Stats with Sparrows - 14

Julia Schroeder

All the models

If you've made it to here, - you're done really really well. You will know by now more than many biologists do. You have all the skills, and all the tools, to run your own analyses. This is the last hand out, and all it provides is a few ideas that you can use to hone your skills - but do not stop at my suggestions! Use the datasets provided and explore them statistically. Think about relationships you can explore. Think about data structure, and run models accordingly. This sort of explorative data analysis is important to better understand your data, find patterns and practise your skills.

While you do this, discuss your finding, and approaches, with the group, GTA and professor.

Here are a few research questions as suggestions to get you started:

1) What is the observer repeatability of the house sparrow ornament measurement? datafile: "OrnamentAge.txt"
2) In Telostylinus angusticollis, are females larger than males? In which traits? Wylde dataset.
3) Does house sparrow body mass change with age? Use the SparrowSize dataset, and assume the first year they are observed is the year they are born. This requires some data wrangling - you need to create a variable age for this dataset.

Here is a past exam question for you to practice. Discuss solutions in the group, and ask your GTA and the professors for feedback.

Exam question:
You are interested in the effects of climate change on bird's timing of breeding. You spend the last 4 years collecting data on the date the birds lay their first egg of the first clutch for an individual female in a given year. Your study species is the climate-change sensitive golden phoenix (*Phoenix potterus fawkes*), whose eggs burst into flames and smoke when they have not hatched by April 20. So with ongoing climate change the hope is that more of them may be able to survive if they could only breed earlier.
You collected data on individual birds, recording the egg laying data in days from 1st March. This way, 14 is March 14, and 36 is April 6, and so forth. You collected this data over the course of four years, between 2006 and 2009. You want to analyse whether laying date changed over the course of the years, in particular, whether it decreased. You use two main approaches for data analysis, and you can find below the R command, and the R output. With the below R output, you will deduce the analysis strategy, and write a methods and results section as you would for your thesis
Use all your knowledge from the R course on how to communicate statistics!

```
> length(PhoenixData$LayingDate)
[1] 108
```

```
> var(PhoenixData$LayingDate)
[1] 539.0041

> summary(PhoenixData$LayingDate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.00   31.75   43.00   48.62   65.25  114.00

> table(PhoenixData$year)
2006 2007 2008 2009
  46   33   10   19

> summary(lm(LayingDate~as.factor(year), data=PhoenixData))
Call:
lm(formula = LayingDate ~ as.factor(year), data = PhoenixData)

Residuals:
    Min      1Q  Median      3Q     Max
-29.804 -16.000  -2.452  12.397  61.697

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           56.804      3.076  18.467  < 2e-16 ***
as.factor(year)2007   -4.501      4.759  -0.946  0.34643
as.factor(year)2008  -20.704      7.279  -2.844  0.00536 **
as.factor(year)2009  -27.804      5.689  -4.887 3.73e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.86 on 104 degrees of freedom
Multiple R-squared:  0.2152,  Adjusted R-squared:  0.1925
F-statistic: 9.505 on 3 and 104 DF,  p-value: 1.325e-05

> summary(lm(LayingDate~year, data=PhoenixData))
Call:
lm(formula = LayingDate ~ year, data = PhoenixData)

Residuals:
    Min      1Q  Median      3Q     Max
-31.247 -15.118  -4.021  11.753  65.205

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 19017.846   3635.037   5.232 8.52e-07 ***
year           -9.451      1.811  -5.218 9.03e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20.81 on 106 degrees of freedom
Multiple R-squared:  0.2044,  Adjusted R-squared:  0.1969
F-statistic: 27.23 on 1 and 106 DF,  p-value: 9.026e-07
```