First_Year_Project

2023-02-07

Question one

Begin with describing and fitting a full model in which the intercepts and slopes of the extinction times versus numbers of pairs may be different in all four combinations of size and migratory status.

We start by importing the dataset directly from the csv file and saving it to the data variable:

```
head(data <- read.csv('Factors Affecting Extinction.csv', header=T))
```

```
##              Species Time Pairs Size Status
## 1        Sparrowhawk 3.03  1.00    L      R
## 2            Buzzard 5.46  2.00    L      R
## 3            Kestrel 4.10  1.21    L      R
## 4          Peregrine 1.68  1.13    L      R
## 5     Grey_partridge 8.85  5.17    L      R
## 6              Quail 1.49  1.00    L      M
```

There are four different combinations of Size and Status, LR, LM, SR and SM. If we want to find the correlation between *extinction time* as a function of *pairs*, we can make a regression line with *pairs* as the predictor value and *extinction time* as the predicted value.

```
ggplot(data, aes(x = Pairs, y = Time)) +
  geom_point() +
  facet_grid(Size ~ Status) +
  theme(legend.position = "top") +
  geom_smooth(method = "lm", formula = y ~ x) +

  stat_poly_eq(formula = y ~ x,
  aes(label = paste(after_stat(eq.label), after_stat(rr.label), sep = "~~~")),
  parse = TRUE)
```

The four panels are labelled M and R (columns) and L and S (rows). The fitted equations shown are:

- M, L: $y = 0.777 + 1.24\,x \quad R^2 = 0.26$
- R, L: $y = 4.83 + 1.77\,x \quad R^2 = 0.07$
- M, S: $y = 0.953 + 0.334\,x \quad R^2 = 0.50$
- R, S: $y = -4.4 + 2.87\,x \quad R^2 = 0.63$

Axes: x-axis labelled "Pairs", y-axis labelled "Time".