Instructon

Consulting Group

2022-05-03

Import the target dara

```
# Get the data
df<-read.csv("consulting project.csv",header=TRUE)
```

Get the target diagnosis code and target procedure

```
diagnosistar<-c('S02.85', 'S02.3', 'S02.12', 'S02.84', 'S02.83','802.6','802.7')

proceduretar<-c('"16.0"', "16.01", '16.09', '16.8', '16.81', '16.89','0N5P', '0N5Q',
                '0N8P', '0N8Q', '0N9P', '0N9Q', '0NBP', '0NBQ', '0NCP', '0NCQ', '0NDP',
                '0NDQ', '0NHP', '0NHQ', '0NNP', '0NNQ', '0NQP', '0NQN', '0NRP', '0NRQ',
                '0NSP', '0NSQ', '0NTP', '0NTQ', '0NUP', '0NUQ')
```

Create the function

```
## Function for finding data with specific target code
get_candi<-function(target,df){
  tmp=startsWith(c(as.matrix(df)),target) # Use StartsWith function to detect all the observation with
  tmp2=data.frame(matrix(tmp,ncol=369))  # Transfer the result from previous step to data.frame
  tmp_index=apply(tmp2,1,sum,na.rm = T)>0 # Find which observation content target code and set them to
  re=df[tmp_index,] # Get all patient with TRUE index
  return(re) # Return the result
}
```

Get the number of patient with specific condition

Get the total number of patient

```
Num_of_patient<-nrow(df)
Num_of_patient
```

```
## [1] 318
```

Get the number for patient with target diagnosis code

```
l<-lapply(diagnosistar,get_candi,df=df)
Onlywithdiagnosis<-distinct(bind_rows(l))
Num_of_patients_with_diagnosis<-nrow(Onlywithdiagnosis)
Num_of_patients_with_diagnosis
```

```
## [1] 55
```

Get the number for with both target diagnosis code and procedure code

```
w<-lapply(proceduretar,get_candi,df=Onlywithdiagnosis)
Bothdiagnosis_and_procedure <-distinct(bind_rows(w))
Num_of_patients_with_diagnosis_and_procedure<-nrow(Bothdiagnosis_and_procedure)
Num_of_patients_with_diagnosis_and_procedure
```

```
## [1] 2
```

Get the number of patient with target procedure code

```
v<-lapply(proceduretar,get_candi,df=df)
Onlywithprocedure<-distinct(bind_rows(v))
Num_of_patients_with_procedure<-nrow(Onlywithprocedure)
Num_of_patients_with_procedure
```

```
## [1] 7
```

Get the number of patient with target procedure code but no target diagnosis code

```
Num_of_patient_with_procedure_no_daignosis<-Num_of_patients_with_procedure-Num_of_patients_with_diagnosi
Num_of_patient_with_procedure_no_daignosis
```

```
## [1] 5
```

Get the number of patient with target diagnosis but not target procedure

```
Num_of_patient_with_diagnosis_no_procedure<-Num_of_patients_with_diagnosis-Num_of_patients_with_diagnosi
Num_of_patient_with_diagnosis_no_procedure
```

```
## [1] 53
```

Get the number of patient with no target diagnosis

```
Num_of_patient_with_no_target_diagnosis<-Num_of_patient-Num_of_patients_with_diagnosis
Num_of_patient_with_no_target_diagnosis
```

```
## [1] 263
```

Get the number of patient with no target diagnosis and no target procedure

```
Num_of_patient_with_no_target_diagnosis_no_procedure<-Num_of_patient_with_no_target_diagnosis-Num_of_pa
Num_of_patient_with_no_target_diagnosis_no_procedure
```

```
## [1] 258
```

Get the number of patient with no target procedure

```
Num_of_patient_with_no_procedure<-Num_of_patient-Num_of_patients_with_procedure
Num_of_patient_with_no_procedure
```

```
## [1] 311
```

Create contingency table

```
data.frame('With target procedure' = c(2,5,7),
           'With no target procedure' = c(53,258,311),
           'total '=c(55,263,318),row.names = c('With target diagnosis',
                                                'With no target diagnosis','total'))
```

```
##                          With.target.procedure With.no.target.procedure total.
## With target diagnosis                        2                         53     55
## With no target diagnosis                     5                        258    263
## total                                        7                        311    318
```

Conditional Probability

```
prob<-2/55
prob
```

```
## [1] 0.03636364
```

Z-test part

We want to test if the population proportion (true proportion of orbital fracture patients who need surgery) is greater than a threshold 5%. For here$\beta_0 = 0.05$ and $\beta_1 > 0.05$

```
ztest<-prop.test(2, 55, p = 0.05, alternative = "greater", conf.level = 0.95, correct = TRUE)
ztest$p.value
```

```
## [1] 0.5614601
```

Because $0.56146 > 0.05$, so we can say this test is not significant, so cannot reject the null hypothesis.