RECITATION 4

Question 1 up to 5

- Any Question?

Question 6

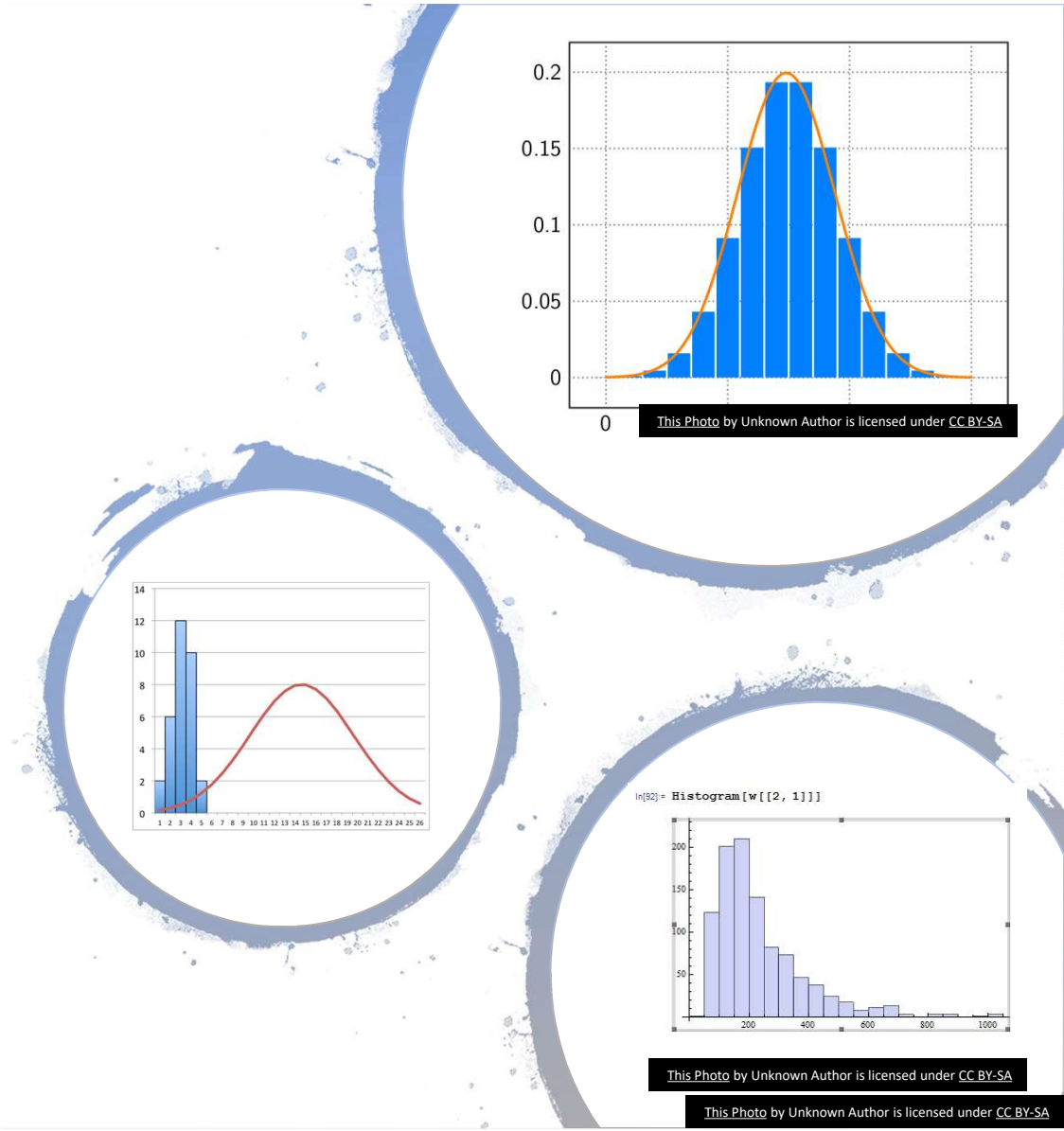
- Follow the steps in the assignment

Question 6

- Follow the steps in the assignment
- **Load data**

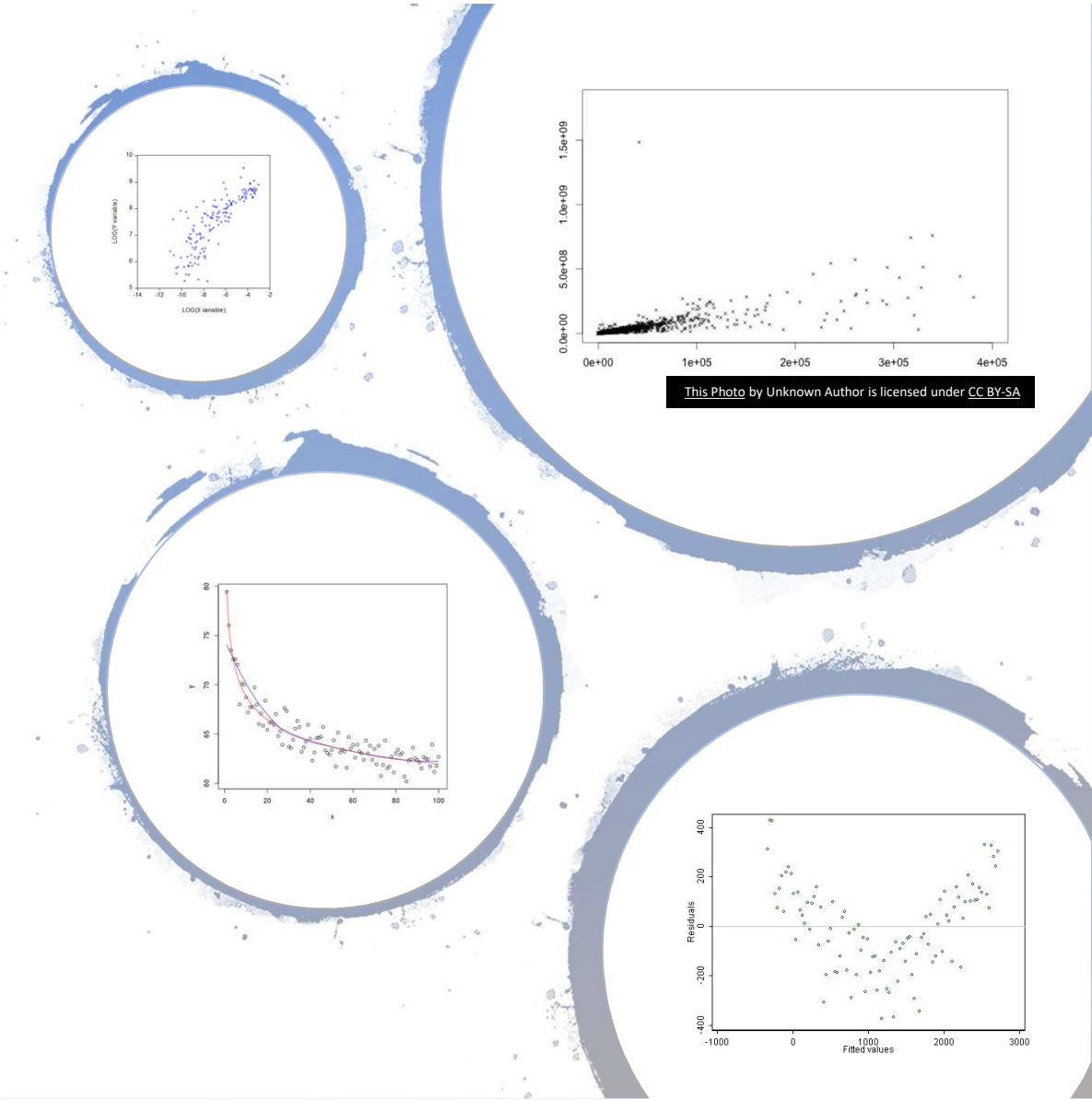
Question 6

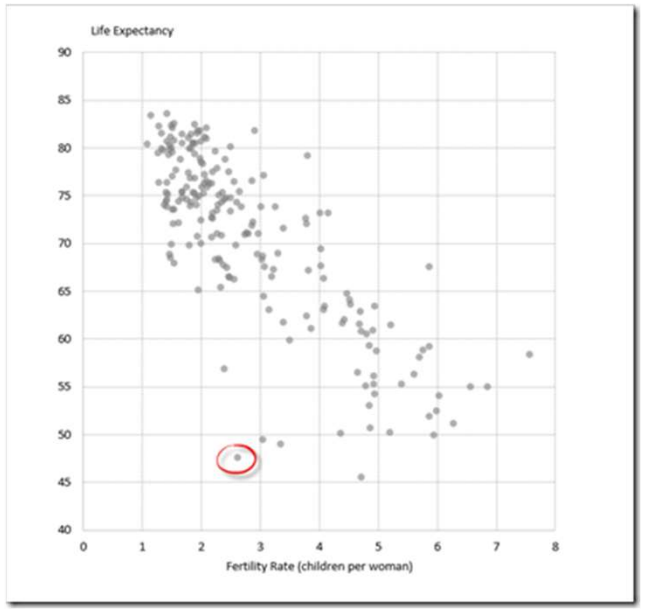
- Follow the steps in the assignment
- Load data
- **Plot the histogram of the explanatory variables and the dependent variable**
- **Are the variables normally distributed?**



Question 6 continued

- **scatter plots of the mpi (dependent variable) vs each of the features**
- **What is the relationship between each feature and dependent variables?**
- **Is it a linear relationship?**





Question 6 continued

- scatter plots of the mpi (dependent variable) vs each of the features
- What is the relationship between each feature and dependent variables?
- **Are they significant outliers?**

Question 6 continued

- **Calculate the correlation between:**

  i. X vs y

  ii. log X vs y

  iii. X vs log y

  iv. log X vs log y v

Question 6 continued

- Calculate the correlation between:
  - i. X vs y
  - ii. log X vs y
  - iii. X vs log y
  - iv. log X vs log y
- **What gives you the best correlation?**

Question 7

- **Create new feactures**

Question 7

- **Create new feactures:**

  nightlight_per_capita: nightlight_sum / landscan_pop

  population_density: landscan_pop / area

Question 7

- Create new feactures:

  nightlight_per_capita: nightlight_sum / landscan_pop

  population_density: landscan_pop / area

- **Plot histograms of each of the features and the dependent variable**

  i. Are the features normally distributed?

Question 7

- Create new feactures:

  nightlight_per_capita: nightlight_sum / landscan_pop

  population_density: landscan_pop / area

- Plot histograms of each of the features and the dependent variable i. Are the features normally distributed?
- **Calculate the following correlations for each feature (Xi) with the MPI (y)**
- **Which are the strongest correlations for each feature?**

Question 8

- **Variables from previous question – Strongest Correlation**
- **Create them**

Question 8

- Variables from previous question – Strongest Correlation
- Create them
- **Use backward-stepwise**
- **Which variables were selected? What was there p-values? What is the overall p-value of the model? Is it significant?**
- **Matlab, you just use a function, python → DA**

- **Ridge regression**
- **Penalty → eg: L2**
- **To obtain the p-value, this may involve steps**
- **Answer the question in the assignment**

- **Lasso regression**
- **Penalty → eg: L1**
- **To obtain the p-value, this may involve steps**
- **Answer the question in the assignment**

Question 9

- **calculate the estimated MPI (log yhat) for each sector using LASSO**

Question 9

- calculate the estimated MPI (log yhat) for each sector using LASSO
- **Calculate the correlation of log yhat to log y.**
- **What is the nature of the correlation?**
- **What does it tell you?**

Question 9

- calculate the estimated MPI (log yhat) for each sector using LASSO
- Calculate the correlation of log yhat to log y.
- What is the natural of the correlation?
- What does it tell you?
- **Calculate R-squared of this result. What does the R-squared tell us about the model?**

Question 10

- Just follow the steps specified
- Remember to differentiate log(y_hat) and y_hat
- **Questions?**