STA160-Indep_odd

Cindy Chen

4/13/2022

Hypotheses:
H_0: There is no association between the two variables.
H_a: There is an association between the two variables.

High blood pressure

Pearson χ^2 test can be used if there are more observations (e.g. lager than 10) for individual cells. Also since it is 2X2 table, we can use Fisher's exact test to compare both p-value

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have high blood pressure are more likely (4.592099 times) to have heart disease than people who do not have high blood pressure. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##        HeartDis
## HighBP      0      1
##      0 138886   5965
##      1  90901  17928
```

High cholesterol

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have high cholesterol are more likely (3.589073 times) to have heart disease than people who do not have high cholesterol. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##          HeartDis
## Highchol      0      1
##        0 138949   7140
##        1  90838  16753
```

Diabetes

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

Odd ratio for no diabetes:
$Odd_{Heart/0} = \frac{15351/23893}{1-15351/23893} = 0.2849789$ $Odd_{NoHeart/0} = \frac{198352/229787}{1-198352/229787} = 6.309909$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.2849789}{6.309909} = 0.04516371$
From the odd ratio for diabetes 0 (no diabetes), it shows that in the sample, people are less likely(0.04516371 times) to have heart disease when they do not have diabetes.

Odd ratio for pre-diabetes:
$Odd_{Heart/1} = \frac{664/23893}{1-664/23893} = 0.02858496$
$Odd_{NoHeart/1} = \frac{3967/229787}{1-3967/229787} = 0.01756708$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.02858496}{0.01756708} = 1.627189$
From the odd ratio for diabetes 1 (Pre-diabetes), it shows that in the sample, people are more likely(1.627189 times) to have heart disease when they have pre-diabetes.

Odd ratio for diabetes:
$Odd_{Heart/1} = \frac{7878/23893}{1-7878/23893} = 0.4919138$
$Odd_{NoHeart/1} = \frac{27468/229787}{1-27468/229787} = 0.1357658$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.4919138}{0.1357658} = 3.623253$
From the odd ratio for diabetes 2 (diabetes), it shows that in the sample, people are more likely(3.623253 times) to have heart disease when they have diabetes.

```
##          HeartDis
## Diabetes       0      1
##        0 198352  15351
##        1   3967    664
##        2  27468   7878
```

Difficulty walking

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have serious difficulty walking are more likely (4.266085 times) to have heart disease than people who do not have serious difficulty walking. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##          HeartDis
## DiffWalk       0      1
##        0 197027  13978
##        1  32760   9915
```

Phys activity

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who doing physical activity or exercise during the past 30 days other than their regular job are less likely (0.5359804 times) to have heart disease than people who do not do the physical activity. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##              HeartDis
## PhysActivity      0      1
##            0  53167   8593
##            1 176620  15300
```

Smoking

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have smoke are more likely (2.203943 times) to have heart disease than people who do not smoke. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##       HeartDis
## Smoker      0      1
##      0 132165   9092
##      1  97622  14801
```

Stroke

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have stroke are more likely (6.936202 times) to have heart disease than people who do not stroke. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

```
##       HeartDis
## Stroke      0      1
##      0 223432  19956
##      1   6355   3937
```

Age

The resulting p-value obtained from the Pearson χ^2 and Fisher's exact test are very small, which means that there is stronger evidence in favor of the alternative hypothesis.

From the odd ratio obtained from the sample, it shows that people who have stroke are more likely (6.936202 times) to have heart disease than people who do not stroke. In addition, since the odd ratio is not equal to 1, it means that two variables are not independent, which is same as the results obtained by Pearson Test and Fisher's Exact Test.

Odd ratio for age group 1 (18-24):
$Odd_{Heart/1} = \frac{29/23893}{1-29/23893} = 0.00121522$
$Odd_{NoHeart/1} = \frac{5671/229787}{1-5671/229787} = 0.02530386$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.00121522}{0.02530386} = 0.04802508$
From the odd ratio for age group 1, it shows that in the sample, people are less likely(0.04802508 times) to have heart disease for people age are between 18 and 24.

Odd ratio for age group 2 (25-30):
$Odd_{Heart/2} = \frac{54/23893}{1-54/23893} = 0.002265196$

$Odd_{NoHeart/2} = \frac{7544/229787}{1-7544/229787} = 0.03394483$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.002265196}{0.03394483} = 0.06673169$
From the odd ratio for age group 2, it shows that in the sample, people are less likely(0.06673169 times) to have heart disease for people age are between 25 and 30.

Odd ratio for age group 3 (31-35):
$Odd_{Heart/2} = \frac{126/23893}{1-126/23893} = 0.005301468$
$Odd_{NoHeart/2} = \frac{10997/229787}{1-10997/229787} = 0.05026281$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.005301468}{0.05026281} = 0.105475$
From the odd ratio for age group 3, it shows that in the sample, people are less likely(0.105475 times) to have heart disease for people age are between 31 and 35.

Odd ratio for age group 4 (36-40):
$Odd_{Heart/2} = \frac{193/23893}{1-193/23893} = 0.00814346$
$Odd_{NoHeart/2} = \frac{13630/229787}{1-13630/229787} = 0.06305602$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.00814346}{0.06305602} = 0.1291464$
From the odd ratio for age group 4, it shows that in the sample, people are less likely(0.1291464 times) to have heart disease for people age are between 36 and 40.

Odd ratio for age group 5 (41-45):
$Odd_{Heart/2} = \frac{351/23893}{1-351/23893} = 0.014909520$
$Odd_{NoHeart/2} = \frac{15806/229787}{1-15806/229787} = 0.07386637$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.014909520}{0.07386637} = 0.2018445$
From the odd ratio for age group 5, it shows that in the sample, people are less likely(0.2018445 times) to have heart disease for people age are between 41 and 45.

Odd ratio for age group 6 (46-50):
$Odd_{Heart/2} = \frac{712/23893}{1-712/23893} = 0.03071481$
$Odd_{NoHeart/2} = \frac{19107/229787}{1-19107/229787} = 0.09069204$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.03071481}{0.09069204} = 0.3386715$
From the odd ratio for age group 6, it shows that in the sample, people are less likely(0.3386715 times) to have heart disease for people age are between 46 and 50.

Odd ratio for age group 7 (51-55):
$Odd_{Heart/2} = \frac{1425/23893}{1-1425/23893} = 0.06342354$
$Odd_{NoHeart/2} = \frac{24889/229787}{1-24889/229787} = 0.1214702$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.06342354}{0.1214702} = 0.5221325$
From the odd ratio for age group 7, it shows that in the sample, people are less likely(0.5221325 times) to have heart disease for people age are between 51 and 55.

Odd ratio for age group 8 (56-60):
$Odd_{Heart/2} = \frac{2253/23893}{1-2253/23893} = 0.1041128$
$Odd_{NoHeart/2} = \frac{28579/229787}{1-28579/229787} = 0.1420371$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.1041128}{0.1420371} = 0.7329972$
From the odd ratio for age group 8, it shows that in the sample, people are less likely(0.7329972 times) to have heart disease for people age are between 56 and 60.

Odd ratio for age group 9 (61-65):
$Odd_{Heart/2} = \frac{3358/23893}{1-3358/23893} = 0.1635257$
$Odd_{NoHeart/2} = \frac{29886/229787}{1-29886/229787} = 0.149504$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.1635257}{0.149504} = 1.093788$
From the odd ratio for age group 9, it shows that in the sample, people are more likely(1.093788 times) to have heart disease for people age are between 61 and 65.

Odd ratio for age group 10 (66-70):
$Odd_{Heart/2} = \frac{4193}{1-4193/23893} = 0.2128426$
$Odd_{NoHeart/2} = \frac{28001/229787}{1-28001/229787} = 0.1387658$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.2128426}{0.1387658} = 1.533826$
From the odd ratio for age group 10, it shows that in the sample, people are more likely(1.533826 times) to have heart disease for people age are between 66 and 70.

Odd ratio for age group 11 (71-75):
$Odd_{Heart/2} = \frac{3947}{1-3947/23893} = 0.1978843$
$Odd_{NoHeart/2} = \frac{19586/229787}{1-19586/229787} = 0.09317748$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.1978843}{0.09317748} = 2.123735$
From the odd ratio for age group 11, it shows that in the sample, people are more likely(2.123735 times) to have heart disease for people age are between 71 and 75.

Odd ratio for age group 12 (76-80):
$Odd_{Heart/2} = \frac{3093}{1-3093/23893} = 0.1487019$
$Odd_{NoHeart/2} = \frac{12887/229787}{1-12887/229787} = 0.05941448$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.1487019}{0.05941448} = 2.502789$
From the odd ratio for age group 12, it shows that in the sample, people are more likely(2.502789 times) to have heart disease for people age are between 76 and 80.

Odd ratio for age group 13 (over 80):
$Odd_{Heart/2} = \frac{4159}{1-4159/23893} = 0.210753$
$Odd_{NoHeart/2} = \frac{13204/229787}{1-13204/229787} = 0.06096508$
Therefore, $Odd_{Heart/NoHeart} = \frac{0.210753}{0.06096508} = 3.456946$
From the odd ratio for age group 13, it shows that in the sample, people are more likely(3.456946 times) to have heart disease for people age are over 80.

From above different age groups odd ratio, it can be seen that the odd ratio are increase as age increase. Interestingly, it can be see that the odd ratio becomes large than 1 after the age group 8. Therefore, the age over 60 will have the high chance to get the heart disease especially those people who are over 80. On the other hands, the odd ratio then to close to zero as age decrease. Therfore, the young people have less chance to get the heart disease.

```
##     HeartDis
## Age      0     1
##   1   5671    29
##   2   7544    54
##   3  10997   126
##   4  13630   193
##   5  15806   351
##   6  19107   712
##   7  24889  1425
##   8  28579  2253
##   9  29886  3358
##   10 28001  4193
##   11 19586  3947
##   12 12887  3093
##   13 13204  4159
```

Conclusion:

	High BP	High Chol	Diabetes	Diff walk	Phys act	Smoking	Stroke	Age
Test Statistics (Person Test)	11119.3	8289.27	8244.889	11477.75	1933.324 (smallest)	3322.39	10454.1	13731.04 (largest)
Odd ratio	4.5921	3.58907	0: 0.045164 1: 1.627189 2: 3.623253	4.266085	0.53598	2.20394	6.936	1: 0.04802508 13:3.456946
Disease chance	High BP more	High Chol more	Diabetes more No diabetes less	Diff walk more	Physical activity less	Smoking more	Stroke more	Age group 1 less Age group 13 more

From the table above, it shows that Stroke has the highest odd ratio, which is 6.936. This means that people who have stroke have the highest chance to get heart disease than other factors since the highest odd ratio is obtained in this group. In addition, the second highest odd ratio occurs on the HighBP variable, which is 4.5921. Therefore, the people who have high blood pressure are more likely to get heart disease than others.

Since p-values are too small to show in the R, test statistics will be used to compare the test result. The test result shows that Age and Diff walk variables have the largest test statistics. Therefore, there is a strong evidence to conclude that these two variables are not independent from heart disease. They are strong associated with heart disease.

Appendix Code

```
#read data
heart <- read.csv("~/Downloads/STA160/heart_disease_health_indicators_BRFSS2015.csv")
#contingency table (HighBP vs Heart Disease)
counthp=table(HighBP=heart$HighBP,HeartDis=heart$HeartDiseaseorAttack)
counthp

#pearson test
pearsonStatistic_hp <- chisq.test(counthp, correct=FALSE)$stat
pearsonpVal_hp <- chisq.test(counthp, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(counthp)$p.val

#odd ratio
odd_hp=(138886*17928)/(5965*90901)
#contingency table (Highchol vs Heart Disease)
countchol=table(Highchol=heart$HighChol,HeartDis=heart$HeartDiseaseorAttack)
countchol
```

```r
#pearson test
pearsonStatistic_chol <- chisq.test(countchol, correct=FALSE)$stat
pearsonpVal_chol <- chisq.test(countchol, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(countchol)$p.val

#odd ratio
odd_chol=(138949*16753)/(7140*90838)
#contingency table (Diabetes vs Heart Disease)
countdiab=table(Diabetes=heart$Diabetes,HeartDis=heart$HeartDiseaseorAttack)
countdiab

#pearson test
pearsonStatistic_diab <- chisq.test(countdiab, correct=FALSE)$stat
pearsonpVal_diab <- chisq.test(countdiab, correct=FALSE)$p.val

#contingency table (DiffWalk vs Heart Disease)
countdiff=table(DiffWalk=heart$DiffWalk,HeartDis=heart$HeartDiseaseorAttack)
countdiff

#pearson test
pearsonStatistic_diff <- chisq.test(countdiff, correct=FALSE)$stat
pearsonpVal_diff <- chisq.test(countdiff, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(countdiff)$p.val

#odd ratio
odd_diff=(197027*9915)/(13978*32760)
#contingency table (PhysActivity vs Heart Disease)
countphysA=table(PhysActivity=heart$PhysActivity,HeartDis=heart$HeartDiseaseorAttack)
countphysA

#pearson test
pearsonStatistic_physA <- chisq.test(countphysA, correct=FALSE)$stat
pearsonpVal_physA <- chisq.test(countphysA, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(countphysA)$p.val

#odd ratio
odd_physA=(53167*15300)/(8593*176620)
#contingency table (smoker vs Heart Disease)
countsmok=table(Smoker=heart$Smoker,HeartDis=heart$HeartDiseaseorAttack)
countsmok

#pearson test
pearsonStatistic_smok <- chisq.test(countsmok, correct=FALSE)$stat
pearsonpVal_smok <- chisq.test(countsmok, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(countsmok)$p.val
```

```r
#odd ratio
odd_smok=(132165*14801)/(9092*97622)
#contingency table (Stroke vs Heart Disease)
countstrok=table(Stroke=heart$Stroke,HeartDis=heart$HeartDiseaseorAttack)
countstrok

#pearson test
pearsonStatistic_strok <- chisq.test(countstrok, correct=FALSE)$stat
pearsonpVal_strok <- chisq.test(countstrok, correct=FALSE)$p.val

#fisher exact test
fisherPval <- fisher.test(countstrok)$p.val

#odd ratio
odd_strok=(223432*3937)/(19956*6355)
#contingency table (age vs Heart Disease)
countage=table(Age=heart$Age,HeartDis=heart$HeartDiseaseorAttack)
countage

#pearson test
pearsonStatistic_age <- chisq.test(countage, correct=FALSE)$stat
pearsonpVal_age <- chisq.test(countage, correct=FALSE)$p.val

#odd ratio
age1_no=(5671/sum(countage[,1]))/(1-(5671/sum(countage[,1])))
age1_yes=(29/sum(countage[,2]))/(1-(29/sum(countage[,2])))
age2_no=(7544/sum(countage[,1]))/(1-(7544/sum(countage[,1])))
age2_yes=(54/sum(countage[,2]))/(1-(54/sum(countage[,2])))
age3_no=(10997/sum(countage[,1]))/(1-(10997/sum(countage[,1])))
age3_yes=(126/sum(countage[,2]))/(1-(126/sum(countage[,2])))
age4_no=(13630/sum(countage[,1]))/(1-(13630/sum(countage[,1])))
age4_yes=(193/sum(countage[,2]))/(1-(193/sum(countage[,2])))
age5_no=(15806/sum(countage[,1]))/(1-(15806/sum(countage[,1])))
age5_yes=(351/sum(countage[,2]))/(1-(351/sum(countage[,2])))
age6_no=(19107/sum(countage[,1]))/(1-(19107/sum(countage[,1])))
age6_yes=(712/sum(countage[,2]))/(1-(712/sum(countage[,2])))
age7_no=(24889/sum(countage[,1]))/(1-(24889/sum(countage[,1])))
age7_yes=(1425/sum(countage[,2]))/(1-(1425/sum(countage[,2])))
age8_no=(28579/sum(countage[,1]))/(1-(28579/sum(countage[,1])))
age8_yes=(2253/sum(countage[,2]))/(1-(2253/sum(countage[,2])))
age9_no=(29886/sum(countage[,1]))/(1-(29886/sum(countage[,1])))
age9_yes=(3358/sum(countage[,2]))/(1-(3358/sum(countage[,2])))
age10_no=(28001/sum(countage[,1]))/(1-(28001/sum(countage[,1])))
age10_yes=(4193/sum(countage[,2]))/(1-(4193/sum(countage[,2])))
age11_no=(19586/sum(countage[,1]))/(1-(19586/sum(countage[,1])))
age11_yes=(3947/sum(countage[,2]))/(1-(3947/sum(countage[,2])))
age12_no=(12887/sum(countage[,1]))/(1-(12887/sum(countage[,1])))
age12_yes=(3093/sum(countage[,2]))/(1-(3093/sum(countage[,2])))
age13_no=(13204/sum(countage[,1]))/(1-(13204/sum(countage[,1])))
age13_yes=(4159/sum(countage[,2]))/(1-(4159/sum(countage[,2])))
```