

Chapter 10: Bivariate Analysis

Exercises - Alternative with SIMD

Brian Fogarty

26 August 2018

Contents

EXERCISE I	1
ANSWERS FOR EXERCISE I	1
Question 1.1.a	2
Question 1.1.b	2
Question 1.1.c	2
Question 1.1.d	3
Question 1.2.1	3
Question 1.2.2	4
Question 1.2.3	5
EXERCISE II	6
ANSWERS FOR EXERCISE II	7
Question 2.1.a	7
Question 2.1.b	7
Question 2.1.c	7
Question 2.1.d	8
Question 2.1.e	8
Question 2.1.f	8
Question 2.2.a	9
Question 2.2.b	10
Question 2.2.c	11
Question 2.2.d	12
Question 2.2.e	13
EXERCISE III	13
Question 3.1	14
Question 3.2	14
Question 3.3	15

EXERCISE I

Using an abbreviated version of the 2011 UK Census dataset (`2011 UK Census.csv`), perform the following exercises. Note: the dataset is a sample of the UK Census consisting of nearly 570,000 observations and 21 variables.

1. Perform recoding and labeling on the following variables:
 - (a) `Sex` - rename as `gender` and label the values as `1='0. Male';2='1. Female'`.
 - (b) `Marital.Status` - recode to create a dummy variable where `1 = 'married'` and all other values equal 0. Name the dummy variable `married` and label the values as `1='1. Married'; 0='0. Not Married'`.

- (c) `Ethnic.Group` - recode to create a dummy variable where 1 = 'white' and all other values equal 0; also get rid of any missing values and non-responses. Name the dummy variable `white` and label the values as 1='1. White'; 0='0. Not White'.
 - (d) `Approximated.Social.Grade` - rename as `class`, get rid of any missing values and non-responses, and label the values as 1='1. AB'; 2='2. C1'; 3='3. C2'; 4='4. DE'. The letters refer to the respondents approximated social class where AB is the highest class.
2. Using crosstabs, perform Chi-Square analysis with an appropriate measure of association between the outcome variable `class` and the other three recoded variables from the previous question. Discuss whether the relationships are statistically significant, interpret the measure of association, and include a plain language discussion.

ANSWERS FOR EXERCISE I

```
setwd("C:/QSSD/Exercises/Chapter 10 - Exercises")
getwd()
```

```
[1] "C:/QSSD/Exercises/Chapter 10 - Exercises"
```

```
census <- read.csv("2011 UK Census.csv", na="NA")
names(census)
```

```
[1] "Person.ID"           "Region"
[3] "Residence.Type"     "Family.Composition"
[5] "Population.Base"    "Sex"
[7] "Age"                "Marital.Status"
[9] "Student"            "Country.of.Birth"
[11] "Health"              "Ethnic.Group"
[13] "Religion"           "Economic.Activity"
[15] "Occupation"         "Industry"
[17] "Hours.worked.per.week" "Approximated.Social.Grade"
```

Question 1.1.a

Gender

```
table(census$Sex)
```

```
      1      2
280569 289172
```

```
library(car)
```

Loading required package: carData

```
census$gender <- recode(census$Sex, "1='0. Male';2='1. Female'")
table(census$gender)
```

```
0. Male 1. Female
280569   289172
```

Question 1.1.b

Married Dummy

```
table(census$Marital.Status)
```

```
      1      2      3      4      5  
270999 214180 11951 40713 31898
```

```
census$married <- recode(census$Marital.Status, "1='0. Not Married';2='1. Married';  
3:5='0. Not Married'")  
table(census$married)
```

```
0. Not Married      1. Married  
      355561          214180
```

Question 1.1.c

White Dummy

```
table(census$Ethnic.Group)
```

```
      -9      1      2      3      4      5  
6804 483477 12209 42712 18786 5753
```

```
census$white <- recode(census$Ethnic.Group, "1='1. White';2:5='0. Not White';  
-9=NA")  
table(census$white)
```

```
0. Not White      1. White  
      79460          483477
```

Question 1.1.d

Social Grade

```
table(census$Approximated.Social.Grade)
```

```
      -9      1      2      3      4  
124103 82320 159642 79936 123740
```

```
census$class <- recode(census$Approximated.Social.Grade, "1='1. AB';2='2. C1';3='3. C2';  
4='4. DE';-9=NA")  
table(census$class)
```

```
1. AB 2. C1 3. C2 4. DE  
82320 159642 79936 123740
```

Question 1.2.1

Class and Gender. Since `gender` is a nominal-level variable we need to use Cramer's V for our measure of association.

```
library(descr)
CrossTable(census$class, census$gender, prop.c=TRUE,
           prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
           total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

Cell Contents

```
|-----|
|                N |
|      N / Col Total |
|-----|
```



```
=====
census$gender
census$class  0. Male  1. Female
-----
1. AB          44240    38080
              0.204    0.167
-----
2. C1          67435    92207
              0.311    0.403
-----
3. C2          51181    28755
              0.236    0.126
-----
4. DE          54262    69478
              0.250    0.304
=====
```

Statistics for All Table Factors

Pearson's Chi-squared test

Chi² = 12183.8 d.f. = 3 p <2e-16

```
library(DescTools)
```

Attaching package: 'DescTools'

The following object is masked from 'package:car':

Recode

```
CramerV(census$gender, census$class)
```

```
[1] 0.1653485
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between `gender` and `class`. The Cramer's V value is .165 indicating a weak relationship between `gender` and `class`. Combining the Chi-Squared test and Cramer's V, we can now say *there is a weak, but statistically significant relationship between individuals' gender and social class*. For a plain language discussion, we may say *there is a higher percentage of men (20.4%) than women (16.7%) who are at the highest social class. At the other extreme,*

there is a higher percentage of women (30.4%) than men (25%) who are at the lowest social class. But, a large plurality of women make up the higher middle class (C1) relative to men.

Question 1.2.2

Class and Married. Since `married` is a nominal-level variable we need to use Cramer's V for our measure of association.

```
CrossTable(census$class, census$married, prop.c=TRUE,
            prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
            total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```
Cell Contents
|-----|
|                N |
|      N / Col Total |
|-----|

=====
census$class      census$married
                  0. Not Married  1. Married
-----
1. AB                32229         50091
                   0.138         0.235
-----
2. C1                92921         66721
                   0.399         0.313
-----
3. C2                36439         43497
                   0.157         0.204
-----
4. DE                71210         52530
                   0.306         0.247
=====
```

Statistics for All Table Factors

Pearson's Chi-squared test

```
-----
Chi^2 = 10746.32      d.f. = 3      p <2e-16
```

```
CramerV(census$married,census$class)
```

```
[1] 0.1552883
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between `married` and `class`. The Cramer's V value is .165 indicating a weak relationship between `married` and `class`. Combining the Chi-Squared test and Cramer's V, we can say *there is a weak, but statistically significant relationship between individuals' marital status and social class*. For a plain language discussion, we may say *there is a higher percentage of married (23.5%) than unmarried individuals (13.8%) who are at the highest social class. At the other extreme, there is a higher percentage of unmarried (30.6%) than unmarried individuals (24.7%) who are at the lowest social class. The middle classes have somewhat similar percentages of married and unmarried individuals.*

Question 1.2.3

Class and White. Since `white` is a nominal-level variable we need to use Cramer's V for our measure of association.

```
CrossTable(census$class, census$white, prop.c=TRUE,
            prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
            total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```
Cell Contents
|-----|
|              N |
|      N / Col Total |
|-----|

=====
              census$white
census$class  0. Not White  1. White
-----
1. AB                9862      72458
                   0.180      0.185
-----
2. C1                22012     137630
                   0.402      0.352
-----
3. C2                 6640      73296
                   0.121      0.187
-----
4. DE                16196     107544
                   0.296      0.275
=====
```

Statistics for All Table Factors

Pearson's Chi-squared test

```
-----
Chi^2 = 1591.084      d.f. = 3      p <2e-16
```

```
CramerV(census$white,census$class)
```

```
[1] 0.05975241
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between `white` and `class`. The Cramer's V value is .060 indicating a very weak relationship between `white` and `class`. Combining the Chi-Squared test and Cramer's V, we can say *there is a very weak, but statistically significant relationship between race and social class*. For a plain language discussion, we may say *there are only small differences between whites and non-whites across the classes with a slightly higher percentage of whites who are at the highest social class and a slightly higher percentage of non-whites who are the lowest social class*. You may be surprised that the relationship is significant given the small number of differences across the classes for race. This is because our N is so large that we are more likely to find statistical significance, but the weakness of the relationship indicates that we do not have substantive significance.

EXERCISE II

Using an abbreviated version of 2015 UK Millennium Cohort survey dataset (`mcs.dta`), perform the following exercises. Note: the survey was carried out in 2015 to 14 year-old pupils in the UK. The dataset consists of nearly 12,000 observations and 52 variables. You need to use the `haven` package to read-in the data.

1. Perform recoding and labeling on the variables below. Note: for all the variables, you need to first convert them to factors using the `as.factor()` function. (This is the same as Exercise I from Chapter 5).
 - (a) `mths` - rename as `maths` and label the values as 1='1. Strongly Disagree'; 2='2. Disagree'; 3='3. Agree'; 4='4. Strongly Agree'. This variable includes pupils' responses to whether they were good at mathematics.
 - (b) `scien` - rename as `science` and label the values as 1='1. Strongly Disagree'; 2='2. Disagree'; 3='3. Agree'; 4='4. Strongly Agree'. This variable includes pupils' responses to whether they were good at science.
 - (c) `sex` - rename as `gender` and label the values as 0='0. Female'; 1='1. Male'.
 - (d) `best` - rename as `bestsch` and label the values as 1='1. Never'; 2='2. Sometimes'; 3='3. Most Times'; 4='4. Always'. This variable includes pupils' responses to how often they do their best at school.
 - (e) `games` - rename as `vidgames` and label the values as 1='1. Never'; 2='2. Less Half Hr'; 3='3. Half Hr to Hr'; 4='4. 1-2 Hrs'; 5='5. 2-3 Hrs'; 6='6. 3-5 Hrs'; 7='7. 5-7 Hrs'; 8='8. More 7 Hrs'. This variable includes pupils' responses for many hours per weekday do they play video games.
 - (f) `sibl_f1` - rename as `siblings` and recode to create a dummy variable which equals 0 if the pupil has no siblings and equals 1 if the pupil has 1 or more siblings - 0='0. No Siblings'; 1:10='1. Siblings'.
2. Using crosstabs, perform Chi-Square analysis with an appropriate measure of association between the outcome variable `maths` and the other five recoded variables from the previous question. Discuss whether the relationships are statistically significant, interpret the measure of association, and include a plain language discussion.

ANSWERS FOR EXERCISE II

Read-in 2015 Millennium Cohort Study data.

```
library(haven)
mcs <- read_dta("mcs.dta")
```

Question 2.1.a

```
library(car)
mcs$mths <- as.factor(mcs$mths)
table(mcs$mths)
```

```
 1    2    3    4
598 1827 5958 3118
```

```
mcs$maths <- recode(mcs$mths, "1='1. Strongly Disagree';2='2. Disagree';
                             3='3. Agree';4='4. Strongly Agree'")
table(mcs$maths)
```

1. Strongly Disagree	2. Disagree	3. Agree
598	1827	5958
4. Strongly Agree		
3118		

Question 2.1.b

```
mcs$scien <- as.factor(mcs$scien)
table(mcs$scien)
```

1	2	3	4
500	1993	6166	2834

```
mcs$science <- recode(mcs$scien, "1='1. Strongly Disagree';2='2. Disagree';
                                 3='3. Agree';4='4. Strongly Agree'")
table(mcs$science)
```

1. Strongly Disagree	2. Disagree	3. Agree
500	1993	6166
4. Strongly Agree		
2834		

Question 2.1.c

```
mcs$sex <- as.factor(mcs$sex)
table(mcs$sex)
```

0	1
5926	5946

```
mcs$gender <- recode(mcs$sex, "0='0. Female';1='1. Male'")
table(mcs$gender)
```

0. Female	1. Male
5926	5946

Question 2.1.d

```
mcs$best <- as.factor(mcs$best)
table(mcs$best)
```

1	2	3	4
---	---	---	---

35 1058 6469 3937

```
mcs$bestsch <- recode(mcs$best, "1='1. Never';2='2. Sometimes';  
3='3. Most Times';4='4. Always'")  
table(mcs$bestsch)
```

1. Never	2. Sometimes	3. Most Times	4. Always
35	1058	6469	3937

Question 2.1.e

```
mcs$games <- as.factor(mcs$games)  
table(mcs$games)
```

1	2	3	4	5	6	7	8
2160	1565	1343	1751	1583	1478	770	862

```
mcs$vidgames <- recode(mcs$games, "1='1. Never';2='2. Less Half Hr';3='3. Half Hr to Hr';  
4='4. 1-2 Hrs'; 5='5. 2-3 Hrs';6='6. 3-5 Hrs';7='7. 5-7 Hrs';  
8='8. More 7 Hrs'")  
table(mcs$vidgames)
```

1. Never	2. Less Half Hr	3. Half Hr to Hr	4. 1-2 Hrs
2160	1565	1343	1751
5. 2-3 Hrs	6. 3-5 Hrs	7. 5-7 Hrs	8. More 7 Hrs
1583	1478	770	862

Question 2.1.f

```
table(mcs$sibl_fl)
```

0	1	2	3	4	5	6	7	8	9	10
1678	5203	2990	1322	417	178	50	21	10	2	1

```
mcs$siblings <- recode(mcs$sibl_fl, "0='0. No Siblings';1:10='1. Siblings'")  
table(mcs$siblings)
```

0. No Siblings	1. Siblings
1678	10194

Question 2.2.a

Maths and Science. Since science is an ordinal-level variable with the same number of values as maths, we will use Kendall's Tau-B for our measure of association.

```
library(descr)  
CrossTable(mcs$maths, mcs$science, prop.c=TRUE,  
prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,  
total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```

Cell Contents
|-----|
|                N |
|      N / Col Total |
|-----|
=====
mcs$maths      mcs$science
1. Strngly Dsg  230      144      189      33
                0.460      0.072      0.031      0.012
-----
2. Disagree    97       628      915      187
                0.194      0.315      0.148      0.066
-----
3. Agree      145      1025     3782     999
                0.290      0.515      0.614      0.353
-----
4. Strngly Agr 28       195      1278     1615
                0.056      0.098      0.207      0.570
=====

```

Statistics for All Table Factors

Pearson's Chi-squared test

```
-----
Chi^2 = 3871.416      d.f. = 9      p <2e-16
```

```
library(DescTools)
KendallTauB(mcs$science,mcs$maths)
```

```
[1] 0.3919121
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between `science` and `maths`. The Kendall's Tau-B value is .392 indicating a positive moderate relationship between `science` and `maths`. Combining the Chi-Squared test and Kendall's Tau-B, we can say *there is a positive moderate statistically significant relationship between pupils' self-evaluation of their science and mathematics ability*. For a plain language discussion, we may say *pupils' self-evaluation of their science ability is moderately related to their self-evaluation of their mathematics ability. Pupils tended to give similar responses on their ability in science and maths, though not identical. For example, 46% of pupils who strongly disagreed they were good in science also responded they strongly disagreed they were good in maths. While 57% of pupils who strongly agreed they were good in science also responded they strongly agreed they were good in maths.*

Question 2.2.b

Maths and Gender. Since `gender` is a nominal-level variable we need to use Cramer's V for our measure of association.

```
CrossTable(mcs$maths, mcs$gender, prop.c=TRUE,
           prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
           total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```

Cell Contents
|-----|

```

	N	
	N / Col Total	
=====		
	mcs\$gender	
mcs\$maths	0. Female	1. Male

1. Strongly Disagree	364	234
	0.063	0.041

2. Disagree	1104	723
	0.191	0.126

3. Agree	3012	2946
	0.522	0.514

4. Strongly Agree	1293	1825
	0.224	0.319
=====		

Statistics for All Table Factors

Pearson's Chi-squared test

Chi² = 199.0432 d.f. = 3 p <2e-16

```
CramerV(mcs$gender,mcs$maths)
```

```
[1] 0.1315545
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between **gender** and **maths**. The Cramer's V value is .132 indicating a weak relationship between **gender** and **maths**. Combining the Chi-Squared test and Cramer's V, we can now say *there is a weak, but statistically significant relationship between gender and pupils' self-evaluation of their mathematics ability*. For a plain language discussion, we may say *male pupils have a somewhat better self-evaluation of their mathematics ability compared to female pupils. The differences are not dramatic, but we find that 31.9% of male pupils strongly agreed they were good at maths, while only 22.4% of female pupils strongly agreed*.

Question 2.2.c

Maths and Bestsch. Since **bestsch** is an ordinal-level variable with a different number of values than **maths**, we will use Goodman and Kruskal's Gamma for our measure of association.

```
CrossTable(mcs$maths, mcs$bestsch, prop.c=TRUE,
            prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
            total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

Warning in chisq.test(tab, correct = FALSE, ...): Chi-squared approximation may be incorrect

Cell Contents

	N
	N / Col Total

```
=====
```

mcs\$maths	mcs\$bestsch			
	1. Never	2. Sometimes	3. Most Times	4. Always
1. Strongly Disagree	6 0.176	105 0.100	288 0.045	196 0.050
2. Disagree	9 0.265	309 0.293	1085 0.168	424 0.108
3. Agree	12 0.353	494 0.469	3515 0.544	1928 0.490
4. Strongly Agree	7 0.206	145 0.138	1577 0.244	1386 0.352

```
=====
```

Statistics for All Table Factors

Pearson's Chi-squared test

```
Chi^2 = 457.5432      d.f. = 9      p <2e-16
```

```
chisq.test(mcs$maths,mcs$bestsch, simulate.p.value = TRUE)
```

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

```
data: mcs$maths and mcs$bestsch
X-squared = 457.54, df = NA, p-value = 0.0004998
```

```
GoodmanKruskalGamma(mcs$bestsch,mcs$maths)
```

```
[1] 0.2597167
```

In this instance, we need to use the `chisq.test()` function. Since the p -value is below .05, we conclude there is a statistically significant relationship between `bestsch` and `maths`. The Goodman and Kruskal's value is .260 indicating a positive weak relationship between `bestsch` and `maths`. Combining the Chi-Squared test and Goodman and Kruskal's Gamma, we can say *there is a positive weak statistically significant relationship between pupils' effort in school and self-evaluation of their mathematics ability*. For a plain language discussion, we may say *pupils' effort level in school is weakly related to their self-evaluation of their mathematics ability. As we might expect, pupils who responded they never or only sometimes give their best effort had the highest percentages of strongly disagree/disagree they were good in maths and vice-versa. However, only a few pupils responded they never give their best effort and thus the percentages might not be the most reliable for this group.*

Question 2.2.d

Maths and Vidgames. Since `vidgames` is an ordinal-level variable with a different number of values than `maths`, we will use Goodman and Kruskal's Gamma for our measure of association.

```
CrossTable(mcs$maths, mcs$vidgames, prop.c=TRUE,
            prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
            total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```

Cell Contents
|-----|
|                N |
|      N / Col Total |
|-----|
=====
mcs$m          mcs$vidgames
1. Nv  2. L  3. H  4. 1-  5. 2-  6. 3-  7. 5-  8. M
-----
1. S      123    85    56    91    80    68    40    54
          0.057  0.054  0.042  0.052  0.051  0.046  0.052  0.063
-----
2. Ds      377    265   193   269   243   202   137   139
          0.175  0.170  0.144  0.154  0.154  0.137  0.178  0.161
-----
3. Ag     1085    794    738   934   807   798   363   435
          0.504  0.508  0.550  0.534  0.511  0.541  0.473  0.505
-----
4. S       569    419    354   455   449   408   228   233
          0.264  0.268  0.264  0.260  0.284  0.276  0.297  0.271
=====

```

Statistics for All Table Factors

Pearson's Chi-squared test

```
-----
Chi^2 = 34.53909      d.f. = 21      p = 0.0317
```

```
GoodmanKruskalGamma(mcs$vidgames,mcs$maths)
```

```
[1] 0.020809
```

Since the p -value is below .05, we conclude there is a statistically significant relationship between `vidgames` and `maths`. The Goodman and Kruskal's value is .021 indicating a positive very weak relationship between `vidgames` and `maths`. Combining the Chi-Squared test and Goodman and Kruskal's Gamma, we can say *there is a positive very weak statistically significant relationship between the number of hours pupils play video games during week days and self-evaluation of their mathematics ability*. For a plain language discussion, we may say *pupils who play video games for more hours during week days tend to have very slightly higher self-evaluation of their maths ability*. This is another example where because our N is so large that we are more likely to find statistical significance, but the weakness of the relationship indicates that we do not have substantive significance.

Question 2.2.e

Maths and Siblings. Since `siblings` is a nominal-level variable we need to use Cramer's V for our measure of association.

```
CrossTable(mcs$maths, mcs$siblings, prop.c=TRUE,
            prop.t=FALSE, prop.r=FALSE, prop.chisq = FALSE,
            total.r=FALSE, total.c=FALSE, chisq=TRUE)
```

```
Cell Contents
|-----|
```

	N	
	N / Col Total	
	=====	
	mcs\$siblings	
mcs\$maths	0. No Siblings	1. Siblings
-----	-----	-----
1. Strongly Disagree	88	510
	0.055	0.052
-----	-----	-----
2. Disagree	266	1561
	0.166	0.158
-----	-----	-----
3. Agree	809	5149
	0.505	0.520
-----	-----	-----
4. Strongly Agree	439	2679
	0.274	0.271
	=====	

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 1.584153 d.f. = 3 p = 0.663

`CramerV(mcs$siblings,mcs$maths)`

[1] 0.01173629

Since the p -value is above .05, we conclude there is NOT a statistically significant relationship between siblings and maths.

EXERCISE III

Using the 2016 SIMD data (`simd.csv`), perform the following exercises.

1. Perform a correlation analysis between `pct_depress` and `pct_school_attend`. Is the correlation statistically significant? If yes, interpret the correlation coefficient.
2. Perform a correlation analysis between `pct_depress` and `pct_low_bw`. Is the correlation statistically significant? If yes, interpret the correlation coefficient.
3. Perform a correlation analysis between `pct_depress` and `pct_no_heating`. Is the correlation statistically significant? If yes, interpret the correlation coefficient.

Question 3.1

```
simd <- read.csv("simd.csv")
cor.test(simd$pct_depress, simd$pct_school_attend)
```

Pearson's product-moment correlation

```
data: simd$pct_depress and simd$pct_school_attend
t = -59.269, df = 6972, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5942211 -0.5630014
sample estimates:
      cor
-0.5788233
```

Since, the p -value is below .05, we can reject the null and conclude that there is a *statistically significant correlation between datazones' school attendance and psychiatric prescriptions percentages*. We find a correlation of -.579, which indicates a *moderate negative relationship between the percentage of school attendance and the percentage of psychiatric prescriptions in Scottish datazones*.

Question 3.2

```
cor.test(simd$pct_depress, simd$pct_low_bw)
```

Pearson's product-moment correlation

```
data: simd$pct_depress and simd$pct_low_bw
t = 19.511, df = 6971, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2051775 0.2496922
sample estimates:
      cor
0.2275537
```

Since, the p -value is below .05, we can reject the null and conclude that there is a *statistically significant correlation between datazones' low birth weight and psychiatric prescriptions percentages*. We find a correlation of .228, which indicates a *weak positive relationship between the percentage of low birth weight and the percentage of psychiatric prescriptions in Scottish datazones*.

Question 3.3

```
cor.test(simd$pct_depress, simd$pct_no_heating)
```

Pearson's product-moment correlation

```
data: simd$pct_depress and simd$pct_no_heating
t = -3.0617, df = 6972, p-value = 0.002209
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.06006239 -0.01318456
sample estimates:
      cor
-0.03664363
```

Since, the p -value is below .05, we can reject the null and conclude that there is a *statistically significant correlation between datazones' homes with no central heating and psychiatric prescriptions percentages*. We find a correlation of -.037, which indicates a *very weak negative relationship between the percentage of homes with no central heating and the percentage of psychiatric prescriptions in Scottish datazones*.