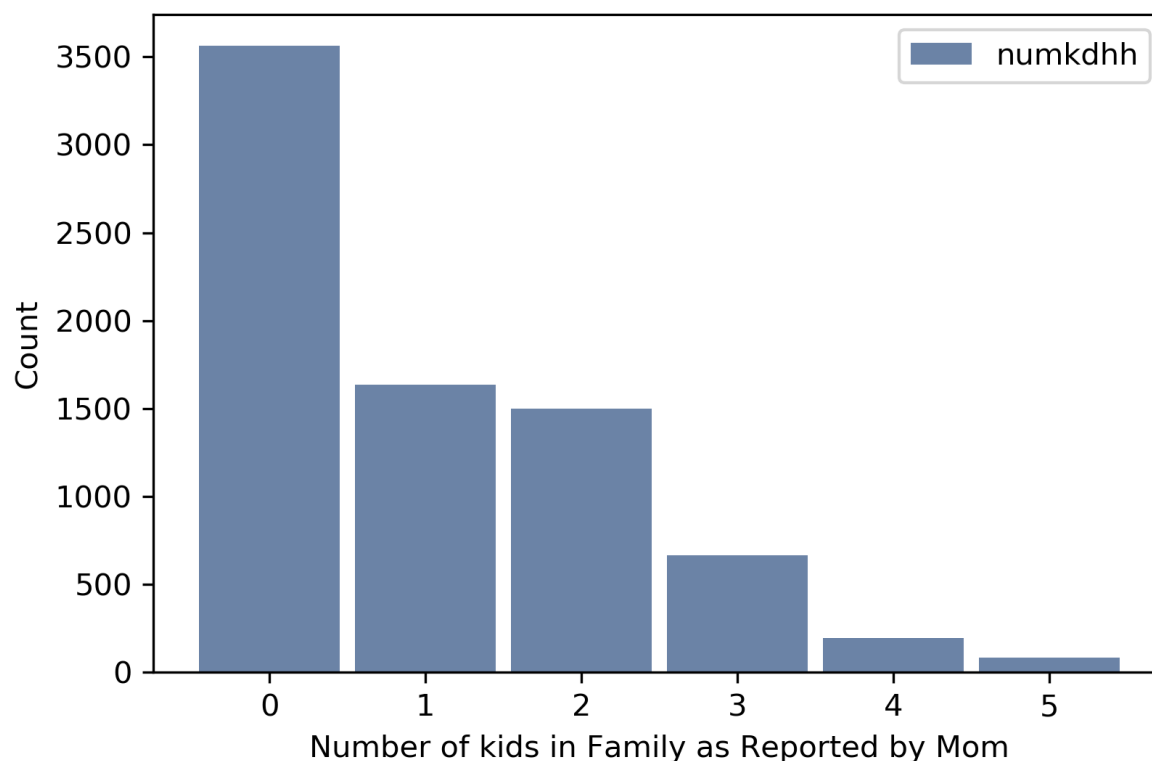


[Think Stats Chapter 3 Exercise 1](#) (actual vs. biased)

Exercise 3.1 Something like the class size paradox appears if you survey children and ask how many children are in their family. Families with many children are more likely to appear in your sample, and families with no children have no chance to be in the sample.

Use the NSFG respondent variable NUMKDHH to construct the actual distribution for the number of children under 18 in the household.

```
%matplotlib inline
import nsfg
import thinkstats2
import thinkplot
resp = nsfg.ReadFemResp()
num_kids_reported_by_mother = resp.numkdhh
hist = thinkstats2.Hist(num_kids_reported_by_mother, label='numkdhh')
thinkplot.Hist(hist)
thinkplot.Config(xlabel='Number of kids in Family as Reported by Mom',
                  ylabel='Count')
```



Now compute the biased distribution we would see if we surveyed the children and asked them how many children under 18 (including themselves) are in their household.

```

from collections import defaultdict
hist2 = defaultdict(int)
for n in hist:
    if n > 0:
        hist2[n] = hist[n]*n
hist2 = thinkstats2.MakeHistFromDict(hist2, label='numkdhh_modified')

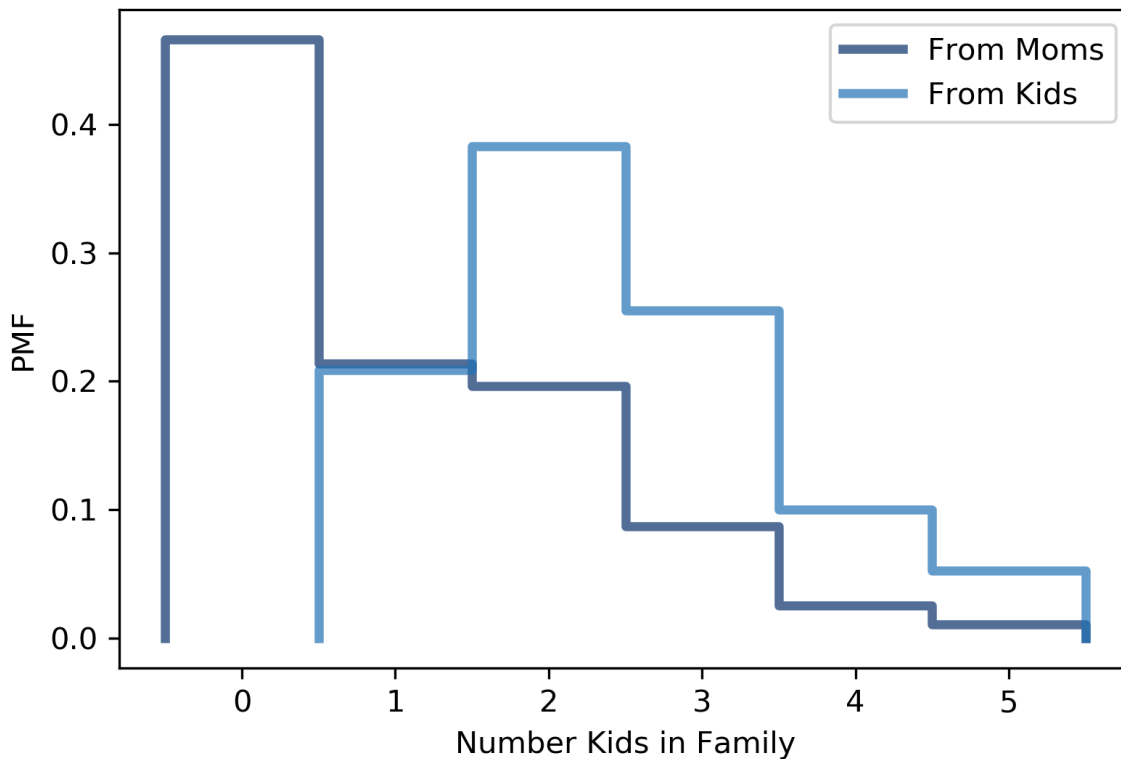
```

Plot the actual and biased distributions, and compute their means.

```

num_kids_pmf = thinkstats2.MakePmfFromHist(hist, label='From Moms')
num_kids_biased_pmf = thinkstats2.MakePmfFromHist(hist2, label="From Kids")
thinkplot.Pmfs([num_kids_pmf, num_kids_biased_pmf])
thinkplot.Config(xlabel='Number Kids in Family', ylabel='PMF')

```



```

print('Mean # of kids as reported by Mom: {0:1.2f}'.format(num_kids_pmf.Mean()))
print('Biased mean # of kids as reported by kids: {0:1.2f}'.format(num_kids_biased_pmf.Mean()))

```

Mean # of kids as reported by Mom: 1.02 Biased mean # of kids as reported by kids: 2.40

All the above work is captured in the notebook: [chap03ex1.ipynb](#).