Chapter 10 - Inferential Tools for Multiple Regression

April 9, 2018

Setting:

- There is a proposed MLR model of the mean of a response (Y) as a function of predictors and/or factors X_1, X_2, ...:

$$\mu\{Y|X_1, X_2, ...\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + ...$$

  Using the above notation implies that any factors have already been encoded as Dummy (indicator) variable(s) X_j that either take on 0 or a 1 as described in Chapter 9.

- There are $p + 1$ parameters: $\beta_0, \beta_1, \beta_2, ..., \beta_p$.

- *Least squares regression* is used (e.g., via `lm()`) to find:

  1. estimates of the $p + 1$ model parameters: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p$; and
  2. standard errors: $SE(\hat{\beta}_0), SE(\hat{\beta}_1), SE(\hat{\beta}_2), ..., SE(\hat{\beta}_p)$.

  "Least squares" means that comparing the regression fit

$$\hat{\mu}\{Y|X_1, X_2, ...\} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + ...$$

  to the data y_i results in the *smallest* possible residual sum of squares

$$\text{residual sum of squares} = \sum_i (y_i - \hat{\mu}_i)^2 = \sum_i residual_i^2.$$

  Any other values used for the coefficients in the regression model results in a larger residual sum of squares.

- If the assumptions of **linearity, normality, constant variance and independence** are met, Chapter 10 gives the theoretical basis for conducting hypothesis tests and constructing CIs for:

  – one β_j (in Chapter 9 notes we have already conducted tests for a single β_j)

  – many β_j's simultaneously

  – a linear combination of β_j's

  – the mean $\mu\{Y|X_1, X_2, ...\}$ over all individuals in a subpopulation who have the same fixed values of X_1, X_2, ...

  – a future single individual with a fixed value of X_1, X_2, ..., (i.e., a PI)

  There are no surprises. We will perform these tasks using R just as learned to do in earlier chapters.

10.2 Inferences about regression coefficients

10.2.1 Least squares estimation

Matrix equations must be used to show how the method least squares regression finds $\hat{\beta}_i$ and $SE(\hat{\beta}_i)$. An outline (exercises 20-22 on pages 302-3): Let:

$\mathbf{Y}$ be a vector of all responses (y_i);

$\mathbf{X_1}$ be a vector of all the values of X_1;

$\mathbf{X_2}$ be a vector of all the values of X_2;

etc ...

ε be a vector with zero mean and constant variance σ^2.

Then under the assumptions of **linearity, constant variance and independence**, the MLR relationship between the response with the explanatory variables is

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X_1} + \beta_2 \mathbf{X_2} + ... + \varepsilon$$

(this equation leaves out possible interactions for simplicity). One more simplification makes the notation even simpler. Let

X be the *model matrix* (or *design matrix*) that has ones in the first column, $\mathbf{X_1}$ in second column, $\mathbf{X_2}$ in third column, etc. Every linear model you fit has an associated model matrix - check it out in R via `model.matrix()`.

Then the MLR can be rewritten as

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

where β is a vector of β_i's. Taking the mean of both sides shows that

$$\mu\{\mathbf{Y}|\mathbf{X}\} = \mathbf{X}\beta.$$

The residual sum of squares that compares the data y_i to this true model is

$$\text{true residual sum of squares} = \sum_i (y_i - \mu_i\{Y|\mathbf{X}\})^2 = (\mathbf{Y} - \mu\{\mathbf{Y}|\mathbf{X}\})^T(\mathbf{Y} - \mu\{\mathbf{Y}|\mathbf{X}\}) = (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta).$$

To find the values of β_i's that minimize these residual sum of squares (=$\hat{\beta}_j$'s):

1. Use calculus to take the derivative of the residual sum of squares with respect to each of the β_i's. Organizing these derivatives into a vector (called the *gradient*) is $\mathbf{Y} - \mathbf{X}\beta$.
2. Now set the derivatives equal to 0 and try to solve for β to get $\hat{\beta}$. This is equivalent to solving a system of linear equations: $\mathbf{Y} = \mathbf{X}\hat{\beta}$. So in a very real sense, we are modeling the individual responses (in $\mathbf{Y}$) as a function of $X_1, X_2, ...$ (in $\mathbf{X}$).
3. Usually, it is not possible to solve this system exactly (because usually the model matrix $\mathbf{X}$ has no inverse). Instead, multiply both sides of the equation by $\mathbf{X}^T$ to get the *normal equation*
$$\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{Y}. \tag{1}$$
4. Now solve this equation (if the matrix $\mathbf{X}^T\mathbf{X}$ has an inverse):
$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{2}$$

Even in the mathematical field of linear algebra, the solution of this equation is called the *least squares solution*.
When fitting a regression model using `lm()`, R is doing these matrix computations to produce the least squares estimates for the regression model.

A couple examples that you know well:

- The simplest case regression model is the null model with only an intercept term: $\mathbf{Y} = \beta_\mathbf{0} + \varepsilon$ (or $\mathbf{Y} = \mu + \varepsilon$ if you'd like). In this case $\mathbf{X} = \mathbf{1}$, simply a vector of 1's. So the least squares solution is the sample mean:
$$\hat{\beta}_0 = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = (\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\mathbf{Y} = \frac{1}{n}\sum Y_i = \bar{Y}.$$

- SLR (from Chapter 7) is the next simplest case, $Y = \beta_0 + \beta_1 X_1 + \varepsilon$. We have seen that the least squares solution reduces to $\hat{\beta}_1 = \frac{s_y}{s_x} r$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ (see Chapter 7 notes). If you know matrix algebra, start with equation (2) to derive the least squares solution for SLR for EXTRA CREDIT.

There are many beautiful things to point out about the least squares solution $\hat{\beta}$.

- When the assumptions of **linearity, constant variance and independence** are satisfied, it's a beautiful thing that the least squares solution $\hat{\beta}$:
  - is the *Best Linear Unbiased Estimator* (BLUE) of β. "Best" means that $\hat{\beta}$ has the smallest variance compared to ANY other linear unbiased estimator of β. This result is the famous *Gauss Markov Theorem.*
  - has variance-covariance matrix $\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ (derived from equation (2)).
  - for **large sample sizes**, has a multivariate normal distribution with mean β and variance $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. This result is the even more famous *Central Limit Theorem.* This is why we can use z and t tests for the regression parameters β even when the data are not normal.
- When the assumptions of **normality**, linearity, constant variance and independence are satisfied, it's a beautiful thing that the least squares solution $\hat{\beta}$:
  - is the best *Minimum Variance Unbiased Estimator* (MVUE) for β. "Best" means that $\hat{\beta}$ has the smallest variance compared to ANY other unbiased estimator of β (not just linear estimators).
  - is the *Maximum Likelihood Estimator* (MLE) for β.
  - has variance-covariance matrix $\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$.
  - **for any sample size**, has a multivariate normal distribution with mean β (i.e., $\hat{\beta}$ is unbiased), and variance $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. This is why we use z and t tests for the regression parameters β when the data are normal.

10.2.1 (still) Constant Variance and the Standard error of the least squares estimator

We want to estimate $SE(\hat{\beta}_j)$ for each $\hat{\beta}_j$ and also the constant variance σ^2. The $SE(\hat{\beta}_j)$'s are straightforward to get. Because $\hat{\beta}$ is a vector that contains all of the $\hat{\beta}_j$'s, the variance-covariance matrix

$$\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

contains the variances $\text{Var}(\hat{\beta}_j) = \left(SE(\hat{\beta}_j)\right)^2$'s along the diagonal. Importantly, $SE(\hat{\beta}_j)$ is the square root of the $j+1$ entry on the diagonal of $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. Off the diagonal are the covariances between any two $\hat{\beta}_j$'s (that we will use later in these notes and that your book considers on p.293). After fitting a regression model using `lm()`, the variance-covariance matrix can be accessed using `vcov()`.

As usual, the constant variance is estimated by the sum of the squares of the residuals:

$$\hat{\sigma}^2 = \sum_i residuals_i^2 = \sum_i (y_i - \hat{\mu}\{Y_i|X_1, ...\})^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

EXAMPLE from Exercise 14 on p. 264 (HW8 #1) regarding the *Pace of Life and Heart Disease*:

```
# Check out the data
library(Sleuth3)
head(ex0914)    # Look at first 6 rows, p=3 regression variables
```

```
##   Bank Walk Talk Heart
## 1   31   28   24    24
## 2   30   23   23    29
## 3   29   24   18    31
## 4   28   28   23    26
## 5   27   22   30    26
## 6   26   25   24    20
```

```r
dim(ex0914)  # n=36,
```

```
## [1] 36  4
```

```r
Y=ex0914$Heart  # vector of responses

# Construct the model matrix X by hand
ones=as.numeric(matrix(1,36,1))  # set up a column of 1's
X=cbind(ones,ex0914$Bank,ex0914$Walk,ex0914$Talk)
colnames(X)=c("Intercept","Bank","Walk","Talk")
head(X)  # Look at first 6 rows
```

```
##      Intercept Bank Walk Talk
## [1,]         1   31   28   24
## [2,]         1   30   23   23
## [3,]         1   29   24   18
## [4,]         1   28   28   23
## [5,]         1   27   22   30
## [6,]         1   26   25   24
```

```r
# Get ls solution
XX = t(X)%*%X   # This is the matrix X'*X, '%*%' indicates matrix multiplication
beta_hat = solve(XX,t(X)%*%Y)   # solves eqn (1) to get ls solution in (2)

# Now get SEs for ls_solution
mu_hat = X%*%beta_hat           # predictions mu_hat = X*beta_hat
res = ex0914$Heart - mu_hat     # residuals y_i - mu_hati
resSS = sum(res^2)              # residual sum of squares
sigma_hat = sqrt(resSS/(36-4))  # (residual SS)/(residual DF), residual DF = n-p-1
sigma_hat                       # estimate of sqrt(constant variance)
```

```
## [1] 4.804986
```

```r
VarCov.beta_hat = sigma_hat^2*solve(XX) # solve(matrix) = inverse of the matrix
VarCov.beta_hat                         # variance-covariance matrix of ls solutions
```

```
##            Intercept         Bank         Walk        Talk
## Intercept 40.1568840 -0.751920767 -0.619420123 -0.30979145
## Bank      -0.7519208  0.038849220  0.002139729 -0.01532596
## Walk      -0.6194201  0.002139729  0.040350169 -0.01451333
## Talk      -0.3097914 -0.015325962 -0.014513332  0.04937966
```

```r
SE = sqrt(diag(VarCov.beta_hat))   # The SE^2s are on the diagonal
cbind(beta_hat,SE)   #table of estimates and SEs - just like lm()?
```

```
##                           SE
## Intercept  3.1786957 6.3369459
## Bank       0.4052170 0.1971021
## Walk       0.4516011 0.2008735
## Talk      -0.1796096 0.2222154
```

```r
# Check the above calculations with R's lm()
m1 = lm(Heart ~ Bank + Walk + Talk,data=ex0914)
summary(m1)
```

```
##
## Call:
## lm(formula = Heart ~ Bank + Walk + Talk, data = ex0914)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4014 -3.0263  0.0602  2.6748  8.4646
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.1787     6.3369   0.502   0.6194
## Bank          0.4052     0.1971   2.056   0.0480 *
## Walk          0.4516     0.2009   2.248   0.0316 *
## Talk         -0.1796     0.2222  -0.808   0.4249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.805 on 32 degrees of freedom
## Multiple R-squared:  0.2236, Adjusted R-squared:  0.1509
## F-statistic: 3.073 on 3 and 32 DF,  p-value: 0.04162
```

```r
head(model.matrix(m1))  # First 6 rows of the model matrix X
```

```
##   (Intercept) Bank Walk Talk
## 1           1   31   28   24
## 2           1   30   23   23
## 3           1   29   24   18
## 4           1   28   28   23
## 5           1   27   22   30
## 6           1   26   25   24
```

```r
vcov(m1)   # variance-covariance matrix, Var(beta_hat) = sigma_hat^2*inv(X'*X)
```

```
##             (Intercept)         Bank         Walk        Talk
## (Intercept)  40.1568840 -0.751920767 -0.619420123 -0.30979145
## Bank         -0.7519208  0.038849220  0.002139729 -0.01532596
## Walk         -0.6194201  0.002139729  0.040350169 -0.01451333
## Talk         -0.3097914 -0.015325962 -0.014513332  0.04937966
```

10.2.2 Test and CI of a single coefficient

Summary output of the linear model gives t-statistics for testing hypotheses about the intercept β_0 and about the other p coefficients $\beta_1, ..., \beta_p$.

- **Test** whether the **intercept** is zero or not:

$$H_0 : \beta_0 = 0 \text{ vs } H_a : \beta_0 \neq 0.$$

  These hypotheses can be re-phrased in terms of the mean of Y when $X_1 = X_2 = X_3 = ... = X_p = 0$:

$$H_0 : \mu\{Y|X_1 = X_2 = X_3 = ... = X_p = 0\} = 0 \text{ vs } H_a : \mu\{Y|X_1 = X_2 = X_3 = ... = X_p = 0\} \neq 0.$$

If H_0 is rejected, use non-technical language in the **conclusion** to describe that the evidence suggests that "the mean response is not 0 when $X_1 = X_2 = X_3 = ... = X_p = 0$".

- Test each of the p **coefficients**, for $j = 1, ..., p$:

$$H_0 : \beta_j = 0 \text{ vs } H_a : \beta_j \neq 0.$$

If H_0 is rejected, use non-technical language in the **conclusion** to describe that the evidence suggests that:

  - "the mean change in the response is not 0 when X_j changes by one unit with other X_i held fixed"
  - "there is an effect on the response due to X_j after accounting for the other X_i"

- If the MLR assumptions are met then the **test statistic** for tests of $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$ for any $j = 0, 1, ..., p$ is

$$t = \frac{\widehat{\beta}_j - 0}{SE\left(\widehat{\beta}_j\right)}$$

which follows a t-distribution with $n - p - 1$ degrees of freedom (i.e., *residual degrees of freedom*):

  - $\widehat{\beta}_j$ is the least squares estimate of β_j
  - $SE\left(\widehat{\beta}_j\right)$ is the standard error of $\widehat{\beta}_j$
  - $p + 1$ is the number parameters $\beta_0, \beta_1, ..., \beta_p$ in the model

- The two-sided p**-value** is calculated by comparing the test statistic to a t-distribution with $n - p - 1$ degrees of freedom.
- A two-sided $100 \times (1 - \alpha)\%$ CI for β_j is calculated by the following equation:

$$\hat{\beta}_j \pm t_{1-\alpha/2, df=n-p-1} SE(\hat{\beta}_j)$$

- Software packages like R automatically test $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0$. You need to do some "hand calculations" if you want to test other hypotheses, e.g.:

  - When testing $H_0 : \beta_j = c$ vs. $H_a : \beta_j \neq c$, the test statistic is $t = \frac{\widehat{\beta}_j - c}{SE\left(\widehat{\beta}_j\right)}$.
  - When testing H_0 vs. a one-sided H_a, use the same test statistic as for a two-sided test to calculate a one-sided p-value.

10.1 Energy costs of echolocation in bats

Chapter 10 in the book and in these notes use the case study of energy costs of echolocation in bats as an example for conducting relevant tests and CIs. The main question of interest in this study:

Is energy expenditure (W) the same for echolocating bats as for non-echolocating bats and non-echolocating birds, after accounting for body mass (g)?

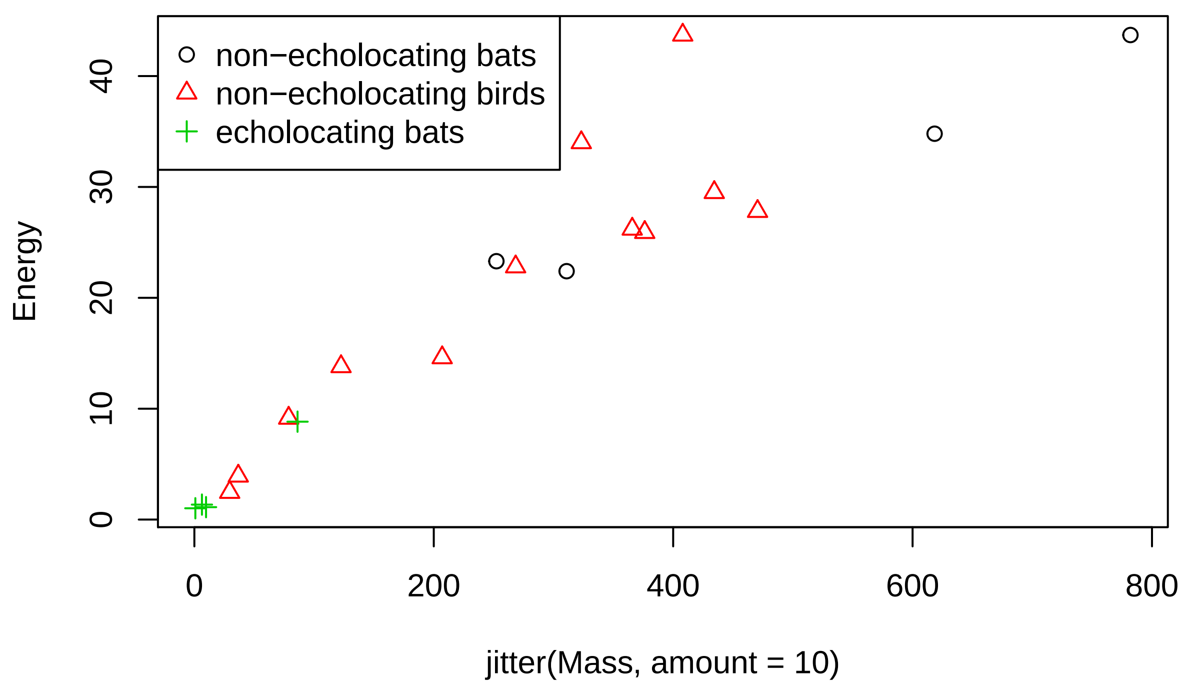
```r
# Get data - compare with Display 10.3
# Make R arrange categories as in Display 10.4 and 10.6
case1002$Type=factor(case1002$Type,levels=c("non-echolocating bats","non-echolocating birds",
                                            "echolocating bats"))

summary(case1002) # p = 3 regression variables (Mass, Dummy_birds(Type), Dummy_echo.bats(Type))
```

```
##        Mass                                Type        Energy
##  Min.    :  6.70   non-echolocating bats : 4   Min.    : 1.020
##  1st Qu.: 63.35   non-echolocating birds:12   1st Qu.: 7.605
##  Median :266.50   echolocating bats      : 4   Median :22.600
##  Mean   :262.68                               Mean   :19.518
##  3rd Qu.:391.00                               3rd Qu.:28.225
##  Max.   :779.00                               Max.   :43.700
```

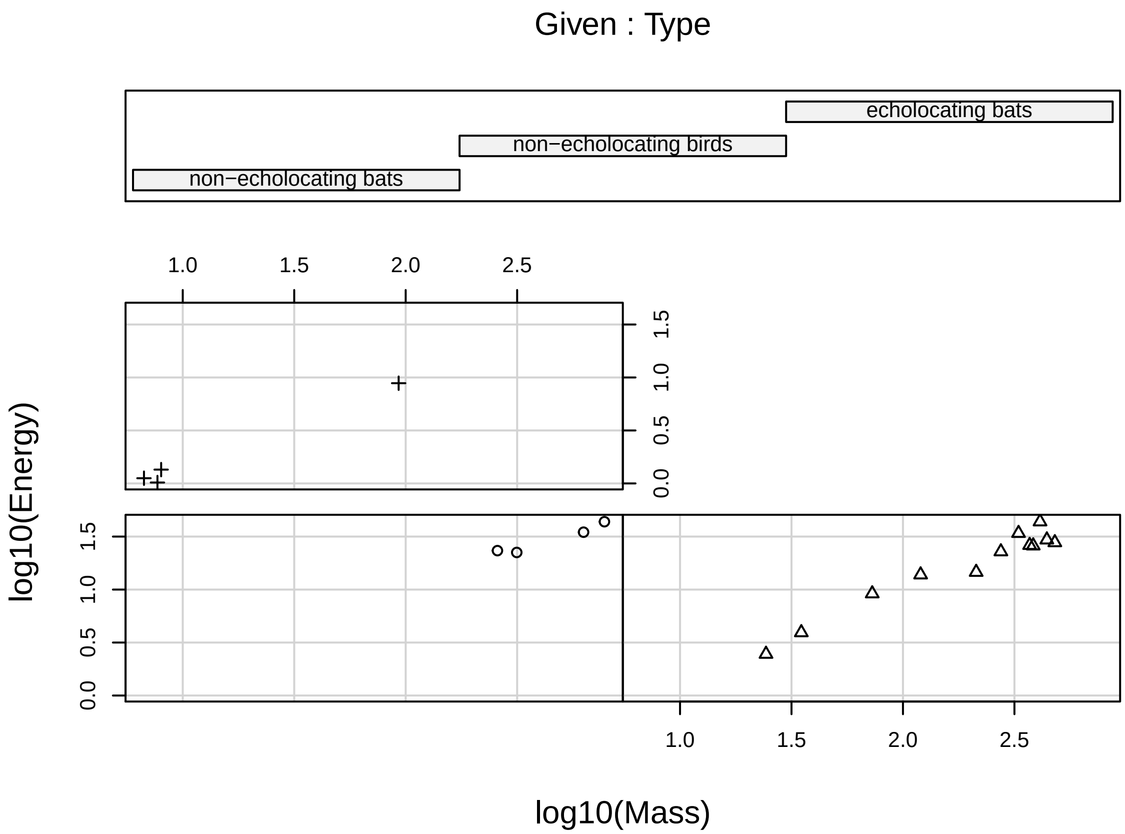
```
dim(case1002)    # n=20
```

```
## [1] 20  3
```

```
# Scatterplot - compare with Display 10.4
# Note: jitter(x,amount=a) jitters by x <- x + runif(-amount,amount) (see 9.5.3)
plot(Energy ~ jitter(Mass,amount=10),pch=as.numeric(Type),col=as.numeric(Type),data=case1002)
legend("topleft",legend=levels(case1002$Type),pch=1:3,col=1:3)
```




```
# This plot shows that you must still log-transform Mass!
#plot(log10(Energy) ~ Mass,pch=as.numeric(Type),data=case1002)

# Only 1 predictor, matrix plot not any more informative than scatterplot
# pairs(case1002)

# Trellis plot
coplot(log10(Energy) ~ log10(Mass)|Type,pch=as.numeric(case1002$Type),data=case1002)
```



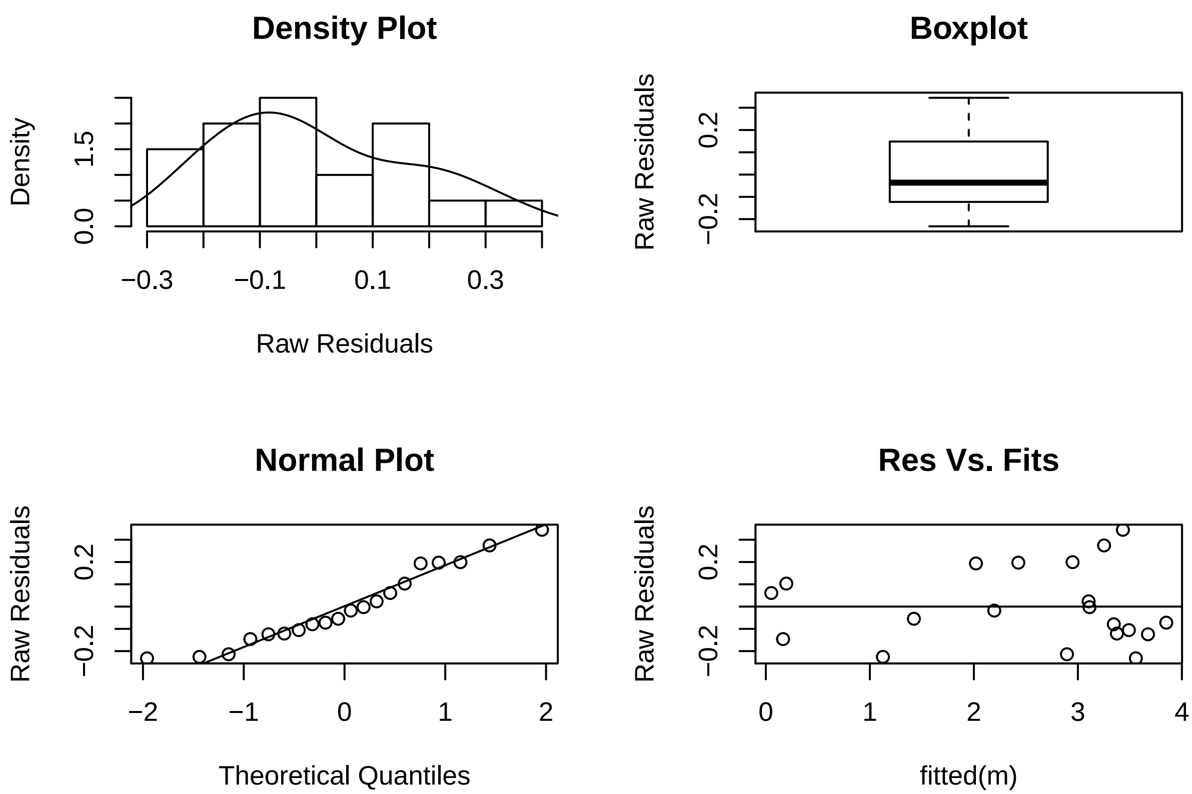
QUESTION:

1. Based on the above plots, write out the MLR model. Use Dummy variables for the factor `Type`. Use natural logs (instead of $\log_{10}$'s) to make it easier to compare to the textbook's analysis (Display 10.6).

2. Give the parameter(s) to be tested that addresses the question of interest: *Is energy expenditure the same for echolocating bats as for non-echolocating bats and non-echolocating birds, after accounting for body mass?*

Fit this model in R:

```
source("http://www.math.montana.edu/parker/courses/STAT411/diagANOVA.r")
#A naive model to start that does not fit the data well
#m.BAD = lm(Energy ~ Mass + Type,data=case1002)
#diagANOVA(m.BAD)

# The model suggested by scatterplot and Trellis plot
m2 = lm(log(Energy) ~ log(Mass) + Type,data=case1002)
diagANOVA(m2)
```




```r
# Extra sum of squares test for interaction log(Mass)*Type
m3 = lm(log(Energy) ~ log(Mass)*Type,data=case1002)
anova(m2,m3)
```

```
## Analysis of Variance Table
##
## Model 1: log(Energy) ~ log(Mass) + Type
## Model 2: log(Energy) ~ log(Mass) * Type
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     16 0.55332
## 2     14 0.50487  2   0.04845 0.6718 0.5265
```

```r
# Compare with Display 10.6
summary(m2)
```

```
##
## Call:
## lm(formula = log(Energy) ~ log(Mass) + Type, data = case1002)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23224 -0.12199 -0.03637  0.12574  0.34457
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -1.57636    0.28724  -5.488 4.96e-05 ***
## log(Mass)                     0.81496    0.04454  18.297 3.76e-12 ***
```

```
## Typenon-echolocating birds  0.10226    0.11418   0.896    0.384
## Typeecholocating bats       0.07866    0.20268   0.388    0.703
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 16 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9781
## F-statistic: 283.6 on 3 and 16 DF,  p-value: 4.464e-14
```

```r
confint(m2)
```

```
##                                 2.5 %     97.5 %
## (Intercept)                -2.1852742 -0.9674462
## log(Mass)                   0.7205339  0.9093811
## Typenon-echolocating birds -0.1397945  0.3443183
## Typeecholocating bats      -0.3509972  0.5083245
```

```r
#Two-sided p-values in the regression output above
2*pt(c(5.488,18.297,0.896,0.388), 20-3-1, lower.tail = FALSE)
```

```
## [1] 4.960624e-05 3.756376e-12 3.835355e-01 7.031296e-01
```

```r
# 95% CIs for beta3 by hand
0.0787  + c(-1,1)*qt(.975,20-3-1)*0.2027
```

```
## [1] -0.3510048  0.5084048
```

```r
exp(0.0787  + c(-1,1)*qt(.975,20-3-1)*0.2027)  # Transform to get 95% CI for ratio of medians
```

```
## [1] 0.7039804 1.6626368
```

QUESTIONS:

1. Are we justified in using the equal slopes model? Report the TWO plots and the test that supports your answer.

2. Address part of the research question of interest: *Is energy expenditure the same for echolocating bats as for non-echolocating bats, after accounting for body mass?* Report an estimate of the difference and the associated p-value.

3. Report and interpret a 95% CI pertinent to the research question in #2.

10.2.3 and 10.4.3 Tests and CIs for a linear combination of coefficients

A linear combination of the regression parameters is written as

$$\gamma = C_0\beta_0 + C_1\beta_1 + \ldots + C_p\beta_p$$

where $C_0, C_1, \ldots, C_p$ are fixed *coefficients* chosen by the researcher.

The linear combination γ is specified by a Greek letter (for "g") to emphasize that it is a parameter that we want to estimate. We estimate γ in the obvious way - replace the regression parameters with their corresponding least squares estimators. This estimator is g:

$$g = C_0\hat{\beta}_0 + C_1\hat{\beta}_1 + \ldots + C_p\hat{\beta}_p$$

QUESTIONS:

1. Consider the other part of the research question of interest: *Is energy expenditure the same for echolocating bats as for non-echolocating birds, after accounting for body mass?* Give the linear combination γ that quantifies this difference.

2. Report g, an estimate of γ.

In Chapter 6 when we considered linear combinations of group means, there was a straightforward formula for the standard error because the group means were independent. In contrast, the regression coefficients are **dependent** so the standard error for a linear combination of regression coefficients is more complicated. The dependencies are specified in the *covariances* in the off-diagonal terms of the variance-covariance matrix $\text{Var}(\hat{\beta})$ (considered earlier in these notes). The formula for $SE(g)$ that accounts for this dependence (derived on p. 293-4):

$$\begin{aligned}(SE(g))^2 &= C_0^2 SE(\hat{\beta}_0)^2 + C_1^2 SE(\hat{\beta}_1)^2 + \ldots + C_p^2 SE(\hat{\beta}_p)^2 + \\ &\quad 2C_0C_1Cov(\hat{\beta}_0,\hat{\beta}_1) + 2C_0C_2Cov(\hat{\beta}_0,\hat{\beta}_2) + \ldots + 2C_{p-1}C_pCov(\hat{\beta}_{p-1},\hat{\beta}_p).\end{aligned}$$

For example, for the case where $\gamma = \beta_3 - \beta_2$, then $g = \widehat{\beta}_3 - \widehat{\beta}_2$, and

$$(SE(g))^2 = \text{Var}\left(\widehat{\beta}_3 - \widehat{\beta}\right) = SE\left(\widehat{\beta}_3\right)^2 + SE\left(\widehat{\beta}_2\right)^2 - 2Cov\left(\widehat{\beta}_3, \widehat{\beta}_2\right).$$

A $100 \times (1-\alpha)\%$ CI for the linear combination γ is:

$$g \pm t_{1-\alpha/2, df=n-p-1} SE(g).$$

To test $H_0: \gamma = \gamma_0$ vs. $H_a: \gamma \neq, <, > \gamma_0$, the test statistic is the t-ratio:

$$t = \frac{g - \gamma_0}{SE(g)}$$

with $df = n - p - 1$.

In R, we are going to calculate $SE(g)$ FOUR different ways:

```r
# I. Calculate g by hand from coefficients
LC.vec=c(0,0,-1,1)
g = sum(coef(m2)*LC.vec)
g
```

```
## [1] -0.02359824
```

```r
# Get the variance-covariance matrix Var(beta_hat)
vcov.mat=vcov(m2)
rownames(vcov.mat)=c("Int","log(Mass)","birds","ech.bats")   # shortening names for nicer display
colnames(vcov.mat)=c("Int","log(Mass)","birds","ech.bats")   # shortening names for nicer display
vcov.mat
```

```
##                   Int    log(Mass)        birds     ech.bats
## Int        0.08250476 -0.012105044 -0.019207033 -0.050561463
## log(Mass) -0.01210504  0.001983939  0.001730953  0.006869742
## birds     -0.01920703  0.001730953  0.013037676  0.014639321
## ech.bats  -0.05056146  0.006869742  0.014639321  0.041078883
```

```r
# Calculate SE(g) by hand
se.g=sqrt(0.04107888+0.013037675-2*0.014639320)
se.g
```

```
## [1] 0.1576005
```

```r
# A two-sided test for gamma
2*(1-pt(abs(g/se.g),20-3-1))   # p-value for test beta3-beta2 = 0
```

```
## [1] 0.8828453
```

```r
# A 95% CI for gamma
g+c(-1,1)*qt(.975,20-3-1)*se.g
```

```
## [1] -0.3576964  0.3104999
```

```r
# II. An easier approach using matrix multiplication
LC.vec<-as.matrix(LC.vec)
as.numeric(sqrt(t(LC.vec)%*%vcov.mat%*%LC.vec))  # SE(g)
```

```
## [1] 0.1576005
```

```r
# III. Redefine the reference level as refit the regression model
d=case1002
d$Type = relevel(d$Type,ref="non-echolocating birds")
summary(lm(log(Energy) ~ log(Mass) + Type,data=d))
```

```
##
## Call:
## lm(formula = log(Energy) ~ log(Mass) + Type, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23224 -0.12199 -0.03637  0.12574  0.34457
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -1.47410    0.23902  -6.167 1.35e-05 ***
## log(Mass)                  0.81496    0.04454  18.297 3.76e-12 ***
## Typenon-echolocating bats -0.10226    0.11418  -0.896    0.384
```

```
## Typeecholocating bats     -0.02360    0.15760  -0.150    0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 16 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9781
## F-statistic: 283.6 on 3 and 16 DF,  p-value: 4.464e-14
```

```
# IV. Easiest way to do it:
library(gmodels)
estimable(m2,as.numeric(LC.vec),conf.int=0.95)
```

```
##                Estimate Std. Error    t value DF  Pr(>|t|)   Lower.CI
## (0 0 -1 1) -0.02359824  0.1576005 -0.1497345 16 0.8828453 -0.3576964
##              Upper.CI
## (0 0 -1 1) 0.3104999
```

QUESTION: *Is energy expenditure the same for echolocating bats as for non-echolocating birds, after accounting for body mass?* Report results of the test regarding γ and interpret the CI for γ in terms of the problem.

10.2.3 (again) Estimating the mean response (Y) at fixed values of $X_1, X_2, ...$

Centering trick for a few values of X_1, X_2, ...

Suppose you do not really care whether the mean response is different than 0 at $X_1 = X_2 = ... = X_p = 0$, i.e. you do not care about the hypotheses $H_0 : \mu\{Y|X_1 = X_2 = X_3 = ... = X_p = 0\} = 0$ vs $H_a : \mu\{Y|X_1 = X_2 = X_3 = ... = X_p = 0\} \neq 0$. Instead, you would like to test the mean of Y at some other values of X_j, e.g.,

$$H_0 : \mu\{Y|X_1 = 15, X_2 = 200\} = 0 \text{ vs } H_a : \mu\{Y|X_1 = 15, X_2 = 200\} < 0.$$

For the echollocation data, as an example consider:

$$H_0 : \mu\{\ln(\text{Energy})|\ln\text{Mass} = \ln(50), \text{Type} = \text{“echolating bats''}\} = \ln(4) \text{ vs}$$
$$H_a : \mu\{\ln(\text{Energy})|\ln\text{Mass} = \ln(50), \text{Type} = \text{“echolating bats''}\} > \ln(4).$$

```
# Test whether the mean of Y  is different than zero at log(Mass)=log(200)
m.center = lm(log(Energy) ~ I(log(Mass)-log(50)) + Type,data=case1002)
summary(m.center)
```

```
##
## Call:
## lm(formula = log(Energy) ~ I(log(Mass) - log(50)) + Type, data = case1002)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23224 -0.12199 -0.03637  0.12574  0.34457
##
## Coefficients:
```

```
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   1.61177    0.13475  11.962 2.16e-09 ***
## I(log(Mass) - log(50))        0.81496    0.04454  18.297 3.76e-12 ***
## Typenon-echolocating birds  0.10226    0.11418   0.896    0.384
## Typeecholocating bats       0.07866    0.20268   0.388    0.703
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 16 degrees of freedom
## Multiple R-squared:  0.9815, Adjusted R-squared:  0.9781
## F-statistic: 283.6 on 3 and 16 DF,  p-value: 4.464e-14
```

```r
t=(1.61177-log(4))/0.13475   # test stat
1-pt(t,20-3-1)               # p-value
```

```
## [1] 0.0568528
```

QUESTION:

1. *Is energy expenditure different than 1W for non-echolocating bats that weigh 50g?*

2. *Is energy expenditure greater than 1W for non-echolocating bats that weigh 50g?*

3. *Is energy expenditure more than 4W for non-echolocating bats that weigh 50g?*

Confidence band for mean of Y for many different values of X_1, X_2, ...

Given values of X_1, X_2, $\dots$, we will estimate the mean response using

$$\hat{\mu}\{Y|X_1, X_2, ...\} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_1 X_2 + ...$$

Using matrix notation, this can be rewritten succinctly as (with $\mathbf{X}$ and $\hat{\beta}$ defined earlier in these notes):

$$\widehat{\mu}\{Y|X_1, X_2, ...\} = \mathbf{X}\hat{\beta}$$

These notes showed earlier that the $\hat{\beta}_j$'s are not independent because there are (non-zero) covariances in the variance-covariance matrix $\text{Var}(\hat{\beta})$. There is also variance-covariance matrix for $\widehat{\mu}\{Y|X_1, X_2, ...\}$ that shows that the mean responses are not independent because there are (non-zero) covariances in the variance-covariance matrix $\text{Var}(\widehat{\mu}\{Y|X_1, X_2, ...\})$. The variance-covariance matrix is

$$\text{Var}\left(\widehat{\mu}\{Y|X_1, X_2, ...\}\right) = \text{Var}(\mathbf{X}\hat{\beta}) = \hat{\sigma}^2 \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T.$$

The square root of the diagonal of this matrix contains the the standard errors that we need to generate a band of CIs.

From the estimates $\widehat{\mu}\{Y|X_1, X_2, ...\}$ and the standard errors, an individual t-test of the true mean μ can be conducted, or a $100(1-\alpha)\%$ t-CI can be constructed using $n-p-1$ degrees of freedom

$$\widehat{\mu}\{Y|X_1, X_2, ...\} \pm t_{1-\alpha, df=n-p-1} \times SE[\widehat{\mu}\{Y|X_1, X_2, ...\}].$$

A family of *Workman-Hotelling* CIs maintain a family-wise confidence level of $100(1-\alpha)\%$ for the mean value of Y over as many values of $X_1, X_2, ...$ as you would like

$$\widehat{\mu}\{Y|X_1, X_2, ...\} \pm \sqrt{(p+1)F_{p+1,n-p-1}(1-\alpha)} \times SE[\widehat{\mu}\{Y|X_1, X_2, ...\}].$$

```
# Set up a new data.frame of predictor and factor values that we want to construct CIs for
minMass=min(case1002$Mass[case1002$Type==levels(case1002$Type)[1]])
maxMass=max(case1002$Mass[case1002$Type==levels(case1002$Type)[1]])
new <- expand.grid(Mass = seq(minMass, maxMass),Type=levels(case1002$Type)[1])

# Get the SEs and individual CIs for the mean Y at specific values of X1, X2, ...
est.mean.ses <- predict(m2, newdata=new, se.fit=TRUE,interval="confidence")
head(data.frame(new,est.mean.ses$fit))
```

```
##   Mass                 Type      fit      lwr      upr
## 1  258 non-echolocating bats 2.949066 2.745262 3.152870
## 2  259 non-echolocating bats 2.952218 2.748507 3.155930
## 3  260 non-echolocating bats 2.955359 2.751739 3.158979
## 4  261 non-echolocating bats 2.958487 2.754958 3.162017
## 5  262 non-echolocating bats 2.961604 2.758164 3.165044
## 6  263 non-echolocating bats 2.964708 2.761357 3.168060
```

```
# Workman Hotelling band
conf.BAND.WH.low <- est.mean.ses$fit[,1] - sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit
conf.BAND.WH.hi <- est.mean.ses$fit[,1] + sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit

# Three panels, one for each species
par(mfrow=c(1,3))

# scatterplot of data for non-echolocating bats
plot(Energy ~ Mass, pch=1,data=case1002[case1002$Type==levels(case1002$Type)[1],],
     main=levels(case1002$Type)[1],ylim=c(0,45),xlim=c(0,800))
lines(new$Mass, exp(est.mean.ses$fit[,1]), lty=1, lwd=2)        # add the fitted line
lines(new$Mass, exp(est.mean.ses$fit[,2]), lty=2, lwd=2, col=2) # lower ind. CL
lines(new$Mass, exp(est.mean.ses$fit[,3]), lty=2, lwd=2, col=2) # upper ind. CL
lines(new$Mass, exp(conf.BAND.WH.low), lty=4, lwd=2, col=4)     # lower WH CL
lines(new$Mass, exp(conf.BAND.WH.hi), lty=4, lwd=2, col=4)      # upper WH  CL

# birds
minMass=min(case1002$Mass[case1002$Type==levels(case1002$Type)[2]])
maxMass=max(case1002$Mass[case1002$Type==levels(case1002$Type)[2]])
new <- expand.grid(Mass = seq(minMass, maxMass),Type=levels(case1002$Type)[2])
est.mean.ses <- predict(m2, newdata=new, se.fit=TRUE,interval="confidence")
head(data.frame(new,est.mean.ses$fit))
```

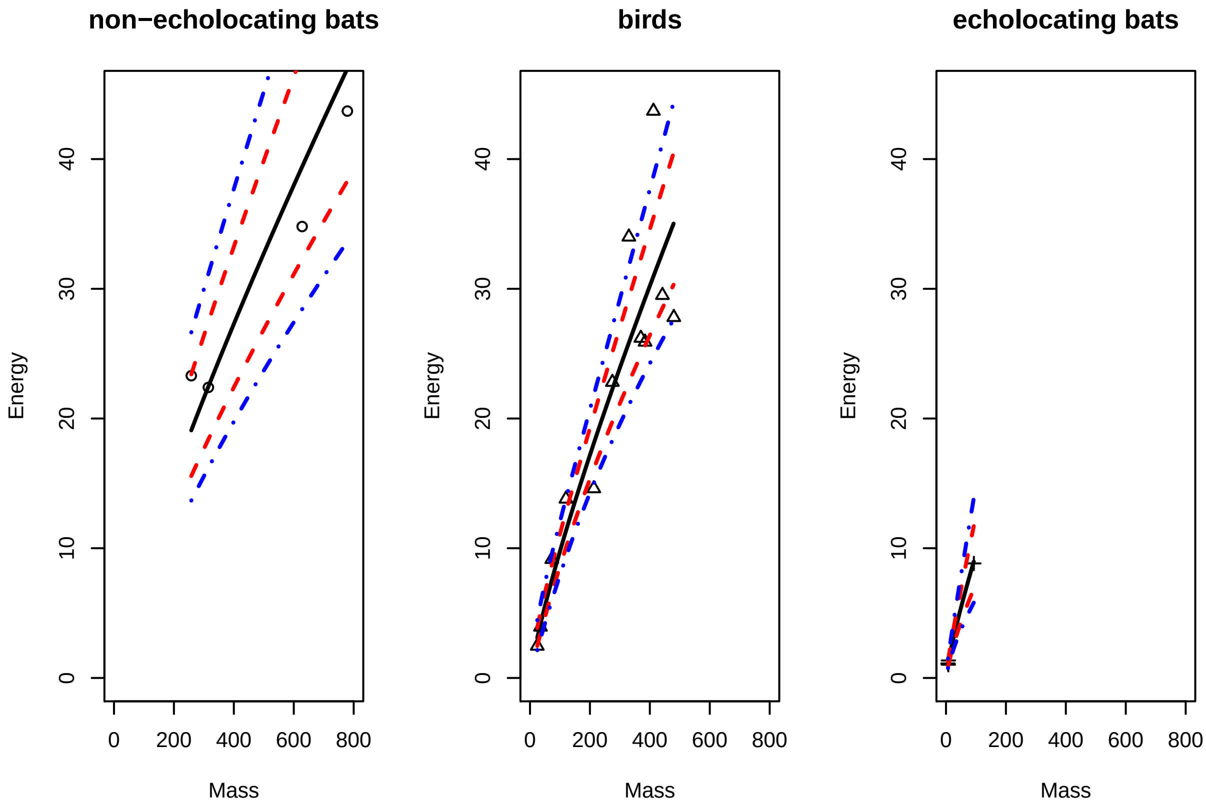
```
##   Mass                  Type      fit       lwr      upr
## 1 24.3 non-echolocating birds 1.126004 0.9023912 1.349617
## 2 25.3 non-echolocating birds 1.158870 0.9385262 1.379214
## 3 26.3 non-echolocating birds 1.190461 0.9732437 1.407679
## 4 27.3 non-echolocating birds 1.220874 1.0066495 1.435098
## 5 28.3 non-echolocating birds 1.250192 1.0388380 1.461546
## 6 29.3 non-echolocating birds 1.278492 1.0698935 1.487090
```

```
conf.BAND.WH.low <- est.mean.ses$fit[,1] - sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit
conf.BAND.WH.hi <- est.mean.ses$fit[,1] + sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit
plot(Energy ~ Mass, pch=2,data=case1002[case1002$Type==levels(case1002$Type)[2],],
```

```r
     main="birds",ylim=c(0,45),xlim=c(0,800))
lines(new$Mass, exp(est.mean.ses$fit[,1]), lty=1, lwd=2)        # add the fitted line
lines(new$Mass, exp(est.mean.ses$fit[,2]), lty=2, lwd=2, col=2) # lower ind. CL
lines(new$Mass, exp(est.mean.ses$fit[,3]), lty=2, lwd=2, col=2) # upper ind. CL
lines(new$Mass, exp(conf.BAND.WH.low), lty=4, lwd=2, col=4)     # lower WH CL
lines(new$Mass, exp(conf.BAND.WH.hi), lty=4, lwd=2, col=4)      # upper WH  CL

# echolocating bats
minMass=min(case1002$Mass[case1002$Type==levels(case1002$Type)[3]])
maxMass=max(case1002$Mass[case1002$Type==levels(case1002$Type)[3]])
new <- expand.grid(Mass = seq(minMass, maxMass),Type=levels(case1002$Type)[3])
est.mean.ses <- predict(m2, newdata=new, se.fit=TRUE,interval="confidence")
head(data.frame(new,est.mean.ses$fit))
```

```
##   Mass              Type        fit         lwr       upr
## 1  6.7 echolocating bats 0.05244027 -0.15658739 0.2614679
## 2  7.7 echolocating bats 0.16581129 -0.03921933 0.3708419
## 3  8.7 echolocating bats 0.26531980  0.06315827 0.4674813
## 4  9.7 echolocating bats 0.35398950  0.15385890 0.5541201
## 5 10.7 echolocating bats 0.43395139  0.23521416 0.6326886
## 6 11.7 echolocating bats 0.50676384  0.30892620 0.7046015
```

```r
conf.BAND.WH.low <- est.mean.ses$fit[,1] - sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit
conf.BAND.WH.hi <- est.mean.ses$fit[,1] + sqrt(4*qf(.95,4,16))*est.mean.ses$se.fit
plot(Energy ~ Mass, pch=3,data=case1002[case1002$Type==levels(case1002$Type)[3],],
     main=levels(case1002$Type)[3],ylim=c(0,45),xlim=c(0,800))
lines(new$Mass, exp(est.mean.ses$fit[,1]), lty=1, lwd=2)        # add the fitted line
lines(new$Mass, exp(est.mean.ses$fit[,2]), lty=2, lwd=2, col=2) # lower ind. CL
lines(new$Mass, exp(est.mean.ses$fit[,3]), lty=2, lwd=2, col=2) # upper ind. CL
lines(new$Mass, exp(conf.BAND.WH.low), lty=4, lwd=2, col=4)     # lower WH CL
lines(new$Mass, exp(conf.BAND.WH.hi), lty=4, lwd=2, col=4)      # upper WH  CL
```



The red bands were drawn using individual (or pointwise) CIs. At any given value of Mass any specific species, we are 95% confident that the true median Energy expenditure lies in the corresponding CI. But we are not controlling for a family of CIs. Each CI individually has a confidence level of 95% but the combined level for all of them is much less than 95%.

The blue bands were drawn using Workman-Hotelling CIs. These control the family-wise confidence level at 95% and hence we can be 95% confident that all median Energy expenditures (corresponding to both species of bats and birds between all Masses observed in the study) lie inside these CIs.

10.2.4 Prediction of a future response

You may not want to estimate the true mean from a group of individuals. Instead, you may want to estimate a future response for a single individual. Because estimating the mean is much different than estimating a future response, most statisticians say that we are *predicting* the future response as opposed to estimating it.

Before, we were keen to estimate the true mean response over a group of individuals at fixed values of $X_1, X_2, ...$ Here, we will predict the future response for a single individual at fixed values of $X_1, X_2, ...$

We will predict the response using

$$\text{Pred}\{Y|X_1, X_2, ...\} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 + ...$$

Just as we saw in Chapter 7 for SLR, the standard error is:

$$SE[\text{Pred}\{Y|X_1, X_2, ...\}] = \widehat{\sigma}\sqrt{1 + SE(\hat{\mu}\{Y|X_1, X_2, ...\})^2}.$$

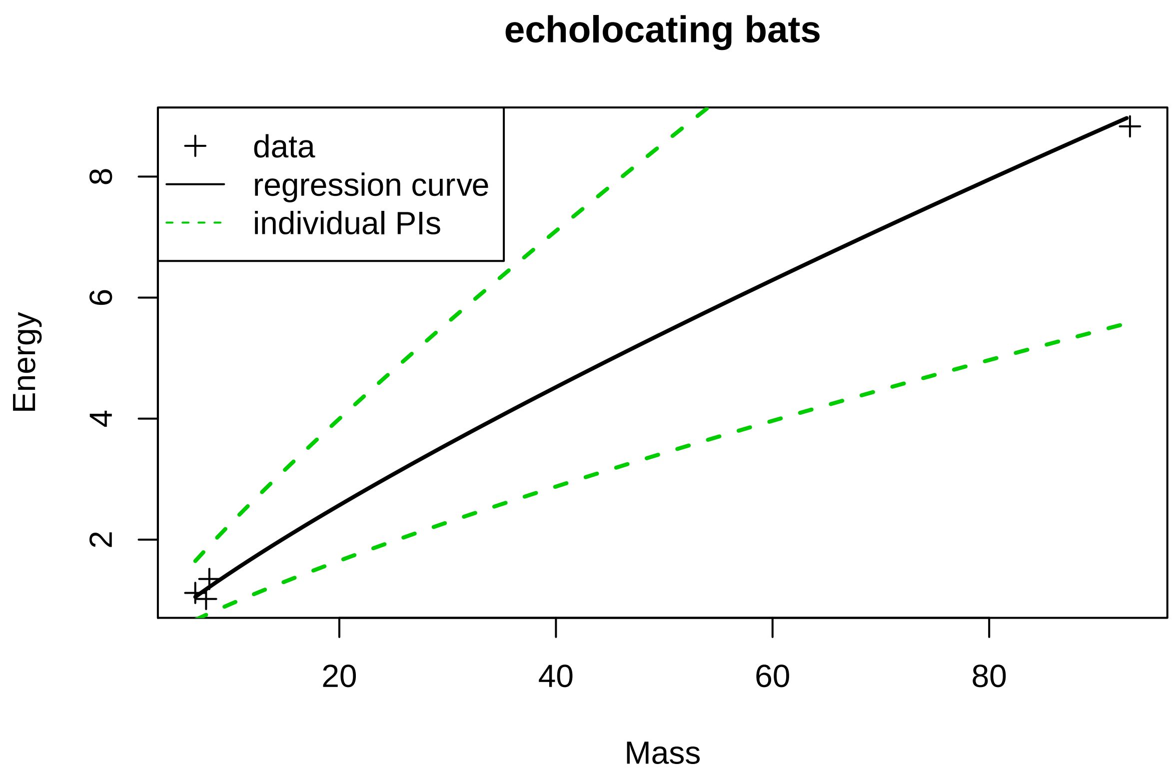
An individual $100(1-\alpha)\%$ *prediction interval* (PI) for the future value of Y at fixed values of $X_1, X_2, ...$ is

$$\text{Pred}\{Y|X_1, X_2, ...\} \pm t_{1-\alpha/2, df=n-p-1} SE[\text{Pred}\{Y|X_1, X_2, ...\}]$$

```r
# Get the individual PIs for the individual predicted responses Y at each X0
pred.ses <- predict(m2, newdata=new, se.fit=TRUE,interval="prediction")
head(data.frame(new,pred.ses$fit))
```

```
##   Mass              Type        fit         lwr       upr
## 1  6.7 echolocating bats 0.05244027 -0.39377221 0.4986527
## 2  7.7 echolocating bats 0.16581129 -0.27854281 0.6101654
## 3  8.7 echolocating bats 0.26531980 -0.17771778 0.7083574
## 4  9.7 echolocating bats 0.35398950 -0.08812504 0.7961040
## 5 10.7 echolocating bats 0.43395139 -0.00753417 0.8754369
## 6 11.7 echolocating bats 0.50676384  0.06568251 0.9478452
```

```r
plot(Energy ~ Mass, data=case1002[case1002$Type==levels(case1002$Type)[3],],main="echolocating bats",pch
lines(new$Mass, exp(pred.ses$fit[,1]), lty=1, lwd=2)        # add the fitted line
lines(new$Mass, exp(pred.ses$fit[,2]), lty=2, lwd=2, col=3) # lower ind. PL
lines(new$Mass, exp(pred.ses$fit[,3]), lty=2, lwd=2, col=3) # upper ind. PL
legend("topleft",legend=c("data","regression curve","individual PIs"),pch=c(3,-1,-1),lty=c(0,1,2),col=c
```



The green bands were drawn using individual (or pointwise) PIs. At any given Mass, we are 95% confident that the median energy expenditure for a single non-echolocating bat lies in the corresponding PI. But we are not controlling for a family of PIs. Each PI individually has a confidence level of 95% but the combined level for all of them is much less than 95%. Like your book, we will not be covering PIs for MLR with a family-wise confidence level.

Comparing these PIs to the CIs generated earlier, it is clear that it is much easier to estimate the mean response $\mu\{Y|X_1, X_2, ...\}$ compared to predicting a response $Y|X_1, X_2, ...$ for a single individual.

10.3 Extra-sums-of-squares F-tests

We have been utilizing extra sum of squares F-tests throughout the course:

- to compare the equal means (reduced) model to ANOVA (full) model (section 5.3)
- to compare the equal means (reduced) model to a regression (full) model (section 8.5)
- to compare SLR (reduced) model to ANOVA (full) model (section 8.5)
- to compare MLR without interactions (reduced) to MLR with interactions (full) model (Chapter 9 and 10 notes).

In any of these scanrios, the hypotheses being tested are

$$H_0\text{: The reduced model is the true model vs. } H_a\text{: The full model is the true model}$$

This makes clear that the extra sum of squares test (also called a **partial F-test** because only partial or extra sum of squares are tested) is a very general tool to compare a reduced (simpler) model to a full (more complicated) model that includes the reduced model plus some extra terms. Statisticians say that the simpler model is **nested** in the full model. Because of the nesting, it is common to rewrite the hypotheses in terms of the additional regression coefficients being added when using the full model.

The hypotheses are tested by fitting both models; calculating the residuals and the Residual sum of squares; then calculating the extra sum of squares:

$$\text{ESS} = \text{SS(reduced model) - SS(full model)}$$

Recall that the Residual sum of squares for a model is the variability that is not explained by the model. So ESS is the amount that the unexplained variation decreases when the extra terms are added to the reduced model. The extra sum of squares F-test determines whether ESS (i.e., the decrease in the unexplained variability of the response) is statistically significant.

The 6 steps of the extra sum of squares F-test:

1. H_0: The reduced model is the true model vs. H_a: The full model is the true model

   or equivalently

   H_0: $\beta_j = 0$ for all of the additional terms in the full model vs. H_a: $\beta_j \neq 0$ for at least one of th additional terms in the full model

2. Check the assumptions for the reduced and full models

3. Test statistic is $F_{stat} = \frac{\frac{ESS}{DF_{reduced} - DF_{full}}}{\hat{\sigma}^2_{full}}$ where $DF_{reduced}$ and DF_{full} are the residual degrees of freedom for the two models and $\hat{\sigma}^2_{full} = SS_{full}/DF_{full} = MSF$. Your book refers to $DF_{reduced} - DF_{full}$ as "the number of β_j's being tested". F_{stat} compares this drop in the Residual sums of squares (ESS) to the unexplained variability in the full model ($\hat{\sigma}^2_{full}$).

4. p-value is $P(F > F_{stat})$ where $F \sim F(DF_{reduced} - DF_{full}, DF_{full})$

5. Make a decision

6. State a conclusion - a small p-value leads you to conclude that the model under H_a is "better"

QUESTIONS:

1. Recall the (reduced) MLR model with no interaction that we fit to the energy expenditure data on birds and bats:

$$\mu\{\ln(Energy)|Mass, Type\} = \beta_0 + \beta_1 \ln(Mass) + \beta_2 \text{Dummy}_{bird} + \beta_3 \text{Dummy}_{echo-bat}$$

.

(a) Write out the (full) MLR model that includes the interaction terms.

(b) Write out the hypotheses that compare the reduced and full models in terms of the additional β_j's that were added in the full model.

(c) R code earlier in these notes implemented the extra sum of squares test

(cf. `anova(lm(log(Energy)~ log(Mass)+Type,data=d),lm(log(Energy)~ log(Mass)*Type,data=case1002)`

From the R output what do you conclude about the additional coefficients?

2. Consider the following hypotheses:

H_0: The equal means model is the true model vs. H_a: The MLR with no interactions is the true model.

(a) Write out the equal means model.

(b) Rewrite the hypotheses in terms of the additional β_j's that were added in the full model.

(c) R code earlier in these notes fit the MLR without an interaction. The output provided a test of the hypotheses in the previous question.

(cf. `anova(lm(log(Energy)~1,data=case1002),lm(log(Energy)~log(Mass)+Type,data=case1002))`

What do you conclude re: the β_j's?

10.4.2 Improving a study with replication

- Replication = taking repeated observations at the same X values.
- When there is no replication the estimate of σ is based on variability of the response about the fitted regression. This is a *model-based* estimate and its validity depends on the validity of the model. If the assumption(s) of the model are violated then the estimate of σ may be badly biased.
- From a designed experiment with replication σ can be estimated without referring to the regression model at all, the pooled estimator from Chapter 5 can be used. This does not require any assumptions about a functional relationship between a response and explantory variables. It is called a *design-based* or *pure error* estimate and will be unbiased.

10.4.6 The Principle of Occam's Razor

- Simple models are preferred over complicated models.
- Founded in common sense and successful experience.
- Often called the *Principle of Parsimony*
- Adopted as the *KISS* design principle (**K**eep **I**t **S**imple **S**tupid) by the US Navy in 1960. As explained by Wikipedia: the KISS principle states that most systems work best if they are kept simple rather than made complicated; therefore simplicity should be a key goal in design and unnecessary complexity should be avoided.