Consider two sub-populations of a larger population $\Omega$. For example, let $\Omega = \text{adults}$, which is broken down into $M = \text{Males}$ and $F = \text{Females}$. Then consider another trait or characteristic. For example $S = \text{Smoker}$, $S' = \text{Nonsmoker}$.

We wish to see if the trait is independent or dependent on the sub-populations. To do so, we will label and compute 8 possible sub-population percentages, which are formally called *conditional probabilities*. In general, the probability of trait $C$ among just sub-population $D$ is given by $P(C \mid D) = \frac{P(C \cap D)}{P(D)}$.

As an example, consider the following data about all adults:

- 52% are female and 35.2% of all adults smoke. But only 30% of males smoke.

<table>
<thead>
<tr>
<th></th>
<th>$S = \text{Smokers}$</th>
<th>$S' = \text{Nonsmoker}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F = \text{Female}$</td>
<td>35.2% - 14.4% = 20.8%</td>
<td>31.2%</td>
</tr>
<tr>
<td>$M = \text{Male}$</td>
<td>0.48 x 0.30 = 0.144 or 14.4%</td>
<td>33.6%</td>
</tr>
</tbody>
</table>

Among the whole population, the percentage of smokers is $P(S) = 35.2\%$.

(i) Among just males, what pct. smoke? In other words, what is the probability of smoking given that one is male?

Here we still want the probability of smoking $S$, but only among sub-population $M$. We now use the notation $P(S \mid M)$, where the first set $S$ is the pct. we want to measure and the second set denotes the restricted sub-population.

Here, $P(S \mid M) = \frac{P(S \cap M)}{P(M)} = \frac{14.4}{48} = 0.30$, or 30%. Thus, 30% of males smoke.

(ii) What is the probability of smoking given that one is female?

Now we want $P(S \mid F)$, the pct. of smokers $S$ among just females $F$. Here $P(S \mid F) = P(S \cap F) / P(F) = 20.8/52 = 0.40$. So 40% of females smoke.

(iii) and (iv) What are the probabilities of not smoking given that one is male and given that one is female? (That is, what pct. are nonsmokers just among males and what pct. are nonsmokers just among females?)

Given that one is male $M$, the chance of not smoking is $P(S' \mid M) = 33.6/48$, or 70%.

Given that one is female $F$, the chance of not smoking is $P(S' \mid F) = 31.2/52$, or 60%.
Observing Dependence

Which group is more likely to smoke, males or females?

Here $P(S \mid M) = 30\%$ and also $P(S \mid F) = 40\%;$ thus, females are more likely to smoke than males. We now can see that smoking is dependent on the sub-population because the percentages of smokers are different among each sub-population.

(If smoking were independent of sex, then each sub-population would have the same percentage of smokers.)

The Remaining Sub-Population Percentages

We can also find four other sub-population percentages. We can find the percentage of males/females just among smokers/non-smokers. In other words, we can now think of the traits $S$ and $S'$ as being the restricted sub-populations.

(v) Among just smokers, what pct. are male? In other words, what is the probability of being male given that one smokes?

(vi) What is the probability of being female given that one smokes?

Now we want the percentage of males $M$ among just smokers $S$: $P(M \mid S) = 14.4/35.2 \approx 40.9\%$. And the percentage of females $F$ among just smokers is $P(F \mid S) = 20.8/35.2 \approx 59.1\%$. Thus among smokers, 40.9% are male and 59.1% are female.

(vii) and (vii) What are the probabilities of being male and of being female given that one is a nonsmoker?

Here, $P(M \mid S') = 33.6/64.8 \approx 51.85\%$ and $P(F \mid S') = 31.2/64.8 \approx 48.15\%$; thus, 51.85% of nonsmokers are male and 48.15% of nonsmokers are female.

More Dependence

Now let’s compare the percentage of females among smokers and among nonsmokers: $P(F \mid S) = 59.1\%$ and $P(F \mid S') = 48.15\%$, which are not the same, so again we see the dependent traits.

(If smoking were independent of the sex, then we would have the same percentage of females in both categories.)
Exercise

Suppose 15% of all students are freshmen and 20% of all students live on campus. But 90% of freshmen live on campus.

Let $F =$ freshmen and $C =$ lives on campus.

(a) Show the information on a Block Diagram.

Put in symbols and compute the probability that one

(b) lives on campus given that one is a freshman
(c) does not live on campus given that one is a freshman

(d) lives on campus given that one is not a freshman
(e) does not live on campus given that one is not a freshman

(f) is a freshman given that one lives on campus
(g) is not a freshman given that one lives on campus

(h) is a freshman given that one does not live on campus,
(i) is not a freshman given that one does not live on campus.

(j) Compare the percentages of freshmen among those living on campus and among those not living on campus.

(k) Compare the percentages of those living on campus among freshmen and among non-freshmen. Who is more likely to live on campus, a freshman or a non-freshman?

(l) Is living on campus independent or dependent on whether or not one is a freshman? Explain completely.
Solution

\[ F = \text{freshmen} \quad F' = \text{non-frosh} \quad C = \text{live on campus} \quad C' = \text{live off campus} \]

90\% of Freshmen live on campus; so 0.15 \times 0.90 = 0.135 or 13.5\% are Freshmen who live on campus

\[
\begin{array}{c|c|c}
F & C & C' \\
\hline
13.5\% & 1.5\% & 15\% \\
6.5\% & 78.5\% & 85\% \\
20\% & 80\% & 100\%
\end{array}
\]

(b) \( P(C \mid F) = \frac{13.5}{15} \rightarrow 90\% \)

(d) \( P(C \mid F') = \frac{6.5}{85} \rightarrow 7.65\% \)

(c) \( P(C' \mid F) = \frac{1.5}{15} \rightarrow 10\% \)

(e) \( P(C' \mid F') = \frac{78.5}{85} \rightarrow 92.35\% \)

(f) \( P(F \mid C) = \frac{13.5}{20} \rightarrow 67.5\% \)

(h) \( P(F \mid C') = \frac{1.5}{80} \rightarrow 1.875\% \)

(g) \( P(F' \mid C) = \frac{6.5}{20} \rightarrow 32.5\% \)

(i) \( P(F' \mid C') = \frac{78.5}{80} \rightarrow 98.125\% \)

(j) \( P(F \mid C) = 67.5\% \) and \( P(F \mid C') = 1.875\% \).

(k) \( P(C \mid F) = 90\% \) and \( P(C \mid F') = 7.65\% \).

(l) Living on campus is dependent on whether or not one is a freshman. By Part (j), those living on campus are much more likely to be freshmen than those living off campus. By Part (k), freshmen are more likely to live on campus than non-freshmen.