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

1. Sixteen hundred draws are made at random with replacement from a box of numbered tickets containing 80 $\boxed{1}$s and 20 $\boxed{0}$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?

$$SE_{\%} = \frac{SD(box)}{\sqrt{n}} \times 100\% = \frac{\sqrt{0.2 \times 0.8}}{\sqrt{1600}} \times 100\% = 1\%.$$

(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 $\frac{78\%-80\%}{1\%} = -2$ *and* $\frac{82\%-80\%}{1\%} = 2$*. I.e., the probability is about* 95%.

2. A market research firm surveyed a simple random sample of 2500 households from a large metropolitan area of more than $100,000$ households.

(a) (4 pts) Of the sample households, 1960 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 $(1960/2500) \times 100\% = 78.4\%$ *and (using the sample data to estimate the SD of the population) the standard error for percentage is approximately*

$$SE_{\%} \approx \frac{\sqrt{0.784 \times 0.216}}{\sqrt{2500}} \times 100\% \approx 0.82\%.$$

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

$$\textit{sample } \% \pm 2SE_{\%} = 78.4\% \pm 1.64\% \qquad (\textit{ or } (76.76\%, 80.04\%)\ )$$

(b) (2 pts) The 2500 households in the sample included 3600 children age 12 or younger. Of these 3600 children, 2160 watched at least two hours of TV per day.

True or false, and explain briefly: The percentage of all children age 12 or younger in the metropolitan area who watch at least two hours of TV per day is likely to be 60% give-or-take 0.8%.

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 $\$49,500$ with an SD of $18,750.

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 $\$49,500 \pm \$750$.

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* $SE = \$18,750/\sqrt{2500} = 375$ *so* $2SE = 750$.

(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* $\$49,500 \pm \$750$ *contains the sample average of* $\$49,500$.

(c) (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.*

(d) (2 pts) About 95% of the households in the metropolitan area have incomes between $\$48,750$ and $\$50,250$.

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* $SD = \$18,750$, *so the incomes in the metropolitan area are much more spread out than indicated in this question.*