# Analytics Homework 1

Arif Ali

09/16/15

**Note: R code is located after the written solutions**

## Problem 1

- a = 9
- b = 28.4605
- c = 17
- d = 23409

## Problem 2

As p increases, myattempts system times tend to go down whereas rgeom remained about the same regardless of the p input. This is probably because rgeom is a built in function within R.

## Problem 3

Please see R code on the last page.

## Problem 4

Please see R code on the last page

## Problem 5

The main argument against SAS is the cost of a license. While many larger companies can afford to pay for the license, individuals in smaller organizations are at a disadvantage. Since R is open sourced and free, many individuals across different disciplines and organizations can collaborate to make it better whereas the updates to SAS are controlled mainly by one company.

However, because of this one organization, the syntax behind SAS is more uniformed when compared to R, especially with respect to packages. In addition, in R there are multiple ways to do the same thing, which can be confusing when collarborating. This actually can impede learning R because the lack of central control of syntax means basic structures are just good practice instead of a requirement.

In terms of technique, R is more robust. While analysis done in SAS can be done in R, the converse isn't true. However, R is still not as favored among big companies as SAS is. So while R is both free and more robust, SAS has more industry backing. The one noticable conclusion is that most of these blog posts actually tend to conclude both are worth knowing; the comment sections tend to be where the debate of which is better is.

**Sources:**

- http://thomaswdinsmore.com/2014/12/15/sas-versus-r-part-two/
- http://www.r-bloggers.com/sas-vs-r-the-right-answer-to-the-wrong-question/

# Problem 6

## Part A

$$P(in\ Sample) = \frac{1000}{100 * 10^6} = 1 * 10^{-5}$$

## Part B

$$P(not\ In\ Any\ Of\ 2000\ Samples) = \left(\frac{100 * 10^6 - 1000}{100 * 10^6}\right)^{2000} \approx 0.98$$

## Part C

$$P(at\ Least\ One\ Sample) = 1 - P(not\ In\ Any\ Sample) = 1 - \left(\frac{100 * 10^6 - 1000}{100 * 10^6}\right)^{n} = 0.50 \Longrightarrow$$

$$\left(\frac{100 * 10^6 - 1000}{100 * 10^6}\right)^{n} = 0.50 \Longrightarrow n = 69315$$

# Problem 7

The phat estimate for p =0.3 has a positive bias of 0.05359303

# Problem 8

$$E(\hat{X}_6) = 0.01560057$$

# Bonus

$$E(X_i) = \left(\frac{1}{2}\right)^{i} \therefore E(X_6) = \left(\frac{1}{2}\right)^{6} and\ E(X_4) = \left(\frac{1}{2}\right)^{4}$$