Philosophy 12: Introduction to Causal Reasoning
Answers to study questions for Lecture 6: "Conditional Relative Frequency"

1. Consider the following properties:

   - Property B: Having a beard
   - Property F: Being Female
   - Property EM: Living east of the Mississippi River

   Consider the conditional relative frequencies:

   - $\text{Fr}(B \mid F)$
   - $\text{Fr}(B \mid EM)$

   Which of the following are almost certainly true?

   [(a)] $\text{Fr}(B \mid F) < \text{Fr}(B)$

   (b) $\text{Fr}(B \mid EM) < \text{Fr}(B)$

   [(c)] $\text{Fr}(B \mid F) < \text{Fr}(B \mid EM)$

   The reason that both A and C are correct is that it is almost definitely true that the frequency of beards among females is less than the frequency of bears in the general population. However, there is no reason to think that the frequency of bears among people living east of the Mississippi is any different from the frequency of beards in the general population. Given that, the frequency of beards among females would also be less than the frequency of beards among those (both male and female) who live east of the Mississippi.

   The reason B is not correct is that there is no reason to think that the frequency of beards among people living east of the Mississippi is any different from the frequency of beards in the general population.

2. Consider the following statement about the relationship between Education and Income:

   > Overall, only 3% of Americans ever earn over $60,000/year. Among college graduates, however, the frequency is 11%.

   Let the variable INCOME range over the values { Over $60K/Year, Under $60K/Year } and the variable EDUCATION range over the values { College Graduate, Not College Graduate }. The statement above claims that:

   (a) Fr(EDUCATION = College Graduate) < Fr(EDUCATION = College Graduate & INCOME = Over $60K/Year)

   (b) Fr(INCOME = Over $60K/Year) < Fr(INCOME = Over $60K/Year) & Fr(EDUCATION = College Graduate)

   (c) Fr(EDUCATION = College Graduate) < Fr(EDUCATION = College Graduate | INCOME = Over $60K/Year)

   [(d)] Fr(INCOME = Over $60K/Year) < Fr(INCOME = Over $60K/Year | EDUCATION = College Graduate)

   The sentence "Among college graduates, however, the frequency is 11%" refers to the relative frequency of Americans who earn over $60,000/year conditional on (among) those Americans who are college graduates.

3. Consider the following description of the relationship between consuming garlic and speeding up recovery from colds:

> Approximately 48% of people surveyed said they recovered from their last cold in under a week. Among those who report eating garlic regularly, the proportion of early recovery was virtually no different (48.1%).

Let the variable COLD RECOVERY have the values { Recovers in over a week, Recovers in under a week } and the variable DIET have the values { Eats garlic regularly, Does not eat garlic regularly }. The statement above claims that:

(a) Fr(DIET = Eats garlic regularly) is approximately equal to Fr(DIET = Eats garlic regularly & COLD RECOVERY = Recovers in under a week)

(b) Fr(COLD RECOVERY = Recovers in under a week) is approximately equal to Fr(COLD RECOVERY = Recovers in under a week & DIET = Eats garlic regularly).

(c) Fr(DIET = Eats garlic regularly) is approximately equal to Fr(DIET = Eats garlic regularly | COLD RECOVERY = Recovers in under a week)

[(d)] Fr(COLD RECOVERY = Recovers in under a week) is approximately equal to Fr(COLD RECOVERY = Recovers in under a week | DIET = Eats garlic regularly).

The sentence "Among those who report eating garlic regularly, the proportion of early recovery was virtually no different (48.1%)" refers to the relative frequency of people who have recovered from their last cold in under a week conditional on (among) those who eat garlic regularly.

4. Consider the following statement about the likelihood of Sally being called for jury duty.

> Given that Sally has registered to vote, the chance that she is summoned for jury duty in 2001 is 1 in 10,000.

Let the variable REGISTERED have the values { Registered to vote, Not registered to vote } and the variable JURY DUTY have the values { Called in 2001, Not called in 2001 }. The statement above claims that:

(a) Fr(REGISTERED = Registered to vote & JURY DUTY = Called in 2001) = $\frac{1}{10{,}000}$

(b) Fr(REGISTERED = Registered to vote | JURY DUTY = Called in 2001) = $\frac{1}{10{,}000}$

[(c)] Fr(JURY DUTY = Called in 2001 | REGISTERED = Registered to vote) = $\frac{1}{10{,}000}$.

This statement refers to the frequency with which people are called for Jury Duty conditional on (among) those who are registered to vote.

5. Consider the following statement above sex and hair color:

> 52% of the undergraduates at UCSD are female, and 20% of the undergraduates are blond females.

Let the variable SEX have the values { Male, Female } and the variable HAIR COLOR have the values { Blond, Dark-haired }. The statement above claims that:

[(a)] Fr(SEX = Female) > Fr(SEX = Female & HAIR COLOR = Blond)

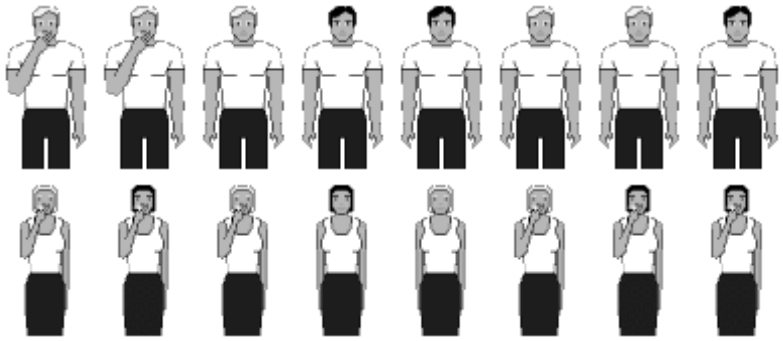
(b) Fr(HAIR COLOR = Blond) > Fr(HAIR COLOR = Blond & SEX = Female)

(c) SEX = Female) > SEX = Female | Fr(HAIR COLOR = Blond)

(d) Fr(HAIR COLOR = Blond) > Fr(HAIR COLOR = Blond | SEX = Female)

The first percentage (52%) states the relative frequency of female undergraduates at UCSD. The second part of the state, "20% of the undergraduates are blond females" refers to the relative frequency of undergraduates who are both blond and female.

Use the population below to answer questions 6 through 13:



6. Fr(SEX = Male)?

   (a) 8

   (b) $\frac{5}{16}$

   [(c)] $\frac{8}{16}$

   (d) None of the above

   Answer: Remember that the relative frequency of males is the number of males in the sample divided by the total number in the sample.

7. Fr(SMOKES = Smoker)?

   (a) 8

   (b) $\frac{5}{16}$

   [(c)] $\frac{8}{16}$

   (d) None of the above

   Answer: Remember that the relative frequency of smokers is the number of smokers in the sample divided by the total number in the sample.

8. Fr(SMOKES = Smoker | SEX = Male)?

   [(a)] $\frac{2}{8}$

   (b) $\frac{2}{16}$

   (c) 2

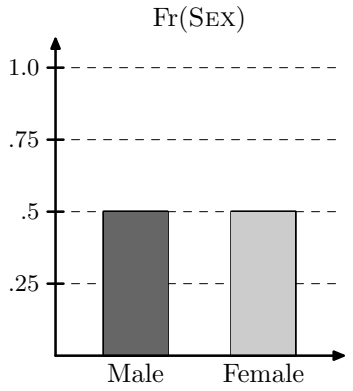
   (d) $\frac{8}{16}$

   (e) $\frac{6}{8}$

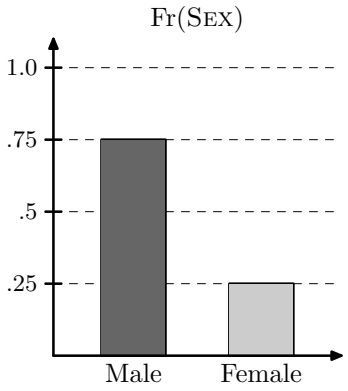
   Answer: The relative frequency of smokers conditional on SEX=Male is the number of people who both smoke and are male, divided by the total number of males.

9. Which histogram correctly represents the frequency of SEX in the sample?

(a)

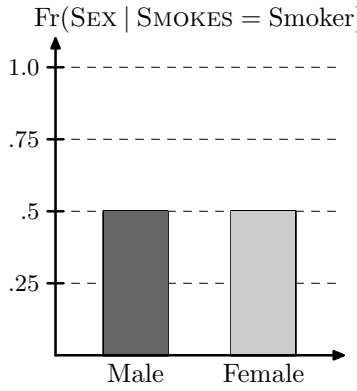


(b)

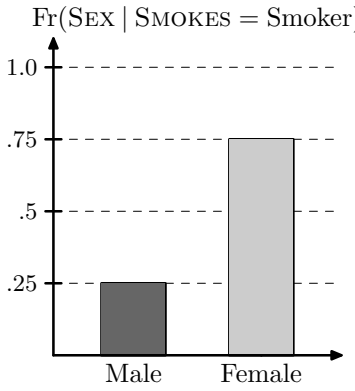


10. Which histogram correctly represents the relative frequency of SEX conditional on SMOKES = Smoker?

(a)

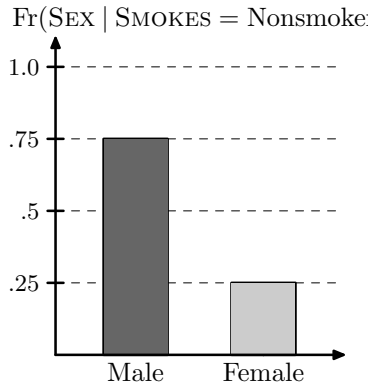


(b)

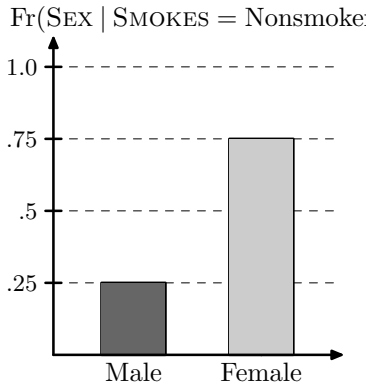


11. Which histogram correctly represents the relative frequency of SEX conditional on SMOKES = Nonsmoker?

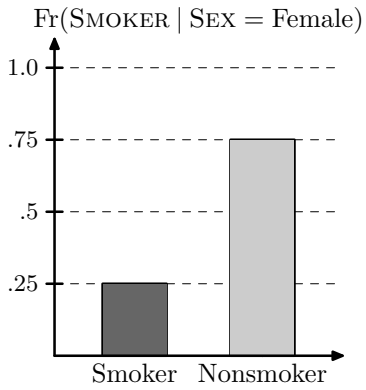
(a)



(b)



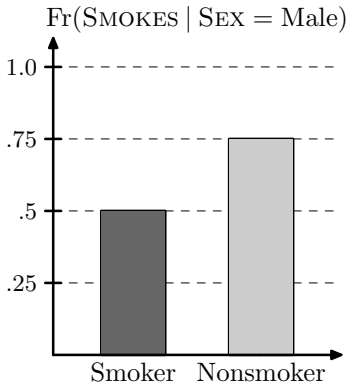
12. Does the following histogram correctly represent the relative frequency of SMOKES conditional on SEX = Female?



(a) Yes

(b) No

13. The histogram below incorrectly represents the relative frequency of SMOKES conditional on SEX = Male because



(a) The height of the bar representing Smokers is too high.

(b) The height of the bar representing Smokers is too low.

(c) The height of the bar representing Nonsmokers is too high.

(d) The height of the bar representing Nonsmokers is too low.

(e) The sum of the bar heights is too low.

(f) The sum of the bar heights is too high.

Use the following data table to answer questions 14 and 15:

Individual	AGE	POLITICAL PARTY
1	Old	Democrat
2	Old	Republican
3	Young	Republican
4	Old	Republican
5	Young	Republican
6	Young	Democrat
7	Old	Republican
8	Young	Republican

14. What is Fr(PARTY = Democrat | AGE = Old)?

    Answer: The relative frequency of PARTY = Democrat conditional on AGE = Old is the number of people in the sample who are both democrats and old, divided by the total number who are old. In this case, it is $\frac{1}{8}$.

15. What is Fr(AGE = Young | PARTY = Republican)?

    Answer: The relative frequency of AGE = Young conditional on PARTY = Republican is the number of people in the sample who are both Republican and young, divided by the total number who are Republicans. In this case, it is $\frac{3}{6} = \frac{1}{2}$.

Use the following contingency table to answer questions 16 through 18.

SEX	PH.D. PLANS=Yes	PH.D. PLANS=No	Total
Male	5	12	17
Female	6	5	11
Total	11	17	28

16. What is Fr(PH.D. PLANS = No | SEX = Female)?

    (a) $\frac{5}{6}$
    (b) $\frac{12}{17}$
    (c) $\frac{5}{17}$
    [(d)] $\frac{5}{11}$

17. What is Fr(SEX = Female | PH.D. PLANS = Yes)?

    [(a)] $\frac{6}{11}$
    (b) $\frac{5}{11}$
    (c) $\frac{5}{17}$
    (d) $\frac{6}{28}$
    (e) None of the above

18. What is Fr(PH.D. PLANS = Yes?

    (a) $\frac{5}{17}$
    (b) $\frac{6}{11}$
    [(c)] $\frac{11}{28}$
    (d) $\frac{6}{28}$
    (e) $\frac{5}{28}$
    (f) None of the above

In 1988, the Department of Highway Safety and Motor Vehicles compiled statistics on the frequency of injuries in car crashes and the frequency of seat belt use. Here is the contingency table they constructed:

SEAT BELT	INJURY=Fatal	INJURY=Non-fatal	Total
No	1601	162,527	164,128
Yes	510	412,368	412,878
Total	2111	574,895	577,006

Since the causal question at hand was whether seat belts prevent fatalities, the authors were interested in whether there was a negative association between seat belt use and fatal injuries. That is, their hypothesis was that the frequency of fatal injuries conditional on wearing a seat belt was lower than the frequency of fatal injuries in the total sample.

19. Which of the following best represents the authors' hypothesis?

(a) $\text{Fr}(\textsc{Seat Belt} = \text{Yes} \mid \textsc{Injury} = \text{Fatal}) < \text{Fr}(\textsc{Seat Belt} = \text{Yes})$

(b) $\text{Fr}(\textsc{Seat Belt} = \text{Yes} \mid \textsc{Injury} = \text{Non-fatal}) < \text{Fr}(\textsc{Seat Belt} = \text{Yes})$

(c) $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belth} = \text{Yes}) < \text{Fr}(\textsc{Injury} = \text{Fatal})$

(d) $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belth} = \text{Yes}) < \text{Fr}(\textsc{Injury} = \text{Non-fatal})$

(e) $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belth} = \text{No}) < \text{Fr}(\textsc{Injury} = \text{Fatal})$

The frequency of fatal injuries conditional on wearing a seat belt is represented by $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belt} = \text{Yes})$, and the frequency of fatal injuries in the total sample is represented by $\text{Fr}(\textsc{Injury} = \text{Fatal})$.

20. What is $\text{Fr}(\textsc{Injury} = \text{Fatal})$?

(a) $\frac{2{,}111}{577{,}006}$

(b) $\frac{1{,}601}{164{,}128}$

(c) $\frac{510}{577{,}006}$

(d) $\frac{510}{412{,}878}$

The $\text{Fr}(\textsc{Injury} = \text{Fatal})$ is given by the number of fatal injuries divided by the total number in the sample.

21. What is $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belt} = \text{Yes})$?

(a) $\frac{1{,}601}{2{,}111}$

(b) $\frac{1{,}601}{164{,}527}$

(c) $\frac{510}{2{,}111}$

(d) $\frac{510}{412{,}878}$

The $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belt} = \text{Yes})$ is given by the number of individuals with $\textsc{Injury} = \text{Fatal}$ and $\textsc{Seat Belt} = \text{Yes}$ divided by the total number of individuals with $\textsc{Seat Belt} = \text{Yes}$.

22. According to the data, is the hypothesis of the author correct?

(a) Yes

(b) No

The hypothesis was $\text{Fr}(\textsc{Injury} = \text{Fatal} \mid \textsc{Seat Belt} = \text{Yes}) < \text{Fr}(\textsc{Injury} = \text{Fatal})$.