

5. (Data clean-up) Download the CPI dataset I have uploaded on the course website. This dataset gives you the consumer-product-index as a function of year and month. Answer these questions about this dataset.

Hint: Look up the following commands and see if they can help you in this problem: `lm`, `abline`, `summary`, `is.na`, `which`,

- (a) Read the CPI dataset in *R*. You may use `'read.csv'` for this purpose. What is the type of the data that you get?
- (b) Since the value of CPI was not available for some years, you see some `'na'` instead of the actual CPI values. Construct two vectors. One for CPIs excluding all the cells that have `'na'` and one for the corresponding years. Call them *year.month.vec* and *cpi.vec.ym*. For this purpose you may use `'is.na'` command in *R*. What is the size of these two vectors.
- (c) Use the `plot` command to plot *cpi.vec.ym* in terms of *year.month.vec*. Why do you see multiple points for each year? Construct two more vectors `'year.vec'` and `'cpi.vec'` in the following way. For each year *i* you take the average of CPI in all the months of that year to get the corresponding CPI. Then put the average CPI value in `"cpi.vec"` and the corresponding year in `"year. vec"`. Again plot *cpi.vec* as a function of *year.vec*.
- (d) Suppose that we want to model the inflation rate and see how much inflation we should expect per year (this is the minimum amount of raise you should expect in your salaries in the future. Do you see why?). What type of model do you suggest for this problem? Explain your model very clearly and then find the parameters of your model? Plot the predictions of your model in the same figure that you have your data points. We will explain how you can quantitatively evaluate the accuracy of your fit later, but for the moment based on the graph you have, do you think the fit is good? Which years seem to deviate from your predictions more?
- (e) Suppose that the salary of a person is 100K per year. Given your model, what is the minimum amount of raise he/she should expect after 1 year of working for the company? How about after 10 years.
- (f) Great depression started in US around 1929 and lasted for 10 years. I encourage you to read a little bit about the great depression. Can you see any sign of the great depression in the CPI graph? Also, great recession happened in 2009 in US. Do you see any sign of that in CPI?
- (g) Now suppose that I claim, because of great depression we should not use the data points of the years 1929 to 1940 for model fitting. They are in some sense outliers and do not represent the usual behavior of the inflation. Then use the same model you used in part (d), but this time only fit it to the CPI data that are collected after 1940. How much difference do you see in your predictions? Again suppose that the salary of a person is 100K. Based on your new predictions, what is the minimum raise he/she should expect per year?