Week5_Assignement_Data607_Chunjie_Nan

Chunjie Nan

9/25/2021

Import data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
vaccine<-read.csv("https://raw.githubusercontent.com/nancunjie4560/Data607/master/israeli_vaccination_da

vaccine<-data.frame(vaccine,stringsAsFactors = FALSE)
head(vaccine)
```

```
##    Age Population..            X            Severe.Cases              X.1
## 1       Not Vax\n% Fully Vax\n% Not Vax\nper 100K\n\n\np Fully Vax\nper 100K
## 2 <50    1,116,834    3,501,118                      43                   11
## 3            23.3%       73.0%
## 4 >50      186,078    2,133,516                     171                  290
## 5             7.9%       90.4%
## 6
##              Efficacy
## 1 vs. severe disease
## 2
## 3
## 4
## 5
## 6
```

```
vaccine<-vaccine[1:5,]

vaccine<-rename(vaccine,'Age Range'='Age', 'Unvaccinated Population' = Population.. , 'Vaccinated Popula
```

Tidy data

```
vaccine<-vaccine[-1,]
vaccine$`Percentage of Unvaccinated`[1]=vaccine$`Unvaccinated Population`[2]
```

```r
vaccine$`Percentage of Unvaccinated`[3]=vaccine$`Unvaccinated Population`[4]
vaccine$`Percentage of Vaccinated`[1]=vaccine$`Vaccinated Population`[2]
vaccine$`Percentage of Vaccinated`[3]=vaccine$`Vaccinated Population`[4]
vaccine<-vaccine[-c(2,4),]
vaccine[] <- lapply(vaccine, gsub, pattern =",", replacement = "")
vaccine[] <- lapply(vaccine, gsub, pattern ="%", replacement = "")


data <- as.data.frame(sapply(vaccine, as.numeric),na.omit=T)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```r
data$`Age Range`=vaccine$`Age Range`
str(data)
```

```
## 'data.frame':    2 obs. of  8 variables:
##  $ Age Range                    : chr  "<50" ">50"
##  $ Unvaccinated Population       : num  1116834 186078
##  $ Vaccinated Population         : num  3501118 2133516
##  $ Unvaccinated Hospitalize per 100k: num  43 171
##  $ Vaccinated Hospitalize per 100k  : num  11 290
##  $ Efficacy vs Severe disease    : num  NA NA
##  $ Percentage of Unvaccinated    : num  23.3 7.9
##  $ Percentage of Vaccinated      : num  73 90.4
```

Caculation

(1) Do you have enough information to calculate the total population? What does this total population represent?

```r
population<-data%>%
  group_by(data[,c(1:3)])%>%
  summarise(total = sum(`Unvaccinated Population`+`Vaccinated Population`))
```

```
## `summarise()` has grouped output by 'Age Range', 'Unvaccinated Population'. You can override using th
```

```r
pop_lower50<-(population$total[1])/0.963
pop_lower50
```

```
## [1] 4795381
```

```r
pop_50above<-(population$total[2])/0.983
pop_50above
```

```
## [1] 2359709
```

```r
total_population<-sum(pop_lower50,pop_50above)
total_population
```

```
## [1] 7155090
```

Yes, I believe there are enough information to calculate the total of Israel population, and the total population is 7155090. first, I found that the population of age less than 50. The given information shows that 96.3% of Israel people either get fully vaccinated or not vaccinated at all. I figure out that 3.7% of population of whose age is less than 50 may get one vaccine shot. From these info I can calculate the total population of whose age is less than 50. The same work has done to calculate the population of whose age is more than 50. sum up these to rage of age, I can calculate the total population of Israel.

(2) Calculate the Efficacy vs. Disease; Explain your results. Efficacy vs. severe disease = 1 - (% fully vaxed severe cases per 100K / % not vaxed severe cases per 100K)

```
# Efficacy age less than 50
vac_hos_rate_less50<-data$`Vaccinated Hospitalize per 100k`[1]/data$`Vaccinated Population`[1] *100000
vac_hos_rate_less50
```

```
## [1] 0.3141854
```

```
unvac_hos_rate_less50<-data$`Unvaccinated Hospitalize per 100k`[1]/data$`Unvaccinated Population`[1] *1
unvac_hos_rate_less50
```

```
## [1] 3.850169
```

```
data$`Efficacy vs Severe disease`[1] = 1-vac_hos_rate_less50/unvac_hos_rate_less50
data$`Efficacy vs Severe disease`[1]
```

```
## [1] 0.918397
```

```
# Efficacy age 50 above
vac_hos_rate_50above<-data$`Vaccinated Hospitalize per 100k`[2]/data$`Vaccinated Population`[2] *100000
vac_hos_rate_50above
```

```
## [1] 13.59259
```

```
unvac_hos_rate_50above<-data$`Unvaccinated Hospitalize per 100k`[2]/data$`Unvaccinated Population`[2] *
unvac_hos_rate_50above
```

```
## [1] 91.89695
```

```
data$`Efficacy vs Severe disease`[2] = 1-vac_hos_rate_50above/unvac_hos_rate_50above
data$`Efficacy vs Severe disease`[2]
```

```
## [1] 0.8520888
```

Use the above Formula: Efficacy vs. severe disease = 1 - (% fully vaxed severe cases per 100K / % not vaxed severe cases per 100K)

Found that the Efficacy for age less than 50 is 91.837%, and the Efficacy for age more than 50 is 85.208%. These two numbers are pretty high that show the vaccinations are very effective according to the Efficacy vs. severe disease rate calculations.

(3) From your calculation of efficacy vs. disease, are you able to compare the rate of severe cases in unvaccinated individuals to that in vaccinated individuals?

Yes, from the Efficacy calculation, I can see following hospitalized patterns.

for those whose age is less than 50, 0.3% of them who has vaccinated are hospitalized. for those whose age is less than 50, 3.8% of them who hasn't vaccinated are hospitalized.

for those whose age is more than 50, 13.59% of them who has vaccinated are hospitalized. for those whose age is more than 50, 91.89% of them who hasn't vaccinated are hospitalized.

```
barplot(c('Hospital Rate age > 50'=unvac_hos_rate_50above, 'Hospital Rage age < 50'=unvac_hos_rate_less5
```
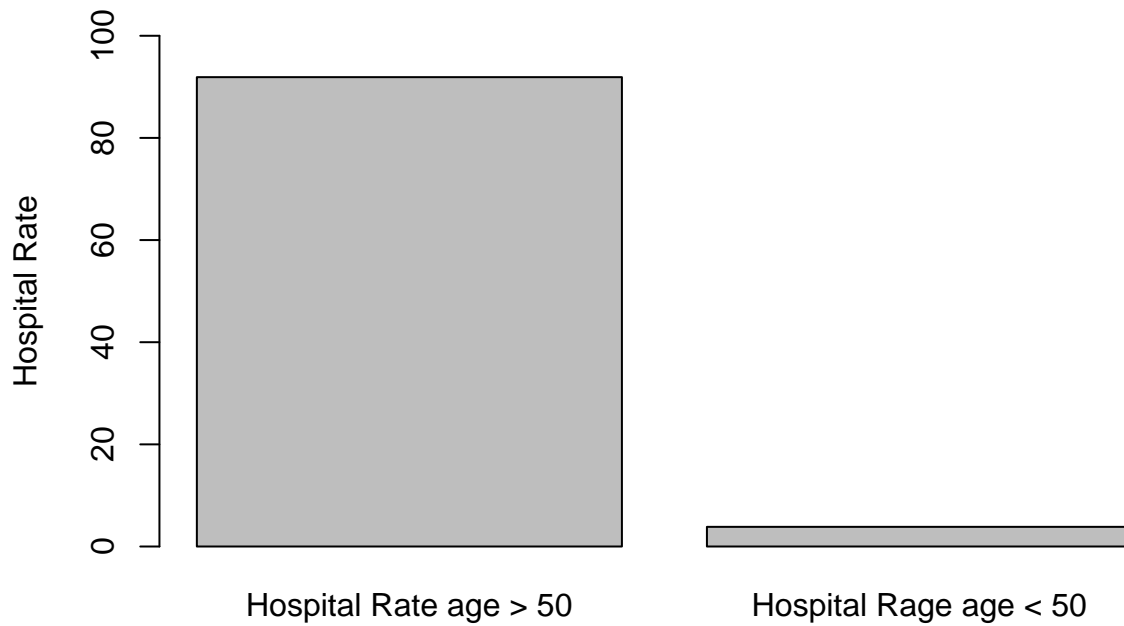
The vaccination works very effective for whose age more than 50, since hospitalization rate is dramatically decreased for those who vaccinated. However, the vaccination shows less difference of hospitalization for those whose age is less than 50. Even though the vaccination shows less difference for those whose age is less than 50, I believe everyone should get the vaccination to avoid spread to the people whose age is more than 50 because it is huge risk for older than 50 year old people.