# Statistical Methods for Data Science

# DATA7202

# Semester 1, 2021

# Lab 2

**Objectives**

On completion of this laboratory session you should be able to understand regression model and Bayesian inference.

1. A common task in machine design is the shaping of steel components for a given role. This requires blocks of steel to be cut into complex shapes using steel cutting tools in a turning operation, in which the cutting tool obviously may not last long. It is of interest to model how long the cutting tool lasts for different speeds of operation. Consequently, the tool life in a turning operation for metal-cutting was observed for 20 speeds (the data is in a csv file called "toollife.csv" on the home page). Speed is in 100 sfpm and life is in minutes.

   (a) Plot TooLife (y-axis) versus Speed (x-axis). Comment on the pattern that you see. Next, plot log(TooLife) versus log(Speed) and comment on the pattern you now see, including any unusual values.

   (b) Compute the linear regression of log(TooLife) on log(Speed). Find the parameter estimates, overall F-value and analysis of variance table. Add the fitted regression line to the scatterplot of the previous question. Discuss these results.

   (c) Plot residuals versus fitted values. Display a normal probability plot of the residuals. Comment on what these indicate about the adequacy of the linear regression assumptions.

   (d) Use the influence.measures() command to obtain leverages for this regression. Discuss which observations may be a concern from their leverage values.

2. Conjugate Normal analysis (with known variance): Given the parameter $\sigma^2$, consider $n$ iid Normal random variables $Y_i$, $(i = 1, \ldots, n)$, each with p.d.f.

$$\mathbb{P}(y \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}.$$

   Suppose that the prior for $\mu$ is also a normal distribution with parameters $\mu_0$ and $\sigma_0^2$. Our objective is to derive the posterior distribution of $\mu$.

3. Conjugate Poisson random variable analysis: Consider $n$ iid Poisson random variables $Y_i$, $(i = 1, \ldots, n)$, each with p.d.f.

$$\mathbb{P}(y \mid \theta) = \mathrm{e}^{-\theta} \frac{\theta^y}{y!}.$$

Suppose that the prior for $\theta$ is the Gamma distribution $\mathsf{Gamma}(\alpha_0, \beta_0)$. Namely

$$p(\theta \mid \alpha_0, \beta_0) = \frac{\beta_0^{\alpha}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1} \mathrm{e}^{-\beta_0 \theta}.$$

Derive the posterior distribution of $\theta$.

4. A ticket inspector has the option of taking three different routes for inspection of parking violations. Each route is characterized by the time it takes to complete the route and the intensity of ticket violations. Suppose the time t spent on route k is exponentially distributed with mean $k/2$ (hours), $k = 1, 2, 3$. For example, route 2 takes on average 1 h to complete. Suppose further that the number of traffic violations encountered, $x$, say, has a Poisson distribution with mean $10kt$. So if route 3 takes 2h, an average of 60 tickets will be issued. Suppose that on a particular day the ticket inspector has issued 60 tickets. Which route has she/he taken?