Solution to homework 4

Problem 1

Central Limit Theorem (CLT):

$Y_1, Y_2, \dots, Y_n$ are independent random observations from distribution with mean μ and variance σ^2. If the sample size n is large, the sample mean follows normal distribution with mean μ and variance σ^2/n.

CLT provides a distribution theory on the sample mean. According to CLT, the mean of a sample large enough is an unbiased estimate of the population mean μ, which is often the parameter of interest. CLT applies regardless of the population distribution. A larger sample size is needed in case of strong non-normal population distribution.

There exists similar distribution theories on sample SD, Q1, IQR, and in fact, any sample quantile. These are widely used results since quantiles are more robust/resistant statistic measures than sample mean.

Problem 2

The number of Children living in poverty, denoted as random variable X, has the distribution: X ~ Binomial (700, 0.22).

So, E(X) = 700*0.22 = 154 and VAR(X) = 700*0.22*0.78 = 120.12.
Approximately, X ~Normal (154, 120.12)

The probability that at least 250 Children are in poverty is
$\text{Prob}(X>=250) = P(Z>(250-154)/(120.12)^{0.5}) = 0$.

Note: It's equivalent to find Prob($\hat{P}$ >250/700), while sample percentage $\hat{P}$ follows distribution $\hat{P}$ ~ Normal(0.22, 0.22*0.78/700) by CLT.

Problem 3

The compression procedure compress 1,000,000 bits into 100,000 bits, therefore the compression ratio is 10:1. This message fails to transfer when more than (12/10,000)*100,000 =120 bits are corrupted during transmission.

The number of bits corrupted X has the distribution X~ Binomial (100,000, 10^{-4}). So E(X) = np = 10 and SD(X) = $[100{,}000 * 10^{-4}*(1-10^{-4})]^{0.5}$ = 3.16. A normal approximation suggests X ~ Normal (10, 3.16).

Therefore, the probability that this message fails to be transferred is
P(X>120) = P(Z>(120-10)/3.16) = 0.