7.SP.2 – Valentine Marbles

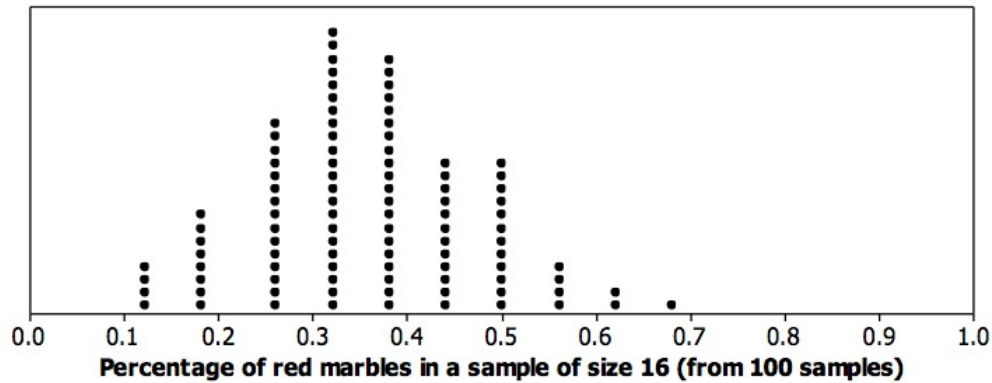## Alignments to Content Standards

- Alignment: 7.SP.A.2

Tags

• *This task is not yet tagged.*

A hotel holds a Valentine's Day contest where guests are invited to estimate the percentage of red marbles in a huge clear jar containing both red marbles and white marbles. There are 11,000 total marbles in the jar: 3696 are red, 7304 are white. The actual percentage of red marbles in the entire jar ($33.6\% = \frac{3696}{11000}$) is known to some members of the hotel staff.

Any guest who makes an estimate that is within 9 percentage points of the true percentage of red marbles in the jar wins a prize, so any estimate from 24.6% to 42.6% will be considered a winner. To help with the estimating, a guest is allowed to take a random sample of 16 marbles from the jar in order to come up with an estimate. (Note: when this occurs, the marbles are then returned to the jar after counting.)

One of the hotel employees who does not know that the true percentage of red marbles in the jar is 33.6% is asked to record the results of the first 100 random samples. A table and dotplot of the results appears below.

Percentage of red marbles in the sample of size 16	Number of times the percentage was obtained
12.50%	4
18.75%	8
25.00%	15
31.25%	22
37.50%	20
43.75%	12
50.00%	12
56.25%	4
62.50%	2
68.75%	1
Total:	100

Percentage of red marbles in a sample of size 16 (from 100 samples)

For example, 15 of the random samples had exactly 25.00% red marbles; only 2 of the random samples had exactly 62.50% red marbles, and so on.

a. Assuming that each of the 100 guests who took a random sample used their random sample's red marble percentage to estimate the whole jar's red marble percentage. Based on the table above, how many of these guests would be "winners"?

b. How many of the 100 guests obtained a sample that was *more than* half red marbles?

c. Should we be concerned that none of the samples had a red marble percentage of exactly 33.6% even though that value is the true red marble percentage for the whole jar? Explain briefly why a guest can't obtain a sample red marble percentage of 33.6% for a random sample of size 16.

d. Recall that the hotel employee who made the table and dotplot above didn't know that the real percentage of red marbles in the entire jar was 33.6%. If another person thought that half of the marbles in the jar were red, explain briefly how the hotel employee could use the dotplot and table results to challenge this person's claim. Specifically, what aspects of the table and dotplot would encourage the employee to challenge the claim?

e. Design a simulation that takes a large number of samples of size 16 from a population in which 65% of the members of the population have a particular characteristic. For each sample of size 16, compute the percentage of red items in the sample. Record these percentages, and then summarize all of your sample percentages using a table and dotplot similar to those shown above. In what ways is your dotplot similar to the dotplot used in this task? In what ways does it differ?

Commentary

For this task, Minitab software was used to generate 100 random samples of size 16 from a population where the probability of obtaining a success in one draw is 33.6% (Bernoulli). Given that multiple samples of the same size have been generated, students should note that there can be quite a bit of variability among the estimates from random samples and that on average, the center of the distribution of such estimates is at the actual population value and most of the estimates themselves tend to cluster around the actual population value. Although formal inference is not covered in Grade 7 standards, students may develop a sense that the results of the 100 simulations tell them what sample proportions would be expected for a sample of size 16 from a population with about $\frac{1}{3}$ successes.

Regarding the last question, students may need some guidance with the devices used for simulation (and if needed, you can consider adjusting the probability of success to $\frac{2}{3}$ if dice or some similar media are used). Ideally, students should attempt several simulations and note that similarly to the example provided in the task, the center of the distribution appears to be close to the actual population percentage, that most of the estimates are near this value, and that there is considerable variability in the estimates. In this case, the dotplot should be close to a mirror image of the one that appears in the task since the new probability of success is roughly the complement of the probability of success used in the earlier portion of the task. Note: consider having several groups of students each record several simulation results (e.g., 20 results) on their own and then having each group contribute these results to a larger class dotplot and table of everyone's results (e.g., the work of 105 groups could then be recorded on a master classroom dotplot which shows the results of 2500 simulations).

Solutions

Solution: Solution

a. Since any estimate from 24.6% to 42.6% will be considered a winner, all of the estimates of 25%, 31.25%, and 37.5% would be the "winners." From the table or dotplot, that would be 15 + 22 + 20 respectively, which would make for a total of 57 winning estimates.

b. From the table or dotplot, an estimate that is "more than half" would be any estimate of 56.25%, 62.5%, or 68.75%. That would be 4 + 2 + 1 respectively, which would make for a total of 7 estimates which have values that correspond to "more than half red."

c. In this case, the estimate from any random sample of 16 marbles will always be of the form $\frac{X}{16}$ where X is a discrete value from $0, 1, 2, \ldots 16$. None of those possible proportions can equal exactly 33.6%; the closest possible sample percentage is 31.25% ($= \frac{5}{16}$) which happens to be the most frequent estimate in the table. Alternatively, $0.336 \cdot 16 = 5.376$, and it is not possible to draw 5.376 red marbles as the number of reds drawn must be a whole number between 0 and 16 (inclusive).

d. Under the assumption that the random samples are representative of the population from which they were selected, if half of the marbles in the glass jar were red, then the dotplot and table would most likely show a majority of sample values centered near 50%. We notice from the collected data that most of the sample estimates are in the 25% to 43.75% range, and the graph seems to be centered more in the low-30% values. Other arguments that discuss how the graph is centered at a value far from 50% and/or how a small number of the estimates are actually near 50% would also be appropriate.

e. Regardless of simulation method, the distribution of the sample percentages should be centered between 60 and 70% and have a spread and approximate shape that are similar to the graph shown earlier in this task. In some respects, the distribution should resemble a "mirror image" of the distribution shown in this task.