Data Science HW5 Total: 20 points Due: February 8th, 2019, NOON 1.

1. In your own words, specify the research question(s) (2 sentences or less): one sentence for overall RQ, one for Study S1.

* Overall: Did children see a difference in self-advantaged vs self-disadvantaged inequities. Will older children be more likely to create inequity that inconveniences themselves than younger children, when given the option to keep or throwaway the extra item vs give or throwaway?
* S1: Will older children be more likely to create inequity that inconveniences themselves than younger children when they are presented with the options to keep, give or throwaway the extra item?

1. Include and explain IVs, DVs for Study S1 [4]

* IV: age group (older vs younger), sex (not analyzed)

\*Condition is also an IV in the first experiment

* DV: choice (throwaway, give, keep)

1. R section (please complete the following and include your script and outputs as in-line text below
   1. CrossTable [4] (your table needs to show the following to get full credit; please take a snapshot of your crosstable and include as an image)
      1. ‘Choice’ & ‘Agegroup’ as the labels for the column and row variables respectively
      2. For ‘Choice’: level names are ‘Give to Other’, ‘Throw Away’, and ‘Give to Self’
      3. For ‘Agegroup’: level names are ‘younger’, and ‘older’ iv. Each cell should only show these values: N, Expected N and Row proportion (make sure values match those reported in the supplementary paper)
   2. Chi-square goodness of fit (younger children) [3.5]
      1. Bonus question: write a code that calculates the phi coefficient without using any R package [1]
   3. Chi-square goodness of fit (older children) [3.5]
      1. Bonus question: write a code that calculates the phi coefficient without using any R package [1]
   4. In your own words, why was the goodness of fit an appropriate test to run? [1]

Goodness of fit is appropriate because we are trying to determine how far the results deviate from chance.

hw\_05

Abby Bergman

2/6/2019

library(vcd)

## Loading required package: grid

library(gmodels)  
library(tidyverse)

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

## ✔ ggplot2 3.1.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.7  
## ✔ tidyr 0.8.2 ✔ 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()

shaw\_s1 <- read\_csv("Shaw\_S1.csv") %>%  
 select(-X5, -X6, -X7, -X8,-X9)

## Warning: Missing column names filled in: 'X5' [5], 'X6' [6], 'X7' [7],  
## 'X8' [8], 'X9' [9]

## Parsed with column specification:  
## cols(  
## `DV (Y= You, T= Trash, M= Mark)` = col\_character(),  
## `Age (1= younger, 2 = older)` = col\_integer(),  
## Age = col\_double(),  
## `Sex (1=Male, 2= Female)` = col\_integer(),  
## X5 = col\_character(),  
## X6 = col\_character(),  
## X7 = col\_character(),  
## X8 = col\_character(),  
## X9 = col\_character()  
## )

names(shaw\_s1)[c(1,2,4)] = c('choice', 'agegroup', 'sex')  
  
shaw\_s1$choice = factor(shaw\_s1$choice, levels = c("T", "Y", "M"), labels = c("Throw Away", "Give to Self", "Give to Other"))  
shaw\_s1$choice = factor(shaw\_s1$choice)  
shaw\_s1$agegroup = factor(shaw\_s1$agegroup, levels = c(1, 2), labels = c("Younger", "Older"))  
shaw\_s1$sex = factor(shaw\_s1$sex)  
  
#CrossTable  
CrossTable(shaw\_s1$agegroup, shaw\_s1$choice,  
 expected=TRUE, prop.r = TRUE, prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | Expected N |  
## | N / Row Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 127   
##   
##   
## | shaw\_s1$choice   
## shaw\_s1$agegroup | Throw Away | Give to Self | Give to Other | Row Total |   
## -----------------|---------------|---------------|---------------|---------------|  
## Younger | 16 | 46 | 20 | 82 |   
## | 23.244 | 33.575 | 25.181 | |   
## | 0.195 | 0.561 | 0.244 | 0.646 |   
## -----------------|---------------|---------------|---------------|---------------|  
## Older | 20 | 6 | 19 | 45 |   
## | 12.756 | 18.425 | 13.819 | |   
## | 0.444 | 0.133 | 0.422 | 0.354 |   
## -----------------|---------------|---------------|---------------|---------------|  
## Column Total | 36 | 52 | 39 | 127 |   
## -----------------|---------------|---------------|---------------|---------------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 22.35745 d.f. = 2 p = 1.396825e-05   
##   
##   
##

# phi = (chi^2/n)^(1/2)  
  
phi <- function(x,n){  
 ((x/n)^(1/2))  
}  
  
#goodness of fit younger  
shaw\_young <- shaw\_s1 %>%  
 filter(agegroup == "Younger")  
shawtab = xtabs(~choice, data = shaw\_young)  
chi <- chisq.test(shawtab)  
#phi  
  
phi(chi[["statistic"]][["X-squared"]],length(shaw\_young$choice))

## [1] 0.4865838

chi <- chisq.test(shawtab)  
  
#goodness of fit older  
shaw\_old <- shaw\_s1 %>%  
 filter(agegroup == "Older")  
shawtab1 = xtabs(~choice, data = shaw\_old)  
chi1<- chisq.test(shawtab1)  
#phi  
phi(chi1[["statistic"]][["X-squared"]],length(shaw\_old$choice))

## [1] 0.4251361

shawtab2 = xtabs(~agegroup + choice, data = shaw\_s1)  
chisq.test(shawtab2)

##   
## Pearson's Chi-squared test  
##   
## data: shawtab2  
## X-squared = 22.357, df = 2, p-value = 1.397e-05

