BDM600 - Lab 6 - Group 8

Ran Arino; Zubeka Dane Dang; Solmaz Heidar Nassab

2024-03-24

Credentials

All members participated in this lab assignment. The file is the merged version.

Start

```r
# install libraries
library(stats)
library(vcd)
```

```
## Loading required package: grid
```

```r
library(vcdExtra)
```

```
## Loading required package: gnm
```

4.1

```r
fat <- matrix(c(6, 4, 2, 11), 2, 2)
dimnames(fat) <- list(diet = c("LoChol", "HiChol"), disease = c("No", "Yes"))

# test
result <- chisq.test(fat)
```

(a)

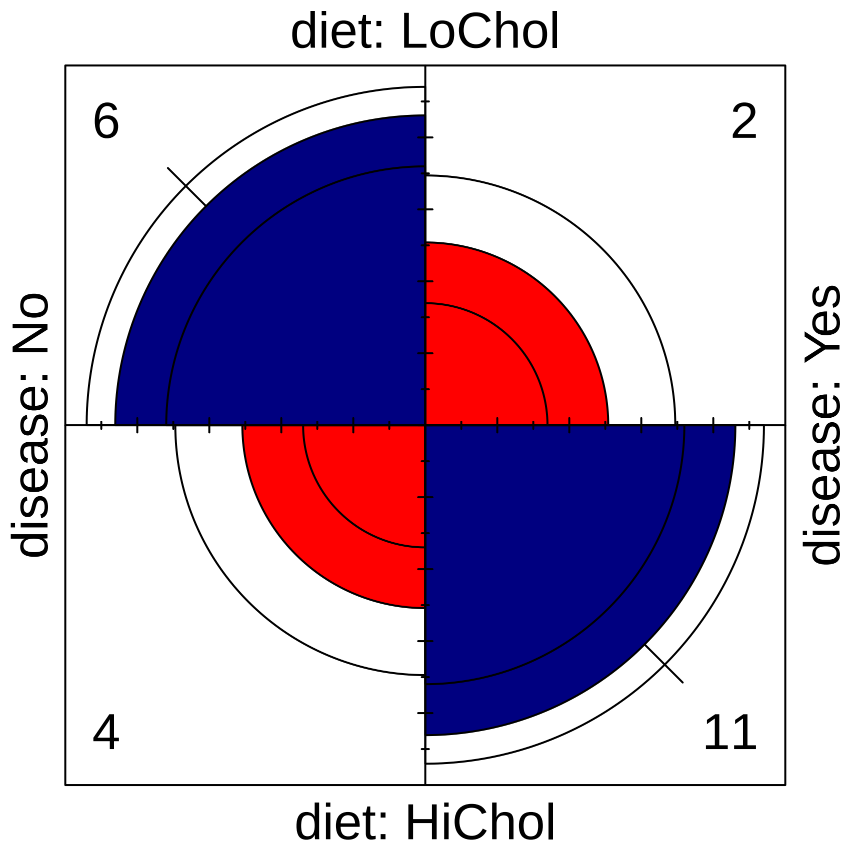
```
## Warning in chisq.test(fat): Chi-squared approximation may be incorrect
```

```r
result
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  fat
## X-squared = 3.1879, df = 1, p-value = 0.07418
```

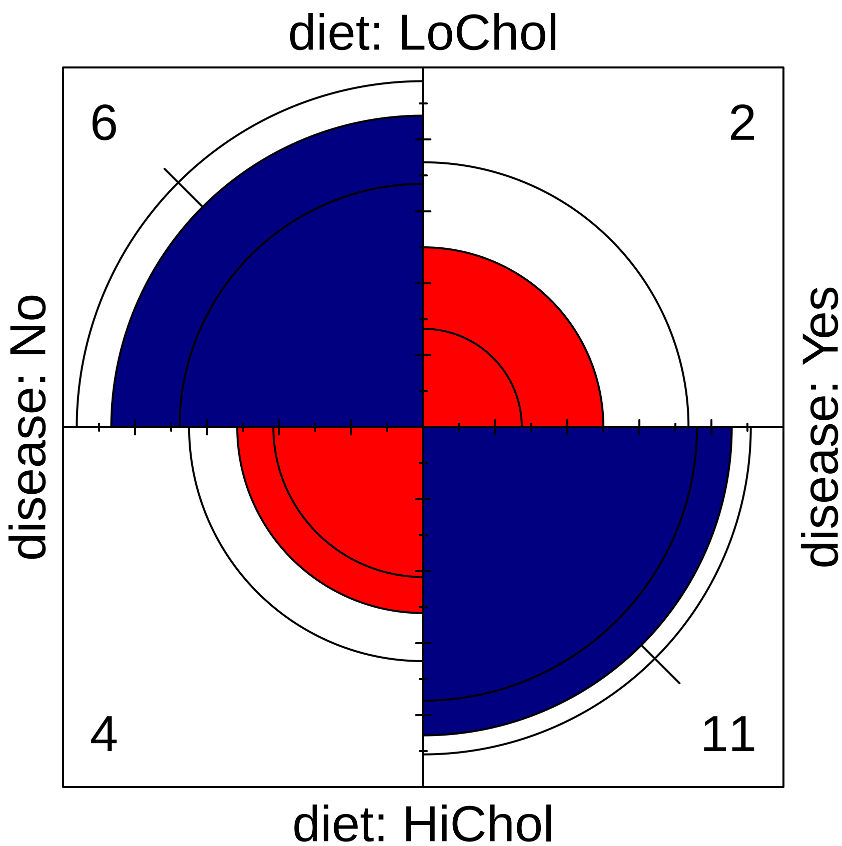
- According to the 95% confidence internal and greater than 0.05 p-value, we cannot prove the significance between diet and disease.
- Since one of the frequencies in the table is less than 5, we may have to use the alternative test, like Fisher's exact test.

```r
# default margins
fourfold(fat)
```

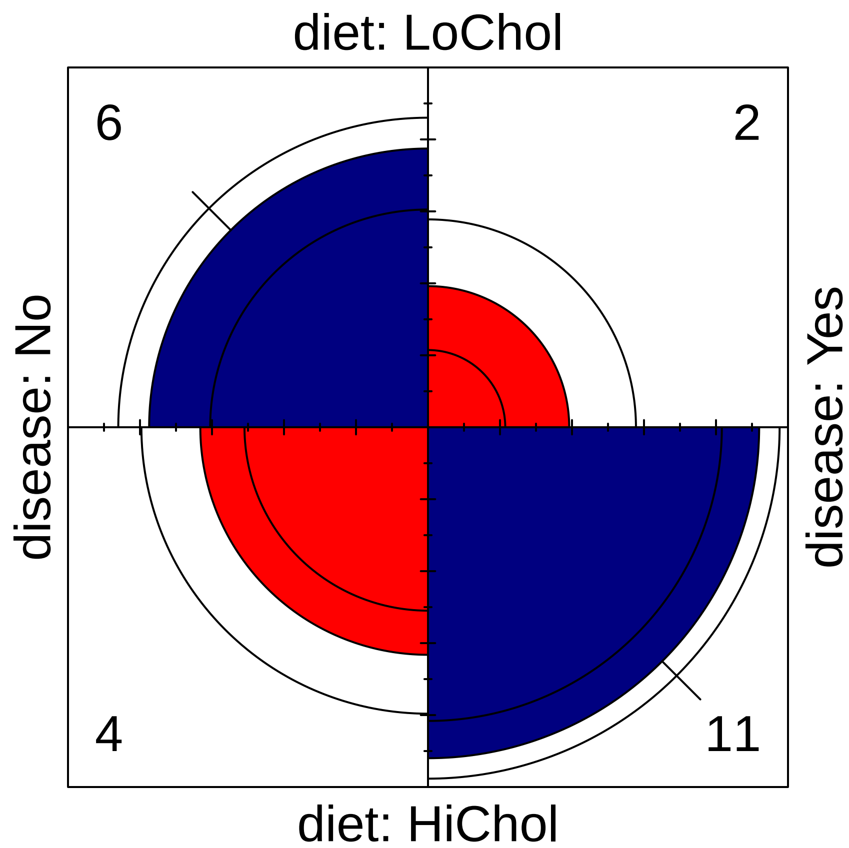


(b)

```r
# standardize row margins
fourfold(fat, margin = 1)
```




```r
# standardize column margins
fourfold(fat, margin = 2)
```



- For default margin, all bands for different cells do not overlap with the diagonal line, which indicates the odds ratio is significantly different from 1. - For row margin, only the pair between "LoChol & No disease" and "LoCho & Yes Disease" overlaps with the diagnal line, which shows no association within a particular combination. - For column margin, only the pair between "LoChol & No disease" and "HiChol & No Disease" overlaps with the diagnal line, which shows no association within a particular combination.

```r
oddsratio(fat, log = FALSE)
```

(c)

```
##  odds ratios for diet and disease
##
## [1] 8.25
```

- Odds ratio shows the 8.25, which significantly different from 1.
- It suggests a strong association between diet and disease, and indicates an increased likelihood of disease occurrence associated with the high cholesterol diet.
- Hence, this number also supports the non-overlapped circle lines when it's default margin.

```r
fisher.test(fat)
```

(d)

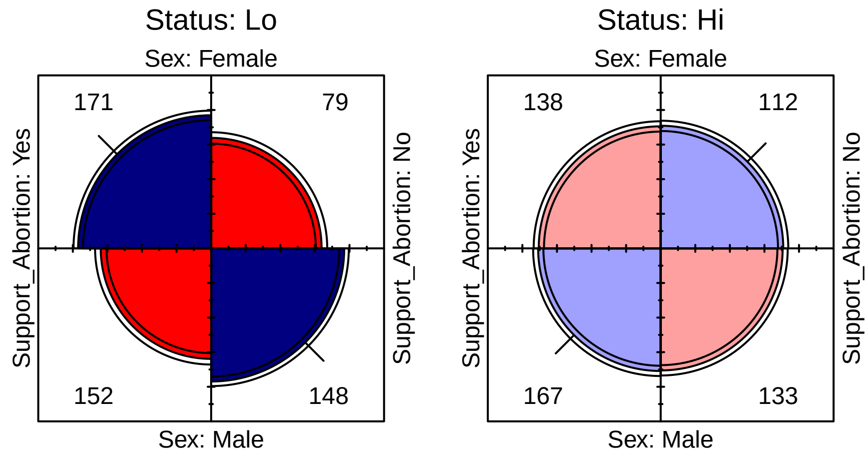
```
##
##  Fisher's Exact Test for Count Data
##
## data:  fat
## p-value = 0.03931
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.8677432 105.5669391
## sample estimates:
## odds ratio
##    7.40185
```

- The p-value shows less than 0.05, which shows the evidence to reject the null hypothesis of no association between diet and disease.
- This is the different results from chisq.test().

(e) To conclude, we can interpret the association between disease and high or low chol although the data sample is not enough (small expected frequencies). The odds ratio computed for diet and disease, at approximately 8.25, indicates a substantial increase in the likelihood of experiencing heart disease symptoms among individuals with a high cholesterol diet compared to those with a low cholesterol diet. Fisher's exact test further confirms the significance of this association, with a p-value of 0.03931 and a confidence interval for the odds ratio excluding 1. These findings underscore the importance of dietary considerations in assessing the risk factors associated with heart disease, emphasizing the potential impact of dietary cholesterol levels on cardiovascular health.

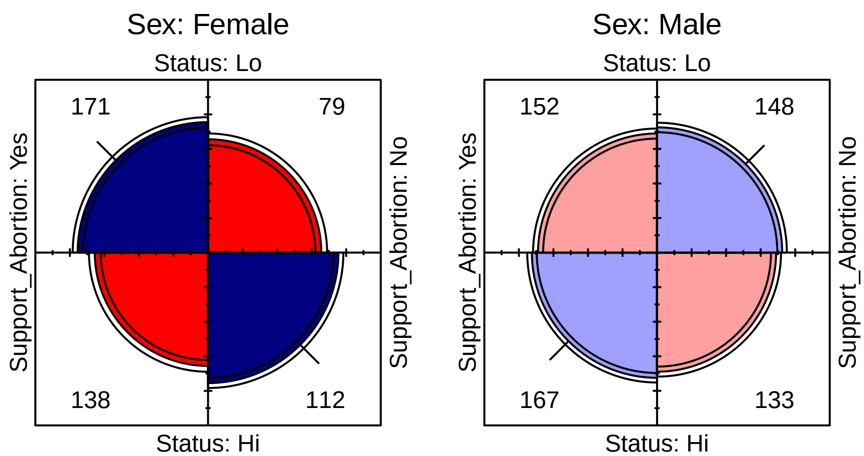
4.2

```r
# load data
data("Abortion", package = "vcdExtra")
# stratification
Abortion2 <- aperm(Abortion, c(1, 3, 2))
# plot
fourfold(Abortion2)
```



(a)

```
# stratification
Abortion3 <- aperm(Abortion, c(2, 3, 1))
fourfold(Abortion3)
```



(b)

```r
# Sex by support for abortion, stratified by status
summary(oddsratio(Abortion2))
```

(c)

```
##
## z test of coefficients:
##
##    Estimate Std. Error z value Pr(>|z|)
## Lo  0.74555    0.17844  4.1781 2.94e-05 ***
## Hi -0.01889    0.17228 -0.1096   0.9127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Status by support for abortion, stratified by sex
summary(oddsratio(Abortion3))
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error z value Pr(>|z|)
## Female  0.56346    0.18623  3.0256 0.002481 **
```

```
## Male   -0.20098    0.16384 -1.2267 0.219941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(d)

- Among low status, there is a high odds ratio compared to 1. According to the fourfold, we can see the significance of the association between Sex and support for Abortion among low status.
- Among high status, there is a slightly lower odds ratio (less than 1). Also, the fourfold plot shows a perfect circle, which means no significant association between Sex and support for abortion among high status.
- Among female, there is a higher odds ratio compared to 1. Also, the fourfold plot shows the significance of the relationship between Status and support for abortion.
- Among male, there is a low odds ratio compared to 1. However, the fourfold plot is closer to a perfect circle. So, we cannot see the significant relationship between status and support for abortion among male.

4.3

```r
# load data
data("JobSat", package = "vcdExtra")
# Standard chi-squared test
chisq_test_standard <- chisq.test(JobSat)
```

(a)

```
## Warning in chisq.test(JobSat): Chi-squared approximation may be incorrect
```

```r
chisq_test_standard
```

```
##
##  Pearson's Chi-squared test
##
## data:  JobSat
## X-squared = 5.9655, df = 9, p-value = 0.7434
```

```r
# apply simulaiton
chisq_test_simulation <- chisq.test(JobSat, simulate.p.value = TRUE)
chisq_test_simulation
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  JobSat
## X-squared = 5.9655, df = NA, p-value = 0.7856
```

- A standard chi-squared test shows no significant relationship between job satisfaction and income.
- Also, because of having low data sample in each frequency and ordinal data, this test is not appropriate option.
- After applying the Monte Carlo test, the p-value slightly increases, so there was no significant difference.

```r
CMHtest(JobSat)
```

(b)

```
## Cochran-Mantel-Haenszel Statistics for income by satisfaction
##
##                  AltHypothesis  Chisq Df     Prob
## cor        Nonzero correlation 2.9830  1 0.084144
## rmeans  Row mean scores differ 4.4774  3 0.214318
## cmeans  Col mean scores differ 3.1036  3 0.375931
## general    General association 5.9034  9 0.749549
```

- For nonzero correlation, although the p-value is lower, it is still above the common alpha evel of 0.05. So, we do not have enough evidence to reject the null hypothesis of zero correlation.
- For both row and column mean score shows slightly higher p-value of 0.21 and 0.38, respectively. These results suggest that both mean socres do not differ significantly across the strata.
- For the General association, the p-value shows a high value of 0.75, which indicates no significance relationship between job satisfaction and income.

4.6

```r
# laod data
data('Mammograms', package = 'vcdExtra')

# Unweighted kappa
unweighted_kappa <- Kappa(Mammograms)
unweighted_kappa
```

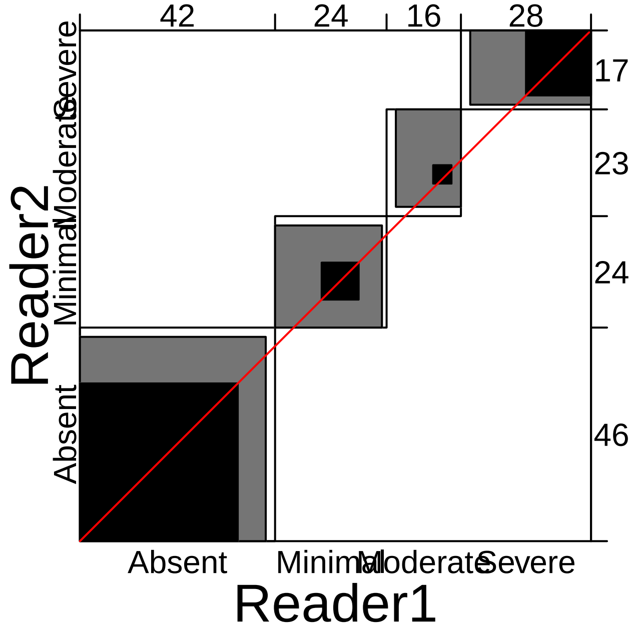
(a)

```
##             value     ASE      z  Pr(>|z|)
## Unweighted 0.3713 0.06033  6.154 7.560e-10
## Weighted   0.5964 0.04923 12.114 8.901e-34
```

```r
# Weighted kappa
weighted_kappa <- Kappa(Mammograms, weights= "Fleiss-Cohen")
weighted_kappa
```

```
##             value     ASE      z  Pr(>|z|)
## Unweighted 0.3713 0.06033  6.154 7.560e-10
## Weighted   0.7641 0.03996 19.122 1.667e-81
```

```r
# agreement plot
agreementplot(Mammograms)
```



(b)

```r
# association stats
association_stats <- assocstats(Mammograms)
association_stats
```

(c)

```
##                     X^2 df   P(> X^2)
## Likelihood Ratio 92.619  9 4.4409e-16
## Pearson          83.516  9 3.2307e-14
##
## Phi-Coefficient   : NA
## Contingency Coeff.: 0.657
## Cramer's V        : 0.503
```

- For both unweighted and weighted Kappa, the p-values indicate statistically significant values.
- The assocstats() results with a Chi-Squared value of 83.516 and a Cramer's V of 0.503 also indicate a moderate to strong association between the raters' scores.
- However, the contingency coefficient and Cramer's V assess only association, so these could be large when there is little agreement. The values of these statistics are not directly comparable

4.7

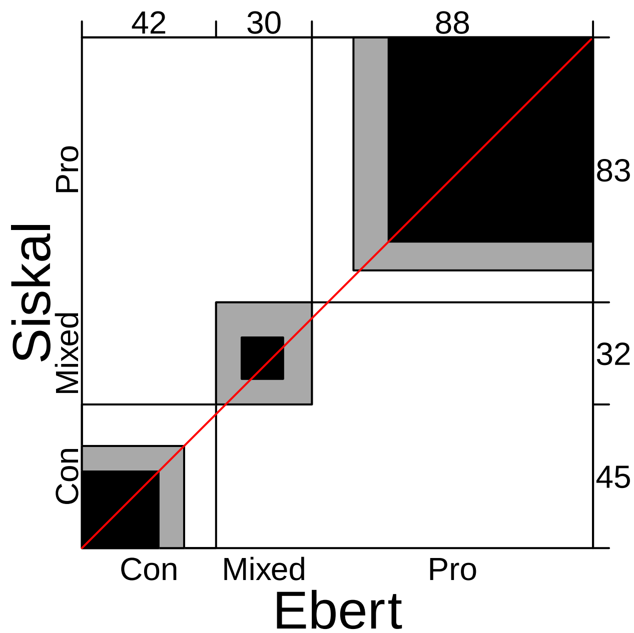
```r
ratings <- matrix(c(24, 8, 13, 8, 13, 11, 10, 9, 64), 3, 3, byrow=TRUE)
dimnames(ratings) <- list(Siskal=c("Con", "Mixed", "Pro"),Ebert =c("Con", "Mixed", "Pro"))

# kappa
Kappa(ratings)
```

(a)

```
##              value     ASE     z  Pr(>|z|)
## Unweighted 0.3888 0.05979 6.503 7.870e-11
## Weighted   0.4269 0.06350 6.723 1.781e-11
```

```r
# agreement plot
agreementplot(ratings)
```



(b)

```r
# test
T <- (ratings + t(ratings))/2
(Chisq <- sum((ratings - T)^2 / T))
```

(c)

```
## [1] 0.5913043
```

```
df <- nrow(T) * (nrow(T)-1) /2
pchisq(Chisq, df, lower.tail = FALSE)
```

```
## [1] 0.8984201
```

```
mcnemar.test(ratings)
```

(d)

```
##
##  McNemar's Chi-squared test
##
## data:  ratings
## McNemar's chi-squared = 0.5913, df = 3, p-value = 0.8984
```

- McNemar's test assesses whether the marginal frequencies of the ratings are equal. The chi-squared statistic from McNemar's test is 0.5913 with 3 degrees of freedom, resulting in a p-value of 0.8984. Since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis of equal marginal frequencies.
- Overall, there is moderate agreement between the raters according to Cohen's kappa statistics. However, the ratings are not symmetric, indicating potential differences in how the raters assign ratings. Additionally, McNemar's test does not detect significant differences in marginal frequencies between the raters. Therefore, while there is some agreement between the raters, there may be systematic differences in how they assign ratings.

4.8

```
data("VisualAcuity")

VA.tab <- xtabs(~ left + right + gender, data = VisualAcuity)
dimnames(VA.tab)[1:2] <- list(c("high", 2, 3, "low"))
names(dimnames(VA.tab))[1:2] <- paste(c("Right", "Left"), "eye grade")
VA.tab
```

(a)

```
## , , gender = male
##
##                  Left eye grade
## Right eye grade high 2 3 low
##            high    1 1 1   1
##            2       1 1 1   1
```

```
##            3       1 1 1   1
##            low     1 1 1   1
##
## , , gender = female
##
##                Left eye grade
## Right eye grade high 2 3 low
##            high    1 1 1   1
##            2       1 1 1   1
##            3       1 1 1   1
##            low     1 1 1   1
```

```r
CMHtest(VA.tab)
```

(b)

```
## $`gender:male`
## Cochran-Mantel-Haenszel Statistics for Right eye grade by Left eye grade
##  in stratum gender:male
##
##                  AltHypothesis Chisq Df Prob
## cor        Nonzero correlation     0  1    1
## rmeans  Row mean scores differ     0  3    1
## cmeans  Col mean scores differ     0  3    1
## general    General association     0  9    1
##
##
## $`gender:female`
## Cochran-Mantel-Haenszel Statistics for Right eye grade by Left eye grade
##  in stratum gender:female
##
##                  AltHypothesis Chisq Df Prob
## cor        Nonzero correlation     0  1    1
## rmeans  Row mean scores differ     0  3    1
## cmeans  Col mean scores differ     0  3    1
## general    General association     0  9    1
```

The CMH test results indicate that there is no significant association found between left and right eye acuity, with a chi-squared statistic of 0 and a p-value of 1.

```r
woolf_test(VA.tab)
```

(c)

```
##
##  Woolf-test on Homogeneity of Odds Ratios (no 3-Way assoc.)
##
## data:  VA.tab
## X-squared = 0, df = 1, p-value = 1
```

- The Woolf test results also indicate no significant difference in the association between left and right eye acuity for men and women, with a chi-squared statistic of 0 and a p-value of 1.
- In conclusion, both the CMH test and the Woolf test suggest that there is no significant association between left and right eye acuity, and there is no difference in this association between men and women in the VisualAcuity data set.