

DSC 241 – Homework 2

Problem 1.

- Perform some simulations to check that the least squares coefficients are indeed normally distributed under the standard assumptions. To simplify things, assume a situation with only one predictor, x , say uniformly distributed in $[-1, 1]$, and that given x the response, y , is generated as $y \sim \mathcal{N}(1+2x, \sigma^2)$, with $\sigma = 0.5$. Generate a sample of size $n \in \{50, 100, 200, 500\}$ from it, and then use that sample to fit the model by least squares; then repeat this procedure $N = 1000$ times. (There are 4 settings, one for each n .) Each time record the intercept and slope. Then produce a plot or two providing (visual) evidence that the least squares intercept and slope are marginally normal. Produce another plot or two providing evidence that they are in fact jointly normal.
- Repeat, now with errors that have the t-distribution with $k \in \{2, 5, 10, 20, 50\}$ degrees of freedom. Do that for $n \in \{50, 100, 200, 500\}$. (There are $5 \times 4 = 20$ settings now.) Offer some brief comments.

[To visualize the distribution of the least squares coefficient vector, you can compute a 2D histogram and plot it in 3D; or you can compute a kernel density estimate and either plot it in 3D (as a function of 2 variables) or plot its level lines.]

Problem 2. This problem is meant as a practice for performing diagnostics. Consider the **Boston** dataset in the package **MASS**. Fit a linear model with `medv` as response, omitting all the discrete predictor variables.

- Check the standard assumptions one at a time, except for independence. Offer some brief comments on what you observe.
- Check for outliers in predictor. Comment on the most significant one if there is any. (What makes that observation unusual?)
- Same with outliers in response.
- Same with influential observations.
- Finally, check for multicollinearity.

Each time, apply all the methods that were introduced in lecture. (This is for you to get familiar with these various tools. In practice, people often have a workflow (routine) they stick to, and only use their preferred methods.)