**Name:** AADITI ROKADE **Date: 11/09/2018**

**SCORE:**

**Exercise 1: Measures of Association for Categorical Variables.**

Use the GSS06 dataset and answer the following questions. It is important that you set up our cross tabulation tables (ROW BY Column; Dependent BY Independent). This means put your dependent variable in the Row(s) box and your independent variable in the Column(s) box. Otherwise it may be difficult for you to know which values to report and how to describe the relationship.

Indicate the strength of the association for the following tables. If a measure does not apply, indicate so by writing N/A for your answer. Remember to consider the size and the shape of the table when deciding which measures to apply.

1. **Whether anti-religionists should be allowed to teach *(‘colath’)* by gender *(‘sex’)***

**Copy and paste all appropriate R output here. *Please add space as needed.***

**> # Create a new data frame selecting only 'colath’ and ‘sex':**

**> col\_sex\_2006 <- GSS2006 %>%**

**+ select(sex, colath) %>%**

**+ filter(!is.na(colath)) %>%**

**+ filter(!is.na(sex))**

**Warning message:**

**package ‘bindrcpp’ was built under R version 3.4.4**

**> t1\_06 <- table(col\_sex\_2006$colath, col\_sex\_2006$sex) %>%**

**+ print()**

**female male**

**allowed 678 539**

**not allowed 429 302**

**> oii.xtab(t1\_06, col = TRUE, row = TRUE,stats = TRUE,**

**+ varnames = c("anti-religioinsts teach", "Sex"))**

Cross-tabulation of anti-religioinsts teach (rows) and Sex (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1948

|

| female | male | Row Total |

-------------|-----------|-----------|-----------|

allowed | 678 | 539 | 1217 |

| 55.71% | 44.29% | 62.47% |

| 61.25% | 64.09% | |

-------------|-----------|-----------|-----------|

not allowed | 429 | 302 | 731 |

| 58.69% | 41.31% | 37.53% |

| 38.75% | 35.91% | |

-------------|-----------|-----------|-----------|

Column Total | 1107 | 841 | 1948 |

| 56.83% | 43.17% | |

-------------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 1.648577 d.f. = 1 p = 0.1991527

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

------------------------------------------------------------

Chi^2 = 1.529508 d.f. = 1 p = 0.2161863

Minimum expected frequency: 315.5909

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.028

Contingency coefficient: 0.028

Cramer's V: 0.028

Ordinal measures of association:

Total number of pairs: 1896378

Concordant pairs: 204756 ( 10.8 %)

Discordant pairs: 231231 ( 12.19 %)

Tied on first variable: 495000 ( 26.1 %)

Tied on second variable: 453640 ( 23.92 %)

Tied on both variables: 511751 ( 26.99 %)

Goodman-Kruskal Gamma: -0.061

Somers' d (col dep.): -0.030

Kendall's tau-b: -0.029

Stuart's tau-c: -0.028

Goodman-Kruskal Lambda:

Row dependent: 0.000

Column dependent: 0.000

**What conclusions can you make from your output?**

Since it is a 2x2 table we look at the value of phi. From the value 0.028 of phi obtained for a alpha of 0.1991527 (i.e. at around 99% confidence level), we can confidently say that there’s an association between an attitude towards ‘anti-religionists should be allowed to teach’ and the ‘gender’ of the survey participant. Although, from the value it can be stated that this association is extremely weak (not generally useful)

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>

1. **How close do you feel to blacks *(‘closeblk’)* by race *(‘race’)?***

**Copy and paste all appropriate R output here. *Please add space as needed.***

**> # Create a new data table selecting only 'closeblk'’ and ‘race’**

**> cblk\_race\_2006 <- GSS2006 %>%**

**+ select(closeblk, race) %>%**

**+ filter(!is.na(closeblk)) %>%**

**+ filter(!is.na(race))**

**> unique(GSS2006$race)**

**[1] "black" "other" "white"**

**> unique(GSS2006$closeblk)**

**[1] "7" NA "very close"**

**[4] "neither one or the other" "8" "6"**

**[7] "4" "3" "not at all close"**

**[10] "2"**

**> cblk\_race\_2006$closeblk <- cblk\_race\_2006$closeblk %>%**

**+ recode("not at all close" = "1 not at all close",**

**+ "neither one or the other" = "5 neither one or the other",**

**+ "very close" = "9 very close")**

**> t2\_06 <- table(cblk\_race\_2006$closeblk, cblk\_race\_2006$race) %>%**

**+ print()**

black other white

1 not at all close 0 35 72

2 1 10 31

3 9 4 41

4 3 8 58

5 neither one or the other 48 130 713

6 4 15 145

7 22 21 217

8 19 5 69

9 very close 165 29 118

**> oii.xtab(t2\_06, col = TRUE, row = TRUE, stats = TRUE,**

**+ varnames = c("Close to black people", "Race"))**

Cross-tabulation of Close to black people (rows) and Race (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1992

|

| black | other | white | Row Total |

---------------------------|-----------|-----------|-----------|-----------|

1 not at all close | 0 | 35 | 72 | 107 |

| 0.00% | 32.71% | 67.29% | 5.37% |

| 0.00% | 13.62% | 4.92% | |

---------------------------|-----------|-----------|-----------|-----------|

2 | 1 | 10 | 31 | 42 |

| 2.38% | 23.81% | 73.81% | 2.11% |

| 0.37% | 3.89% | 2.12% | |

---------------------------|-----------|-----------|-----------|-----------|

3 | 9 | 4 | 41 | 54 |

| 16.67% | 7.41% | 75.93% | 2.71% |

| 3.32% | 1.56% | 2.80% | |

---------------------------|-----------|-----------|-----------|-----------|

4 | 3 | 8 | 58 | 69 |

| 4.35% | 11.59% | 84.06% | 3.46% |

| 1.11% | 3.11% | 3.96% | |

---------------------------|-----------|-----------|-----------|-----------|

5 neither one or the other | 48 | 130 | 713 | 891 |

| 5.39% | 14.59% | 80.02% | 44.73% |

| 17.71% | 50.58% | 48.70% | |

---------------------------|-----------|-----------|-----------|-----------|

6 | 4 | 15 | 145 | 164 |

| 2.44% | 9.15% | 88.41% | 8.23% |

| 1.48% | 5.84% | 9.90% | |

---------------------------|-----------|-----------|-----------|-----------|

7 | 22 | 21 | 217 | 260 |

| 8.46% | 8.08% | 83.46% | 13.05% |

| 8.12% | 8.17% | 14.82% | |

---------------------------|-----------|-----------|-----------|-----------|

8 | 19 | 5 | 69 | 93 |

| 20.43% | 5.38% | 74.19% | 4.67% |

| 7.01% | 1.95% | 4.71% | |

---------------------------|-----------|-----------|-----------|-----------|

9 very close | 165 | 29 | 118 | 312 |

| 52.88% | 9.29% | 37.82% | 15.66% |

| 60.89% | 11.28% | 8.06% | |

---------------------------|-----------|-----------|-----------|-----------|

Column Total | 271 | 257 | 1464 | 1992 |

| 13.60% | 12.90% | 73.49% | |

---------------------------|-----------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 567.2916 d.f. = 16 p = 0.00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000001958888

Minimum expected frequency: 5.418675

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.534

Contingency coefficient: 0.471

Cramer's V: 0.377

Ordinal measures of association:

Total number of pairs: 1983036

Concordant pairs: 218567 ( 11.02 %)

Discordant pairs: 444780 ( 22.43 %)

Tied on first variable: 179292 ( 9.04 %)

Tied on second variable: 813055 ( 41 %)

Tied on both variables: 327342 ( 16.51 %)

Goodman-Kruskal Gamma: -0.341

Somers' d (col dep.): -0.153

Kendall's tau-b: -0.203

Stuart's tau-c: -0.171

Goodman-Kruskal Lambda:

Row dependent: 0.106

Column dependent: 0.089

**What observations can you make?**

Since it is **NOT** a 2x2 table we look at the value of Cramer’s V. From the value 0.377 of Cramer’s, we can very confidently (p/alpha is very low , hence almost 99.99% confidence level) say that there’s a very strong positive association between the variables ‘how close people feel to black’ and ‘race’.

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>

1. **What is the relationship between health *(‘health’)* and social class *(‘class’)*?**

**Copy and paste all appropriate R output here. *Please add space as needed.***

 > **# Create a new data table selecting only 'health'’ and ‘class':**

**> hlth\_cls\_2006 <- GSS2006 %>%**

**+ select(health, class) %>%**

**+ filter(!is.na(health)) %>%**

**+ filter(!is.na(class))**

**> unique(hlth\_cls\_2006$class)**

**[1] "working class" "middle class" "lower class" "upper class"**

**> unique(hlth\_cls\_2006$health)**

**[1] "fair" "excellent" "good" "poor"**

**> health\_levels <- c("poor", "fair", "good", "excellent")**

**> hlth\_cls\_2006$health <- factor(hlth\_cls\_2006$health, levels = health\_levels)**

**> class\_levels <- c("lower class", "working class", "middle class", "upper class")**

**> hlth\_cls\_2006$class <- factor(hlth\_cls\_2006$class, levels = class\_levels)**

**> t3\_06 <- table(hlth\_cls\_2006$class, hlth\_cls\_2006$health) %>%**

**+ print()**

poor fair good excellent

lower class 19 39 44 13

working class 41 192 457 223

middle class 48 160 407 285

upper class 1 11 21 26

**> options(scipen=999)**

**> oii.xtab(t3\_06, col = TRUE, row = TRUE,stats = TRUE,**

**+ varnames = c("Class", "Health"))**

Cross-tabulation of Class (rows) and Health (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1987

|

| poor | fair | good | excellent | Row Total |

--------------|-----------|-----------|-----------|-----------|-----------|

lower class | 19 | 39 | 44 | 13 | 115 |

| 16.52% | 33.91% | 38.26% | 11.30% | 5.79% |

| 17.43% | 9.70% | 4.74% | 2.38% | |

--------------|-----------|-----------|-----------|-----------|-----------|

working class | 41 | 192 | 457 | 223 | 913 |

| 4.49% | 21.03% | 50.05% | 24.42% | 45.95% |

| 37.61% | 47.76% | 49.19% | 40.77% | |

--------------|-----------|-----------|-----------|-----------|-----------|

middle class | 48 | 160 | 407 | 285 | 900 |

| 5.33% | 17.78% | 45.22% | 31.67% | 45.29% |

| 44.04% | 39.80% | 43.81% | 52.10% | |

--------------|-----------|-----------|-----------|-----------|-----------|

upper class | 1 | 11 | 21 | 26 | 59 |

| 1.69% | 18.64% | 35.59% | 44.07% | 2.97% |

| 0.92% | 2.74% | 2.26% | 4.75% | |

--------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 109 | 402 | 929 | 547 | 1987 |

| 5.49% | 20.23% | 46.75% | 27.53% | |

--------------|-----------|-----------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 74.02014 d.f. = 9 p = 0.000000000002466496

Minimum expected frequency: 3.236537

Cells with Expected Frequency < 5: 1 of 16 (6.25%)

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.193

Contingency coefficient: 0.190

Cramer's V: 0.111

Ordinal measures of association:

Total number of pairs: 1973091

Concordant pairs: 454906 ( 23.06 %)

Discordant pairs: 309737 ( 15.7 %)

Tied on first variable: 541574 ( 27.45 %)

Tied on second variable: 379304 ( 19.22 %)

Tied on both variables: 287570 ( 14.57 %)

Goodman-Kruskal Gamma: 0.190

Somers' d (col dep.): 0.127

Kendall's tau-b: 0.119

Stuart's tau-c: 0.098

Goodman-Kruskal Lambda:

Row dependent: 0.064

Column dependent: 0.005

Warning messages:

1: In chisq.test(t, correct = FALSE, ...) :

Chi-squared approximation may be incorrect

2: running command '/usr/libexec/java\_home' had status 1

**What observations can you make?**

Since it is **NOT** a 2x2 table we look at the value of Cramer’s V. From the value 0.111 of Cramer’s, we can very confidently (p/alpha is very low , hence almost 99.99% confidence level) say that there’s a very weak positive association between the variables ‘health’ and ‘social class’.

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>

**Exercise 2: Measures of Association for Categorical Variables. Repeat analysis with the GSS16 dataset.**

Use the GSS16 dataset and answer the following questions. It is important that you set up our crosstabulation tables (ROW BY Column; Dependent BY Independent). This means put your dependent variable in the Row(s) box and your independent variable in the Column(s) box. Otherwise it may be difficult for you to know which values to report and how to describe the relationship.

Indicate the strength of the association for the following tables. If a measure does not apply, indicate so by writing N/A for your answer. Remember to consider the size and the shape of the table when deciding which measures to apply.

1. **Whether anti-religionists should be allowed to teach *(‘colath’)* by gender *(‘sex’)***

**Copy and paste all appropriate R output here. *Please add space as needed.***

**> # Create a new data frame selecting only 'colath’ and ‘sex':**

**> col\_sex\_2016 <- GSS2016 %>%**

**+ select(sex, colath) %>%**

**+ filter(!is.na(colath)) %>%**

**+ filter(!is.na(sex))**

**Warning message:**

**package ‘bindrcpp’ was built under R version 3.4.4**

**> t1\_06 <- table(col\_sex\_2016$colath, col\_sex\_2016$sex) %>%**

**+ print()**

**female male**

**allowed 667 563**

**not allowed 360 250**

**> oii.xtab(t1\_06, col = TRUE, row = TRUE,stats = TRUE,**

**+ varnames = c("anti-religioinsts teach", "Sex"))**

Cross-tabulation of anti-religioinsts teach (rows) and Sex (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1840

|

| female | male | Row Total |

-------------|-----------|-----------|-----------|

allowed | 667 | 563 | 1230 |

| 54.23% | 45.77% | 66.85% |

| 64.95% | 69.25% | |

-------------|-----------|-----------|-----------|

not allowed | 360 | 250 | 610 |

| 59.02% | 40.98% | 33.15% |

| 35.05% | 30.75% | |

-------------|-----------|-----------|-----------|

Column Total | 1027 | 813 | 1840 |

| 55.82% | 44.18% | |

-------------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 3.791721 d.f. = 1 p = 0.05150668

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

------------------------------------------------------------

Chi^2 = 3.60003 d.f. = 1 p = 0.05777853

Minimum expected frequency: 269.5272

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.044

Contingency coefficient: 0.044

Cramer's V: 0.044

Ordinal measures of association:

Total number of pairs: 1691880

Concordant pairs: 166750 ( 9.86 %)

Discordant pairs: 202680 ( 11.98 %)

Tied on first variable: 465521 ( 27.52 %)

Tied on second variable: 380870 ( 22.51 %)

Tied on both variables: 476059 ( 28.14 %)

Goodman-Kruskal Gamma: -0.097

Somers' d (col dep.): -0.048

Kendall's tau-b: -0.045

Stuart's tau-c: -0.042

Goodman-Kruskal Lambda:

Row dependent: 0.000

Column dependent: 0.000

**What conclusions can you make from your output?**

Since it is a 2x2 table we look at the value of phi. From the value 0.044 of phi obtained for a alpha of p = 0.05150668 (i.e. at around 99.5% confidence level), we can confidently say that there’s an association between an attitude towards ‘anti-religionists should be allowed to teach’ and the ‘gender’ of the survey participant. Although, from the value it can be stated that this association is extremely weak (not generally useful)

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>

1. **How close do you feel to blacks *(‘closeblk’)* by race *(‘race’)?***

**Copy and paste all appropriate R output here. *Please add space as needed.***

 > # Create a new data table selecting only 'closeblk'’ and ‘race':

**> cblk\_race\_2016 <- GSS2016 %>%**

**+ select(closeblk, race) %>%**

**+ filter(!is.na(closeblk)) %>%**

**+ filter(!is.na(race))**

**> unique(GSS2016$race)**

**[1] "white" "other" "black"**

**> unique(GSS2016$closeblk)**

**[1] "7" NA "neither one or the other"**

**[4] "2" "very close" "6"**

**[7] "8" "not at all close" "3"**

**[10] "4"**

**> cblk\_race\_2016$closeblk <- cblk\_race\_2016$closeblk %>%**

**+ recode("not at all close" = "1 not at all close",**

**+ "neither one or the other" = "5 neither one or the other",**

**+ "very close" = "9 very close")**

**> t2\_06 <- table(cblk\_race\_2016$closeblk, cblk\_race\_2016$race) %>%**

**+ print()**

black other white

1 not at all close 3 14 58

2 2 6 31

3 3 3 55

4 5 7 53

5 neither one or the other 61 75 550

6 11 13 112

7 28 19 212

8 29 15 96

9 very close 174 31 212

**> oii.xtab(t2\_06, col = TRUE, row = TRUE, stats = TRUE,**

**+ varnames = c("Close to black people", "Race"))**

Cross-tabulation of Close to black people (rows) and Race (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1878

|

| black | other | white | Row Total |

---------------------------|-----------|-----------|-----------|-----------|

1 not at all close | 3 | 14 | 58 | 75 |

| 4.00% | 18.67% | 77.33% | 3.99% |

| 0.95% | 7.65% | 4.21% | |

---------------------------|-----------|-----------|-----------|-----------|

2 | 2 | 6 | 31 | 39 |

| 5.13% | 15.38% | 79.49% | 2.08% |

| 0.63% | 3.28% | 2.25% | |

---------------------------|-----------|-----------|-----------|-----------|

3 | 3 | 3 | 55 | 61 |

| 4.92% | 4.92% | 90.16% | 3.25% |

| 0.95% | 1.64% | 3.99% | |

---------------------------|-----------|-----------|-----------|-----------|

4 | 5 | 7 | 53 | 65 |

| 7.69% | 10.77% | 81.54% | 3.46% |

| 1.58% | 3.83% | 3.84% | |

---------------------------|-----------|-----------|-----------|-----------|

5 neither one or the other | 61 | 75 | 550 | 686 |

| 8.89% | 10.93% | 80.17% | 36.53% |

| 19.30% | 40.98% | 39.88% | |

---------------------------|-----------|-----------|-----------|-----------|

6 | 11 | 13 | 112 | 136 |

| 8.09% | 9.56% | 82.35% | 7.24% |

| 3.48% | 7.10% | 8.12% | |

---------------------------|-----------|-----------|-----------|-----------|

7 | 28 | 19 | 212 | 259 |

| 10.81% | 7.34% | 81.85% | 13.79% |

| 8.86% | 10.38% | 15.37% | |

---------------------------|-----------|-----------|-----------|-----------|

8 | 29 | 15 | 96 | 140 |

| 20.71% | 10.71% | 68.57% | 7.45% |

| 9.18% | 8.20% | 6.96% | |

---------------------------|-----------|-----------|-----------|-----------|

9 very close | 174 | 31 | 212 | 417 |

| 41.73% | 7.43% | 50.84% | 22.20% |

| 55.06% | 16.94% | 15.37% | |

---------------------------|-----------|-----------|-----------|-----------|

Column Total | 316 | 183 | 1379 | 1878 |

| 16.83% | 9.74% | 73.43% | |

---------------------------|-----------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 265.9969 d.f. = 16 p = 0.00000000000000000000000000000000000000000000002674867

Minimum expected frequency: 3.800319

Cells with Expected Frequency < 5: 1 of 27 (3.703704%)

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.376

Contingency coefficient: 0.352

Cramer's V: 0.266

Ordinal measures of association:

Total number of pairs: 1762503

Concordant pairs: 186345 ( 10.57 %)

Discordant pairs: 411091 ( 23.32 %)

Tied on first variable: 148513 ( 8.43 %)

Tied on second variable: 783629 ( 44.46 %)

Tied on both variables: 232925 ( 13.22 %)

Goodman-Kruskal Gamma: -0.376

Somers' d (col dep.): -0.163

Kendall's tau-b: -0.221

Stuart's tau-c: -0.191

Goodman-Kruskal Lambda:

Row dependent: 0.095

Column dependent: 0.000

Warning messages:

1: In chisq.test(t, correct = FALSE, ...) :

Chi-squared approximation may be incorrect

2: running command '/usr/libexec/java\_home' had status 1

**What observations can you make?**

Since it is **NOT** a 2x2 table we look at the value of Cramer’s V. From the value 0.266 of Cramer’s, we can very confidently (p/alpha is very low , hence almost 99.99% confidence level) say that there’s a moderately strong positive association between the variables ‘how close people feel to black’ and ‘race’.

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>

1. **What is the relationship between health *(‘health’)* and social class *(‘class’)***

**Copy and paste all appropriate R output here. *Please add space as needed.***

**> # Create a new data table selecting only 'health'’ and ‘class’:**

**> hlth\_cls\_2016 <- GSS2016 %>%**

**+ select(health, class) %>%**

**+ filter(!is.na(health)) %>%**

**+ filter(!is.na(class))**

**> unique(hlth\_cls\_2016$class)**

**[1] "middle class" "working class" "lower class" "upper class"**

**> unique(hlth\_cls\_2016$health)**

**[1] "good" "excellent" "poor" "fair"**

**> health\_levels <- c("poor", "fair", "good", "excellent")**

**> hlth\_cls\_2016$health <- factor(hlth\_cls\_2016$health, levels = health\_levels)**

**> class\_levels <- c("lower class", "working class", "middle class", "upper class")**

**> hlth\_cls\_2016$class <- factor(hlth\_cls\_2016$class, levels = class\_levels)**

**> t3\_06 <- table(hlth\_cls\_2016$class, hlth\_cls\_2016$health) %>%**

**+ print()**

poor fair good excellent

lower class 36 58 65 17

working class 53 226 450 160

middle class 28 135 371 225

upper class 1 5 27 13

**> options(scipen=999)**

**> oii.xtab(t3\_06, col = TRUE, row = TRUE,stats = TRUE,**

**+ varnames = c("Class", "Health"))**

Cross-tabulation of Class (rows) and Health (cols)

Cell Contents

|-------------------------|

| Count |

| Row Percent |

| Column Percent |

|-------------------------|

Total Observations in Table: 1870

|

| poor | fair | good | excellent | Row Total |

--------------|-----------|-----------|-----------|-----------|-----------|

lower class | 36 | 58 | 65 | 17 | 176 |

| 20.45% | 32.95% | 36.93% | 9.66% | 9.41% |

| 30.51% | 13.68% | 7.12% | 4.10% | |

--------------|-----------|-----------|-----------|-----------|-----------|

working class | 53 | 226 | 450 | 160 | 889 |

| 5.96% | 25.42% | 50.62% | 18.00% | 47.54% |

| 44.92% | 53.30% | 49.29% | 38.55% | |

--------------|-----------|-----------|-----------|-----------|-----------|

middle class | 28 | 135 | 371 | 225 | 759 |

| 3.69% | 17.79% | 48.88% | 29.64% | 40.59% |

| 23.73% | 31.84% | 40.64% | 54.22% | |

--------------|-----------|-----------|-----------|-----------|-----------|

upper class | 1 | 5 | 27 | 13 | 46 |

| 2.17% | 10.87% | 58.70% | 28.26% | 2.46% |

| 0.85% | 1.18% | 2.96% | 3.13% | |

--------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 118 | 424 | 913 | 415 | 1870 |

| 6.31% | 22.67% | 48.82% | 22.19% | |

--------------|-----------|-----------|-----------|-----------|-----------|

Statistics for All Table Factors

Pearson's Chi-squared test

------------------------------------------------------------

Chi^2 = 133.3343 d.f. = 9 p = 0.00000000000000000000000244297

Minimum expected frequency: 2.902674

Cells with Expected Frequency < 5: 1 of 16 (6.25%)

Unable to find any JVMs matching version "(null)".

No Java runtime present, try --request to install.

Likelihood ratio chi-square: NaN df: NaN p-value: NaN

Chi-square-based measures of association:

Phi: 0.267

Contingency coefficient: 0.258

Cramer's V: 0.154

Ordinal measures of association:

Total number of pairs: 1747515

Concordant pairs: 459617 ( 26.3 %)

Discordant pairs: 239018 ( 13.68 %)

Tied on first variable: 450068 ( 25.75 %)

Tied on second variable: 350068 ( 20.03 %)

Tied on both variables: 248744 ( 14.23 %)

Goodman-Kruskal Gamma: 0.316

Somers' d (col dep.): 0.210

Kendall's tau-b: 0.201

Stuart's tau-c: 0.168

Goodman-Kruskal Lambda:

Row dependent: 0.066

Column dependent: 0.000

Warning messages:

1: In chisq.test(t, correct = FALSE, ...) :

Chi-squared approximation may be incorrect

2: running command '/usr/libexec/java\_home' had status 1

**What observations can you make?**

Since it is **NOT** a 2x2 table we look at the value of Cramer’s V. From the value 0.154 of Cramer’s, we can very confidently (p/alpha is very low , hence almost 99.99% confidence level) say that there’s a weak positive association between the variables ‘health’ and ‘social class’.

|  |  |
| --- | --- |
| Level of association | Verbal description |
| 0.00 | No relationship |
| 0.00 to 0.15 | Not generally useful |
| 0.10 to 0.20 | Weak |
| 0.20 to 0.25 | Moderate |
| 0.25 to 0.30 | Moderately strong |
| 0.30 to 0.35 | Strong |
| 0.35 to 0.40 | Very strong |
| 0.40 to 0.45 | Worrisomely strong |
| 0.45 to 0.99 | Redundant (two variables are probably measuring the same concept) |
| 1.00 | Perfect |

Table *\*\*source:* <http://homes.chass.utoronto.ca/~josephf/pol242/LM-3A#STANDARDS>