**> ##### R Lab 6: Measures of Association for Categorical Variables -- 2016 GSS Data ####**

**> install.packages("tidyverse")**

Error in install.packages : Updating loaded packages

Restarting R session...

**> install.packages("tidyverse")**

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/tidyverse\_1.2.1.tgz'

Content type 'application/x-gzip' length 77756 bytes (75 KB)

==================================================

downloaded 75 KB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//RtmpED9Gni/downloaded\_packages

**> install.packages("oii")**

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/oii\_1.0.2.1.tgz'

Content type 'application/x-gzip' length 33412 bytes (32 KB)

==================================================

downloaded 32 KB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//RtmpED9Gni/downloaded\_packages

**> install.packages("DescTools")**

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/DescTools\_0.99.25.tgz'

Content type 'application/x-gzip' length 3966766 bytes (3.8 MB)

==================================================

downloaded 3.8 MB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//RtmpED9Gni/downloaded\_packages

**> library(DescTools)**

Warning message:

package ‘DescTools’ was built under R version 3.4.4

**> library(oii)**

Warning message:

package ‘oii’ was built under R version 3.4.2

**> library(tidyverse)**

── Attaching packages ─────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

✔ ggplot2 3.0.0 ✔ purrr 0.2.5

✔ tibble 1.4.2 ✔ dplyr 0.7.6

✔ tidyr 0.8.1 ✔ stringr 1.3.1

✔ readr 1.1.1 ✔ forcats 0.3.0

── Conflicts ────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──

✖ dplyr::filter() masks stats::filter()

✖ dplyr::lag() masks stats::lag()

Warning messages:

1: package ‘tidyverse’ was built under R version 3.4.2

2: package ‘ggplot2’ was built under R version 3.4.4

3: package ‘tibble’ was built under R version 3.4.3

4: package ‘tidyr’ was built under R version 3.4.4

5: package ‘purrr’ was built under R version 3.4.4

6: package ‘dplyr’ was built under R version 3.4.4

7: package ‘stringr’ was built under R version 3.4.4

8: package ‘forcats’ was built under R version 3.4.3

> # Importing GSS 2016 Data

> GSS2016 <- read\_csv("GSS\_2016.csv")

Parsed with column specification:

cols(

.default = col\_character(),

agekdbrn = col\_integer(),

babies = col\_integer(),

bigbang = col\_logical(),

bigbang1 = col\_logical(),

bigbang2 = col\_logical(),

boyorgrl = col\_logical(),

cohort = col\_integer(),

cohrs2 = col\_integer(),

condrift = col\_logical(),

coninc = col\_double(),

conrinc = col\_double(),

copres10 = col\_integer(),

copres105plus = col\_integer(),

cosei10 = col\_double(),

cosei10educ = col\_double(),

cosei10inc = col\_double(),

dateintv = col\_integer(),

earnrs = col\_integer(),

educ = col\_integer(),

electron = col\_logical()

# ... with 81 more columns

)

See spec(...) for full column specifications.

Warning: 2 parsing failures.

row # A tibble: 2 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 1210 old1 no trailing characters " or older" 'GSS\_2016.csv' file 2 2760 earnrs an integer eight or more 'GSS\_2016.csv'

**> # 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

Warning message:

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

**> # 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

**> # 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