

Question 4: What is the difference between marginal and joint probability?

In Question 3 we introduced the idea of joint probability. Joint probabilities are the likelihoods associated with compound events using “and”. The joint probability of  $A$  and  $B$  is the likelihood that both events will occur simultaneously. Such probabilities are often encountered in consumer surveys. In this context, a consumer answers many questions about their behavior. In addition, the consumer also gives other information about themselves like gender, income, or educational attainment.

In several examples, we have looked at the results of a survey of 3743 consumers regarding data usage on cell phones. This survey also examined the gender of cell phone users with the results shown below.

Amount of Data Used	Male Users	Female Users	Total for each row
Less than 200 MB	650	641	1291
200 MB up to, but not including 500 MB	442	316	758
500 MB up to, but not including 1 GB	291	132	423
1 GB up to, but not including 2 GB	172	152	324
2 GB or more	507	440	947
Total for each column	2062	1681	3743

Suppose we are interest in the joint probability that user is male and uses 2GB or more of data. The inclusion of the word “and” indicates we are interested in a compound event  $A$  and  $B$  where

$A$ : the user is male

$B$ : 2 GB or more of data is used

To find the joint probability  $P(A \text{ and } B)$ , we need to calculate the relative frequency of the event  $A$  and  $B$ . Since there are 507 male users who use 2 GB or more of data and 3743 total users,

$$P(A \text{ and } B) = \frac{507}{3743} \approx 0.135$$

In this context, the probability of either of the individual events is called the marginal probability. The marginal probability of A is calculated by finding the relative frequency of the event A.

Amount of Data Used	Male Users	Female Users	Total for each row
Less than 200 MB	650	641	1291
200 MB up to, but not including 500 MB	442	316	758
500 MB up to, but not including 1 GB	291	132	423
1 GB up to, but not including 2 GB	172	152	324
2 GB or more	507	440	947
Total for each column	2062	1681	3743

Since there is a total of 2062 male users out of 3743 total users, the marginal probability is

$$P(A) = \frac{2062}{3743} \approx 0.551$$

The total number of male users is found by summing the shaded column values. In the example below, we use joint and marginal probabilities to find the likelihood of a compound event using “or”.

### Example 6 Joint and Marginal Probability

In this example, we are interested in the events

*C*: the user is female

*D*: less than 200 MB of data is used

Use the information in the table below to calculate probabilities in each part.

Amount of Data Used	Male Users	Female Users	Total for each row
Less than 200 MB	650	641	1291
200 MB up to, but not including 500 MB	442	316	758
500 MB up to, but not including 1 GB	291	132	423
1 GB up to, but not including 2 GB	172	152	324
2 GB or more	507	440	947
Total for each column	2062	1681	3743

- a. Find the likelihood that a female user will use less than 200 MB of data.

**Solution** In terms of these events, we must find the probability that the user is female and less than 200 MB of data is used,  $P(C \text{ and } D)$ . From the table, we recognize that there are 641 female users who use less than 200 MB of data. Since the total number of users is 3743, the relative frequency may be calculated,

$$P(C \text{ and } D) = \frac{641}{3743} \approx 0.171$$

The likelihood that a user in the survey is female and used less than 200 MB of data is approximately 17.1%.

- b. Find the probability that a user is female.

**Solution** To use calculate the relative frequency of the event, we must divide the number of female users by the total number of users. The total number female users is at the bottom of the third column. Dividing this by the total number of users in the bottom of the last column gives,

$$P(C) = \frac{1681}{3743} \approx 0.449$$

The probability that a user in the survey is female is approximately 44.9%.

- c. Find the probability that a user in the survey will use less than 200 MB of data.

**Solution** According to the survey, 1291 users of the total 3743 users used less than 200 MB of data. This means the probability of using less than 200 MB of data is

$$P(D) = \frac{1291}{3743} \approx 0.345$$

The likelihood of using less than 200 MB of data is approximately 34.5%.

- d. Find the probability that the user is female or less than 200 MB of data is used.

**Solution** The event “the user is female or less than 200 MB of data is used” corresponds to the compound event  $C$  or  $D$ . We calculate the probability of this event by using the probabilities of the events in parts a through c,

$$\begin{aligned} P(C \text{ or } D) &= P(C) + P(D) - P(C \text{ and } D) \\ &= \frac{1681}{3743} + \frac{1291}{3743} - \frac{641}{3743} \\ &= \frac{2331}{3743} \\ &\approx 0.623 \end{aligned}$$

The probability that a user is female or less than 200 MB of data is used is approximately 62.3%.

