Statistics and Data Analysis Regression Analysis (1)

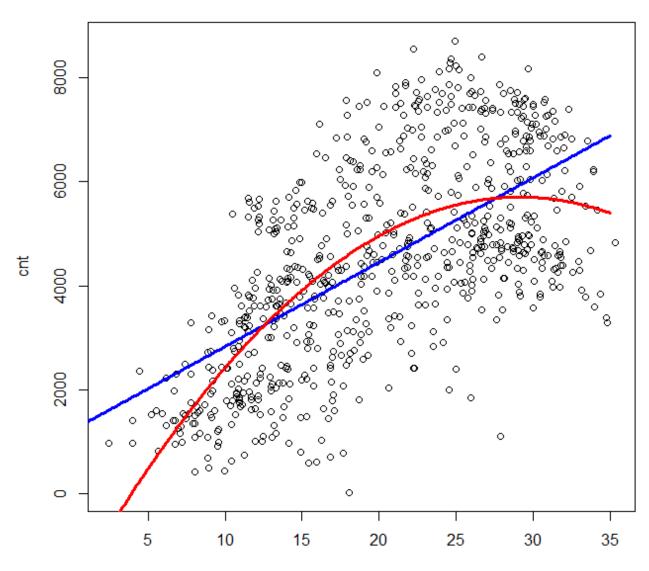
Instructor: Ling-Chieh Kung Department of Information Management National Taiwan University

- 1. The "Bike_Month" sheet contains the numbers of monthly public bike rentals in a city and other related information. The data were collected for two years.
 - (a) Draw a line chart to depict the trend and fluctuation for monthly rentals. Qualitatively describe the trend and seasonal effect.
 - (b) Construct a simple linear regression model for *instant* and *cnt*. What is your regression line?
 - (c) Check R^2 and the *p*-value and make some interpretations.
 - (d) Predict monthly rentals for the next year with the model.

- 2. Consider the daily rental data contained in the "Bike_day" sheet.
 - (a) Construct a regression model for *instant* and *cnt*. Do you still see an increasing trend?
 - (b) Recall that we have a regression line for *instant* and *cnt* for monthly rentals. Is the line for daily rentals flatter or steeper than that for monthly rentals? Why?
 - (c) Add the column *holiday* into the regression model in (a). In average what is the impact of being a holiday?
 - (d) Remove holiday and add the column workingday into the regression model in (a). In average what is the impact of being a working day? Compare the result with holiday.

- 3. Consider the daily rental data contained in the "Bike_day" sheet.
 - (a) How do *temp*, *atemp*, *hum*, and *windspeed* affect *cnt*?
 - (b) If you used a regression model with the five variables listed in (a), what are the potential drawbacks?
 - (c) Try to take away temp and do the analysis again.
 - (d) Try to add *instant* and do the analysis again.

- 4. Consider the daily rental data contained in the "Bike_day" sheet.
 - (a) Construct a regression model with *temp* as the only independent variable and *cnt* as the dependent variable. Interpret and validate your model.
 - (b) Some people suggest that *temp* should have a nonlinear impact on *cnt*. Does this fit your intuition? Draw a scatter plot to help you judge the intuition.
 - (c) To capture the nonlinear relationship, let's add a variable $temp^2$ as our second independent variable. Construct the regression model, interpret it, and validate it.
 - (d) Does adding $temp^2$ improves the model?
 - (e) Let's visualize the two regression models.



temp