Homework Assignment 4

Data Analysis & Statistics, Winter 2024/25

Lecturer: Dario Paape

Group Members:

Daniel Lösel
Martín E. Iribarren

1. Setup and Data Preparation

Loading Libraries

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(readr)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack
```

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x tidyr::pack()   masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
```

```r
library(here)
```

```
## here() starts at C:/Users/emart/Backup/Osnabrück/Jahre/Zweites Jahr/3. Drittes Semest
```

Loading Dataset

```r
dataset <- read.csv(here("data", "data_wdi", "data.csv"))
```

Data Pre-processing

```r
#Select Columns
dataset <- dataset %>%
  select(-Country.Code, -Series.Code)

data_long <- dataset %>%
  pivot_longer(
    cols = starts_with("X"),
    names_to = "Year",
    values_to = "Value"
  ) %>%
  mutate(Year = gsub("X|\\.YR", "", Year))
```

```r
data_clean <- data_long %>%
  filter(Series.Name %in% c("Gross fixed capital formation (% of GDP)",
                            "General government final consumption expenditure (% of GDP)
                            "GDP (current US$)",
                            "GDP growth (annual %)",
                            "Inflation, GDP deflator (annual %)",
                            "Control of Corruption: Estimate")) %>%


  pivot_wider(names_from = Series.Name, values_from = Value) %>%

  mutate(Year = gsub("[^0-9]", "", Year),
         Year = as.numeric(Year)) %>%

  rename(
    Total_Investment = "Gross fixed capital formation (% of GDP)",
    Government_Expenditure = "General government final consumption expenditure (% of GDP
    GDP = "GDP (current US$)",
    GDP_Growth = "GDP growth (annual %)",
    Inflation = "Inflation, GDP deflator (annual %)",
```

```
    Control_of_Corruption = "Control of Corruption: Estimate"
  )
```

```
data_clean <- data_clean %>%
  mutate(
    Control_of_Corruption = ifelse(Control_of_Corruption == "..", 0.0, Control_of_Corrup
    across(c(Total_Investment, Government_Expenditure, GDP, Inflation, Control_of_Corrup
  ) %>%
  mutate(GDP_Growth = as.numeric(GDP_Growth)) %>%
  drop_na()
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(...)`.
## Caused by warning:
## ! NAs introducidos por coerción
```

2. Exploratory Data Analysis

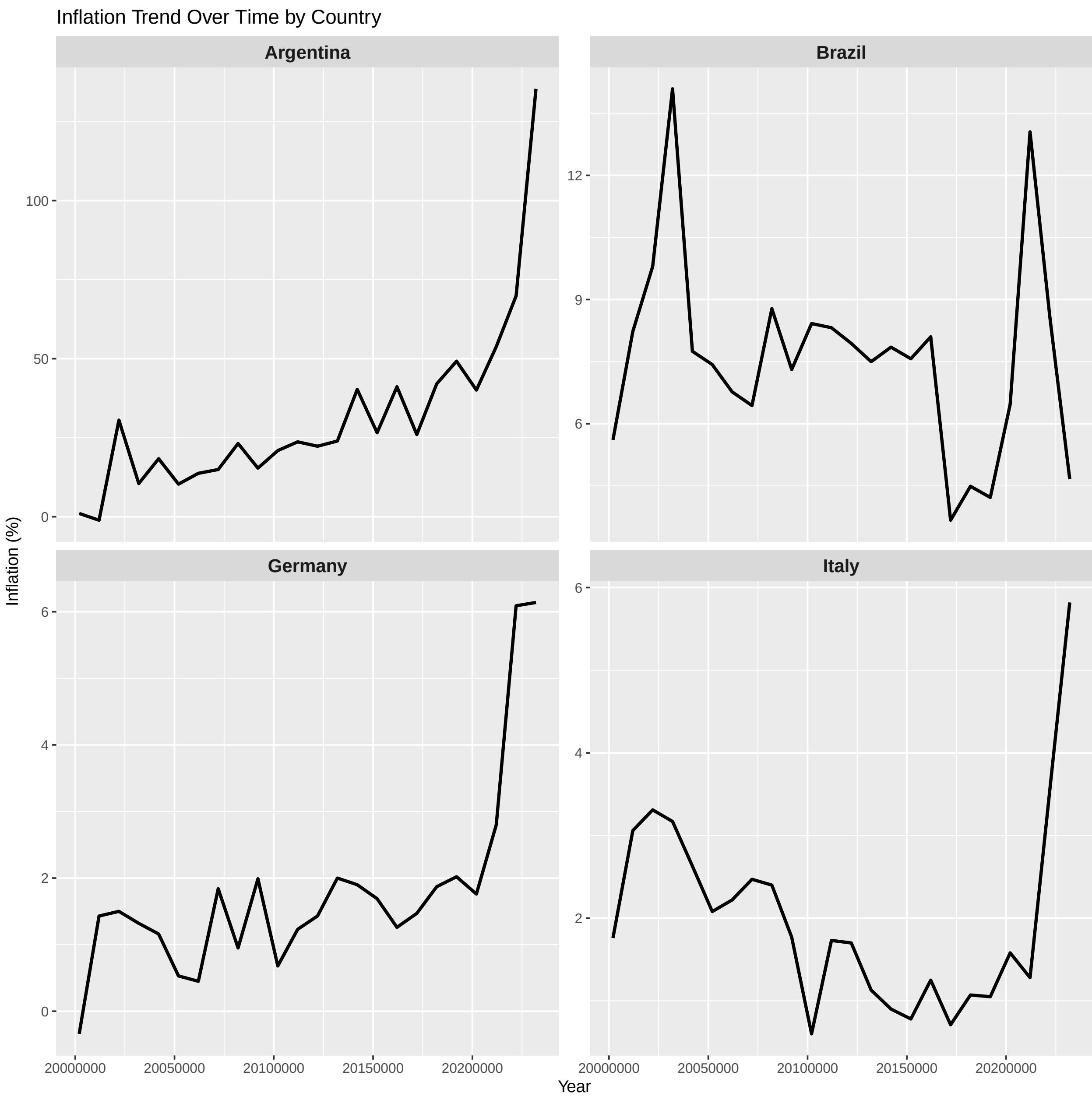
Inflation Distribution by Country

```r
inflation_data <- data_clean %>%
  filter(Country.Name %in% c("Germany", "Argentina", "Brazil", "Chile", "Italy")) %>%
  select(Country.Name, Year, Inflation) %>%
  group_by(Country.Name, Year) %>%
  summarise(Inflation = mean(Inflation, na.rm = TRUE)) %>%
  ungroup()
```

```
## `summarise()` has grouped output by 'Country.Name'. You can override using the
## `.groups` argument.
```

```r
ggplot(inflation_data, aes(x = Year, y = Inflation)) +
  geom_line(size = 1, color = "black") +
  facet_wrap(~ Country.Name, scales = "free_y", ncol = 2) +
  labs(title = "Inflation Trend Over Time by Country",
       x = "Year",
       y = "Inflation (%)") +
  theme(
    legend.position = "none",
    strip.text = element_text(size = 12, face = "bold"),  #
  )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Government Expenditure by Country (% of GDP)

```
investment_data <- data_clean %>%
  select(Country.Name, Year, Government_Expenditure) %>%
  group_by(Country.Name) %>%
  summarise(Average_Government_Expenditure = mean(Government_Expenditure, na.rm = TRUE))
  mutate(Average_Government_Expenditure = round(Average_Government_Expenditure, 2)) %>%
  arrange(desc(Average_Government_Expenditure)) %>%
  ungroup()


investment_data
```

```
## # A tibble: 20 x 2
##    Country.Name       Average_Government_Expenditure
##    <chr>                                       <dbl>
##  1 Netherlands                                 24.0
##  2 France                                      23.4
##  3 Saudi Arabia                                23.3
##  4 United Kingdom                              19.8
##  5 Germany                                     19.4
##  6 Australia                                   19.4
##  7 Brazil                                      19.2
##  8 Italy                                       19.2
##  9 Japan                                       19.1
## 10 Spain                                       18.8
## 11 Russian Federation                          17.9
## 12 China                                       15.7
## 13 Argentina                                   15.1
## 14 United States                               14.8
## 15 Korea, Rep.                                 14.7
## 16 Turkiye                                     13.6
## 17 Switzerland                                 11.3
## 18 Mexico                                      10.9
## 19 India                                       10.8
## 20 Indonesia                                    8.62
```

```r
corruption_data <- data_clean %>%
  select(Country.Name, Year, Control_of_Corruption) %>%
  group_by(Country.Name) %>%
  summarise(Average_Control_of_Corruption = mean(Control_of_Corruption, na.rm = TRUE)) %
  mutate(Average_Control_of_Corruption = round(Average_Control_of_Corruption, 2)) %>%
  arrange(desc(Average_Control_of_Corruption)) %>%
  ungroup()

corruption_data
```

```
## # A tibble: 20 x 2
##    Country.Name       Average_Control_of_Corruption
##    <chr>                                      <dbl>
##  1 Switzerland                                 1.96
##  2 Netherlands                                 1.91
##  3 Australia                                   1.79
##  4 Germany                                     1.73
##  5 United Kingdom                              1.69
##  6 Japan                                       1.34
##  7 United States                               1.31
##  8 France                                      1.27
##  9 Spain                                       0.92
```

```
## 10 Korea, Rep.                                    0.5 
## 11 Italy                                          0.31
## 12 Saudi Arabia                                   0.05
## 13 Turkiye                                       -0.16
## 14 Brazil                                        -0.21
## 15 China                                         -0.34
## 16 Argentina                                     -0.35
## 17 India                                         -0.38
## 18 Mexico                                        -0.58
## 19 Indonesia                                     -0.64
## 20 Russian Federation                            -0.92
```

3. Hypothesis Testing

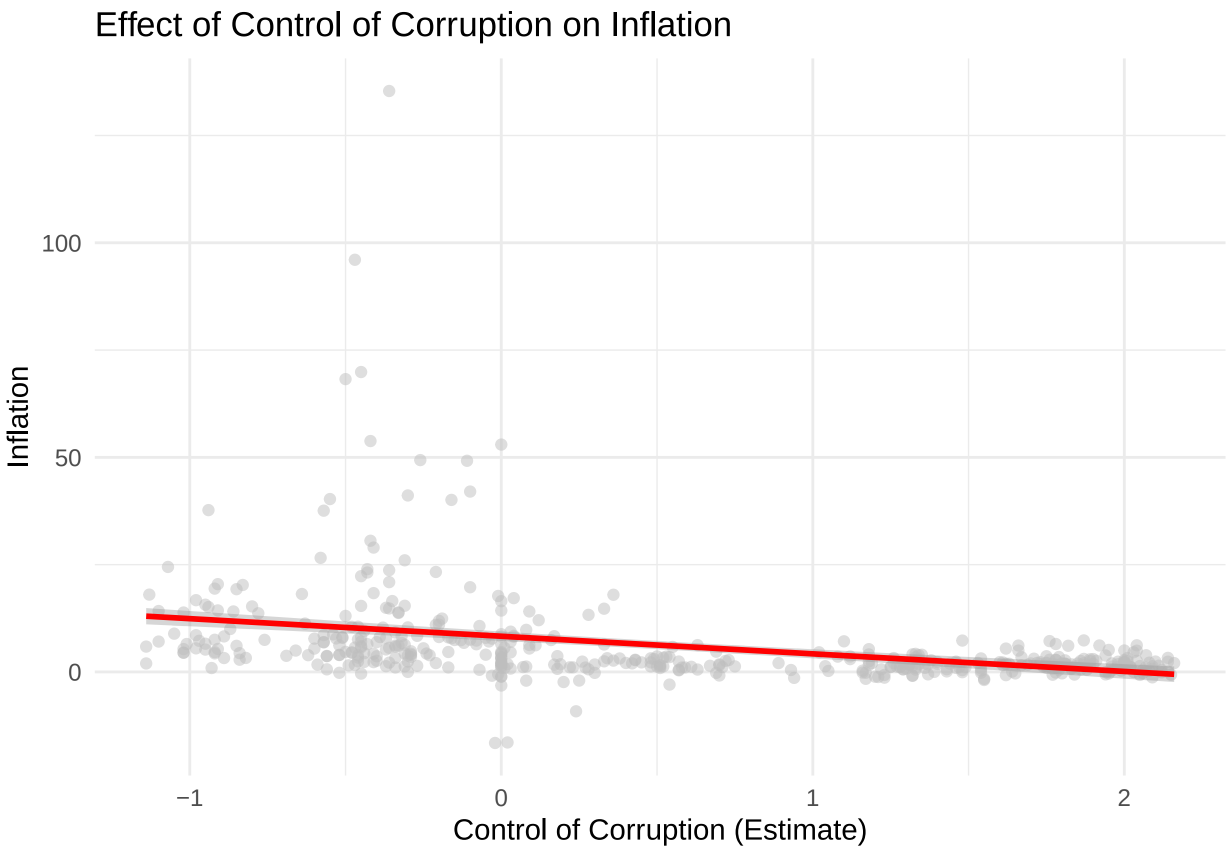
Hypothesis 1: Effect of Control of Corruption on Inflation

```
model_h1 <- lmer(Inflation ~ Control_of_Corruption + (1 | Country.Name),
                                    data = data_clean)
summary(model_h1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Inflation ~ Control_of_Corruption + (1 | Country.Name)
##    Data: data_clean
##
## REML criterion at convergence: 3495.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3502 -0.2161 -0.0378  0.1292 11.9185
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  Country.Name (Intercept) 41.54    6.445
##  Residual                 78.56    8.863
## Number of obs: 479, groups:  Country.Name, 20
##
## Fixed effects:
##                       Estimate Std. Error t value
## (Intercept)              7.780      1.605   4.849
## Control_of_Corruption   -3.167      1.032  -3.068
##
## Correlation of Fixed Effects:
##             (Intr)
## Cntrl_f_Crr -0.360
```

```
ggplot(data_clean, aes(x = Control_of_Corruption, y = Inflation)) +
  geom_point(alpha = 0.5, color = "gray") +
  geom_smooth(method = "lm", aes(group = 1), color = "red", size = 1, se = TRUE) +
  labs(title = "Effect of Control of Corruption on Inflation",
       x = "Control of Corruption (Estimate)",
       y = "Inflation") +
  theme_minimal()
```

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



Interpretation Hypothesis 1:

The model indicates an average inflation of 7.7%, when the control of corruption is zero. The coefficient for Control_of_Corruption is -3.167 with a t-value of -3.068, meaning that for each unit increase in the control of corruption, inflation decreases by approximately 3.167%.

This negative relationship is statistically significant, supporting the hypothesis that better control of corruption is associated with lower inflation. However, as we will observe in the next models, the random effects show high variability across countries, with a variance of 41.54 and a standard deviation of 6.445.

Hypothesis 2: The effect of Government expenditure on GDP growth

```
model_h2 <- lmer(GDP_Growth ~ Government_Expenditure + (1 + Government_Expenditure || Co
summary(model_h2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: GDP_Growth ~ Government_Expenditure + ((1 | Country.Name) + (0 +
##     Government_Expenditure | Country.Name))
##    Data: data_clean
##
```

```
## REML criterion at convergence: 2440.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.1452 -0.3173  0.0675  0.4657  3.1625
##
## Random effects:
##  Groups          Name                    Variance Std.Dev.
##  Country.Name    (Intercept)             7.666702 2.76888
##  Country.Name.1 Government_Expenditure 0.002129 0.04614
##  Residual                                8.349488 2.88955
## Number of obs: 479, groups:  Country.Name, 20
##
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)             14.80679    1.54547   9.581
## Government_Expenditure -0.70876    0.08423  -8.414
##
## Correlation of Fixed Effects:
##             (Intr)
## Gvrnmnt_Exp -0.905
```

Interpretation Hypothesis 2:

The model indicates a significant negative relationship between government expenditure and GDP growth, with a coefficient of -0.70876 (t-value = -8.414). This suggests that higher government expenditure is associated with lower GDP growth. In terms of fixed effects, when there is zero government expenditure, the model predicts a baseline GDP growth of 14%. Additionally, the residuals appear reasonable, and indicate that the model fit might not be perfect for all countries, suggesting that some country-specific factors could influence the relationship between government expenditure and GDP growth.

Hypothesis 3: Effect of Total Investment on GDP

```
model_h3 <- lmer(GDP_Growth ~ Total_Investment + (1 | Country.Name),
                 data = data_clean)
summary(model_h3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: GDP_Growth ~ Total_Investment + (1 | Country.Name)
##    Data: data_clean
##
## REML criterion at convergence: 2478.7
##
```

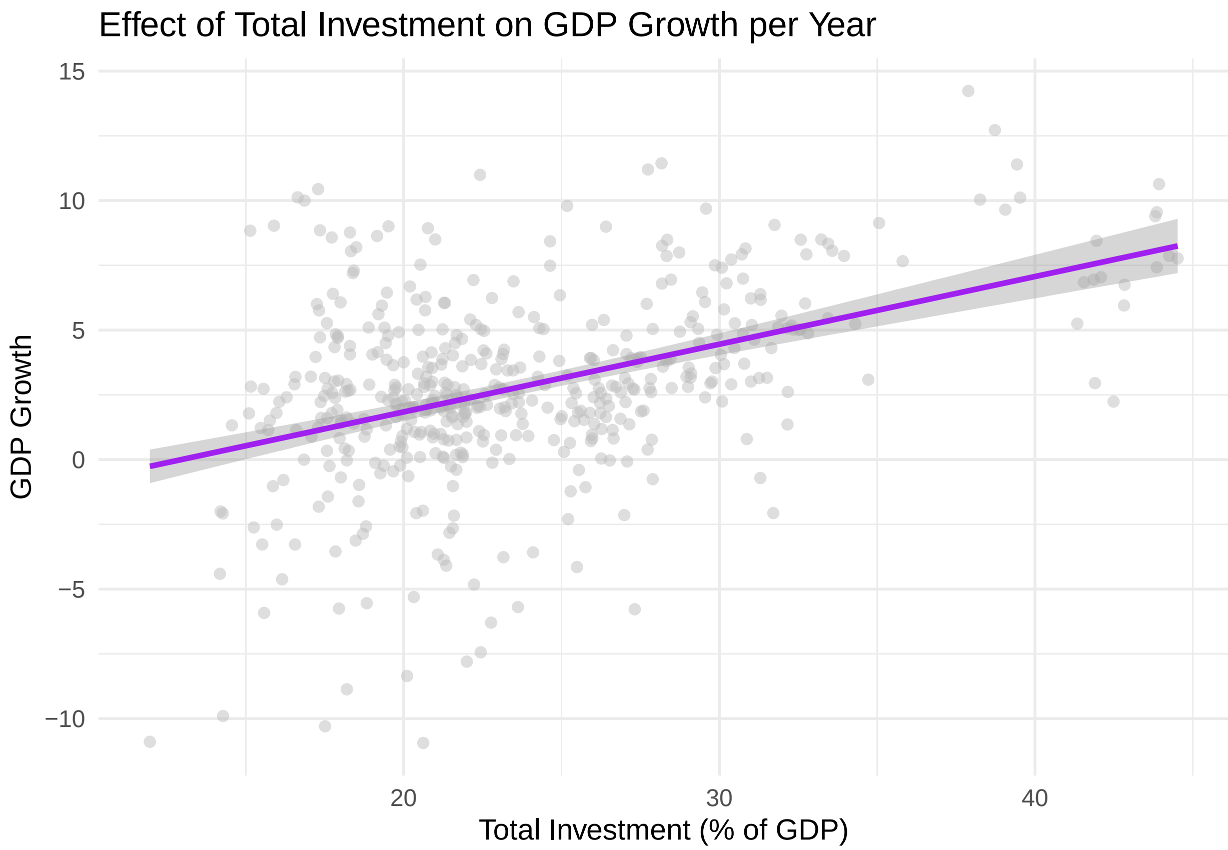
```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9789 -0.3498  0.0683  0.4547  2.7434
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  Country.Name (Intercept) 1.090    1.044
##  Residual                 9.729    3.119
## Number of obs: 479, groups:  Country.Name, 20
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)      -2.70131    0.95450  -2.830
## Total_Investment  0.23245    0.03837   6.058
##
## Correlation of Fixed Effects:
##             (Intr)
## Ttl_Invstmn -0.958
```

```r
gdp_investment <- data_clean %>%
  group_by(Country.Name, Year) %>%
  summarise(Total_Investment = mean(Total_Investment),
            GDP_Growth = mean(GDP_Growth))
```

```
## `summarise()` has grouped output by 'Country.Name'. You can override using the
## `.groups` argument.
```

```r
ggplot(gdp_investment, aes(x = Total_Investment, y = GDP_Growth)) +
  geom_point(alpha = 0.5, color = "gray") +
  geom_smooth(method = "lm", aes(group = 1), color = "purple", size = 1, se = TRUE) +
  labs(title = "Effect of Total Investment on GDP Growth per Year",
       x = "Total Investment (% of GDP)",
       y = "GDP Growth") +
  theme_minimal()
```

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



Interpretation Hypothesis 3: The model shows a positive correlation between total investment and GDP growth (coefficient of 0.23245, t = 6.058). The variability in the random effects suggests that some countries exhibit atypical behavior compared to the rest. This can be understood in light of the high negative correlation between the intercept and Total_Investment (-0.958), which could indicate differences in the initial growth levels between countries with varying investment levels.