Problem H-301. A SNP has alleles A and G.

In group Y, the genotypes occur in proportions 30% AA, 60% AG, and 10% GG.
In group Z, the genotypes occur in proportions 20% AA, 30% AG, and 50% GG.
A study sample S is comprised 75% from group Y and 25% from group Z.

(a) What is the probability of each genotype in S?
(b) If a randomly selected individual in S has genotype GG, what is the probability that they are in group Y?
(c) Genotype GG results in plain fur, while AA and AG result in spotted fur. If a randomly selected individual in S has spotted fur, what is the probability that they are in group Z?

Problem H-302.

(a) Write down the likelihood L defined by n i.i.d. random variables $Y_1, \ldots, Y_n$, each having the geometric distribution $P(Y = y) = (1 - p)^{y-1} p, y = 0, 1, 2, \ldots$ (This is the version whose interpretation is Y is the number heads before the first tails, where p is probability of heads.)

(b) For this geometric distribution, determine the Maximum Likelihood Estimate $\hat{p}$ of p if $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$.

Problem H-303. Assume crossovers occur as a Poisson process with rate $\lambda = 1 \text{ M}^{-1} = 0.1 \text{ cM}^{-1}$. Number the crossovers on a chromosome 1, 2, 3, \ldots in order of increasing coordinates. What is the expected value, in cM, of the distance between the third and ninth crossovers, and what is the standard deviation?

Be careful: the positions of these crossovers are dependent.

Problem H-304. Let X have a normal distribution with mean 25 and variance 10. Compute $P(X < 22)$, $P(X > 29)$, and $P(22 < X < 29)$. To do this, convert to the standard normal distribution using Z-scores, and then use Matlab, R, or other suitable software to compute $\Phi(z)$. (In a previous statistics class, you may have done this with a lookup table in your book.)

This question is about the normal distribution, not the binomial distribution, so the continuity correction isn’t applicable.

In Matlab, normcdf(z) gives $\Phi(z)$ and norminv(p) is the inverse function. So p=normcdf(z) when z=norminv(p), to within numerical precision errors. In R, pnorm(z) gives $\Phi(z)$ and qnorm(p) is the inverse function. So p=pnorm(z) when z=qnorm(p), to within numerical precision errors.

See the slides for examples. See the Software page of the class website for Matlab and R help info.

Problem H-305. Let X have the binomial distribution for $n = 100$, $p = .4$ and Y have the binomial distribution for $n = 100$, $p = .3$.

(a) Plot the probability density functions (pdfs) of X and Y on one graph, using different colors or different marker symbols to distinguish X and Y.

(b) Plot the cumulative distribution function (cdf) of X.