

## Домашнее задание 1 Рубанов Владислав БПТ 201

### Задание 1.

Преобразуем переменную коррупции (*cpi*), чтобы большему значению соответствовало более высокое значение коррупции. Сейчас переменная принимает значения от 0 до 10. Для того, чтобы сохранить изначальную шкалу измерения индекса, перезапишем переменную *cpi*: вычтем из 10 каждое (старое) значение *cpi*.

Перед этим также очистим данные и выведем их в более удобном порядке.

```
library(foreign)
library(dplyr)
library(ggplot2)
library(GGally)
library(margins)
library(psych)
library(memisc)
library(haven)

df <- read.dta("tvims3_hw1.dta")

labhw <- df %>% dplyr::select(country, cpi, dem, fp, loggdppc, stab, britcol) %>% na.omit()
labhw <- mutate(labhw, cpi = 10 - cpi)

head(labhw)
```

##	country	cpi	dem	fp	loggdppc	stab	britcol
## 1	Afghanistan	8.300000	10.766666	27.66667	6.216606	-2.495173931	0
## 2	Albania	7.016667	24.666666	50.00000	8.216088	-0.243981183	0
## 3	Algeria	7.033333	10.100000	37.83333	8.427268	-1.127763748	0
## 5	Angola	7.983333	2.383333	37.16667	8.371011	-0.505494356	0
## 7	Argentina	7.116667	28.466665	51.66667	9.118225	0.007387357	0
## 8	Armenia	7.166667	22.616667	34.50000	8.016318	-0.009873525	0

Теперь приведем **описательные статистики** по нашим переменным:

```
describe(labhw)
```

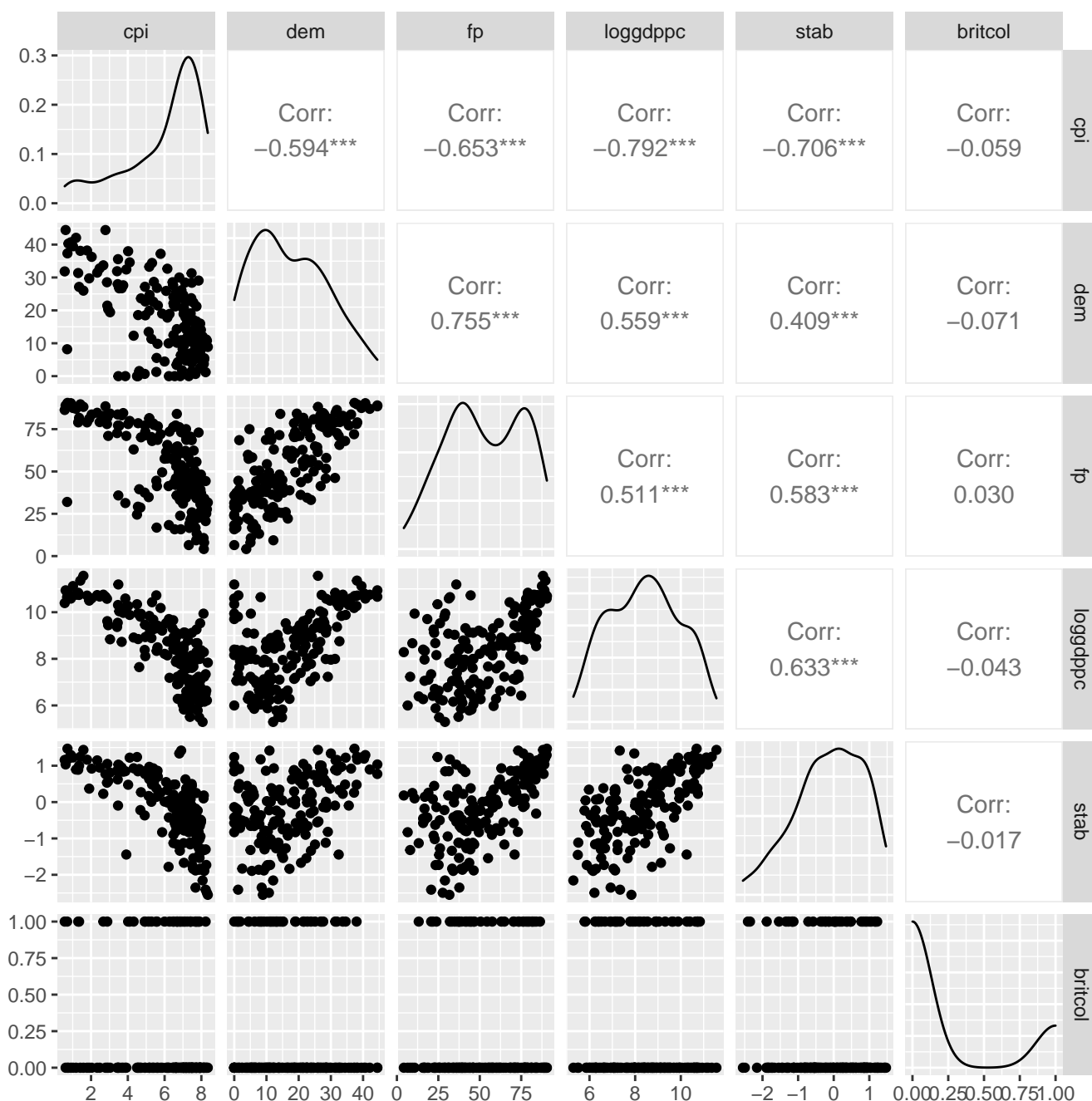
##	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew
## country*	1	170	85.50	49.22	85.50	85.50	63.01	1.00	170.00	169.00	0.00
## cpi	2	170	5.96	2.08	6.71	6.27	1.52	0.57	8.35	7.78	-1.12
## dem	3	170	17.26	11.41	15.57	16.71	13.21	0.00	44.43	44.43	0.35
## fp	4	170	52.94	22.83	51.08	53.44	29.03	4.17	90.50	86.33	-0.09
## loggdppc	5	170	8.40	1.53	8.40	8.39	1.87	5.29	11.56	6.27	0.01
## stab	6	170	-0.12	0.94	0.00	-0.06	1.05	-2.55	1.47	4.02	-0.45
## britcol	7	170	0.22	0.42	0.00	0.15	0.00	0.00	1.00	1.00	1.32
##	kurtosis	se									

```
## country*      -1.22 3.77
## cpi           0.18 0.16
## dem          -0.82 0.88
## fp           -1.14 1.75
## loggdppc     -0.99 0.12
## stab         -0.49 0.07
## britcol      -0.27 0.03
```

```
summary(labhw)
```

```
##      country          cpi          dem          fp
## Length:170      Min.    :0.5667      Min.    : 0.000      Min.    : 4.167
## Class :character 1st Qu.:4.9375      1st Qu.: 7.896      1st Qu.:35.917
## Mode  :character Median :6.7083      Median :15.575      Median :51.083
##                      Mean  :5.9585      Mean  :17.256      Mean  :52.939
##                      3rd Qu.:7.4787      3rd Qu.:25.979      3rd Qu.:73.500
##                      Max.   :8.3500      Max.   :44.433      Max.   :90.500
##      loggdppc      stab      britcol
## Min.    : 5.293      Min.    :-2.550327      Min.    :0.0000
## 1st Qu.: 7.111      1st Qu.: -0.717995      1st Qu.:0.0000
## Median : 8.396      Median :-0.001243      Median :0.0000
## Mean    : 8.399      Mean    :-0.118161      Mean    :0.2235
## 3rd Qu.: 9.531      3rd Qu.: 0.628244      3rd Qu.:0.0000
## Max.    :11.562      Max.    : 1.468562      Max.    :1.0000
```

```
ggpairs(labhw[, -c(1)])
```



Предварительные **описательные статистики** имеют особую важность в контексте регрессионного анализа с переменными взаимодействия, поскольку здесь, как нигде иначе, нам необходимо понимать **распределения переменных**, с которыми мы будем работать. Теперь гипотеза, которую мы хотим проверить, носит условный характер. Выдача регрессии больше не будет интерпретироваться, как в безусловной линейно-зависимой модели, в терминах среднего эффекта.

Несмотря на то, что оценки коэффициентов в модели теперь будет показывать лишь эффект или значение при всех предикторах, равных 0, либо при некоторых из переменных, равных 0 (например, при модераторе и переменной взаимодействия, равных 0), для нас **критически важно** знать две вещи для верной интерпретации результатов регрессии. *Во-первых*, что означает фраза „при X (или любой другой переменной(-ых)), равном(-ой) нулю“, а, *во-вторых*, что скрывается за фразой „с увеличением на одну единицу измерения“ — ведь смысл будет сильно отличаться, если мы имеем дело с дамми-переменной; непрерывной шкалой от 0 до 1, от -10 до 10, от 0 до 100; с логарифмом и т.д. В некоторых случаях это может быть изменением при переходе от теоретического минимума к теоретическому максимуму, а где-то — теоретически такого значения может даже и не быть, либо не наблюдаться в выборке. В особенности это важно, если мы применяем процедуру центрирования: чтобы при интерпретации можно было понимать, сколько составляет среднее значение переменной.

Также важно иметь представление о **распределении** переменных, с которыми мы работаем (см. корреляционную матрицу с распределением по главной диагонали) и их корреляцию между собой. То же самое касается интерпретации с помощью предельных эффектов. **Важно понимать, с какими данными мы работаем.**

Итак, из выдачи мы можем увидеть, что после очистки данных мы работаем со 170 наблюдениями. Мы видим, что теперь значение *cpi* действительно принимает большее значение при более высоком значении коррупции (все так же непрерывная шкала от 0 до 10), причем важно, что „0“ в настоящий момент будет означать полное отсутствие коррупции, в отличие от того, если бы мы не совершили преобразование с данной переменной. Однако важно отметить, что минимальное значение в датасете, очевидно, не 0, а 0.57. Максимальное же: 8.35. Среднее значение зависимой переменной равно 5.96, а медиана: 6.71.

Несмотря на то, что Vanhanen's democratization index (*dem*) является непрерывной шкалой от 0 до 100, в представленном датасете переменная *dem* принимает значения от 0 до 44.43, т.е. нет наблюдений со значением даже больше половины (среднее = 17.26).

Переменная *fp* (свобода прессы) в массиве принимает значения от 4,17 до 90,50 — практически весь возможный диапазон (переменная приведена к непрерывной шкале от 0 до 100). Для будущей интерпретации важно заметить, что переменная свободы прессы не принимает значения 0 в выборке. Также разброс этой переменной между наблюдениями самый большой.

Очевидно, что натуральный логарифм ВВП на душу населения (*loggdppc*) принимает значения от 5.29 до 11.56. Необходимо быть осторожным с интерпретацией его коэффициента и эффекта: *значение „0“ очевидным образом является лишь теоретическим.*

Уровень политической стабильности (*stab*) варьируется от -2,55 до 1,47 — в данном случае при интерпретации „0“ *будет медианой (приблизительно).*

И, очевидно, что для дамми-переменной *britcol* „0“ означает то, что страна не является бывшей британской колонией, а „1“ — что является бывшей британской колонией. По описательным статистикам можно понять, что большинство стран в датасете логичным образом ими не являются.

Также можно отметить, что между ключевыми предикторами *fp* и *dem* прослеживается сильная корреляция (0.755), которую можно логически объяснить. Это может создать некоторые трудности в ходе работы.

Теперь мы готовы идти дальше.

## Задание 2.

Разделим страны на 2 группы: „not free“ и „free“ по показателю „свобода прессы“ (значение 70 в качестве критерия).

Для этого в начале **создадим тестовый датасет**, в котором создадим новую дамми-переменную *freedom\_of\_press* (для того, чтобы не удалять существующую переменную *fp* и иметь возможность сравнить успешность работы алгоритма по разделению данных на подвыборки). Затем заменим значения дамми-переменной „0“ и „1“ на „not free“ и „free“, а также **установим переменную *freedom\_of\_press* в качестве фактора.**

```
# создание подгрупп
```

```
test_data <- mutate(labhw, freedom_of_press = as.numeric(labhw$fp > 70))
test_data$freedom_of_press <- dplyr::recode(test_data$freedom_of_press, "1" = "free", "0" = "not free")
test_data <- mutate(test_data, freedom_of_press = factor(freedom_of_press))
```

```
head(test_data)
```

```
##      country      cpi      dem      fp loggdppc      stab britcol
## 1 Afghanistan 8.300000 10.766666 27.66667 6.216606 -2.495173931      0
```

```
## 2      Albania 7.016667 24.666666 50.00000 8.216088 -0.243981183      0
## 3      Algeria 7.033333 10.100000 37.83333 8.427268 -1.127763748      0
## 5        Angola 7.983333  2.383333 37.16667 8.371011 -0.505494356      0
## 7    Argentina 7.116667 28.466665 51.66667 9.118225  0.007387357      0
## 8      Armenia 7.166667 22.616667 34.50000 8.016318 -0.009873525      0
##  freedom_of_press
## 1          not free
## 2          not free
## 3          not free
## 5          not free
## 7          not free
## 8          not free
```

Итак, по первым значениям мы можем видеть, что у нас получилось корректно разделить страны на 2 группы.

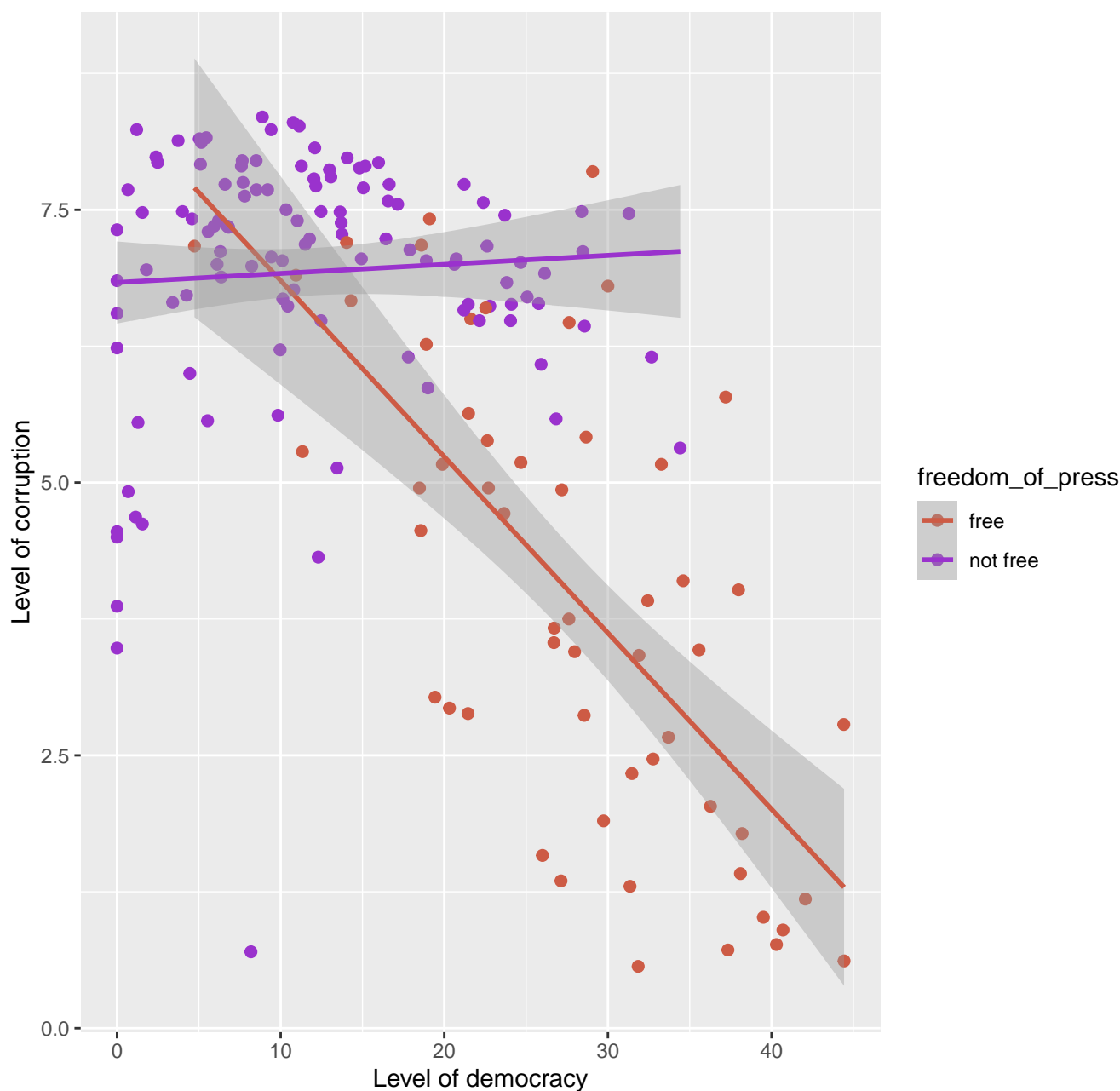
Теперь **визуализируем взаимосвязь коррупции и уровня демократии** в разных группах „not free“ и „free“.

```
# график

ggplot(data = test_data, aes(x = dem, y = cpi, color = freedom_of_press)) +
  geom_point(size=2) + labs(x = "Level of democracy",
                           y = "Level of corruption",
                           title = "The effect of democracy on corruption") +
  geom_smooth(method=lm)+ scale_colour_manual(values = c("coral3", "darkorchid3"))

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

## The effect of democracy on corruption



На графике мы можем видеть **предварительную картину**, показывающую положение дел еще до оценивания полноценной модели.

На графике мы можем увидеть две регрессионной модели (и соответствующие им линии). **Факторным цветом** на графике изображены государства, для которых индекс свободы прессы составляет менее 70 из 100 („not free“, фиолетовый цвет), а также те, у которых данный индекс больше 70 („free“, рыжий цвет). По оси абсцисс находится уровень демократии (*dem*: *большее значение означает рост уровня демократии*), по оси ординат — уровень коррупции (*cpi*: *более высокое значение соответствует большему значению коррупции*).

Мы сразу можем заметить **нетипичные наблюдения** в данных: в основном они относятся к группе „not free“: в левом нижнем углу, а также в левой части графика по оси ординат при значении демократии, равным 0, либо близким к нему с нетипично низкими значениями уровня коррупции для такого значения уровня демократии.

На графике отчетливо видно, что для государств из группы „free“ с высоким значением свободы прессы прослеживается **сильная отрицательная взаимосвязь переменных *dem* и *cpi***: с увеличением демократии индекс коррупции заметно уменьшается (то есть **уровень коррупции снижается**).

Для стран из группы „not free“ с низким уровнем значения свободы прессы, наоборот, в среднем прослеживается **слабая положительная связь** между демократией и индексом коррупции: то есть, согласно графику, в среднем, в странах со значением *fp* меньше 70, **увеличение уровня демократии приводит к увеличению уровня коррупции**.

Однако **важно заметить**, что в случае для стран из группы „not free“ много точек (наблюдений) выходит за рамки доверительного интервала: ранее уже было сказано о нетипичных наблюдениях, да и не только они выходят за рамки ДИ. Следовательно, **можно предположить**, что у коэффициента для демократии в данной модели будут большие стандартные ошибки и, следовательно, незначимый коэффициент — слишком много выбивающихся точек. Эффект может быть слабым и статистически незначимым.

Также, в любом случае, если говорить об этих данных, мы уже можем сказать о наличии гетероскедастичности, поскольку у нас есть **переменная-условие** *freedom\_of\_press*, которая делит выборку на 2 части. Следовательно, у нас уже не единая выборка, а **две выборки в одном датасете**.

Таким образом, после некоторой разведывательной работы, **можно предположить**, что для стран с низким уровнем значения свободы прессы эффект демократии на увеличение коррупции будет слабым положительным (не сильно выраженным) и незначимым, а для стран, в которых уровень свободы прессы высокий, можно ожидать сильной отрицательной взаимосвязи между демократией и коррупцией в государстве.

Тем не менее, нам необходимо оценить модель и протестировать различия во взаимосвязи коррупции и демократии в разных группах стран, а также проинтерпретировать их.

**Важно отметить**, что в данном случае оценивание двух регрессионных моделей на двух группах стран вместо оценивания модели по всей выборке при подобной спецификации, не приведет к смещению коэффициентов, поскольку в модели нет контрольных переменных. Однако все равно **делать так не рекомендуется**, поскольку, в любом случае, *n* в каждой из моделей по подгруппам будет меньше, чем в полной модели, а, следовательно, будут больше стандартные ошибки оценок и больше вероятность того, что коэффициенты могут стать незначимыми (особенно, если одна из подвыборок сильно меньше другой по объему), а оценки — несостоятельными. К тому же, мы можем столкнуться с проблемой большего веса нетипичных наблюдений и не сможем явно сравнить оценки регрессий, т.к. они будут получены на разных подвыборках.

```
# модели

free <- subset(test_data, freedom_of_press == 'free')
m1_1 <- lm(cpi ~ dem, data = free)
summary(m1_1)

##
## Call:
## lm(formula = cpi ~ dem, data = free)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7562 -0.9903 -0.3070  0.8928  4.0777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.46526    0.69822  12.124 < 2e-16 ***
## dem          -0.16145    0.02392  -6.748 1.05e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.581 on 54 degrees of freedom
## Multiple R-squared:  0.4575, Adjusted R-squared:  0.4475
## F-statistic: 45.54 on 1 and 54 DF,  p-value: 1.055e-08

not_free <- subset(test_data, freedom_of_press == 'not free')
m1_2 <- lm(cpi ~ dem, data = not_free)
summary(m1_2)

##
## Call:
## lm(formula = cpi ~ dem, data = not_free)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.2011 -0.4040  0.2593  0.7780  1.4431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.833337   0.189787  36.005  <2e-16 ***
## dem          0.008284   0.012831   0.646    0.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.172 on 112 degrees of freedom
## Multiple R-squared:  0.003708, Adjusted R-squared:  -0.005188
## F-statistic: 0.4168 on 1 and 112 DF,  p-value: 0.5199

mtable(m1_1, m1_2)

##
## Calls:
## m1_1: lm(formula = cpi ~ dem, data = free)
## m1_2: lm(formula = cpi ~ dem, data = not_free)
##
## =====
##              m1_1          m1_2
## -----
## (Intercept)  8.465***    6.833***
##              (0.698)    (0.190)
## dem          -0.161***    0.008
##              (0.024)    (0.013)
## -----
## R-squared     0.458        0.004
## N             56          114
## =====
## Significance: *** = p < 0.001;
##               **  = p < 0.01;
##               *   = p < 0.05
```

В последней таблице мы можем явно увидеть сравнение результатов оценивания двух моделей. Проинтерпретируем их:

**Первая модель** (подгруппа „free“,  $fp > 70$ ):



- $\hat{\beta}_0 = 8.465$  (Intercept): В странах, в которых значение свободы прессы ( $fp$ ) больше 70, среднее значение индекса коррупции, при условии равенства уровня демократии 0, равно 8.465 (минимум 0, максимум 10, большему значению соответствует больший уровень коррупции), при прочих равных условиях.
  - Примечательно, что данное значение оценки константы лишь теоретическое: среди стран из группы „free“ нет ни одного наблюдения, у которого значение демократии было бы равно 0.
- $\hat{\beta}_1 = -0.161$  (dem): В странах, в которых значение свободы прессы больше 70, с увеличением значения индекса демократии на 1 единицу измерения, значение индекса коррупции в среднем уменьшается на 0.161, при прочих равных условиях.

**Вторая модель** (подгруппа „not free“,  $fp < 70$ ):

- $\hat{\beta}_0 = 6.833$  (Intercept): В странах, в которых значение свободы прессы ( $fp$ ) меньше 70, среднее значение индекса коррупции, при условии равенства уровня демократии 0, равно 6.833 (минимум 0, максимум 10, большему значению соответствует больший уровень коррупции), при прочих равных условиях.
  - А вот среди стран из группы „not free“ наблюдений с индексом демократии = 0 уже сразу несколько.
- $\hat{\beta}_1 = 0.008$  (dem): В странах, в которых значение свободы прессы меньше 70, с увеличением значения индекса демократии на 1 единицу измерения, значение индекса коррупции в среднем увеличивается на 0.008, при прочих равных условиях (коэффициент статистически незначим).

Можно заметить, что наши **предположения подтвердились**: коэффициент при предикторе  $dem$  значим на очень малом уровне значимости для подгруппы стран „free“ и действительно является незначимым по подгруппе стран „not free“ (p-value составляет целых 0.52, это очень много).

Опять же, в данном случае эти оценки можно было получить из модели с переменной взаимодействия, выведя их по отдельности, поскольку в приведенной спецификации нет контрольных переменных.

### Задание 3.

Оценим модель с переменной взаимодействия, в которой  $fp$  (свобода прессы) будет выступать в качестве **модератора**, но **без деления** стран на группы „free“ и „not free“.

```
m2 <- lm(cpi ~ dem + fp + dem:fp, data = labhw)
summary(m2)

##
## Call:
## lm(formula = cpi ~ dem + fp + dem:fp, data = labhw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.4638 -0.6189  0.0900  0.7843  3.1928
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.9463178  0.3808200  15.615  < 2e-16 ***
## dem          0.2229018  0.0287393   7.756 8.42e-13 ***
## fp           0.0153127  0.0084706   1.808  0.0725 .
##
```

```
## dem:fp      -0.0041878  0.0004077 -10.272  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.217 on 166 degrees of freedom
## Multiple R-squared:  0.664, Adjusted R-squared:  0.658
## F-statistic: 109.4 on 3 and 166 DF,  p-value: < 2.2e-16
```

Для удобства интерпретации я предлагаю совершить процедуру **центрирования** предиктора  $dem$ . Мы вычтем из каждого значения предиктора  $dem_i$  его среднее значение. Теперь  $dem_i = 0$ , когда  $dem_i = \overline{dem}$ , то есть равно своему среднему значению по выборке. Это сильно упростит интерпретацию для нас.

По моему мнению, предиктор  $fp$  лучше не центрировать, иначе в связке с центрированием  $dem$  это может привести к усложнению интерпретации. В любом случае, это не приведет к изменению предельного эффекта (как и коэф. при перем. взаимодействия), поскольку это лишь алгебраическое преобразование.

```
#центрирование
labhw_cen <- mutate(labhw, dem_c = dem - mean(dem))
labhw_cen$dem <- NULL

head(labhw_cen)

##      country      cpi      fp loggdppc      stab britcol      dem_c
## 1 Afghanistan 8.300000 27.66667 6.216606 -2.495173931      0 -6.488981
## 2   Albania 7.016667 50.00000 8.216088 -0.243981183      0  7.411019
## 3   Algeria 7.033333 37.83333 8.427268 -1.127763748      0 -7.155647
## 5    Angola 7.983333 37.16667 8.371011 -0.505494356      0 -14.872314
## 7  Argentina 7.116667 51.66667 9.118225  0.007387357      0  11.211018
## 8   Armenia 7.166667 34.50000 8.016318 -0.009873525      0  5.361020
```

И теперь снова оценим регрессионную модель с переменной-модератором  $fp$  и переменной взаимодействия.

```
m3 <- lm(cpi ~ dem_c + fp + dem_c:fp, data = labhw_cen)
summary(m3)

##
## Call:
## lm(formula = cpi ~ dem_c + fp + dem_c:fp, data = labhw_cen)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.4638 -0.6189  0.0900  0.7843  3.1928
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.7926330  0.3752115  26.099  < 2e-16 ***
## dem_c         0.2229018  0.0287393   7.756 8.42e-13 ***
## fp          -0.0569507  0.0063929  -8.908 8.84e-16 ***
## dem_c:fp     -0.0041878  0.0004077 -10.272  < 2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.217 on 166 degrees of freedom
## Multiple R-squared:  0.664, Adjusted R-squared:  0.658
## F-statistic: 109.4 on 3 and 166 DF,  p-value: < 2.2e-16
```

**Спецификация** модели будет выглядеть следующим образом:

$$C\hat{P}I_i = 9.793 + 0.223(dem_i - \overline{dem}) - 0.057fp_i - 0.004(dem_i - \overline{dem}) \times fp_i$$

Интерпретация:

- $\hat{\beta}_0 = 9.7926330$  (*Intercept*) — среднее значение зависимой переменной (уровня коррупции) при всех предикторах, равных, 0; то есть при уровне демократии, равном среднему значению и при уровне прессы, равному 0 (где большее значение индекса коррупции означает больший уровень коррупции) и при прочих равных условиях;
- $\hat{\beta}_1 = 0.2229018$  ( $dem\_c$ ) — при увеличении значения предиктора уровень демократии на 1 единицу измерения, значение зависимой переменной (уровень коррупции) в среднем будет увеличиваться на 0.2229018 при условии того, что уровень свободы прессы равен 0 (минимальному значению) и при прочих равных условиях;
- $\hat{\beta}_2 = -0.0569507$  ( $fp$ ) — при увеличении свободы прессы на 1 единицу измерения, уровень коррупции в стране в среднем уменьшается на 0.0569507 при условии того, что центрированный предиктор — уровень демократии — равен 0, т.е. уровень демократии равен среднему значению и при прочих равных условиях;
- $\hat{\beta}_3 = -0.0041878$  ( $dem\_c \times fp$ ) — при увеличении свободы прессы на 1 единицу измерения и при прочих равных условиях взаимосвязь уровня демократии и уровня коррупции в начале становится менее выраженной положительной, а затем — становится более выраженной отрицательной: коэф. взаимосвязи в среднем уменьшается на 0.0041878.

Итак, мы уже можем сделать некоторые **выводы**.

1. Все наши оценки являются **значимыми** на очень малом уровне значимости.
2. Казалось бы, даже при уровне демократии, равному среднему значению (17.26), если уровень свободы прессы равен 0 (минимальное значение, которого нет в датасете), уровень коррупции в среднем равен 9.79, что *практически является теоретическим максимумом данного индекса*. Мы видим, что **без свободы прессы уровень демократии никак не способствует снижению коррупции**.
3. Более того, при том же (теоретическом) уровне свободы прессы, равному 0, **увеличение демократии** на 1 единицу измерения **приводит к увеличению уровня коррупции на 0.22** (Corruption Perception Index принимает значения от 0 до 10). Данный вывод звучит еще более красноречиво.
4. Благодаря центрированию переменной  $dem$  мы можем увидеть, что при уровне демократии, равному среднему, **увеличение уровня свободы прессы** на 1 единицу измерения **приводит к снижению уровня коррупции** на 0.06.
5. Также можно увидеть, что с определенного момента **взаимосвязь демократии и коррупции меняет свой знак с положительного на отрицательный** с ростом свободы прессы. То есть, когда уровень индекс свободы прессы становится больше некоторого значения (ниже мы узнаем, какого), рост демократии действительно приводит к уменьшению коррупции.

**Можно предположить**, что увеличение свободы прессы **всегда** приводит к снижению уровня коррупции (даже если уровень демократии очень низкий), в то время как увеличение уровня демократии приводит к снижению уровня коррупции лишь начиная с определенного уровня свободы прессы.

Также необходимо отметить, что алгебраическое преобразование, возникающее в результате центрирования переменных, привело к различным коэффициентам и стандартным ошибкам (не всем) в модели с центрированным предиктором *dem* по сравнению с таковыми в нецентрированной модели. Однако это происходит потому, что они измеряют разные существенные величины в каждой модели, а не потому, что одна модель дает лучшие оценки, чем другая (Brambor et al, 2006). Важно, что **предельный эффект** (интересующий нас в большей степени в модели взаимодействия) **не изменится**.

Однако можно заметить, что значение оценок коэффициентов при предикторе *dem* и при переменной взаимодействия *dem* и *fp* не изменились по сравнению с моделью без центрирования переменной *dem*. Это обусловлено особенностями центрирования.

#### Задание 4.

Для того, чтобы понять, как изменяется предельный эффект демократии на зависимую переменную и его значимость в зависимости от „условия“ (свободы прессы), в начале необходимо рассчитать его в общем виде, вспомнив, что **предельный эффект — это то, на сколько будет изменяться зависимая переменная *cpi* при минимальном изменении *dem*** и взять частную производную по переменной *dem*. Еще раз запишем спецификацию модели:

$$\hat{CPI}_i = 9.793 + 0.223(dem_i - \overline{dem}) - 0.057fp_i - 0.004(dem_i - \overline{dem}) \times fp_i$$

Предельный эффект будет выглядеть следующим образом:

$$\frac{\partial cpi_i}{\partial dem_i} = \hat{\beta}_1 + \hat{\beta}_3 \cdot fp_i$$

$$\frac{\partial cpi_i}{\partial dem_i} = 0.2229018 - 0.0041878 \cdot fp_i$$

Мы видим, что **предельный эффект демократии на уровень коррупции зависит от значения свободы прессы**, о чем говорилось ранее. Также можно заметить, что при уровне свободы прессы, равному 0, предельный эффект составляет 0.2229018 (сам коэф.  $\hat{\beta}_1$ ), о чем также говорилось ранее. Однако постепенно он уменьшается с ростом *fp* и в определенный момент становится отрицательным.

Можно сразу найти „точку перегиба“: значение свободы прессы, после которого предельный эффект становится отрицательным, т.е. меняет знак, **приравняв производную к 0**:

$$0.2229018 - 0.0041878 \cdot fp_i = 0$$

$$fp_i \approx 53.23$$

Следовательно, после того, как *fp* становится больше 53.23 (это немного больше среднего (52.94)), **предельный эффект демократии на уровень коррупции становится отрицательным**.

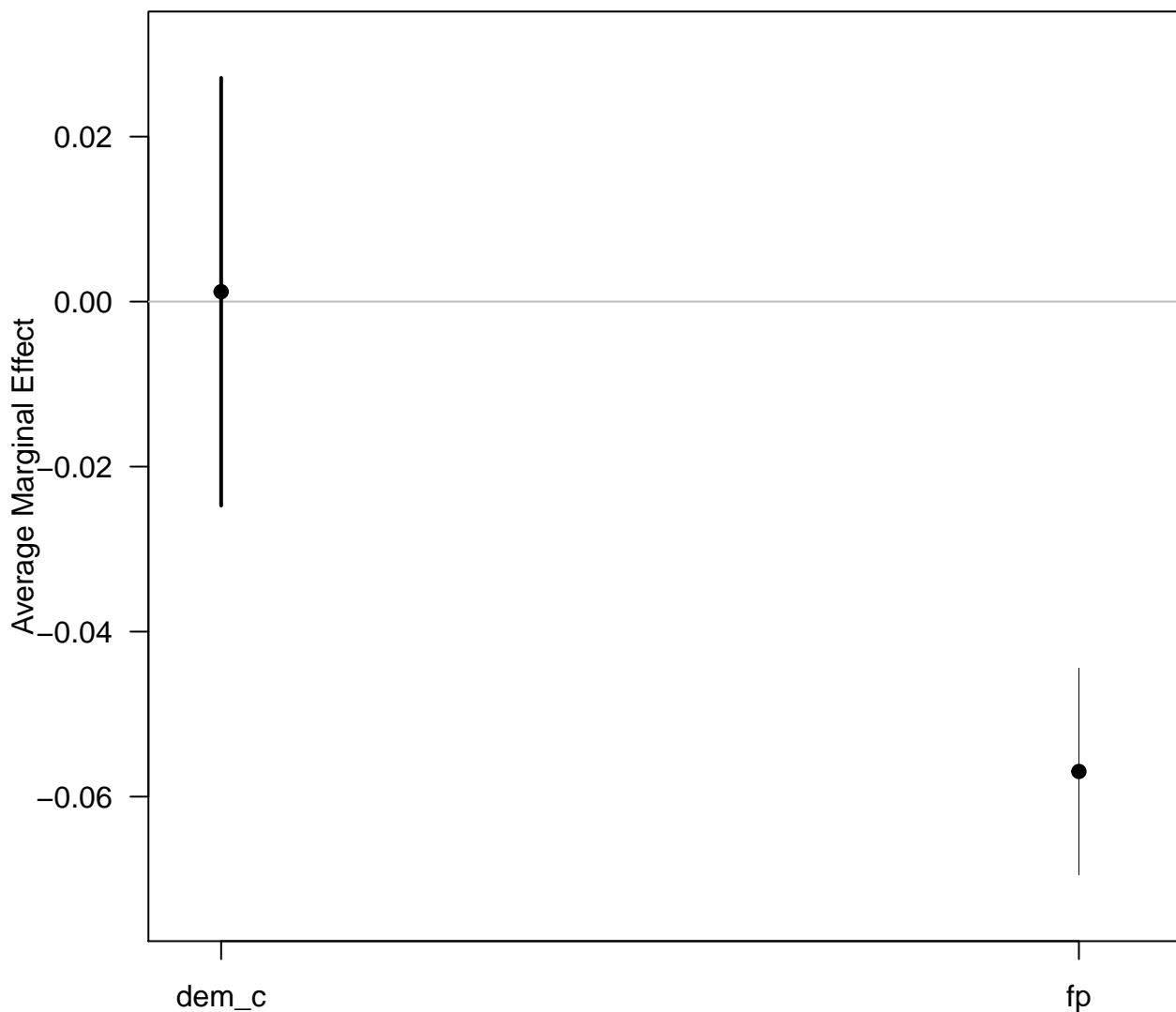
Далее проведем расчеты в R:

```
# предельный эффект

margins_m3 <- margins(m3)
summary(margins_m3)

## factor      AME      SE      z      p    lower    upper
## dem_c    0.0012 0.0132  0.0910 0.9275 -0.0247  0.0271
## fp    -0.0570 0.0064 -8.9084 0.0000 -0.0695 -0.0444
```

```
plot(margins_m3)
```



В данной выдаче мы можем увидеть **AME** — Average Marginal Effect — усредненный предельный эффект.

Однако АМЕ не является столь емким в интерпретации, мы можем лишь понять, как в среднем одна переменная влияет на другую и представить общую картинку — предварительную и описательную.

Так, в среднем предельный эффект демократии на коррупцию равен 0.0012 (к тому же, в среднем, он статистически незначим: доверительный интервал накрывает значение 0, а также p-value очень высок; также это можно явно заметить на графике). Однако мы уже видели, что в определенный момент предельный эффект меняет знак при достижении переменной, отвечающей за свободу прессы, значения 53.23, поэтому нам было бы интереснее посмотреть на значения, которые принимает предельный эффект демократии на коррупцию при различных интересующих нас значениях свободы прессы. При этом можно отметить, что, в случае с *fp*, *предельный эффект свободы прессы на уровень демократии в среднем статистически значим* и отрицателен: это добавляет аргумент в пользу того, что увеличение свободы прессы всегда приводит к снижению уровня коррупции.

Уйдем от понятия среднего предельного эффекта и посчитаем МЕ в разных **интересующих нас**

**точках.** Выберем значения индекса свободы прессы от 0 до 100 с шагом в 10 и посмотрим, как это отразится на предельном эффекте демократии на коррупцию:

```
s <- margins(m3, at = list(fp = seq(0, 100, by = 10)))

## Warning in check_values(data, at): Some 'at' values for 'fp' are outside observed data
## range (4.16666650772095,90.5)!

summary(s)
```

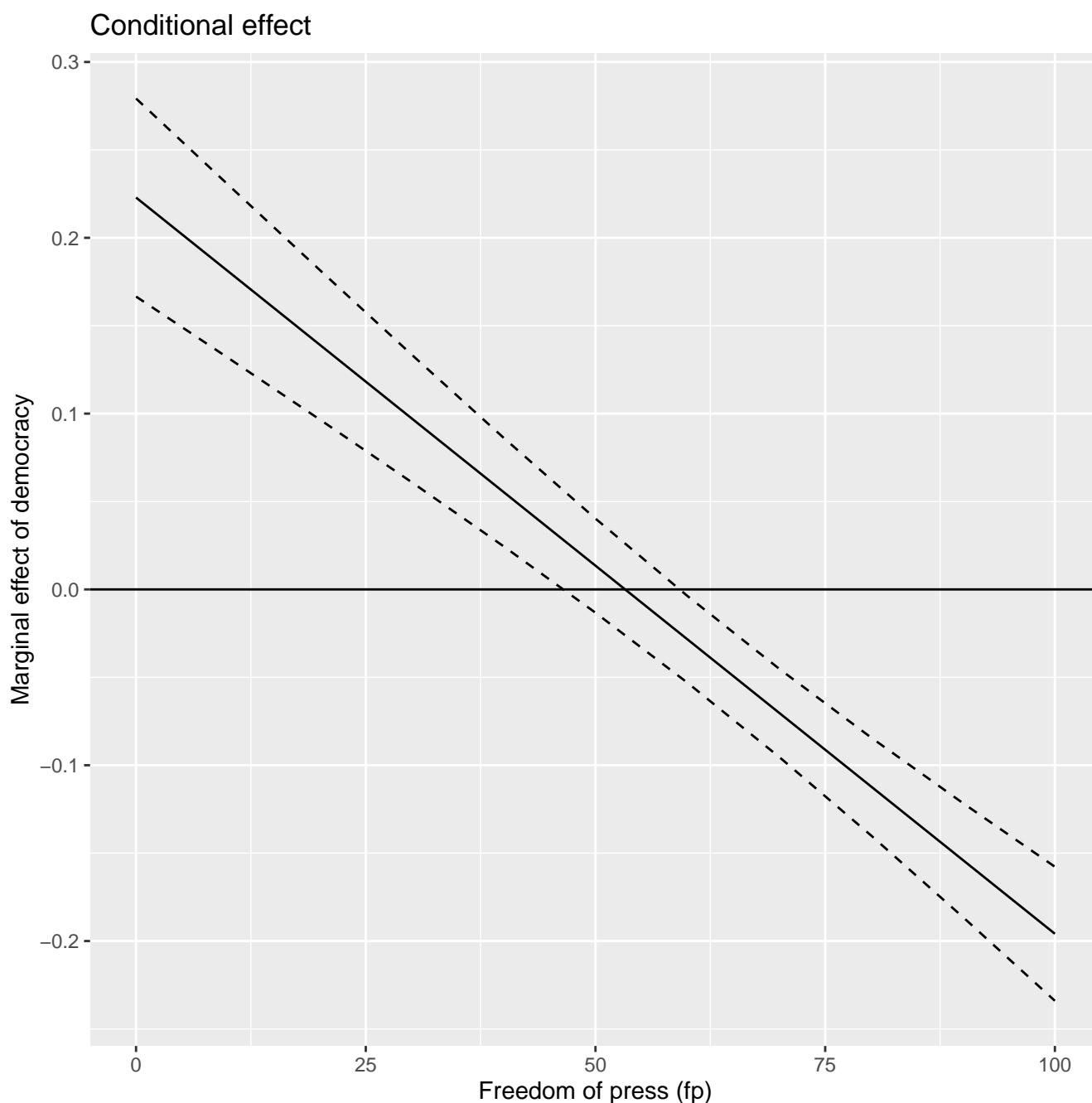
##	factor	fp	AME	SE	z	p	lower	upper
##	dem_c	0.0000	0.2229	0.0287	7.7560	0.0000	0.1666	0.2792
##	dem_c	10.0000	0.1810	0.0251	7.2030	0.0000	0.1318	0.2303
##	dem_c	20.0000	0.1391	0.0217	6.4150	0.0000	0.0966	0.1817
##	dem_c	30.0000	0.0973	0.0185	5.2548	0.0000	0.0610	0.1335
##	dem_c	40.0000	0.0554	0.0157	3.5172	0.0004	0.0245	0.0863
##	dem_c	50.0000	0.0135	0.0137	0.9890	0.3227	-0.0133	0.0403
##	dem_c	60.0000	-0.0284	0.0126	-2.2532	0.0242	-0.0530	-0.0037
##	dem_c	70.0000	-0.0702	0.0128	-5.4916	0.0000	-0.0953	-0.0452
##	dem_c	80.0000	-0.1121	0.0142	-7.8894	0.0000	-0.1400	-0.0843
##	dem_c	90.0000	-0.1540	0.0165	-9.3106	0.0000	-0.1864	-0.1216
##	dem_c	100.0000	-0.1959	0.0195	-10.0692	0.0000	-0.2340	-0.1578
##	fp	0.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	10.0000	-0.0570	0.0064	-8.9083	0.0000	-0.0695	-0.0444
##	fp	20.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	30.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	40.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	50.0000	-0.0570	0.0064	-8.9083	0.0000	-0.0695	-0.0444
##	fp	60.0000	-0.0570	0.0064	-8.9083	0.0000	-0.0695	-0.0444
##	fp	70.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	80.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	90.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444
##	fp	100.0000	-0.0570	0.0064	-8.9084	0.0000	-0.0695	-0.0444

В выдаче нас интересуют строки, в которых фактор равен 'dem\_c'. Мы также видим, что между  $50 < fp_i < 60$  предельный эффект демократии на коррупцию меняет знак и становится отрицательным (мы уже знаем, что это происходит при  $fp_i > 53.23$ . Теперь мы можем увидеть более детальную информацию про то, какое значение принимает предельный эффект в зависимости от разных значений  $fp$ . В максимальной точке  $fp_i = 100$  (таких примеров нет в датасете и, вероятно, не может быть), предельный эффект демократии равен  $-0.1959$ . Также можно заметить, что в некоторый момент (грубо: при  $fp$  около 50), предельный эффект будет незначимым. Ниже этот сюжет будет развит.

Поскольку мы работаем с двумя непрерывными переменными, мы можем построить **график**, на котором можно увидеть зависимость предельного эффекта уровня демократии на коррупцию от значений свободы прессы в более явном виде.

```
# график
ggplot(s, aes(x = fp)) +
  geom_line(aes(y = dydx_dem_c)) +
  geom_line(aes(y = dydx_dem_c+1.96*sqrt(Var_dydx_dem_c)), linetype = 2) +
  geom_line(aes(y = dydx_dem_c-1.96*sqrt(Var_dydx_dem_c)), linetype = 2) +
  geom_hline(yintercept = 0) +
  ggtitle("Conditional effect") +
```

```
xlab("Freedom of press (fp)") + ylab("Marginal effect of democracy")
```



Перед нами график, демонстрирующий предельный эффект уровня демократии (*dem*) на уровень коррупции (*cpi*) при разных значениях свободы прессы (*fp*), а также 95%-й доверительный интервал для предельного эффекта.

1. На нем мы явно видим, что **предельный эффект изменяется линейно** (становится меньше на 0.0041878).
2. Как уже говорилось, он **меняет знак** примерно после значения  $fp = 53.23$ .
3. Предельный эффект статистически значим при большей части значений, принимаемых индексом свободы прессы: ДИ накрывает значение „0“ примерно при значениях *fp* от 47 до 59. То есть **предельный эффект уровня демократии на коррупцию значим при значениях свободы прессы от 0 до 47 и от 59 до 100**.
4. Можно заметить, что **рамки 95%-го доверительного интервала ожидаемо расширяются ближе к концам графика**. Это означает, что в этих областях у нас меньше всего данных,

что логично, поскольку в данном массиве переменная  $fp$  принимает значения от 4.17 до 90.50. К тому же, в случае с минимальными значениями уровня демократии, мы можем вспомнить, что на графике было сразу несколько наблюдений с уровнем демократии, равному 0, либо около 0, но, при этом, с достаточно низким уровнем коррупции (возможно, это страны Персидского залива). Однако в целом **ДИ относительно узкий**.