

Solutions: Inference on Proportions in R Exercises

BIOST 514/517

Week 6

Question 1

```
inflamm <- read.table("../inflamm.txt",header=TRUE)
dat <- inflamm[!is.na(inflamm$smoker),]
```

Question 2

Expected code:

```
xNosmok <- sum(!dat$smoker & dat$death)
nNosmok <- sum(!dat$smoker)
pNosmok <- xNosmok/nNosmok

xSmok <- sum(dat$smoker & dat$death)
nSmok <- sum(dat$smoker)
pSmok <- xSmok/nSmok
```

Advanced code:

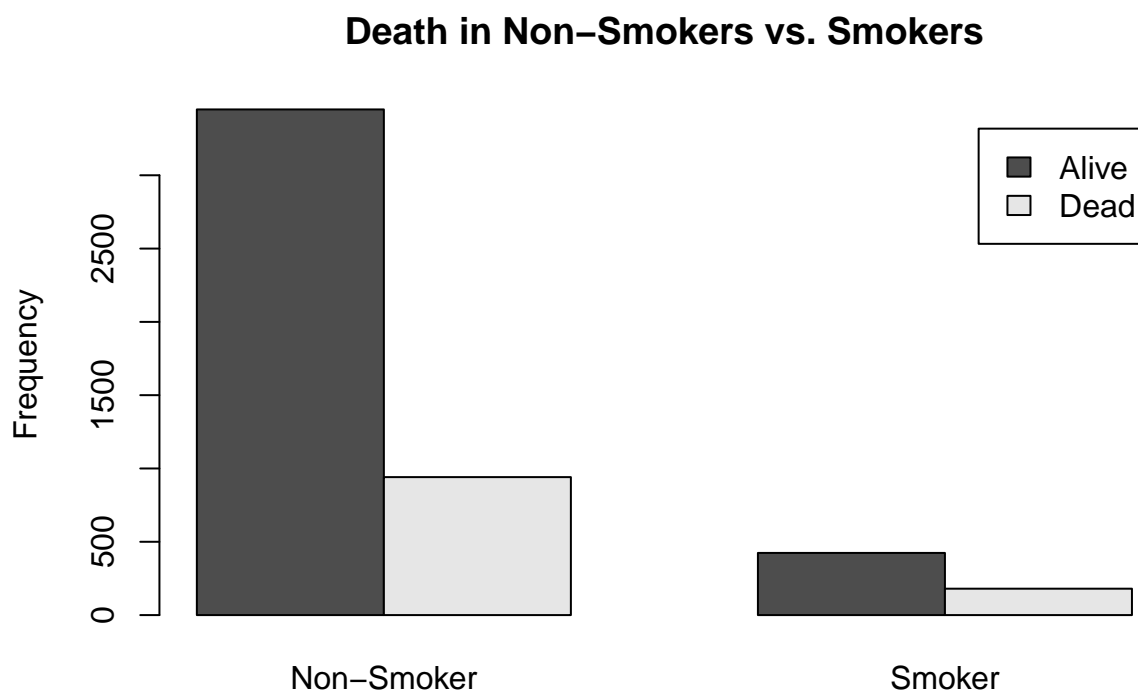
```
tab1 <- table(dat$death,dat$smoker) # rows correspond to death, columns to smoking status
tab1
```

```
##
##      0      1
## 0 3449  424
## 1   94   180
```

The frequency and sample proportion of death among non-smokers is 941 and 0.21, respectively, and the frequency and sample proportion of death among smokers is 180 and 0.3.

Bonus visualization:

```
barplot(tab1,beside=TRUE,
        main=c("Death in Non-Smokers vs. Smokers"),
        ylab="Frequency",
        names.arg = c("Non-Smoker","Smoker"),
        legend.text=c("Alive","Dead"))
```



Question 3

```
(nSmok * pSmok) > 10 & (nSmok * (1-pSmok)) > 10
```

```
## [1] TRUE
```

By the rule of thumb, we seem to have a sufficient sample size to approximate the distribution of the sample proportion with a Normal distribution.

Question 4

```
hiTestCorrect <- prop.test(x=xSmok,n=nSmok)
hiTestCorrect$conf.int
```

```
## [1] 0.2621128 0.3365245
## attr(,"conf.level")
## [1] 0.95
```

Our results are consistent with the true proportion of death among smokers being between 0.26 and 0.34.

Question 5

```
diffTestCorrect <- prop.test(x=c(xNosmok,xSmok),
                             n=c(nNosmok,nSmok))
```

```
diffTestCorrect$p.value
```

```
## [1] 4.91493e-06
```

With the small p-value, $P < 0.001$, we have strong evidence to reject the null hypothesis that the proportion of death is the same between smokers and non-smokers.