# Numerical tests*

```
shapiro.test(experiment$density)
```

```
##
## Shapiro-Wilk normality test
##
## data:  experiment$density
## W = 0.70312, p-value = 8.683e-05
```

Shapiro-Wilk test REJECTS normality of data...

*1/2  
lower*

5.18.

```
strength <- c(9.60, 11.28, 9.00,
              9.69, 10.10, 9.57,
              8.43, 11.01, 9.03,
              9.98, 10.44, 9.80)
pressure <- rep(c("120", "130", "140", "150"), each=3) #row
```

*0/3*

```
temp <- rep(c("250","260","270"),4) #column
experiment <- data.frame(strength,pressure,temp)

experiment.aov <- aov(strength~pressure+temp,data=experiment)
summary(experiment.aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## pressure    3  0.581   0.1936   0.539 0.6727
## temp        2  4.658   2.3288   6.487 0.0316 *
## Residuals    6  2.154   0.3590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

X

At a significance level of  $\alpha=0.05$ , we FTR equality of treatment effects for pressure and REJECT for temperature.

5.25.

```
rate <- c(2.29, 2.06, 1.90,
          2.47, 2.05, 1.93,
          2.48, 2.23, 1.75,
          2.12, 2.03, 2.06,
          2.65, 3.20, 3.10,
          2.68, 3.18, 3.24,
          2.06, 3.96, 3.98,
          2.38, 3.64, 3.24,
          2.24, 11.00, 9.96,
          2.71, 11.00, 10.01,
          2.81, 9.06, 9.36,
          2.08, 11.30, 10.40)
freq <- rep(c("10","1","0.1"),each=12) #row
env <- rep(c("Air","H2O","Salt H2O"),12) #column

experiment <- data.frame(rate,freq,env)
```

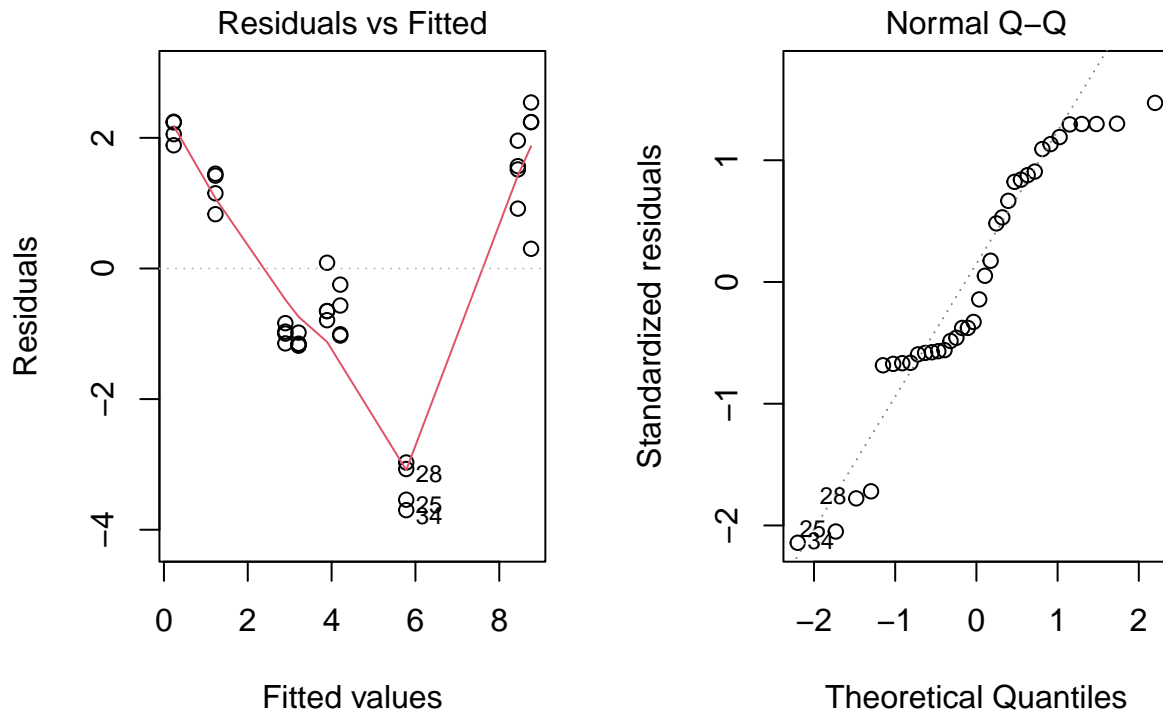
```
experiment.aov <- aov(rate~freq+env, data=experiment)
summary(experiment.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## freq        2 209.89  104.95  30.295 5.1e-08 ***
## env         2  64.25   32.13   9.274 0.000697 ***
## Residuals   31 107.39    3.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

interaction?

At a significance level of  $\alpha=0.05$ , we REJECT the null hypothesis of equality of row & column treatment effects.

```
par(mfrow=c(1,2))
plot(experiment.aov,1)
plot(experiment.aov,2)
```



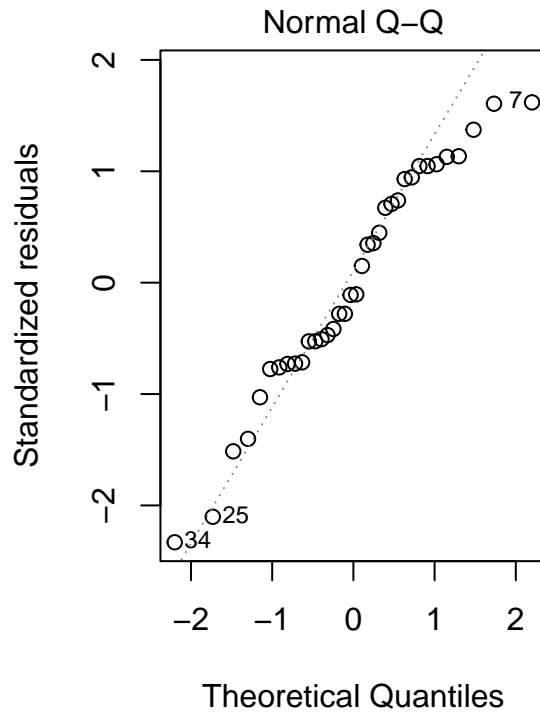
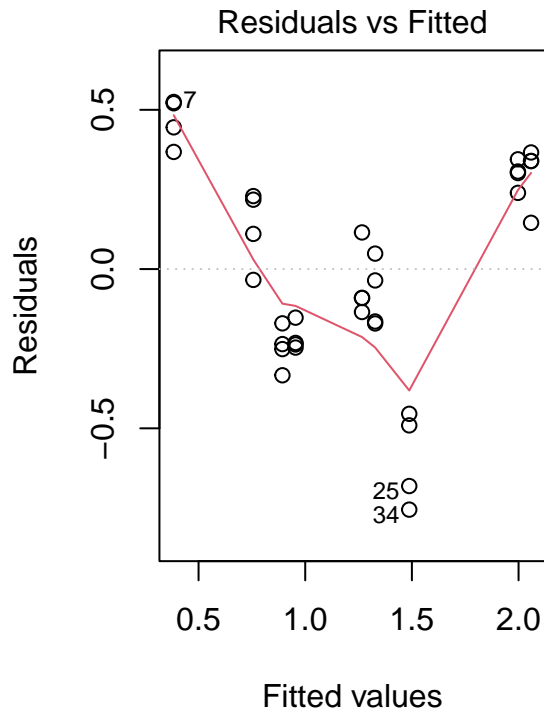
**Residuals vs Fitted** Homogeneity of variance (homoscedasticity) assumption is definitely violated.  
**Normal Q-Q** Normality of data is questionable – requires formal testing.

```
experiment.log.aov <- aov(log(rate)~freq+env, data=experiment)
summary(experiment.log.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## freq       2  7.570   3.785   31.031 3.98e-08 ***
## env        2  2.358   1.179    9.664 0.000547 ***
## Residuals  31  3.781   0.122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At a significance level of  $\alpha=0.05$ , we REJECT the null hypothesis of equality of row & column treatment effects.

```
par(mfrow=c(1,2))
plot(experiment.log.aov,1)
plot(experiment.log.aov,2)
```



**Residuals vs Fitted** Homogeneity of variance (homoscedasticity) assumption is still violated, but it is better.

**Normal Q-Q** Normality of data is still questionable – requires formal testing.

*# Numerical tests*

```
shapiro.test(experiment$rate)
```

```
##
## Shapiro-Wilk normality test
##
## data:  experiment$rate
## W = 0.67915, p-value = 1.352e-07
```

```
shapiro.test(log(experiment$rate))
```

```
##
## Shapiro-Wilk normality test
##
## data:  log(experiment$rate)
## W = 0.78983, p-value = 1.029e-05
```

The Shapiro-Wilk test rejects normality of data for both unmodified and log-modified data.