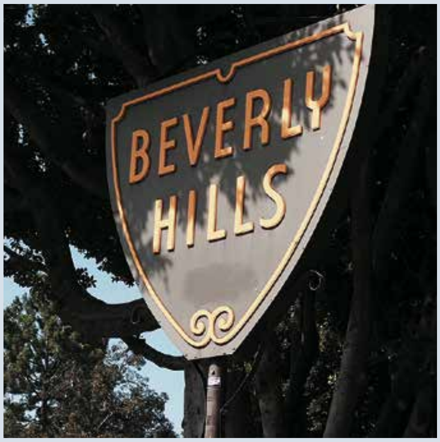
MEDIA clips

Cris Wellington//Anne Quinn

High-Priced Homes



Los Angeles has always been home to some of the world's most expensive real estate. But forget Beverly Hills, 90210: The new hot spot for multimillion-dollar mansions is Duarte, 91008. . . .

A scant 1,391 people live in 91008 ZIP code, and only 12 homes are currently on the market. So a single high-priced listing (like the mammoth nine-bedroom, built this year, that's selling for $19.8 million) is enough to skew the median [sale] price skyward. . . .

Source: Francesca Levy, "America's Most Expensive ZIP Codes 2010," *Forbes*, Sept. 27, 2010, http://www.msnbc.msn.com/id/39521539/ns/business-forbescom/

1. In data related to the article, more than 80 home prices are listed. We will consider 11 prices in hypothetical community A in Los Angeles. In this community, a set of homes for sale is advertised at the following prices in millions of dollars: 1.125, 5.95, 4.68, 0.63, 5.88, 2.98, 1.299, 5.888, 5.85, 1.875, and 5.98.

   (*a*) Construct a box plot of these sale prices.

   (*b*) Calculate the mean sale price.

2. Consider the set of 11 sale prices above, with the addition of a new house listed at $19.8 million.

   (*a*) Construct a box plot of these sale prices.

   (*b*) How is this box plot similar to and different from that in question 1?

   (*c*) Calculate the mean sale price.

3. Consider the data you have collected in both scenarios.

   (*a*) Compare the change in the median sale price with the change in the mean sale price after the $19.8 million house was added.

   (*b*) With respect to the clip about the town of Duarte, California, how would you reword the last sentence?

4. In this example, we have seen how an outlier could possibly influence the mean or median.

   (*a*) Make your own data set in which the inclusion of an outlier does not change the median but drastically changes the mean.

   (*b*) Make your own data set in which the mean and median are equal to each other before and after including an outlier.

Media Clips appears in every issue of *Mathematics Teacher*, offering readers contemporary, authentic applications of quantitative reasoning based on print or electronic media. All submissions should be sent to the editors. For information on the department and guidelines for submitting a clip, visit http://www.nctm.org/mtcalls.

Department editors
Louis Lim, *louis.lim1@gmail.com, Bayview Secondary School, Richmond Hill, ON, Canada; and* ***Chris Bolognese,*** *bolognesechris@gmail.com, Columbus Academy, Columbus, OH*

Conditional Probability and Crime

Source: *Dilbert*, Sept. 29, 1992, http://dilbert.com/strips/comic/1992-09-29/. Reprinted with permission.

Although the 2012 U.S. Census does not include data relating to unmarried male criminals specifically, it does have separate data for male criminals and unmarried criminals. According to the 2012 census, there were 613 married and 2,560 unmarried (never married, divorced, or widowed) death row inmates in the year 2009. Further 2012 census data shows that in the year 2009 there were 233,090,322 adults and that 56.4% of the general adult population was married.

Table 1 ("Conditional Probability and Crime")

	On Death Row	Not on Death Row	Total
Married	613		
Unmarried	2,560		
Total			233,090,322

1. (*a*) Complete **table 1 ("Conditional Probability and Crime")**, using the 2012 census information that 56.4% of the general population was married and the total number of adults was 233,090,322.

   (*b*) A *conditional probability* written as $P(A|B)$ measures the probability of an event, A, given that another event, B, is occurring or has already occurred. We can compute this probability as follows:

$$P(A|B) = \frac{n(A \text{ and } B)}{n(B)}$$

   Let U denote the event that someone is unmarried. Let D denote the event that someone is on death row. Use the data in the table to calculate $P(U|D)$ and $P(D|U)$. Explain why the two probabilities, $P(U|D)$ and $P(D|U)$, are so different.

2. (*a*) From the 2012 census, we know that 50.8% of the U.S. population is female. In addition, 667,039 men and 93,176 females are in jail. Complete **table 2 ("Conditional Probability and Crime")**.

   (*b*) Let M denote the event that someone is male. Let J denote the event that someone is in jail. Without actually calculating the conditional probabilities, predict whether the two calculations, $P(M|J)$ and $P(J|M)$, are likely to be similar or different in this context. Explain your thinking.

   (*c*) For A and B as in part (*b*), use the data in **table 2 ("Conditional Probability and Crime")** to calculate $P(M|J)$ and $P(J|M)$.

   (*d*) Are the final answers in question 2(*b*) equal? Why or why not?

   (*e*) Does it appear that men commit 90% of violent crime? Does your answer support or contradict the claim in the cartoon?

Table 2 ("Conditional Probability and Crime")

	In Jail	Not in Jail	Total
Male	667,039		
Female	93,176		
Total			233,090,322

Table 3 ("Conditional Probability and Crime")

	B	Not *B*	Total
A			
Not *A*			
Total			

3. (*a*) Create your own situation in the two-way table (**table 3 ["Conditional Probability and Crime"]**) so that $P(A|B) = P(B|A)$.

   (*b*) In general, if $P(A|B) = P(B|A)$ for two events A and B, what must be true about events A and B?

MEDIA CLIPS solutions

"High-Priced Homes" answers

The website http://en.wikipedia.org/wiki/Quartile describes multiple ways that we could compute the median. We use method 1, in which the median is not included in the upper or lower half of the data.

1. (*a*) We can reorder the given list of values (in $ million) in ascending size: 0.63, 1.125, 1.299, 1.875, 2.98, 4.68, 5.85, 5.88, 5.888, 5.95, and 5.98. The minimum and maximum values are immediately discernible. Because the number of values is odd, the median is the [(11 + 1)/2]th, or the 6th, value—namely, 4.68 (million). Less than this median, we have the 5 values—0.63, 1.125, 1.299, 1.875, and 2.98—from which the [(5 + 1)/2]th, or the 3rd, value 1.299 of this sublist is the first quartile. In a similar manner, the 5 values greater than the median—namely, 5.85, 5.88, 5.888, 5.95, and 5.98—give the 3rd value, 5.888, as the third quartile.

   The five-number summary is therefore given by 0.63 (minimum), 1.299 (first quartile), 4.68 (second quartile, or median), 5.888 (third quartile), and 5.98 (maximum). The box plot is shown in **fig. 1 ("High-Priced Homes")**.

   (*b*) The mean sale price is approximately $3.83 million. Recall that the mean is the average of the values—in this case, (0.63 + 1.125 + 1.299 + 1.875 + 2.98 + 4.68 + 5.85 + 5.88 + 5.888 + 5.95 + 5.98)/11. Note that here the mean, 3.83, is less than the median, 4.68.

2. (*a*) We can append 19.8 ($ million) to the ordered list of 11 values to get the minimum, 0.63, and the maximum, 19.8. Since this list has an even number of values, the median is the average of the (12/2)th, or 6th, value and the [(12 + 2)/2]th, or 7th, value—that is, (4.68 + 5.85)/2 = 5.265. Now split the list of 12 values into two sublists. The left sublist of 6 values is {0.63, 1.125, 1.299, 1.875, 2.98, 4.68} from which the first quartile is the average of the (6/2)th, or 3rd, value and the [(6 + 2)/2]th, or 4th, value. That is, Q_1 = (1.299 + 1.875)/2 = 1.587.

   Similarly, we can use the right sublist—{5.85, 5.88, 5.888, 5.95, 5.98, 19.8}—to determine the third quartile using the average of the 3rd and 4th values. Then we have (5.888 + 5.95)/2 = 5.919 as our third quartile.

   Thus, the 5-number summary (in $ million) is given by 0.63 (minimum), 1.587 (first quartile), 5.265 (second quartile, or median), 5.919 (third quartile), and 19.8 (maximum). The box plot would look like the lower graph in **figure 2 ("High-Priced Homes")**. Note that the maximum value in this case is an outlier, as represented by a single point.

   (*b*) Both box plots have the same minimum value and have similar values for quartiles 1, 2, and 3. However, their maximum values are quite different.

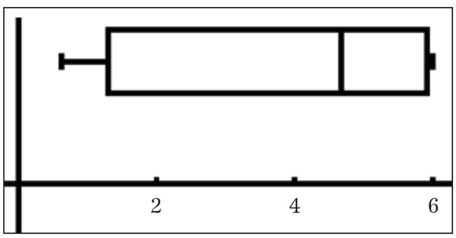


Fig. 1 ("High-Priced Homes")

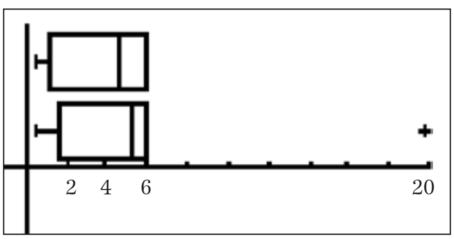


Fig. 2 ("High-Priced Homes")

   (*c*) The mean sale price with the additional home is $5.161 million, far greater than the mean before inclusion of the new, expensive home.

3. (*a*) The median sale price increased by only about a half million dollars (that is, $585,000) from the 11-number data set to the 12-number data set, whereas the mean sale price jumped by more than $1,300,000 million:

   Median2 – Median1 = $5,265,000 – $4,680,000 = $585,000
   Mean2 – Mean1 = $5,161,000 – $3,831,000 = $1,330,000

   We see that outliers do not affect the medians of data sets as much as they affect the means.

   (*b*) Extreme values, such as $19.8 million, do not affect the medians of data sets as much as they affect the means. The sentence in the article should have read: "So a single high-priced listing (like the mammoth nine-bedroom, built this year, that's selling for $19.8 million) is enough to skew the *mean* [sale] price skyward."

4. (*a*) Consider an initial data set {5, 5}. The median and the mean are both equal to 5. Now consider the expanded data set {5, 5, 200}, in which 200 is the included outlier. The median is still equal to 5, but the mean is equal to 70. Comparing these statistics before and after the outlier was included, we see that the median has not changed but the mean has increased substantially.

   (*b*) Consider the initial data set {2, 2, 20, 20}. The median and the mean are equal to 11. Now consider the set {2, 2, 20, 20, 56}, in which 56 is the included outlier. The median and the mean are now equal to 20.

"Conditional Probability and Crime" answers

1. (*a*) Complete the table relating marriage and crime by working in the rightmost column first (see **table 4 ["Conditional Probability and Crime"]**).

   (*b*) Find the conditional probability:

$$P(U \mid D) = P(\text{unmarried} \mid \text{on death row}) = \frac{n(\text{unmarried and on death row})}{n(\text{on death row})} = \frac{2560}{3173} \approx .8068$$

   Therefore, approximately 81% of people on death row are unmarried.

   Conversely,

$$P(D \mid U) = P(\text{on death row} \mid \text{unmarried}) = \frac{n(\text{on death row and unmarried})}{n(\text{unmarried})} = \frac{2560}{101{,}627{,}380} \approx .00002519.$$

   Thus, approximately 0.0025% of unmarried people are on death row.

   (*c*) The two answers are so different because the number of people on death row is much smaller than the number of unmarried people. In calculations, the number 2,560 representing the people who are unmarried and on death row is divided by two very different denominators (3,173 and 101,627,380). Further, when we consider the context, just because someone is unmarried does not imply that he or she would be on death row. Conversely, however, someone on death row would be much more likely to be unmarried.

2. (*a*) Complete the table relating gender and crime by working in the rightmost column first (see **table 5 ["Conditional Probability and Crime"]**).

   (*b*) $P(M|J)$ and $P(J|M)$ should be quite different because the total number of males in the U.S. population is much greater than those in jail:

$$P(M \mid J) = P(\text{male} \mid \text{in jail}) = \frac{n(\text{male and in jail})}{n(\text{in jail})} = \frac{667{,}039}{760{,}215} \approx .8774$$

   Therefore, approximately 88% of those in jail are male.

   (*c*) Conversely,

$$P(J \mid M) = P(\text{in jail} \mid \text{male}) = \frac{n(\text{in jail and male})}{n(\text{is male})} = \frac{667{,}039}{114{,}680{,}438} \approx .00582.$$

   Thus, approximately 0.58% of males are in jail.

   (*d*) These are very different probabilities. As with question 1, the population in jail is very small compared with the total population. Interested readers may wish to further investigate the base-rate fallacy.

   (*e*) From these calculations, we see that close to 90% of those in jail are male. However, those in jail most likely committed a nonviolent crime. Jail is mostly for those awaiting trial or serving a term for less than 1 year, whereas prison is for those serving longer terms for more severe crimes. Thus, we cannot determine whether 90% of those who committed violent crimes were male. Further, the cartoon claims something even more specific about these men being unmarried. We do not have the data necessary to accept or refute the cartoon's claim.

3. Our goal is to construct a situation in which $P(A|B) = P(B|A)$.

Table 4 ("Conditional Probability and Crime")

	On Death Row	Not on Death Row	Total
Married	613	131,462,942 – 613 = 131,462,329	(0.564)(233,090,322) ≈ 131,462,942
Unmarried	2,560	101,627,380 – 2,560 = 101,624,820	(0.436)(233,090,322) ≈ 101,627,380
Total	613 + 2,560 = 3,173	131,462,329 + 101,624,820 = 233,087,149	233,090,322

Table 5 ("Conditional Probability and Crime")

	In Jail	Not in Jail	Total
Male	667,039	114,680,438 – 667,039 = 114,013,399	(0.492)(233,090,322) ≈ 114,680,438
Female	93,176	118,409,884 – 93,176 = 118,316,708	(0.508)(233,090,322) ≈ 118,409,884
Total	667,039 + 93,176 = 760,215	114,013,399 + 118,316,708 = 232,330,107	233,090,322

Table 6 ("Conditional Probability and Crime")

	B	Not B	Total
A	10	5	15
Not A	5	18	23
Total	15	23	38

(a) Consider the example in **table 6 ("Conditional Probability and Crime")**. In this case, $P(A|B) = 10/15$, and $P(B|A) = 10/15$ as well.

(b) From **table 6 ("Conditional Probability and Crime")**, it appears that $P(A|B) = P(B|A)$ only when events A and B have the same number or size. We can see this algebraically as well:

Since the numerators themselves are equal, these ratios are equal only if the denominators are equal. Stated more explicitly, $P(A|B) = P(B|A)$ if and only if $P(A) = P(B)$.

CRIS WELLINGTON, wellingtoncris@yahoo.com, who submitted "High-Priced Homes," teaches sixth-grade mathematics at Middle Village Prep Charter School in Middle Village, New York. He is interested in problem solving and student thinking and enjoys researching new ideas and trying new ideas in his classroom. **ANNE QUINN,** quinna@edinboro.edu, who submitted "Conditional Probability and Crime," is chair of the Mathematics and Computer Science Department at Edinboro University in Pennsylvania.