If we had only been given the acceptance rates and the application rates, we could still recombine this information to obtain the overall acceptance rates for each gender. The program acceptance rates would have to be "weighted" by the application rate for that gender:

Acceptance rate for women =

$$\left(\frac{\#\,women\,applied\ \ program\ A}{total\,\#\,female\,applicants}\right)\left(\begin{array}{c}program\ A\,accept\\ rate\,for\,women\end{array}\right)+\left(\frac{\#\,women\,applied\ \ program\ F}{total\,\#\,female\,applicants}\right)\left(\begin{array}{c}program\ F\ accept\\ rate\,for\,women\end{array}\right)$$

(f) Verify that this formula produces the overall acceptance rate from (a).

(g) Repeat (f) for the men.

Simpson's Paradox Revisited

We will now use this relationship to discover mathematical conditions under which Simpson's Paradox can and can not occur. First define the following notation:

A_m = acceptance rate for men in program A
A_w = acceptance rate for women in program A
F_m = acceptance rate for men in program F
F_w = acceptance rate for women in program F
P_m = proportion of men applying to program A (so 1-P_m is the proportion of men applying to program F)
P_w = proportion of women applying to program A (so 1-P_w is the proportion of women applying to program F)

If we assume that women are accepted at a higher rate then men in each program, this implies $A_w > A_m$ and $F_W > F_m$. The question is under what conditions can the overall acceptance rate for men be higher than the overall acceptance rate for women. Asked another way, under certain conditions, is the paradox impossible?

To find these overall acceptance rates, we need the weighted averages:

Men: $A_mP_m + F_m(1\text{-}P_m)$ Women: $A_wP_w + F_m(1\text{-}P_w)$

Simpson's paradox is not present if we are guaranteed that the overall acceptance rate for women must be at least as large as the overall acceptance rate for men when $A_w > A_m$ and $F_W > F_m$.

(h) Consider the scenario where men and women apply to program A at the same rate (so $P_m = P_w$). Use the weighted averages to show that Simpson's paradox cannot occur in this scenario.

(i) Now consider the scenario where women and as likely to get into Program A as Program F and similarly for the men (so $A_w = F_w$ and $A_m = F_m$). Use the weighted averages to show that Simpson's paradox can not occur in this scenario.

Discussion: Simpson's Paradox does not occur whenever the sample sizes in the two groups are similar or the acceptance rates in the two groups in similar. In either of these situations, the weighted average will simplify to be the average of the two acceptance rates. When we do have unequal sample sizes and success rates, then we should use a weighted average to combine the results across groups.

SECTION 1-3 SUMMARY

The principle of confounding reveals why one cannot draw cause and effect conclusions from observational studies: we are not able to eliminate the possibility of another variable that impacts the response variable in a way that is not distinguishable from the explanatory variable. A special case of this arises with Simpson's Paradox, and you worked through a few examples mathematically to see how the "paradox" can be explained.