

Input data summary

It is important to understand the nature of the data you are working with: how large is the sample, how appropriate is the sample for asking the science questions you have, how well did you sample across a gradient of values for each covariate, what correlations exist among your covariates, is there any evidence of outliers or errors in the dataset, etc. To do this, you'll need to work with the data in a stats package to obtain summary statistics and to view some plots of the covariates.

In R, we can easily import a text file containing the data and view some summary statistics.

Start with a text file that contains:

```
eh      freq  area  sex   mass  length
11      1     1     0    34.0  126
11      1     1     0    33.4  128.5
11      1     1     0    29.9  130
11      1     1     0    31.0  113
11      1     1     0    26.7  116
11      1     1     0    23.9  124
...
```

Use R to input the file, drop unnecessary columns, manipulate labels for categorical variables, & summarize. The file "fawnsdata.r" does this with a version of the dataset that just has the variable names and the data, i.e., the comments, semicolons, etc. are stripped out.

```
      eh      area      sex      mass      length
die :68  cntrl:59  female:57  Min.   :22.80  Min.   :108.0
live:47  trmt :56   male :58   1st Qu.:31.90  1st Qu.:120.0
              Median :33.60  Median :124.0
              Mean   :34.38  Mean   :123.2
              3rd Qu.:36.60  3rd Qu.:127.0
              Max.   :72.00  Max.   :135.5
```

Note: there appears to be one very heavy fawn; is it an error in the data?

Next, you might want to look at simple cross-tabulations for fate versus sex or area. The idea here is to understand your data. If you are working with *a priori* models, you are **not** looking at the data to get ideas for what model structures to consider.

```
> prop.table(table(dat$eh))
      die      live
0.5913043 0.4086957

> prop.table(table(dat$eh, dat$area), margin=2)
      cntrl      trmt
die  0.6101695 0.5714286
live 0.3898305 0.4285714

> prop.table(table(dat$eh, dat$sex), margin=2)
      female      male
die  0.4912281 0.6896552
live 0.5087719 0.3103448
```

Next, you might want to take a look at plots of pairs of covariates to gain understanding of the sample in hand.

```
f=subset(fawns,select=c(status,AREA,SEX,weight,length))  
pairs(f)
```

