

FLS 6415: Class 4 Homework

September 14, 2017

Remember to answer all the questions in R markdown and produce a PDF. Email your completed homework (R markdown file and PDF) to jonnyphillips@gmail.com by midnight the night before class.

The homework is to analyse the data from Whitt (2014). From the class website, open the data file *replication.dta* in R. Each row of the dataset represents one ‘game’, i.e. an allocation decision by one subject between two recipients. The key variables are:

- **ethnicity** - the *subject*’s ethnicity
- **d4ethnicity1** - the first recipient’s ethnicity
- **d4ethnicity2** - the second recipient’s ethnicity
- **d4amountsent1** - the amount the subject allocated to the first recipient
- **d4amountsent2** - the amount the subject allocated to the second recipient

The analysis may be easier if you use specific functions to assist in data manipulation, so be sure to check the example code on the course website for ideas while you’re doing the homework - the hints below reference these functions and should directly correspond to the examples in the `example_code`.

1. Briefly describe what the ‘treatment’ is in this experiment.

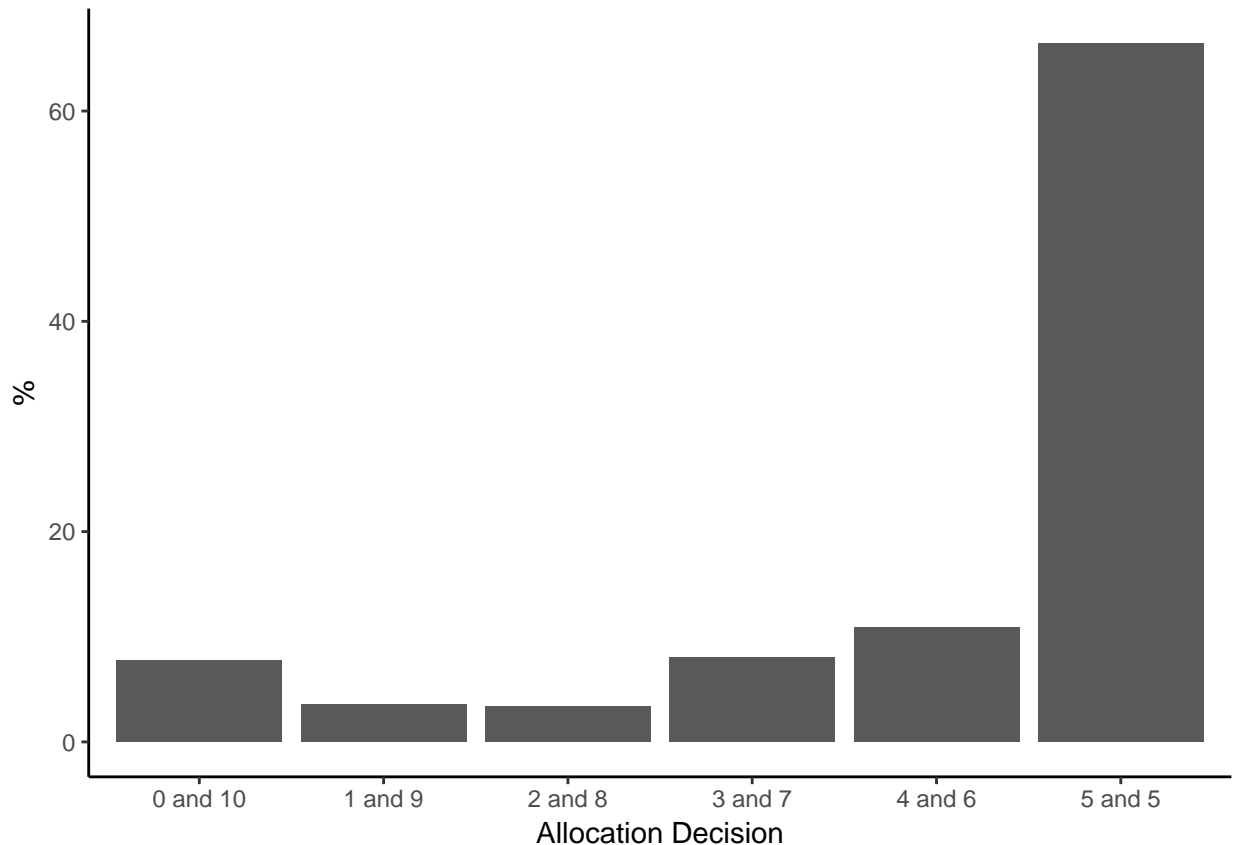
The treatment is the ethnicity of the two recipients in the version of the dictator game that is presented to the subject. For the specific test of whether subjects treat co-ethnics differently, treatment is the presence of a recipient of the same ethnicity among the two recipients (while control is the absence of a co-ethnic).

2. To describe the distribution of outcomes for the full sample, replicate Figure 1. *Hint:* Try converting the outcome variable to a factor and then using the `fct_collapse` function in the `forcats` package to recode the specific number to each of the groups (eg. “5 and 5”). To quickly get the percentage in each factor category, see the example code for how to summarise factor variables.

```
library(tidyverse)
library(forcats)
library(dplyr)
library(knitr)

fig1 <- d %>% mutate(d4amountsent1_fig1 = fct_collapse(factor(d4amountsent1),
  `0 and 10` = c("0", "10"), `1 and 9` = c("1", "9"), `2 and 8` = c("2", "8"),
  `3 and 7` = c("3", "7"), `4 and 6` = c("4", "6"), `5 and 5` = c("5", "5"))) %>%
  group_by(d4amountsent1_fig1) %>% summarise(n = n()) %>% mutate(pct = (n/sum(n)) *
  100)

ggplot(fig1) + geom_col(aes(x = d4amountsent1_fig1, y = pct)) + theme_classic() +
  xlab("Allocation Decision") + ylab("%")
```



3. Calculate the average amount that each subject allocated to a co-ethnic (a respondent of the same ethnicity as the subject). What can we conclude from this? *Hint*: Create a new variable to record the amount transferred to co-ethnics. Use the `case_when` function to combine the values for respondent one and respondent two into a single column.

```
co_ethnic_amount <- d %>% mutate(co_ethnic_amount = case_when(ethnicity == d4ethnicity1 ~
  d4amountsent1, ethnicity == d4ethnicity2 ~ d4amountsent2)) %>% summarise(co_ethnic_amount = mean(co_
  na.rm = TRUE))
```

The average amount subjects transferred to co-ethnics is 5.746. Since this is greater than an equal distribution of 5, this implies that there is a bias towards co-ethnics.

4. Produce a 3x3 table showing the average allocation decision, with the *recipient's* ethnicity in the columns and the *subject's* ethnicity in the rows. *Hint*: Create a new variable for each ethnicity showing the amount transferred to that ethnicity when they participated in the game, and then summarise these values.

```
library(knitr)
d <- d %>% mutate(amount_bosniak = case_when(d4ethnicity1 == "Bosnjak" ~ d4amountsent1,
  d4ethnicity2 == "Bosnjak" ~ d4amountsent2))
d <- d %>% mutate(amount_croat = case_when(d4ethnicity1 == "Croat" ~ d4amountsent1,
  d4ethnicity2 == "Croat" ~ d4amountsent2))
d <- d %>% mutate(amount_serb = case_when(d4ethnicity1 == "Serb" ~ d4amountsent1,
  d4ethnicity2 == "Serb" ~ d4amountsent2))

ethnicity_table <- d %>% group_by(ethnicity) %>% summarise(mean_bosniak = round(mean(amount_bosniak,
  na.rm = TRUE), 3), mean_croat = round(mean(amount_croat, na.rm = TRUE),
  3), mean_serb = round(mean(amount_serb, na.rm = TRUE), 3))
```

```
colnames(ethnicity_table) <- c("Ethnicity", "Bosnjak", "Croat", "Serb")
kable(ethnicity_table, caption = "Allocations from Subject Ethnicity (rows) to Recipient Ethnicity (columns)")
```

Table 1: Allocations from Subject Ethnicity (rows) to Recipient Ethnicity (columns)

Ethnicity	Bosnjak	Croat	Serb
Bosnjak	5.714	4.920	4.374
Croat	4.446	5.913	4.612
Serb	4.601	4.732	5.629

5. From the table you produced in Q4, which ethnicity exhibits the greatest co-ethnic/in-group bias? Which ethnic pairing is most asymmetric (i.e. where A treats B better than B treats A)?

```
ethnicity_table_long <- ethnicity_table %>% gather("Bosnjak", "Croat", "Serb",
  key = "Recipient_Ethnicity", value = "Allocation")
most_biased <- ethnicity_table_long %>% filter(Recipient_Ethnicity == Ethnicity) %>%
  filter(Allocation == max(Allocation)) %>% select(Ethnicity)

most_asymm <- ethnicity_table_long %>% filter(Recipient_Ethnicity != Ethnicity) %>%
  mutate(Pairing = paste(pmin(Ethnicity, Recipient_Ethnicity), pmax(Ethnicity,
    Recipient_Ethnicity))) %>% group_by(Pairing) %>% arrange(Allocation) %>%
  summarise(Difference = diff(Allocation)) %>% filter(Difference == max(Difference)) %>%
  select(Pairing)
```

The amount transferred to co-ethnics is greatest among the ethnicity Croat.

The most ‘asymmetric’ relationship is between Bosnjak Croat. Bosniak subjects give much more to Croat recipients than Croat subjects give to Bosniak recipients.

6. Another way to analyse the data is with a regression. Conduct an OLS regression to assess whether *Bosniak* subjects allocate more or less if the first recipient (use variables *d4ethnicity1*, *d4amountsent1*) is a co-ethnic (ignore the second recipient and amount for now). Interpret the results of this regression. *Hint*: Create a new dataset including data only for Bosniak subjects.

```
library(xtable)
d_bosniak <- d %>% filter(ethnicity == "Bosnjak")
print(xtable(summary(lm(d4amountsent1 ~ d4ethnicity1, data = d_bosniak))), comment = FALSE)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.8375	0.1826	31.98	0.0000
d4ethnicity1Croat	-0.9966	0.2522	-3.95	0.0001
d4ethnicity1Serb	-1.2953	0.2558	-5.06	0.0000

The regression shows that, relative to the base category of a Bosniak co-ethnic, Bosniak subjects allocate less if the first recipient is Croat (by about 1 unit on average) and even less if they are Serb (by about 1.3 units). This further confirms the existence of a co-ethnic bias for Bosniaks.

7. Now conduct a regression that includes subjects of *all* ethnicities and includes an interaction term between subject ethnicity and recipient ethnicity. Again, only use the data for the first recipient (*d4ethnicity1*, *d4amountsent1*), ignoring the second recipient. Interpret the results of this regression.

```
print(xtable(lm(d4amountsent1 ~ d4ethnicity1 * ethnicity, data = d)), comment = FALSE)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.8375	0.1936	30.15	0.0000
d4ethnicity1Croat	-0.9966	0.2675	-3.73	0.0002
d4ethnicity1Serb	-1.2953	0.2713	-4.77	0.0000
ethnicityCroat	-1.2637	0.2943	-4.29	0.0000
ethnicitySerb	-1.3243	0.2774	-4.77	0.0000
d4ethnicity1Croat:ethnicityCroat	2.1728	0.4002	5.43	0.0000
d4ethnicity1Serb:ethnicityCroat	1.3037	0.4093	3.19	0.0015
d4ethnicity1Croat:ethnicitySerb	0.9201	0.3914	2.35	0.0190
d4ethnicity1Serb:ethnicitySerb	2.1061	0.3940	5.34	0.0000

With interaction terms in our regression, we need to define the base category carefully: It is for the Bosniak subject and the Bosniak first recipient. We also need to be careful to look at the net effect of the base terms and the interaction to understand the differences. Relative to the baseline:

1. At the baseline, the estimated allocation from Bosniak subjects to Bosniak recipients is the intercept (5.8375);
2. Bosniak subjects allocate less to Croats and Serbs (-0.99, -1.29), as in Q6;
3. Croat and Serb subjects allocate less to Bosniak recipients (-1.26, -1.32);
4. Croat subjects allocate about the same to co-ethnic Croats (2.17 - 1.26 - 0.99 ~ 0);
5. Croat subjects allocate substantially less to Serbs (1.30 - 1.26 - 1.32 < 0);
6. Serb subjects allocate less to co-ethnic Serbs (2.10 - 1.32 - 1.30 < 0);
7. Serb subjects allocate substantially less to Croats (0.92 - 1.32 - 0.99 < 0);

So far we have assumed that the identity of the second recipient does not affect the amount allocated to the first recipient. But it might be that, for example, subjects prefer to give more to Croats when the alternative is Serbs compared to when the alternative is Bosniaks.

8. To see if the exact pairings in the game make a difference, first measure the average allocation that Bosniak subjects make when they are forced to choose between a Bosniak and a Croat (this is the first line of Table 3 in Whitt 2014). *Hint:* Use your dataset that focuses only on Bosniak subjects. And remember that the Bosniak-Croat pairing of subjects occurs in two ways, once with Bosniak as the first recipient and once as the second recipient.

```
d_bosniak_b_c <- d_bosniak %>% filter(d4ethnicity1 == "Bosnjak" & d4ethnicity2 ==
  "Croat" | d4ethnicity1 == "Croat" & d4ethnicity2 == "Bosnjak") %>% summarise(mean_bosniak = mean(amount_bosniak, na.rm = TRUE), mean_croat = mean(amount_croat, na.rm = TRUE))
```

For Bosniaks, and considering both the first and second recipients, co-ethnic Bosniaks consistently receive a larger allocation, 5.55 while Croats receive 4.45.

9. Replicate all the rows of Table 3 for all subject and recipient ethnicity combinations. Report the mean to each recipient and the t-test for the difference in means (the first 5 columns and the last column). Don't worry about the exact order or structure of the table as long as it contains the right results. *Hint:* This question probably requires a bit more (repetitive) coding. One approach is to create a separate dataset for each possible combination of subject and recipient ethnicities and summarise each dataset separately. Then combine all your estimates into a single table at the end, for example using rbind.

```
d <- d %>% filter(d4ethnicity2 != "") #Annoying missing value needs to be removed

# The tidyverse method (see code below for alternative method):

Q9_table <- d %>% mutate(Pairing = paste(pmin(d4ethnicity1, d4ethnicity2), pmax(d4ethnicity1,
```

```

d4ethnicity2))) %>% separate(Pairing, into = c("Ethnicity1", "Ethnicity2"),
sep = " ", remove = FALSE) %>% mutate(amount1 = case_when(d4ethnicity1 ==
Ethnicity1 ~ d4amountsent1, d4ethnicity2 == Ethnicity1 ~ d4amountsent2)) %>%
mutate(amount2 = case_when(d4ethnicity1 == Ethnicity2 ~ d4amountsent1, d4ethnicity2 ==
Ethnicity2 ~ d4amountsent2)) %>% group_by(ethnicity, Ethnicity1, Ethnicity2) %>%
summarise(Amount1 = round(mean(amount1), 3), Amount2 = round(mean(amount2),
3), t.test_p.value = round(t.test(amount1, amount2)$p.value, 3)) %>%
select(Subject_Ethnicity = ethnicity, Ethnicity1, Amount1, Ethnicity2, Amount2,
t.test_p.value)

kable(Q9_table, caption = "Allocations by Ethnicity across Both Recipients")

```

Table 2: Allocations by Ethnicity across Both Recipients

Subject_Ethnicity	Ethnicity1	Amount1	Ethnicity2	Amount2	t.test_p.value
Bosnjak	Bosnjak	5.550	Croat	4.450	0.000
Bosnjak	Bosnjak	5.864	Serb	4.136	0.000
Bosnjak	Croat	5.373	Serb	4.627	0.006
Croat	Bosnjak	3.953	Croat	6.047	0.000
Croat	Bosnjak	4.924	Serb	5.076	0.653
Croat	Croat	5.808	Serb	4.192	0.000
Serb	Bosnjak	4.925	Croat	5.075	0.583
Serb	Bosnjak	4.316	Serb	5.684	0.000
Serb	Croat	4.427	Serb	5.573	0.000

10. Using your findings in Q9 (or Table 3 in Whitt 2014), how do subjects allocate the money when there is no co-ethnic present in the experiments?

Strikingly, Croats and Serbs exhibit no bias when allocating between two non-co-ethnics. However, Bosniaks are strongly biased towards Croats over Serbs. In fact, their bias towards Croats against Serbs is almost as large as their bias towards co-ethnic Bosniaks over Croats. This seems consistent with the political and military alignments during the war.

11. Whitt’s Table 5 investigates whether subjects’ attitudes to ethnicity explain their allocations. To conduct a similar analysis (not a full replication as some variables are missing/messy), first filter the data to cases where the subject has a co-ethnic (in-group member) among the recipients. Next create a binary variable ‘bias’ which equals 1 if the allocation is unequal and 0 if the allocation is 5:5. Then conduct a regression of this bias variable on the survey measure “How important is your ethnicity to you?” (variable *ethnicities* [sic]). Include control variables for the subject’s age (*age*), education (*education*) and ethnicity (*ethnicity*). Interpret briefly the results of the regression.

```

# Subset to games where in-group present
d_in_group <- d %>% mutate(in_group = ifelse(ethnicity == d4ethnicity1 | ethnicity ==
d4ethnicity2, 1, 0)) %>% filter(in_group == 1) %>% mutate(bias = ifelse(d4amountsent1 ==
5, 0, 1))

print(xtable(summary(glm(bias ~ ethnicities + ethnicity + age + education, data = d_in_group,
family = "binomial")), caption = "How Ethnic Attitudes affect Allocations"),
comment = FALSE)

```

The regression shows that - relative to the base category of “not important at all” (and controlling for age, education and subject ethnicity) - respondents how report that ethnicity is “somewhat” important to them are more likely to bias their allocations towards co-ethnics (about 0.7 units more; statistically significant

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.6664	0.5954	-1.12	0.2630
ethnicitiesSomewhat important	0.7083	0.3437	2.06	0.0393
ethnicitiesVery important	1.5132	0.3632	4.17	0.0000
ethnicityCroat	0.0822	0.2558	0.32	0.7479
ethnicitySerb	0.2343	0.2471	0.95	0.3431
age	-0.0150	0.0082	-1.84	0.0659
educationSome secondary	-0.9570	0.6263	-1.53	0.1265
educationCompleted secondary	-0.5701	0.4077	-1.40	0.1620
educationHigher education	0.2626	0.4442	0.59	0.5544

Table 3: How Ethnic Attitudes affect Allocations

at $p=0.05$), and respondents reporting that ethnic ties are “very” important to them exhibit even greater allocation bias towards co-ethnics (about 1.5 units more; statistically significant at $p=0.001$)