

Домашнее задание №1

Агалаян Роберт

2023-10-10

Задание 1

1.

Загрузка данных и очистка NA's.

```
df0 <- read_dta("lab1_3.dta")
df0 <- na.omit(df0)
summary(df0)
```

```
##      year      country      pol_stab      con_cor
##  Min.   :2006  Length:339    Min.   :-0.4738  Min.   :-0.2673
##  1st Qu.:2008    Class :character  1st Qu.: 0.5635  1st Qu.: 0.4338
##  Median :2010    Mode  :character  Median : 0.9288  Median : 1.3360
##  Mean   :2010                      Mean  : 0.8344  Mean   : 1.2375
##  3rd Qu.:2013                      3rd Qu.: 1.1273  3rd Qu.: 1.9978
##  Max.   :2015                      Max.   : 1.5255  Max.   : 2.4700
##  herfgov_DPI  govt_consump_WDI  countrygroup
##  Min.   :0.1842  Min.   :10.18  Min.   :1.000
##  1st Qu.:0.4315  1st Qu.:18.00  1st Qu.:2.000
##  Median :0.5865  Median :19.63  Median :3.000
##  Mean   :0.6508  Mean   :19.75  Mean   :2.469
##  3rd Qu.:0.9501  3rd Qu.:21.42  3rd Qu.:3.000
##  Max.   :1.0000  Max.   :27.94  Max.   :3.000
```

LSDV-модель:

```
LSDV <- lm(pol_stab ~ con_cor + factor(countrygroup) + herfgov_DPI + govt_consump_WDI, data = df0)
summary(LSDV)
```

```
##
## Call:
## lm(formula = pol_stab ~ con_cor + factor(countrygroup) + herfgov_DPI +
##   govt_consump_WDI, data = df0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09235 -0.17980  0.02141  0.19724  0.74640
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.580173    0.120857   4.800 2.39e-06 ***
## con_cor         0.436876    0.026565  16.445 < 2e-16 ***
## factor(countrygroup)2  0.359913    0.071144   5.059 6.98e-07 ***
## factor(countrygroup)3  0.014535    0.054693   0.266  0.79059
## herfgov_DPI     -0.140675    0.069692  -2.019  0.04434 *
## govt_consump_WDI  -0.014622    0.005239  -2.791  0.00556 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2841 on 333 degrees of freedom
## Multiple R-squared:  0.5277, Adjusted R-squared:  0.5206
## F-statistic: 74.41 on 5 and 333 DF, p-value: < 2.2e-16
```

$$\hat{y}_{it} = 0.58 + 0.36 * countrygroup2 + 0.015 * countrygroup3 + 0.437 * con_cor$$

2.

Intercept: 0.58 - среднее значение уровня политической стабильности в первой группе стран при нулевом значении показателя контроля коррупции. Оценка коэффициента значима на 0.1% уровне значимости.

countrygroup2: При прочих равных условиях значение уровня политической стабильности во второй группе стран в среднем больше на 0.36 в сравнении с базовой - первой группой стран. Оценка коэффициента значима на 0.1% уровне значимости.

countrygroup3: При прочих равных условиях значение уровня политической стабильности в третьей группе стран в среднем больше на 0.015 в сравнении с базовой - первой группой стран. Оценка коэффициента незначима.

con_cor: В среднем, при прочих равных при увеличении показателя контроля коррупции на одну единицу измерения уровень политической стабильности увеличивается на 0.437. Оценка коэффициента значима на 0.1% уровне значимости. Коэффициент отражает изменения воо временной перспективе. (т.к. изменения между группами стран отражаются коэффициентами при *countrygroupX*)

3.

Проверим качество работы модели для разных групп стран:

```
y_hat <- LSDV$fitted.values
panel0 <- data.frame(df0, y_hat)
merged <- panel0 %>% group_by(countrygroup)%>% summarize(., cor(pol_stab, y_hat))%>% merge(panel0, ., by="countrygroup")
merged$new <- ifelse(abs(merged$`cor(pol_stab, y_hat)`)<0.5,1,0)
unique(merged$country[merged$new == 1])
```

```
## [1] "Croatia"      "Bulgaria"      "Estonia"       "Czech Republic"
## [5] "Poland"        "Hungary"       "Slovakia"      "Slovenia"
```

```
merged$`cor(pol_stab, y_hat)`[merged$countrygroup == 1][1]
```

```
## [1] 0.7069278
```

```
merged$`cor(pol_stab, y_hat)`[merged$countrygroup == 2][1]
```

```
## [1] 0.4160657
```

```
merged$`cor(pol_stab, y_hat)`[merged$countrygroup == 3][1]
```

```
## [1] 0.7567848
```

Для второй группы стран модель работает хуже, чем для 1-й и 3-й: *cor* = 0.416

Возможно, эта группа слишком сильно отличается от остальных. Вполне допустимо, что вторая группа не похожа на первую и третью, поскольку последние состоят из представителей стран Запада.

Задание 2

1.

Для проверки различий во взаимосвязи предиктора и отклика в выделенных группах стран можно воспользоваться *varying-slope* моделью.

```
varslope <- lm(pol_stab ~ con_cor*factor(countrygroup), data = df0)
summary(varslope)
```

```
##
## Call:
## lm(formula = pol_stab ~ con_cor * factor(countrygroup), data = df0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.15044 -0.18293  0.02867  0.20777  0.69907
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.03723    0.20887  -0.178  0.85865
## con_cor         0.55734    0.11266   4.947 1.2e-06 ***
## factor(countrygroup)2  0.66217    0.21379   3.097  0.00212 **
## factor(countrygroup)3  0.24841    0.21293   1.167  0.24420
## con_cor:factor(countrygroup)2  -0.24209    0.13681  -1.770  0.07772 .
## con_cor:factor(countrygroup)3  -0.11911    0.11553  -1.031  0.30329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2884 on 333 degrees of freedom
## Multiple R-squared:  0.5134, Adjusted R-squared:  0.5061
## F-statistic: 70.27 on 5 and 333 DF, p-value: < 2.2e-16
```

Спецификация предложенной модели:

$$\hat{y}_{it} = -0.03723 + 0.55734 * con_cor + 0.66217 * countrygroup2 + 0.24841 * countrygroup3 - 0.24209 * con_cor * countrygroup2 - 0.11911 * con_cor * countrygroup3$$

2.

Интерпретация:

Intercept: среднее значение показателя политической стабильности в первой группе стран (базовой категории) при предикторах равных 0 составляет -0.03723.

con_cor: в среднем при прочих равных при увеличении значения показателя контроля коррупции на 1 значение показателя политической стабильности в группе стран 1 (базовой категории) изменяется на 0.55734.

countrygroup2: отклонение значение показателя политической стабильности в среднем во 2-ой группе стран от значения показателя политической стабильности в первой группе стран (базовой категории) при всех предикторах равных 0 равно 0.66217.

countrygroup3: отклонение значение показателя политической стабильности в среднем в 3-й группе стран от значения показателя политической стабильности в первой группе стран (базовой категории) при всех предикторах равных 0 составляет 0.24841.

con_cor:countrygroup2: в среднем взаимосвязь значения показателя контроля коррупции и значения показателя политической стабильности во 2-ой группе стран отличается от первой группы стран (базовой категории) при прочих равных на -0.24209.

con_cor:countrygroup3: в среднем взаимосвязь значения показателя контроля коррупции и значения показателя политической стабильности в 3-й группе странотличается от первой группы стран (базовой категории) при прочих равных на -0.11911.

Исходя из оценок коэффициентов при предикторах, а точнее на оценке при *con_cor:countrygroup2*, значимой (но только на 0.1% уровне значимости, в то время как оценка при *con_cor:countrygroup3* не значима), мы можем сказать, что различия во взаимосвязи есть, по крайней мере между уже выделевшейся ранее второй группой и первой.

Задание 3

Первая модель определённо имеет проблемы в виде её относительной громоздкости, а также в связи с тем, что она не справляется одинаково и достаточно хорошо со всеми группами стран.

Вторая модель учитывает различия во взаимосвязи в группах стран, однако ввиду того, что мы имеем достаточно похожие группы (1 и 3) - представителей развитых стран Запада, и отличающуюся от них (2) группу стран Восточной Европы (посткоммунистических), нам стоит учесть этот фактор и, возможно, стоит пересобрать группы (1+3 и 2).

Задание 4

1.

LSDV-модель с фиксированными эффектами на годы:

```
LSDV_year <- lm(pol_stab ~ con_cor + factor(year) + herfgov_DPI + govt_consump_WDI, data = df0)
summary(LSDV_year)
```

```
##
## Call:
## lm(formula = pol_stab ~ con_cor + factor(year) + herfgov_DPI +
##   govt_consump_WDI, data = df0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13247 -0.19660  0.02381  0.21475  0.68752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.849212    0.124420   6.825 4.29e-11 ***
## con_cor         0.325912    0.021966  14.837 < 2e-16 ***
## factor(year)2007  -0.025563    0.075425  -0.339  0.734890
## factor(year)2008  -0.027601    0.075437  -0.366  0.714688
## factor(year)2009  -0.069669    0.076125  -0.915  0.360770
## factor(year)2010  -0.030807    0.075801  -0.502  0.616060
## factor(year)2011  -0.008808    0.075716  -0.116  0.907461
## factor(year)2012  -0.034506    0.075711  -0.456  0.648869
## factor(year)2013   0.011880    0.076222   0.156  0.876240
## factor(year)2014  -0.032865    0.075657  -0.434  0.664289
## factor(year)2015  -0.067394    0.075520  -0.892  0.372839
## herfgov_DPI     -0.231750    0.065394  -3.544  0.000452 ***
## govt_consump_WDI -0.012048    0.005573  -2.162  0.031352 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3109 on 326 degrees of freedom
## Multiple R-squared:  0.4464, Adjusted R-squared:  0.426
## F-statistic: 21.91 on 12 and 326 DF, p-value: < 2.2e-16
```

$$\hat{y}_{it} = 0.849 - 0.026 * year2007 - 0.028 * year2008 - 0.07 * year2009 - 0.038 * year2010 - 0.009 * year2011 - 0.035 * year2012 + 0.12 * year2013 - 0.033 * year2014 - 0.067 * year2015 + 0.326 * con_cor$$

2.

Intercept: В среднем значение показателя политической стабильности в 2006 году (базовая категория) равно 0.849, при равенстве предиктора нулю.

year2010: В среднем при прочих равных значение показателя политической стабильности в 2010 году отклоняется от базовой категории (2006 год) на 0.038.

con_cor: В среднем при прочих равных при увеличении показателя контроля коррупции на одну единицу измерения уровень политической стабильности увеличивается на 0.326. Коэффициент отражает изменения между группами стран. (т.к. изменения между временными периодами отражают коэффициентам при *yearXXXX*)

В данном случае дамми поглощают все изменяющиеся во времени характеристики, и неизменяющиеся в пространстве. Поскольку в данном случае мы не хотим сравнить показатель переменной до и после какого-либо события, поэтому использование такой модели кажется нецелесообразным. Помимо этого, сама по себе LSDV модель достаточно объёмная и стоит рассмотреть вариант модели с внутригрупповым преобразованием, это более экономный вариант. Также стоит заметить, что оценки коэффициентов при годах не значимы. Также, модель с фиксированными эффектами на года имеет более низкий R-квадрат = 0.45 < 0.5 и 0.52 у *varying-slope* и LSDV (на пространственные единицы) моделей соответственно.