STAT 702/J702
September 2nd, 2004

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

- Homework Solutions
- The Hypergeometric
- Binomial vs. Hypergeometric

Ch.1 # 38) A child has six blocks, three of which are red and three of which are green. How many patterns can she make by placing them in a line?

What if three white blocks are added?

Ch. 1 # 42) How many ways can 11 boys on a soccer team be grouped into 4 forwards, 3 midfielders, 3 defenders, and 1 goalie?

Ch. 1 # 57) Cabinets A, B, and C each have two drawers with one coin per drawer. A has two gold, B has two silver, and C has one gold and one silver.

A cabinet is chosen at random and a drawer is opened showing a silver. What is the chance the other is silver too?

Last time… <u>Binomial Experiment</u>

1. n identical trials
2. Each trial has only two possible outcomes ("Success" or "Failure")
3. Probability of "Success" is a constant p for every trial
4. Trials are independent

$$P[k \text{ successes in } n \text{ trials}] = \binom{n}{k} p^k (1-p)^{n-k}$$

Hypergeometric Experiment

1. Population of size n
2. r are "successes" and n-r are "failures"
3. A random sample of size m is taken without replacement

Example) An assembly line produced n =2000 parts, of which r =40 were defective. (Note that this is a 0.02 defective rate).

A random sample of size m =20 is chosen. What is the probability that exactly 10 of these 20 will be defectives?

The first "trick" is to realize that, since we are taking a random sample, every possible sample of size 20 has the same probability. (e.g. all of the sample points have the same probability.)

In the binomial case we figured out the probability of each sample point and then multiplied that by the number of sample points in our event.

Another way of calculating the probability of an event when all sample points are equally probable is:

P(A) = $\frac{\text{number of sample points in A}}{\text{total number of sample points}}$

STAT 702/J702 B.Habing Univ. of S.C. 10

In general, for a population of size *n* with *k* successes and a sample of size *m* we get:

$$P[k \text{ successes out of } m] = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

STAT 702/J702 B.Habing Univ. of S.C. 11

Example – Capture/Recapture)

Goal: To estimate the size *n* of a population.

Method: "Randomly" capture, tag, and release *r* of them. Then "randomly capture" *m* of them and see how many are tagged.

STAT 702/J702 B.Habing Univ. of S.C. 12

Now the probability of a certain number being captured will be hypergeometric!

$$P[k \text{ tagged out of } m] = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

The problem is that we know *r*, *k*, and *m*, but we are looking for n!

Since we can't find *n* exactly, we will attempt to estimate it by choosing the value of *n* that "seems most likely". That is, what value of *n* would give us the largest probability of observing the *k* that we did.

Mathematically then, we need to find the *n* that maximizes

$$L_n = P[k \text{ tagged out of } m] = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

If *n* was continuous we could try taking the derivative with respect to *n* and setting it equal to zero.

We will use a similar logic here and take the ratio L_n/L_{n-1} .

So n should be the greatest integer not exceeding mr / k .

So if r =10, m =20, and k =4 we estimate n to be 50.

When are the binomial and hypergeometric similar?

What do you lose if you sample with replacement instead? (e.g. why not always use binomial?)