Multiple Linear Regression Case Study

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

February 5, 2008

Birds and Bats

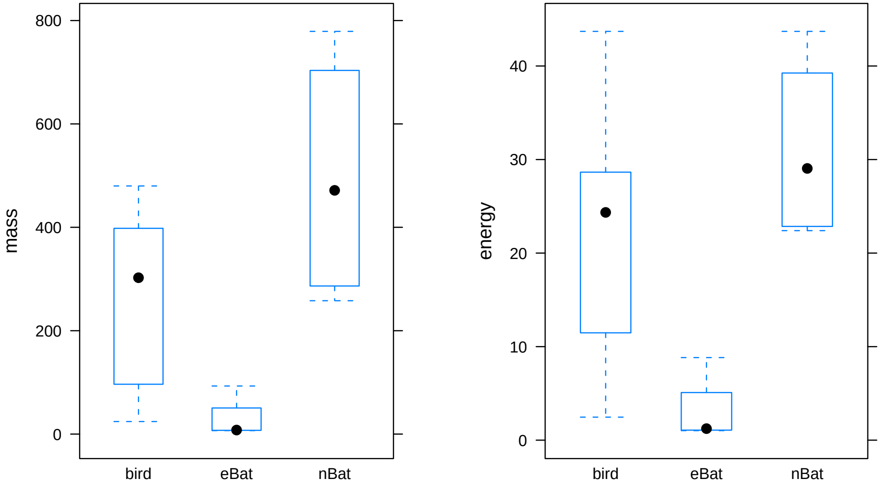
- Birds and bats must expend considerable energy to fly.
- Some bats use echolocation in flight which also requires energy.
- Other bats eat fruit and do not have the ability to echolocate.
- Scientists studied energy use of several species of birds and bats to examine the relationship between mass and energy expenditure during flight to see if echolocating bats had a higher cost.
- Variables are `mass` (grams), `type` (factor with levels `bird`, `eBat`, and `nBat`, latter two for echolocating and non-echolocating), and the response `energy` (Watts).

Data

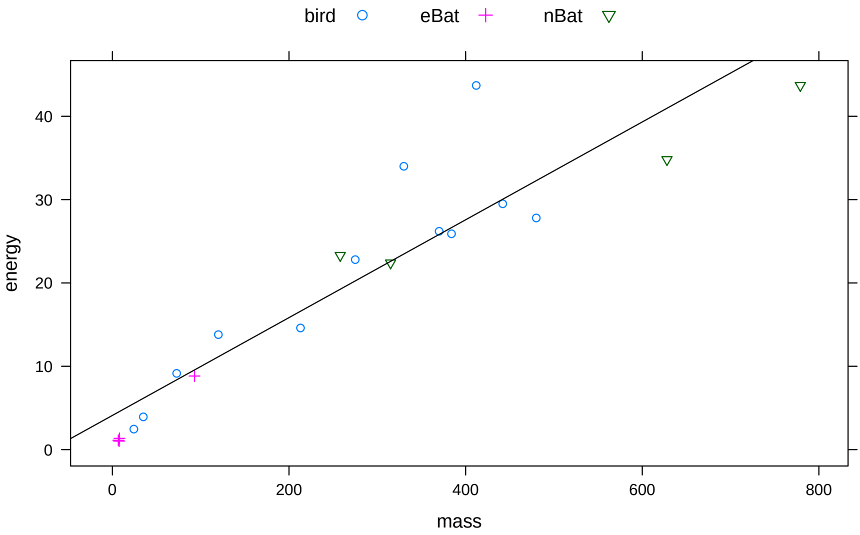
```
> bats = read.table("bats.txt", header = T)
> bats

                   species  mass type energy
1            PteropusGouldi 779.0 nBat  43.70
2      PteropusPoliocephalus 628.0 nBat  34.80
3    HypsignathusMonstrosus 258.0 nBat  23.30
4             EidolonHelvum 315.0 nBat  22.40
5        MeliphagaVirescens  24.3 bird   2.46
6    MelipsittacusUndulatus  35.0 bird   3.93
7           SturmisVulgaris  72.8 bird   9.15
8            FalcoSpaverius 120.0 bird  13.80
9          FalcoTinnunculus 213.0 bird  14.60
10         CorvusOssifragus 275.0 bird  22.80
11          LarusAtricilla 370.0 bird  26.20
12             ColumbaLivia 384.0 bird  25.90
13             ColumbaLivia 442.0 bird  29.50
14             ColumbaLivia 412.0 bird  43.70
15             ColumbaLivia 330.0 bird  34.00
16        CorvusCrytoleucos 480.0 bird  27.80
17      PhyllostomasHastatus  93.0 eBat   8.83
18          PlecotusAuritus   8.0 eBat   1.35
19 PipistrellusPipistrellus   6.7 eBat   1.12
20          PlecotusAuritus   7.7 eBat   1.02
```

- Notice that both `mass` and `energy` span different orders of magnitude.
- The two bat types are quite different in mass.
- Birds fill the gap.
- Each observation corresponds to a single study.
- Some studies are on the same species.

Box-and-Whisker Plots



Scatterplot



Observations

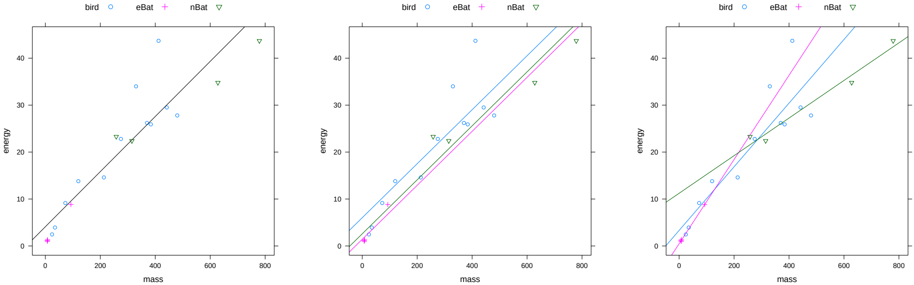
- The scatterplot reveals potential problems with fitting a standard regression model:
  - Two bird observations appear to be potential outliers;
  - There is some apparent curvature;
  - Points with high mass have more variable energy measurements than points with low mass;
- We will, however, fit a few models to illustrate the method and to show how these potential problems can be identified more readily with *residual plots*.

Fitting Models

```
> fit0 = lm(energy ~ mass,
+      data = bats)
> fit1 = lm(energy ~ mass +
+      type, data = bats)
> fit2 = lm(energy ~ mass *
+      type, data = bats)
```

- `fit0` is a simple linear regression of energy on mass
- `fit1` adds type as an input variable. This has the effect of allowing the intercept to be different for each type.
- `fit2` has `mass` and `type` and an interaction between them. This has allows each type to have its own slope and intercept.

Plots of Fitted Models



Estimated Coefficients

```
> coef(fit0)

(Intercept)        mass
 4.09991727  0.05869642
```

- `fit0` shows the intercept and parameter for `mass` which is the slope.

Estimated Coefficients

```
> coef(fit1)

(Intercept)        mass    typeeBat    typenBat
 6.02197707  0.05749542 -4.60071984 -3.43220829
```

- `fit1` shows an intercept for all predictions, a parameter for `mass` which is the common slope, and then adjustments to be made if the type is `eBat` or `nBat`.
- In effect, these are estimated differences of the intercept relative to `bird`.
- For birds, the intercept is 6.02.
- For echolocating bats the intercept is $6.02 + (-4.6) = 1.42$.
- For non-echolocating bats the intercept is $6.02 + (-3.43) = 2.59$.
- The three lines are parallel and share the common slope 0.0575.

Estimated Coefficients

```
> coef(fit2)
  (Intercept)          mass      typeeBat      typenBat
   3.31674159    0.06777464   -2.82275855    7.91064213
mass:typeeBat mass:typenBat
   0.02186199   -0.02772895
```

- `fit2` shows six estimated coefficients, the intercept and slope for `bird` and then adjustments to each of these for the other types.
- For birds, the intercept is 3.32 and the slope is 0.0678.
- For echolocating bats the intercept is $3.32 + (-2.82) = 0.494$ and the slope is $0.0678 + (0.0219) = 0.0896$
- For non-echolocating bats the intercept is $3.32 + (7.91) = 11.2$ and the slope is $0.0678 + (-0.0277) = 0.04$

Interpretation of Coefficients

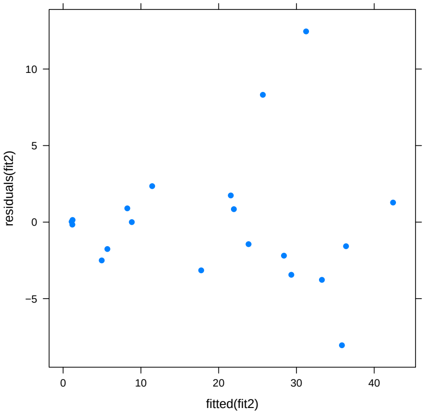
```
> coef(fit2)
  (Intercept)           mass       typeeBat       typenBat
   3.31674159     0.06777464    -2.82275855     7.91064213
mass:typeeBat mass:typenBat
   0.02186199    -0.02772895
```

- The intercept is the predicted energy of a bird at mass 0 — no biological relevance.
- The third coefficient is the estimated difference between the predicted energies for echolocating bats and birds at mass 0.
- Notice that the predicted difference is not the same at all masses.
- This parameter has no biological significance also.
- Similar comments can be made about the non-echolocating bats — in particular, even though the intercept for non-echolocating bats is higher than for birds, at the range of mass where there are both birds and non-echolocating bats, the bird line is *higher*.

Residual Plot

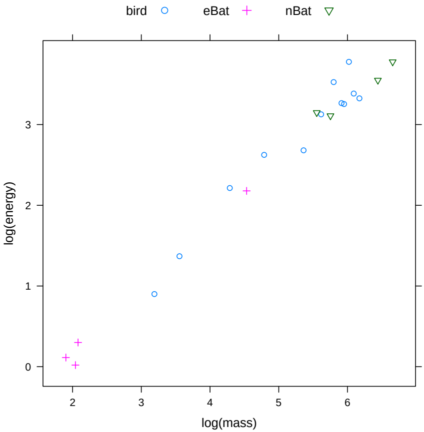
```
> plot(xyplot(residuals(fit2) ~
+      fitted(fit2), pch = 16))
```

- Residual plot from last fit.
- Notice the fan-shaped pattern.
- Residuals are larger for large mass.
- A transformation may help.



Log Transformed Data

- Log transformation of both variables leads to data that better fits linear model assumptions.



More Analysis

- Do remaining analysis live in R.