1. (One possible answer)

(a) $H_0 : \alpha\beta_{12} - \alpha\beta_{11} - (\alpha\beta_{22} - \alpha\beta_{21}) = 7$.
For $C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 \end{bmatrix}$ there exists $A = \begin{bmatrix} -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$ such that $C = AX$. Therefore, $C\underline{\beta}$ is estimable. Since $rank(C) = dim(C) = 1$, $C\underline{\beta} = \underline{d}$ is testable, where $\underline{d} = 7$.

(b) $H_0 : \alpha\beta_{12} - \alpha\beta_{11} - (\alpha\beta_{22} - \alpha\beta_{21}) = 0$ is testable. $C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 \end{bmatrix}$ and $\underline{d} = 0$.
(From Christensen p. 54) If we impose the constraint $-\alpha\beta_{11} + \alpha\beta_{12} + \alpha\beta_{21} = \alpha\beta_{22}$, we get

$$Y = \mu \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} + \alpha_1 \begin{bmatrix}1\\1\\1\\1\\0\\0\\0\\0\end{bmatrix} + \alpha_2 \begin{bmatrix}0\\0\\0\\0\\1\\1\\1\\1\end{bmatrix} + \beta_1 \begin{bmatrix}1\\1\\0\\0\\1\\1\\0\\0\end{bmatrix} + \beta_2 \begin{bmatrix}0\\0\\1\\1\\0\\0\\1\\1\end{bmatrix} + \alpha\beta_{11} \begin{bmatrix}1\\1\\0\\0\\0\\0\\0\\0\end{bmatrix} + \alpha\beta_{12} \begin{bmatrix}0\\0\\1\\1\\0\\0\\0\\0\end{bmatrix} + \alpha\beta_{21} \begin{bmatrix}0\\0\\0\\0\\1\\1\\0\\0\end{bmatrix} + (-\alpha\beta_{11} + \alpha\beta_{12} + \alpha\beta_{21}) \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\1\end{bmatrix} + e.$$

$$Y = \mu \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} + \alpha_1 \begin{bmatrix}1\\1\\1\\1\\0\\0\\0\\0\end{bmatrix} + \alpha_2 \begin{bmatrix}0\\0\\0\\0\\1\\1\\1\\1\end{bmatrix} + \beta_1 \begin{bmatrix}1\\1\\0\\0\\1\\1\\0\\0\end{bmatrix} + \beta_2 \begin{bmatrix}0\\0\\1\\1\\0\\0\\1\\1\end{bmatrix} + \alpha\beta_{11} \left(\begin{bmatrix}1\\1\\0\\0\\0\\0\\0\\0\end{bmatrix} - \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\1\end{bmatrix} \right) + \alpha\beta_{12} \left(\begin{bmatrix}0\\0\\1\\1\\0\\0\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\1\end{bmatrix} \right) + \alpha\beta_{21} \left(\begin{bmatrix}0\\0\\0\\0\\1\\1\\0\\0\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\1\end{bmatrix} \right) + e.$$

$$Y = X_0\underline{\gamma} + e, \text{ where } X_0 = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & -1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & -1 & 1 & 1 \end{bmatrix}.$$

(c) $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha\beta_{21} = \alpha\beta_{22} = 0$. For $\underline{a}' = (-3, 1, 1, 1, 1, 1, 1, 1, 1)$, $X\underline{a} = \underline{0}$, but $\underline{c}_i'\underline{a} = 1 \neq 0$, where $\underline{c}_i'$ is the i^{th}- row of the C matrix. Therefore, $\underline{c}_i'\underline{\beta}$ is not estimable, for $i = 1, \ldots, 4$ and thus not testable.

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(d) $H_0 : \mu + \alpha_1 = 7$. For $\underline{a}' = (-3, 1, 1, 1, 1, 1, 1, 1, 1)$, $X\underline{a} = 0$, but for $C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $C\underline{a} = -2 \neq 0$. Therefore, $\mu + \alpha_1$ is not estimable, and thus not testable.

(e) $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha\beta_{21} = \alpha\beta_{22} = 0$ and $\alpha_1 - \alpha_2 = 7$ is not testable.

(f) $H_0 : \alpha\beta_{11} = \alpha\beta_{12} = \alpha\beta_{21} = \alpha\beta_{22} = 0$ and $\alpha_1 - \alpha_2 = 0$ is not testable.

2. Graphs were omitted in this solution.

(a) Which single predictor do you think is probably the best predictor of `biomass`?
`pH` or `Zn` could be best single predictors of `biomass`.

(b) Do you see any evidence of multicollinearity?
Yes, `Na` and `Zn` seem to have a strong positive correlation.

```
> round(cor(biomass[-(1:2)]),4)
          biomass salinity      pH       K      Na      Zn
biomass    1.0000  -0.1032  0.7742 -0.2051 -0.2721 -0.6244
salinity  -0.1032   1.0000 -0.0513 -0.0205  0.1623 -0.4208
pH         0.7742  -0.0513  1.0000  0.0187 -0.0378 -0.7222
K         -0.2051  -0.0205  0.0187  1.0000  0.7921  0.0740
Na        -0.2721   0.1623 -0.0378  0.7921  1.0000  0.1171
Zn        -0.6244  -0.4208 -0.7222  0.0740  0.1171  1.0000
```

(c) What does `fin=c(6.0,6.0)` do?
From `help(par)`: "fin: A numerical vector of the form 'c(x, y)' which gives the size of the figure region in inches".

(d) What happens when you change the smoothing parameter in `loess.smooth()` from 0.90 to 0.50?
We obtain a less smooth line.

(e) Create a matrix that has $\underline{b}_{OLS}$ in the first column and a vector of corresponding standard error in the second.

```
MSE <- (t(e)%*%e)/(length(Y)-qr(X)$rank)
cov.b <- as.numeric(MSE)*solve(t(X)%*%X)
labels <- c("Intercept","Salinity","pH","K","Na","Zn")
results <- round(cbind(b,sqrt(diag(cov.b))),4)
temp <- cbind(labels,results)
colnames(results) <- c("Estimate","Std Error")
colnames(temp) <- c("         ","Estimate ","Std Error")
library(MASS)
write.matrix(temp,file="c:/My Documents/TA511/hw04.out")
```

```
          Estimate  Std Error
Intercept 1252.4546 1234.7101
Salinity  -30.2849  24.03
pH        305.488   87.881
K         -0.2853   0.3483
Na        -0.0087   0.0159
Zn        -20.6768  15.0541
```

In the file `hw04.out` the numbers are not aligned to the right This is because all entries in a matrix are coerced to be of the same type and since we included a column with labels, then all columns are considered names and therefore aligned to the left. Furthermore, it is no longer possible to use the numeric columns of the matrix `temp` for numeric operations.

```
> results
          Estimate Std Error
X0       1252.4546 1234.7101
salinity  -30.2849   24.0300
pH        305.4880   87.8810
K          -0.2853    0.3483
Na         -0.0087    0.0159
Zn        -20.6768   15.0541
```

On the other hand, the matrix `results` is still numeric and could be used for further operations if we needed to.

3.
```
fit1 <- lm(biomass~salinity+pH+K+Na+Zn,data=biomass)
> summary(fit1)
lm(formula = biomass ~ salinity + pH + K + Na + Zn, data = biomass)

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.252e+03  1.235e+03   1.014  0.31666
salinity    -3.028e+01  2.403e+01  -1.260  0.21505
pH           3.055e+02  8.788e+01   3.476  0.00126 **
K           -2.853e-01  3.483e-01  -0.819  0.41776
Na          -8.666e-03  1.593e-02  -0.544  0.58953
Zn          -2.068e+01  1.505e+01  -1.373  0.17744
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 398.3 on 39 degrees of freedom
Multiple R-Squared: 0.6773,     Adjusted R-squared: 0.636
F-statistic: 16.37 on 5 and 39 DF,  p-value: 1.081e-08
```

4.
```
V1 <- c(1,1,4,1,4,4)
X <- matrix(c(rep(c(1,0,0),3),0,1,0,rep(c(0,0,1),2)),6,3,byrow=T)
Y <- c(2,1,3,17,10,12)
fit2 <- lm(Y~X-1,weights=1/V1)
# -1 is included to omit the intercept in order to have a full rank cell means model

> fit2$fitted
        1         2         3         4         5         6
 1.666667  1.666667  1.666667 17.000000 11.000000 11.000000
```