LAB-03 test for proportion

Moushreeta Debroy(22122132)

2022-11-08

<AIM:Our> objective is to test the percentage of smokers among the patients. ABOUT THE DATASET:Here we are working with blood pressure dataset.It consists of columns “patient number”,“blood pressure abnormality”,“level of hemoglobin”,“sex”,“smoking”,“age” etc.we found this dataset from kaggle. INTRODUCTION: TEST OF PROPORTION;A test of proportion will assess whether or not a sample from a population represents the true proportion from the entire population. Z TEST:A Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Z-tests test the mean of a distribution. Here we are using Z test as the sample size is greater than 30.

data<-read.csv('data.csv')

View(data)

head(data)

## Patient\_Number Blood\_Pressure\_Abnormality Level\_of\_Hemoglobin  
## 1 1 1 11.28  
## 2 2 0 9.75  
## 3 3 1 10.79  
## 4 4 0 11.00  
## 5 5 1 14.17  
## 6 6 0 11.64  
## Genetic\_Pedigree\_Coefficient Age BMI Sex Pregnancy Smoking Physical\_activity  
## 1 0.90 34 23 1 1 0 45961  
## 2 0.23 54 33 1 NA 0 26106  
## 3 0.91 70 49 0 NA 0 9995  
## 4 0.43 71 50 0 NA 0 10635  
## 5 0.83 52 19 0 NA 0 15619  
## 6 0.54 23 48 0 NA 1 27042  
## salt\_content\_in\_the\_diet alcohol\_consumption\_per\_day Level\_of\_Stress  
## 1 48071 NA 2  
## 2 25333 205 3  
## 3 29465 67 2  
## 4 7439 242 1  
## 5 49644 397 2  
## 6 7513 NA 3  
## Chronic\_kidney\_disease Adrenal\_and\_thyroid\_disorders  
## 1 1 1  
## 2 0 0  
## 3 1 0  
## 4 1 0  
## 5 0 0  
## 6 0 0

set.seed(200)

s<-sample(1:nrow(data),200,replace=F)  
s

## [1] 166 1394 1775 232 727 1756 1580 181 1562 1088 1106 1590 1883 1592 539  
## [16] 1708 280 445 1044 1398 1499 328 1252 1055 454 1005 1958 350 1404 902  
## [31] 643 939 867 1253 1921 1462 235 1640 1127 1597 120 274 1305 982 100  
## [46] 102 900 1854 1331 308 1631 1139 1749 984 597 1989 334 1019 1469 534  
## [61] 543 216 1686 1385 13 53 752 1859 1885 1647 1504 1556 1938 1263 421  
## [76] 286 1531 1917 212 317 342 1432 1018 1815 1904 370 91 710 1918 1770  
## [91] 1699 96 702 1757 1272 1266 1665 307 1135 747 184 129 703 915 1013  
## [106] 292 1910 645 1970 528 1673 506 517 52 1688 98 844 78 947 1690  
## [121] 1934 1841 1992 45 238 714 1834 1001 10 443 1222 659 360 998 1169  
## [136] 1800 1855 1551 2 551 1752 861 173 373 535 1544 441 1198 758 1743  
## [151] 1491 1620 1351 1535 595 1412 1265 1353 1516 632 314 940 282 1928 1137  
## [166] 883 1233 851 263 1837 1243 1442 1347 1907 524 1357 1249 886 1986 741  
## [181] 666 1948 1495 43 1666 473 420 1821 1477 55 1909 1890 1669 1964 843  
## [196] 1532 1027 811 766 478

new<-data[s,]  
head(new)

## Patient\_Number Blood\_Pressure\_Abnormality Level\_of\_Hemoglobin  
## 166 166 1 8.64  
## 1394 1394 1 13.63  
## 1775 1775 1 9.84  
## 232 232 0 14.17  
## 727 727 1 9.87  
## 1756 1756 1 8.88  
## Genetic\_Pedigree\_Coefficient Age BMI Sex Pregnancy Smoking  
## 166 0.72 32 10 1 1 0  
## 1394 0.40 58 23 1 NA 1  
## 1775 0.20 43 39 0 NA 1  
## 232 0.63 40 41 0 NA 1  
## 727 0.84 36 22 0 NA 0  
## 1756 0.68 74 15 1 NA 0  
## Physical\_activity salt\_content\_in\_the\_diet alcohol\_consumption\_per\_day  
## 166 4171 42413 464  
## 1394 37094 23324 20  
## 1775 22886 24279 237  
## 232 1858 13974 94  
## 727 43654 40983 180  
## 1756 43264 9096 281  
## Level\_of\_Stress Chronic\_kidney\_disease Adrenal\_and\_thyroid\_disorders  
## 166 3 1 0  
## 1394 1 0 0  
## 1775 1 0 1  
## 232 3 1 1  
## 727 1 0 1  
## 1756 2 1 0

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.2

##   
## 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(new,Smoking)

## Smoking n  
## 1 0 102  
## 2 1 98

P0:proportion of patient smoking

H0:P0=0.50 H1:P0!=0.50

x=98  
n=200  
prop.test(x,n,p=.50,alternative = "two.sided",conf.level = 0.95)

##   
## 1-sample proportions test with continuity correction  
##   
## data: x out of n, null probability 0.5  
## X-squared = 0.045, df = 1, p-value = 0.832  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4191175 0.5612729  
## sample estimates:  
## p   
## 0.49

Here we see that our p-value>alpha which is 0.05. Therefore, we accept our null hypothesis. Hence, amongst the patients we have 50% smokers.