R Question

1. First, I simulated 100,000 samples and plotted the resulting histogram to get a sense of the distribution. The resulting histogram resembled a binomial distribution, with slightly less than 20% of the observations equal to -50, and the remaining ~80% equal to 15.
Next, I found the mean value of 100 samples, and replicated this 10,000 times. This vector was also plotted, and revealed a roughly normal distribution. The mean calculated from this vector was ~4.15.
I think this answer is reasonably close, given the number of samples simulated.

My willingness to play this game is dependent on the amount of money I have to play this game. Over the long term, I'd expect to gain money, but would need to be able to absorb a few losses.

2. I chose to use the Poisson distribution and the rpois() and dpois() methods. The comments included in the .R file support the exploration of the Central Limit Theorem.

3. The population mean is typically unknown, as its oftentimes infeasible, or outright impossible, to measure the desired parameter across an entire population. Even if it can be done, it is likely prohibitively expensive.
Instead, the sample mean can be studied in order to infer the population mean.

4.
a. This distribution appears to be skewed to the right, as it is asymmetric about its mean. Outliers are more probable on the right side of the distribution.
b. The mean and median are not likely the same value, as in a symmetric normal distribution. In this case, large but improbable observations skew the mean to a larger value than the median.
c. This distribution appears to have a single mode, and is hence unimodal. While its not possible for a distribution to have more than one mean or median, by definition of those terms, it is possible to have more than one mode.

5. X1, X2, X3 and s, 2s, s => X = 4X1 + 0.5X2 + X3
a. The variance is a measure of the typical distance from the mean that an observation exhibits. Simply summing values would result in values lower than the mean cancelling out values higher than the mean, and would result in an inaccurate variance. Squaring the deviation from the mean preserves distance.
b. Using absolute variations to describe variance would prevent comparisons between different sample sizes, as the more samples collected, the higher the variance would appear.
c. One advantage of using variance is computational efficiency: to compute the standard deviation, you must first compute the variance. Using variance avoids the additional step.
Furthermore, the variability of linear combinations of independent random variables is simply a sum, which is not the case for standard deviations.
d. The variance of X2 is 4 times the variance of X1, indicating much more "spread" in the distribution.
e. The variance of X is the variability of a linear combination of X1, X2 and X3. In this case, that would be the sum: 16s^2 + s^2 + s^2 = 18s^2, 18 times the variance of X1.

6. Frost identical cakes, calculate mean cake frosting times. 10 random bakers chosen, mean time is 8.73 min. 1000 random bakers: 10.19 min. 100,00: 9.97 min.
I would guess the population mean to be approximately 10 minutes, as the Law of Large Numbers dictates that the sample mean approaches the population mean as the sample size increases.