LAB-04 TEST FOR SINGLE POPULATION PROPORTION

Nilanjana Dey

2023-12-15

#INTRODUCTION

library(readr)

## Warning: package 'readr' was built under R version 4.2.3

diabetes <- read\_csv("C:/Users/NILANJANA/Downloads/diabetes.csv")

## Rows: 768 Columns: 9  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## dbl (9): Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, D...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

View(diabetes)

# CHOOSING THE TARGET VARIABLE AS ‘OUTCOME’

target\_var=diabetes$Outcome

target\_var

## [1] 1 0 1 0 1 0 1 0 1 1 0 1 0 1 1 1 1 1 0 1 0 0 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0  
## [38] 1 1 1 0 0 0 1 0 1 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 1 0 1 0  
## [75] 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1  
## [112] 1 0 0 1 1 1 0 0 0 1 0 0 0 1 1 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## [149] 0 0 0 0 1 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 1 0 1 0 1 0 0 0 0 0  
## [186] 1 1 1 1 1 0 0 1 1 0 1 0 1 1 1 0 0 0 0 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 1 1  
## [223] 0 0 0 0 0 1 0 0 1 1 0 0 0 1 1 1 1 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0  
## [260] 1 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 0 1 0 0 0 1 0 0 1 1 0 0 1 0 0 0 1 1 1 0 0  
## [297] 1 0 1 0 1 1 0 1 0 0 1 0 1 1 0 0 1 0 1 0 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 0 1  
## [334] 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1  
## [371] 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 1 1 0 0 1 0 0 1 0 0 1 0 1 1 0 1 0 1 0 1  
## [408] 0 1 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1  
## [445] 1 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1  
## [482] 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 1 1 0  
## [519] 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 0  
## [556] 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 1 0 1 0 1 0  
## [593] 1 0 0 1 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0  
## [630] 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 1 0 1 1 1 1 0  
## [667] 1 1 0 0 0 0 0 0 0 1 1 0 1 0 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 1 1 0 0 0 0 1 1  
## [704] 0 0 0 1 0 1 1 0 0 1 0 0 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1  
## [741] 1 0 0 1 0 0 1 0 1 1 1 0 0 1 1 1 0 1 0 1 0 1 0 0 0 0 1 0

# A SAMPLE OF 300 IS DRAWN FROM THE NO. OF OUTCOMES WITH REPLACEMENT

sample1=sample(target\_var,300,replace=TRUE)  
sample1

## [1] 0 1 0 1 0 1 1 0 1 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0  
## [38] 0 1 0 0 0 0 1 1 0 0 1 0 1 0 0 0 1 1 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 1 1 0  
## [75] 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0  
## [112] 1 1 1 1 1 0 1 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1 1  
## [149] 1 0 1 0 0 0 0 0 1 1 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## [186] 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 1 1 0 0  
## [223] 0 1 0 1 0 1 1 1 0 0 0 0 1 1 0 0 0 0 0 1 1 0 1 0 1 0 0 0 1 0 1 1 1 0 0 0 1  
## [260] 0 1 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1  
## [297] 1 0 0 0

# THE sample drawn is converted into a dataframe for further analysis  
s=as.data.frame(sample1)  
s

## sample1  
## 1 0  
## 2 1  
## 3 0  
## 4 1  
## 5 0  
## 6 1  
## 7 1  
## 8 0  
## 9 1  
## 10 0  
## 11 1  
## 12 0  
## 13 0  
## 14 0  
## 15 0  
## 16 1  
## 17 0  
## 18 0  
## 19 0  
## 20 0  
## 21 1  
## 22 0  
## 23 0  
## 24 0  
## 25 0  
## 26 0  
## 27 0  
## 28 0  
## 29 0  
## 30 0  
## 31 1  
## 32 0  
## 33 1  
## 34 0  
## 35 0  
## 36 0  
## 37 0  
## 38 0  
## 39 1  
## 40 0  
## 41 0  
## 42 0  
## 43 0  
## 44 1  
## 45 1  
## 46 0  
## 47 0  
## 48 1  
## 49 0  
## 50 1  
## 51 0  
## 52 0  
## 53 0  
## 54 1  
## 55 1  
## 56 0  
## 57 0  
## 58 1  
## 59 0  
## 60 0  
## 61 1  
## 62 0  
## 63 0  
## 64 0  
## 65 0  
## 66 0  
## 67 0  
## 68 1  
## 69 0  
## 70 0  
## 71 0  
## 72 1  
## 73 1  
## 74 0  
## 75 1  
## 76 0  
## 77 0  
## 78 0  
## 79 0  
## 80 0  
## 81 0  
## 82 0  
## 83 0  
## 84 0  
## 85 0  
## 86 1  
## 87 0  
## 88 1  
## 89 0  
## 90 0  
## 91 0  
## 92 0  
## 93 0  
## 94 1  
## 95 0  
## 96 0  
## 97 0  
## 98 0  
## 99 0  
## 100 0  
## 101 0  
## 102 0  
## 103 0  
## 104 0  
## 105 1  
## 106 0  
## 107 0  
## 108 1  
## 109 0  
## 110 0  
## 111 0  
## 112 1  
## 113 1  
## 114 1  
## 115 1  
## 116 1  
## 117 0  
## 118 1  
## 119 1  
## 120 1  
## 121 0  
## 122 1  
## 123 0  
## 124 1  
## 125 0  
## 126 0  
## 127 0  
## 128 0  
## 129 0  
## 130 0  
## 131 0  
## 132 0  
## 133 0  
## 134 0  
## 135 0  
## 136 0  
## 137 0  
## 138 0  
## 139 0  
## 140 0  
## 141 0  
## 142 1  
## 143 1  
## 144 0  
## 145 0  
## 146 1  
## 147 1  
## 148 1  
## 149 1  
## 150 0  
## 151 1  
## 152 0  
## 153 0  
## 154 0  
## 155 0  
## 156 0  
## 157 1  
## 158 1  
## 159 0  
## 160 0  
## 161 0  
## 162 0  
## 163 1  
## 164 0  
## 165 1  
## 166 0  
## 167 1  
## 168 1  
## 169 0  
## 170 0  
## 171 0  
## 172 0  
## 173 0  
## 174 0  
## 175 0  
## 176 0  
## 177 0  
## 178 0  
## 179 1  
## 180 0  
## 181 0  
## 182 0  
## 183 0  
## 184 0  
## 185 0  
## 186 0  
## 187 0  
## 188 0  
## 189 0  
## 190 0  
## 191 0  
## 192 0  
## 193 1  
## 194 1  
## 195 0  
## 196 0  
## 197 0  
## 198 0  
## 199 0  
## 200 1  
## 201 0  
## 202 0  
## 203 0  
## 204 0  
## 205 0  
## 206 1  
## 207 0  
## 208 0  
## 209 1  
## 210 0  
## 211 0  
## 212 0  
## 213 0  
## 214 0  
## 215 0  
## 216 0  
## 217 1  
## 218 0  
## 219 1  
## 220 1  
## 221 0  
## 222 0  
## 223 0  
## 224 1  
## 225 0  
## 226 1  
## 227 0  
## 228 1  
## 229 1  
## 230 1  
## 231 0  
## 232 0  
## 233 0  
## 234 0  
## 235 1  
## 236 1  
## 237 0  
## 238 0  
## 239 0  
## 240 0  
## 241 0  
## 242 1  
## 243 1  
## 244 0  
## 245 1  
## 246 0  
## 247 1  
## 248 0  
## 249 0  
## 250 0  
## 251 1  
## 252 0  
## 253 1  
## 254 1  
## 255 1  
## 256 0  
## 257 0  
## 258 0  
## 259 1  
## 260 0  
## 261 1  
## 262 0  
## 263 0  
## 264 0  
## 265 0  
## 266 1  
## 267 0  
## 268 1  
## 269 0  
## 270 0  
## 271 0  
## 272 1  
## 273 0  
## 274 0  
## 275 0  
## 276 0  
## 277 0  
## 278 0  
## 279 0  
## 280 0  
## 281 1  
## 282 1  
## 283 0  
## 284 0  
## 285 0  
## 286 0  
## 287 0  
## 288 1  
## 289 0  
## 290 0  
## 291 0  
## 292 0  
## 293 1  
## 294 0  
## 295 0  
## 296 1  
## 297 1  
## 298 0  
## 299 0  
## 300 0

# We create a contingency table to count the no. of females having diabetes   
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.2.3

##   
## 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(s,sample1)

## sample1 n  
## 1 0 214  
## 2 1 86

##HYPOTHESIS TESTING

## BOTH-TAILED TEST

H0: P0=O. VS H1: P1!=0.56

n=300  
x=89  
prop.test(x, n, p = 0.56, alternative = "two.sided",  
 correct = TRUE)

##   
## 1-sample proportions test with continuity correction  
##   
## data: x out of n, null probability 0.56  
## X-squared = 83.364, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.56  
## 95 percent confidence interval:  
## 0.2462485 0.3523909  
## sample estimates:  
## p   
## 0.2966667

## RIGHT TAILED

H0: P0=O. VS H1: P1>0.56

n=300  
x=89  
prop.test(x, n, p = 0.56, alternative = "greater",  
 correct = TRUE)

##   
## 1-sample proportions test with continuity correction  
##   
## data: x out of n, null probability 0.56  
## X-squared = 83.364, df = 1, p-value = 1  
## alternative hypothesis: true p is greater than 0.56  
## 95 percent confidence interval:  
## 0.2536789 1.0000000  
## sample estimates:  
## p   
## 0.2966667

## LEFT TAILED

H0: P0=O. VS H1: P1<0.56

n=300  
x=89  
prop.test(x, n, p = 0.34, alternative = "less",  
 correct = TRUE)

##   
## 1-sample proportions test with continuity correction  
##   
## data: x out of n, null probability 0.34  
## X-squared = 2.321, df = 1, p-value = 0.06382  
## alternative hypothesis: true p is less than 0.34  
## 95 percent confidence interval:  
## 0.0000000 0.3434281  
## sample estimates:  
## p   
## 0.2966667