Lab 05

Jyosna Philip

2023-12-19

## AIM

To test the equality of two population proportion

## DATASET DESCRIPTION

This dataset contains information on bone tumors. The data was collected from patients at the Memorial Sloan Kettering Cancer Center (MSKCC).

Bone\_Tumor <- read.csv("C:/Users/jyosn/Downloads/archive (1)/Bone Tumor Dataset.csv")

## ANALYSIS

head(Bone\_Tumor)

## Patient.ID Sex Age Grade Histological.type  
## 1 STS\_001 Female 63 High pleiomorphic leiomyosarcoma  
## 2 STS\_002 Female 54 Intermediate pleiomorphic leiomyosarcoma  
## 3 STS\_003 Male 22 Intermediate synovial sarcoma  
## 4 STS\_004 Female 54 Intermediate synovial sarcoma  
## 5 STS\_005 Male 63 High pleiomorphic spindle cell undifferentiated  
## 6 STS\_006 Female 58 High poorly differentiated synovial sarcoma  
## MSKCC.type Site.of.primary.STS Status..NED..AWD..D.  
## 1 Leiomyosarcoma parascapusular NED  
## 2 MFH left thigh D  
## 3 MFH right buttock D  
## 4 Leiomyosarcoma right thigh NED  
## 5 MFH right parascapusular AWD  
## 6 MFH right thigh D  
## Treatment  
## 1 Radiotherapy + Surgery  
## 2 Surgery + Chemotherapy  
## 3 Radiotherapy + Surgery  
## 4 Radiotherapy + Surgery  
## 5 Radiotherapy + Surgery + Chemotherapy  
## 6 Surgery + Chemotherapy

tail(Bone\_Tumor)

## Patient.ID Sex Age Grade Histological.type  
## 495 STS\_495 Male 54 High pleomorphic sarcoma  
## 496 STS\_496 Male 31 Intermediate epithelioid sarcoma  
## 497 STS\_497 Male 41 High poorly differentiated synovial sarcoma  
## 498 STS\_498 Female 60 Intermediate pleomorphic sarcoma  
## 499 STS\_499 Male 43 High synovial sarcoma  
## 500 STS\_500 Female 73 Intermediate pleiomorphic leiomyosarcoma  
## MSKCC.type Site.of.primary.STS Status..NED..AWD..D.  
## 495 MFH right thigh AWD  
## 496 MFH left thigh D  
## 497 MFH right parascapusular AWD  
## 498 Synovial sarcoma right thigh NED  
## 499 MFH right thigh D  
## 500 Leiomyosarcoma left biceps NED  
## Treatment  
## 495 Radiotherapy + Surgery + Chemotherapy  
## 496 Surgery + Chemotherapy  
## 497 Radiotherapy + Surgery + Chemotherapy  
## 498 Radiotherapy + Surgery  
## 499 Radiotherapy + Surgery + Chemotherapy  
## 500 Radiotherapy + Surgery

## CREATING 2 POPULATIONS

##Population 1  
male<- subset(Bone\_Tumor, Sex =="Male")

#Population 2  
female<- subset(Bone\_Tumor, Sex =="Female")

## TO CHECK HYPOTHESIS

Taking the Sample

sm=sample(male$Grade,150,replace = FALSE)  
S1=as.data.frame(sm) #to convert to dataframe  
  
  
sf=sample(female$Grade,150,replace = FALSE)  
S2=as.data.frame(sf)

## COUNT OF THE SAMPLE

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

count(S1,sm)

## sm n  
## 1 High 111  
## 2 Intermediate 39

count(S2,sf)

## sf n  
## 1 High 35  
## 2 Intermediate 115

## FIXING PROPORTION

P0: proportion of High Grade Tumors in Male

P1: proportion of High Grade Tumors in Female

## assume the population propotion

H0: P0 = P1

H1: P0 != P1

Success = 110 High Grade Tumors in Male

Success = 28 High Grade Tumors in Female

x1=c(110,28) #x1 has the success rates  
n1=c(150,150) #n1 has the sample sizes  
prop.test(x1, n1, alternative = "two.sided", conf.level = 0.99)

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: x1 out of n1  
## X-squared = 88.043, df = 1, p-value < 2.2e-16  
## alternative hypothesis: two.sided  
## 99 percent confidence interval:  
## 0.4160426 0.6772907  
## sample estimates:  
## prop 1 prop 2   
## 0.7333333 0.1866667

We get a very small p-value<0.01. Hence we have to reject H0. Therefore the true proportion of males and females with high grade bone tumors is not equal. We can now move forward to the one sided test.

## ONE SIDED TEST

## RIGHT-SIDED TEST

H0: P0 = P1 H1: P0 > P1 Alternative hypothesis:True Proportion of High grade tumors in males is greater than that of females.

prop.test(x1, n1, alternative = "greater", conf.level = 0.99)

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: x1 out of n1  
## X-squared = 88.043, df = 1, p-value < 2.2e-16  
## alternative hypothesis: greater  
## 99 percent confidence interval:  
## 0.4280485 1.0000000  
## sample estimates:  
## prop 1 prop 2   
## 0.7333333 0.1866667

We get a p-value which is very less than 0.01. Hence, we can reject H0 and conclude that P0>P1.

## Conclusion

Our rejection of null hypothesis in the two-sided test led us to carrying out the right-sided test. P-value obtained was less than 0.01, and so we rejected our null hypothesis. As a result our alternative hypothesis is accepted. Thus, the true proportion of males with high grade bone tumor is more than the true proportion of females with high grade bone tumor.