***COURSERA: STATS W/ R SPECIALIZATION***

***COURSE 1 - Introduction to Probability and Data***

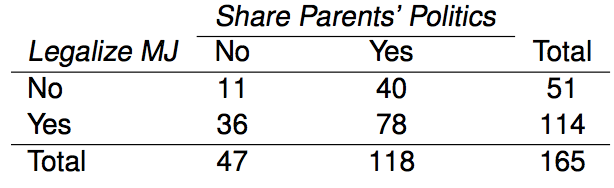
**WEEK 3 - Introduction to Probability**

***Defining Probability***

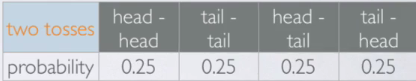
* **Random process** 🡪 know what outcomes *could* happen but not which particular one *will* happen
* Ex: coin toss, die roll, shuffle on music player, stock market
* Sometime helpful to model process as random even if it truly isn’t so
* P(A) = probability of event A
* Several interpretations of probability, but almost all agree on mathematical rule: 0 <= P(A) <= 1
* **Frequentist interpretation =** a **relative frequency =** proportion of times an outcome would occur if we ran the process an infinity number of times
* **Bayesian interpretation** = **subject degree of belief** = for same event, 2 people could have different viewpoints + as such assign different probabilities to it
* Allows for prior info to be integrated into the inferential framework
* Largely popularized by revolutionary advances in computation technology + methods in past 20 years
* **Law of Large Numbers =** as more observations are collected, proportion of occurrences of a particular outcome converges to the probability of that outcome (1/2 for coin toss, 1/6 for die roll)
* More surprising to see 3 heads in 100 coin flips compared to 10, + even more so for 1k flips
* **Independence** 🡪 coin toss P(H on toss 10) = P(H on toss 11) 🡪 coin is not **due** a heads
* Common misunderstanding of law of large numbers = **Law of Averages (Gambler’s fallacy:** random processes are supposed to compensate for what happened in the past (I’m due a good roll/hand/spin, etc.)
* But say you get 100’s of heads in a row on a coin flip, coin is most likely not fair

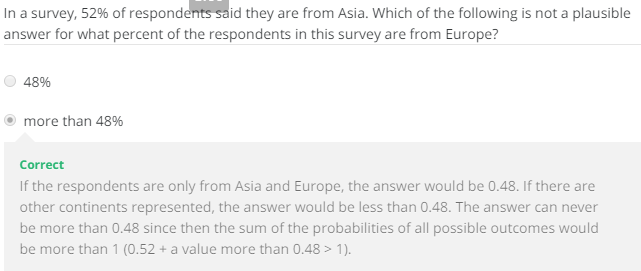
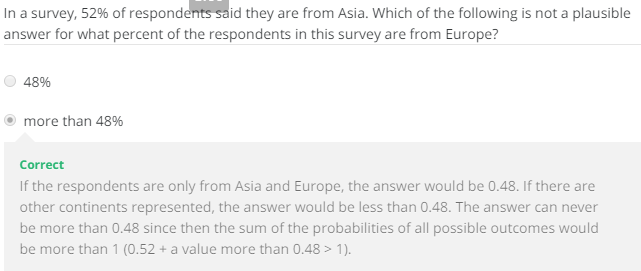
**Disjoint Events + General Addition Rule**

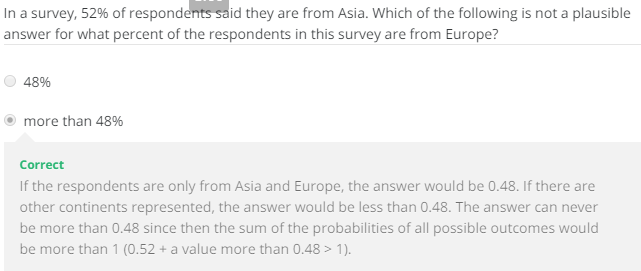
* **Disjoint events =** mutually exclusive = both events cannot happen (coin toss cannot be H AND T)
* P(A and B) = 0 🡺 P(A U B) = 0
* **Non-disjoint events** CAN happen at the same time (get an A in stats + in economics
* P(A and B) != 0 🡺 P(A U B) != 0 🡪 some # between 0 and 1
* **Union** of DISJOINT events 🡪 Probability of one event happening or the other happening?
* Drawing a Jack or a 3 from a deck? 🡪 P(J or 3) = P(J) + P(3) = 4/52 + 4/52 = 2/13
* So for disjoint events, **P(A OR B) = P(A) + P(B)**
* **Union** of NON-disjoint events:
* Probability of drawing Jack or red (could red jack) 🡪 **P(J or Red) = P(J) + P(Red) – P(J U Red)**  = 4/52 + 26/52 – 2/52 = 28/52 = 14/26 = 7/13
* So for NON-disjoint events, **P(A OR B) = P(A) + P(B) – P(A U B)**
* *Note that they can be the same formula b/c for disjoint events P(A U B) = 0*
* 



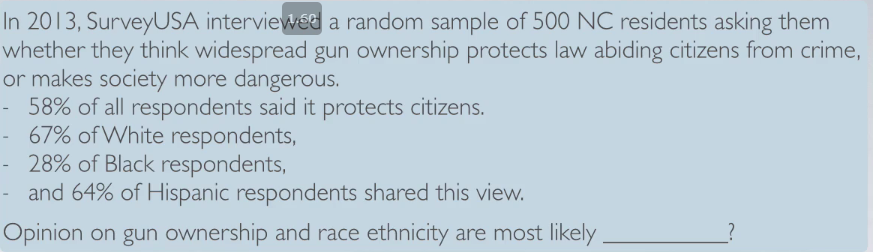
* NON-disjoint 🡪 P(M) + P(P) – P(M U P)
* 114/165 + 118/165 – 78/165 = **(114 + 118 – 78) / 156**
* The above encapsulates the **General Addition Rule**
* **Sample space (SS) =** all possible outcomes of a trial/experiment
* Couple has 2 kids 🡪 SS of genders = **S = {MM, FF, MF, FM}**
* **Probability distribution** = lists out all possible outcomes in a SS + probabilities w/ which they occur

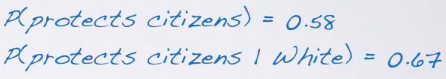
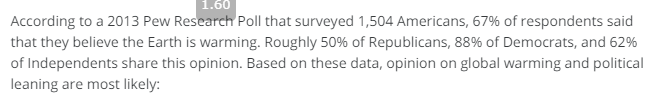
* Can create similar probability distribution for any *discrete* event of interest
* 3 Broad Rules of Probability Distributions
* Events listed must be disjoint
* Each P is between 0-1
* Sum of all probabilities w/in the distribution sum up to 1
* **Complementary events** = 2 disjoint events whose P’s sum to 1 🡪 P(A) + P(A(c)) = 1
* Coin 🡪 P(H(c)) = P(T) b/c tail is the complement of head
* For 2 coins, P(HH(c)) = sum of P(HT) + P(TH) + P(TT) 🡪 sum of all other events to 1st event so that the sum of all events equals 1
* Dividing the SS into 2 such that the sum of the 2 probabilities equals 1 (even though there’s 4 total)
* *Disjoint + Complementary are NOT the same* 🡪 not all disjoint events sum up to 1 (more than 2 events in a SS), while all complementary events sum up to 1
* So, complementary events are always disjoint, while disjoint events are not always complementary
* 

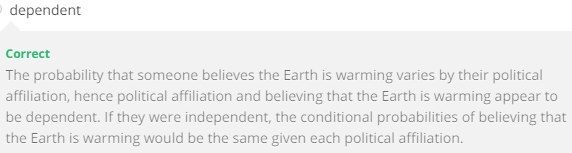


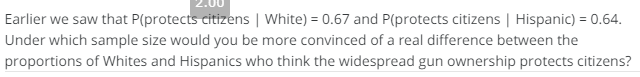
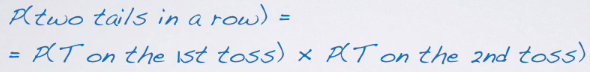
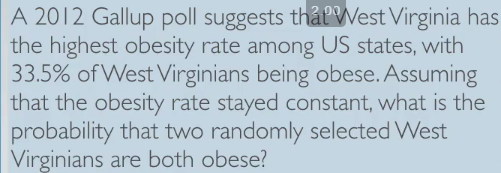
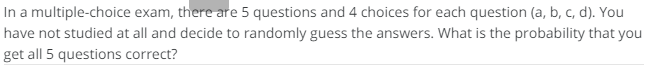
**Independence**

* 2 processes are **independent** if knowing the outcome of 1 provides no useful info about the outcome of the other
* Knowing a coin lands on H does nothing to help us know what they 2nd coin toss will land on
* But knowing 1st card drawn from a deck helps us update P(any other card), if w/out replacement
* Checking for independence rule:
* If probability of Event A happening given Event B happened is the same as the original probability of event A, they are independent 🡪 **P(A | B) = P(A), then** A + B are independent
* Knowing B tells us nothing about A
* 



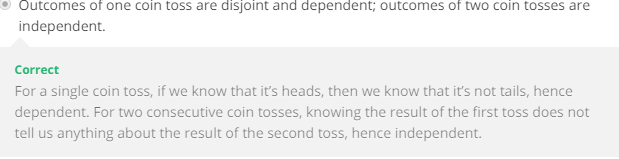
* This is b/c knowing P(protects) = 0.58, but knowing their white gives P(protects | w) = 0.67
* 
* The same goes for Blacks and Hispanics, so the opinions on gun ownership vary greatly by race, so they can be thought to most likely be **dependent**
* Knowing someone’s race may give us helpful info about their opinion on gun ownership
* 



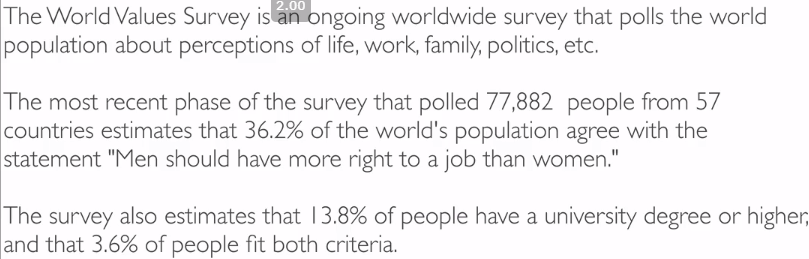
* Using phrase “most likely” b/c this is all *sample data* + we’re not using inference tools yet
* If we observe a difference between **conditional probability** based on the sample, we can say “the data *suggest* dependence”
* Then the next step would be to conduct a hypothesis test to see if this observed difference was due to change/natural random sampling, or if there’s some real effect/difference in the population
* Can do some speculation based on the size of the observed difference as well as based on the sample size
* If difference is large/varies greatly = stronger evidence difference is real
* If sample size if large = even SMALL difference in conditional probabilities provides strong evidence difference is real
* 
* 
* Product Rule for Independence
* **If A + B are independent, P(A U B) = P(A) \* P(B)**
*  **= 0.5 \* 0.5 = 0.25**
* Can be expanded to as many independent events as we have
* 
* 
* Given P(obese) = 0.335
* Given 2 individuals are randomly selected 🡺 they are therefore *independent*
* So, P(obese1 U Pobese2) = P(obese1) \* P(obese2) = P(o) + P(o) = 0.335^2 = 0.11 = 11% chance 2 randomly selected West Virginians are obese
* 

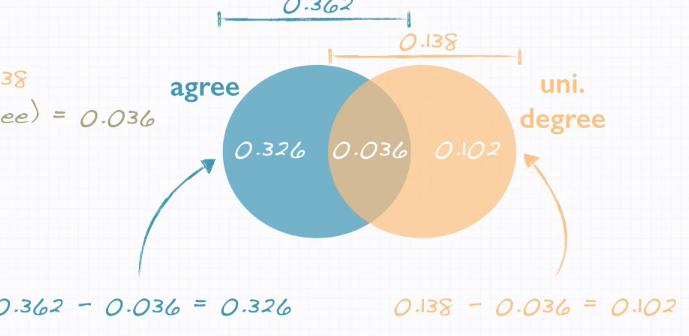
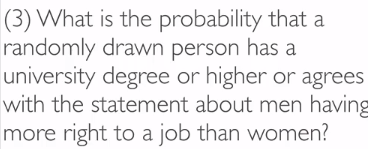
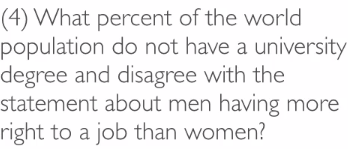


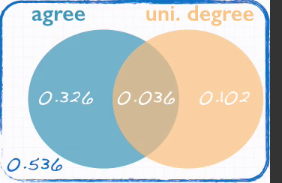
**Spotlight: Disjoint vs. Independent**

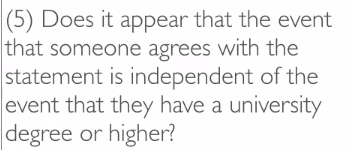
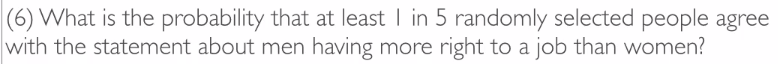
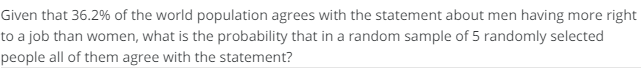
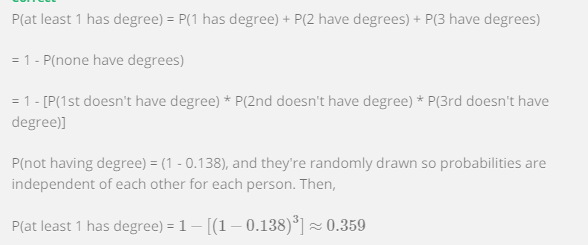
* **Disjoint** = mutually exclusive = both cannot happen at same time = P(A U B) = 0
* **Independent** = knowing outcome of 1 event gives no knowledge of outcome of another = P(A | B) = P(A) if independent
* Ex: Babies have 3 possible eye colors, Bl, Gr, Br 🡪 disjoint (cannot happen at same time)
* 2 babies 🡪 now 1st has Bl eyes 🡪 also disjoint
* if not related, gives no knowledge of outcome for 2nd baby = independent 🡪
* if related, may give some knowledge = dependent (may be more likely to have bl eyes
* Also note that outcomes for eye color for 1 baby are dependent 🡪 knowing a baby’s eye color to be blue means we have knowledge they aren’t brown or green
* Can generalize this to say Disjoint events w/ non-zero probability are always dependent on each other (if we know 1 happened we know the other cannot)
* 

**Probability Examples**



* What do we know?
* P(agree) = 0.362 P(degree) = 0.138 P(A U D) = .036
* 
* To be disjoint P(A U D) should = 0, and it does not 🡪 NOT disjoint
* 
* 32.6% agree w/ no degree, 10% don’t agree w/ a degree
* 
* P(D) = .138, P(A) = .362 🡪 P(D) OR P(A) for NON-disjoint events **= P(A) + P(B) – P(A U B)** from **general addition rule/additivity axiom**
* 0.138 + 0.362 - 0.036 = 0.464 🡪 **46.4%** of randomly drawing a person w/ a degree or a person who agrees w/ the statement
*  🡺 P(D(c)) = .862, P(A(c)) = .638 🡪 P(no degree nor agree) 🡪 complement of P(A or D) 🡪 1 - .464 = .536 🡪 area in Venn diagram outside both circles

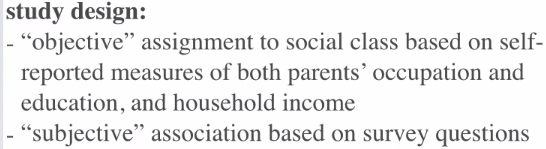


*  🡺 If A + B are independent, P(A U B) = P(A) \* P(B)
* This is NOT true so they are NOT likely to be independent
* 
* P(A) = .362, SS = {0,1,2,3,4,5} where it’s possible none agree, 1 agrees, 2 agrees, up to all agree
* Only care if at least 1 does, so divide SS into {0, >= 1}
* Subtract complement of at least 1 agrees = none agree 🡪 P(>1 agree) = 1 – P(none) 🡪 P(Dis, Dis, Dis, Dis, Dis)
* P(Dis) = .638 (the complement of agreeing(, and these are independent, so .638^5 = 0.1057
* So we have 1 - 0.1057 = **.8943 = 89.43% at least 1 in 5 randomly selected people agree**
* 
* P(A) = .362, SS = {0,1,2,3,4,5} 🡪 split SS = {all agree, >= 1 disagree} 🡪 P(all) = P(A)^5 b/c they’re independent = .362^5 = **.006 = 0.6% chance all of 5 randomly sampled persons agree**
* 
* P(Deg) = .138, SS = {0,1,2,3} 🡪 split into SS = {none, at least 1}
* P(>= 1) = 1 - P(none) 🡪 1 – P(No,No,No) 🡪 P(No) = 1 – P(Deg) = 1 - .138 = .862
* P(no)^3 = .862^3 = .641 🡪 1 - .641 = **.359**
* 

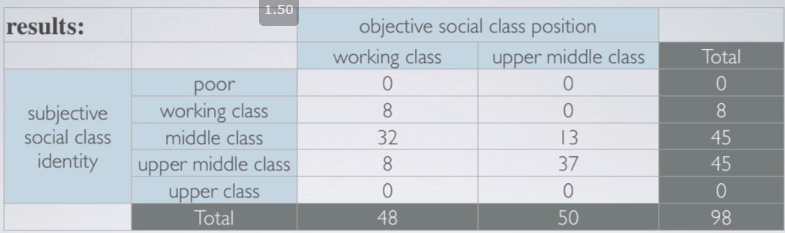
**Conditional Probabilty**

* 

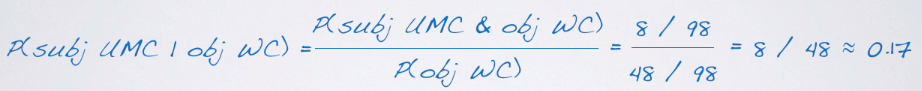


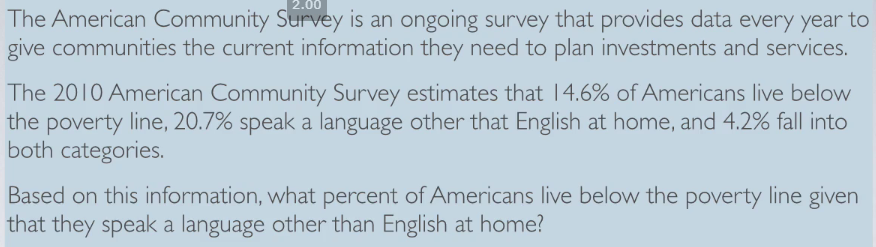


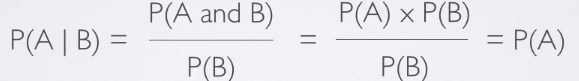
* The social class here is what we’re taking to be the “truth”
* Given 2 categorical variables, we can create a **contingency table** of objective + subjective social class

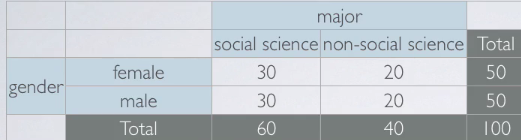


* **Marginal Probability** 🡪 counts used to calculate P() come from the **margins** of the contingency table 🡪 bottom row, last col (the TOTAL row + TOTAL col)
* P(objective = upper-middle class) 🡺 P(obj UMC)
* Look at objective upper-middle class column 🡪 see 50 students in this category out of 98 total = 50/98 = ~1/2 = ~51%
* P(sub UMC) = see 45 students in this row out of 98 total = 45/98 = .46
* **Join Probability** 🡪 P() at intersection of 2 events of interest (intersection of Venn diagram)
* P(sub UMC &obj UMC) 🡪 look for cell intersection of specified row + col 🡪 37/98 = .38
* P(sub WC & obj WC) = 8/98 = .08
* **Conditional Probability**
* P() a student objectively in WC associates w/ UMC 🡪 P() they believe they’re UMC given that they’re WC 🡪 **P(sub UMC | obj WC) 🡪 8 WC who think UMC /48 WC = .17**
* P(sub MC | obj UMC) = 13/50 = .26
* Calculate conditional probabilities w/ Bayes’ Theorem 🡪 **P(A | B) = P(A & B) / P(B)**

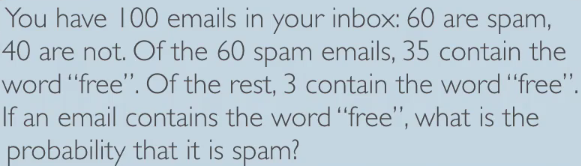
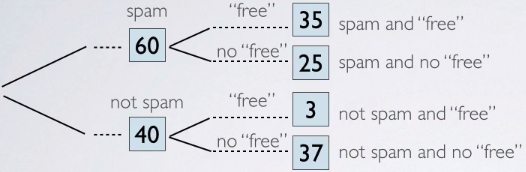
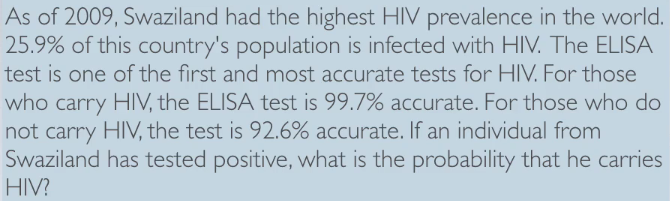
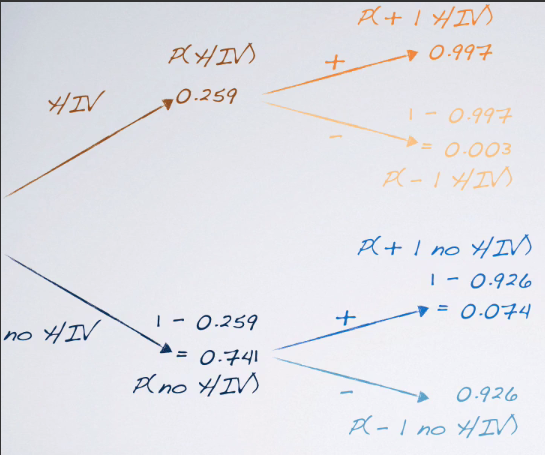
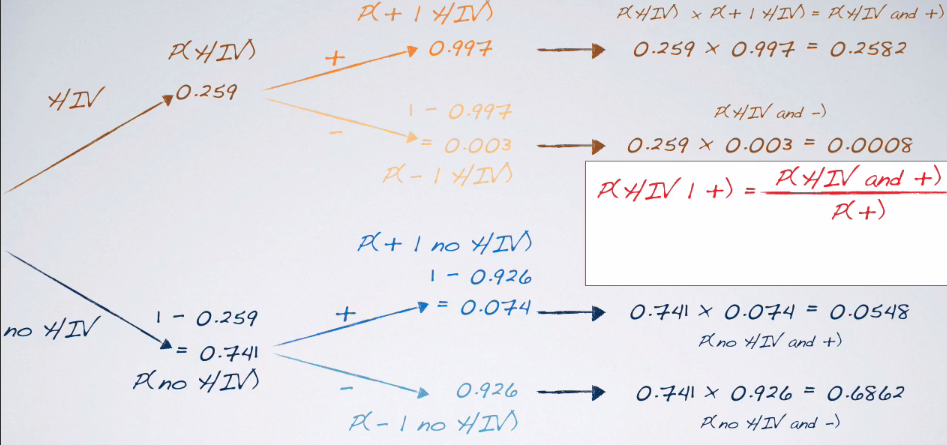


* 
* P(Below | Non-English) = P(Below & Non-English) / P(Non-English) **= 4.2 / 20.7 = ~.2**
* Can use this info to compare to the general public.
* Since we know 14.6% live below the poverty line, it seems that living below the poverty line is more prevalent for those who speak a language other than English at home
* This suggest language spoken at home and living below poverty level MAY be dependent
* P(non-English | below) = P(below & non-English) / P(below) = .042 / .146 = .29
* Remember Product Rule for Independent Events
* 
* If we believe the events are NOT independent, or we cannot check if they are, the joint probability must be calculated differently
* Bayes’s DOESN’T have a independence condition 🡪 rearrange it 🡪 P(A & B) = P(A | B) \* P(B)
* This is **general product rule**
* Generally, if P(A|B) = P(A), events A + B are said to be independent 🡪 knowing event B occurred did nothing to change the probability of A occurring
* Mathematically 🡪 if they’re independent, P(A & B) = P(A) \* P(B)



* 
* P(SS) = 60/100 = .6
* Find P(SS) if we know the randomly sampled student is a female = P(SS | G) = P(SS & F) / P(F) = (30/100) / (50/100) = 30 / 50 = .6
* Also note that **P(SS) = P(SS | M) = P(SS | F) = .6** 🡪 *all are the same P() 🡪 if we know P(A | B) = P(A), then the events are independent*
* Here, P(SS) = P(SS | either gender) 🡪 can determine gender and major are independent

**Probability Trees**

* These are helpful for solving conditional probabilities, especially when the probability we’re asked for is the opposite of what we’re given: P(A|B) 🡪 P(B|A)
* 
* P(S) = .6, P(NS) = .4, P(F|S) = 35/60, P(F|NS) = 3/40
* 
* P(S|F) = P(S & F) / P(F) 🡪 35 / (35 + 3) = 35/38 = .92
* i.e. Bayes’ 🡪 numerator = joint, denominator = marginal of what to condition on
* Not to do the same w/ probabilities and not actual counts (don’t know sample size or populations)
* 
* P(P|H) = .997, P(N|NH) = .926, P(H) = .259
* P(P|H) = P(P & H) / .259
* P(H|P) = P(H & P) / P(P) 🡪 reverse of what we’ve given
* 1st branch = always made up of marginal probabilities b/c we’re splitting up the population w/out conditioning based on any other attributes 🡪 have HIV or not
* 
* 
*  🡺 
* These events are dependent, so to get the joint probabilities (i.e. P(H & P)) we multiply them together in a new set of branches
* 
* P(H|P) = P(H & P) / P(P)
* P(P) is made up of both sets of (+) joint probabilities 🡪 P(H+ OR NH+) 🡪 disjoint 🡪 add them
* .2582 / (.2582 + .0548) = .82