# MTH225 Fall2016 Final Problem 11

Data from a bacteria culture gives the approximate cell count (in thousands) by day. It is assumed that the growth is exponential, that is, the population $P(t)$ after $t$ days is given by:

$$P(t) = P(0) \cdot \exp(a + bt)$$

The data in `MTH225_Fall2016_Final_Problem11.csv` contains daily measurements of $P(t)$. The objective of this exercise is to fit a simple regression to estimate $a$ and $b$ and use them to compute a confidence interval for the count on day 40.

The variable names are:

- `date` number of days since the start of the culture
- `count` approximate cell count on (in thousands) on that day

Data with an exponential growth pattern can be linearized using a log transform:

$$z = \ln(y)$$

Then we fit a simple regression model for $z$:

$$z = a + bx + e$$

Once we have the estimates for $a$ and $b$, the fitted values for $y$ are:

$$\hat{y} = e^{a+bx}$$

- 2 points: Write R code to read the data and convert it to an R data frame, and perform the log transform on $y$.
- 1 point: Write the data block of a STAN model file that extracts the data from the R workspace.
- 1 point: Write the parameters block of a STAN model file that declares the paramter(s) of your model.
- 2 points: Write the model block of a STAN model file that specifies the priors and likelihood for your model.
- 1 point: Write R code to apply the `extract` function to the data structure output from the `stan` function.

- 1 point: Use the `extract()` function of the RSTAN package to obtain the values for the parameters from the posterior draw.
- 1 point: Use the posterior draw to construct to construct a simulated $\hat{y}$ value at day 40 as $\exp(a + 40b + sigma)$ for each $(a, b, sigma)$ triple in the posterior draw.
- 1 point: Use the `quantile` function on the simulated $\hat{y}$ values to compute a 95% confidence interval for $\hat{y}$.

(10 points possible)