Martin Losert and Katharina Naumann
Research Methods and Mathematical Psychology
Department of Psychology
University of Tuebingen

Problems in R

R Language and Programming

a) Write a for loop that creates the vector $\mathbf{x} = (3\ 5\ 7\ 9\ 11\ 13\ 15\ 17)'$. Hint: Use the formula $n \cdot 2 + 1$. Employ two different approaches:

   (a) Pre-allocate enough memory: Start with a zero vector of appropriate length and iteratively replace its elements. Hint: `?numeric`

   (b) Grow an object: Start with a `NULL` object and iteratively append new results.

   Verify with a larger vector and the `system.time()` function that the first approach is more efficient.

b) Implement the following function in R:

$$f(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

c) Write a function that cumulatively adds the values of the vector $\mathbf{x} = (1, 2, 3, 4, \ldots, 20)'$. The result should look like $\mathbf{y} = (1, 3, 6, 10, \ldots, 210)'$. Solve this problem by

   (a) using a for loop,

   (b) matrix multiplication (vectorized solution),

   (c) an R function (search for it with `??`).

   Which solution is the most efficient?

d) Write a function that implements some statistical procedure. This could either be

   - a simple linear regression, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; your function should estimate the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ using, e. g., $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$; or
   - a test for independence in a two-by-two contingency table that computes both the Pearson-chi-square statistic

$$X^2 = \sum_{cells} \frac{(x_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}},$$

     and the likelihood ratio statistic

$$G^2 = 2 \sum_{cells} x_{ij} \log \frac{x_{ij}}{\hat{\mu}_{ij}},$$

     where x_{ij} and $\hat{\mu}_{ij} = x_{i+}x_{+j}/x_{++}$ are the observed and expected frequencies, respectively; or

- any other statistical test that you find interesting (maybe search at Wikipedia for the formulae).

Assign the return value of your function a new class. Write a print method for this class.