

https://classroom.github.com/a/GfxTqd_W

The dataset `carnivora` is a sample of 112 species from the Order Canivora. You can find the CSV of the dataset at the following link:

<https://raw.githubusercontent.com/roualdes/data/master/carnivora.csv>

You can find the help file for this dataset at the following link:

<https://github.com/roualdes/data/blob/master/carnivora.txt>

1. Read in the dataset using the function `read.csv`. Use `dplyr` to create a new dataset that consists of no-missing data (no NAs) and only the variables `Family`, `LS`, `BW`, and `SW`. Further, filter your dataset down to only the families `Canidae`, `Felidae`, and `Mustelidae`.

```
library(dplyr)
carnivora <- "https://raw.githubusercontent.com/roualdes/data/master/carnivora.csv"
  read.csv %>%
  select(Family, LS, BW, SW) %>%
  na.omit %>%
  filter(Family %in% c("Canidae", "Felidae", "Mustelidae")) %>%
  droplevels
```

2. Using `ggplot2`, make a scatter plot with body weight `SW` as the response variable and litter size `LS` as the explanatory variable and with points colored by `Family`. Title your plot. Edit the axis labels to include the proper units. Put the units in parentheses, (units). Hint:
3. Using `ggplot2`, make a scatter plot with body weight `SW` as the response variable and birth weight `BW` as the explanatory variable and with points colored by `Family`. Title your plot. Edit the axis labels to include the proper units. Put the units in parentheses, (units).

```
library(ggplot2)
carnivora %>%
  ggplot(aes(LS, SW, color=Family)) +
  geom_point() +
  labs(title="Body weight on litter size by Family",
       x="Litter size", y="Body weight (kg)")

carnivora %>%
  ggplot(aes(BW, SW, color=Family)) +
  geom_point() +
  labs(title="Body weight on birth weight by Family",
       x="Birth weight (g)", y="Body weight (kg)")
```



4. Use the likelihood method together with `optim` to predict `SW` using a multiple linear regression model with unique intercepts by `Family`, one slope across `LS`, and one slope across `BW`.

```
ll <- function(beta, y, mX) {
  yhat <- apply(mX, 1, function(row) {sum( row * beta )})
  sum( (y - yhat)^2 )
}

X <- model.matrix( ~ Family + LS + BW, data=carnivora)

(beta_hat <- optim(rnorm(ncol(X)), ll,
  method="L-BFGS-B", y=carnivora$SW, mX=X)$par)
```

```
## [1] -9.53012933  8.08336592  8.59991042 -0.05083347
## [5]  0.10475837
```

5. Write 1 complete English sentence describing the estimated intercept for Mustelidae. The expected body weight for a member of the family Mustelidae when all numerical explanatory variables are equal to 0 is -0.93 kilograms.

```
round(beta_hat[1] + beta_hat[3], 2)
```

```
## [1] -0.93
```

6. Write 1 complete English sentence describing the estimated slope for LS. For an animal from one of the Families Canidae, Felidae, or Mustelidae, the expected increase in body weight for each next offspring in a litter is -0.05 kilograms, holding all else constant.

```
round(beta_hat[4], 2)
```

```
## [1] -0.05
```

7. Write 1 complete English sentence describing the estimated slope for BW. For an animal from one of the Families Canidae, Felidae, or Mustelidae, the expected increase in body weight for each 1 gram increase in birth weight is 0.1 kilograms, holding all else constant.

```
round(beta_hat[5], 2)
```

```
## [1] 0.1
```

8. Use the bootstrap method to calculate $R = 999$ bootstrapped estimated coefficients from your model.

```
library(boot)

breg <- function(data, idx) {
  y <- data[idx, 1]
  X <- data[idx, -1]
  optim(beta_hat, ll, # warm starts
        method="L-BFGS-B", y=y, mX=X)$par
}
```

```
b <- boot(cbind(carnivora$SW, X), R=999, breg,
          ncpus=3, parallel="multicore")
```

9. Write 1 complete English sentence describing a 90% confidence interval for intercept for Felidae.

```
f_int <- b$t[,1] + b$t[,2]
(f_ci <- round(quantile(f_int, probs=c(0.05, 0.95)), 2))

##      5%   95%
## -9.37  4.83
```

We are 90% confident that the expected body weight for a member of the Family Felidae, when all numerical explanatory variables are equal to 0, is between $(-9.37, 4.83)$ kilograms.

10. Write 1 complete English sentence describing a 90% confidence interval for the estimated slope for LS.

```
ls_s <- b$t[,4]
(ls_ci <- round(quantile(ls_s, probs=c(0.05, 0.95)), 2))

##      5%   95%
## -1.28  1.37
```

For an animal from one of the Families Canidae, Felidae, or Mustelidae, the expected increase in body weight for each next offspring in a litter is between $(-1.28, 1.37)$ kilograms, holding all else constant.

11. Write 1 complete English sentence describing a 90% confidence interval for the estimated slope for BW.

```
bw_s <- b$t[,5]
(bw_ci <- round(quantile(bw_s, probs=c(0.05, 0.95)), 2))

##      5%   95%
## 0.09 0.12
```

For an animal from one of the Families Canidae, Felidae, or Mustelidae, the expected increase in body weight for each 1 gram increase in birth weight is between $(0.09, 0.12)$ kilograms, holding all else constant.

12. Write 1 complete English sentence describing a 90% confidence interval for the predicted SW for Felidae when BW and LS are equal to their median.

```
x_bar <- carnivora %>%
  group_by(Family) %>%
  summarise(ls_bar = median(LS), bw_bar = median(BW)) %>%
  slice(2) %>%
  unlist %>%
  .[2:3] %>%
  c(1, 1, 0, .) # int, F, 0, ls_bar, bw_bar

yhat <- apply(b$t, 1, function(row) {sum(x_bar * row)})
(yhat_ci <- round(quantile(yhat, probs=c(0.05, 0.95)), 2))

##      5%   95%
## 20.70 32.11
```

For an animal from the Family Felidae with birth weight and litter size equal to their mean, the expected body weight is between (20.7, 32.11) kilograms.