- For full credit: ***show/explain your work/calculations on every question***.
- Please write clearly — if we can't read it, we won't give you credit for it.
- Please do ***not*** write your answers between the lines of the questions. You can write your answers on the bottom and back of this sheet and/or on a separate piece of paper.

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

(a) (3 pts) What is the ***expected number*** of $\boxed{1}$s in the sample?

Expected number = (proportion of $\boxed{1}$s in box)×(number of draws)=$\frac{80}{100} \times 400 = 320$.

(b) (3 pts) What is the ***standard error*** (SE) for the ***number*** of $\boxed{1}$s in the sample?

SE(count)=$SD_{box} \times \sqrt{\text{number of draws}} = \sqrt{0.2 \times 0.8} \times \sqrt{400} = 0.4 \times 20 = 8$.

(c) (4 pts) What is the (approximate) probability that the number of $\boxed{1}$s in the sample is between 320 and 330?

The SE is 8, and the expected value is 320, so $330 - 320 = 10 = 1.25 \times SE$. According to the *Normal Approximation*, the probability that the number of $\boxed{1}$s is between 320 and 330 is (approximately) equal to the area under the normal curve between 0 and 1.25. This area is equal to *half* the area between −1.25 and 1.25 under the normal curve, which (according to the table) is 78.87%. So the probability that we seek is $78.87\%/2 \approx 39.44\%$.

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.

Sample percentage of households with 2+ **cars**= $\frac{960}{1600} \times 100\% = 60\%$.

Standard Error for percentage= $\frac{SD_{pop}}{\sqrt{1600}} \times 100\% \approx \frac{\sqrt{0.6 \times 0.4}}{40} \times 100\% \approx 1.22\%$.

Comment: We use the (known) SD of the sample as an estimate for the (unknown) SD of the population.

95-% confidence interval: (Sample% ±2SE)= (60% ± 2.44%).

(b) (4 pts) Of the sample households, 640 had annual incomes of $65,000 or more. Use this data to construct a 95%-confidence interval for the percentage of households in the metropolitan area with incomes of $65,000 or more, or explain why this is not possible.

Sample percentage of households with income of $65,000+= $\frac{640}{1600} \times 100\% = 40\%$.

Standard Error for percentage= $\frac{SD_{pop}}{\sqrt{1600}} \times 100\% \approx \frac{\sqrt{0.4 \times 0.6}}{40} \times 100\% \approx 1.22\%$.

Comment: We use the (known) SD of the sample as an estimate for the (unknown) SD of the population.

95-% confidence interval: (Sample% ±2SE)= (40% ± 2.44%).

(c) (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: This data shows that the percentage of all children age 6 or younger in the metropolitan area who watch at least three hours of TV per day is about 55% give-or-take 2%.

False: The calculations for sample percentage and standard error are correct,

$$\frac{1375}{2500} \times 100\% 55\% \quad \text{and} \quad \frac{\sqrt{0.55 \times 0.45}}{\sqrt{2500}} \times 100\% \approx 1\%$$

Which might lead to the conclusion that (55% ± 2%) is a 95%-confidence interval for the percentage of children in the population, age 6 and under, who watch 3+ hours of TV per day.

The problem is that this is ***not*** a simple random sample of such children — it is a *cluster sample*, and the correct standard error in this case will generally be larger and the normal approximation is not appropriate.