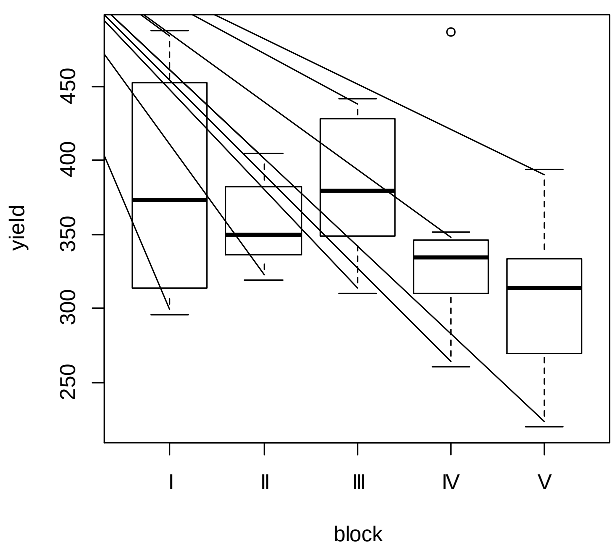
1.

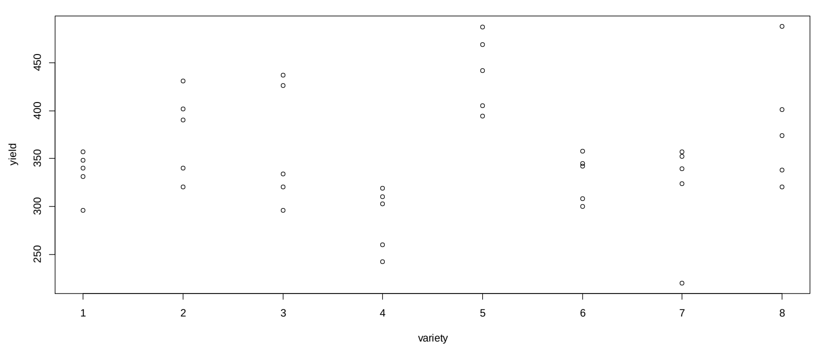
(a)

Block Designs

(b)

Interaction exists between few couples





(c)

H0:There is no significant differentiating between samples.

H1: There is significant differentiating between samples.

F(7.32) at 0.05 level of significance=2.3127

F=5.1057 which is larger than 2.3127, so H0 is rejected-there is significant differentiating between samples.

```
Analysis of Variance Table

Response: yield
          Df Sum Sq Mean Sq F value  Pr(>F)
block      4  33396  8348.9  2.4777 0.06252 .
variety    1    388   388.1  0.1152 0.73640
Residuals 34 114569  3369.7
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:
lm(formula = yield ~ ., data = oat)

Residuals:
   Min     1Q Median     3Q    Max
-89.40 -37.96  -2.64  38.15 145.32

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  376.882     27.316  13.797 1.72e-15 ***
blockII      -25.500     29.024  -0.879    0.386
blockIII       0.125     29.024   0.004    0.997
blockIV      -42.000     29.024  -1.447    0.157
blockV       -77.000     29.024  -2.653    0.012 *
variety        1.359      4.006   0.339    0.736
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 58.05 on 34 degrees of freedom
```

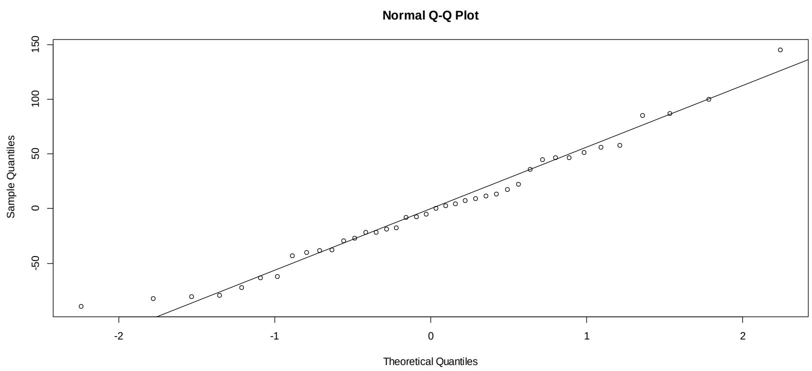
Multiple R-squared: 0.2277,   Adjusted R-squared: 0.1142

F-statistic: 2.005 on 5 and 34 DF, p-value: 0.1029

(d)

Residual plots of the original full regression model:

Indicating that it does not satisfied the 3+1 assumptions on residuals



Try doing some transformation on the first model:

> oat.2<-lm(log(yield)~block+variety,oat)

> anova(oat.2)

Analysis of Variance Table

Response: log(yield)

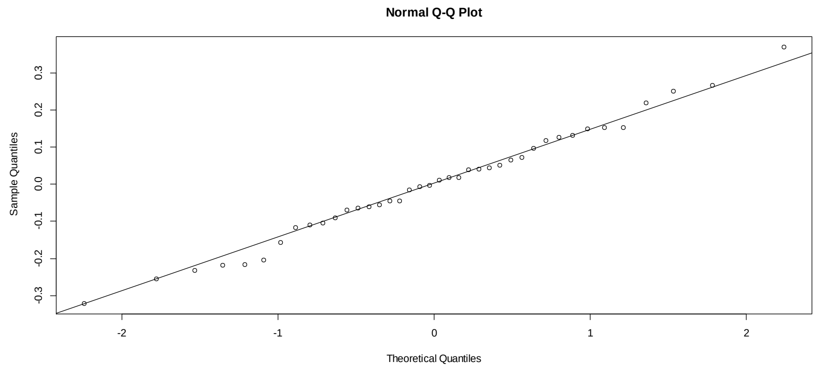
```
          Df  Sum Sq  Mean Sq F value  Pr(>F)
block      4 0.29185 0.072963  2.7308 0.04504 *
variety    1 0.00126 0.001262  0.0472 0.82925
Residuals 34 0.90843 0.026718
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> summary(oat.2)

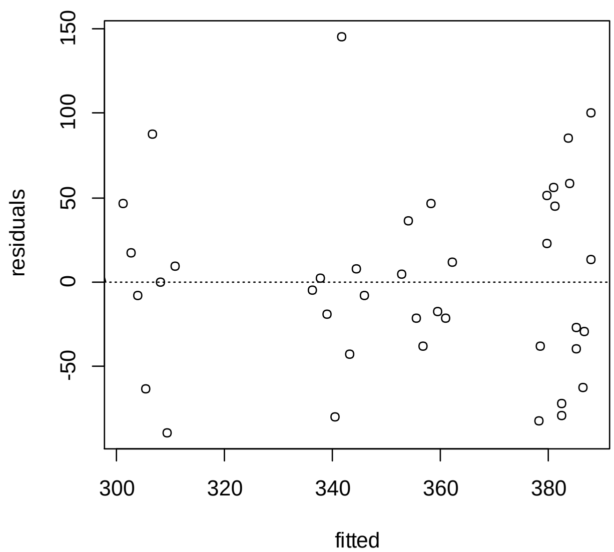
Call:

```
lm(formula = log(yield) ~ block + variety, data = oat)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32103 -0.09405  0.00364  0.10133  0.36969

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.919679   0.076917  76.962   <2e-16 ***
blockII     -0.054564   0.081729  -0.668   0.5089
blockIII     0.010490   0.081729   0.128   0.8986
blockIV     -0.113362   0.081729  -1.387   0.1745
blockV      -0.222183   0.081729  -2.719   0.0102 *
variety      0.002451   0.011280   0.217   0.8292
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1635 on 34 degrees of freedom
Multiple R-squared:  0.2439,    Adjusted R-squared:  0.1328
F-statistic: 2.194 on 5 and 34 DF,  p-value: 0.0778
```

QQ plot of transforemd model:



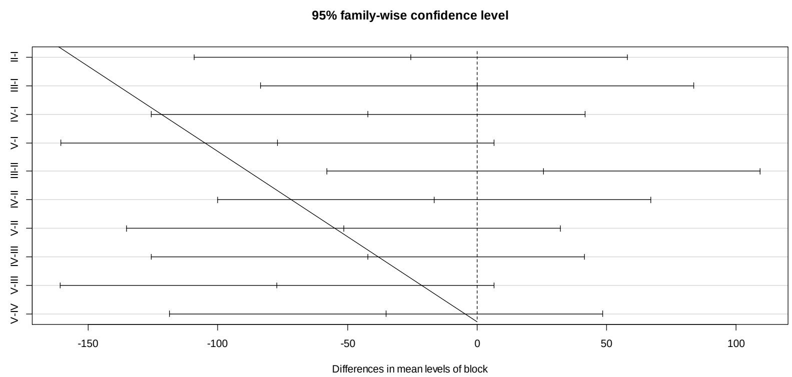
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles



residuals
fitted

(e)

Yes. it is necessary to perform multiple comparisons



2.

(a)

Latin square design

(b)

Linear model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}$$
$$i, j, k = 1, 2, 3, 4$$

(c)

Conducting null hypothesis for positions, run and fabric area. by doing so we will accept all of them at 5% degree of significance. Our null hypothesis: positions don't differ, run don't differ, and fabric area don't differ.

ANOVA table					
source	d.f	sum of squares	Mean SS	F	F3,9 0.05
position	3	6596	2198.667	0.239622	3.86
run	3	6404	2134.667	0.232647	

fabric Area	3	6420	2140	0.233228	
error	9	82580	9175.556		
Total	15	102000			

3.

(a)

```
Call:
lm(formula = Sales ~ ., data = tp)

Residuals:
     1       2       3       4       5       6       7       8       9
 1.4444 -1.8889  0.4444  0.4444  1.4444 -1.8889 -1.8889  0.4444  1.4444

Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     -12.444      9.950  -1.251   0.3375
Week              5.333      1.210   4.409   0.0478 *
StoreLocationS   22.667      4.361   5.198   0.0351 *
StoreLocationU   49.000      7.650   6.405   0.0235 *
StoreTypeDI      13.667      2.095   6.524   0.0227 *
StoreTypeGR          NA         NA      NA       NA
DisplayB         14.000      2.419   5.787   0.0286 *
DisplayC          7.667      2.419   3.169   0.0868 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.963 on 2 degrees of freedom
```

Multiple R-squared: 0.9865,   Adjusted R-squared: 0.9459

F-statistic: 24.29 on 6 and 2 DF, p-value: 0.04006

At least one type of each variable, from the table above, has a p-value of less than 0.10. Hence, for all three factors, there is a substantial difference between the groups. So, we wouldn't exclude a variable if we were using the backward stepwise regression with alpha to leave 0.1.

(b)

From the data, it is observed that store type DE, location U and display B together appear to maximize sales value:67 toothpastes.

(c)

We are not able to estimate the probability that 80 or more toothpaste will be sold during

a week based on the given information.

(d)

The model ran into perfect multicollinearity. ''Alias'' refers to the variables that are linearly dependent on others (i.e. cause perfect multicollinearity).

```
> vif(tp.0)
Error in vif.default(tp.0) : there are aliased coefficients in the model
> alias(tp.0)
Model :
Sales ~ Week + Locat + Type + Display


Complete :
       (Intercept) Week LocatS LocatU TypeDI DisplayB DisplayC
TypeGR  9/2        -1/2 -3/2    -3   -1/2     0        0
```