

STAT 509 2017 Summer HW15

Instructor: Shiwen Shen

Lecture Day: June 7

1. For the `teengamb` dataset, use R to calculate the 95% two-sided confidence interval and prediction interval for gamble when income is 2, and make a detailed interpretation about these two intervals by the context of the problem. Show me your R code and R output.
2. There is a `gala` dataset in `faraway` package. It concerns the number of species of tortoise on the various Galapagos Islands. There are 30 cases (Islands) and 7 variables in the dataset, including
 - **Species** The number of species of tortoise found on the island
 - **Endemics** The number of endemic species
 - **Elevation** The highest elevation of the island (m)
 - **Nearest** The distance from the nearest island (km)
 - **Scruz** The distance from Santa Cruz island (km)
 - **Adjacent** The area of the adjacent island (km²)

Fit a simple linear regression model with **Species** as response and **Elevation** as explanatory variable. Show me the output.

- (a) Calculate \hat{Y} (a vector) and \bar{Y} (a number).
- (b) Calculate SSTO and SSE.
- (c) Draw the scatter plot (with the regression line) and residual plot. Do you think the equal variance assumption holds?
- (d) Use qq plot to check whether the normality assumption holds.
- (e) Re-fit the model with the transformation $\log Y$, and draw the scatter plot, residual plot, and qq plot. Make comments to each plot. Does the transformation make your model better?
- (f) Re-fit the model with the transformation \sqrt{Y} , and draw the scatter plot, residual plot, and qq plot. Make comments to each plot. Does the transformation make your model better?
- (g) Compare the coefficient of determination in the original regression model and the model with \sqrt{Y} transformation. Make comments.

Note: if you have problem loading faraway package, download the gala dataset from the course webpage and save it in D drive. Run the following code to load.

```
gala <- read.table("D:/galadata.txt", sep="\t")
```