1. 100 hundred draws are made at random with replacement from a box of numbered tickets containing 9 $\boxed{1}$s and 1 $\boxed{0}$s.

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

$$\textit{Expected percentage} = \textit{percentage of } \boxed{1} \textit{ s in box} = \frac{9}{10} \times 100\% = 90\%$$

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

$$\textit{SE for the sample } \% = \frac{\sqrt{(\textit{proportion of } \boxed{1})(\textit{proportion of } \boxed{0})}}{\sqrt{\textit{sample size}}} \times 100\% = \frac{\sqrt{0.9 \cdot 0.1}}{\sqrt{100}} \times 100\% = 3\%$$

(c) (2 pts) What is the (approximate) probability that the sample percentage of $\boxed{1}$s is between 87% and 93%? Why?

Sample percentages approximately follow the normal curve when n is big enough (which it is in this case), so

$$P(87\% \leq \textit{sample } \% \leq 93\%) \approx 68\%$$

because this is the probability that the sample % falls in the range $EV(\%) \pm 1 \cdot SE(\%)$, which is approximately equal to the area under the normal curve from -1 to 1.

2. A market research firm surveyed a simple random sample of 2500 households from a large metropolitan area of more than 500,000 households.

(a) (4 pts) Of the sample households, 1960 owned two or more computers. Use this data to construct a 95%-confidence interval for the percentage of all households in the metropolitan area who own two or more computers. *Show your work.*

A 95%-confidence interval for the percentage of all households who own two or more computers is given by (sample %) $\pm\, 2SE(\%)$. In this case, we have:

$$\textit{sample } \% = \frac{1960}{2500} \times 100\% = 78.4\% \quad \textit{and} \quad SE(\%) \approx \frac{\sqrt{0.784 \times 0.216}}{\sqrt{2500}} \times 100\% \approx 0.823\%$$

so $2SE(\%) \approx 1.646\%$ and the confidence interval is $(78.4\% \pm 1.646\%)$.

(b) (2 pts) The 2500 households in the sample included 4900 children age 6 or younger. Of these 4900 children, 2940 watched at least three hours of TV per day.

True or false, and explain (briefly): A 95%-confidence interval for the percentage of all children age 6 or younger in the metropolitan area who watch at least three hours of TV per day is 60%±1.4%.

False. *The sample % of children age 6 or younger who watch at least three hours of TV daily is $(2940/4900) \times 100\% = 60\%$ and $(\sqrt{0.6 \times 0.4}/\sqrt{4900}) \times 100\% \approx 0.7\%$... So, **if this were a simple random sample** of children then the confidence interval would be correct. But it **isn't** a simple random sample of children — it is a **cluster sample**, so the calculation is incorrect.*

3. The average household income for all the households in the sample in the problem above was $\$48,500$ with an SD of $\$15,000$.

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 $\$48,500 \pm \$600$.

True. *This is a simple random sample of households, the sample average is* $\$48,500$ *and the SE (for the average) is* $15000/\sqrt{2500} = \$300$*, so the given confidence interval is correct (because* $2 \times 300 = 600$*).*

(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 in (a) contains the average household income* ***in the sample****, because the sample average is the middle of the interval.*

(c) (2 pts) About 95% of the households in the metropolitan area have incomes between $\$47,900$ and $\$49,100$.

False. *The standard error (*300 *in this case) measures the variation between different sample averages,* **not** *between different household incomes. The variation in income between households is estimated by the sample* $SD = 15000$*. Moreover, there is no reason to believe that household income in this region has a normal distribution.*

(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 precisely how a* 95%*-confidence interval for a population average is interpreted. More precisely, we expect that about* 95% *of all intervals constructed in this way will contain the population average and therefore any one of them (like the one in (a)) has a* 95% *chance of containing the population average.*