pharynx.R

riserate

Sat Apr 1 20:17:11 2017

pharynx <- read.csv("pharynx.csv", header = TRUE)  
  
#1   
summary(pharynx$TIME)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 11.0 240.5 445.0 558.7 778.5 1823.0

#Want to see the pharynx variable   
pharynx$TIME1<-factor(NA,levels=c("Greater than or equal to 500", "less than 500 days"))   
#Create a variable that will say true when the person survived 500 times or more or less than 500 days   
pharynx$TIME1[pharynx$TIME>=500]<-"Greater than or equal to 500"   
#One level is that they are greater than 500   
pharynx$TIME1[pharynx$TIME<500]<-"less than 500 days"   
#One level is less than 500 days   
table(pharynx$TIME1)

##   
## Greater than or equal to 500 less than 500 days   
## 88 107

prop.test(table(pharynx$TIME1),0.50,correct=F)

##   
## 1-sample proportions test without continuity correction  
##   
## data: table(pharynx$TIME1), null probability 0.5  
## X-squared = 1.8513, df = 1, p-value = 0.1736  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3830507 0.5213957  
## sample estimates:  
## p   
## 0.4512821

# 1-sample proportions test without continuity correction  
#   
# data: table(pharynx$TIME1), null probability 0.5  
# X-squared = 1.8513, df = 1, p-value = 0.1736  
# alternative hypothesis: true p is not equal to 0.5  
# 95 percent confidence interval:  
# 0.3830507 0.5213957  
# sample estimates:  
# p   
# 0.4512821   
  
#(1a) A one sample z-test  
#(1b) The parameter of interest is seeing whether or not the proportion of those who survive past 500 days differs between this study and the European one (given the proportion is 0.5).  
#(1c) H0 : p= 0.5 vs. H A : p 0.5  
#(1d) p-value = 0.1736  
#(1e) Fail to reject null hypothesis.  
#(1f) The 95% confidence interval is (0.3830507, 0.5213957)  
#(1g) The conclusion is that we are 95% confident that the true proportion of patients that survive more than 500 days lies between 38.31% and 52.14%  
  
#2(a) The two-sample z-test and chi squared test.  
 # For two-sample z-test: H0 : p 1 = p 2 H A : p 1 ≠ p 2 (where p 1 is proportion of survival past 500 days in the standard treatment group and p 2 is the proportion of survival past 500 days in the test treatment group)  
 # For chi-square test: H0 : The proportion of survival past 500 days in the standard treatment group is independent of (or not associated with) the proportion of survival past 500 days in the test treatment group. H A : The proportion of survival past 500 days in the standard treatment group is dependent on (or associated with) the proportion of survival past 500 days in the test treatment group.  
#2(b)   
#chi-squared test   
table<-table(pharynx$TX,pharynx$TIME1)   
table

##   
## Greater than or equal to 500 less than 500 days  
## 1 50 50  
## 2 38 57

addmargins(table)

##   
## Greater than or equal to 500 less than 500 days Sum  
## 1 50 50 100  
## 2 38 57 95  
## Sum 88 107 195

prop.table(table, margin=1)

##   
## Greater than or equal to 500 less than 500 days  
## 1 0.5 0.5  
## 2 0.4 0.6

chisq.test(pharynx$TX,pharynx$TIME1,correct=F)

##   
## Pearson's Chi-squared test  
##   
## data: pharynx$TX and pharynx$TIME1  
## X-squared = 1.9674, df = 1, p-value = 0.1607

#The p-value is 0.1607, thus at the 0.05 level of significance the null hypothesis is failed to be rejected.  
  
#(3a) The expected cell count in each cell needs to be at least 5 or greater.  
#(3b)  
#checking cell count assumption   
treat.test<-chisq.test(pharynx$TX,pharynx$TIME1,correct=F)   
treat.test$expected

## pharynx$TIME1  
## pharynx$TX Greater than or equal to 500 less than 500 days  
## 1 45.12821 54.87179  
## 2 42.87179 52.12821

#The expected cell count assumption is met since each cell of the table has a value greater than 5.  
#(3c)  
#chi-squared test for stage of tumor   
chisq.test(pharynx$T\_STAGE,pharynx$TIME1, correct=F)

## Warning in chisq.test(pharynx$T\_STAGE, pharynx$TIME1, correct = F): Chi-  
## squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: pharynx$T\_STAGE and pharynx$TIME1  
## X-squared = 9.6507, df = 3, p-value = 0.02178

stage.test<-chisq.test(pharynx$T\_STAGE,pharynx$TIME1)

## Warning in chisq.test(pharynx$T\_STAGE, pharynx$TIME1): Chi-squared  
## approximation may be incorrect

stage.test$expected

## pharynx$TIME1  
## pharynx$T\_STAGE Greater than or equal to 500 less than 500 days  
## 1 4.061538 4.938462  
## 2 11.733333 14.266667  
## 3 41.969231 51.030769  
## 4 30.235897 36.764103

#The expected cell count assumption was not met since 2 cells of the table had value less than 5.  
  
#An appropriate alternative test is the Fisher’s exact test.   
#fisher's test for stage of tumor   
fisher.test(pharynx$T\_STAGE,pharynx$TIME1)

##   
## Fisher's Exact Test for Count Data  
##   
## data: pharynx$T\_STAGE and pharynx$TIME1  
## p-value = 0.02034  
## alternative hypothesis: two.sided

#4(a) 0.3173  
1-pchisq(1,df=1)

## [1] 0.3173105

#4(b) 0.0833  
1-pchisq(3,df=1)

## [1] 0.08326452

#4(b) 0.0253   
1-pchisq(5,df=1)

## [1] 0.02534732

#4(c) 0.6065  
1-pchisq(1,df=2)

## [1] 0.6065307

#4(d) 0.2231   
1-pchisq(3,df=2)

## [1] 0.2231302

#4(e) 0.0821  
1-pchisq(5,df=2)

## [1] 0.082085