**Process of generating conclusions about a population from a noisy sample; generating new knowledge to get answers about a population; Think of probability as being able to repeat an experiment over and over and over and the percentage of times that something happens defines that population parameter;**

**Probability**

**Given a random experiment (say rolling a die) a probability measure is a population quantity (an intrinsic property of the die) that summarizes the randomness. When we talk about probability we aren’t taking about something in the data we have, but as a conceptual thing that exists in the population that we would like to estimate; P(A||B)=P(A)+P(B)-P(A&&B);**

**Densities and mass functions for random variables are all we will need for probabilities of numeric outcomes of experiments; The density is not a statement about what occurs in the data, you’re talking about population quantities; We will collect data that will be used to estimate properties of the population; discrete probabilities: we will assign a probability to each value it can take; continuous probabilities: we will assign a probability to the ranges that it can take; The web site traffic on a given day is a discrete random variable, but it is interesting because there is no upper bound. We will likely use the Poisson distribution to model that;**

**A probability mass function evaluated at a value corresponds to the probability that a random variable takes that value; To be valid a pmf function must satisfy that it’s greater than zero and the sum of all possible values add to one; Binomial pmf: used for flipping coins. Poisson: used for modeling counts. Bernoulli: flip one coin;**

**EX: X = 0 represents tails and X=1 represents heads; upper case represents a potentially unrealized value of the random variable; so we can talk about the probability that capital X equals zero or one;**

**p(x) = (1/2)x (1/2)1-x for x = 0,1; p(x) = (θ)x (1- θ)1-x for x = 0,1; data can be used to estimate population proportion theta; An example is the probability that someone has hypertension or not;**

**A probability density function evaluates the probability in a range of values for continuous variables; All areas must be at least probability zero, and the total area under it must be one;**

**Beta continuous distribution is a triangle with width 1 and height 2, with it’s highest point being at point (1,2); The area under the curve can be found by the triangle area function; Also you can use pbeta(0.75, 2, 1) to find the area under the curve; The “p” stands for probability from a beta density of being less than 0.75; This “p” gives a result similar to the cumulative distribution function; “p” functions can take arrays as the first argument;**

**The cumulative distribution function of a random variable X returns the probability that the random variable is less than or equal to the value x; F(x) = P(X<=x); works for discrete and continuous; The survival function of a random variable X is defined as the probability that the random variable is greater than the value x; S(x) = P(X>x); S(x) = 1 – F(x); the cdf is found by integrating the pdf;**

**The quantiles align the observations from least to greatest and it finds the score such that 95% score worse and 5% score better than you; The αth quantile of a distribution function F is the point xα so that F(xα)= α; median is the 50th percentile; nearly always just use the common distributions to find quantiles; qbeta(-.5, 2, 1) gives the quantile at the 50th quantile; “q” stands for quantiles;**

**We treat the median of our population as the median of our population, and certain assumptions are made; The sample median is the estimator and the population median is the estimand; every estimator has an associated estimand, and this is the formal process of statistical inference, linking your sample to a population;**

**conditional probability P(B|A) = P(A&&B) / P(A); If you can count the outcomes and they are all equally likely then this next formula is correct: P(B|A) = (# of outcomes in A && B) / (# of outcomes in event A);**

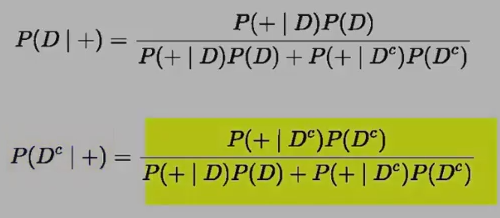
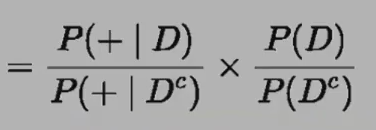
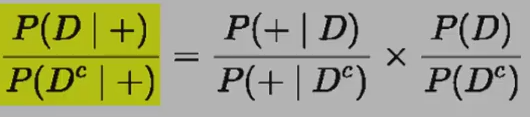
**Joint probability P(A&&B) = P(A)P(B|A) = P(B)P(A|B); If you can count the outcomes and they are all equally likely then this next formula is correct: P(A&&B) = (# of outcomes in A && B) / (Total # of outcomes);**

**Mutually exclusive events can’t occur at the same time; P(B|A) = 0; P(A&&B) = 0; P(B||A) = P(A) + P(B);**

**Independent events don’t affect each other’s probabilities; P(B|A) = P(B); P(A&&B) = P(A)P(B); P(A||B) = P(A) + P(B) + P(A)P(B);**

**Bayes’ rule finds P(B|A) given P(A|B); P(B|A) = P(A|B)P(B) / ( P(A|B)P(B) + P(A|Bc)P(B c) );**

**D: has disease; Sensitivity = P(+|D); Specificity = P(-|D c); Want both of these high for the test to be good; P(D|+) found by Bayes’ rule is the positive predictive value; For our example P(+|D) = .997 and P(-|D c) = .985; Given a positive test result, here are the probabilities of having the disease or not having the disease, only the numerator changes; we will divide these two results to work with the likelihood ratios;**

****

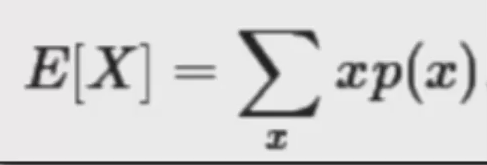
**If you take the probability and divide it by one minus the probability you get the odds; So we have the odds of the disease given a positive test result; The post-test odds of disease are equal to the diagnostic likelihood ratio times the pre-test odds of disease; post-test odds of D = DLR+ \* pre-test odds of D; For our example, DLR+ = .997/(1-.985)= about 66, DLR- = (1 - .997) / .985 = about .003**

**If a positive test, whatever your pre-test odds are, you multiply them by 66 to get your post-test odds; If a negative test, you multiply the pre-test odds by .003 to get your post-test odds;**

**We will assume the collection of random variables used are independent and from the same distribution;**

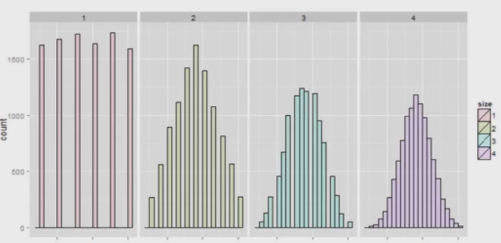
**Expected Values**

**The density and mass functions have characteristics and these are also characteristics of the random variables drawn from them. One of the most useful characterizations is called expected values, another type is quantiles covered above; The mean is the most useful of the expected values; The mean describes the center of the distribution and the variance and standard deviation tell us how spread out it is; Our sample mean will be an estimation of our population mean; Our sample variance will be an estimation of our population variance; Our sample standard deviation will be an estimation of our population standard deviation; The expected value is the mean;**

** We multiply x times the probability of each value;**

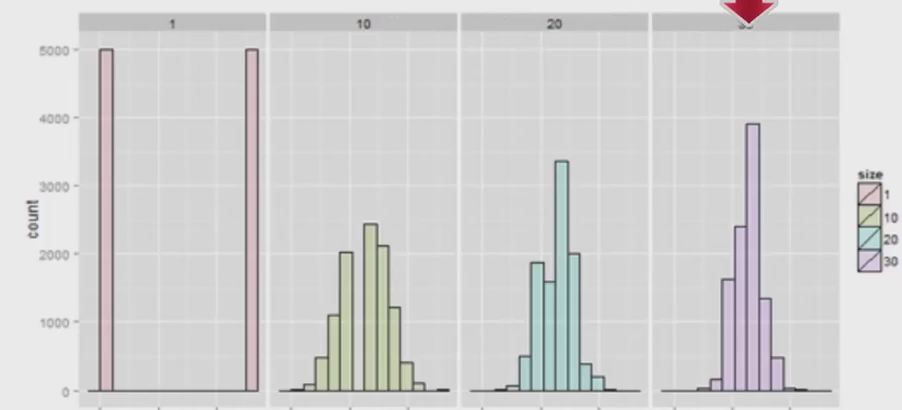
**The sample mean is the center of mass if we treat each sample point as equally likely;**

**The mean for a continuous rv. Is still the center of mass;**

**The average of random variables is itself a random variable; ie: Repeatedly rolling 6 dice and finding their average; Because it is a random variable it also has a distribution and that distribution has characteristics like an expected value and standard deviation; the center of this distribution is the same as the original distribution; This is useful when finding the population mean from the sample mean because they should be about the same (the sample mean is unbiased because its distribution is centered at what it’s trying to estimate);**

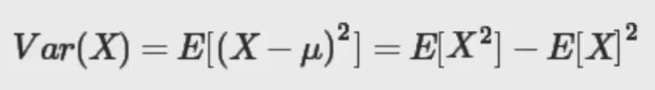
**This is a distribution of die rolls and averages of die rolls; the four figures are the averages of 1, 2, 3, and 4 die rolls;**

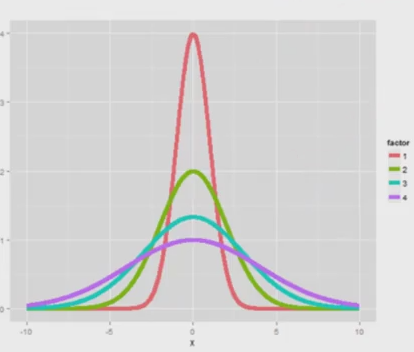
**With multiple rolls it becomes more Gaussian; notice that all four are centered at the same location;**

**This is the same but using flips of the coin one, ten, twenty, or thirty times; As the average is comprised of more coin flips, its distribution gets more concentrated around the mean, however the distribution is always centered at the same place;**

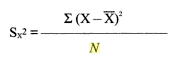
**Variance**

**The variance is how fat or how thin the distribution is around the mean; variance is exactly the expected square distance that the random variable is from the mean; A shortcut is to use the expected value of the random variable squared minus the square of the expected value of the random variable, as shown below; Below is the population variance;**

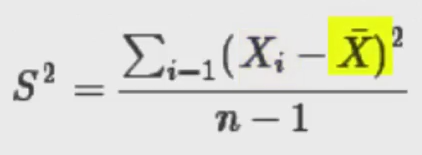
****

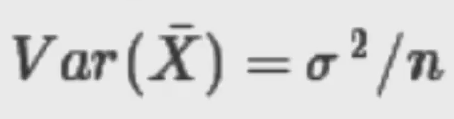
**Higher variance creates a more wide distribution;**

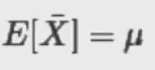
**The sample variance is:**

****

**The estimated population variance is:**

****

**The sample variance itself is a random variable, and the expected value is centered at the population variance; As you collect more data, the distribution of the sample variance will get more concentrated around the population variance it is trying to estimate; We use n-1 instead of n because the sample variance tends to underestimate the population variance; n-1 is called the degrees of freedom;**

****

**The average of the numbers that are sampled from a population (X bar) itself is a random variance, it has its own population mean and population variance; The variance of the sample mean decreases to zero as n accumulates more data; The sample variance S squared, estimates the population variance, sigma squared; standard deviation is the standard error; The variance of the sample mean is σ2 / n, the logical estimate is s2 / n, the logical estimate of the standard error is S / sqrt(n);**

**Sample standard error of the mean:**

****

**S, the standard deviation, is how variable the population is;**

**S / sqrt(n), the standard error, is how variable averages of random samples of size n from the population are;**

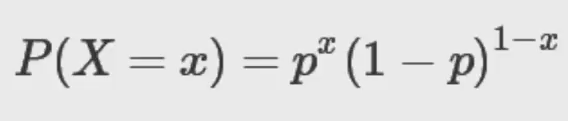
**Useful functions for simulations**

**sample(0 : 1, n, replace =TRUE) is for coin flips; rnorm(n) for normal distribution; runif(n) for uniform; rpois(n, lambda) is for Poisson distribution; sd(apply(matrix(sample(0:1, numsim\*n, replace=TRUE), numsim), 1, mean)) will simulate “numsim” simulations with n coinflips each and find the standard deviation of the means; choose(8, 7) will do 8 choose 7; rbinom for binomial;**

**var(x) the variance; var(x) / n the variance of the averages; sd(x) is the standard deviation; sd(x) / sqrt(n) is the standard deviation of the averages; The n refers to the number of observations each average is comprised of, not the “numsim”;**

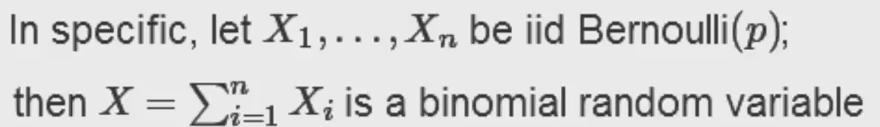
**Distributions**

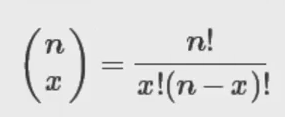
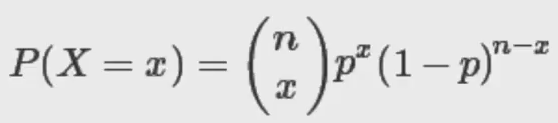
**Bernoulli arises as a result of a binary outcome like a coin flip; values 1 or 0, probabilities of p and 1-p;**

****

**mean: p; variance: p(1-p)**

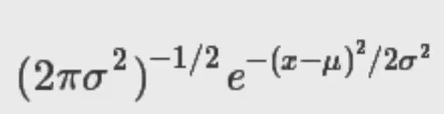
**binomial is the number of heads from flipping n coins with probabilities p and 1-p;**

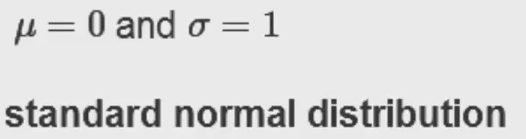
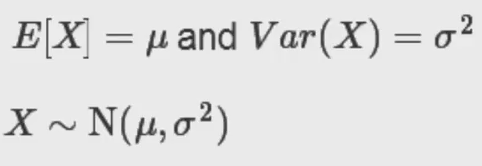
****

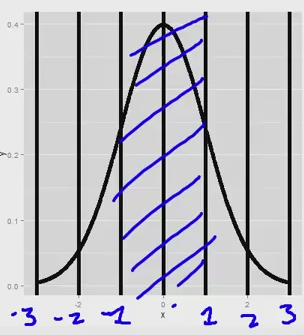
** **

**n choose x counts the number of ways of selecting x items out of n without replacement disregarding the order of the items;**

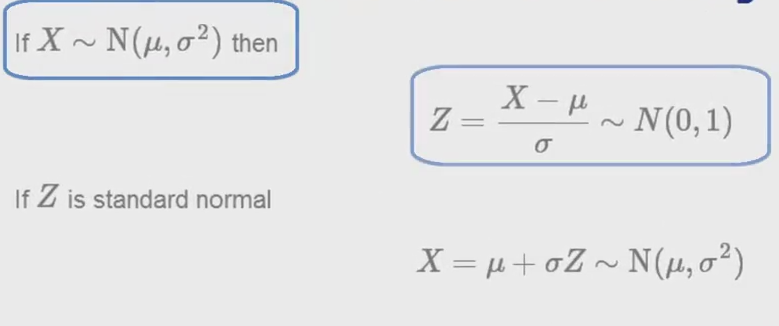
**Normal distribution is the most important distribution; it has mean μ and variance σ2 if it’s associated density is formula is:**

****

****

**All normal are the same in respect to their probabilities in terms of mu plus or minus sigma; What changes is the numbers on the x axis**

**How to move from standard normal to normal or vise versa:**

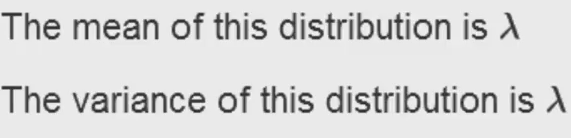
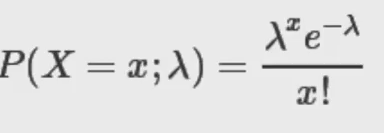
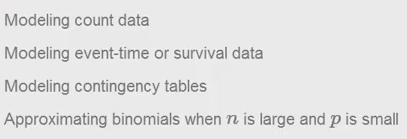
****

**Approximately 68%, 95%, and 99% of the normal density lies within 1, 2 and 3 standard deviations;**

**-1.28, -1.645, -1.96, and -2.33 are the 10th, 5th, 2.5th, and 1st percentiles of the standard normal;**

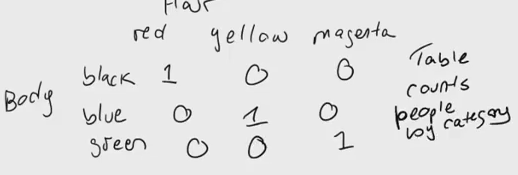
**qnorm(.95, mean=, sd=); pnorm(1.96, mean=, sd=)**

**The Poisson distribution is probably the second most useful distribution; it is used to model counts; the density formula is below, and x is a non-negative integer;**

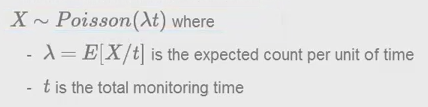
****

**Especially useful if count data is unbounded**

**Modeling the time before reoccurrence of symptoms**

**an example contingency table is shown below;**

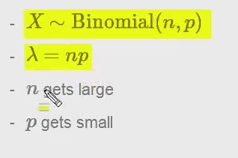
**it is a count by categories;**

**Poisson is used to model rates; **

**Example: The # of people who show up at a bus stop is Poisson with mean 2.5; If watching the bus stop for 4 hours, what is the probability that 3 or fewer people show up the whole time;**

**Ppois(3, lambda = 2.5 \* 4)**

**Poisson is an approximation of the binomial when ni s large and p is small;**

****

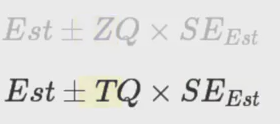
**Asymptotes is the term for the behavior when the sample size limits to infinity**

**As your same size approaches infinity you get the correct mean and variance. As your collect more observations of averages your distribution of averages becomes like the normal distribution. Remember when working with averages we use the standard error, sigma divided by square root n. If we take in enough observations of averages we know that the true average will fall between Xbar plus or minus 2\*sigma/sqrt(n) with 95% confidence.**

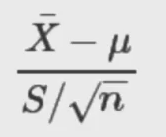
**For binomial confidence intervals working with p hat, a proof shows that the variance is greatest at p=.5. For p=.5 we know that the 95% confidence interval is 1/sqrt(n) so a pretty good confidence interval for p hat is p plus or minus 1/sqrt(n). binom.test(num\_heads,coin\_flips)$conf.int also works well, it doesn’t rely on the central limit formula.**

**If you don’t have a lot of observations for a binomial averaging experiment, adding two successes and two failures helps.**

**Poisson.test(x, T = 94.32)$conf**

**T quantile can be used with small samples**

**If you have a choice between the T interval and Z interval, use the T interval, because when you collect more data it acts like the Z interval anyway. The Z interval has two parameters, the mean and the variance; but the T distribution is only really talked about as it is centered around zero, with the standard normal as the scale, it’s only indexed by one parameter, the degrees of freedom. As these degrees of freedom increase it becomes more like a standard normal.**

**is in fact not Gaussian distributed. If we replace S with sigma it would be exactly standard normal, however when there is an s it doesn’t have a distribution of a standard normal; instead it has a T distribution. As n increases this distinction is irrelevant. If you use the normal distribution with small sample sizes you’ll get confidence intervals that are too narrow.**

**The interval is** 

