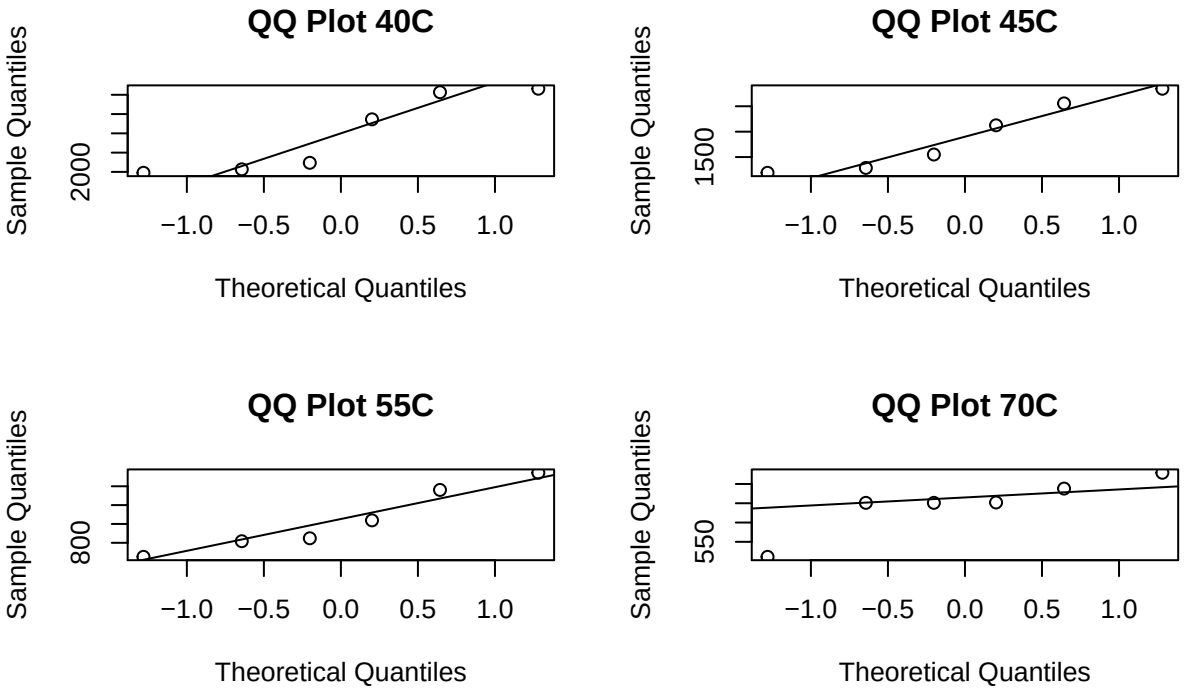
Homework 04
Joseph Blubaugh
jblubau1@tamu.edu
STAT 642-720

1.

a) The data from all 4 temperatures appear to be from different distributions with different variances. Running the shapiro test on each temperature group separately we can conclude that the data are normal. The Brown-Forsythe-Levene test supports the statement that the data have different variances and the shapiro test run on all groups together support that the data are not normal.




```
                          C40     C45       C55       C70
Shapiro Test p.value 0.1141215 0.44899 0.5764368 0.1348967

Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  3   23.43 9.404e-07 ***
      20
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


    Shapiro-Wilk normality test

data:  dt.melt$failure
W = 0.75293, p-value = 5.497e-05
```

b) Based on the regression model, an estimate for a power transformation is 3.459

```
(dt.sum = ddply(dt.melt, .(temp), summarize,
                S = var(failure),
                Y.bar = mean(failure)))
```

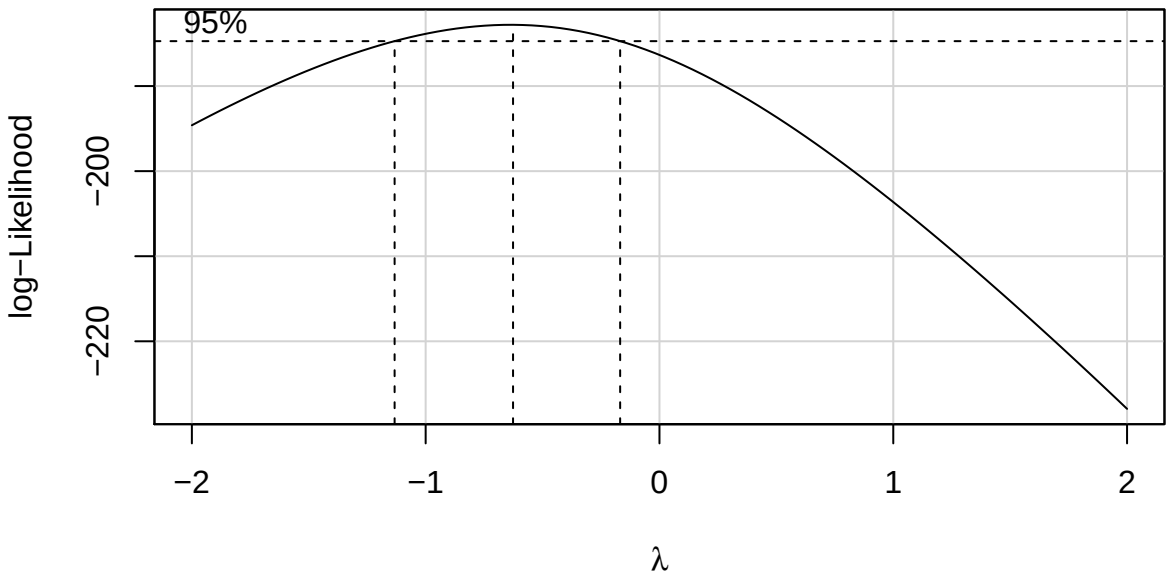
```
  temp         S    Y.bar
1   40 4086312.7 3955.667
2   45  474998.7 1925.500
3   55  118782.8 1043.000
4   70    5391.6  647.000
```

```
(mdl = lm(log(S) ~ log(Y.bar), data = dt.sum))
```

```
Call:
lm(formula = log(S) ~ log(Y.bar), data = dt.sum)

Coefficients:
(Intercept)   log(Y.bar)
    -13.166        3.459
```

c) The transformation estimate according to BoxCox is -.8



d) It can be determined by the SW test that using the box cox transformation makes the data normal. The BFL test also shows that the variance is stablized by the transformation and is acceptable to use in AOV.

```
    Shapiro-Wilk normality test

data:  dt.melt$fail.trans
W = 0.93694, p-value = 0.1393

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.8664 0.1678
      20
```

e) The AOV on the original data has a much higher F Statistic and slightly higher pvalue than the AOV for the transformed data. Since the data has been transformed, SS for both models are on two different scales and cannot be compared.

```
summary(aov(failure ~ temp, data = dt.melt))
```

```
            Df   Sum Sq  Mean Sq F value   Pr(>F)
temp         3 39183995 13061332   11.15 0.000162 ***
Residuals   20 23427429  1171371
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(fail.trans ~ temp, data = dt.melt))
```

```
            Df    Sum Sq   Mean Sq F value   Pr(>F)
temp         3 5.868e-05 1.956e-05   32.72 6.57e-08 ***
Residuals   20 1.196e-05 5.980e-07
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

f) All grouped pairs printed below are significant except for 45-40 which is not because the 95% confidence interval crosses 0.

```
TukeyHSD(aov(fail.trans ~ temp, data = dt.melt))
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = fail.trans ~ temp, data = dt.melt)

$temp
              diff           lwr         upr     p adj
45-40 0.0009800059 -0.0002694609 0.002229473 0.1586993
55-40 0.0025331376  0.0012836708 0.003782604 0.0000818
70-40 0.0041188274  0.0028693605 0.005368294 0.0000001
55-45 0.0015531317  0.0003036648 0.002802598 0.0116586
70-45 0.0031388214  0.0018893546 0.004388288 0.0000045
70-55 0.0015856898  0.0003362230 0.002835157 0.0099149
```

g) Based on the p-value it can be determined that only the linear trend is significant.

```
aovRes = aov(fail.trans ~ temp, data = dt.melt)
cntrMat = t(contr.poly(4,scores=c(40,45,55,70)))
summary(glht(aovRes, linfct=mcp(temp=cntrMat)))
```

```
     Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = fail.trans ~ temp, data = dt.melt)

Linear Hypotheses:
          Estimate Std. Error t value Pr(>|t|)
.L == 0  3.101e-03  3.157e-04   9.825   <1e-06 ***
.Q == 0 -4.011e-04  3.157e-04  -1.271     0.51
.C == 0  2.378e-05  3.157e-04   0.075     1.00
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

2.

a) Based on the Kruskal-Wallis test there are significant differences between the average failure time of the 4 temperatures

```
    Kruskal-Wallis rank sum test

data:  failure by temp
Kruskal-Wallis chi-squared = 18.278, df = 3, p-value = 0.0003853
```

b) With the Dunn procedure there are significant differences between the following pairs (40-45), (40-70), (45-70)

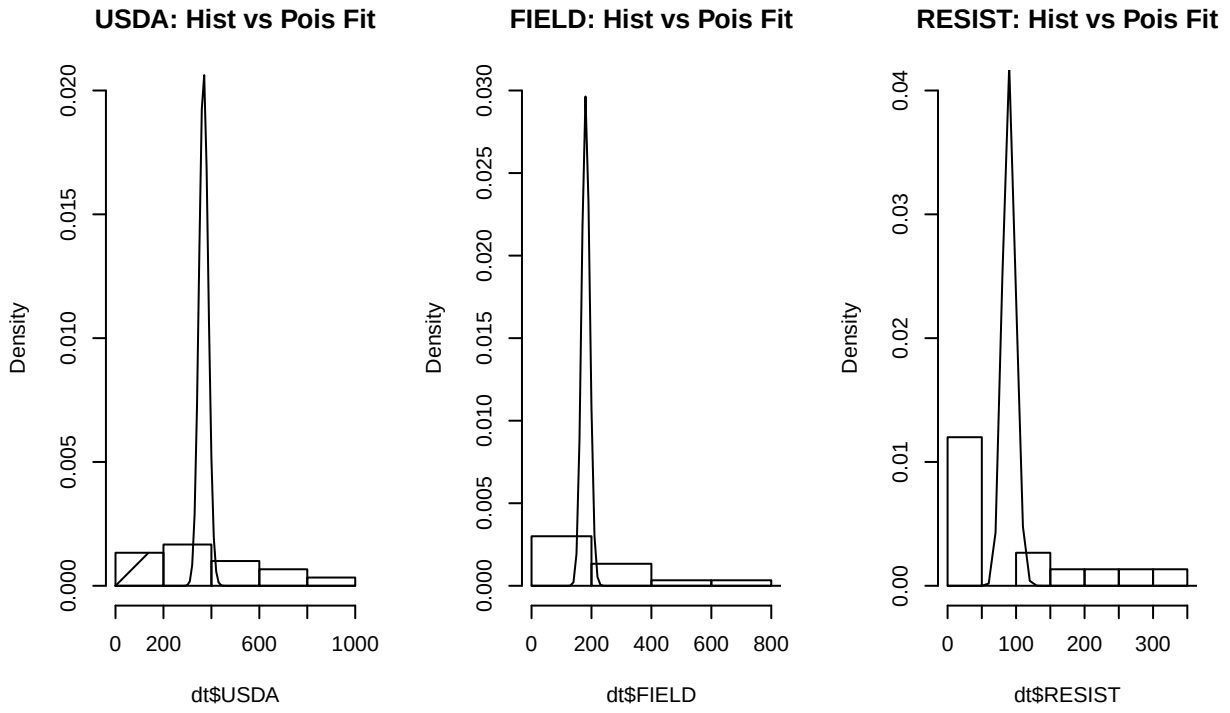
```
Multiple comparison test after Kruskal-Wallis
p.value: 0.05
Comparisons
        obs.dif critical.dif difference
40-45  4.333333     10.77064      FALSE
40-55 10.833333     10.77064       TRUE
40-70 16.166667     10.77064       TRUE
45-55  6.500000     10.77064      FALSE
45-70 11.833333     10.77064       TRUE
55-70  5.333333     10.77064      FALSE
```

c) AOV on the transformed data also detected a significant difference in the mean failure rates based on temperature, however the Tukey method for multiple comparison found more significant differences than the Dunn procedure.

Pairs	Tukey (transformed)	Dunn (original)
45-40		
55-40	***	***
70-40	***	***
55-45	***	
70-45	***	***
70-55	***	

3.

a) Poisson appears not to be a good distribution for these datasets after fitting a poisson model and plotting it against each data set.



b) Because the scaled deviance is so high, there is over dispersion and a transformation is needed. A log transformation works well in this case.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	42	8413	200.3
Scaled Deviance	42	42.71	1.01
Pearson Chi-Square	42	8273	196.9
Scaled Pearson X2	42	42	1
Log Likelihood		219.7	
Full Log Likelihood		-4332	
AIC (smaller is better)		8671	
AICC (smaller is better)		8672	
BIC (smaller is better)		8677	

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept	1	5.9081	0.1889	5.5378 6.2783	978.16	<.0001
type FIELD	1	-0.7081	0.3288	-1.3526 -0.0636	4.64	0.0313
type RESIST	1	-1.3994	0.4246	-2.2317 -0.5672	10.86	0.0010

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
type USDA	0	0.0000	0.0000	0.0000 0.0000		
Scale	0	14.0350	0.0000	14.0350 14.0350		

c) Based on the Durbin-Watson test of Autocorrelation, there is no evidence of correlation in the residuals for either of the 3 datasets

```
 lag Autocorrelation D-W Statistic p-value
   1      -0.1275745      2.121726   0.978
 Alternative hypothesis: rho != 0

 lag Autocorrelation D-W Statistic p-value
   1     -0.09605484      2.146113    0.97
 Alternative hypothesis: rho != 0

 lag Autocorrelation D-W Statistic p-value
   1      -0.1974338      2.322507    0.77
 Alternative hypothesis: rho != 0
```

4.

a) Model: $Y_{ij} = \mu + a_i + e_{ij}$ μ is the mean of all samples and a_i is the efect of t_i. $\mu + a_i$ is the mean of $treatment_i$. The expected mean squares are 1956.6 + 231.2 = 2187.8

```
Linear mixed model fit by REML ['lmerMod']
Formula: Aflatoxin ~ 1 + (1 | Lot.Number)
   Data: dt

REML criterion at convergence: 275.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.26902 -0.77902 -0.02608  0.72451  1.56485

Random effects:
 Groups     Name        Variance Std.Dev.
 Lot.Number (Intercept) 431.4    20.77
 Residual               231.2    15.20
Number of obs: 32, groups:  Lot.Number, 8

Fixed effects:
            Estimate Std. Error t value
(Intercept)    40.09       7.82   5.127
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
Lot.Number   7  13696  1956.6   8.464 3.2e-05 ***
Residuals   24   5548   231.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b) Variance within lots is 431.1, variance within sample differences is 231.2

c) 431.1 + 231.2 = 662.3

d) Proportion of variance in lots: 431.1 / 662.3 = .65, Proportion of variance in samples: 231.2 / 662.3 = .349

e) sqrt(431.1) = 20.76

f) You could continue to measure the increase in variance from adding additional samples until variance stabalizes, then you would know that you have an adequate number of samples

5.

a) Yes because it is assumed that the treatments are independent.

b) 5 reps for each of the 8 samples would be required to reach the desired power of .9

```
t = 8; r = 2:10; alpha = .01; gamma = .9; tau = 2; u1 = t-1
u2 = t*(r-1); lambda = sqrt(1 + (r*tau)); Fcr = qf(1-alpha, u1, u2)
C = (1/lambda)^2 * Fcr; power = 1 - pf(C, u1, u2)

data.frame(t, r, alpha, gamma, tau, u1, u2, lambda, Fcr, C, power)
```

```
  t  r alpha gamma tau u1 u2   lambda      Fcr         C     power
1 8  2  0.01   0.9   2  7  8 2.236068 6.177624 1.2355249 0.3834528
2 8  3  0.01   0.9   2  7 16 2.645751 4.025947 0.5751352 0.7659997
3 8  4  0.01   0.9   2  7 24 3.000000 3.495928 0.3884364 0.8999179
4 8  5  0.01   0.9   2  7 32 3.316625 3.258338 0.2962125 0.9503449
5 8  6  0.01   0.9   2  7 40 3.605551 3.123757 0.2402890 0.9724452
6 8  7  0.01   0.9   2  7 48 3.872983 3.037188 0.2024792 0.9833860
7 8  8  0.01   0.9   2  7 56 4.123106 2.976845 0.1751085 0.9893302
8 8  9  0.01   0.9   2  7 64 4.358899 2.932385 0.1543361 0.9928010
9 8 10  0.01   0.9   2  7 72 4.582576 2.898270 0.1380128 0.9949468
```

6. At minimum 5 treatments would be required to reach the desired power of .9

```
t = 2:10; r = 5; alpha = .01; gamma = .9; tau = (2.1/2.0)^2; u1 = t-1
u2 = t*(r-1); lambda = sqrt(1 + (r*tau)); Fcr = qf(1-alpha, u1, u2)
C = (1/lambda)^2 * Fcr; power = 1 - pf(C, u1, u2)

data.frame(t, r, alpha, gamma, tau, u1, u2, lambda, Fcr, C, power)
```

	t	r	alpha	gamma	tau	u1	u2	lambda	Fcr	C	power
1	2	5	0.01	0.9	1.1025	1	8	2.55196	11.258624	1.7287715	0.2250060
2	3	5	0.01	0.9	1.1025	2	12	2.55196	6.926608	1.0635867	0.3756262
3	4	5	0.01	0.9	1.1025	3	16	2.55196	5.292214	0.8126240	0.5053992
4	5	5	0.01	0.9	1.1025	4	20	2.55196	4.430690	0.6803363	0.6136614
5	6	5	0.01	0.9	1.1025	5	24	2.55196	3.895070	0.5980913	0.7017287
6	7	5	0.01	0.9	1.1025	6	28	2.55196	3.527559	0.5416597	0.7719918
7	8	5	0.01	0.9	1.1025	7	32	2.55196	3.258338	0.5003206	0.8271899
8	9	5	0.01	0.9	1.1025	8	36	2.55196	3.051726	0.4685951	0.8700052
9	10	5	0.01	0.9	1.1025	9	40	2.55196	2.887560	0.4433874	0.9028619

7.

a) The assistant is not correct, positive correlation in the residuals will understate variance and lower the power of the F-test
b) Variance is different between the treatments, the data needs to be transformed to stablize variance. A box-cox transformation is suggested.
c) The Hsu procedure works by looking at the confidence intervals of an estimator, when you change the scale you no longer have a symmetrical confidence interval, it is likely right skewed.
d) The data must have small differences in variance. A log transformation should be applied to the data if the variation between treatments is large. The F test assumes normality, KW only does need normality, but does need stable variance
e) To control variation within each manufacturer, blocking should be done on manufacturer
f) The two treatments have different sample sizes and different degrees of freedom which are accounted for in the MS of each statistic so the ratio is correct.
g) Since the range of errors are all within the range of the MSE, I would not consider any of the residuals outliers
h) If you have equal sample sizes you are more likely to have the same variance if your samples comes from the same population.