Homework 8

Mark Tao

```
library(knitr)
opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
knitr::opts_chunk$set(error = TRUE)
```

Analysis of the 1932 German Election during the Weimar Republic

Who voted for the Nazis? Researchers attempted to answer this question by analyzing aggregate election data from the 1932 German election during the Weimar Republic.

This exercise is based on the following article: King, Gary, Ori Rosen, Martin Tanner, Alexander F. Wagner. 2008. "Ordinary Economic Voting Behavior in the Extraordinary Election of Adolf Hitler." *Journal of Economic History* 68(4): 951-996.

We analyze a simplified version of the election outcome data, which records, for each precinct, the number of eligible voters as well as the number of votes for the Nazi party. In addition, the data set contains the aggregate occupation statistics for each precinct.

```
nazi <- read.csv("/data/qss/UNCERTAINTY/nazis.csv")
```

Name	Description
shareself	Proportion of self-employed potential voters
shareblue	Proportion of blue-collar potential voters
sharewhite	Proportion of white-collar potential voters
sharedomestic	Proportion of domestically employed potential voters
shareunemployed	Proportion of unemployed potential voters
nvoter	Number of eligible voters
nazivote	Number of votes for Nazis

The goal of analysis is to investigate which types of voters (based on their occupation category) cast ballots for the Nazis. One hypothesis says that the Nazis received much support from blue-collar workers. Since the data do not directly tell us how many blue-collar workers voted for the Nazis, we must infer this information using a statistical analysis with certain assumptions. Such an analysis where researchers try to infer individual behaviors from aggregate data is called ecological inference.

Question 1

1.1. We exploit the linear relationship between the Nazis' vote share and the proportion of blue-collar voters by regressing the former on the latter. That is, fit the following linear regression model:

$$E(nazishare_i|shareblue_i) = \alpha + \beta shareblue_i$$

Note: vote share is the number voters for a given candidate or party divided by the total number of voters.

```
nazi$nazivoteshare <- (nazi$nazivote/nazi$nvoter)
(nazi.test <- lm(nazivoteshare ~ shareblue, data = nazi))
```

```
##
## Call:
## lm(formula = nazivoteshare ~ shareblue, data = nazi)
##
## Coefficients:
## (Intercept)    shareblue
##     0.39558      0.06518
```

1.2. Compute the estimated slope coefficient, its standard error, and the 95% confidence interval.

```r
confint(nazi.test)
```

```
##                   2.5 %    97.5 %
## (Intercept)  0.36296607 0.4282031
## shareblue   -0.03730872 0.1676687
```

```r
summary(nazi.test)
```

```
##
## Call:
## lm(formula = nazivoteshare ~ shareblue, data = nazi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30151 -0.07133 -0.00092  0.06986  0.33037
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.39558    0.01661  23.812   <2e-16 ***
## shareblue    0.06518    0.05220   1.249    0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.108 on 679 degrees of freedom
## Multiple R-squared:  0.002291,   Adjusted R-squared:  0.0008218
## F-statistic: 1.559 on 1 and 679 DF,  p-value: 0.2122
```

```r
print("Beta = 0.07, (95% CI = [-0.04, 0.17])")
```

```
## [1] "Beta = 0.07, (95% CI = [-0.04, 0.17])"
```

1.3. Give a substantive interpretation of each quantity.

The estimated slope coefficient is .07. This means that the slope for the regression is closer to horizontal. Additionally, the 95% confidence interval does include zero, which in turn means that the p-value is greater than .05 and thus fail to reject the null hypothesis. No satistically significant relationship can be established between the nazi vote share and the proportion of blue collar workers.

Question 2

2.1. Based on the fitted regression model from the previous question, predict the average Nazi vote share Y_i given various proportions of blue-collar voters X_i. Specifically, predict the value of Y_i for each value of X_i from 0 to 1 by 0.01.

```r
blue.range <- seq(from = 0, to = 1, by = 0.01)
(blue.pred <- predict(nazi.test, interval = "confidence", newdata = data.frame(shareblue = blue.range))
```

```
##          fit       lwr       upr
```

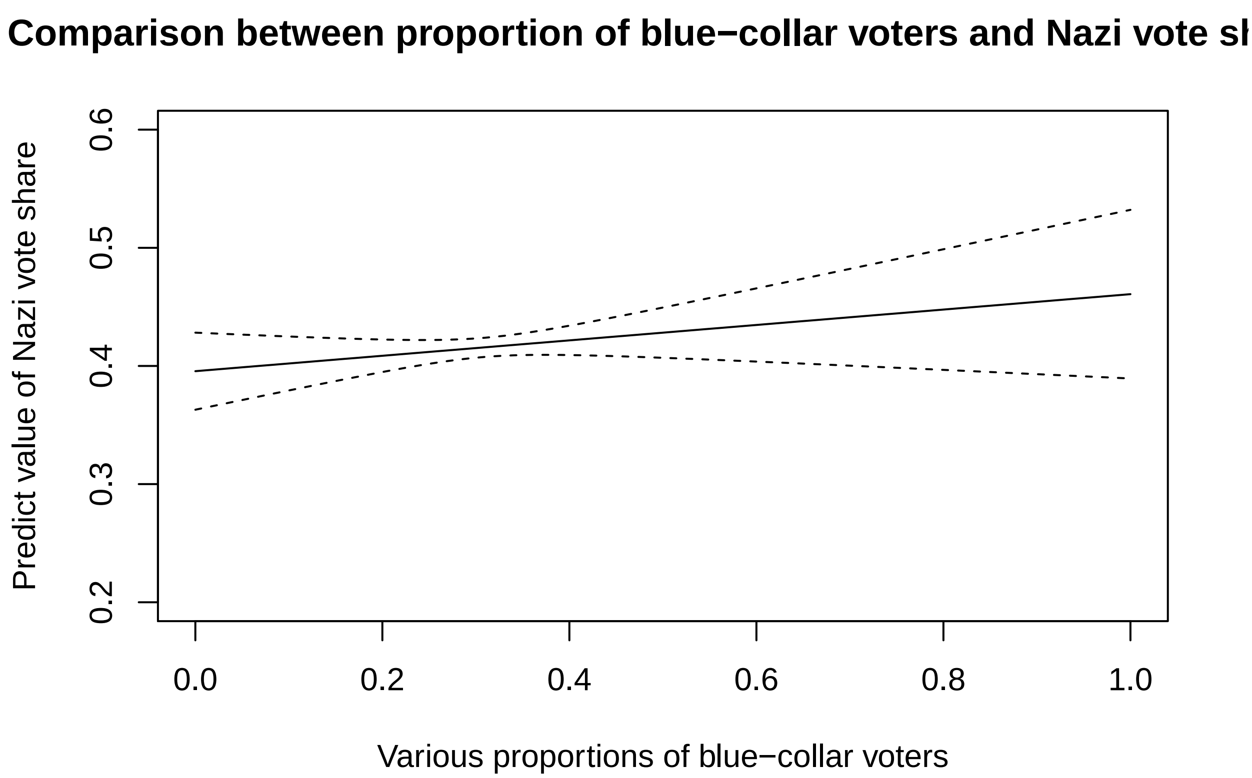
```
## 1   0.3955846 0.3629661 0.4282031
## 2   0.3962364 0.3646094 0.4278634
## 3   0.3968882 0.3662506 0.4275258
## 4   0.3975400 0.3678893 0.4271907
## 5   0.3981918 0.3695254 0.4268582
## 6   0.3988436 0.3711586 0.4265286
## 7   0.3994954 0.3727884 0.4262024
## 8   0.4001472 0.3744147 0.4258797
## 9   0.4007990 0.3760368 0.4255612
## 10  0.4014508 0.3776544 0.4252472
## 11  0.4021026 0.3792668 0.4249384
## 12  0.4027544 0.3808734 0.4246354
## 13  0.4034062 0.3824734 0.4243390
## 14  0.4040580 0.3840658 0.4240502
## 15  0.4047098 0.3856495 0.4237701
## 16  0.4053616 0.3872231 0.4235001
## 17  0.4060134 0.3887852 0.4232416
## 18  0.4066652 0.3903336 0.4229968
## 19  0.4073170 0.3918660 0.4227680
## 20  0.4079688 0.3933797 0.4225579
## 21  0.4086206 0.3948709 0.4223703
## 22  0.4092724 0.3963353 0.4222095
## 23  0.4099242 0.3977677 0.4220807
## 24  0.4105760 0.3991613 0.4219907
## 25  0.4112278 0.4005082 0.4219474
## 26  0.4118796 0.4017986 0.4219606
## 27  0.4125314 0.4030213 0.4220415
## 28  0.4131832 0.4041633 0.4222032
## 29  0.4138350 0.4052108 0.4224592
## 30  0.4144868 0.4061503 0.4228233
## 31  0.4151386 0.4069706 0.4233066
## 32  0.4157904 0.4076641 0.4239167
## 33  0.4164422 0.4082289 0.4246555
## 34  0.4170940 0.4086690 0.4255190
## 35  0.4177458 0.4089935 0.4264981
## 36  0.4183976 0.4092147 0.4275805
## 37  0.4190494 0.4093463 0.4287525
## 38  0.4197012 0.4094019 0.4300005
## 39  0.4203530 0.4093939 0.4313121
## 40  0.4210048 0.4093331 0.4326765
## 41  0.4216566 0.4092286 0.4340846
## 42  0.4223084 0.4090878 0.4355290
## 43  0.4229602 0.4089169 0.4370035
## 44  0.4236120 0.4087209 0.4385031
## 45  0.4242638 0.4085039 0.4400237
## 46  0.4249156 0.4082691 0.4415621
## 47  0.4255674 0.4080192 0.4431156
## 48  0.4262192 0.4077565 0.4446819
## 49  0.4268710 0.4074827 0.4462593
## 50  0.4275228 0.4071994 0.4478462
## 51  0.4281746 0.4069078 0.4494414
## 52  0.4288264 0.4066090 0.4510438
## 53  0.4294782 0.4063039 0.4526525
## 54  0.4301300 0.4059932 0.4542668
```

```
## 55  0.4307818 0.4056775 0.4558861
## 56  0.4314336 0.4053575 0.4575097
## 57  0.4320854 0.4050335 0.4591373
## 58  0.4327372 0.4047061 0.4607683
## 59  0.4333890 0.4043754 0.4624026
## 60  0.4340408 0.4040420 0.4640396
## 61  0.4346926 0.4037060 0.4656792
## 62  0.4353444 0.4033676 0.4673212
## 63  0.4359962 0.4030271 0.4689653
## 64  0.4366480 0.4026847 0.4706113
## 65  0.4372998 0.4023405 0.4722591
## 66  0.4379516 0.4019947 0.4739085
## 67  0.4386034 0.4016474 0.4755594
## 68  0.4392552 0.4012987 0.4772117
## 69  0.4399070 0.4009487 0.4788653
## 70  0.4405588 0.4005976 0.4805200
## 71  0.4412106 0.4002454 0.4821758
## 72  0.4418624 0.3998922 0.4838326
## 73  0.4425142 0.3995380 0.4854904
## 74  0.4431660 0.3991830 0.4871490
## 75  0.4438178 0.3988271 0.4888085
## 76  0.4444696 0.3984705 0.4904687
## 77  0.4451214 0.3981132 0.4921296
## 78  0.4457732 0.3977552 0.4937912
## 79  0.4464250 0.3973966 0.4954534
## 80  0.4470768 0.3970374 0.4971162
## 81  0.4477286 0.3966776 0.4987796
## 82  0.4483804 0.3963173 0.5004435
## 83  0.4490322 0.3959566 0.5021078
## 84  0.4496840 0.3955953 0.5037727
## 85  0.4503358 0.3952336 0.5054380
## 86  0.4509876 0.3948716 0.5071036
## 87  0.4516394 0.3945091 0.5087697
## 88  0.4522912 0.3941462 0.5104362
## 89  0.4529430 0.3937830 0.5121030
## 90  0.4535948 0.3934195 0.5137701
## 91  0.4542466 0.3930556 0.5154376
## 92  0.4548984 0.3926915 0.5171053
## 93  0.4555502 0.3923270 0.5187734
## 94  0.4562020 0.3919623 0.5204417
## 95  0.4568538 0.3915973 0.5221103
## 96  0.4575056 0.3912321 0.5237791
## 97  0.4581574 0.3908666 0.5254482
## 98  0.4588092 0.3905009 0.5271175
## 99  0.4594610 0.3901350 0.5287870
## 100 0.4601128 0.3897688 0.5304568
## 101 0.4607646 0.3894025 0.5321267
```

2.2. Plot the predicted value of Nazi votes against various values of blue-collar share within its observed range (the horizontal axis) as a solid line. Add 95% confidence intervals as dashed lines.

```
plot(blue.range, blue.pred[, "fit"], xlab = "Various proportions of blue-collar voters",
    ylab = "Predict value of Nazi vote share", main = "Comparison between proportion of blue-collar vote
    type = "l", ylim = c(0.2, 0.6))
abline(h = 0, )
```

```r
lines(x = blue.range, y = blue.pred[, 2], lty = "dashed")
lines(x = blue.range, y = blue.pred[, 3], lty = "dashed")
```



2.3. Give a substantive interpretation of the plot.

The slope is positive, meaning that as proportions of blue-collar voters increases, the predicted value of nazi-vote share increases. The line is also between the bounds of the 95% confidence interval, and approaches it most closely at a value between 0.2 and 0.4

Question 3

3.1. Fit a linear regression model where the overall Nazi vote share (`nazishare`) is regressed on the the occupation variables. The model should contain no intercept and five predictors (`shareself`,`shareblue`, `sharewhite`, `sharedomestic`, and `shareunemployed`), each representing the proportion of potential voters in a certain occupation type. Present the estimated coefficients, the standard errors, and the 95% confidence intervals for each of the predictors.

```r
(testnazi <- lm(nazivoteshare ~ -1 + shareself + shareblue +
    sharewhite + sharedomestic + shareunemployed, data = nazi))
```

```
##
## Call:
## lm(formula = nazivoteshare ~ -1 + shareself + shareblue + sharewhite +
##     sharedomestic + shareunemployed, data = nazi)
##
## Coefficients:
```

```
##       shareself        shareblue       sharewhite    sharedomestic 
##         1.11426          0.54038          0.28509          0.05221 
## shareunemployed 
##        -0.02816
```

```r
confint(testnazi)
```

```
##                      2.5 %    97.5 %
## shareself        0.7868165 1.4417096
## shareblue        0.4648233 0.6159423
## sharewhite       0.1378070 0.4323667
## sharedomestic   -0.1268549 0.2312741
## shareunemployed -0.1658852 0.1095646
```

```r
summary(testnazi)
```

```
## 
## Call:
## lm(formula = nazivoteshare ~ -1 + shareself + shareblue + sharewhite + 
##     sharedomestic + shareunemployed, data = nazi)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.28271 -0.06847 -0.00055  0.06790  0.32369 
## 
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)    
## shareself        1.11426    0.16677   6.681 4.95e-11 ***
## shareblue        0.54038    0.03848  14.042  < 2e-16 ***
## sharewhite       0.28509    0.07501   3.801 0.000157 ***
## sharedomestic    0.05221    0.09120   0.572 0.567181    
## shareunemployed -0.02816    0.07014  -0.401 0.688202    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1024 on 676 degrees of freedom
## Multiple R-squared:  0.9435, Adjusted R-squared:  0.9431 
## F-statistic:  2259 on 5 and 676 DF,  p-value: < 2.2e-16
```

3.2. Interpret the estimate of each coefficient and its 95% confidence interval. What is the model's estimate of Nazi vote share when 100% of a precinct is self-employed? What is the estimate when 100% of a precinct is unemployed? Explain these result.

shareself: Beta = 1.11, (95% CI = [0.79, 1.44]) shareblue: Beta = 0.54, (95% CI = [0.46, 0.62]) sharewhite: Beta = 0.29, (95% CI = [0.14, 0.43]) sharedomestic: Beta = 0.05 (95% CI = [-0.13, 0.23]) shareunemployed: Beta = -0.03 (95% CI = [-0.17, 0.11])

The self-employed, blue-collar, and white-colalr variables are all statistically signficiant, meaning that we can conclude that a relationship exists between these variables in a precinct and the Nazi vote share, while the last two variables are not statistically significant and thus we fail to reject the null hypothesis. The model's estimate of Nazi vote share is actually 1.11 when 100% of a precinct is self-employed, which techincally doesn't make sense but goes to show how important of a variable it is.

Question 4

4.1. Finally, we consider a model-free approach to ecological inference. That is, we ask how much we can learn from the data alone without making an additional modeling assumption. For each precinct, obtain the minimum number of blue-collar Nazi voters that is logically possible given a scenario in which all non-blue-collar voters in precincts vote for the Nazis. Then sum this minimum number over all precincts and present the result. This is the minimum number of blue-collar Nazi votes in the election.

```
min.nazi <- (nazi$nazivoteshare + nazi$shareblue - 1)/nazi$shareblue
min.nazi[min.nazi < 0] <- 0

blue.votes <- nazi$shareblue * nazi$nvoter
min.nazi.votes <- blue.votes * min.nazi

(tot.min.nazi.votes <- sum(min.nazi.votes))
```

```
## [1] 3738.263
```

4.2. Similarly, what is the largest possible value for blue-collar Nazi voters? Calculate these bounds, keeping in mind that the value for blue-collar Nazi voters cannot be negative or greater than 1.

```
max.nazi <- nazi$nazivoteshare/nazi$shareblue

max.nazi[max.nazi < 0] <- 0
max.nazi[max.nazi > 1] <- 1

max.nazi.votes <- max.nazi * blue.votes

(tot.max.nazi.votes <- sum(max.nazi.votes))
```

```
## [1] 13266736
```

4.3. Present these numbers as a proportion of blue-collar voters. Give a brief substantive interpretation of the results. Does this method of finding the min and max Nazi vote share provide much additional information about the Nazi support among blue-collar workers?

```
tot.min.nazi.votes/sum(blue.votes)
```

```
## [1] 0.0002675665
```

```
tot.max.nazi.votes/sum(blue.votes)
```

```
## [1] 0.9495677
```

```
print("Because the range of the vote share is so wide, this isn’t really helpful in providing additional
```

```
## [1] "Because the range of the vote share is so wide, this isn’t really helpful in providing additiona
```

Question 5

5.1. What would happen if all non-blue-collar citizens opposed the Nazi’s? Would it have been possible for the Nazis to have won the election? Use the actual threshold in the election: 37.7% Nazi voteshare. Assume everyone voted.

```
(total.blue <- sum(nazi$shareblue * nazi$nvoter))
```

```
## [1] 13971343
```

```r
total.blue/sum(nazi$nvoter)
```

```
## [1] 0.3150558
```

```r
print("If all non-blue-collar citizens opposed the Nazi’s, they only would’ve gotten 32% of the votes, 
```

```
## [1] "If all non-blue-collar citizens opposed the Nazi’s, they only would’ve gotten 32% of the votes,
```

5.2. How concentrated were Nazi voters? Identify the precincts with a population of Nazi voters in the top decile. If these votes are removed would the Nazis have still won the election?

```r
quantile(nazi$nazivote, probs = 0.9)
```

```
##   90% 
## 50241
```

```r
cutnazi <- nazi[nazi$nazivote <= 50241, ]

sum(cutnazi$nazivote)/sum(cutnazi$nvoter)
```

```
## [1] 0.4090871
```

5.3. What do these results suggest about the magnitude of the Nazi victory?

The magnitude was pretty big, because they still carried the vote (above 37.7% of the votes) even in the last decile of precincts of nazi share.