Examples 3 and 4 from September 21st

3) Five of 20 presses are being tested for excessive wear.
   a) If four of the 20 are actually defective, what is the probability that none of the defectives will be found?

So this is a hypergeometric distribution with N=20, n=5, r=4, and the question is seeking P[X=0].

\[
\frac{\binom{4}{0} \binom{16}{5}}{\binom{20}{5}} = \frac{1 \cdot \frac{16!}{5!11!}}{\frac{20!}{5!15!}} = \frac{15 \cdot 14 \cdot 13 \cdot 12}{20 \cdot 19 \cdot 18 \cdot 17} \approx 0.2817
\]

b) Give a reasonable “95% range” for how many defectives you would expect to find in the five sampled.

A rough estimate would be the mean +/- 2 standard deviations. In this case the mean is np=5*(4/20)=1 and the variance can be written np(1-p)(N-n)/(N-1)=5*.2*.8*15/19≈.631579, making the sd≈.7947194.

The rough interval would thus be 1 +/- 1.593. Since only integers can happen, this becomes 0 to 2.

4) Approximately 3.1% of the South Carolina population is of Hispanic or Latino origin.
   a) How many South Carolinians would you need to sample on average before you arrive at your first Hispanic?

If we act like this is a binomial experiment (assuming any sample we take is much smaller than the entire population) then this is a geometric distribution, and the problem is asking for the mean.

\[\mu = \frac{1}{p} = \frac{1}{.031} \approx 32.258\]

b) Give a reasonable “95% range” of how many you would need to sample to have 10 Hispanics included?

Again, assuming a binomial is a good approximation, this is now a negative binomial problem. As in 3b, we need to find the mean and variances. The mean is r(1/p)=10*(1/.031)≈322.58 and the variance is r(1-p)/p^2=10*(1-.031)/.031^2= 10,083.25 (making the sd=100.4154).

A rough interval would thus be 322.58 +/- 200.83, which rounds to 122 to 523.