# Solutions to exercises

Listed below are the solutions to the exercises.

All solutions are found using RStudio, though **you should only do the exercises in RStudio if indicated in the list of exercises.** This may result in slight differences in numerical answers, which is due to rounding errors.

The solutions may often be computed in different ways and when two solutions are given it does not necessarily mean that more solutions does not exist. However, when two solutions are given we encourage you to think about why these two solutions are equivalent.

```
library(mosaic)
```

### Exercise 8.5

**a)**

```
table8.19 <- matrix(c(860, 11800, 140, 87120), 2)
rownames(table8.19) <- c("BC.Yes", "BC.No")
colnames(table8.19) <- c("DT+", "DT-")
table8.19
```

```
##          DT+   DT-
## BC.Yes   860   140
## BC.No  11800 87120
```

```
round(prop.table(table8.19, margin = 1), 2)
```

```
##         DT+  DT-
## BC.Yes 0.86 0.14
## BC.No  0.12 0.88
```

The probability of DT+ given BC.Yes is:

$$\frac{860}{860+140} = 0.86,$$

and the probability of DT- given BC.No is:

$$\frac{87120}{11800+87120} = 0.88.$$

It seems a fair tool, as it catches a fairly large percentage of `BC.Yes`, while excluding most of the `BC.No`.

**b)**

```
round(prop.table(table8.19, margin = 2), 2)
```

```
##         DT+ DT-
## BC.Yes 0.07   0
## BC.No  0.93   1
```

Probability of BC.No given DT+ is

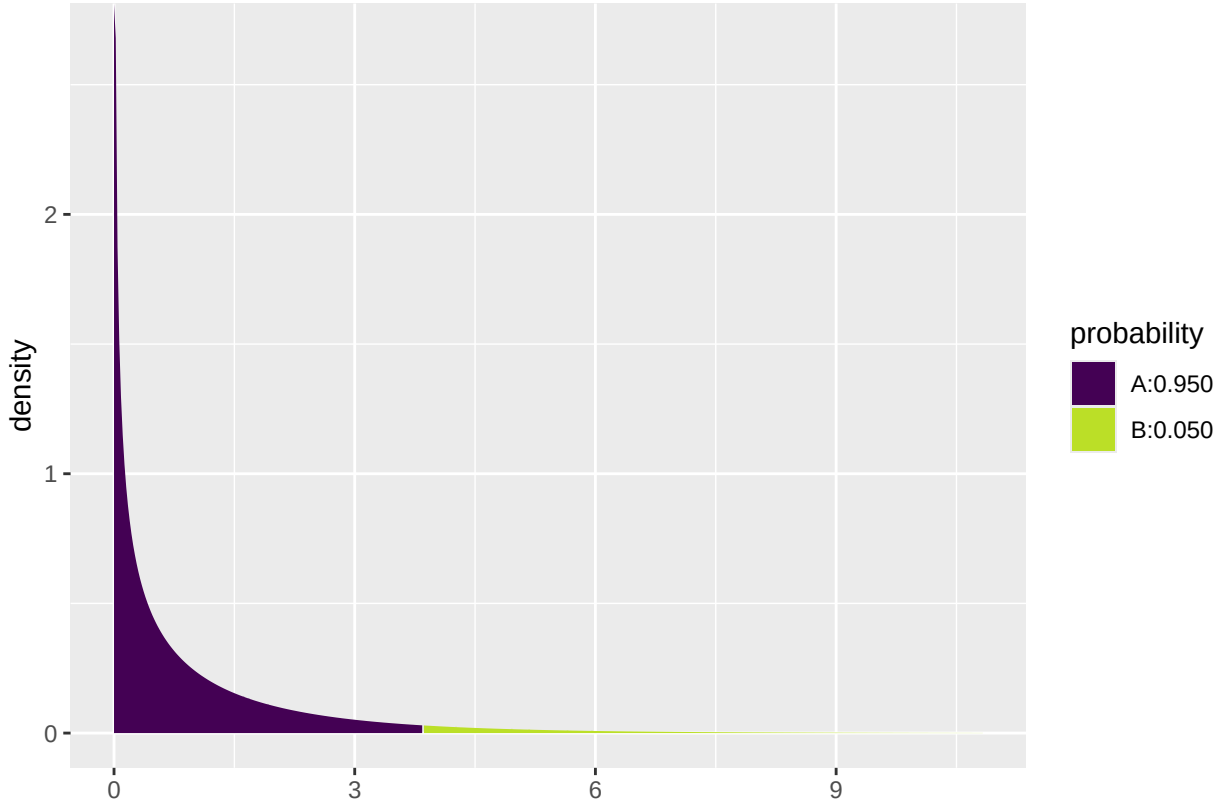
$$\frac{11800}{11800+860} = 0.93.$$

This is high due to the rarity of the disease, and will create a lot of unnessesary anxiety. However, it may be unaviodable due to the rarity.

**Exercise 8.7**

**a)**

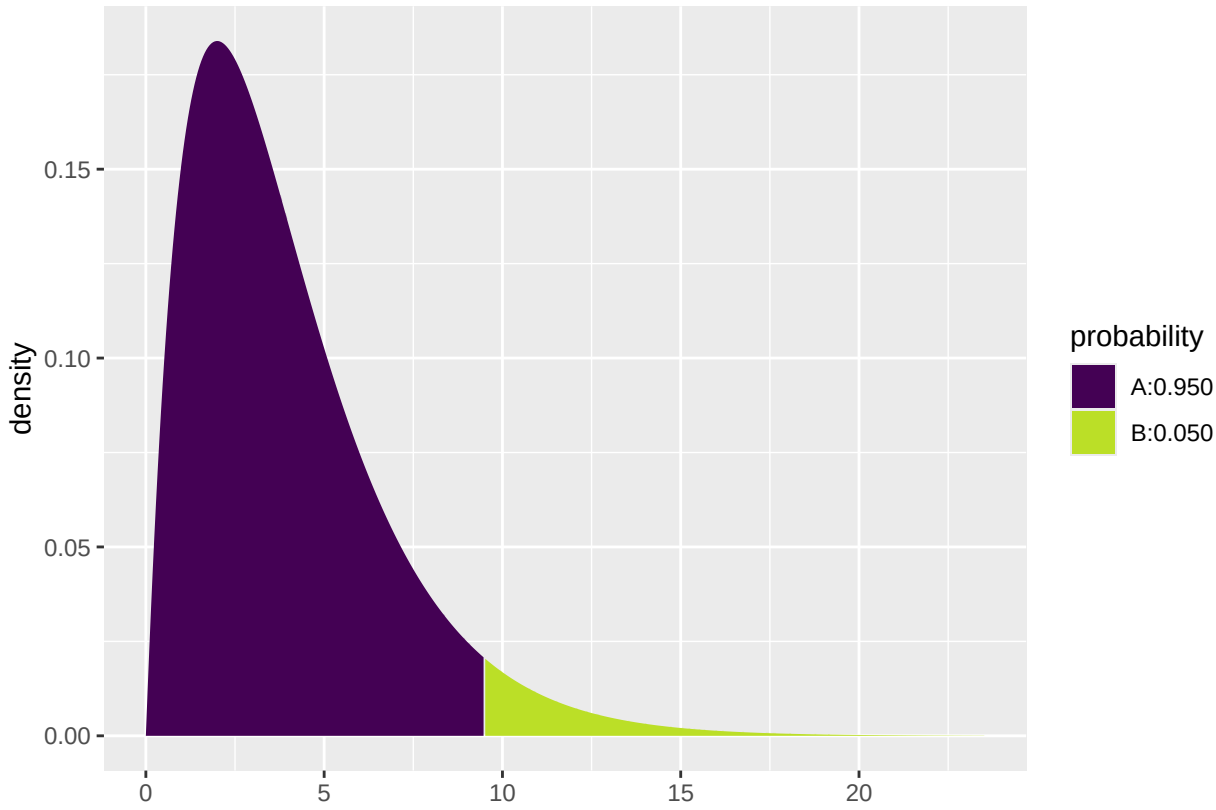
```
a <- c(2, 2)
qdist("chisq", p = 0.95, df = (a[1] - 1) * (a[2] - 1))
```




```
## [1] 3.841459
```

**b)**

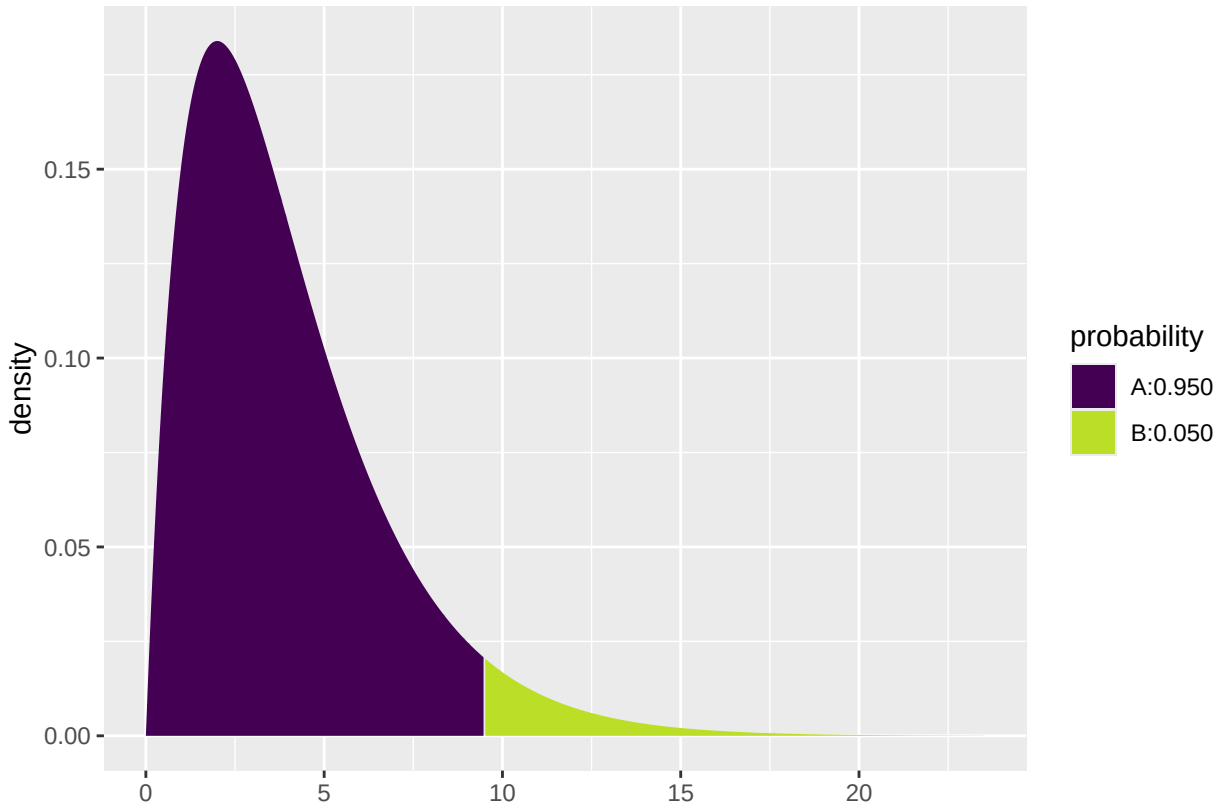
```
b <- c(3, 3)
qdist("chisq", p = 0.95, df = (b[1] - 1) * (b[2] - 1))
```




```
## [1] 9.487729
```

c)

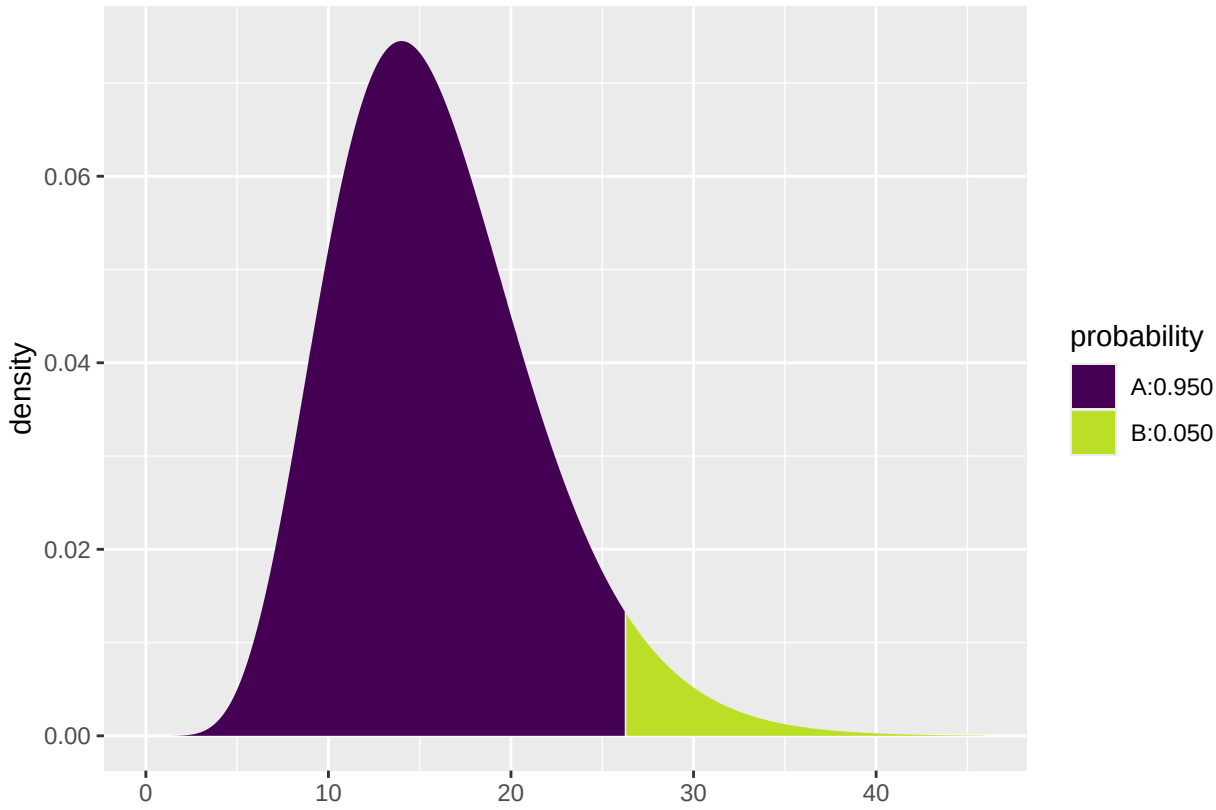
```
c <- c(2, 5)
qdist("chisq", p = 0.95, df = (c[1] - 1) * (c[2] - 1))
```




```
## [1] 9.487729
```

**d)**

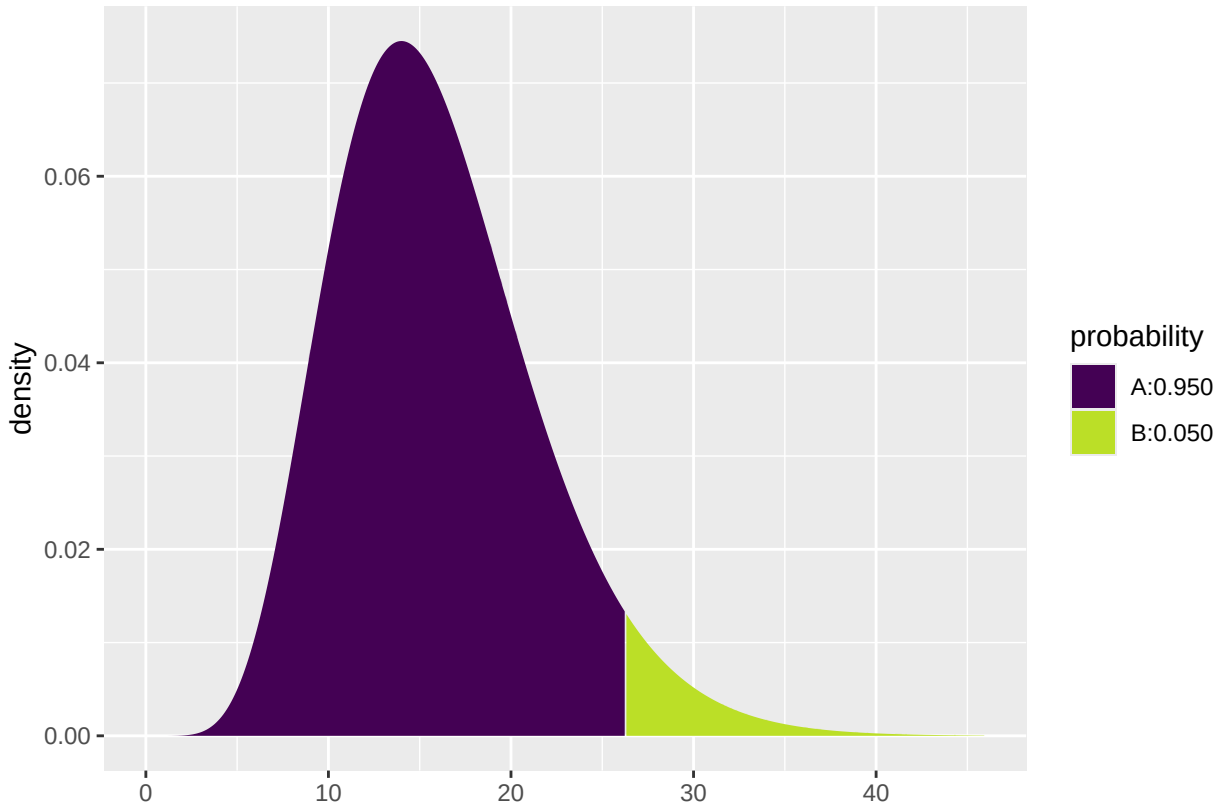
```
d <- c(5, 5)
qdist("chisq", p = 0.95, df = (d[1] - 1) * (d[2] - 1))
```




```
## [1] 26.29623
```

e)

```
e <- c(3, 9)
qdist("chisq", p = 0.95, df = (e[1] - 1) * (e[2] - 1))
```

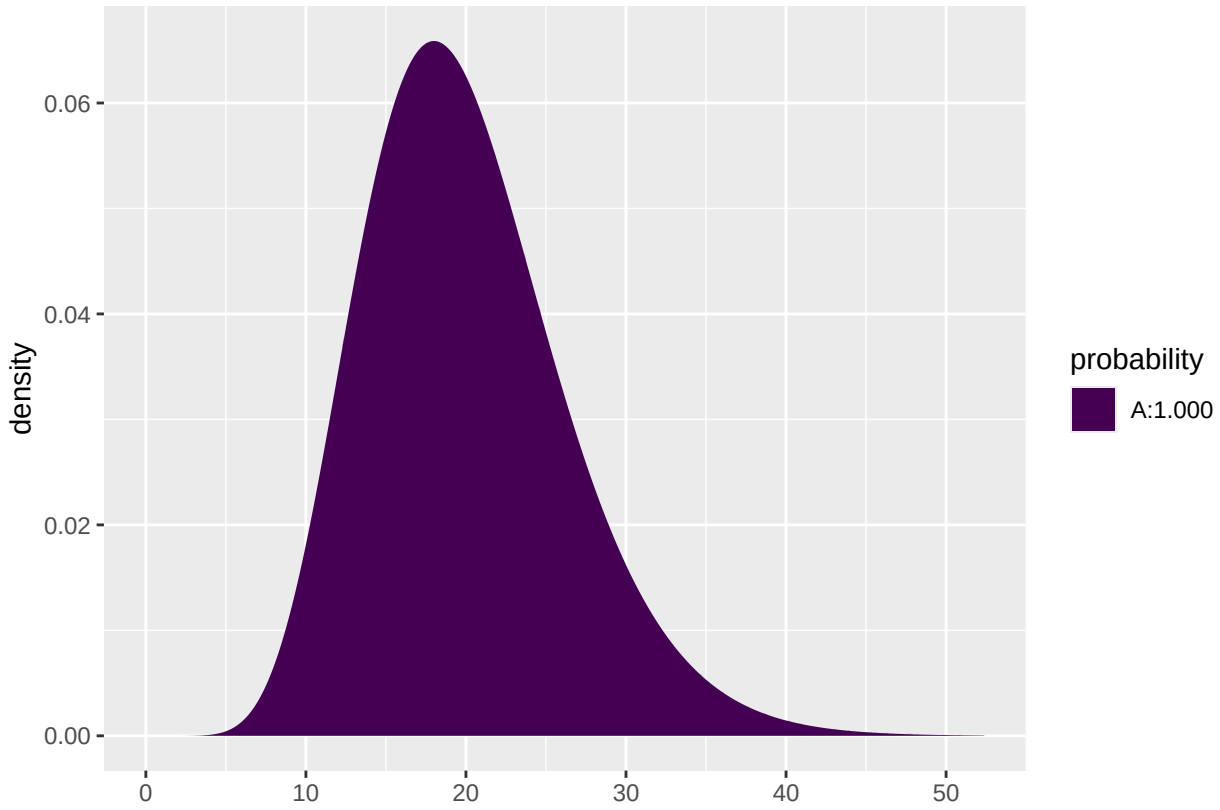



```
## [1] 26.29623
```

**Exercise 8.16**

**a)**

```
1 - pdist("chisq", q = 5889.27, df = (6 - 1) * (5 - 1))
```




```
## [1] 0
```

Which is extremely significant at a 5% significance level.

**b)** Standardized residuals (z-scores) outside $-3$ and $+3$ show highly significant diffence between observed and expected frequency.

(i) `None` are addicted to the internet.
(ii) `Partner died` abandon the internet.

**c)**

```
data <- matrix(c(93,311,977-93,3257-195), 2)
rownames(data) <- c("Married", "Divorced")
colnames(data) <- c("most+", "most-")
data
```

```
##          most+ most-
## Married     93   884
## Divorced   311  3062
```

```
testStat <- chisq.test(data)
testStat
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 0.048673, df = 1, p-value = 0.8254
```

```
testStat$res
```

```
##               most+       most-
```

```
## Married   0.2375203 -0.07599988
## Divorced -0.1278321  0.04090270
```

**Exercise "Flight anxiety"**

```
data <- matrix(c(860, 830, 140, 170), 2)
rownames(data) <- c("2005", "2007")
colnames(data) <- c("Safe+", "Safe-")
data
```

```
##      Safe+ Safe-
## 2005   860   140
## 2007   830   170
```

```
testStat <- chisq.test(data)
testStat$expected
```

```
##      Safe+ Safe-
## 2005   845   155
## 2007   845   155
```

```
testStat
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  data
## X-squared = 3.2105, df = 1, p-value = 0.07317
```

No significant change from 2005 to 2007, at a 5% significance level, in the amount of anxiety experienced due to flying.