

HW1

Intermediate Statistics

Frank Edwards

1/25/2019

Question 1

For this problem, we will use data built into R. Load the data using the console command `data(anscombe)`.

Describe the relationships between y_1 and x_1 , y_2 and x_2 , y_3 and x_3 , and y_4 and x_4 in sequence. Construct linear regression models and any descriptive statistics or visuals that you find helpful. Compare results across the y - x pairs. Do you notice anything interesting across your results? If so, describe the implications of your findings.

Question 2

Repeat the analysis from Question 1, but this time, use a loop in your code to run a regression model on each y - x pair. Hint: you can either use column position or create two vectors of column names as your loop index for regression variables.

Question 3

Use the data in `revenue_dat.csv` in the course repository.

1. Load in the data
2. Use the `mutate()` function in the `dplyr` package to construct a new measure for property tax revenue per capita called `prop_tax_pc` using the variables `rev_prop_tax` and `tot_pop`.
3. Transform `prop_tax_pc` into a binary variable equal to `TRUE` if property tax revenues per capita are greater than or equal to the median value of the variable, `FALSE` otherwise
4. Run a linear regression model of the proportion of the population in deep poverty (`pop_pct_deep_pov`) as a function of property tax revenues per capita (your binary measure).
5. Add an interaction term. Interact your property tax indicator with the racial composition of the county (`pct_blk = pop_blk / pop_tot`)
6. Display your results

Question 4

Describe and interpret the results from the interaction model. Explain the meaning of the coefficients. Use graphs to help explain the findings.