snəẃeyəɬ leləm̓. Langara.
THE COLLEGE OF HIGHER LEARNING. THE COLLEGE OF HIGHER LEARNING.

Langara College
Department of Mathematics and Statistics
DANA 4820
Activity 1
Fall 2023

Last Name: ____Gupta_____

First Name: _____Rahul_____

Student #: _____100405295_____

Question 1 – The dataset "Analgesic" is based on a study of 100 women suffering from excessive menstrual bleeding, which considers whether a new analgesic provides greater relief than the standard analgesic. Use all the methods explained in chapter 1 and explain whether you meet the condition(s).

a. [5] Test the hypothesis that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic. Report and interpret the P-value.

Wald Test

H0 : π = 0.5
Ha : π != 0.5

```
> y <- sum(first_column == "Yes")
> y
[1] 60
> phat=y/n
> phat
[1] 0.6
> SE=sqrt((phat*(1-phat))/n)
> SE
[1] 0.04898979
> z=(phat-0.5)/SE
> z
[1] 2.041241
> pvalue1=2*pnorm(z,lower.tail = FALSE)
> #or chi-square test statistic
> z^2
[1] 4.166667
> pvalue2=pchisq(z^2,1,lower.tail = FALSE)
> pvalue2
[1] 0.04122683
>
```

As P-value is much less than the significant value of 0.05 (0.04122683), it's not statistically significant and hence we can reject the null hypothesis. The evidence is statistically sufficient to say that the value for pi is either more or less than 0.5.

Score Test

```
> ztest=prop.test(y,n, p = 0.5, alternative = "two.sided",
+                 correct = FALSE,conf.level=0.95)
> ztest

        1-sample proportions test without continuity correction

data:  y out of n, null probability 0.5
X-squared = 4, df = 1, p-value = 0.0455
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5020026 0.6905987
sample estimates:
  p
0.6
```

As P-value is much less than the significant value of 0.05 (0.0.0455), it's not statistically significant and hence we can reject the null hypothesis. The evidence is statistically sufficient to say that the value for pi is either more or less than 0.5.

Likelihood ratio Test

```
> #likelihood-ratio test
> likelihood_ratio_test <- function(phat, pnull, y,n)
+ {
+   LL <- c("LRTstatistic"=2*log(dbinom(x = y, prob = phat, size = n)/dbinom(x = y, prob = pnull, size = n)),
+           "p-value"=pchisq(2*log(dbinom(x = y, prob = phat, size = n)/dbinom(x = y, prob = pnull, size = n)),
1,lower.tail = FALSE))
+   return(LL)
+ }
> likelihood_ratio_test(phat=0.6,pnull=0.5,y,n)
LRTstatistic      p-value
  4.02710271   0.04477477
```

As P-value is much less than the significant value of 0.05 (0.0.04477477), it's not statistically significant and hence we can reject the null hypothesis. The evidence is statistically sufficient to say that the value for pi is either more or less than 0.5.

b. [2] Construct and interpret a 95% confidence interval for the probability of greater relief with the new analgesic.

```
> ztest=prop.test(y,n, p = 0.5, alternative = "two.sided",
+                 correct = FALSE,conf.level=0.95)
> ztest

        1-sample proportions test without continuity correction

data:  y out of n, null probability 0.5
X-squared = 4, df = 1, p-value = 0.0455
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5020026 0.6905987
sample estimates:
  p
0.6
```

As per the confidence of 95%, the probability that the women who experiences of greater relief with new analgesic is between the confidence interval of 0.5020026 & 0.6905987 with a p-value of 0.0455.

c. [2] Use the binomial distribution directly for part (a) and compare your conclusion with the results concluded in part (a).

```
> #Binomial Exact test
> binom.test(y,n,p=0.5, conf.level = 0.95, alternative = "two.sided")

        Exact binomial test

data:  y and n
number of successes = 60, number of trials = 100, p-value = 0.05689
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.4972092 0.6967052
sample estimates:
probability of success
                   0.6

> library(exactci)
Loading required package: ssanv
Loading required package: testthat
> binom.exact(60, 100, 0.50, alternative="two.sided", midp=TRUE) # mid P-value

        Exact two-sided binomial test (central method), mid-p version

data:  60 and 100
number of successes = 60, number of trials = 100, p-value = 0.04604
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.5017318 0.6925190
sample estimates:
probability of success
                   0.6

>
```

As the above p-value is closer to our assumed value of 0.05 we have failed to reject the null hypothesis.

Question 2 – The University of Michigan Health and Retirement (HRS) surveys more than 22,000 American over the age of 50 every two years. A subsample of the HRS participated in 2009 Internet-based survey that collected information on a number of topical areas, including health (physical and mental health behaviors), psychosocial items, economics (income, assets, expectations, and consumption), and retirement. Two of the questions asked were, "Would you say your health is excellent, very good, fair, or poor?" and "Do you smoke cigarettes now?" The two-way table summarizes the answers on these two questions.

		Current Smoker	
		Yes	No
Health	Excellent	25	484
	Very good	115	1557
	Good	145	1309
	Fair	90	545
	Poor	29	11

a) [2] Create a two-way table using RStudio.

```
> Health <- c(25,115,145,90,29,484,1557,1309,545,11)
> table <- as.table(matrix(Health, nrow = 5, byrow = FALSE,
+                                dimnames = list(Health= c('Excellent', 'Very good', 'Good', 'Fair', 'Poor'),
+                                                Current.smoker = c('Yes', 'No'))))
> table
           Current.smoker
Health       Yes   No
  Excellent   25  484
  Very good  115 1557
  Good       145 1309
  Fair        90  545
  Poor        29   11
> addmargins(table)
           Current.smoker
Health       Yes   No  Sum
  Excellent   25  484  509
  Very good  115 1557 1672
  Good       145 1309 1454
  Fair        90  545  635
  Poor        29   11   40
  Sum        404 3906 4310
>
```

b) [4] Does the data give you sufficient evidence that self-evaluation of health is associated with smoking status? ($\alpha = 0.05$)

```
        Pearson's Chi-squared test

data:  table
X-squared = 229.66, df = 4, p-value < 2.2e-16
```

As the p-value is significantly lower than our alpha value of 0.05, we have now sufficient evidence to reject the null hypothesis and it's statistically significant too to reject the null hypothesis. Hence, it shows sufficient evidence that self-evaluation of health does associate with smoking status.

c) [7] Can a linear trend be detected in this two-way table? If yes, detect a linear and a positive linear (separately) trend between the variables?

```
> library(vcdExtra)
> A=CMHtest(table, rscores = c(1,2,3,4,5), cscores=0:1, types="cor")
> A
Cochran-Mantel-Haenszel Statistics for Health by Current.smoker

          AltHypothesis  Chisq Df      Prob
cor Nonzero correlation 92.461  1 6.866e-22

> #Sample correlation R
> n=sum(table)
> n
[1] 4310
> M=sqrt(A$table[1])
> M
[1] 9.615671
> R=M/sqrt(n-1)
> R
[1] 0.1464844
> P_value=pnorm(M, lower.tail = FALSE)
> P_value
[1] 3.433014e-22
>
```

As the sample, correlation value is close to 0 (0.1464844), it seems much weaker. The trend is positive with the value more than 0 and has p-value much lesser than 0.05, our alpha value, it can be said that the evidence is too strong that we can reject null hypothesis with a stronger non zero correlation.

d) [3] Compare the odds of being a current smoker within the two groups: those who say their health is excellent and those who say their health is poor, using odds ratio and interpret the value.

```
> oddsratio(c(29,11,25,484), method="wald", conf=0.95, correct=FALSE)
$data
          Outcome
Predictor  Disease1 Disease2 Total
  Exposed1       29       11    40
  Exposed2       25      484   509
  Total          54      495   549

$measure
          odds ratio with 95% C.I.
Predictor  estimate   lower    upper
  Exposed1     1.00      NA       NA
  Exposed2    51.04 22.8868 113.8246

$p.value
          two-sided
Predictor  midp.exact fisher.exact  chi.square
  Exposed1         NA           NA          NA
  Exposed2          0 1.536149e-24 1.89706e-43

$correction
[1] FALSE
```

The result from above illustrates that the group with health excellent is 51.04 times higher than the group who has health poor in terms of current smokers.