R assignment-3

Math 361

1. The following information is found on Catalina Foothills School disctrict webpage: http://www.cfsd16.org/public/_distinfo/dist_testing.aspx

"Students in our district are required to take the Stanford 10 and AIMS (Arizona Instrument to Measure Standards) exams, which measure academic performance in specific areas. Students in grades 2 and 9 take the Stanford 10, a national norm-referenced assessment created by Pearson covering language arts and mathematics. Students in grades 3-8 take the AIMS DPA (Dual Purpose Assessment), a combination of AIMS criterion-referenced assessment items based on the Arizona Academic Standards and items from the Stanford 10.

Beginning with the class of 2006, 10th grade students must pass the AIMS in order to graduate. Students in grade 10 take the AIMS HS (High School) assessment and continue to test twice annually in grades 11-12 until they have met or exceeded the standard in each area tested - writing, reading, and mathematics."

The table, called table-3, shows the Catalina Foothills School disctrict's AIMS mathematics score. Consider the percentage scores for meeting the criteria ($\%Meets$) for 2011. These data are also found in an excel-file called data-3.

(a) Import the excel-file called data-3 into R by using the scan command (see page 104 in Verzani's book and the lecture notes.)
(b) Use R to calculate the overall mean, variance, and standard deviation of the percentage scores for meeting the criteria for 2011.
(c) How does the mean in part (b) compare to the overall mean (for grades 3,4,5,6,7,8,10) for the state in table-3?

Problem 2, see next page.

Written by John Palmer

2. I thought it would be interesting to use R to simulate a "game show" problem that frequently confuses people the first time they hear it. A game show contestant is faced with 3 doors (which we will just call 1, 2, and 3). Behind one of the doors is riches beyond the dreams of algebra. The other two doors obscure 100 lbs of bricks, and an annoying Parrot. The contestant chooses one the three doors and will win whatever is behind that door. To make things more interesting the inevitable game show host intrudes after the contestant has made a choice and opens one of the two doors not chosen by the contestant to reveal either the bricks or the Parrot. The question is: *Should the contestant stay with his or her original choice or should they switch?*

We will use R to simulate playing this game many times and try to determine if there is an advantage to switching. We first need a way to make a random choice of a door. It is not completely necessary but we will first make a random choice of a door to put the riches behind, then we will have the contestant make a second random choice of a door. In both cases we suppose that each door is picked with probability $\frac{1}{3}$. To generate a random string of 1, 2 and 3's we will use the random number generator *runif*. The R command,

$$x = runif(500, min = .5, max = 3.5),$$

generates a sequence of 500 numbers picked between .5 and 3.5 with the *uniform distribution*. For this distribution a number between .5 and 1.5 is as likely to occur as a number between 1.5 and 2.5 or a number between 2.5 and 3.5. Now execute,

$$x = round(x).$$

The function *round* takes a number between .5 and 1.5 and rounds it off to 1. It rounds numbers between 1.5 and 2.5 to 2 and etc. Type x and press enter and you should see a sequence of 500 numbers 1, 2, or 3. I claim that the probability of getting a 1 or a 2 or a 3 is $\frac{1}{3}$. One simple way to test if this is plausible is to enter $hist(x)$. This will graphically give the frequencies with which 1, 2, and 3 occur.

Problem 1. *Print out histograms for the frequency of 1, 2, and 3 for $N = 500$,$N = 10,000$ and $N = 100,000$ where N is the number of random numbers generated. If things are working properly the histograms should show that the number of 1's, 2's and 3's are all roughly equal. Be careful not to print out x for $N = 10,000$ or $N = 100,000$ or you will overwhelm the screen.*

Now we are ready to do the experiment. Fix a sample size (maybe 500 trials) and generate two sequences of random 1, 2, and 3's named x and y. You can think of x as the choice behind which lie the riches and y is the sequence of contestant guesses. If the contestant chooses the strategy of *always staying with the original guess* then the number of times that this strategy will win for him or her is precisely the number of places at which x and y match up exactly. To see how often this takes place I need to explain that when you take the difference $x - y$ it gives a vector of the same length as x and y (number of entries) and the j^{th} entry of $x - y$ is $x[j] - y[j]$ (the subtraction takes place entry by entry).

Problem 2. *How does $hist(x - y)$ gives you an idea of the number of matches in x and y ?*

To be more precise about this number we make use of another feature of R. Type,

$$comp = (x == y)$$

This generates a vector with the same number of entries as x and y but with each entry either TRUE or FALSE. The j^{th} entry is TRUE if $x[j] = y[j]$ and FALSE otherwise. Note: you can also use $<$ or $>$ and etc, instead of the *exactly equal* sign, ==, to get different and sometimes useful logical vectors. Now type,

$$sum(comp)$$

When TRUE and FALSE are forced to numerical values $TRUE = 1$ and $FALSE = 0$. Thus $sum(comp)$ counts the number of exact matches between x and y.

Problem 3 *Run this simulation often enough to get a feeling for the probability that the strategy of not switching wins. Record the results and make a guess what this probability is.*

What about the strategy of *always switching?* I claim that those times which you win with the strategy of always switching are *exactly* those times that you lose with the strategy of never switching.

Problem 4. *Explain this and use the data to estimate the probability of success with the strategy of always switching. Make an analysis of this problem that does not rely on simulations.*