Exercise: Analyzing Survey Experiments

Download the datasets from the website containing responses to survey experiments. Answer the following questions:

```
survey <- read_csv("Survey_data.csv")
```

Priming Experiment

`Survey_data.csv`

1. Respondents were primed with one of two questions (variable `Primed`, prime 'hope'=0 and prime 'anger'=1). How does this prime affect subsequent answers to the question about the need for constitutional reform (variable `Constitution_Reform_Support`)? Calculate the difference-in-means estimate of the average treatment effect. Interpret the result.

```
survey %>% t.test(Constitution_Reform_Support~Primed, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  Constitution_Reform_Support by Primed
## t = -5.2416, df = 2050.8, p-value = 1.756e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.15704462 -0.07152551
## sample estimates:
## mean in group 0 mean in group 1
##       0.4970986       0.6113837
```

2. Perform the same analysis as in Q1, but using a simple OLS regression.

```
survey %>% lm(Constitution_Reform_Support ~ Primed, data=.) %>% stargazer(digits=3, header=F)
```

List Experiment

`Survey_data.csv`

The data is for the list experiment of how many activities respondents have done in the past one year, with three items in control and a fourth item relating to clientelism in the treatment.

The number of items the respondent says they have done are contained in the variable `list_response`. If they were shown a control list (3 items) the variable `list_treated` is equal to zero, and if they were shown a treatment list (4 items) the variable `list_treated` is equal to one.

3. Calculate the average treatment effect by subtracting the mean number of responses between the treated and control lists. Interpret the results in terms of the proportion of respondents who have experienced the sensitive item.

```
survey %>% t.test(list_response~list_treated, data=.)
```

```
##
##  Welch Two Sample t-test
```

Table 1:

	Dependent variable:
	Constitution_Reform_Support
Primed	0.114***
	(0.022)
Constant	0.497***
	(0.015)
Observations	2,053
R^2	0.013
Adjusted R^2	0.013
Residual Std. Error	0.494 (df = 2051)
F Statistic	27.464*** (df = 1; 2051)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
##
## data:  list_response by list_treated
## t = -6.3197, df = 1947.1, p-value = 3.24e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2736258 -0.1440183
## sample estimates:
## mean in group 0 mean in group 1
##        1.264151        1.472973
```

```
#OR
survey %>% group_by(list_treated) %>%
  summarize(mean=mean(list_response,na.rm=T)) %>%
  mutate(ATE=mean-lag(mean)) %>%
  kable(digits=4)
```

list_treated	mean	ATE
0	1.2642	NA
1	1.4730	0.2088

4. Conduct the same analysis using a simple OLS regression. What is the 95% confidence interval of the proportion of respondents who received a clientelist offer?

```
survey %>% lm(list_response~list_treated, data=.) %>%
  tidy() %>%
  mutate(conf.lo=estimate-std.error*1.96,
         conf.hi=estimate+std.error*1.96) %>%
  filter(term=="list_treated") %>%
  select(conf.lo, conf.hi) %>%
  kable(digits=4)
```

conf.lo	conf.hi
0.1438	0.2738

5. Now let's check the assumptions of the list experiment. First, check if there is a design effect using the function `ict.test` in the `list` package (for R only). (Note you will need to remove missing values of the `list_response` variable before running the test). Interpret the results based on the explanation in the outcome of `ict.test`.


```
survey_no_na <- survey %>% filter(is.na(list_response)==F)
```

```
## pi(Y_i(0) = 2, Z_i = 0)  0.2005 0.0172
## pi(Y_i(0) = 3, Z_i = 0)  0.0146 0.0076
##
##  Y_i(0) is the (latent) count of 'yes' responses to the control items. Z_i is the (latent) binary res
##
## Bonferroni-corrected p-value
## If this value is below alpha, you reject the null hypothesis of no design effect. If it is above alpl
##
## Sensitive Item 1
##        0.9467432
```

6. Next, let's check for floor and ceiling effects. There is a complex statistical test for this in the `list` package using the `ictreg` function (in R only): Try the code below and interpret the 'floor' and 'ceiling' parameters to see if they are statistically significant from zero. These are the estimates of whether anyone who should have answered '4' actually lied and answered '3', or who should have answered '1' actually lied and answered '0'.

```r
ictreg(list_response~1,
       data=survey %>% as.data.frame(),
       treat="list_treated",
       J=3,
       method="ml",
       floor=T,
       ceiling=T,
       floor.fit="bayesglm",
       ceiling.fit="bayesglm") %>%
  summary()
```

```
## Warning in ictreg(list_response ~ 1, data = survey %>% as.data.frame(), :
## log-likelihood is not monotonically increasing.
```

```
##
## Item Count Technique Regression
##
## Call: ictreg(formula = list_response ~ 1, data = survey %>% as.data.frame(),
##     treat = "list_treated", J = 3, method = "ml", floor = T,
##     ceiling = T, ceiling.fit = "bayesglm", floor.fit = "bayesglm")
##
## Sensitive item
##                 Est.    S.E.
## (Intercept) -1.61411 0.22439
##
## Control items
##                 Est.    S.E.
## (Intercept) -0.28742 0.03259
##
## Ceiling
##                 Est.    S.E.
## (Intercept) -5.25951 4.96644
##
## Floor
##                 Est.    S.E.
## (Intercept) -5.75978 4.61495
##
## Log-likelihood: -2368.2
##
```

```
## Number of control items J set to 3. Treatment groups were indicated by '1' and the control group by
```

7. The survey also asked people directly whether they had experienced the sensitive clientelism item (Has anyone ever offered you a gift, some food or money in exchange for your vote?) in variable `direct_clientelism`. Compare the non-response rate (NA responses) to the direct and indirect questions. Does this justify the use of a list experiment or not?

```
survey %>% summarize(na_list=(sum(is.na(list_response))/n())*100,
                     na_direct=(sum(is.na(direct_clientelism))/n())*100) %>%
  kable(digits=4)
```

na_list	na_direct
0.4871	8.1831

8. Compare the estimate of the incidence of clientelism from the direct responses to the indirect estimate from the list experiment. What does this suggest about the level of social desirability bias?

```
survey %>% summarize(direct_mean=mean(direct_clientelism,na.rm=T)) %>%
  kable(digits=4)
```

direct_mean
0.1767

9. Are men or women more likely to have experienced the sensitive item (clientelism)? Run an OLS regression with an interaction between the list treatment and gender to find out. Interpret the results.

```
survey %>% lm(list_response~list_treated*gender, data=.) %>%
  tidy() %>%
  kable(digits=4)
```

term	estimate	std.error	statistic	p.value
(Intercept)	1.2407	0.0288	43.1269	0.0000
list_treated	0.1645	0.0404	4.0697	0.0000
gendermale	0.0707	0.0500	1.4146	0.1573
genderother	0.5948	0.7469	0.7964	0.4259
list_treated:gendermale	0.1317	0.0701	1.8773	0.0606

Conjoint Experiment

`Conjoint_data.csv`

Respondents to a household survey were shown pairs of candidate profiles with different characteristics and asked which candidate they would vote for.

The dataset is arranged with one row for every candidate that each respondent assessed (2 experiments * 2 candidates * 4047 respondents). So there are four rows for every respondent - two experiments with two candidates in each. Column `UID` identifies each respondent, column `Round` describes whether it was the first or second experiment, and `Choice` identifies each candidate presented in each experiment.

The first columns (starting `Profile_...`) describe the attributes of each candidate profile. The variable `conjoint_choice` is a binary indicator of which of the two candidate profiles the respondent choose to vote for. There are also columns for the characteristics of the respondent (gender, age, and whether they are a co-ethnic of the candidate profile).

```r
conjoint <- read_csv("Conjoint_data.csv")
```

10. How many possible combinations of attributes for a single profile are there? Identify how many unique values are possible for each of the five attributes and multiply these together.

```r
2*3*5*2*2
```

```
## [1] 120
```

11. Run a simple OLS regression to evaluate how the 'Public Goods (PG)' attribute of a Profile affects the respondent's choice of candidate.

```r
conjoint %>% zelig(conjoint_choice~1+Profile_PG, data=.,model="ls", cite=F) %>%
  from_zelig_model() %>%
  stargazer(header=F)
```

Table 7:

	Dependent variable:
	conjoint_choice
Profile_PGGood Roads	0.221***
	(0.008)
Constant	0.369***
	(0.005)
Observations	16,188
R^2	0.049
Adjusted R^2	0.049
Residual Std. Error	0.487 (df = 16186)
F Statistic	829.426*** (df = 1; 16186)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
conjoint %>% lm(conjoint_choice~Profile_PG, data=.) %>%
  tidy() %>%
  kable(digits=4)
```

term	estimate	std.error	statistic	p.value
(Intercept)	0.3686	0.0054	67.8997	0
Profile_PGGood Roads	0.2206	0.0077	28.7998	0

12. Since our outcome is a binary variable, run the same regression but with a logit model.

```r
conjoint %>% zelig(conjoint_choice~1+Profile_PG, data=.,model="logit")%>%
  from_zelig_model() %>%
  stargazer(header=F)
```

How to cite this model in Zelig: R Core Team. 2007. logit: Logistic Regression for Dichotomous Dependent Variables in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," http://zeligproject.org/

13. Since each respondent participated in two experiments, their answers are likely to be highly correlated. So we have less 'N' than we think. Cluster the standard errors of your OLS regression according to the respondent identifier (UID).

Table 9:

	Dependent variable:
	conjoint_choice
Profile_PGGood Roads	0.899^{***}
	(0.032)
Constant	-0.538^{***}
	(0.023)
Observations	16,188
Log Likelihood	−10,809.020
Akaike Inf. Crit.	21,622.040
Note:	*p<0.1; **p<0.05; ***p<0.01

	Model 1
(Intercept)	0.37^{*}
	[0.36; 0.38]
Profile_PGGood Roads	0.22^{*}
	[0.20; 0.24]
R^2	0.05
Adj. R^2	0.05
Num. obs.	16188
RMSE	0.49

* 0 outside the confidence interval

Table 10: Statistical models

```
conjoint %>% lm_robust(conjoint_choice~1+Profile_PG, data=., clusters=UID)%>%
  texreg()
```

14. Assess the influence of all of the five profile attributes at the same time in an OLS regression. Interpret the results.

```
conjoint %>% zelig(conjoint_choice ~ 1 + Profile_Gender + Profile_Caste + Profile_Party +
    Profile_PG + Profile_Promise, data = ., model = "ls") %>% from_zelig_model() %>%
    stargazer(header = F)
```

How to cite this model in Zelig: R Core Team. 2007. ls: Least Squares Regression for Continuous Dependent Variables in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," http://zeligproject.org/

15. We can also assess how respondents' characteristics affect their choice. Does the importance of the 'Promise' attribute vary by gender? Use an interaction term between the Promise attribute and respondent gender (ignoring all other variables), and interpret the results.

```
conjoint %>% zelig(conjoint_choice~1+Profile_Promise*respondent_gender, data=.,model="logit")%>%
  from_zelig_model() %>%
  stargazer(header=F)
```

How to cite this model in Zelig: R Core Team. 2007. logit: Logistic Regression for Dichotomous Dependent Variables in Christine Choirat, Christopher Gandrud, James Honaker, Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," http://zeligproject.org/

Table 11:

	Dependent variable:
	conjoint_choice
Profile_GenderMale	0.002
	(0.008)
Profile_CasteOBC	0.034***
	(0.009)
Profile_CasteSC	0.021**
	(0.009)
Profile_PartyINC	−0.033***
	(0.012)
Profile_PartyJDU	0.009
	(0.011)
Profile_PartyJMM	−0.011
	(0.012)
Profile_PartyRJD	−0.038***
	(0.011)
Profile_PGGood Roads	0.221***
	(0.008)
Profile_PromisePatronage	0.001
	(0.008)
Constant	0.361***
	(0.011)
Observations	16,188
R^2	0.051
Adjusted R^2	0.050
Residual Std. Error	0.487 (df = 16178)
F Statistic	96.182*** (df = 9; 16178)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 12:

	Dependent variable:
	conjoint_choice
Profile_PromisePatronage	0.088^{*}
	(0.051)
respondent_gendermale	0.125^{***}
	(0.046)
Profile_PromisePatronage:respondent_gendermale	-0.135^{**}
	(0.065)
Constant	-0.162^{***}
	(0.036)
Observations	16,188
Log Likelihood	−11,203.050
Akaike Inf. Crit.	22,414.090
Note:	*p<0.1; **p<0.05; ***p<0.01