UCSC AMS 5-02 **Quiz 4 – White Solutions** **Winter 2017**

1. Four hundred draws are made at random with replacement from a box of numbered tickets containing 8 $\boxed{1}$s and 2 $\boxed{0}$s.

(a) (2 pts) What is the ***expected*** percentage of $\boxed{1}$s in the sample?

The expected percentage of $\boxed{1}$ *s is equal to the percentage of* $\boxed{1}$ *s in the box, which is* 80%*.*

(b) (2 pts) What is the ***standard error*** (SE) for the percentage of $\boxed{1}$s in the sample?

$$SE_{\%} = \frac{SD(box)}{\sqrt{n}} \times 100\% = \frac{\sqrt{0.2 \times 0.8}}{\sqrt{400}} \times 100\% = 2\%.$$

(c) (2 pts) What is the (approximate) probability that the sample percentage of $\boxed{1}$s is between 78% and 82%? Why?

According to the normal approximation, the required probability is (approximately) equal to the area under normal curve between $\frac{78\%-80\%}{2\%} = -1$ *and* $\frac{82\%-80\%}{2\%} = 1$*. I.e., the probability is about* 68%*.*

2. A market research firm surveyed a simple random sample of 1600 households from a large metropolitan area of more than 100, 000 households.

(a) (4 pts) Of the sample households, 960 owned two or more cars. Use this data to construct a 95%-confidence interval for the percentage of all households in the metropolitan area who own two or more cars, or explain why this is not possible.

The sample percentage of two-car households is $(960/1600) \times 100\% = 60\%$ *and (using the sample data to estimate the SD of the population) the standard error for percentage is approximately*

$$SE_{\%} \approx \frac{\sqrt{0.6 \times 0.4}}{\sqrt{1600}} \times 100\% \approx 1.225\%.$$

The 95%*-confidence interval for the percentage of two-car households in the metropolitan area is:*

$$\textit{sample } \% \pm 2SE_{\%} = 60\% \pm 2.45\% \qquad (\textit{ or } (57.55\%, 62.45\%)\ )$$

(b) (2 pts) The 1600 households in the sample included 2500 children age 6 or younger. Of these 2500 children, 1375 watched at least three hours of TV per day.

True or false, and explain briefly: The percentage of all children age 6 or younger in the metropolitan area who watch at least three hours of TV per day is likely to be 55% give-or-take 1%.

False*. The numbers are correct for a* ***simple random sample*** *of children, but the sample described above is* ***not*** *a simple random sample.*†

†The sample described in problem 2(b) is an example of a *cluster sample*.

3. The average household income for all the households in the sample in the problem above was $\$56,500$ with an SD of $\$14,000$.

True or False, and justify your answer briefly:

(a) (2 pts) A 95%-confidence interval for the average household income in the metropolitan area is about $\$56,500 \pm \$700$.

True. *A 95%-confidence interval for the average household income in the metropolitan area is given by* (sample average) $\pm\ 2SE$. *The standard error in this case is* $SE = 14000/\sqrt{1600} = 350$ *so* $2SE = 700$.

(b) (2 pts) The chance is about 95% that the interval in (a) contains the average household income in the sample.

False. *The chance is exactly* 100% *that the interval* $\$56,500 \pm \$700$ *contains the sample average of* $\$56,500$.

(c) (2 pts) About 95% of the households in the metropolitan area have incomes between $\$55,800$ and $\$57,200$.

False. *This is not what the confidence interval for* **average** *income predicts. Moreover the income data for the population is likely to have a similar spread to the sample data which is summarized by the* $SD = \$14,000$, *so the incomes in the metropolitan area are much more spread out than indicated in this question.*

(d) (2 pts) The chance is about 95% that the interval in (a) contains the average household income in the metropolitan area.

True. *This is the correct interpretation of a confidence interval for average income.*