

## 02441 Applied Statistics and Statistical Software

### Exercise 3D - Fish

To study the influence of ocean grazers on regeneration rates of seaweed in the intertidal zone, a researcher scraped rock plots free of seaweed and observed the degree of regeneration when certain types of seaweed-grazing animals were denied access. The grazers were limpets (L), small fishes (f) and large fishes (F). A plot was taken to be a square rock surface, 100 cm on each side. Each plot received one of six treatments, named here by which grazers were allowed access.

symbol	Description
LfF	All three grazers were allowed access
fF	Limpets were excluded by surrounding the plot with caustic paint
Lf	Large fish were excluded by covering the plot with a course net
f	Limpets and large fish were excluded
L	Small and large fish were excluded by covering the plot with a fine net
C	Control: limpets, small fish and large fish were all excluded

Because the intertidal zone is a highly variable environment, the researcher applied the treatments in eight blocks of twelve plots each. Within each block she randomly assigned treatments to plots so that each treatment was applied to two plots. The blocks covered a wide range of tidal conditions.

Symbol	Description
Block 1	just below high tide level, exposed to heavy surf
Block 2	just below high tide level, protected from the surf
Block 3	midtide, exposed
Block 4	midtide, protected
Block 5	just above low tide level, exposed
Block 6	just above low tide level, protected
Block 7	on near-vertical rock wall, midtide level, protected
Block 8	on near-vertical rock wall, above low tide level, protected

### Variables

Variable name	Description
cover	Percentage of regenerated seaweed
block	Blocks for different tidal situations
treat	Treatment, i.e. what fishes are excluded

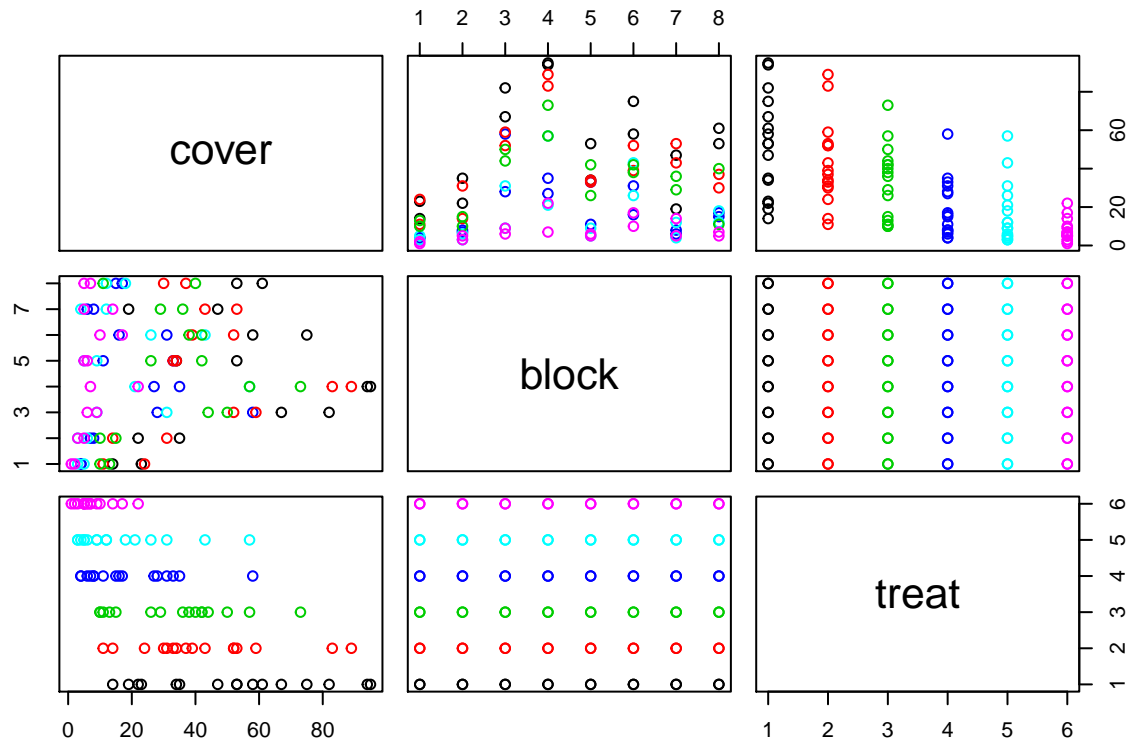
1. Make appropriate plots to investigate whether treatments or blocks have any influence on the percentage of regenerated seaweed

Start by loading and inspecting data

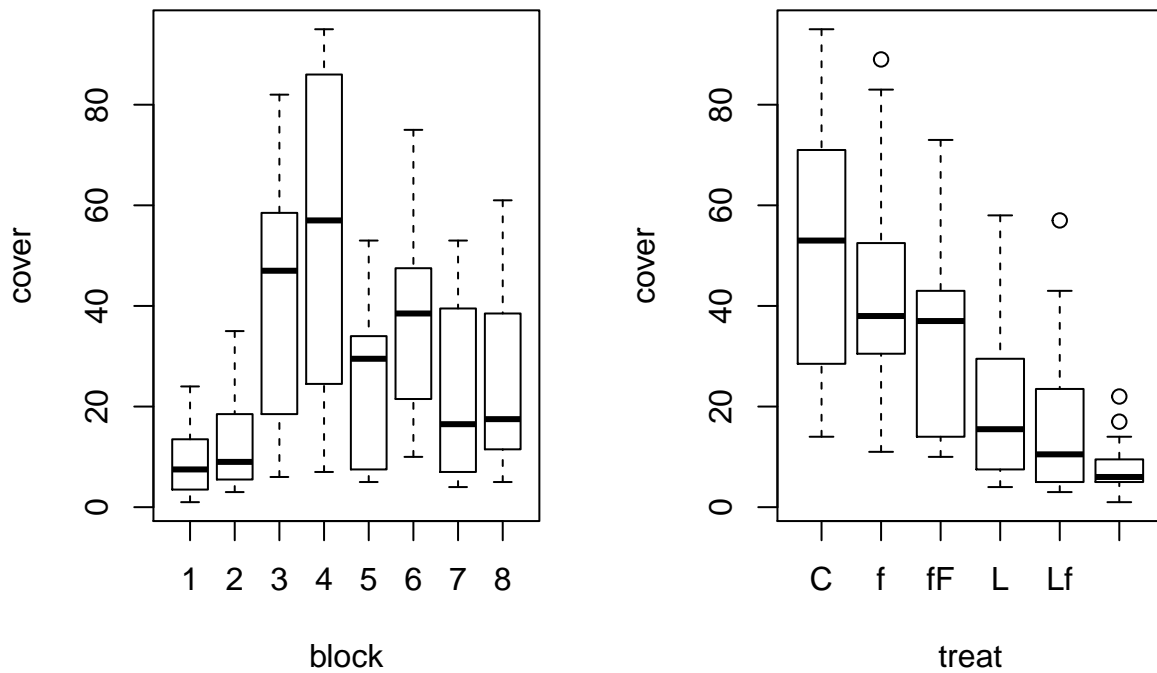
```
fish <- read.table("fishgrazer.txt", header = TRUE,  
                  colClasses = c('numeric', 'factor', 'factor'))
```

## Informative plots

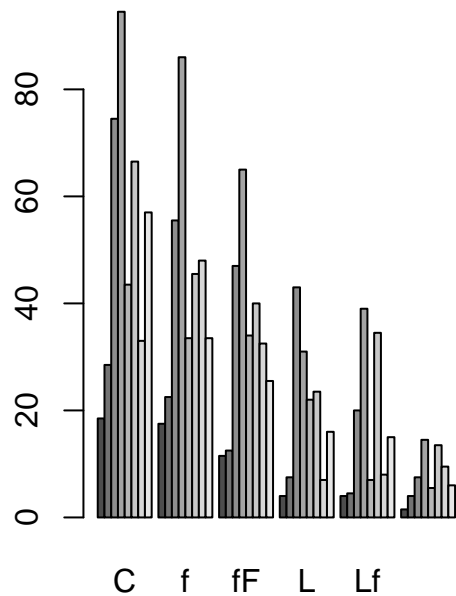
```
pairs(fish, col = fish$treat)
```



```
par(mfrow=c(1,2))
plot(cover~block+treat,fish)
```



```
barplot(tapply(fish$cover,list(fish$block,fish$treat),mean),beside=TRUE)
```



## 2. Test the model using a 5% significance level

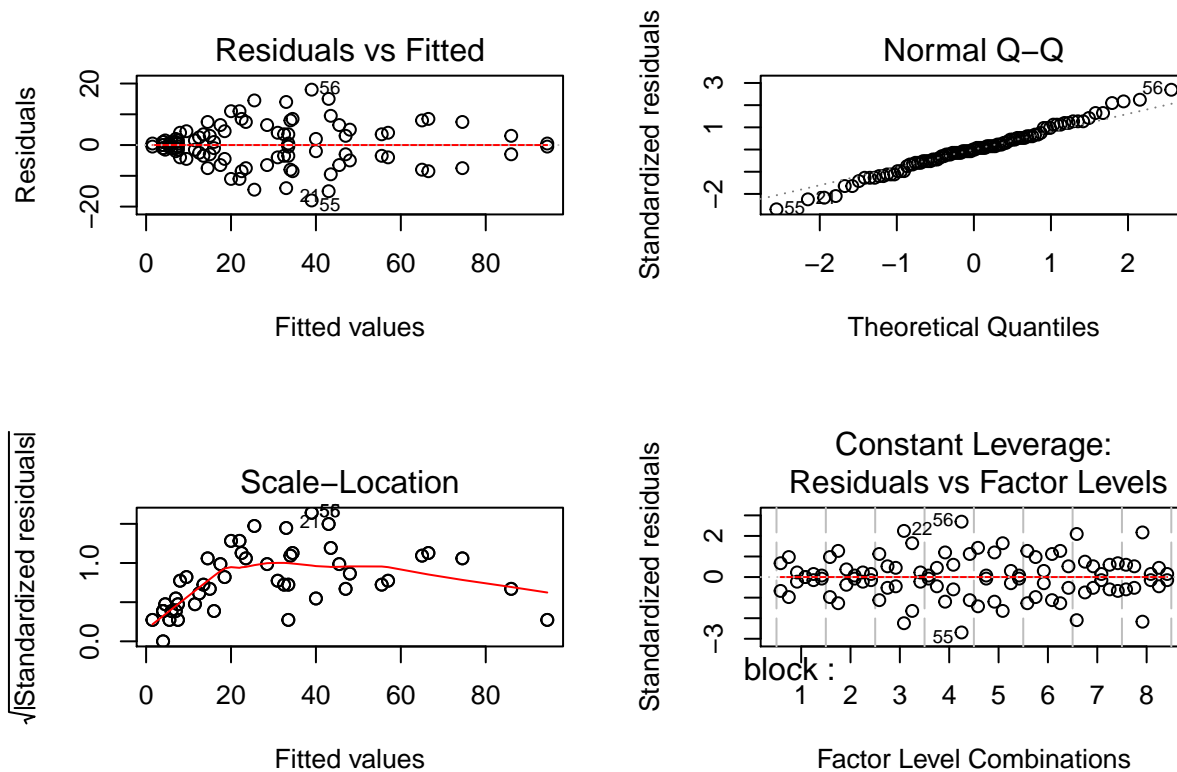
Start with a full model with interactions

```
lm1 <- lm(cover~.^2,fish)
anova(lm1)
```

```
## Analysis of Variance Table
##
## Response: cover
##          Df Sum Sq Mean Sq F value    Pr(>F)
## block      7 19105.5   2729.4  30.5454 1.296e-15 ***
## treat      5 23045.5   4609.1  51.5824 < 2.2e-16 ***
## block:treat 35  6612.5    188.9   2.1144 0.008128 **
## Residuals  48  4289.0     89.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model diagnostics

```
par(mfrow=c(2,2))
plot(lm1)
```



One can observe clear signs of variance inhomogeneity from the Residual vs Fitted and the Scale-Location plot. The seaweed cover (response variable) was measured in % and was bound between 0 and 100. It is therefore desirable to transform the response variable using the logit function  $\log(\frac{y}{1-y})$ .

```
# Convert percentages to values between 0 and 1
fish$logitcover <- fish$cover/100
# Apply logit to bound (back-transformed) predictions between 0 and 1
fish$logitcover <- log(fish$logitcover / (1-fish$logitcover))
lm2a <- lm(logitcover~treat*block,fish)
anova(lm2a)
```

```
## Analysis of Variance Table
##
## Response: logitcover
##           Df Sum Sq Mean Sq F value Pr(>F)
## treat      5  96.993  19.3986  64.0553 <2e-16 ***
## block      7  76.239  10.8912  35.9634 <2e-16 ***
## treat:block 35  15.230   0.4352   1.4369 0.1209
## Residuals  48  14.536   0.3028
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Remove insignificant interaction term
lm2b <- update(lm2a, .~.-treat:block)
anova(lm2b)
```

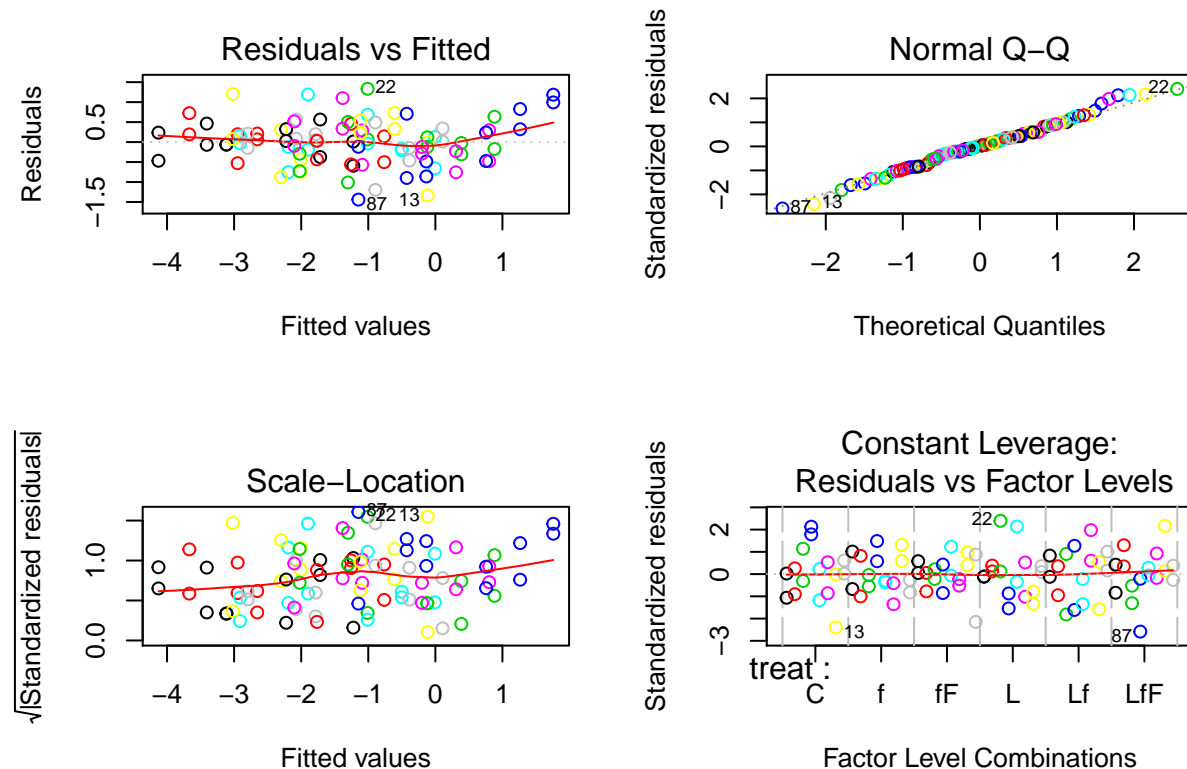
```
## Analysis of Variance Table
##
## Response: logitcover
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## treat      5 96.993 19.3986  54.090 < 2.2e-16 ***
## block      7 76.239 10.8912  30.368 < 2.2e-16 ***
## Residuals 83 29.767  0.3586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The final model includes the two main effects, treat and block. The response variable, cover, has been logit transformed to make sure that the linear model can only predict cover between 0 and 1, which in turn relate to outcomes between 0% and 100% seaweed coverage.

### Model diagnostics

```
par(mfrow=c(2,2))
plot(lm2b, col = fish$block)
```



No systematic patterns are left in the diagnostics plots. The model assumptions are fulfilled.

### Interpret model

```
summary(lm2b)
```

```
##
## Call:
## lm(formula = logitcover ~ treat + block, data = fish)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43960 -0.38863  0.01354  0.32217  1.33326
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.2226     0.2204  -5.548 3.37e-07 ***
## treatf       -0.4941     0.2117  -2.334  0.0220 *
## treatfF      -1.0019     0.2117  -4.732 9.03e-06 ***
## treatL       -1.8925     0.2117  -8.938 8.68e-14 ***
## treatLf      -2.1849     0.2117 -10.319 < 2e-16 ***
## treatLfF     -2.9052     0.2117 -13.721 < 2e-16 ***
## block2        0.4600     0.2445   1.881  0.0634 .
## block3        2.1046     0.2445   8.608 3.97e-13 ***
## block4        2.9807     0.2445  12.192 < 2e-16 ***
## block5        1.2160     0.2445   4.974 3.49e-06 ***
## block6        2.0251     0.2445   8.283 1.77e-12 ***
## block7        1.1085     0.2445   4.534 1.93e-05 ***
## block8        1.3300     0.2445   5.440 5.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5989 on 83 degrees of freedom
## Multiple R-squared:  0.8534, Adjusted R-squared:  0.8322
## F-statistic: 40.25 on 12 and 83 DF, p-value: < 2.2e-16
```

### 3. Is there evidence of any interaction effects?

No. See model lm2b above.

How about trying to split block into two categorical variables, i.e. zone and exposure?

```
fish$exposure <- factor(as.numeric(fish$block)%2==0, labels = c('True', 'False'))
fish$zone <- fish$block
levels(fish$zone) <- c(1,1,2,2,3,3,4,4)
lm3a <- lm(logitcover~treat*exposure*zone, fish)
drop1(lm3a, test = 'F')
```

```
## Single term deletions
##
## Model:
## logitcover ~ treat * exposure * zone
##              Df Sum of Sq    RSS      AIC F value Pr(>F)
## <none>                        14.536 -85.218
## treat:exposure:zone 15      6.5786 21.115 -79.379  1.4482 0.1641
lm3b <- update(lm3a, .~-treat:exposure:zone)
drop1(lm3b, test = 'F')
```

```
## Single term deletions
##
## Model:
```

```

## logitcover ~ treat + exposure + zone + treat:exposure + treat:zone +
##   exposure:zone
##           Df Sum of Sq   RSS   AIC F value Pr(>F)
## <none>                21.115 -79.379
## treat:exposure   5     2.7506 23.866 -77.623  1.6413 0.1621
## treat:zone      15     5.9012 27.016 -85.719  1.1738 0.3151
## exposure:zone    3     1.6951 22.810 -77.966  1.6858 0.1790

lm3c <- update(lm3b, .~-treat:zone)
drop1(lm3c, test = 'F')

## Single term deletions
##
## Model:
## logitcover ~ treat + exposure + zone + treat:exposure + exposure:zone
##           Df Sum of Sq   RSS   AIC F value Pr(>F)
## <none>                27.016 -85.719
## treat:exposure   5     2.7506 29.767 -86.411  1.5883 0.1732
## exposure:zone    3     1.6951 28.711 -85.877  1.6313 0.1889

lm3d <- update(lm3c, .~-exposure:zone)
drop1(lm3d, test = 'F')

## Single term deletions
##
## Model:
## logitcover ~ treat + exposure + zone + treat:exposure
##           Df Sum of Sq   RSS   AIC F value Pr(>F)
## <none>                28.711 -85.877
## zone              3    66.142 94.853  22.846 62.199 <2e-16 ***
## treat:exposure   5     2.751 31.462 -87.095  1.552 0.1831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm3e <- update(lm3d, .~-treat:exposure)
anova(lm3e)

## Analysis of Variance Table
##
## Response: logitcover
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat       5 96.993 19.3986   53.026 < 2.2e-16 ***
## exposure    1  8.402  8.4017   22.966 6.833e-06 ***
## zone        3 66.142 22.0473   60.265 < 2.2e-16 ***
## Residuals  86 31.462  0.3658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lm3e)

##
## Call:
## lm(formula = logitcover ~ treat + exposure + zone, data = fish)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38130 -0.33468 -0.00006  0.38979  1.38907

```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.2885     0.1952  -6.600 3.21e-09 ***
## treatf       -0.4941     0.2138  -2.311  0.0232 *
## treatfF      -1.0019     0.2138  -4.685 1.04e-05 ***
## treatL       -1.8925     0.2138  -8.850 9.95e-14 ***
## treatLf      -2.1849     0.2138 -10.217 < 2e-16 ***
## treatLfF     -2.9052     0.2138 -13.585 < 2e-16 ***
## exposureFalse 0.5917     0.1235   4.792 6.83e-06 ***
## zone2         2.3127     0.1746  13.245 < 2e-16 ***
## zone3         1.3906     0.1746   7.964 6.25e-12 ***
## zone4         0.9893     0.1746   5.666 1.91e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6048 on 86 degrees of freedom
## Multiple R-squared:  0.845, Adjusted R-squared:  0.8288
## F-statistic: 52.1 on 9 and 86 DF, p-value: < 2.2e-16
```

Compare model fit of lm2b and lm3e

```
AIC(lm2b, lm3e)
```

```
##           df           AIC
## lm2b 14 188.0247
## lm3e 11 187.3415
```