HW6:ANLY-511 Homework assigned on 10/2/17

1. (2 points) Let $X_1, ..., X_{12}$ be a random sample of size 12 from the $U(0,1)$ distribution. Explain why $Z = X_1 + X_2 + ... + X_{12} - 6$ has an approximate standard normal distribution. You will have to find or look up the variance of a single X_i.
2. (2 points) Problem 4.4 #12 in Chihara/Hesterberg.
3. (2 points) Problem 3.9 #9 in Chihara/Hesterberg
4. (2 points) Problem 3.9 #25 in Chihara/Hesterberg. Import the dataset Lottery.csv and conduct a test of the null hypothesis that the data in the file come from a multinomial distribution on $\{1, ..., 39\}$ with all $p_i = 1/39$. Report the p-value 10 and state your conclusion. This is similar to the question whether birth dates of soccer players follow a uniform distribution
5. (2 Points) Problem 4.4 #28 in Chihara/Hesterberg, but with sapply() instead of a for loop
6. (5 points) Problem 4.4 #14 in Chihara/Hesterberg. Use results from Appendix A to do this problem.
7. (5 points) Let $X_1, X_2, ..., X_9 \sim U(0,1)$ and let $Y_1, Y_2, ..., Y_{12} \sim U(0.5, 1.5)$, all independent. Let $W = \bar{X} - \bar{Y}$ be the difference of the sample means.
   a) Find the sampling distribution of W approximately.
   b) Simulate the sampling distribution of W in `R` and plot your results (adapt code from the previous exercise). Check that the simulated mean and the standard error are close to the theoretical mean and the standard error.
   c) Use your simulation to find $\Pr(W < .6)$. Calculate an approximate answer using the results of part a) and compare.
8. (5 points) Import the data set `Titanic.csv` which contains survival data (0 = death, 1 = survival) and ages of 658 passengers of the Titanic which sank on April 15, 1912 (the day when Americans had to file income tax returns for the first time). Examine the null hypothesis that the mean ages of survivors and of victims are the same against the alternative that these mean ages are different, using a permutation test. Compute the p-value and state your conclusion. This is a two-sided test - how is the p-value computed in this case?