

11.2: Conditional Probability (Section 10.4)

Let's determine the conditional probability of someone being unhealthy, given that they are over 70 years of age, using the NHANES dataset. Let's create a new data frame that

```
healthDataFrame <-  
  NHANES %>%  
  mutate(  
    Over70 = Age > 70,  
    Unhealthy = DaysPhysHlthBad > 0  
  ) %>%  
  dplyr::select(Unhealthy, Over70) %>%  
  drop_na()  
  
glimpse(healthDataFrame)
```

```
## Observations: 4,891  
## Variables: 2  
## $ Unhealthy <lgl> FALSE, FALSE, FALSE, TRUE, FALSE, TRUE,...  
## $ Over70 <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALS...
```

First, what's the probability of being over 70?

```
pOver70 <-  
  healthDataFrame %>%  
  summarise(pOver70 = mean(Over70)) %>%  
  pull()  
  
# to obtain the specific value, we need to extract it from the data frame  
  
pOver70
```

```
## [1] 0.11
```

Second, what's the probability of being unhealthy?

```
pUnhealthy <-  
  healthDataFrame %>%  
  summarise(pUnhealthy = mean(Unhealthy)) %>%  
  pull()  
  
pUnhealthy
```

```
## [1] 0.36
```

What's the probability for each combination of unhealthy/healthy and over 70/ not? We can create a new variable that finds the joint probability by multiplying the two individual binary variables together; since anything times zero is zero, this will only have the value 1 for any case where both are true.

```
pBoth <- healthDataFrame %>%
  mutate(
    both = Unhealthy*Over70
  ) %>%
  summarise(
    pBoth = mean(both)) %>%
  pull()

pBoth
```

```
## [1] 0.043
```

Finally, what's the probability of someone being unhealthy, given that they are over 70 years of age?

```
pUnhealthyGivenOver70 <-
  healthDataFrame %>%
  filter(Over70 == TRUE) %>% # limit to Over70
  summarise(pUnhealthy = mean(Unhealthy)) %>%
  pull()

pUnhealthyGivenOver70
```

```
## [1] 0.38
```

```
# compute the opposite:
# what the probability of being over 70 given that
# one is unhealthy?
pOver70givenUnhealthy <-
  healthDataFrame %>%
  filter(Unhealthy == TRUE) %>% # limit to Unhealthy
  summarise(pOver70 = mean(Over70)) %>%
  pull()

pOver70givenUnhealthy
```

```
## [1] 0.12
```

This page titled [11.2: Conditional Probability \(Section 10.4\)](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.