

# Homework 3

*Shun-Lung Chang, Dilip Hiremath*

```
library(knitr)
library(vcd)
library(DescTools)
library(psych)
```

```
load("data/creditcard.Rdata")
```

## 1. First of all, tabulate the CreditCard variable.

(a) (half a point) How many applicants was the credit card approved?

```
table(creditcard$CreditCard)
```

```
 0    1
383 307
```

(b) (1 point) What are the odds of getting a credit card application denied?

```
383 / 307
```

```
[1] 1.247557
```

(c) (1 point) What is the “risk” (i.e. probability) of getting a credit card application approved?

```
307 / (383 + 307)
```

```
[1] 0.4449275
```

## 2. Cross-tabulate the variables Children and CreditCard.

(a) (1 point) How many applicants with children in their household was the credit card denied?

```
170
```

```
table(creditcard$Children, creditcard$CreditCard)
```

```
      0    1
0 213 161
1 170 146
```

(b) (1 point) How many applicants without children in their household got the credit card application approved?

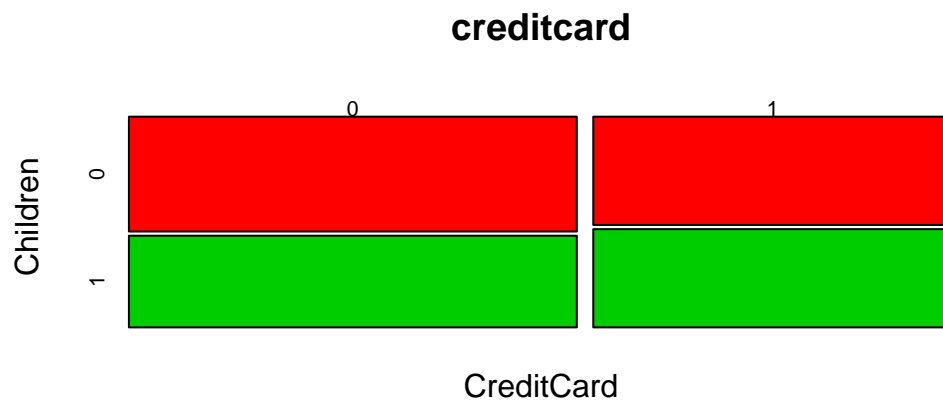
161

```
xtabs(~ creditcard$Children + creditcard$CreditCard)
```

```
      creditcard$CreditCard
creditcard$Children  0    1
0      213 161
1      170 146
```

(c) (half a point) Draw a mosaicplot visualising the contingency table of credit card approval and whether children live in the applicant's household.

```
mosaicplot(CreditCard ~ Children, data = creditcard, color = 2:3)
```



3. You continue with your analysis of the relationship between Children and CreditCard.

(a) (1 point) Are applicants with children in their household less likely (as measured in odds) to get credit card applications approved than others? Calculate the odds ratio for getting a credit card application approved comparing applicants with children in their household with those without.

```
(146 / 170) / (161 / 213)
```

```
[1] 1.136208
```

```
# or
```

```
oddsratio(xtabs(~ creditcard$Children + creditcard$CreditCard), log = FALSE)
```

```
odds ratios for creditcard$Children and creditcard$CreditCard
```

```
[1] 1.136208
```

(b) (1 point) Are applicants with children less likely (as measured in risk) to get credit card applications denied than applicants without children? Calculate the relative risk for getting a credit card application denied comparing applicants with children in their household to those without.

```
(213 / (213 + 161)) / (170 / (170 + 146))
```

```
[1] 1.058635
```

(c) (half a point) Looking at the mosaic plot created in Question 2c, how strong is the relationship between the two variables Children and CreditCard.

4. Now, you assess the relationship between Children and CreditCard using the  $\chi^2$ -statistic.

(a) (1 point) Calculate the  $\chi^2$ -statistic to assess the relationship between Children and CreditCard.

```
chisq.test(creditcard$Children, creditcard$CreditCard)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: creditcard$Children and creditcard$CreditCard
X-squared = 0.56827, df = 1, p-value = 0.4509
```

(b) (1.5 points) Calculate the expected frequencies under the assumption that home ownership has no effect on credit card approval. For which cells are expected frequencies higher than the observed ones?

```
chisq.test(creditcard$Children, creditcard$CreditCard)$expected
```

```

      creditcard$CreditCard
creditcard$Children    0     1
0 207.5971 166.4029
1 175.4029 140.5971
```

```
chisq.test(creditcard$Children, creditcard$CreditCard)$expected > xtabs(~ creditcard$Children + creditcard$CreditCard)
```

```

      creditcard$CreditCard
creditcard$Children    0     1
0 FALSE  TRUE
1  TRUE FALSE
```

5. In the following, perform the analyses separately for female and male applicants.

```
creditcard_f <- creditcard[creditcard$Gender == 0, ]
creditcard_m <- creditcard[creditcard$Gender == 1, ]
```

(a) (half a point) Calculate the  $\chi^2$ -statistic to assess the relationship between Children and CreditCard.

```
chisq.test(creditcard_f$Children, creditcard_f$CreditCard)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: creditcard_f$Children and creditcard_f$CreditCard
X-squared = 1.0837, df = 1, p-value = 0.2979
```

```
chisq.test(creditcard_m$Children, creditcard_m$CreditCard)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: creditcard_m$Children and creditcard_m$CreditCard
X-squared = 3.0337, df = 1, p-value = 0.08155
```

(b) (1.5 point) Calculate the expected frequencies under the assumption that Children has no effect on credit card approval. For which cells are expected frequencies higher than the observed ones?

```
chisq.test(creditcard_f$Children, creditcard_f$CreditCard)$expected
```

```
          creditcard_f$CreditCard
creditcard_f$Children    0         1
0 70.31081 58.68919
1 50.68919 42.31081
```

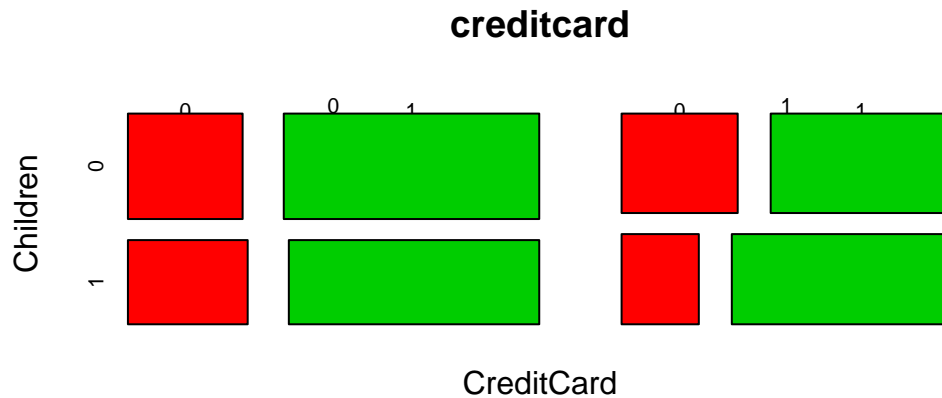
```
chisq.test(creditcard_m$Children, creditcard_m$CreditCard)$expected
```

```
          creditcard_m$CreditCard
creditcard_m$Children    0         1
0 137.1581 107.84188
1 124.8419  98.15812
```

(c) (half a point) Do the results differ for the two sexes?

6. (2.5 points) Visualise the relationships using mosaicplots. Do the differences between females and males in relation to credit card approval and children in the household become visible in the plots? Provide reason for your answer!

```
mosaicplot(CreditCard ~ Children + Gender, data = creditcard, color = 2:5)
```



7. (2.5points) Compute the odds ratios to assess the relationship between Children and CreditCard for males and females separately. Are the results more in line with the  $\chi^2$ -statistic or more in line with the mosaicplot?

```
oddsratio(xtabs(~ creditcard_f$Children + creditcard_f$CreditCard), log = FALSE)
```

odds ratios for creditcard\_f\$Children and creditcard\_f\$CreditCard

```
[1] 0.7238095
```

```
oddsratio(xtabs(~ creditcard_m$Children + creditcard_m$CreditCard), log = FALSE)
```

odds ratios for creditcard\_m\$Children and creditcard\_m\$CreditCard

```
[1] 1.408696
```

8. Now, you assess the relationships between SavingsType and CreditCard.

(a) (half a point) Calculate the  $\chi^2$ -statistic to assess the relationship between SavingsType and CreditCard.

```
chisq.test(creditcard$SavingsType, creditcard$CreditCard)
```

Pearson's Chi-squared test

data: creditcard\$SavingsType and creditcard\$CreditCard  
X-squared = 9.1916, df = 2, p-value = 0.01009

(b) (1 point) Calculate the Phi-coefficient, the Contingency Coefficient and Cramer's V to assess the relationship between SavingsType and CreditCard.

```
# phi(table(creditcard$SavingsType, creditcard$CreditCard)) # cannot be calculated, since the size of c
relation <- list("Contingency Coefficient" =
ContCoef(x = creditcard$SavingsType, y = creditcard$CreditCard),
"Cramer's V" = CramerV(x = creditcard$SavingsType, y = creditcard$CreditCard))
```

```
relation
```

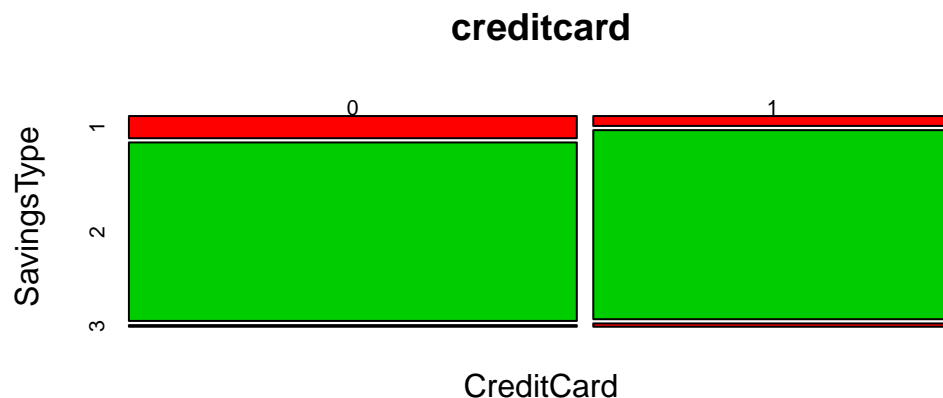
```
$`Contingency Coefficient`  
[1] 0.114656
```

```
$`Cramer's V`  
[1] 0.1154171
```

(c) (half a point) Why is there no result for the Phi-Coefficient?

(d) (half a point) Visualize the relationship between credit card approval and savings type. Which of the statistics used for this data comes closer to the visual representation of the relationship's strength?

```
mosaicplot(CreditCard ~ SavingsType, data = creditcard, color = 2:3)
```



9. (2.5 points) Since there are too few cases for savings type 3 (CDs), you exclude applicants having this savings type and re-run the analysis of Question 8 again. Summarize the results and comment on the differences.

```
# remove savings type 3  
creditcard_remove_3 <- creditcard[!creditcard$SavingsType == 3, ]
```

```
chisq.test(creditcard_remove_3$SavingsType, creditcard_remove_3$CreditCard)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: creditcard_remove_3$SavingsType and creditcard_remove_3$CreditCard  
X-squared = 7.3615, df = 1, p-value = 0.006663
```

```
relation <- list("Phi-coefficient" = phi(table(creditcard_remove_3$SavingsType, creditcard_remove_3$CreditCard)),  
                "Contingency Coefficient" =  
ContCoef(x = creditcard_remove_3$SavingsType, y = creditcard_remove_3$CreditCard),  
          "Cramer's V" = CramerV(x = creditcard_remove_3$SavingsType, y = creditcard_remove_3$CreditCard))  
relation
```

```
$`Phi-coefficient`
```

```
[1] 0.11
```

```
$`Contingency Coefficient`
```

```
[1] 0.1085816
```

```
$`Cramer's V`
```

```
[1] 0.1092274
```

```
mosaicplot(CreditCard ~ SavingsType, data = creditcard_remove_3, color = 2:3)
```

