The Price of Prejudice

Replication by: Pawonee Khadka, University of Alabama

11/28/2022

The paper is based on a field experiment that goes on to investigate ethnic prejudice in the workplace. The authors want to see how potential discriminators respond to changes in the cost of discrimination.

```
library(dplyr)
library(tidyverse)
library(haven)
```

All analysis for the original paper was done using STATA and all data and necessary code instructions was made publicly available. With my replication work, I didn't have much to do with the raw data, but jumped into replication right away. Following is the chuck for replication of Table 2: Team Production Function.

```
Price <- read_dta("Price_of_Prejudice_Stata_data (1).dta") %>%
  mutate( productivity = case_when(ethnicity == 3 | couple == 4 ~ 0, TRUE ~ 1)) %>%
  mutate (LnProd_1 = log(prod_1),
          LnProd_2 = log(own_prod_2),
          muslim_team = 0)
Price_1 <- Price %>%
  mutate(Muslim_sounding_team = case_when(ethnicity == 2 & couple == 1 ~ 1,
                                          TRUE ~ 0))

Price_2 <- Price_1 %>%
  mutate(Danish_sounding_team = case_when(ethnicity == 1 & couple == 1 ~ 1,
                                          TRUE ~ 0))

Price_2 <- Price_2 %>%
  mutate(Alone_2nd_round = case_when(couple == 3 ~ 1, TRUE ~ 0))

Price_2 <- Price_2 %>%
  mutate(team = 1)%>%
  mutate(team = case_when(Danish_sounding_team == 1 ~ 2, Muslim_sounding_team
                          ==1 ~ 3, Alone_2nd_round == 1 ~ 4))

Price_2 <- Price_2 %>%
  mutate(Team_type = case_when(team == 1 ~ "Heterogeneous", team ==2 ~ "Danish",
                               team == 3 ~ "Muslim", team == 4 ~"Alone"))
```

Dummies:

```
Price_2 <- Price_2 %>%
  mutate(danish = case_when(ethnicity == 1 ~ 1, TRUE ~ 0))%>%
  mutate(lnprodpartnertemp = case_when(Alone_2nd_round == 0 ~ log(prod_partner),
```

```
                                        TRUE ~ 0 ),
         decision_maker = case_when(type ==1 & couple != 3 ~ 1, TRUE ~0))
# Here, danish 0 "Muslim-sounding" 1 "Danish-sounding"
#temporary partner variable
```

Interaction -between being a decision-maker and being in a heterogeneous team

```
Price_2 <- Price_2 %>%
  mutate(decision_maker_mixed = case_when(muslim_team == 0 &
                                          Danish_sounding_team == 0 &
                                          Alone_2nd_round == 0 ~
                                          decision_maker, TRUE ~ 0),
         lnprod1alone = LnProd_1 * Alone_2nd_round)
#name first one "D-M in heterogeneous team" and second lnprod1alone "ln(Prod_1) * alone"
```

Estimate production function

```
Price_3 <- Price_2 %>%
  filter(productivity == 1)

A <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male, data = Price_3)
# For vce(robust) :
library(sandwich)
library(lmtest)
AA <- vcovHC(A, type = "HC1")
robust_AA <- sqrt(diag(AA))

B <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+ decision_maker,
        data = Price_3)
BB <- vcovHC(B, type = "HC1")
robust_BB <- sqrt(diag(BB))

C <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+
          Danish_sounding_team + Muslim_sounding_team + Alone_2nd_round,
        data = Price_3)
CC <- vcovHC(C, type = "HC1")
robust_CC <- sqrt(diag(CC))

D <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male+
          Danish_sounding_team + Muslim_sounding_team + Alone_2nd_round+
          decision_maker + decision_maker_mixed, data = Price_3)
DD <- vcovHC(D, type = "HC1")
robust_se <- sqrt(diag(DD))

library (stargazer)
stargazer(A,B,C,D, type ="text",
          title = "Table 2:Team Production Function",
          dep.var.labels = c("Dependent Variable: ln(prod2)"),
          notes.label ="Significance Levels",
          covariate.labels = c("logprod1i","logprod1j","logprod1i * Alone","Male","Danish","Muslim","Al
                               Heterogeneous"),
```

```
          no.space = TRUE,
          se = list(robust_AA,robust_BB,robust_CC,robust_se),
          keep.stat = c("adj.rsq","n")

          )
```

Table 2:Team Production Function

	Dependent variable:			
	Dependent Variable: ln(prod2)			
	(1)	(2)	(3)	(4)
logprod1i	0.416***	0.413***	0.385***	0.387***
	(0.066)	(0.066)	(0.052)	(0.053)
logprod1j	0.416***	0.421***	0.385***	0.387***
	(0.066)	(0.067)	(0.052)	(0.053)
logprod1i * Alone	0.398***	0.401***	0.444	0.443
	(0.064)	(0.065)	(0.285)	(0.287)
Male	0.027	0.026	0.020	0.020
	(0.027)	(0.027)	(0.033)	(0.033)
Danish		-0.008		-0.028
		(0.024)		(0.039)
Muslim				0.038
				(0.047)
Alone			0.047*	0.064**
			(0.025)	(0.032)
Decision maker			-0.024	-0.025
			(0.035)	(0.035)
Heterogeneous			-0.354	-0.335
			(1.411)	(1.422)
Constant	0.849	0.848	1.135***	1.117***
	(0.517)	(0.519)	(0.369)	(0.373)
Observations	147	147	147	147
Adjusted R2	0.465	0.461	0.463	0.456
Significance Levels	*p<0.1; **p<0.05; ***p<0.01			

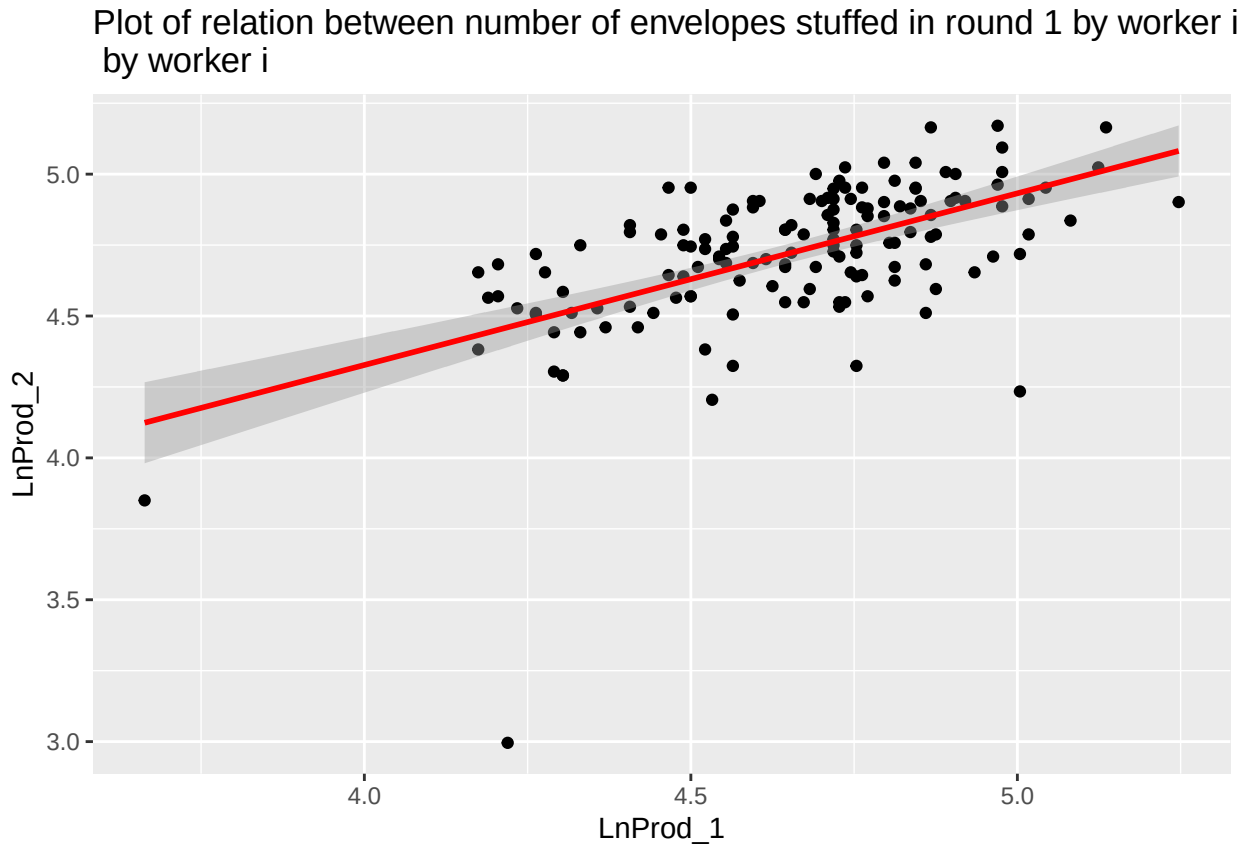
In this table: *the Dependent variable is the log of the number of envelopes stuffed in round 2 by worker i* prod1i is the number of envelopes stuffed in round 1 by worker i *prod1j is the number of envelopes stuffed by i's coworker in round 2* Alone is a dummy set to 1 if worker i works alone in round 2 *Male is worker i's gender* Decision maker indicates if worker i makes a choice of coworker *The remaining dummies characterize team composition in round 2.

- Two main observations can be made here:
- The table replication is not exact. This is because of tiny differences between softwares. I have used the stargazer package which allows for professional tables, yet some differences remain as compared to original table.
- There is a key difference to mark here. The number of observations, in this replication is 147 unlike 140 in the paper. This has led to some minor chane in statistic. The code was explained fairly well, however, authors did not provide specifications on filtering the data. My assumption is that the error is stemming from filtering differences by "Or" vs "And"

A few graphs to show the effect of independent variables on productivity, i.e. the number of envelopes stuffed in round 2 by worker i

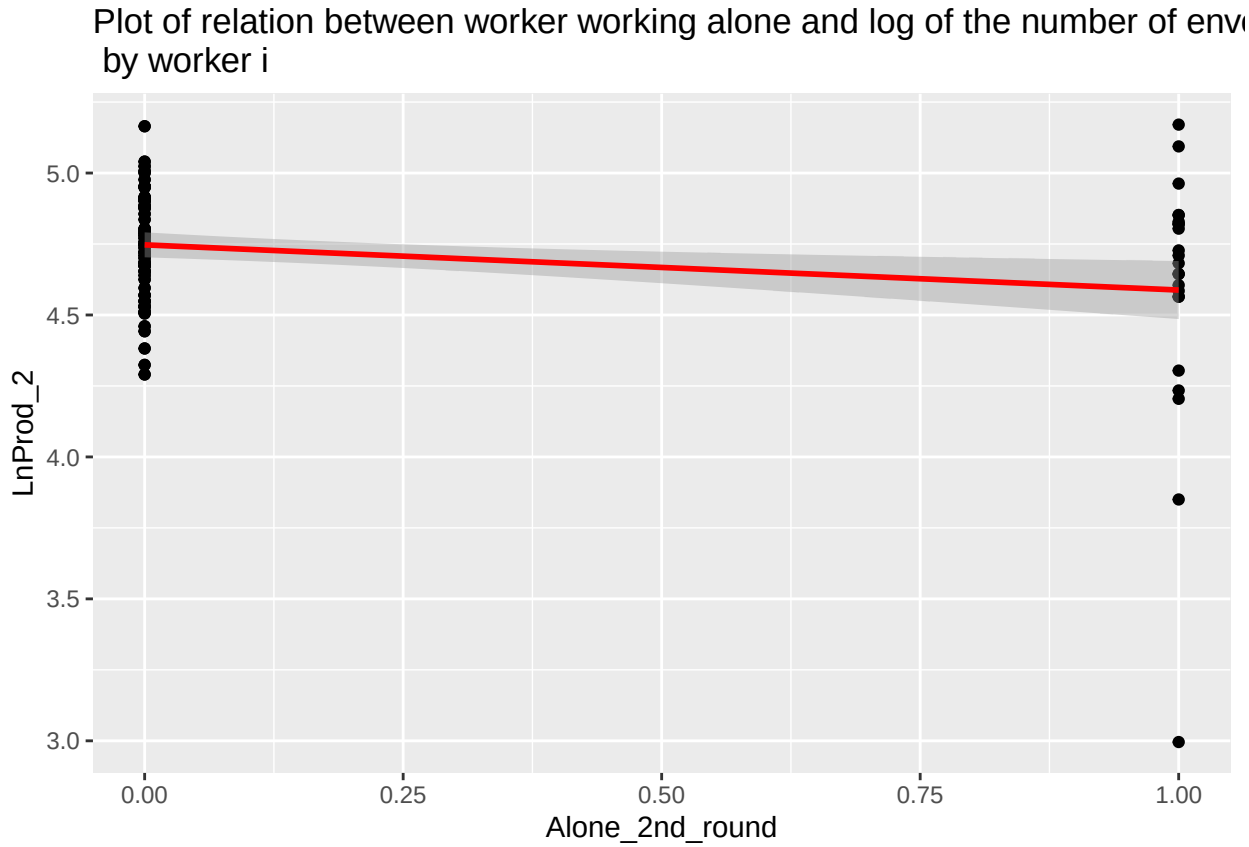
```
library(ggplot2)
ggplot(Price_3, aes(x=LnProd_1, y=LnProd_2))+
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
   ggtitle("Plot of relation between number of envelopes stuffed in round 1 by worker i and log of the n
 by worker i")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



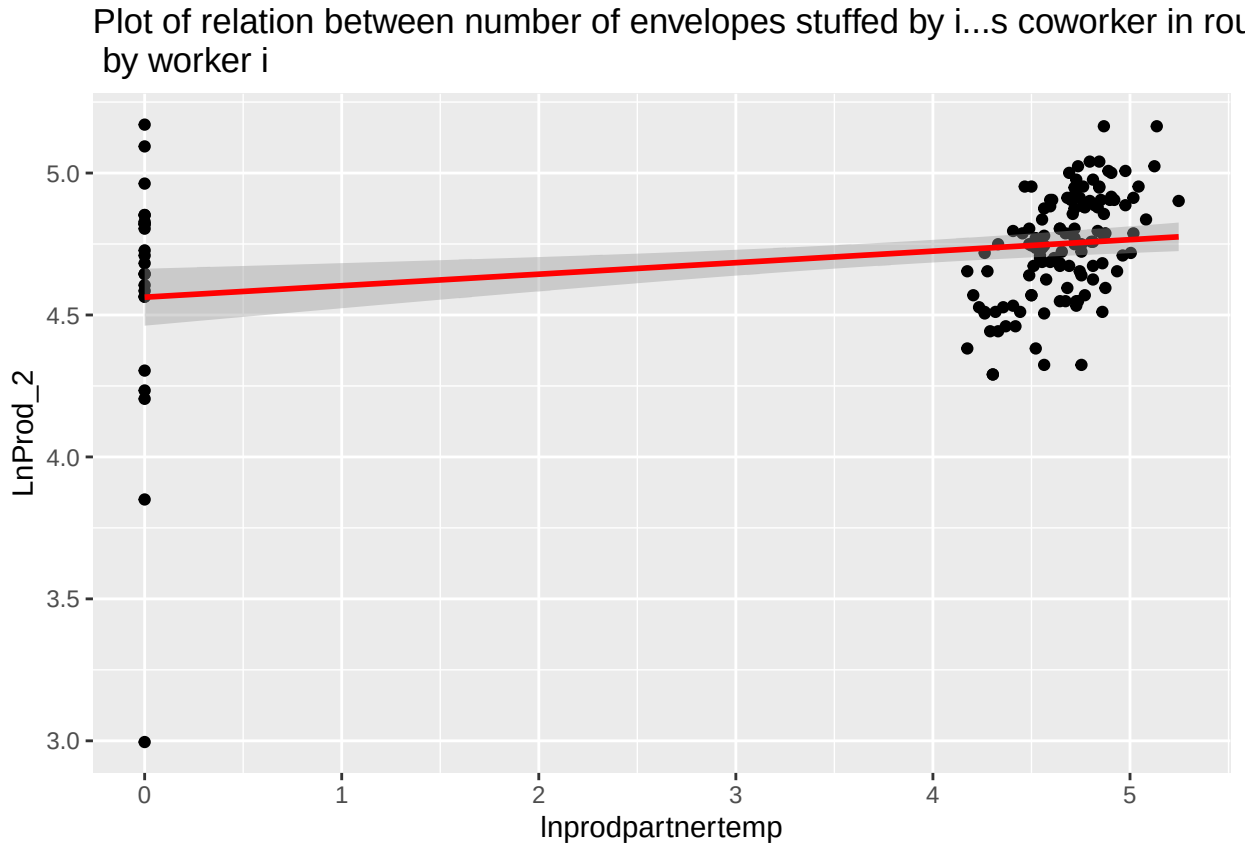

```
library(ggplot2)
ggplot(Price_3, aes(x=Alone_2nd_round, y=LnProd_2))+
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
   ggtitle("Plot of relation between worker working alone and log of the number of envelopes stuffed in
 by worker i")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
library(ggplot2)
ggplot(Price_3, aes(x=lnprodpartnertemp, y=LnProd_2))+
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
  ggtitle("Plot of relation between number of envelopes stuffed by i's coworker in round 2 and log of t
 by worker i")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



The paper moves on two second part where they create a new variable- price which gives the cost of discrimination. Remaining is an attempt to replicate another existing table:

Estimating the cost of discrimination

Part 2:

```
Price_4 <- Price_3 %>%
  filter(type == 1)
Price_4 <- Price_4 %>%
  filter(ethnicity !=3 | type_day_1 != 9999 | type_day_2 != 9999)
Price_4 <- Price_4 %>%
  filter(info ==1)

Price_5 <- Price_4 %>%
  mutate(alone = 0,
         lnprod1alone = 0,
         lnprodpartnertemp = log(prod_own))

m <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male, data = Price_5)
mm <- vcovHC(m, type = "HC1")
robust_m <- sqrt(diag(mm))
n <- predict(m)
nn <- exp(n)
```

```
Price_6 <- Price_4 %>%
  mutate(
         lnprodpartnertemp = log(prod_other))

o <- lm(LnProd_2 ~ LnProd_1+lnprodpartnertemp+lnprod1alone+male, data = Price_6)
oo <- vcovHC(o, type = "HC1")
robust_o <- sqrt(diag(oo))
p <- predict(o)
pp <- exp(p)
```

Additional Contribution to Paper Extension

In the existing dataset, there is a variable discrimination. The second table is a probit based regression on how the cost varies based on ethnicity and checks its impact on discrimination. I am particularly interested to see if there is a link between just gender and Danish_sounding vs Muslim sounding on discrimination:

```
library(stats)
model_pr <- glm(discr~ Danish_sounding_team+male+LnProd_1, family = binomial(link = "probit"), data = Pr
pr <- vcovHC(model_pr, type = "HC1")
robust_pr <- sqrt(diag(pr))

model_pr1 <- glm(discr~ Muslim_sounding_team+male+LnProd_1, family = binomial(link = "probit"), data = P
pr1 <- vcovHC(model_pr1, type = "HC1")
robust_pr1 <- sqrt(diag(pr1))

stargazer(model_pr,model_pr1, type = "text",
          keep.stat = c("adj.rsq","n"),
          se = list(robust_AA,robust_BB,robust_CC,robust_se))
```

```
================================================
                        Dependent variable:
                    ----------------------------
                               discr
                          (1)            (2)
------------------------------------------------
Danish_sounding_team    2.250***


Muslim_sounding_team                    6.224


male                   -0.623***      -0.472***
                        (0.027)        (0.027)

LnProd_1               -1.429***      -0.407***
                        (0.066)        (0.066)

Constant                6.291***      1.456***
                        (0.517)        (0.519)

------------------------------------------------
```

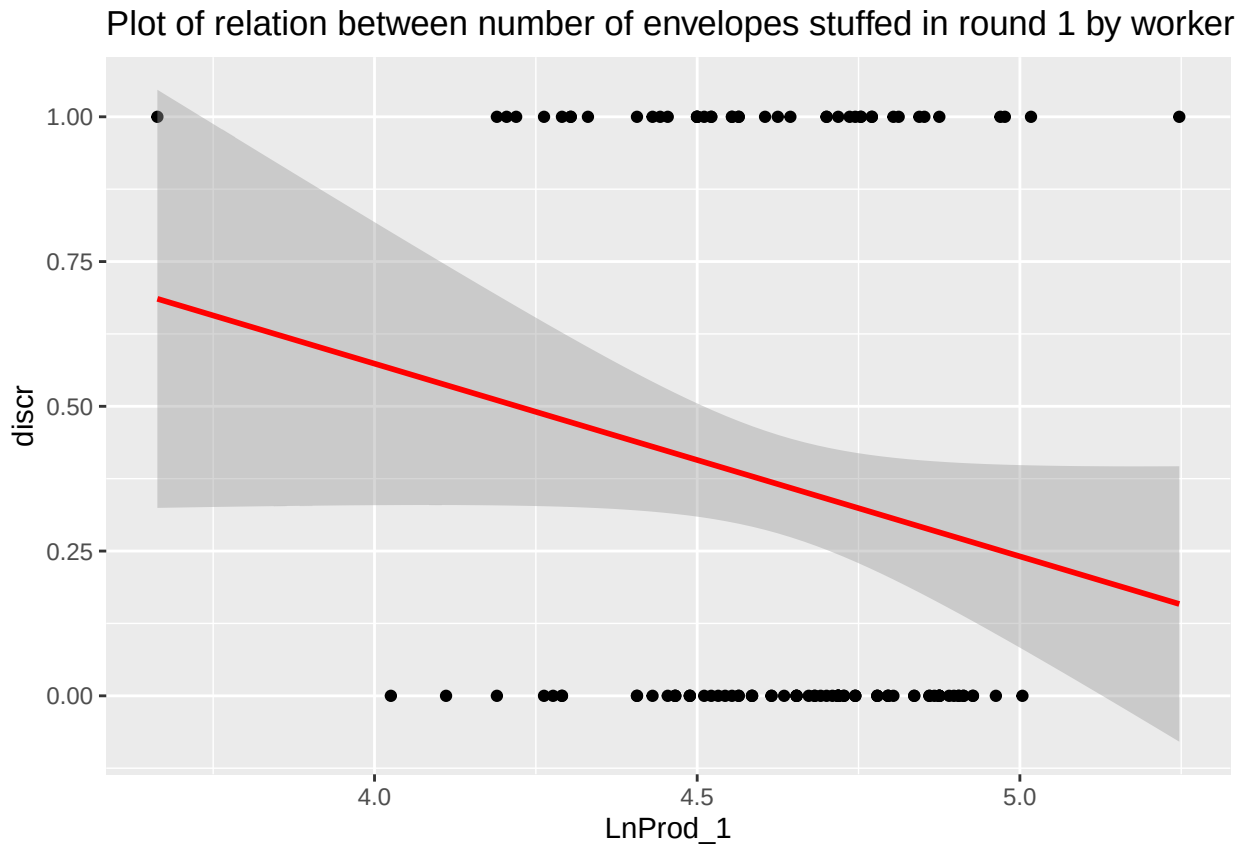
```
Observations              124            124
================================================
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

The paper finds an important result that cost of discrimination is directly proportional to probability of distribution. I didn't replicate the cost portion. * But even without that the paper is questionable if we refer to my added regression table. There is no impact on discrimination from muslim sounding name. * This further brings to question whether the measure was built correctly. It could have been crucial that the researchers made the field question available so the correct interpretation could be taken. * It is interesting to notice how productivity and gender both have a negative relation with workplace discrimination. Following graphs reflect it:

```
library(ggplot2)
ggplot(Price_3, aes(x=LnProd_1, y=discr))+
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
   ggtitle("Plot of relation between number of envelopes stuffed in round 1 by worker i and discriminati
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```

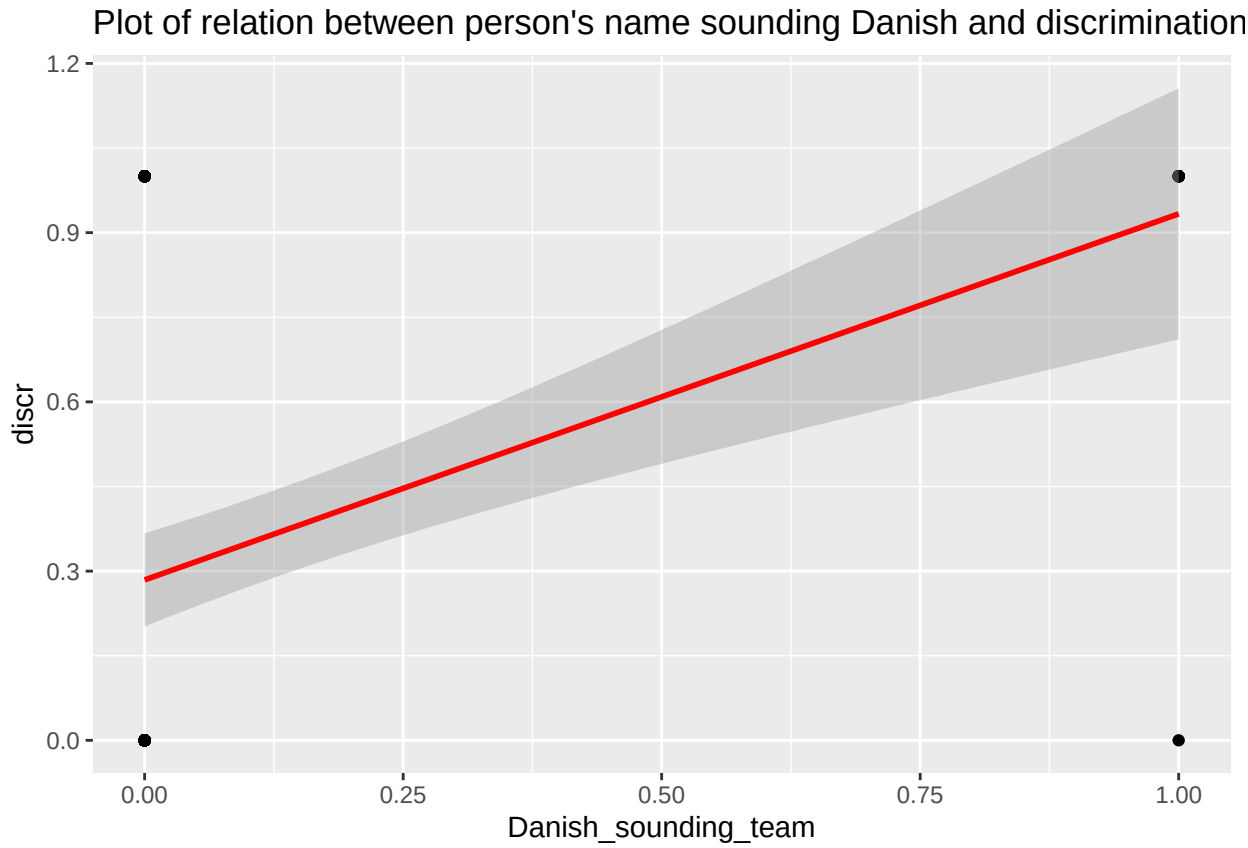
```
## `geom_smooth()` using formula 'y ~ x'
```




```
library(ggplot2)
ggplot(Price_3, aes(x=Danish_sounding_team, y=discr))+
```

```
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
   ggtitle("Plot of relation between person's name sounding Danish and discrimination in workplace")
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
## `geom_smooth()` using formula 'y ~ x'
```

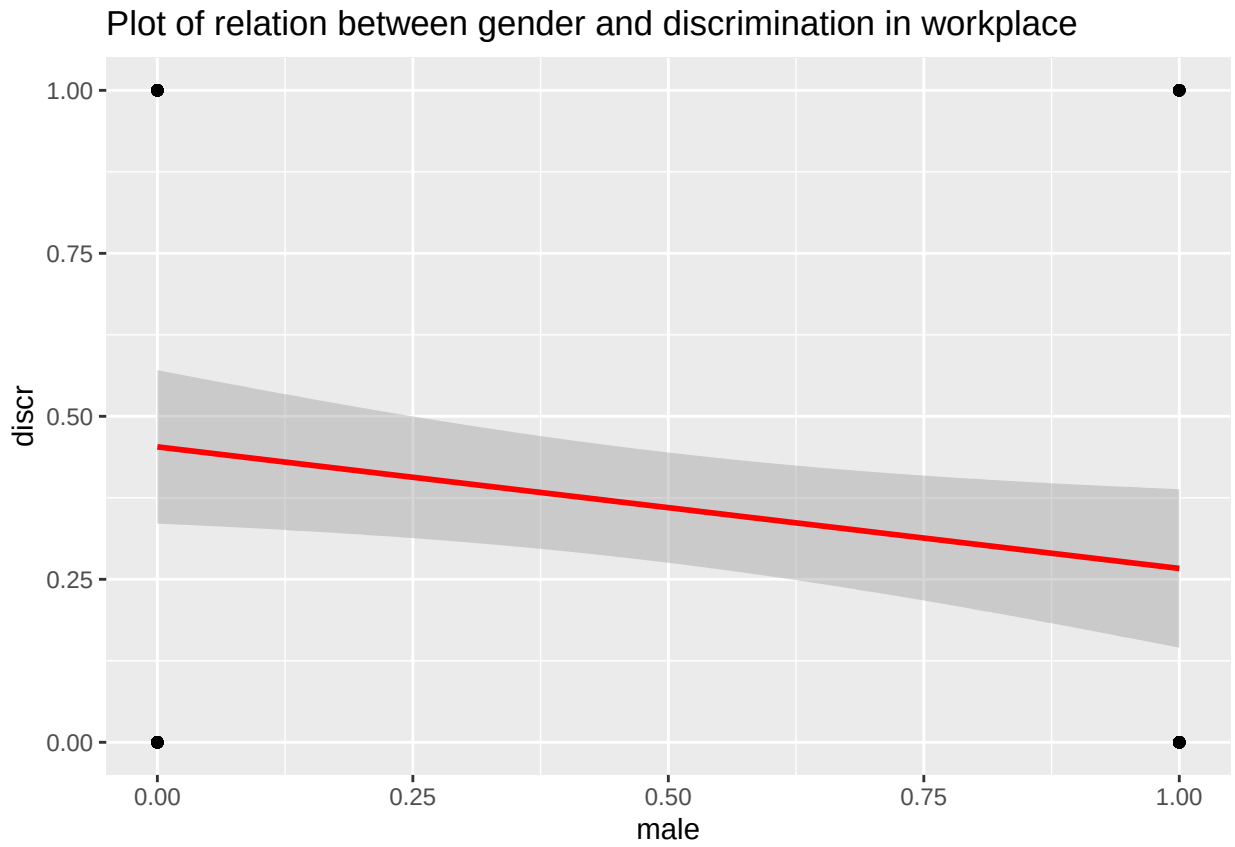



```
library(ggplot2)
ggplot(Price_3, aes(x=male, y=discr))+
  geom_point()+
  stat_smooth(method = "lm", col = "red")+
   ggtitle("Plot of relation between gender and discrimination in workplace")
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau

## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau

## `geom_smooth()` using formula 'y ~ x'
```


Plot of relation between gender and discrimination in workplace
discr
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
male