Programming Assignment 6
STAT 311

Please complete the following problems and submit a file named `STAT311-HW6.R` to Gradescope. You should start from the provided `STAT311-HW6.R` file on Canvas.

Overview

Address each of the following questions. As always, you should start from the provided skeleton code.

Question 1 - (4 points)

Consider the following table of "theft of vehicle" crimes that occurred in Vancouver, Canada and were reported to the Vancouver Police Department's Calls for Service Database. (Data from Intra-week spatial-temporal patterns of crime by Andresen and Malleson)

	Mon	Tue	Wed	Thur	Fri	Sat	Sun
Theft of Vehicle	1031	951	931	911	917	1004	951

We wish to test the hypothesis that the rate of vehicle thefts is constant throughout the week. This data can be found in the variable `vehicleTheft`.

a) Calculate the expected counts for the number of theft of vehicle crimes for each day of the week. Your answer should be in the form of a list of length 7. Save your answer in the variable `q1.a`

b) Calculate the χ^2 test statistic for testing the hypothesis described. Save your answer in the variable `q1.b`

c) Calculate the p-value for the hypothesis. Save your answer in the variable `q1.c`

d) `TRUE` or `FALSE`, testing at a 5% level, we can reject the null hypothesis and conclude that vehicle thefts are more common on Monday and Saturday. Save your answer in the variable `q1.d`

Question 2 - (5 points)

A nervous gambler finds that they keep losing money at the craps table, and suspect the house of cheating. They record a large number of rolls (over which they continue to lose money...) and find the following data.

	2	3	4	5	6	7	8	9	10	11	12
Number of Rolls	15	30	36	50	57	59	55	48	36	20	17

Note that in craps, each roll is a roll of two dice, summed together, taking values from 2 to 12. This data can be found in the variable `gambler`.

a) To find evidence that the dice are unfair, what will be the assumed distribution of each possible value? Your answer should be a list of length 11, with the first value representing $P(X = 2)$, the second $P(X = 3)$, etc. Save your answer in the variable `q2.a` (To check your answer: The variable `q2.a` should sum to 1.)

b) Calculate the expected counts for the number of rolls of each die. Your answer should be in the form of a list of length 11. Save your answer in the variable `q2.b`

c) Calculate the χ^2 test statistic for testing the hypothesis that the dice are unfair. Save your answer in the variable `q2.c`

d) Calculate the p-value for the hypothesis. Save your answer in the variable `q2.d`

e) `TRUE` or `FALSE`, testing at a 5% level, the gambler cannot reject the hypothesis that the dice are fair. Save your answer in the variable `q1.e`

Question 3 - (6 points)

A survey of voters in "Frontline" counties, counties that represent highly contested regions for congressional races, contacted a total of 1000 voters. Poll respondents were split in half, and each half was asked different questions. 322 of 500 respondents agreed with "Making community college tuition-free for all". 341 of 500 respondents agree with "Making community college tuition-free for families with incomes below \$125,000 a year". We wish to test the hypothesis that the percentage of voters in frontline districts supporting free community college for lower income families is higher than the percentage supporting free community college for all families. Specifically, we test against the alternative $p_{lowerincome} - p_{all} > 0$.

Note that these questions should be solved using a normal approximation, not the exact or continuity corrected solutions provided by `prop.test`.

a) If we don't assume the two proportions are the same, what is our best estimate of the standard error of the difference between these two proportions? Save your answer in the variable `q3.a`

b) Using the estimate of the standard error found above, calculate a 95% confidence interval for the difference between the percentage supporting free community college for lower income families and the percentage supporting free community college for all families. (Specifically calculate a CI for $p_{lowerincome} - p_{all}$). Save your answer in the variable `q3.b`

c) Under the assumption needed for our hypothesis test, we can calculate a more accurate estimate of the standard error of the difference between the two proportions; calculate that value. Save your answer in the variable `q3.c`

d) Calculate the test statistic Z for this hypothesis test. Save your answer in the variable `q3.d`

e) Calculate the p-value for this hypothesis test. Save your answer in the variable `q3.e`

f) `TRUE` or `FALSE`, testing at a 5% level, we can reject the alternative that a higher proportion of voters support free community college for lower income families than support free community college for all families. Save your answer in the variable `q3.f`