

The Silicon Valley Dating Game

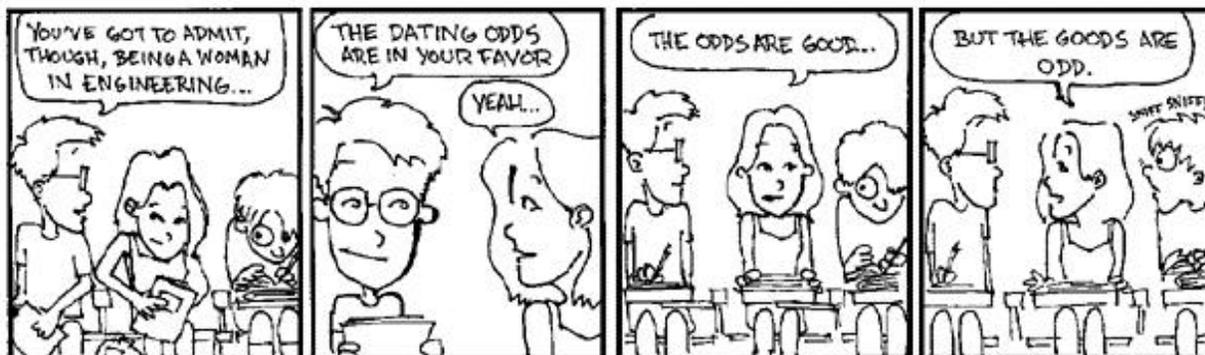
by

Peter Berg

June 11, 2001

Statistics 204 – Final Project

Stanford University



“Piled Higher and Deeper” (by Jorge Cham)
© The Stanford Daily

Introduction

In the San Francisco Bay area one often hears complaints about the large ratio of single men to single women. It is true that the plethora of tech-jobs has attracted a disproportionate number of “nerdy” males, but people often cite largely unsupported “statistics” of 3:1 ratios on the Peninsula, with some cities as high as 5:1. Complaints are especially common among male bar patrons who lament the seemingly scant availability of women in Silicon Valley. However, women who are told that they are lucky to have odds in their favor often counter with a sentiment similar to the one expressed in the above cartoon. It would seem then that the Silicon Valley dating scene is rather poor.

The purpose of this study was to come up with some estimate of the male-to-female ratio of bar patrons in the area immediately surrounding Stanford University and Palo Alto (ranging from Redwood City to Mountain View). In finding a reliable estimate for the percentage of men and women going out to bars one can hope to dispel some of the myths surrounding Silicon Valley dating.

After collecting and analyzing data, it was found that the approximately 62% of bar patrons in Palo Alto and surrounding cities are male (38% female), with a standard deviation of 1.27%. This means that the percentage of male bar patrons lies in the range of (59.46%, 64.54%) with 95% confidence. These numbers very roughly correspond to a 5:3 ratio of men to women.

Design

This study was conducted using a two-stage cluster sample. The primary sampling units were cities near Stanford University, and the secondary sampling units were individual bars within those cities. In the interest of time and simplicity, the “population” from which the clusters were selected consisted of the area spanned by four cities, Mountain View, Palo Alto, Menlo Park and Redwood City. Two of these (Menlo Park and Palo Alto) were selected at random to serve as primary sampling units. Then a list of all the bars in Menlo Park and Palo Alto were compiled and from these another simple random sample was taken (the ssu’s).

A disadvantage to using two-stage cluster sampling is the loss of precision from techniques like random or stratified sampling. However, in this case two-stage cluster sampling seemed best suited for the sampling model. A stratified sample, while preferred, would have been nearly impossible, because to implement it effectively the bars would had to have been classified into certain categories (restaurant/bar, dance club, bar, billiard hall, etc.). These distinctions are considerably subjective, as will be explained later, and in some cases would have yielded only one bar per stratum. Additionally, a pilot study could not guarantee any significant geographical correlation, so stratification by location did not seem worthwhile. Cluster sampling also had the advantage that the bars were already grouped by city.

Once the primary sampling units were determined, the challenge was to create a comprehensive list of bars for both selected cities. This was done with the aid of the telephone book, dining and entertainment guides, and, perhaps most helpful, Yahoo!’s online directory. Most important in compiling a list, however was defining the term “bar”. There are many different establishments, all with different closing times, menus, policies, clientele, and atmospheres. For the purposes of this study, a “bar” is defined as an establishment which has the following characteristics:

- serves alcoholic beverages (beer, wine, and/or hard liquor)
- is open past 10pm at night on Thursdays, Fridays and Saturdays.
- there must be at least 1.5 hours between the time when the establishment stops serving food and the time when it stops serving alcoholic drinks (this does not include hors d’oeuvres or small snacks served at the bar).

Procedure

Due to time constraints, the actual study was conducted over three days – Thursday, Friday and Saturday – June 7, 8, and 9, 2001. These days seemed the most reasonable, since bars traditionally conduct the most business on these evenings. Data was collected by a team of two people over a period of two to three hours each night. The order in which bars were visited was determined largely by their respective closing times and not by geography. For example, if a bar stopped serving food at 9:30 or 10:00pm and shut its doors by 11:00-11:30pm, then a headcount of bar patrons was done earlier in the evening (say 10:30pm) than for bars that closed much later (e.g. 2 am). Additionally, the exact time of the headcount for any given bar differed each evening. This method of counting ensured that most bars were visited towards the middle of their peak times and also reduced bias that might appear as a result of geographical similarities or cyclic trends in bar patronage.

As mentioned above, there were two people (my friend and I) involved in counting at all the bars. Both counters would enter the bar at the same time and slowly move from one end to the other, counting patrons as they went. Handheld mechanical click-counters were used to allow for full concentration on bar patrons. Once the back of the bar was reached, both turned around and counted a second time, moving from the back towards the front of the bar. One counter was assigned men on the way into the bar, while the other counted women. During the second count they switched, and the first counter took women while the second counted men. It should be noted that the number of male and female bar patrons was almost always the same on both passes through each bar. In the few instances where that was not true, the average of the two numbers were used. The difference was never more than 5 percent.

Despite all the planning that was done, there were a few challenges that arose during the course of the experiment. One challenge was how to deal with individuals who entered or exited the bar immediately following the counters' arrival. Many bars are in a continual state of flux, and getting all patrons to stand still during counting is simply not possible. Luckily, this was only a problem at a couple crowded bars. To get around it, the counters waited until movement had stabilized somewhat and then performed a rapid

count. Additionally, it was often the case that as people entered the bar, others exited, keeping the total inside the bar approximately constant. Furthermore, the fact that this problem was only encountered in very large bars meant that small changes in total bar numbers had a very small effect on the overall proportion of males to females.

Another complication arose when trying to count the number of individuals in the restrooms while the headcounts were taking place. To solve this issue, my partner and I (since we were both male) waited outside the men's and women's restrooms for several minutes after our initial headcount to see if anyone exited the restrooms. These additional individuals were then added to the total.

On the last night of data collection (Saturday, June 9), there was a completely unforeseen and rather unfortunate event which caused data collection to halt. A shooting took place outside one of the Palo Alto bars which had been selected as part of the sample. As a result, no data could be collected for that bar, and indeed data collection at all downtown Palo Alto bars was disturbed for reasons ranging from fleeing patrons to police roadblocks which made moving around the city very slow and arduous. Fortunately, the pilot survey that had been conducted the previous Saturday (6/2/01) included all the bars that fell into the random sample (as well as several others). So the data for that pilot Saturday was used instead, and the partial data for the day of the shooting was disregarded entirely to prevent skewing of the results.

Calculations

The survey data describe a two-stage cluster sample with unequal cluster sizes. Also, it is important to keep in mind that the value of interest is a proportion, not a total. From these facts it was concluded that using ratio estimation to find the mean proportion of males for the population would be the best model. From the raw data (see the Appendix), one can see that the same bars (ssu's) were sampled each night for the purposes of minimizing the variation between nights and types of establishments. It was assumed that the first SRS of the secondary sampling units was more or less representative of each cluster.

The data from all three nights for the first psu (Palo Alto) were strung together to form a single vector with 21 data points. In effect, by sampling the same seven bars in Palo Alto each night it was possible to treat the eighteen (18) original bars as a population of $(3 \times 18 = 54)$ fifty-four, from which a sample of size $3 \times 7 = 21$ was taken. Similarly, in Menlo Park three of six bars were sampled each night, for a total of nine out of eighteen. This seemed to make considerably more sense than calculating variances for each night, not only because of the tedium involved, but primarily because concatenating the data lowered the variance while summarizing all the information for each city. The sample sizes of 7 and 3 bars for Palo Alto and Menlo Park, respectively, were determined in part by calculations after the pilot study to reduce the variance, while simultaneously keeping the total time required for nightly data collection around 2.5 hours. It was found in the pilot survey that the upper limit of data collection within the 2.5 hour window of closing times was around ten bars. The higher proportion of Menlo Park bars sampled had to do with the small size of the cluster and the fact that sampling one extra bar required very little extra work and yielded a sample of 50% as opposed to 33%.

In setting up the calculation, the data used in computing the mean and variance were the number of men per bar and the total number of people per bar. The percent of men per bar was calculated for each establishment on each night (so the Palo Alto ratio vector consisted of 21 proportions describing the fraction of male patrons). To aid the reader, definitions and values of the variables essential to the computations are listed below:

Variable and Value	Description
$N = 4$	number of psu's in our population (four cities)
$n = 2$	number of psu's sampled
$M_1 = 54$ $M_2 = 18$	number of ssu's in the i^{th} psu (Palo Alto = M_1 , Menlo Park = M_2)
$m_1 = (3 \times 7) = 21$ $m_2 = (3 \times 3) = 9$	number of elements (bars) in each ssu.
$\bar{y}_1 = 0.613$ $\bar{y}_2 = 0.639$	sample mean (average of all ratios for bars in a given city)

First the mean value of the proportion of males in bars near Palo Alto was calculated using the ratio estimator for population mean:

$$\hat{y}_r = \frac{\sum_i M_i \bar{y}_i}{\sum_i M_i}.$$

Using this, we find that the average proportion of men for the entire population of four cities is 62%.

The variance formula for the case of ratio estimation is based on the Taylor series approximation and has the form

$$\hat{V}(\hat{y}_r) = \frac{1}{\bar{M}^2} \left(1 - \frac{n}{N}\right) \frac{s_r^2}{n} + \frac{1}{nN} \sum_i M_i^2 \left(1 - \frac{m_i}{M_i}\right) s_i^2$$

where \bar{M}^2 is the average cluster size, the s_i^2 's are just the variance of the proportion vector for each city, and s_r^2 is given by

$$s_r^2 = \frac{\sum_i (M_i \bar{y}_i - M_i \hat{y}_r)^2}{n-1}.$$

Plugging in the values for each respective variable yields a variance of $\hat{V}(\hat{y}_r) = 0.00016$. The standard deviation then becomes 0.0127, or 1.27%. A 95% confidence interval is located two standard deviations from either side of the mean proportion, so the interval becomes [59.46%, 64.54%], a margin of error of ± 2.54 . (For more rigorous calculations and the data, please see the appendix).

Conclusion

Based on this study's findings, it is safe to say that the ratio of men to women in bars near Stanford University and Palo Alto is weighted rather heavily towards the male population. At 5 men for every 3 women, the ratios would appear to favor women's chances for finding a man at a bar, though there are several other factors which complicate the issue.

First, this study makes no claims about the ratio of *single* males to *single* females. The data only describes all the individuals in local bars. Furthermore, given the short time frame within which the data was collected, it is difficult to make any sort of conjectures

as to the type of clientele found in local bars on Thursdays, Fridays and Saturdays. For example, it is quite possible that couples are more likely to go out for dinner on Thursday and stay out a while later for drinks. Singles, on the other hand, are probably more likely to frequent bars on Friday and Saturday nights.

There is most likely also a correlation between the type of bar and the male-to-female ratio. For example, bars such as the Oasis in Menlo Park and Antonio's Nut House in Palo Alto are rather similar in atmosphere and clientele. They also both had higher male/female ratios than most other bars, usually hovering around 3:1.

Additional things to be considered are possible sources of error which could influence the variance. These include human error in counting people (it is still noteworthy to mention that only very rarely were the two bar counts unequal) or the fact that only one weekend was picked for data collection (and this happened to be during Stanford University's finals week).

Given more time and resources, it would be good to create a much larger list of bars on the Peninsula. These should then be classified into various types of strata (based on atmosphere or genre) with samples drawn from each. In this way, variance within a cluster (or stratum) is greatly reduced. A further measure of interest could come from examining geographic trends (e.g. do bars in Palo Alto have better or worse ratios than San Jose? than San Francisco?).

In any case, the data do make a compelling argument that the ratio of men to women in local bars is consistently rather high, though not as high as some people claim. Perhaps this knowledge will lead bachelors to hone their skills in making themselves interesting and attractive to women. It is unclear, however, if women will continue to find Silicon Valley's male goods "odd", even if the numbers are in their favor.

Appendix

Bars in Palo Alto and Menlo Park

Palo Alto (M1)

- 1 Gordon Biersch Brewery
- 2 Q Café
- 3 Rose & Crown
- 4 Fanny & Alexander
- 5 Rudy's Pub
- 6 Blue Chalk Café
- 7 Empire Tap Room
- 8 Left At Albuquerque
- 9 Nola's Restaurant & Bar
- 10 Perry's Palo Alto
- 11 La Bodeguita Del Medio
- 12 Antonio's Nut House
- 13 Crowne Plaza Cabana Palo Alto
- 14 Sheraton
- 15 Zibibbo
- 16 Circadia
- 17 Compadres Mexican Bar & Grill
- 18 Island

Menlo Park (M2)

- 1 British Bankers Club
- 2 Dutch Goose
- 3 Duck Club
- 4 Oasis Beer Garden
- 5 Vida Bistro
- 6 Wild Hare