CT5102: Programing for Data Analytics

Assignment 1: Atomic Vectors

Combinations of Dice						**Sum of two dice**					
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(2)	(3)	(4)	(5)	(6)	(7)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(3)	(4)	(5)	(6)	(7)	(8)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)	(4)	(5)	(6)	(7)	(8)	(9)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)	(5)	(6)	(7)	(8)	(9)	(10)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)	(6)	(7)	(8)	(9)	(10)	(11)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)	(7)	(8)	(9)	(10)	(11)	(12)

2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	*2*	*3*	*4*	*5*	*6*	*5*	*4*	*3*	*2*	*1*

The goal of this assignment is to use vectors to perform simulation of two dice rolls, and analyse the data (see table above for combinations, outcomes and expected frequencies). The R features that can be used are:

- The sample() function, to simulate the roll of a dice, assuming each face has an equal chance of being selected.
- Logical subsetting for atomic vectors
- qplot(data, binwidth=1) to view the simulations
- The mean() and sqrt() functions to calculate the root mean square error which captures the distance between the simulation results and the true probabilities (e.g. 6/36 for the number 7 etc)

For sample sizes of 10^1, 10^2, 10^3, 10^4, 10^5 and 10^6:

- Generate the simulation of two dice rolls
- Plot on the screen to visually verify that they distribution looks as expected
- Extract the frequency of the 11 possible outcomes (note the *table* function cannot be used for this analysis, rather a form of logical subsetting should be used), and the results copied into a new atomic vector
- Calculate the RMSE between the simulated dice rolls and the expected outcomes.
- Plot this RMSE value for each of the sample sizes on the x axis, using qplot (x,y)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$