

Exercise 5 - solution

Exercise A

Question 1

```
power.prop.test(p1=0.40,p2=0.60,power=0.85)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##              n = 110.6668
##              p1 = 0.4
##              p2 = 0.6
##      sig.level = 0.05
##              power = 0.85
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

We could include 111 women in each treatment group, hence we could include 222 women in total. This is enough to have 85% power if the chances of pregnancy are indeed 40% and 60% in the two groups. However, the power might be considerably less if the treatment turns out to be less effective than we thought. It would be a pity to make a type-II error, that is, to not conclude that the data show enough evidence to conclude that the treatment works, if the treatment works reasonably well, although maybe less well than we anticipated. Hence the result to the calculation made at the next questions are important to make a good, well-informed, decision about the sample size.

Question 2

```
power.prop.test(n=111,p1=0.40,p2=0.5)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##              n = 111
##              p1 = 0.4
##              p2 = 0.5
```

```
##      sig.level = 0.05
##      power = 0.3210212
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

The power of the study drops to only 32%, if we include 222 women (111 per group) and if the treatment results in only 50% chance of pregnancy. This is important to know. If we think that a treatment that leads to 50% chance of pregnancy is still a “successful” treatment worth prescribing, we might worry about the sample size of 222 women. With this sample size, it is unlikely that we can show that the treatment “works”, if it works but leads to only 50% chance of pregnancy.

```
power.prop.test(n=111,p1=0.40,p2=0.55)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##      n = 111
##      p1 = 0.4
##      p2 = 0.55
##      sig.level = 0.05
##      power = 0.6106339
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

The power of the study drops to 61%, if we include 222 women (111 per group) and if the treatment results only in 55% chance of pregnancy. This is important to know. After all, 55% is not very different from our expectation of 60%, so it might not be completely unlikely. Is it wise, and ethical, to choose a sample size of 222 women knowing this result? Only the investigators can know about that, after balancing the pros and cons, and thinking about the financial, logistical and ethical constraints they have. The investigators are expected to take an informed decision about the sample size accordingly.

```
power.prop.test(n=111,p1=0.40,power=0.75)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##      n = 111
##      p1 = 0.4
##      p2 = 0.5760567
##      sig.level = 0.05
##      power = 0.75
##      alternative = two.sided
##
```

NOTE: n is number in *each* group

The smallest improvement in chance of pregnancy that leads to a decent power of 75% or more, with this sample size of 222 women, is 18%, i.e. 58% chance of pregnancy with treatment versus 40% without treatment (if the chance of pregnancy without treatment is indeed 40%). Again, we believe it is important to know to take an informed decision about the sample size of the study. If we think that a treatment leading to less than 58% chances of pregnancy can still be thought as a “successful” treatment, then we could consider increasing the sample size. Indeed, otherwise we have a somewhat low power (less than 75%) to have a significant result with such a treatment.

To sum-up, several sample size and power calculations are often interesting to perform, to best understand the consequences of the choice of a specific sample size. The main aim of this exercise was to illustrate that.

Note that it is often interesting to choose the sample size which is the result of using not what we expect in the treatment group as input in the sample size calculation (60% in this exercise), but instead the “least good result” that we would still be happy to have (e.g., 50%). But, ethical, financial and logistical constraints are also important to take into account, such that we do not always proceed like this. Planning a study is not all about statistics, although statistics have an important role to play!

Exercise B

Part I

Question 1

We first load the data and look at the “summary”, as always.

```
load(url("http://paulblanche.com/files/smoking.rda"))
d <- smoking
summary(d)
```

```
##           age      smoker    death
##  below 45:628   no :732   no :945
##  45-54   :208   yes:582   yes:369
##  55-64   :236
##  above 65:242
```

We can see that 945 women among the 1314 were still alive 20 years after the initial survey. There were 582 smokers among the 1314.

Question 2 (2.a & 2.b)

We first produce a 2 by 2 table.

```
table(smoker=d$smoker,death=d$death)
```

```
##      death
## smoker no yes
##    no  502 230
##    yes 443 139
```

We observe:

- 502 non-smokers alive at 20 years after the initial survey
- 230 non-smokers dead within 20 years after the initial survey
- 443 smokers alive at 20 years after the initial survey
- 139 smokers dead within 20 years after the initial survey

This gives the following 20-year risk estimates for smokers and non-smokers:

```
230/(502+230) # 20-year risk of death among non smokers
```

```
## [1] 0.3142077
```

```
139/(443+139) # 20-year risk of death among smokers
```

```
## [1] 0.2388316
```

That is, 31.4% for non smokers and 23.9% for smokers. These results seem surprising: we know that smoking is unhealthy, hence we expected a higher risk of death for smokers.

Question 2 (2.c & 2.d)

We now compute the corresponding “exact” binomial confidence intervals.

```
binom.test(x=230,n=502+230) # non smokers
```

```
##
## Exact binomial test
##
## data: 230 and 502 + 230
## number of successes = 230, number of trials = 732, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.2807031 0.3492176
## sample estimates:
## probability of success
## 0.3142077
```

```
binom.test(x=139,n=443+139) # smokers
```

```
##
## Exact binomial test
##
```

```
## data: 139 and 443 + 139
## number of successes = 139, number of trials = 582, p-value < 2.2e-16
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.2047323 0.2756061
## sample estimates:
## probability of success
## 0.2388316
```

We therefore have the following 20-year risk estimate and 95% CI:

- Non-smokers: 31.4% (28.1 ; 34.9)
- Smokers: 23.9% (20.5 ; 27.6)

The 95% confidence intervals do not overlap, hence the confidence intervals “suggest” that the direction of the result, that is, a higher risk of death for non-smokers, is not due to small sample random variation.

Question 3

We now perform a statistical hypothesis test to compare the risk among smokers and non-smokers. We choose to use the Fisher’s exact test because we prefer to obtain “exact” results to approximate results when we can. However, the large sample size can justify to use a large sample method such as the Pearson Chi-square test.

```
fisher.test(table(smoker=d$smoker,death=d$death))
```

```
##
## Fisher's Exact Test for Count Data
##
## data: table(smoker = d$smoker, death = d$death)
## p-value = 0.002989
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.5307485 0.8822128
## sample estimates:
## odds ratio
## 0.6850392
```

We obtain a p-value=0.003, smaller than 5%, hence we conclude to a significant association between smoking and 20-year risk of death.

Just out of curiosity, we also compute the p-value of the Pearson Chi-square test.

```
chisq.test(table(smoker=d$smoker,death=d$death))
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  table(smoker = d$smoker, death = d$death)
## X-squared = 8.7515, df = 1, p-value = 0.003093
```

We see that the p-value is almost the same, which is due to the large sample size.

Questions 4 & 5

We now use the fonction `table2x2()` from the **Publish** package to estimate several association measures, with confidence intervals. We start with the risk difference:

```
d$death<-relevel(d$death, ref="yes")
Tab1 <- table(smoker=d$smoker,death=d$death)
library(Publish)
```

```
## Loading required package: prodlim
```

```
table2x2(Tab1,stats = c("table","rd"))
```

```
## -----
##
## 2x2 contingency table
## -----
##
##           deathyes      deathno      Sum
## smokerno          230          502      732
## smokeryes          139          443      582
## --              --              --
## Sum              369          945     1314
##
## -----
##
## Statistics
## -----
##
## a= 230
## b= 502
## c= 139
## d= 443
##
## p1=a/(a+b)= 0.3142
## p2=c/(c+d)= 0.2388
##
## -----
##
## Risk difference
## -----
```

```
##
## Risk difference = RD = p1-p2 = 0.07538
## Standard error = SE.RD = sqrt(p1*(1-p1)/(a+b)+p2*(1-p2)/(c+d)) = 0.02463
## Lower 95%-confidence limit: = RD - 1.96 * SE.RD = 0.0271
## Upper 95%-confidence limit: = RD + 1.96 * SE.RD = 0.1237
##
## The estimated risk difference is 7.5% (CI_95%: [2.7;12.4]).
```

We can first check that the probability of death estimates match the previous results. We have:

- Non-smokers: 31.4%
- smokers: 23.9%

We finally conclude that the estimated risk difference is 7.5% (95% CI= [2.7;12.4]), with higher risk for non-smokers.

We now compute the ratio of the risk.

```
table2x2(Tab1,stats = c("rr"))
```

```
##
## -----
##
## Risk ratio
## -----
##
## Risk ratio = RR = p1/p2 = 1.3156
## Standard error = SE.RR = sqrt((1-p1)/a+(1-p2)/c)= 0.09197
## Lower 95%-confidence limit: = RR * exp(- 1.96 * SE.RR) = 1.0986
## Upper 95%-confidence limit: = RR * exp(1.96 * SE.RR) = 1.5755
##
## The estimated risk ratio is 1.316 (CI_95%: [1.099;1.575]).
```

The estimated risk ratio is 1.316 (CI_95%: [1.099;1.575]). Equivalently, we can also say that the chance of dying is 31.6% (1-1.316) higher for non smokers than for smokers.

Finally, we now compute the odds ratio.

```
table2x2(Tab1,stats = c("or"))
```

```
##
## -----
##
## Odds ratio
## -----
##
## Odds ratio = OR = (p1/(1-p1))/(p2/(1-p2)) = 1.4602
## Standard error = SE.OR = sqrt((1/a+1/b+1/c+1/d)) = 0.1257
```

```
## Lower 95%-confidence limit: = OR * exp(- 1.96 * SE.OR) = 1.1414
## Upper 95%-confidence limit: = OR * exp(1.96 * SE.OR) = 1.8680
##
## The estimated odds ratio is 1.460 (CI_95%: [1.141;1.868]).
```

The estimated odds ratio is 1.460 (CI_95%: [1.141;1.868]).

Question 6

We now produce a barplot to compare the number of included women in each age group, between smokers and non smokers.

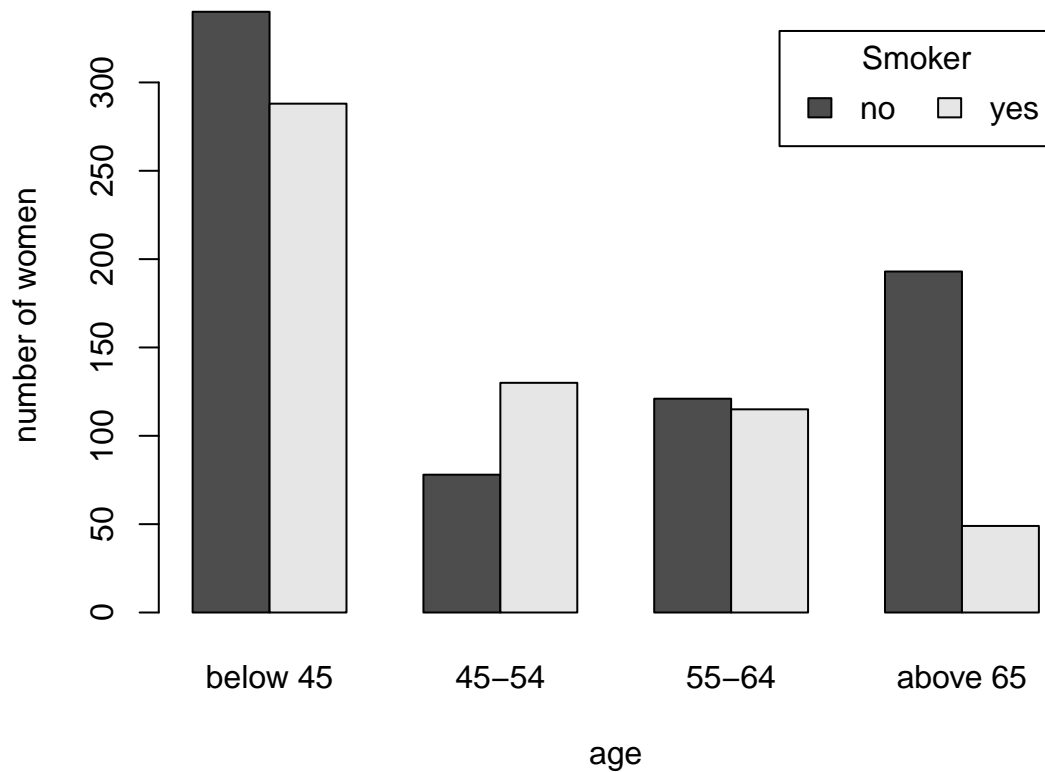
We first compute the frequency table.

```
Tab2 <- table(d$smoker,d$age)
Tab2
```

```
##
##      below 45 45-54 55-64 above 65
## no      340   78   121    193
## yes     288  130   115     49
```

And we now plot these counts.

```
barplot(Tab2,beside=TRUE,
        xlab="age",
        ylab="number of women",
        legend=TRUE,
        args.legend=list(title="Smoker",ncol=2))
```

Interestingly, we observe that very few women above 65 are smokers. Although the number of smokers and non-smokers is not the same also in the other age groups, the differences are not as large.

Question 7

We now produce a barplot to compare the 20-year risk of death in each age group.

We first compute the frequency table, and then compute the risks of death.

```
Tab3 <- table(d$death,d$age) # counts
```

```
Tab3
```

```
##
##      below 45 45-54 55-64 above 65
## yes       32   39   91   207
## no       596  169  145   35
```

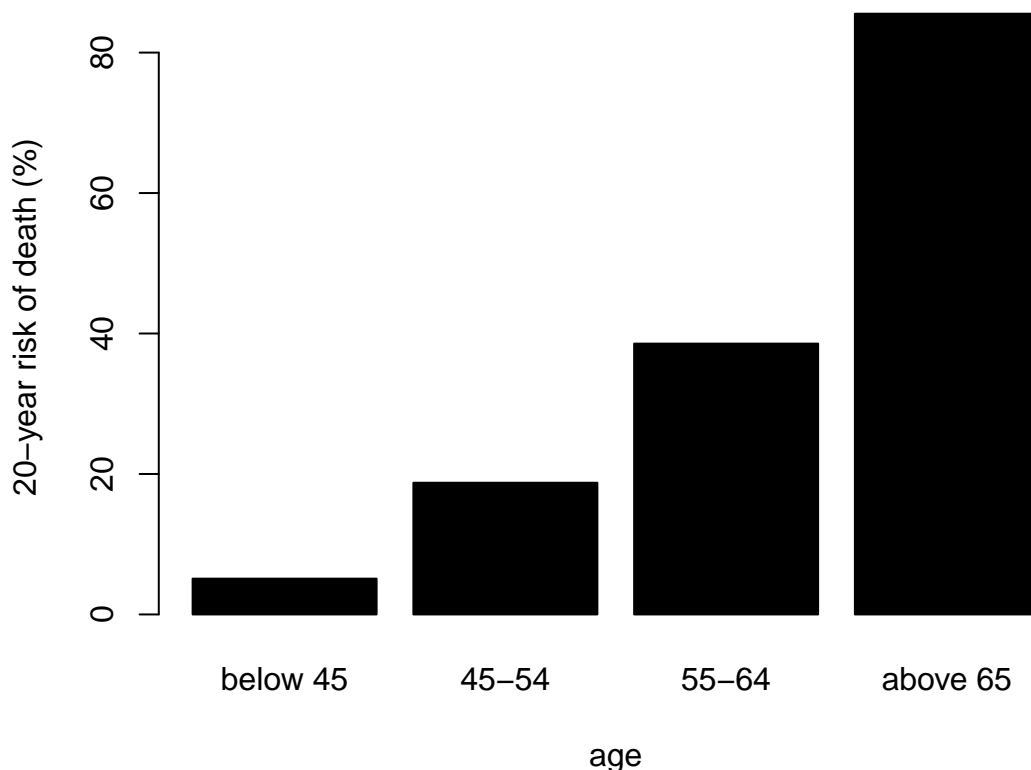
```
Tab4 <- prop.table(Tab3,margin=2) # proportions per age
```

```
Tab4
```

```
##
##      below 45      45-54      55-64      above 65
## yes 0.05095541 0.18750000 0.38559322 0.85537190
## no  0.94904459 0.81250000 0.61440678 0.14462810
```

We are now ready to plot the 20-year risk of death in each age group.

```
barplot(100*Tab4[1,],
        xlab="age",
        ylab="20-year risk of death (%)",
        col="black")
```



We can observe that the older the women, the more likely they die 20 years. This seems to make perfect sense.

Question 8

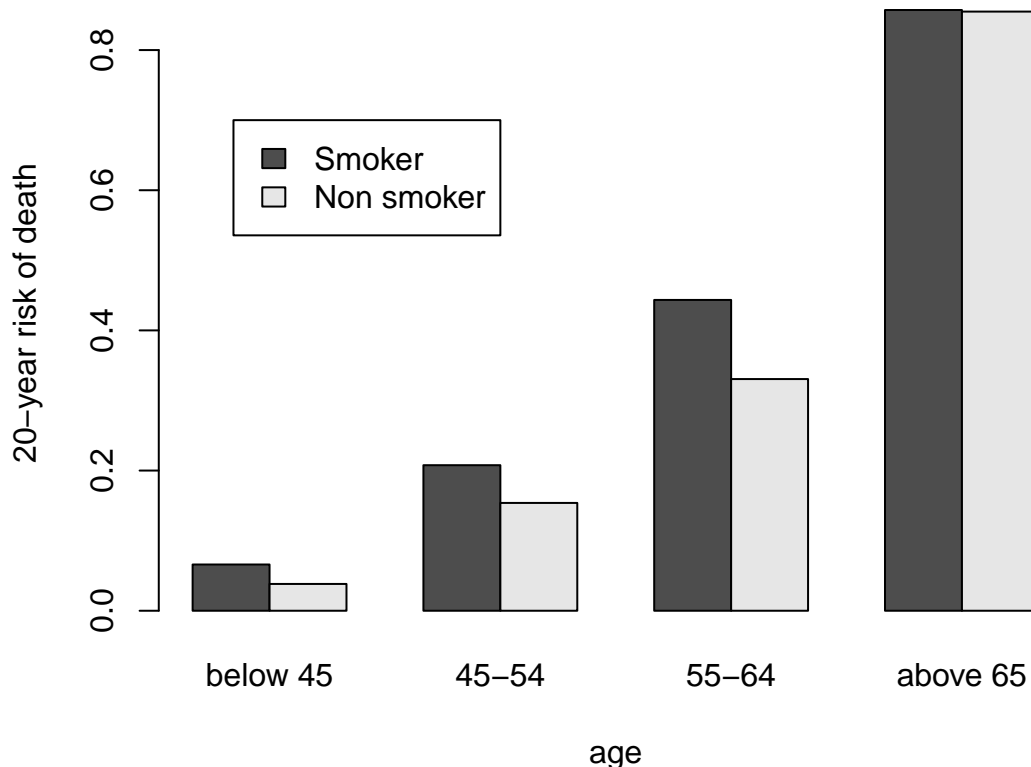
We now produce a barplot to compare the 20-year risk in each age group, between smokers and non smokers. First, we compute the probabilities in each age group.

```
Tab3smokers <- table(d$death[d$smoker=="yes"],
                    d$age[d$smoker=="yes"]) # counts for smokers
Tab4smokers <- prop.table(Tab3smokers,
                          margin=2) # proportions per age for smokers
Tab3NonSmokers <- table(d$death[d$smoker=="no"],
                       d$age[d$smoker=="no"]) # counts for non smokers
Tab4NonSmokers <- prop.table(Tab3NonSmokers,
                             margin=2) # proportions per age for non smokers
# merge the results into one unique matrix
Tab5 <- rbind(Tab4smokers[1,], Tab4NonSmokers[1,])
rownames(Tab5) <- c("Smoker", "Non smoker")
Tab5
```

```
##           below 45    45-54    55-64  above 65
## Smoker    0.06597222 0.2076923 0.4434783 0.8571429
## Non smoker 0.03823529 0.1538462 0.3305785 0.8549223
```

We are now ready to produce the plot.

```
barplot(Tab5,
        beside=TRUE,
        xlab="age",
        ylab="20-year risk of death",
        legend=TRUE, args.legend = list(x=5, y=0.7)) # to locate the legend
```



We see observe that in all age groups, the 20-year risk of death is higher for smokers than for non-smokers. This makes sense, we know that smoking is not healthy. We also observe that there is a substantial difference in 20-year risk between age groups. This makes sense, the older a women, the more likely she cannot survive 20 years. Finally, we observe that the differences in 20-year risk between age groups are substantially larger than the differences in 20-year risk between smoker and non smokers, when comparing women of the same age group. This makes sense: smoking is bad, but not as bad as becoming old (for the 20-year risk of death).

Question 9

We have seen that most smokers are younger than non-smokers in this data set. Especially, most of the women older than 65 are non-smokers. Because we have also seen that the 20-year risk of death is much higher for women older than 65 than the others (as expected), we might

suspect that the risk difference between smokers and non-smokers is mostly driven by the fact that those who do not smoke are older.

Smoking is known to be unhealthy, but it is probably not as life-threatening as becoming “old”, which is what suggests the plot obtained at question 8. Hence the results of question 2: when comparing young smokers to old non-smokers, the data suggest that old non-smokers die more often.

Part II

Question 10

Below we exemplify how to do it for the age group 45-54.

```
Tab1a <- table(smoker=d$smoker[d$age=="45-54"],death=d$death[d$age=="45-54"])
table2x2(Tab1a,stats = c("table","rr"))
```

```
## -----
##
## 2x2 contingency table
## -----
##
##           deathyes      deathno      Sum
## smokerno           12          66       78
## smokeryes          27         103      130
## --              --          --       --
## Sum              39         169      208
##
## -----
##
## Statistics
## -----
##
##
## a= 12
## b= 66
## c= 27
## d= 103
##
## p1=a/(a+b)= 0.1538
## p2=c/(c+d)= 0.2077
##
## -----
##
## Risk ratio
## -----
```

```
##
## Risk ratio = RR = p1/p2 = 0.7407
## Standard error = SE.RR = sqrt((1-p1)/a+(1-p2)/c)= 0.3160
## Lower 95%-confidence limit: = RR * exp(- 1.96 * SE.RR) = 0.3987
## Upper 95%-confidence limit: = RR * exp(1.96 * SE.RR) = 1.3761
##
## The estimated risk ratio is 0.741 (CI_95%: [0.399;1.376]).
```

We can proceed similarly for the other age groups and obtain:

- Age group below 45: estimated risk ratio 0.580 (CI_95%: [0.291;1.153]).
- Age group 45-54: estimated risk ratio 0.741 (CI_95%: [0.399;1.376]).
- Age group 55-64: estimated risk ratio 0.745 (CI_95%: [0.538;1.033]).
- Age group above 65: estimated risk ratio 0.997 (CI_95%: [0.877;1.134]).

There is no significant difference in any age group, although there seems to be a systematic trend towards a lower 20-year risk of death for non-smokers. The fact that the results are not significant is probably due to a lack of power. The sample size of each age group is not very large.

Question 11

The results of Part II can be thought as more interesting because the two groups that we compare (smokers versus non smokers) are expected to be more similar with respect to everything but smoking, as the women of the two groups have a similar age. Hence we can expect that the association between smoking and death that we estimated are closer to causal associations than those of Part I, although we cannot rigourously claim that they are indeed causal.

Exercise C

Question 1

We first load the data and look at the first rows.

```
rm(list=ls())    # clear memory (to start fresh for the new exercise)
load(url("http://paulblanche.com/files/sedative.rda"))
head(sedative)
```

```
##   zolpidem doxepin
## 1      no      no
## 2      no      no
## 3      no      no
## 4      no      no
## 5      no      no
## 6      no     yes
```

```
d <- sedative # data in an object with a shorter name, for convenience
```

Question 2

First, we tabulate the data from the night after receiving zolpidem.

```
table(d$zolpidem)
```

```
##  
## no yes  
## 33 19
```

We can read that 33 out of $33+19=52$ subjects did not wake before receiving the loudest tone (110 dB).

```
prop.table(table(d$zolpidem))
```

```
##  
## no yes  
## 0.6346154 0.3653846
```

This corresponds to $33/(19+33)=63.5\%$ of the subjects. We proceed similarly with the data from the night after receiving doxepin.

```
table(d$doxepin)
```

```
##  
## no yes  
## 9 43
```

```
prop.table(table(d$doxepin))
```

```
##  
## no yes  
## 0.1730769 0.8269231
```

We read that 9 out of 52 subjects did not wake before receiving the loudest tone (110 dB), that is 17.3%.

Question 3

We compute 95% exact binomial confidence intervals for the risk of not waking before receiving the loudest tone (110 dB), after receiving each treatment.

```
binom.test(x=33,n=52)
```

```
##  
## Exact binomial test  
##
```

```
## data: 33 and 52
## number of successes = 33, number of trials = 52, p-value = 0.07039
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.4895590 0.7637564
## sample estimates:
## probability of success
## 0.6346154
```

```
binom.test(x=9,n=52)
```

```
##
## Exact binomial test
##
## data: 9 and 52
## number of successes = 9, number of trials = 52, p-value = 2.038e-06
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.08232569 0.30328011
## sample estimates:
## probability of success
## 0.1730769
```

We read the 95% confidence intervals: [49.0%; 76.4%] for zolpidem and [8.2% ; 30.3%] for doxepin.

Note that to round the values of the upper and lower limits, we can use this code:

```
round(100*binom.test(x=9,n=52)$conf.int,1)
```

```
## [1] 8.2 30.3
## attr(,"conf.level")
## [1] 0.95
```

Question 4

Here we want to compare two proportions obtained from the **same** subjects. In other words, we want to compare two sets of binary data which are **paired**. Although we might first think about using a Fisher test or a Chi-square test when we aim to compare two proportions, here this would not be appropriate. This is because Fisher test and Chi-square tests assume that the two samples that we compare are independent. This is not the case here as the data come from two different nights for the **same** subjects. The data are paired per subject and not independent. Some subjects will tend to be fast asleep no matter the drug they take (or not) and some others tend to be light sleepers, also no matter the drug they take. To analyze such **paired data** the appropriate analogue to the chi-square test is the **McNemar test**.

Question 5

We now compute the p-value from a **McNemar test** test. We prefer the “exact” version of the test to a (maybe more commonly used) version based on a large sample approximation. After all, why approximate when you can get the exact?

The aim when using this test is to understand whether the data of the two nights are sufficiently different to confidently conclude that one treatment leads to a higher risk of not waking than the other.

```
Tabcompare <- table(zolpidem=d$zolpidem,
                    doxepin=d$doxepin)
Tabcompare
```

```
##           doxepin
## zolpidem no yes
##      no    5  28
##      yes    4  15
```

We can now proceed with the `mcnemar.exact` function of the package `exact2x2`.

```
library(exact2x2)

## Warning: package 'exact2x2' was built under R version 4.4.3
## Loading required package: exactci
## Warning: package 'exactci' was built under R version 4.4.3
## Loading required package: ssanv
## Loading required package: testthat
## Warning: package 'testthat' was built under R version 4.4.3
mcnemar.exact(Tabcompare)

##
## Exact McNemar test (with central confidence intervals)
##
## data: Tabcompare
## b = 28, c = 4, p-value = 1.93e-05
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  2.448889 27.465170
## sample estimates:
## odds ratio
##           7
```

We read that the p-value is very small: $p < 0.001$. Hence the data contains enough evidence to confidently conclude that the risk of not waking are different after receiving each treatment.

Usually it is unnecessary to provide the exact value of a p-value when it is smaller than 0.001. Reporting the exact value of the p-value, i.e., 1.93e-05, will often look a bit “foolish” in a scientific paper.

A possible conclusion sentence is: “*The risk of not waking after receiving zolpidem (63.5%, 95%-CI=[49.0%; 76.4%]) was found significantly larger than that after receiving doxepin (17.3%, 95%-CI=[8.2% ; 30.3%], p-value<0.001).*”

Question 6

In the abstract of Drake et al (2017), we can read: “*A significantly greater proportion of participants in the zolpidem condition (63.5%) did not wake until receiving the loudest tone (110 dB) as compared to the doxepin (17.6%) and placebo conditions (17.3%, 5.8%).*” We can notice that 63.5% and the significant result are similar to those we found with our own analyses above. We can be happy to have successfully reproduced their results. There is, however, a tiny difference of 17.6% instead of 17.3% in the result for doxepin. We suspect that this is a minor typo because 17.3% is actually reported in the result section ‘Primary Outcomes’, subsection ‘Auditory Awakening Threshold’, on page 4 of the paper.