

Descriptive Statistics

Prerequisites

none

Descriptive statistics are numbers that are used to summarize and describe data. The word "data" refers to the information that has been collected from an experiment, a survey, an historical record, etc. (By the way, "data" is plural. One piece of information is called a "datum.") If we are analyzing birth certificates, for example, a descriptive statistic might be the percentage of certificates issued in New York State, or the average age of the mother. Any other number we choose to compute also counts as a descriptive statistic for the data from which the statistic is computed. Several descriptive statistics are often used at one time, to give a full picture of the data.

Descriptive statistics are just descriptive. They do not involve **generalizing** beyond the data at hand. Generalizing from our data to another set of cases is the business of [inferential statistics](#), which you'll be studying in the next [section](#). Here we focus on (mere) descriptive statistics.

Some descriptive statistics are shown in Table 1. The table shows the average salaries for various occupations in the United States in 1999. (Click [here](#) to see how much individuals with other occupations earn.)

| Table 1. Average salaries for various occupations in 1999. | |
|--|--|
| \$112,760 | pediatricians |
| \$106,130 | dentists |
| \$100,090 | podiatrists |
| \$ 76,140 | physicists |
| \$ 53,410 | architects, |
| \$ 49,720 | school, clinical, and counseling psychologists |
| \$ 47,910 | flight attendants |
| \$ 39,560 | elementary school teachers |
| \$ 38,710 | police officers |
| \$ 18,980 | floral designers |

Descriptive statistics like these offer insight into American society. It is interesting to note, for example, that we pay the people who educate our children and who protect our citizens a great deal less than we pay people who take care of our feet or our teeth.

For more descriptive statistics, consider Table 2. It shows the number of unmarried men per 100 unmarried women in U.S. Metro Areas in 1990. From this table we see that men outnumber women most in Jacksonville, NC, and women outnumber men most in Sarasota, FL. You can see that descriptive statistics can be useful if we are looking for an opposite-sex partner! (These data come from the [Information Please Almanac web site](#).)

| Cities with mostly men | Men per 100 Women | Cities with mostly women | Men per 100 Women |
|--------------------------------------|-------------------|--------------------------|-------------------|
| 1. Jacksonville, N. C. | 224 | 1. Sarasota, FL | 66 |
| 2. Killeen-Temple, TX | 123 | 2. Bradenton, FL | 68 |
| 3. Fayetteville, NC | 118 | 3. Altoona, PA | 69 |
| 4. Brazoria, TX | 117 | 4. Springfield, IL | 70 |
| 5. Lawton, OK | 116 | 5. Jacksonville, TN | 70 |
| 6. State College, PA | 113 | 6. Gadsden, AL | 70 |
| 7. Clarksville-Hopkinsonville, TN-KY | 113 | 7. Wheeling, WV | 70 |
| 8. Anchorage, Alaska | 112 | 8. Charleston, WV | 71 |
| 9. Salinas-Seaside-Monterey, CA | 112 | 9. St. Joseph, MO | 71 |
| 10. Bryan-College Station, TX | 111 | 10. Lynchburg, VA | 71 |

NOTE: Unmarried includes never-married, widowed, and divorced persons, 15 years or older.

These descriptive statistics may make us ponder why the numbers are so disparate in these cities. One potential explanation, for instance, as to why there are more women in Florida than men may involve the fact that elderly

individuals tend to move down to the Sarasota region and that women tend to out live men. Thus, more women might live in Sarasota than men. However, in the absence of proper data, this is only speculation.

You probably know that descriptive statistics are central to the world of sports. Every sporting event produces numerous statistics such as the shooting percentage of players on a basketball team. For the Olympic marathon (a foot race of 26.3 miles), we possess data that cover more than a century of competition. (The first modern Olympics took place in 1896). The following table shows the winning times for both men and women (the latter have only been allowed to compete since 1984).

| Women | | | |
|-------|--------------------|---------|---------|
| Year | Winner | Country | Time |
| 1984 | Joan Benoit | USA | 2:24:52 |
| 1988 | Rosa Mota | POR | 2:25:40 |
| 1992 | Valentina Yegorova | UT | 2:32:41 |
| 1996 | Fatuma Roba | ETH | 2:26:05 |
| 2000 | Naoko Takahashi | JPN | 2:23:14 |
| Men | | | |
| Year | Winner | Country | Time |
| 1896 | Spiridon Louis | GRE | 2:58:50 |
| 1900 | Michel Thato | FRA | 2:59:45 |
| 1904 | Thomas Hicks | USA | 3:28:53 |
| 1906 | Billy Sherring | CAN | 2:51:23 |
| 1908 | Johnny Hayes | USA | 2:55:18 |
| 1912 | Kenneth McArthur | S. Afr. | 2:36:54 |
| 1920 | Hannes Kolehmainen | FIN | 2:32:35 |
| 1924 | Albin Stenroos | FIN | 2:41:22 |
| 1928 | Boughra El Ouafi | FRA | 2:32:57 |
| 1932 | Juan Carlos Zabala | ARG | 2:31:36 |
| 1936 | Sohn Kee-Chung | JPN | 2:29:19 |
| 1948 | Delfo Cabrera | ARG | 2:34:51 |
| 1952 | Emil Ztopek | CZE | 2:23:03 |

| | | | |
|------|---------------------|---------|---------|
| 1956 | Alain Mimoun | FRA | 2:25:00 |
| 1960 | Abebe Bikila | ETH | 2:15:16 |
| 1964 | Abebe Bikila | ETH | 2:12:11 |
| 1968 | Mamo Wolde | ETH | 2:20:26 |
| 1972 | Frank Shorter | USA | 2:12:19 |
| 1976 | Waldemar Cierpinski | E.Ger | 2:09:55 |
| 1980 | Waldemar Cierpinski | E.Ger | 2:11:03 |
| 1984 | Carlos Lopes | POR | 2:09:21 |
| 1988 | Gelindo Bordin | ITA | 2:10:32 |
| 1992 | Hwang Young-Cho | S. Kor | 2:13:23 |
| 1996 | Josia Thugwane | S. Afr. | 2:12:36 |
| 2000 | Gezahenge Abera | ETH | 2:10.10 |

There are many descriptive statistics that we can compute from the data in the table. To gain insight into the improvement in speed over the years, let us divide the men's times into two pieces, namely, the first 12 races (up to 1948) and the second 13 (starting from 1952). The average winning time for the first 12 races is 2 hours 46 minutes and 8 seconds (written 2:46:08). The average winning time for the second 13 races is 2:09:40. This is quite a difference (almost half an hour). Does this prove that the fastest men are running faster? Or is the difference just due to chance, no more than what often emerges from chance differences in performance from year to year? We can't answer this question with descriptive statistics alone. All we can affirm is that the two averages is "suggestive."

Examining Table 3 leads to many other questions. We note that Takahashi (the lead female runner in 2000) would have beaten the male runner in 1956 and all male runners in the first 12 marathons. This fact leads us to ask whether the gender gap will close or remain constant. When we look at the times within each gender, we also wonder how far they will decrease (if at all) in the next century of the Olympics. Might we one day witness a sub-2 hour marathon? The study of statistics can help you make reasonable guesses about the answers to these questions.