Lab 4

Haroon Popal

April 02, 2021

# Introduction

## Research Question

Interpersonal relationships have significant implications for perception, emotion, memory, motivation, and decision-making. Throughout the lifespan, having strong social relationships are important for personal success, health and well-being. One theory proposes that humans have a unique ability to track a large number of social relationships, which have allowed us to live in large groups and societies. As humans enjoy very diverse social lives across their family life, careers, friends, communities, and social media, the research question this project attempts to address is what is the impact of a sustained decrease in the typical number and variety of social relationships on mental health.

# Data Source

Participants were recruited from Amazon Mechnical Turk (mTurk). mTurk is hosted by Amazon and provides an online platform for individuals (mTurkers) to partake in surveys. Use of mTurk has become popular in psychology research, as it allows for larger sample sizes than what can be typically collected in a lab setting. A link to the survey was posted on mTurk using turkprime.com. The use of “bots” to pose as real participants has been an issue with mTurk, and turkprime has a system that catches and excludes some of these bots. Participants from across the United States were eligible to take the survey at three time points during the beginning of the COVID-19 pandemic, from March to May 2020. An additional eligibility criteria was that participants must have learned English as their first language. This is due to cultural differences in social relationships, which we believe exist but are unable to directly address. Data was collected at three time points. Only individuals who completed the first time point were invited to participate at the second and third time points. All of the survey data for the first time point was collected within one day. At the second and third time points, the survey was opened and participants were invite to complete the survey again. The survey remained open for one week to allow individuals to complete the survey. The second survey was opened three weeks after the first survey was completed, and the third survey was opened four weeks after the second survey was completed. In this way, the time between surveys differed for each participant. Individuals who missed the second time were still allowed to complete the third time point. In total, 767 participants completed the survey at the first time point, 501 at the second time point, and 365 at the third time point. The level 1 unit of analysis is observations and the level 2 unit of analysis is individuals.

# Import data  
responses\_excluded <- read.csv('/Users/Administrator/Google\_Drive/olson\_lab/projects/social\_distancing/survey\_data/responses\_excluded.csv')

socdist\_data <- responses\_excluded %>% select(AmazonIdentifier, AN\_URels, IN\_valw, AN\_reldiv\_sum, dist\_total, anxiety, depression, behav\_emo\_control, pos\_affect\_calm, pos\_affect\_happy, bored, do\_interests, regulation, routine, health, wave)  
  
names(socdist\_data)[names(socdist\_data)=='AmazonIdentifier'] <- 'ID'

# Dropped Cases

# Drop cases that have any missing data  
socdist\_data <- socdist\_data[rowSums(is.na(socdist\_data)) == 0,]  
socdist\_data %>% group\_by(wave) %>% summarise(count = n\_distinct(ID))

## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 3 x 2  
## wave count  
## <dbl> <int>  
## 1 1 467  
## 2 2 289  
## 3 3 230

length(unique(socdist\_data$ID))

## [1] 533

Cases were dropped for two possible non-compliance reasons. The first is if participants failed attention check questions which were designed to easily reveal whether participants are attending to the survey. These questions were simple and had clear answers such as “which of the following is most likely the color red?” (e.g. strawberries). Participants who failed even one attention check question (of two) were excluded. The second way a participant might have been excluded is if they did not follow directions in how to format their responses for free response questions or did not answer all the questions. This was done so that the response coding could more easily be done by the researchers. In total 647 observations were dropped and 234 individuals were dropped because of subject noncompliance.

|  |  |  |  |
| --- | --- | --- | --- |
| Wave | Original responses | Excluding attention check failures | Excluding directions failures |
| 1 | 767 | 753 | 467 |
| 2 | 501 | 463 | 289 |
| 3 | 365 | 360 | 230 |
| Unique individuals | 767 | 754 | 533 |

# Variables

This study includes variables that can be grouped as predictors, outcomes, and nuisance variables. The outcome we measured was a latent factor of mental well-being, derived from measures of anxiety, behavioral/emotional control, depression, calm-positive affect, and happy/positive affect. These five variables were measured by asking participants whether they felt each factor much less, less, about the same, more, or much more in the past week. The predictors of interest were the number of social relationships, the variety of social relationship, and the quality of social relationships they experienced in the past week. Participants were asked to report each person they interacted with in the past week, and their relationship to that person, results in their number of relationships (i.e. 10 relationships), and variety of relationship (i.e. 5 unique relationships) reported. They were also asked to rate how positive or negative each relationship was on a scale of one to five indicating very negative, negative, neutral, positive, or very positive. Nuisance variables were boredom, ability to do their interests (i.e. hobbies), ability to regulate one’s emotions, their ability to do normal routine, and physical well-being, measured by a report of how they feel similar to the mental well-being questions). The observation level variables were the social relationship feature variables and the nuisance variables, which were all person-centered as we wished to know the pure within-person effects. The individual level variables were the participant means for the predictors, the outcome, and the nuisance variables. All variables were standardized.

## Latent variable

A latent variable of well-being was created from the five mental health questions. This latent variable will serve as a predictor of the mental health outcomes, and an outcome of the social relationship variables.

wellb\_mod\_mg <- 'Wellb =~ NA\*anxiety + pos\_affect\_calm + depression + pos\_affect\_happy + behav\_emo\_control  
 Wellb ~~ 1\*Wellb'  
  
wellb\_configural\_res <- sem(wellb\_mod\_mg, socdist\_data, estimator='wlsmv')  
socdist\_data$wellb\_lv <- as.numeric(predict(wellb\_configural\_res))  
  
# Reverse score mental well-being variable so it is easier to interpret (higher score means better mental well-being)  
socdist\_data$wellb\_lv <- socdist\_data$wellb\_lv \* -1  
  
  
# Set wave as a factor variable  
socdist\_data$wave <- as.numeric(socdist\_data$wave)

## Subsetting due to missingness

socdist\_data\_ss <- subset(socdist\_data, !is.na(AN\_URels) & !is.na(IN\_valw) & !is.na(AN\_reldiv\_sum) & !is.na(wellb\_lv))  
nrow(socdist\_data\_ss)

## [1] 986

length(unique(socdist\_data\_ss$ID))

## [1] 533

## Aggregation and centering

The observation level variables will be aggregated as averages of the observations for each participant. Then person-centered variables will be created as this is a longitudinal analysis and we wish to focus on the pure within-subject effects.

socdist\_data\_ss<-group\_by(socdist\_data\_ss, ID)  
socdist\_data\_ss<-mutate(socdist\_data\_ss,   
 AN\_URels.mean = mean(AN\_URels,na.rm=TRUE),  
 IN\_valw.mean = mean(IN\_valw,na.rm=TRUE),  
 AN\_reldiv\_sum.mean = mean(AN\_reldiv\_sum,na.rm=TRUE),  
 wellb\_lv.mean = mean(wellb\_lv, na.rm=TRUE),   
 dist\_total.mean = mean(dist\_total,na.rm=TRUE),  
 bored.mean = mean(bored,na.rm=TRUE),  
 do\_interests.mean = mean(do\_interests,na.rm=TRUE),  
 regulation.mean = mean(regulation,na.rm=TRUE),  
 routine.mean = mean(routine,na.rm=TRUE),  
 health.mean = mean(health,na.rm=TRUE),  
 coobs=n())  
  
socdist\_data\_ss <- ungroup(socdist\_data\_ss)  
socdist\_data\_ss$AN\_URels.pc <- socdist\_data\_ss$AN\_URels - socdist\_data\_ss$AN\_URels.mean  
socdist\_data\_ss$IN\_valw.pc <- socdist\_data\_ss$IN\_valw - socdist\_data\_ss$IN\_valw.mean  
socdist\_data\_ss$AN\_reldiv\_sum.pc <- socdist\_data\_ss$AN\_reldiv\_sum - socdist\_data\_ss$AN\_reldiv\_sum.mean  
socdist\_data\_ss$wellb\_lv.pc <- socdist\_data\_ss$wellb\_lv - socdist\_data\_ss$wellb\_lv.mean  
socdist\_data\_ss$dist\_total.pc <- socdist\_data\_ss$dist\_total - socdist\_data\_ss$dist\_total.mean  
socdist\_data\_ss$bored.pc <- socdist\_data\_ss$bored - socdist\_data\_ss$bored.mean  
socdist\_data\_ss$do\_interests.pc <- socdist\_data\_ss$do\_interests - socdist\_data\_ss$do\_interests.mean  
socdist\_data\_ss$regulation.pc <- socdist\_data\_ss$regulation - socdist\_data\_ss$regulation.mean  
socdist\_data\_ss$routine.pc <- socdist\_data\_ss$routine - socdist\_data\_ss$routine.mean  
socdist\_data\_ss$health.pc <- socdist\_data\_ss$health - socdist\_data\_ss$health.mean  
  
# Standardize  
socdist\_data\_ss$AN\_URels.pc\_z <- as.numeric(scale(socdist\_data\_ss$AN\_URels.pc))  
socdist\_data\_ss$IN\_valw.pc\_z <- as.numeric(scale(socdist\_data\_ss$IN\_valw.pc))  
socdist\_data\_ss$AN\_reldiv\_sum.pc\_z <- as.numeric(scale(socdist\_data\_ss$AN\_reldiv\_sum.pc))  
socdist\_data\_ss$wellb\_lv\_z <- as.numeric(scale(socdist\_data\_ss$wellb\_lv))  
socdist\_data\_ss$dist\_total.pc\_z <- as.numeric(scale(socdist\_data\_ss$dist\_total.pc))  
socdist\_data\_ss$bored.pc\_z <- as.numeric(scale(socdist\_data\_ss$bored.pc))  
socdist\_data\_ss$do\_interests.pc\_z <- as.numeric(scale(socdist\_data\_ss$do\_interests.pc))  
socdist\_data\_ss$regulation.pc\_z <- as.numeric(scale(socdist\_data\_ss$regulation.pc))  
socdist\_data\_ss$routine.pc\_z <- as.numeric(scale(socdist\_data\_ss$routine.pc))  
socdist\_data\_ss$health.pc\_z <- as.numeric(scale(socdist\_data\_ss$health.pc))  
  
  
socdist\_data\_ss[, 16:ncol(socdist\_data\_ss)] <- lapply(16:ncol(socdist\_data\_ss), function(x) as.numeric(socdist\_data\_ss[[x]]))  
  
# Create person-level dataframe  
socdist\_data\_pl <- group\_by(socdist\_data\_ss, ID)  
socdist\_data\_pl <- filter(socdist\_data\_pl[c('ID','AN\_URels.mean', 'IN\_valw.mean', 'AN\_reldiv\_sum.mean', 'wellb\_lv.mean', 'dist\_total.mean', 'bored.mean', 'do\_interests.mean', 'regulation.mean', 'routine.mean', 'health.mean')], row\_number(ID)==1)  
  
socdist\_data\_pl$AN\_URels.mean\_z <- as.numeric(scale(socdist\_data\_pl$AN\_URels.mean))  
socdist\_data\_pl$IN\_valw.mean\_z <- as.numeric(scale(socdist\_data\_pl$IN\_valw.mean))  
socdist\_data\_pl$AN\_reldiv\_sum.mean\_z <- as.numeric(scale(socdist\_data\_pl$AN\_reldiv\_sum.mean))  
socdist\_data\_pl$dist\_total.mean\_z <- as.numeric(scale(socdist\_data\_pl$bored.mean))  
socdist\_data\_pl$bored.mean\_z <- as.numeric(scale(socdist\_data\_pl$dist\_total.mean))  
socdist\_data\_pl$do\_interests.mean\_z <- as.numeric(scale(socdist\_data\_pl$do\_interests.mean))  
socdist\_data\_pl$regulation.mean\_z <- as.numeric(scale(socdist\_data\_pl$regulation.mean))  
socdist\_data\_pl$routine.mean\_z <- as.numeric(scale(socdist\_data\_pl$routine.mean))  
socdist\_data\_pl$health.mean\_z <- as.numeric(scale(socdist\_data\_pl$health.mean))  
  
socdist\_data\_pl\_merge <- socdist\_data\_pl[c('ID', 'AN\_URels.mean\_z', 'IN\_valw.mean\_z', 'AN\_reldiv\_sum.mean\_z', 'dist\_total.mean\_z', 'bored.mean\_z', 'do\_interests.mean\_z', 'regulation.mean\_z', 'routine.mean\_z', 'health.mean\_z')]  
  
socdist\_data\_ss <- merge(socdist\_data\_ss, socdist\_data\_pl\_merge, by="ID")

# Summary Statistics

describe(socdist\_data\_ss)

describe(socdist\_data\_pl)

# Analysis and Estimation Technique

Longitudinal analysis with multilevel regression and full maximum likelihood Time-varying predictor of weeks in pandemic

## Empty model

model1 <- lmer(wellb\_lv\_z ~ 1 + (1|ID), data=socdist\_data\_ss)  
summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: wellb\_lv\_z ~ 1 + (1 | ID)  
## Data: socdist\_data\_ss  
##   
## REML criterion at convergence: 2651.4  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.4417 -0.5260 0.0525 0.4224 3.7163   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.5340 0.7308   
## Residual 0.4804 0.6931   
## Number of obs: 986, groups: ID, 533  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)  
## (Intercept) -0.01156 0.03946 502.30017 -0.293 0.77

## Within- and Between-person model

model2 <- lmer(wellb\_lv\_z ~ AN\_URels.pc\_z + IN\_valw.pc\_z + AN\_reldiv\_sum.pc\_z +   
 wave + wave\*AN\_reldiv\_sum.pc\_z +  
 dist\_total.pc\_z +   
 bored.pc\_z + do\_interests.pc\_z +   
 routine.pc\_z + health.pc\_z +   
 AN\_URels.mean\_z + IN\_valw.mean\_z +   
 AN\_reldiv\_sum.mean\_z + dist\_total.mean\_z +   
 bored.mean\_z + do\_interests.mean\_z +   
 routine.mean\_z + health.mean\_z +   
 (1 | ID),   
 data=socdist\_data\_ss, REML=FALSE)  
  
summary(model2)

## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's  
## method [lmerModLmerTest]  
## Formula: wellb\_lv\_z ~ AN\_URels.pc\_z + IN\_valw.pc\_z + AN\_reldiv\_sum.pc\_z +   
## wave + wave \* AN\_reldiv\_sum.pc\_z + dist\_total.pc\_z + bored.pc\_z +   
## do\_interests.pc\_z + routine.pc\_z + health.pc\_z + AN\_URels.mean\_z +   
## IN\_valw.mean\_z + AN\_reldiv\_sum.mean\_z + dist\_total.mean\_z +   
## bored.mean\_z + do\_interests.mean\_z + routine.mean\_z + health.mean\_z +   
## (1 | ID)  
## Data: socdist\_data\_ss  
##   
## AIC BIC logLik deviance df.resid   
## 2460.5 2563.3 -1209.3 2418.5 965   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.6937 -0.4709 0.0157 0.4421 3.2231   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## ID (Intercept) 0.4064 0.6375   
## Residual 0.3890 0.6237   
## Number of obs: 986, groups: ID, 533  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) -0.323148 0.059341 927.118105 -5.446 6.61e-08 \*\*\*  
## AN\_URels.pc\_z -0.034501 0.025922 450.229108 -1.331 0.18388   
## IN\_valw.pc\_z 0.040215 0.020047 443.388411 2.006 0.04547 \*   
## AN\_reldiv\_sum.pc\_z 0.061879 0.069897 720.914180 0.885 0.37630   
## wave 0.183647 0.028578 600.835238 6.426 2.67e-10 \*\*\*  
## dist\_total.pc\_z -0.019758 0.020160 443.942696 -0.980 0.32759   
## bored.pc\_z -0.059489 0.020784 444.545892 -2.862 0.00441 \*\*   
## do\_interests.pc\_z 0.018212 0.020803 443.053145 0.875 0.38180   
## routine.pc\_z 0.088032 0.021118 446.912230 4.169 3.68e-05 \*\*\*  
## health.pc\_z 0.034399 0.019995 442.754593 1.720 0.08607 .   
## AN\_URels.mean\_z -0.049902 0.043967 497.877006 -1.135 0.25693   
## IN\_valw.mean\_z 0.105761 0.037060 539.956140 2.854 0.00449 \*\*   
## AN\_reldiv\_sum.mean\_z 0.007874 0.044582 520.356228 0.177 0.85988   
## dist\_total.mean\_z -0.187876 0.038654 533.035877 -4.860 1.54e-06 \*\*\*  
## bored.mean\_z 0.018412 0.037299 613.672338 0.494 0.62174   
## do\_interests.mean\_z 0.159478 0.039435 555.359857 4.044 6.00e-05 \*\*\*  
## routine.mean\_z 0.108732 0.040603 558.404600 2.678 0.00763 \*\*   
## health.mean\_z 0.150065 0.036416 516.944073 4.121 4.40e-05 \*\*\*  
## AN\_reldiv\_sum.pc\_z:wave -0.016870 0.031966 768.849893 -0.528 0.59782   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## Correlation matrix not shown by default, as p = 19 > 12.  
## Use print(x, correlation=TRUE) or  
## vcov(x) if you need it

# Results Table

class(model1) <- "lmerMod"  
class(model2) <- "lmerMod"  
  
stargazer(model1, model2, type="text",   
 out = "/Users/Administrator/Google\_Drive/courses/Hierarchical\_Linear\_Modeling/labs/lab\_04/regression\_table.html",star.cutoffs=c(.05, .005, .001),   
 covariate.labels = c("# of unique relationships PC",   
 "IN time-weighted valence PC",   
 "Relation diversity PC",  
 "Time point",  
 "Total distance traveled PC",  
 "Boredom PC",   
 "Ability to do interests PC",   
 "Ability to do normal routine PC",   
 "General physical health PC",  
 "# of unique relationships M",   
 "IN time-weighted valence M",  
 "Relation diversity M",  
 "Total distance traveled M",  
 "Boredom M",   
 "Ability to do interests M",   
 "Ability to do normal routine M",   
 "General physical health M",  
 "Relation diversity PC \* Time point"),  
 dep.var.labels = c("Mental well-being"),   
 title = "Table 3. Regression table for the empty, within and between, and the cross-level interaction models. ",   
 notes="All variables are z-standardized. Standard errors in are parentheses. IN = Immediate network, PC = person-centered, M = person-level mean.")

# Write-up

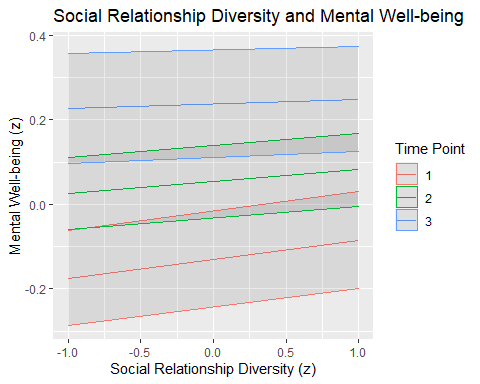
# Graph

# Graph

model2\_eff<-effect("AN\_reldiv\_sum.pc\_z:wave",model2,  
 xlevels=list(AN\_reldiv\_sum.pc\_z=c(-1,1), wave=c(1,2,3)))  
model2\_eff\_df <- data.frame(model2\_eff)  
model2\_eff\_df$AN\_reldiv\_sum.pc\_zf <- factor(model2\_eff\_df$AN\_reldiv\_sum.pc\_z,   
 labels=c("-1 SD", "+1 SD"))  
model2\_eff\_df$wavef <- factor(model2\_eff\_df$wave,   
 labels=c("1", "2", "3"))  
model2\_eff\_df

## AN\_reldiv\_sum.pc\_z wave fit se lower upper  
## 1 -1 1 -0.17469753 0.05720677 -0.286961259 -0.06243381  
## 2 1 1 -0.08468067 0.05840249 -0.199290909 0.02992957  
## 3 -1 2 0.02582000 0.04388686 -0.060304461 0.11194447  
## 4 1 2 0.08209627 0.04407788 -0.004403042 0.16859559  
## 5 -1 3 0.22633754 0.06611294 0.096596169 0.35607891  
## 6 1 3 0.24887322 0.06354231 0.124176493 0.37356994  
## AN\_reldiv\_sum.pc\_zf wavef  
## 1 -1 SD 1  
## 2 +1 SD 1  
## 3 -1 SD 2  
## 4 +1 SD 2  
## 5 -1 SD 3  
## 6 +1 SD 3

ggplot(model2\_eff\_df, aes(x=AN\_reldiv\_sum.pc\_z, y=fit, color=wavef)) +   
 geom\_line() +   
 geom\_ribbon(aes(ymin=lower, ymax=upper), alpha=.1) +   
 ggtitle("Social Relationship Diversity and Mental Well-being") +   
 xlab("Social Relationship Diversity (z)") +   
 ylab("Mental Well-being (z)") +   
 labs(color="Time Point")



# Discussion

# Appendix