> ##### R Lab 6: Measures of Association for Categorical Variables -- 2006 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//RtmphU6Khp/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//RtmphU6Khp/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//RtmphU6Khp/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 2006 Data**

**> GSS2006 <- read\_csv("GSS\_2006.csv")**

Parsed with column specification:

cols(

.default = col\_character(),

prestg10 = col\_integer(),

prestg105plus = col\_integer(),

sppres10 = col\_integer(),

sppres105plus = col\_integer(),

papres10 = col\_integer(),

papres105plus = col\_integer(),

mapres10 = col\_integer(),

mapres105plus = col\_integer(),

sei10 = col\_double(),

spsei10 = col\_double(),

pasei10 = col\_double(),

masei10 = col\_double(),

sei10educ = col\_double(),

spsei10educ = col\_double(),

pasei10educ = col\_double(),

masei10educ = col\_double(),

sei10inc = col\_double(),

spsei10inc = col\_double(),

pasei10inc = col\_double(),

masei10inc = col\_double()

# ... with 110 more columns

)

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

Warning: 7 parsing failures.

row # A tibble: 5 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 1722 physhlth an integer DONT KNOW 'GSS\_2006.csv' file 2 2958 prozfor1 an integer DONT KNOW 'GSS\_2006.csv' row 3 3164 adults no trailing characters " or more" 'GSS\_2006.csv' col 4 3170 sphrs2 no trailing characters + hrs 'GSS\_2006.csv' expected 5 3246 physhlth an integer DONT KNOW 'GSS\_2006.csv'

... ................................. ... ................................................................. ........ ....................................... [... truncated]

Warning message:

In rbind(names(probs), probs\_f) :

number of columns of result is not a multiple of vector length (arg 1)

**> # 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

Warning message:

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

**> # 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

Warning message:

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

> **# 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