Quiz 4 – Solutions

1. (4 pts) A box contains 100 tickets: 60 $\boxed{1}$s and 40 $\boxed{0}$s. Tickets are drawn from the box at random with replacement, and you win a dollar if more $\boxed{0}$s are drawn than $\boxed{1}$s. There are two choices:

(i) 10 draws are made from the box.
(ii) 100 draws are made from the box.

Which choice gives a better chance of winning that dollar, (i) or (ii), or do they both give the same chance of winning?

According to the law of averages, as the number of draws from the box gets bigger, the more likely it is that the percentage of $\boxed{1}$*s in the sample will be close to the percentage of* $\boxed{1}$*s in the box. In the game described above, you win* \$1.00 *if you draw more* $\boxed{0}$*s than* $\boxed{1}$*s . This means that to win, we want the percentage of* $\boxed{1}$*s to be* ***less than*** 50%*, which is* ***not close*** *to the percentage of* $\boxed{1}$*s in the box. Drawing more times from the box makes this less likely to happen (according to the law of averages), so it would be better in this case to draw* 10 *tickets from the box,* ***not*** 100 *tickets.*

2. 100 draws are made at random with replacement from the box $\boxed{\boxed{2}\,\boxed{4}\,\boxed{5}\,\boxed{6}\,\boxed{8}}$.

(a) (4 pts) Find the ***expected value*** and the ***standard error*** for the sum of the draws.

The average of the box is $\dfrac{2+4+5+6+8}{5} = 5$ *and the standard deviation of the box is*

$$SD = \sqrt{\frac{(2-5)^2 + (4-5)^2 + 5-5)^2 + (6-5)^2 + (8-5)^2}{5}} = 2.$$

Therefore the expected value of the sum is

$$EV(sum) = n \cdot Avg = 100 \cdot 5 = 500$$

and the standard error is

$$SE(sum) = \sqrt{n} \cdot SD = \sqrt{100} \cdot 2 = 20.$$

(b) (2 pt) What is the approximate probability that the ***observed sum*** of the draws is between 470 and 510?

According to the central limit theorem, when the sample size is sufficiently large, the probability histogram for the sum of n *draws (taken at random with replacement) is approximated by the normal curve once it is converted to standard units. In this case, it means that*

$$P(470 \le \textit{sum of 100 draws} \le 510) \approx \textit{area under the normal curve between } \frac{470-500}{20} = -1.5 \textit{ and } \frac{510-500}{20} = 0.5.$$

The area under the normal curve between -1.5 *and* 0 *is* 1/2 *the table entry for* $z = 1.5$ *and the area under the normal curve between* 0 *and* 0.5 *is* 1/2 *the entry in the table for* $z = 0.5$*. I.e.,*

$$P(470 \le \textit{sum of 100 draws} \le 510) \approx \frac{1}{2}T(1.5) + \frac{1}{2}T(0.5) \approx 62.465\%.$$

3. (6 pts) 100 draws are made at random with replacement from a box with 36 $\boxed{1}$s and 64 $\boxed{0}$s . Find the approximate probability that the percentage of $\boxed{1}$s in the sample is between 32% and 40%.

The proportion of $\boxed{1}$*s in the box is* 0.36 *so the standard deviation of the box is* $SD = \sqrt{0.36 \times 0.64} = 0.48$*. The expected value for the percentage of* $\boxed{1}$*s in* $n = 100$ *draws from this box (at random, with replacement) is*

$$EV(\%) = 36\% \qquad (= \textit{the percentage of } \boxed{1}\textit{s in the box})$$

and the standard error for the percentage is

$$SE(\%) = \frac{SD}{\sqrt{n}} \times 100\% = \frac{0.48}{10} \times 100\% = 4.8\%.$$

Therefore according to the normal approximation,

$$P(32\% \le \textit{sample } \% \le 40\%) \approx \textit{area under normal curve between } \frac{32\% - 36\%}{4.8\%} \approx -0.833 \textit{ and } \frac{40\% - 36\%}{4.8\%} \approx 0.833.$$

This area is between the table entries for 0.8 *(*57.63%*) and* 0.85 *(*60.47%*), so the probability is about* 59% *or so.*

4. (4 pts) A sample of $n = 400$ adults from a large city is surveyed. The percentage of adult women in the city is 52% and the percentage of women in the sample is 59%. Is it likely that the sample was a ***simple random sample***? Explain why or why not.

The sample percentage of women in a simple random sample of 400 *adults is expected to be close to the population percentage of* 52%*, and* 59% *is not close at all, statistically speaking, and it is unlikely that a simple random sample would yield such a sample percentage.*

Specifically, the standard error for the sample percentage in this case is

$$SE = \frac{SD(box)}{\sqrt{n}} \times 100\% = \frac{\sqrt{0.52 \times 0.48}}{\sqrt{400}} \times 100\% \approx 2.5\%,$$

and the normal approximation tells use that the sample percentage of women in simple random samples of 400 *adults is very likely to be within* 5% $(= 2 \cdot SE(\%))$ *of the population percentage. Going further, the probability that the sample percentage of women in a simple random sample is between* $46\% = 52\% - 6\%$ *and* $58\% = 52\% + 6\%$ *is roughly equal to the area under the normal curve between* -2.4 *and* 2.4 *(because* $6\%/2.5\% = 2.4$*), and this area is* 98.36%*. The given sample percentage of* 59% *is well outside this range, which is possible with a simple random sample, but is very unlikely, so we can say that it is unlikely that the (hypothetical) sample was a simple random sample.*[†]

[†]Another explanation is that the assumption about the percentage of women in the city is wrong. We will explore this idea in the context of ***tests of significance*** next week. In this problem however, we accept this assumption as absolutely true.