Data Exploration: Intergroup Contact

Yanxi Fang

October 28, 2021

This week, we discussed the importance of intergroup contact in settings like neighborhoods, public transportation, and sports leagues. All of these studies ensured or inferred that intergroup contact actually took place. However, some researchers have studied the possibility that even *imagined* contact with a member of an outgroup can change people's perception of that group. In this Data Exploration assignment we will explore two datasets derived from imagined intergroup contact experiments. In part one, you will look at data from a recent study conducted by Dong Wang, Iain Johnston (a professor in the Harvard Government Department) and Baoyu Wang. Wang et al. (2021) conducted an experiment on a group of Chinese students to determine if imagined social contact could reduce antipathy toward Japanese people. In part two, you will look at the results of our in-class survey, which tested whether imagined social contact with a member of one's less-preferred political party would change attitudes toward members of that party. You can do either part first, but you will probably find the exercise most valuable if you do some of each part.

If you have a question about any part of this assignment, please ask! Note that the actionable part of each question is **bolded**.

Part One: Chinese Students and Perception of Japanese People

Data Details:

- File Name: `ChinaJapanData.csv`
- Source: These data are from Wang et al. (2021). Please take some time to skim this paper in order to get a feel for the population they studied, their key hypotheses, and their experimental procedure. Subjects were asked to imagine a bus ride, either one in which they talked to a Japanese person (treatment) or just enjoyed the scenery (control). They were then asked a series of questions to assess their affective feelings toward Japanese and Chinese people, their perceptions of the characteristics of Japanese and Chinese identity, and demographic, policy, and pschological questions to serve as control variables.

Variable Name	Variable Description
`subject`	Anonymized identifier for each experimental subject
`treated`	Binary variable equal to `TRUE` if the subject was told to imagine a bus ride with a Japanese person (the treatment) and `FALSE` if the subject was told to imagine the scenery on a bus ride (control)
`JapanPos`	Affective feeling about Japanese people ranging from 1 (negative) to 7 (positive)
`JapanWarm`	Affective feeling about Japanese people ranging from 1 (cool) to 7 (warm)

Variable Name	Variable Description
JapanAdmire	Affective feeling about Japanese people ranging from 1 (loathing) to 7 (admiration)
JapanRespect	Affective feeling about Japanese people ranging from 1 (contempt) to 7 (respect)
ChinaPos	Affective feeling about Chinese people ranging from 1 (negative) to 7 (positive)
ChinaWarm	Affective feeling about Chinese people ranging from 1 (cool) to 7 (warm)
ChinaAdmire	Affective feeling about Chinese people ranging from 1 (loathing) to 7 (admiration)
ChinaRespect	Affective feeling about Chinese people ranging from 1 (contempt) to 7 (respect)
PosDiff	Difference between the Chinese and Japanese positivity score
WarmDiff	Difference between the Chinese and Japanese warmth score
AdmireDiff	Difference between the Chinese and Japanese admiration score
RespectDiff	Difference between the Chinese and Japanese respect score
JapanID_avg	Average of 30 ratings of Japanese people on identity trait pairs, coded from 1 to 7 where higher numbers are less favorable; see p. 12 of Wang et al. (2021) for details
ChinaID_avg	Average of the same 30 identity ratings of Chinese people
ID_diff_avg	Difference between `ChinaID_avg` and `JapanID_avg`
age	Age in years
gender	Gender, coded 1 for male and 0 for female
jpfriend	Indicator variable for if subject has a Japanese friend (1) or does not (0)
MediaInd	Attitude toward media independence from the government ranging from 1 (strongly oppose) to 5 (strongly support)
freetrade	Indicator variable for if subject supports free trade (1) or does not (0)
school_major	Categorical variable denoting major in school; 1 = social sciences, 2 = humanities, 3 = sciences and engineering, 4 = law
PrejControl	Motivation to Control Prejudice index; an average of 17 items rated from 1 to 7 in which higher scores denote a greater motivation to control the expression of prejudice

```
# load the data
ChinaJapan <- read_csv('ChinaJapanData.csv', show_col_types = FALSE)
```

Question 1

Part a

The new variables `avg_jap`, `avg_chi`, and `avg_diff` are created below to reflect the average affective feeling toward Japanese people, the average affective feeling toward Chinese people, and the average difference between the two, respectively.

```
ChinaJapan <- ChinaJapan %>%
  mutate(avg_jap = (JapanPos + JapanWarm + JapanAdmire + JapanRespect)/4) %>%
  mutate(avg_chi = (ChinaPos + ChinaWarm + ChinaAdmire + ChinaRespect)/4) %>%
  mutate(avg_diff = avg_chi - avg_jap)
```

Part b

Below, I run multiple t-tests to determine the statistical significance associated with the mean values. I did this for all three affect averages from part (a), as well as two individual affect items: `JapanAdmire` and `ChinaAdmire`. The results are printed below.

As shown, the averages are substantially different between the treatment and control groups for the `avg_jap`, `avg_diff`, and `JapanAdmire` variables, while the averages are similar between the two groups for the `avg_chi` and `ChinaAdmire` variables. This shows up in the t-tests: the two variables with similar averages across the treatment and control groups had p-values greater than 0.05, meaning that at the 95% confidence level, we cannot reject the null hypothesis that the average in the treatment group is the same as the average in the control group.

Thus, it seems that imagined social contact did (positively) change subjects' affect toward Japanese people, but not toward Chinese people. Since the average affect score toward Japanese people is lower than the average affect score toward Chinese people for both the treatment and control groups, the fact that there was an increase in the affect toward Japanese people as a result of the treatment means that affective polarization is being reduced, since the gap in the scores is being narrowed as a result of the treatment.

```
q1b <- ChinaJapan %>%
  group_by(treated) %>%
  select(treated, avg_jap, avg_chi, avg_diff, JapanAdmire, ChinaAdmire) %>%
  summarise(avg_jap = mean(avg_jap),
            avg_chi = mean(avg_chi),
            avg_diff = mean(avg_diff),
            jap_adm = mean(JapanAdmire),
            chi_adm = mean(ChinaAdmire))

q1b
```

```
## # A tibble: 2 x 6
##   treated avg_jap avg_chi avg_diff jap_adm chi_adm
##   <lgl>     <dbl>   <dbl>    <dbl>   <dbl>   <dbl>
## 1 FALSE      3.72    5.00    1.28     3.45    4.75
## 2 TRUE       4.25    4.89    0.638    4.28    4.77
```

```
ChinaJapan_treated <- ChinaJapan %>%
  filter(treated == "TRUE")

ChinaJapan_control <- ChinaJapan %>%
  filter(treated == "FALSE")

t.test(ChinaJapan_treated$avg_jap, ChinaJapan_control$avg_jap)
```

```
##
##  Welch Two Sample t-test
##
## data:  ChinaJapan_treated$avg_jap and ChinaJapan_control$avg_jap
## t = 2.9066, df = 117.97, p-value = 0.004365
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1699774 0.8966892
## sample estimates:
## mean of x mean of y
##  4.250000  3.716667
```

```r
t.test(ChinaJapan_treated$avg_chi, ChinaJapan_control$avg_chi)
```

```
##
##  Welch Two Sample t-test
##
## data:  ChinaJapan_treated$avg_chi and ChinaJapan_control$avg_chi
## t = -0.57654, df = 117.99, p-value = 0.5654
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4804340  0.2637673
## sample estimates:
## mean of x mean of y
##  4.887500  4.995833
```

```r
t.test(ChinaJapan_treated$avg_diff, ChinaJapan_control$avg_diff)
```

```
##
##  Welch Two Sample t-test
##
## data:  ChinaJapan_treated$avg_diff and ChinaJapan_control$avg_diff
## t = -2.244, df = 117.91, p-value = 0.0267
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.20791316 -0.07542017
## sample estimates:
## mean of x mean of y
##  0.637500  1.279167
```

```r
t.test(ChinaJapan_treated$JapanAdmire, ChinaJapan_control$JapanAdmire)
```

```
##
##  Welch Two Sample t-test
##
## data:  ChinaJapan_treated$JapanAdmire and ChinaJapan_control$JapanAdmire
## t = 3.4294, df = 117, p-value = 0.0008363
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3520847 1.3145820
## sample estimates:
## mean of x mean of y
##  4.283333  3.450000
```

```r
t.test(ChinaJapan_treated$ChinaAdmire, ChinaJapan_control$ChinaAdmire)
```

```
##
##  Welch Two Sample t-test
##
## data:  ChinaJapan_treated$ChinaAdmire and ChinaJapan_control$ChinaAdmire
## t = 0.075179, df = 116.95, p-value = 0.9402
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.4223870  0.4557204
## sample estimates:
## mean of x mean of y
##  4.766667  4.750000
```

Part c

For each of the above variables, I calculated a Cohen's D value. As shown below, the value is small-medium for `avg_diff`, medium for `avg_jap`, negligible for `avg_chi`, medium-large for `JapanAdmire`, and negligible for `ChinaAdmire`. These interpretations match my intuitive interpretation, which is that the respondents in the treatment group would experience some difference in affect toward the out-group. These interpretations are also consistent with my interpretation from (b), which stated that `avg_diff`, `avg_jap`, and `JapanAdmire` had statistically-significant differences between the treatment and control groups.

```
cohen.d(ChinaJapan_treated$avg_diff, ChinaJapan_control$avg_diff)
```

```
##
## Cohen's d
##
## d estimate: -0.4097052 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.7750252 -0.0443853
```

```
cohen.d(ChinaJapan_treated$avg_jap, ChinaJapan_control$avg_jap)
```

```
##
## Cohen's d
##
## d estimate: 0.5306791 (medium)
## 95 percent confidence interval:
##     lower     upper
## 0.1628238 0.8985343
```

```
cohen.d(ChinaJapan_treated$avg_chi, ChinaJapan_control$avg_chi)
```

```
##
## Cohen's d
##
## d estimate: -0.1052607 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.4670576  0.2565362
```

```
cohen.d(ChinaJapan_treated$JapanAdmire, ChinaJapan_control$JapanAdmire)
```

```
##
## Cohen's d
##
## d estimate: 0.6261119 (medium)
## 95 percent confidence interval:
##     lower     upper
## 0.2558130 0.9964109
```

```r
cohen.d(ChinaJapan_treated$ChinaAdmire, ChinaJapan_control$ChinaAdmire)
```

```
##
## Cohen's d
##
## d estimate: 0.01372575 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.3478251  0.3752766
```

Question 3: Data Science Question

Part a

For this question, I chose `AffectDiff\_avg`, the average difference in affect scores toward the Chinese versus toward the Japanese, as the dependent variable. In addition to the `treated` variable (which indicates whether a respondent was in the treatment or control group), I chose `jpfriend` (whether the respondent has a Japanese friend), `age`, and `gender` as control variables. I hypothesize that respondents with Japanese friends would have a smaller value of `AffectDiff\_avg` (thus indicating a more similar attitude toward Chinese people as toward Japanese people), when holding all other variables constant. I hypothesize the same for younger people and for females – that younger or female people, when holding all other variables constant, would have a lower value of `AffectDiff\_avg`.

Part b

Below, I ran a linear multiple regression model with the three variables I mentioned above, as well as the `treated` variable, as the independent variables. The regression output is printed below. The `treated` variable has an estimated coefficient of -0.64566, which indicates that respondents in the treatment group had a smaller difference in affect scores toward Chinese versus affect scores toward Japanese; this coefficient is also statistically significant at the 95% confidence level, with a p-value of 0.0273 (i.e. less than 0.05). However, the rest of the variables did *not* have statistically significant coefficients: all of the p-values were substantially higher than 0.05, meaning that the coefficient is not statistically different from zero (and thus making the interpretation not meaningful).

```r
model3b <- lm(avg_diff ~ treated + jpfriend + age + gender, data = ChinaJapan)
summary(model3b)
```

```
##
## Call:
## lm(formula = avg_diff ~ treated + jpfriend + age + gender, data = ChinaJapan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3168 -0.8641 -0.0623  0.8426  4.3207
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.13453    1.05625   2.021   0.0456 *
## treatedTRUE -0.64566    0.28874  -2.236   0.0273 *
## jpfriend     0.17804    0.29850   0.596   0.5520
## age         -0.04055    0.04620  -0.878   0.3819
```

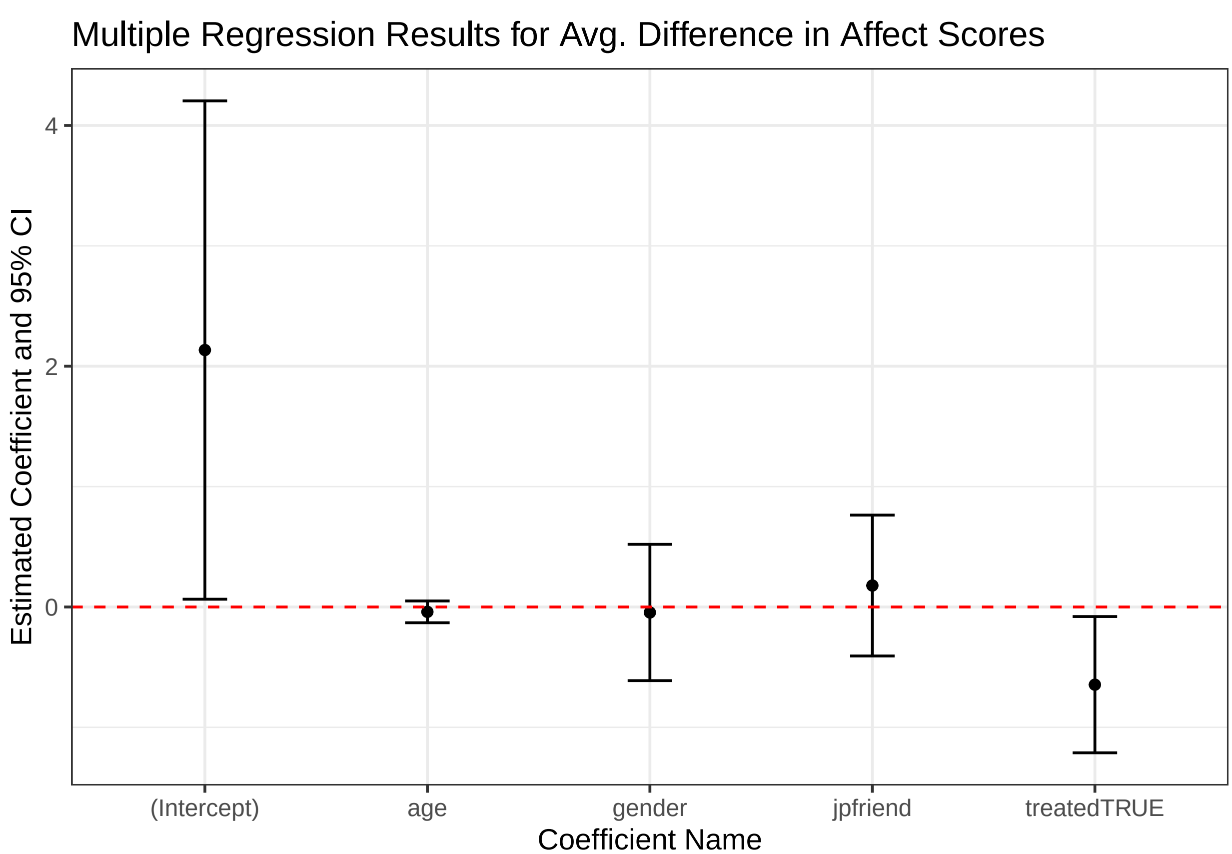
```
## gender      -0.04588    0.28881  -0.159   0.8741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.578 on 115 degrees of freedom
## Multiple R-squared:  0.05091,    Adjusted R-squared:  0.0179
## F-statistic: 1.542 on 4 and 115 DF,  p-value: 0.1946
```

Part c

Below, I created a coefficient regression plot containing the visualized results of the regression from part (b).

```
table3c <- tidy(model3b)
table3c <- table3c %>%
  mutate(lower = estimate - 1.96*std.error) %>%
  mutate(upper = estimate + 1.96*std.error)

table3c %>%
  ggplot(aes(x = term, y = estimate)) +
  geom_point() +
  geom_errorbar(aes(ymin = lower, ymax = upper, width = 0.2)) +
  geom_hline(yintercept = 0, color = "red", lty = 2) +
  xlab("Coefficient Name") + ylab("Estimated Coefficient and 95% CI") +
  ggtitle("Multiple Regression Results for Avg. Difference in Affect Scores") +
  theme_bw()
```



Part Two: Harvard Students and Perception of Members of the Other Party

Data Details:

- File Name: `Oct28ClassData.csv`
- Source: These data are from the Qualtrics survey you all took last week. About half of you were in the control condition and were simply asked to imagine a bus ride with beautiful scenery; the other half were in the treatment condition and were asked to imagine a bus ride next to a member of your non-preferred US political party. All students were then asked a series of questions to assess their affective feelings toward Democrats and Republicans and their perceptions of the characteristics of Democratic and Republican identity. Additional control variables were merged from the class background survey.

Variable Name	Variable Description
`Treated`	Binary variable equal to `TRUE` if the subject was told to imagine a bus ride with a member of the opposing political party (the treatment) and `FALSE` if the subject was told to imagine the scenery on a bus ride (control)
`ClosestParty`	Which of the two major US political parties the subject feels closes to
`strongPARTISAN`	Binary variable coded as `TRUE` if subject self-identifies as a strong partisan and `FALSE` otherwise
`ControlScenario_1`	First text-based reflection of the imagined bus ride for those in the control condition
`ControlScenario_2`	Second text-based reflection of the imagined bus ride for those in the control condition
`ControlScenario_3`	Third text-based reflection of the imagined bus ride for those in the control condition
`TreatmentScenario_1`	First text-based reflection of the imagined bus ride for those in the treatment condition
`TreatmentScenario_2`	Second text-based reflection of the imagined bus ride for those in the treatment condition
`TreatmentScenario_3`	Third text-based reflection of the imagined bus ride for those in the treatment condition
`RepublicanAffect_1`	Affective feeling about Republicans ranging from 1 (negative) to 7 (positive)
`RepublicanAffect_2`	Affective feeling about Republicans ranging from 1 (cool) to 7 (warm)
`RepublicanAffect_3`	Affective feeling about Republicans ranging from 1 (loathing) to 7 (admiration)
`RepublicanAffect_4`	Affective feeling about Republicans ranging from 1 (contempt) to 7 (respect)
`DemocraticAffect_1`	Affective feeling about Democrats ranging from 1 (negative) to 7 (positive)
`DemocraticAffect_2`	Affective feeling about Democrats ranging from 1 (cool) to 7 (warm)
`DemocraticAffect_3`	Affective feeling about Democrats ranging from 1 (loathing) to 7 (admiration)
`DemocraticAffect_4`	Affective feeling about Democrats ranging from 1 (contempt) to 7 (respect)

Variable Name	Variable Description
`RepublicanIdentity_1`	Rating of Republican identity trait from 1 (obstinate) to 7 (open-minded) (note that identity favorability is associated with the higher number, unlike in the data from Part One)
`RepublicanIdentity_2`	Rating of Republican identity trait from 1 (evil) to 7 (moral)
`RepublicanIdentity_3`	Rating of Republican identity trait from 1 (arrogant) to 7 (humble)
`RepublicanIdentity_4`	Rating of Republican identity trait from 1 (cruel) to 7 (kind)
`DemocraticIdentity_1`	Rating of Democratic identity trait from 1 (obstinate) to 7 (open-minded)
`DemocraticIdentity_2`	Rating of Democratic identity trait from 1 (evil) to 7 (moral)
`DemocraticIdentity_3`	Rating of Democratic identity trait from 1 (arrogant) to 7 (humble)
`DemocraticIdentity_4`	Rating of Democratic identity trait from 1 (cruel) to 7 (kind)
`gender`	Character variable reflecting self-identified gender
`college_stats`	Binary variable coded as `TRUE` if subject self-identifies as having taken college-level statistics and `FALSE` otherwise
`year`	Year in college from 1 to 4
`US`	Binary variable coded as `TRUE` if subject self-identifies as having been born in the United States and `FALSE` otherwise
`InPartyAffect_(1-4)`	Affective feelings about the in-party using the same numbering scheme as above
`InPartyAffect_avg`	Average affective feelings about the in-party
`OutPartyAffect_(1-4)`	Affective feelings about the out-party using the same numbering scheme as above
`OutPartyAffect_avg`	Average affective feelings about the out-party
`InPartyIdentity_(1-4)`	Identity ratings about the in-party using the same numbering scheme as above
`InPartyIdentity_avg`	Average identity ratings about the in-party
`OutPartyIdentity_(1-4)`	Identity ratings about the out-party using the same numbering scheme as above
`OutPartyAffect_avg`	Average identity ratings about the out-party
`AffectDiff_(1-4)`	Difference in affective feelings between in-party and out-party using the same numbering scheme as above
`AffectDiff_avg`	Average difference in affective feelings between in-party and out-party
`IdentityDiff_(1-4)`	Difference in identity ratings between in-party and out-party using the same numbering scheme as above
`IdentityDiff_avg`	Average difference in identity ratings between in-party and out-party

```r
# load the data
ClassExperiment <- read_csv('Oct28ClassData.csv', show_col_types = FALSE)

# run models (for blog)
model1 <- lm(AffectDiff_avg ~ Treated, data = ClassExperiment)
model2 <- lm(AffectDiff_avg ~ Treated + ClosestParty, data = ClassExperiment)
model3 <- lm(AffectDiff_avg ~ Treated + strongPARTISAN, data = ClassExperiment)
model4 <- lm(AffectDiff_avg ~ Treated + gender + college_stats + year + US,
             data = ClassExperiment)
model5 <- lm(AffectDiff_avg ~ Treated + ClosestParty + strongPARTISAN + gender +
               college_stats + year + US, data = ClassExperiment)
```

```r
stargazer(model1, model2, model3, model4, model5, header = F,
          title = "Harvard Students' Perception of the Out-Party",
          column.sep.width = "1pt", font.size = "tiny")
```

Table 3: Harvard Students' Perception of the Out-Party

	Dependent variable:				
	AffectDiff_avg				
	(1)	(2)	(3)	(4)	(5)
Treated	0.230	0.224	0.246	0.185	0.170
	(0.339)	(0.338)	(0.314)	(0.347)	(0.321)
ClosestPartythe Republican Party		−0.769			−0.433
		(0.615)			(0.611)
strongPARTISAN			1.130***		1.123***
			(0.313)		(0.366)
genderWoman				−0.130	−0.325
				(0.360)	(0.337)
college_stats				−0.482	−0.320
				(0.378)	(0.352)
year				−0.485**	−0.300
				(0.213)	(0.220)
US				−0.006	0.031
				(0.391)	(0.363)
Constant	2.007***	2.073***	1.426***	3.696***	2.581***
	(0.245)	(0.249)	(0.278)	(0.791)	(0.889)
Observations	73	73	73	68	68
R^2	0.006	0.028	0.162	0.104	0.259
Adjusted R^2	−0.008	0.0004	0.138	0.032	0.172
Residual Std. Error	1.448 (df = 71)	1.442 (df = 70)	1.339 (df = 70)	1.416 (df = 62)	1.309 (df = 60)
F Statistic	0.459 (df = 1; 71)	1.013 (df = 2; 70)	6.769*** (df = 2; 70)	1.440 (df = 5; 62)	2.994*** (df = 7; 60)
Note:					*p<0.1; **p<0.05; ***p<0.01