

STAT 509 2017 Summer HW19

Instructor: Shiwen Shen

Lecture Day: June 15

1. In `gala` dataset:

(a) Find 95% confidence and prediction intervals using two linear regression models:

```
fit1 <- lm(Species ~ Endemics + Area + Elevation +  
           Nearest + Scruz + Adjacent, data=gala)  
fit2 <- lm(Species ~ Endemics, data=gala)
```

with `Endemics=26`, `Area=262`, `Elevation=368`, `Nearest=10`, `Scruz=57`, and `Adjacent=261`.

(b) Compare confidence and prediction intervals in two models, which one is better? Why?

(c) Compare the r^2 and R_{adj}^2 in two models. Make comments.

(d) Check the error term normality assumption in `fit1` using QQ plot.

(e) Check the equal variance assumption in `fit1` using residual plot.

(f) Find the best Box Cox transformation function for the `Species` variable in `fit1`.

(g) Re-fit the model with transformed `Species`, draw the residual plot. Does the Box Cox transformation make the situation better?