Sevilay Sezer Barutcu

6th March 2022

**INTL 601 - HOMEWORK 1**

**Gauss-Markov assumptions**

- The model is linear in the parameters

- No endogeneity in the model (independent variable $X$ and $\epsilon$ are not correlated)

- Errors are normally distributed with constant variance

- No autocorrelation in the errors

- No multicollinearity between variable

**Linearity**

- The relationship between the predictor (x) and the outcome (y) is assumed to be linear

- Non-linearity of the outcome - predictor relationships

- Model plots:

-- Residuals vs Fitted. Used to check the linear relationship assumptions. A horizontal line, without distinct patterns is an indication for a linear relationship, what is good.

-- Normal Q-Q. Used to examine whether the residuals are normally distributed. Itâ€™s good if residuals points follow the straight dashed line.

-- Scale-Location (or Spread-Location). Used to check the homogeneity of variance of the residuals (homoscedasticity). Horizontal line with equally spread points is a good indication of homoscedasticity.

-- Residuals vs Leverage. Used to identify influential cases, that is extreme values that might influence the regression results when included or excluded from the analysis.

In my analysis I found correlation between control variable political orientation and Pragmatic Legitimacy1 so the endogeneity assumption is violated.

I did OLS analysis between PragmatticLeg1 and MoralLeg1. Both of them are independent variable. I should have done it with dependent variable as well but it ewould be too long for this report.

**Variables**

**Dependent Variables :** PragmatticLeg2,PragmatticLeg3,MoralLeg1,MoralLeg2,MoralLeg3,MoralLeg4,CogLeg1,CogLeg2, Gender"

**Independent Variable:**

WillCoProduce1, WillCoProduce2, WillCoProduce3 WillCoProduce4

**Control Variable:**

Age, Employment, Education, Political Orientation, Public Service Motivation

**Console Data**

> data = read.csv("C:/Users/sevil/Desktop/INTL 601/HW1/SEVİLAY/willingness to coproduce/CoProductionDataEnglish.csv")

> #Lets explore the data

> colnames(data)

[1] "interview" "ds.RD02" "ds.RD03" "PragmatticLeg1"

[5] "PragmatticLeg2" "PragmatticLeg3" "MoralLeg1" "MoralLeg2"

[9] "MoralLeg3" "MoralLeg4" "CogLeg1" "CogLeg2"

[13] "CogLeg3" "WillCoProduce1" "WillCoProduce2" "WillCoProduce3"

[17] "OrgLeg" "ds.DE01" "PoliticOrient" "PubServMot1"

[21] "PubServMot2" "PubServMot3" "PubServMot4" "Employment"

[25] "Education" "Age" "ds.DE15"

> head()

Error in checkHT(n, dx <- dim(x)) :

argument "x" is missing, with no default

> head(data)

interview ds.RD02 ds.RD03 PragmatticLeg1 PragmatticLeg2 PragmatticLeg3

1 1 1 2 4 4 4

2 2 2 1 2 2 3

3 3 1 3 3 3 3

4 4 2 3 3 3 3

5 5 1 2 3 4 4

6 6 2 1 3 3 3

MoralLeg1 MoralLeg2 MoralLeg3 MoralLeg4 CogLeg1 CogLeg2 CogLeg3

1 5 5 5 5 4 4 5

2 3 3 3 3 3 3 1

3 2 3 3 3 4 4 2

4 2 4 5 5 5 5 5

5 4 4 4 4 5 5 4

6 3 3 3 3 3 3 3

WillCoProduce1 WillCoProduce2 WillCoProduce3 OrgLeg ds.DE01

1 4 2 1 1 out of 4 Female

2 1 1 1 2 out of 4 Female

3 3 3 3 4 out of 4 Male

4 5 5 5 2 out of 4 Male

5 2 2 3 2 out of 4 Female

6 3 3 3 1 out of 4 Female

PoliticOrient PubServMot1 PubServMot2 PubServMot3 PubServMot4

1 [6] 4 4 4 4

2 [6] 3 1 2 2

3 [5] 3 4 3 4

4 [6] 5 5 5 5

5 [4] 3 3 3 4

6 [6] 3 3 3 3

Employment

1 Employed in a private company.

2 Not working

3 Employed in a private company.

4 Employed in a private company.

5 Employed in a non-profit organization.

6 Employed in the public service.

Education

1 Technical college entrance qualification

2 Main/elementary school certificate

3 Abitur (general or subject-related higher education entrance qualification)

4 Main/elementary school certificate

5 Abitur (general or subject-related higher education entrance qualification)

6 Secondary school certificate, intermediate school leaving certificate or equivalent

Age ds.DE15

1 45 Yes

2 37 Yes

3 31 Yes

4 55 Yes

5 63 Yes

6 32 Yes

> summary(data)

interview ds.RD02 ds.RD03 PragmatticLeg1

Min. : 1.0 Min. :1.000 Min. :1.000 Min. :1.000

1st Qu.: 250.8 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:3.000

Median : 500.5 Median :2.000 Median :2.000 Median :3.000

Mean : 500.5 Mean :1.513 Mean :2.008 Mean :3.412

3rd Qu.: 750.2 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:4.000

Max. :1000.0 Max. :2.000 Max. :3.000 Max. :5.000

PragmatticLeg2 PragmatticLeg3 MoralLeg1 MoralLeg2

Min. :1.000 Min. :1.000 Min. :1.0 Min. :1.000

1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.0 1st Qu.:3.000

Median :3.000 Median :3.000 Median :4.0 Median :4.000

Mean :3.392 Mean :3.398 Mean :3.7 Mean :3.757

3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.0 3rd Qu.:4.000

Max. :5.000 Max. :5.000 Max. :5.0 Max. :5.000

MoralLeg3 MoralLeg4 CogLeg1 CogLeg2

Min. :1.000 Min. :1.000 Min. :1.00 Min. :1.000

1st Qu.:3.000 1st Qu.:3.000 1st Qu.:4.00 1st Qu.:4.000

Median :4.000 Median :3.000 Median :4.00 Median :4.000

Mean :3.849 Mean :3.376 Mean :4.26 Mean :4.135

3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:5.00 3rd Qu.:5.000

Max. :5.000 Max. :5.000 Max. :5.00 Max. :5.000

CogLeg3 WillCoProduce1 WillCoProduce2 WillCoProduce3

Min. :1.000 Min. :1.000 Min. :1.000 Min. :1.000

1st Qu.:3.000 1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000

Median :4.000 Median :3.000 Median :2.000 Median :2.000

Mean :3.751 Mean :2.677 Mean :2.421 Mean :2.393

3rd Qu.:5.000 3rd Qu.:4.000 3rd Qu.:3.000 3rd Qu.:3.000

Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000

OrgLeg ds.DE01 PoliticOrient PubServMot1

Length:1000 Length:1000 Length:1000 Min. :1.00

Class :character Class :character Class :character 1st Qu.:3.00

Mode :character Mode :character Mode :character Median :4.00

Mean :3.67

3rd Qu.:4.00

Max. :5.00

PubServMot2 PubServMot3 PubServMot4 Employment

Min. :1.000 Min. :1.000 Min. :1.000 Length:1000

1st Qu.:3.000 1st Qu.:3.000 1st Qu.:3.000 Class :character

Median :4.000 Median :4.000 Median :4.000 Mode :character

Mean :3.702 Mean :3.679 Mean :3.675

3rd Qu.:4.000 3rd Qu.:4.000 3rd Qu.:4.000

Max. :5.000 Max. :5.000 Max. :5.000

Education Age ds.DE15

Length:1000 Min. :18.00 Length:1000

Class :character 1st Qu.:29.00 Class :character

Mode :character Median :45.00 Mode :character

Mean :43.45

3rd Qu.:56.00

Max. :69.00

> mod = lm( PragmatticLeg1 ~ MoralLeg1, data = data )

> summary(mod)

Call:

lm(formula = PragmatticLeg1 ~ MoralLeg1, data = data)

Residuals:

Min 1Q Median 3Q Max

-2.9071 -0.5263 -0.1454 0.4737 2.6163

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.00279 0.10731 18.66 <2e-16 \*\*\*

MoralLeg1 0.38087 0.02821 13.50 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7898 on 998 degrees of freedom

Multiple R-squared: 0.1545, Adjusted R-squared: 0.1536

F-statistic: 182.3 on 1 and 998 DF, p-value: < 2.2e-16

> library(stargazer) #to make LatexTAble

Please cite as:

Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

> stargezer(mod)

Error in stargezer(mod) : could not find function "stargezer"

> stargazer(mod)

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com

% Date and time: Paz, Mar 06, 2022 - 20:25:27

\begin{table}[!htbp] \centering

\caption{}

\label{}

\begin{tabular}{@{\extracolsep{5pt}}lc}

\\[-1.8ex]\hline

\hline \\[-1.8ex]

& \multicolumn{1}{c}{\textit{Dependent variable:}} \\

\cline{2-2}

\\[-1.8ex] & PragmatticLeg1 \\

\hline \\[-1.8ex]

MoralLeg1 & 0.381$^{\*\*\*}$ \\

& (0.028) \\

& \\

Constant & 2.003$^{\*\*\*}$ \\

& (0.107) \\

& \\

\hline \\[-1.8ex]

Observations & 1,000 \\

R$^{2}$ & 0.154 \\

Adjusted R$^{2}$ & 0.154 \\

Residual Std. Error & 0.790 (df = 998) \\

F Statistic & 182.320$^{\*\*\*}$ (df = 1; 998) \\

\hline

\hline \\[-1.8ex]

\textit{Note:} & \multicolumn{1}{r}{$^{\*}$p$<$0.1; $^{\*\*}$p$<$0.05; $^{\*\*\*}$p$<$0.01} \\

\end{tabular}

\end{table}

> library(car)

Zorunlu paket yükleniyor: carData

> qqPlot(mod) #see error distribution

[1] 443 652

> #Error distribution seems linearly distributed

> plot(cooks.distance(mod))

> par(mfrow = c(2,2))

> plot(mod)

> par(mfrow = c(1,1))

> mod.null = lm(PragmatticLeg1 ~ 1, data = data[!is.na(data$Ethnic),])

Error in lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :

0 (non-NA) cases

> mod.null = lm(PragmatticLeg1 ~ 1, data = data[!is.na(data$MoralLeg1),])

> anova(mod;mod.null) #check the model against null (just controls)

Error: unexpected ';' in "anova(mod;"

> anova(mod, mod.null)

Analysis of Variance Table

Model 1: PragmatticLeg1 ~ MoralLeg1

Model 2: PragmatticLeg1 ~ 1

Res.Df RSS Df Sum of Sq F Pr(>F)

1 998 622.53

2 999 736.26 -1 -113.73 182.32 < 2.2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> mod2 = lm(PragmatticLeg1 ~ MoralLeg1\*PoliticOrient, data = data)

> #may be results changes according to control variables

> summary(mod2)

Call:

lm(formula = PragmatticLeg1 ~ MoralLeg1 \* PoliticOrient, data = data)

Residuals:

Min 1Q Median 3Q Max

-3.0102 -0.5498 -0.1404 0.4502 2.2946

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.03871 0.71115 -0.054 0.95660

MoralLeg1 0.87742 0.21860 4.014 6.43e-05 \*\*\*

PoliticOrient[2] 2.03871 0.94133 2.166 0.03057 \*

PoliticOrient[3] 1.93326 0.79054 2.446 0.01464 \*

PoliticOrient[4] 2.32350 0.86409 2.689 0.00729 \*\*

PoliticOrient[5] 2.26859 0.77230 2.937 0.00339 \*\*

PoliticOrient[6] 1.87435 0.72916 2.571 0.01030 \*

PoliticOrient[7] 2.48962 0.81250 3.064 0.00224 \*\*

PoliticOrient[8th] 2.32885 0.86173 2.703 0.00700 \*\*

PoliticOrient[9] 2.56686 0.82895 3.097 0.00201 \*\*

PoliticOrientleft [1] 1.03871 1.31917 0.787 0.43124

PoliticOrientright [11] 3.03065 0.93359 3.246 0.00121 \*\*

MoralLeg1:PoliticOrient[2] -0.54409 0.27683 -1.965 0.04965 \*

MoralLeg1:PoliticOrient[3] -0.46360 0.23562 -1.968 0.04940 \*

MoralLeg1:PoliticOrient[4] -0.58831 0.25180 -2.336 0.01967 \*

MoralLeg1:PoliticOrient[5] -0.55251 0.23215 -2.380 0.01750 \*

MoralLeg1:PoliticOrient[6] -0.44251 0.22286 -1.986 0.04735 \*

MoralLeg1:PoliticOrient[7] -0.58833 0.24036 -2.448 0.01455 \*

MoralLeg1:PoliticOrient[8th] -0.55848 0.25034 -2.231 0.02592 \*

MoralLeg1:PoliticOrient[9] -0.68861 0.24603 -2.799 0.00523 \*\*

MoralLeg1:PoliticOrientleft [1] -0.18177 0.35818 -0.507 0.61194

MoralLeg1:PoliticOrientright [11] -0.82097 0.27416 -2.994 0.00282 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7856 on 978 degrees of freedom

Multiple R-squared: 0.1801, Adjusted R-squared: 0.1625

F-statistic: 10.23 on 21 and 978 DF, p-value: < 2.2e-16

> #here it seems political orient is correlated with results

> library(margins)

> cplot(mod2, x = 'PoliticOrient', what = 'effect', data = data)

Error in plot.window(...) : sonlu 'ylim' değerleri gerekli

In addition: Warning messages:

1: In min(x) : no non-missing arguments to min; returning Inf

2: In max(x) : no non-missing arguments to max; returning -Inf

> mod3 = lm(PragmatticLeg1 ~ MoralLeg1\*PoliticOrient + I(PoliticOrient > 5), data = data)

> cplot(mod3, x = 'PoliticOrient', what = 'effect', data = data)

Error in plot.window(...) : sonlu 'ylim' değerleri gerekli

In addition: There were 50 or more warnings (use warnings() to see the first 50)

> cplot(mod3, x = 'MoralLeg1', what = 'effect', data = data)

Error in plot.window(...) : sonlu 'ylim' değerleri gerekli

In addition: There were 50 or more warnings (use warnings() to see the first 50)

> mod4 = lm(PragmatticLeg1 ~ MoralLeg1\*PoliticOrient + CogLeg1, data = data)

> cplot(mod4, x = 'polity2\_', what = 'effect', data = data)

Error in seq.default(min(x, na.rm = TRUE), max(x, na.rm = TRUE), length.out = n) :

'from' must be a finite number

In addition: Warning messages:

1: In min(x, na.rm = TRUE) :

no non-missing arguments to min; returning Inf

2: In max(x, na.rm = TRUE) :

no non-missing arguments to max; returning -Inf

> mod5 = lm(mpg ~ wt + I(wt^2), data = mtcars)

> margins(mod5)

Average marginal effects

lm(formula = mpg ~ wt + I(wt^2), data = mtcars)

wt

-5.845

> cplot(mod5, "wt", what = "prediction", main = "Predicted WillingnessToCoProduce, Given Weight")

> cplot(mod5, "wt", what = "effect", main = "Average Marginal Effect of Weight")

> mod6 = lm(mpg ~ hp \* wt, data = mtcars)

> persp(mod6, "wt", "hp", theta = c(45, 135, 225, 315), what = "effect")

Chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart

Description automatically generated with low confidence

Engineering drawing

Description automatically generated