Philosophy 12: Introduction to Causal Reasoning
Answers to study questions for Lecture 7: "Independence"

1. Answer(s): (a)

   Remember that there are 52 cards and 13 diamonds. So, we should draw a diamond $\frac{13}{52}$ of the time. $\frac{13}{52} = \frac{1}{4} = 0.25$.

2. Answer(s): (a)

   There are four aces and only one is a diamond. So, the chances that we'll draw the ace of diamonds is $\frac{1}{4} = 0.25$.

3. Answer(s): (a)

   Remember that choosing an ace will be independent of choosing a diamond if Fr(Diamond) = Fr(Diamond | Ace). In the previous two questions, you computed Fr(Diamond) and Fr(*Diamond* | Ace). Fr(Diamond) = 0.25 and Fr(Diamond | Ace) = 0.25. These are equal so choosing an ace is independence of choosing a diamond.

4. Answer(s): (c)

   There are 26 red cards, but only 13 are diamonds. So the chance that the card is a diamond, knowing that the card is red, is $\frac{13}{26} = \frac{1}{2} = 0.5$

5. Answer(s): (b)

   Remember that red cards and diamonds will be independent only if Fr(Diamond) = Fr(Diamond|Red card). In the previous question, we computed Fr(Diamond) and Fr(Red card). Fr(Diamond) = 0.25 but Fr(Diamond | Red card) = 0.5. These are not equal, so choosing a diamond is not independent of (i.e., it is associated with) choosing a red card.

6. Answer(s): (a), (c)

   Remember that being male and having blond hair are independent properties if

   $$\text{Fr}(\textsc{Sex} = \text{Male}) = \text{Fr}(\textsc{Sex} = \text{Male} \mid \textsc{Hair Color} = \text{Blond}.$$

   Thus, we need one histogram that shows us Fr(Sex = Male), which is histogram A, and one which shows us Fr(Sex = Male | Hair Color = Blond), which is histogram C.
   The reason B is not correct is that Fr(Blond is only relevant information if we can also find out Fr(Blond | Male). But we don't have the histogram we need to get that information.
   The reason D is not correct is that learning the frequency of smokers among the blond individuals doesn't help us determine if being male and being blond are independent.

7. Answer(s): (b)

   Remember that the properties being male and smoking are independent if Fr(Sex = Male = Fr(Sex = Male | Smoker? = Yes). From these histograms we can tell that Fr(Sex = Male) ≠ Fr(Sex = Male | Smoker? = Yes), thus being male is dependent on, or associated with, smoking.

8. Answer(s): (c)

   Being male and being blond are independent if Fr(Sex = Male) = Fr(Sex = Male | Hair Color = Blond). But these histograms only provide Fr(Sex = Male) and Fr(Hair Color = Blond); thus, there is not enough information in these histograms to tell whether being male and blond are independent.

9. Answer(s): (a), (d)

   Remember that A is independent of B if $\text{Fr}(A) = \text{Fr}(A \mid B)$, so to determing whether being male is independent of being over six feet tall, we need to know whether Fr(Sex = Male) is equal to Fr(Sex = Male | Height = 6+ feet).

10. Answer(s): (b)

    Remember that the frequency of males is $\text{Fr}(\textsc{Sex} = \text{Male})$ and is the total number of males, divided by the total number of individuals in the population.

11. Answer(s): (d)

    Remember that the frequency of males among the people over six feet tall is $\text{Fr}(\text{Male} \mid 6+ \text{ feet})$ and is the number of males within the subpopulation of people over six feet tall.

12. Answer(s): (b)

    In the last two questions, we computed $\text{Fr}(\text{Male})$ and $\text{Fr}(\text{Male} \mid 6+ \text{ feet})$. Remember that being male is independent of being over six feet tall only if

    $$\text{Fr}(\textsc{Sex} = \text{Male}) = \text{Fr}(\textsc{Sex} = \text{Male} \mid \textsc{Height} = 6+ \text{ feet})$$

    $\text{Fr}(\textsc{Sex} = \text{Male}) = \frac{1}{2}$, but $\text{Fr}(\textsc{Sex} = \text{Male} \mid \textsc{Height} = 6+ \text{ feet}) = \frac{5}{8}$, so they are not independent.

13. Answer(s): (a), (b) , (d)

    The reason A is correct is that $A \perp\!\!\!\perp B$ if and only if $\text{Fr}(A) = \text{Fr}(A \mid B)$. The reason C is correct is that $A \perp\!\!\!\perp B$ if and only if $\text{Fr}(A \,\&\, B) = \text{Fr}(A) \cdot \text{Fr}(B)$. The reason D is correct is that independence is symmetric, so if male is independent of blond, then blond is independent of male.

14. Answer(s): (d)

    Remember that A is independent of B if $\text{Fr}(A) = \text{Fr}(A \mid B)$. Is right-handed independent of black hair? Right-handed will be independent of black hair only if $\text{Fr}(\text{Right-handed}) = \text{Fr}(\text{Right-handed} | \textsc{Hair Color} = \text{Black})$. According to the data, both frequencies equal 0.9, so they are independent. Then, since $A \perp\!\!\!\perp B$ implies $\text{Fr}(A \,\&\, B) = \text{Fr}(A) \cdot \text{Fr}(B)$, we know that

    $\text{Fr}(\text{Right-handed} \,\&\, \textsc{Hair Color} = \text{Black}) = \text{Fr}(\text{Right-handed}) \cdot \text{Fr}(\textsc{Hair Color} = \text{Black}) = 0.9 \cdot 0.6 = 0.54.$

15. Answer(s): (b)

    The frequency of cheats on taxes is the total number who cheat on taxes divided by the total number in the sample.

16. Answer(s): (c)

    The frequency of tax cheating conditional on high income is the number who have the properties of "Cheats on taxes" and "High income", divided by the total number with High income.

17. Answer(s): (b)

    The frequency of tax cheating conditional on medium income is the number who have both Cheats on taxes and Medium income, divided by the total number with Medium income.

18. Answer(s): (d)

    The frequency of tax cheating conditional on low income is the number with both Cheats on taxes and Low income, divided by the number with Low income.

19. Answer(s): (b)

    The frequency of not tax cheating is the number who do not cheat on their taxes, divided by the total number of people.

20. Answer(s): (d)

    The frequency of not tax cheating conditional on high income is the number who have both Does not cheat on taxes and High Income, divided by the number who have High income.

21. Answer(s): (b)

    The frequency of not tax cheating conditional on Medium income is the number who have both Does not cheat on taxes and Medium income, divided by the number who have Medium income.

22. Answer(s): (c)

    The frequency of not tax cheating conditional on low income is the number who have both "Does no cheat on taxes" and Low income, divided by the number with Low income.

23. Answer(s): (b)

24. Answer(s): (b)

25. Answer(s): (b)

    Remember that two variables are independent if every value of one variable is independent of every value of the other variable.

26. Answer(s): (b), (e)

    The reason A is not correct is that HAIR COLOR is not a binary variable, so one value of HAIR COLOR being independent of a value of HANDED doesn't imply that any other values of HAIR COLOR are independent of that value of HANDED. The reason D is not correct is that two variables are only independent if every value of one variable is independent from every value of the other variable. Is HANDED = Right $\perp\!\!\!\perp$ HAIR COLOR = Blond? If not, then the variables are not necessarily independent.

27. Answer(s): (b)

    Strictly speaking, the answer is no, because of the following:

    - Fr(Male) = $\frac{1144}{2287} = 0.5002$
    - Fr(Male | White) = $\frac{972}{1949} = 0.4987$
    - Fr(Male | Nonwhite) = $\frac{172}{338} = 0.509$
    - Fr(Female) = $\frac{1143}{2287} = 0.4498$
    - Fr(Female | White) = $\frac{979}{1949} = 0.501$
    - Fr(Female | Nonwhite) = $\frac{166}{338} = 0.491$

    But as you can see, in a sense they are almost independent. Many times with real world data, we don't get exact independencies in the calculations. For this reason, statisticians like to define something we might call "approximate independence." Most of the different kind of statistical tests that you might hear or read about are actually ways of telling when two properties are approximately independent.