GV207 2022

Gabriel Edwards

2022-11-17

Table of Contents

# Load Packages

if(!require(pacman))install.packages("pacman")  
  
pacman::p\_load(  
 tidyverse,  
 scales,  
 janitor  
)

# Questions

## Q1. Load Data

world.data <- read\_csv("world.csv")

## Q2. Categorical variable analysis

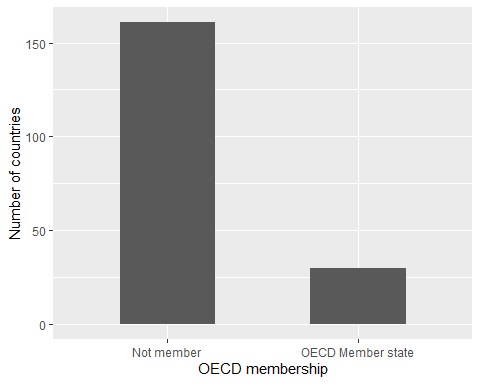
ft.oecd <- tabyl(world.data, oecd) %>%   
 as.data.frame() %>%   
 rename(OECD = oecd,  
 Freq = n,  
 Percentage = percent)

## Q3.

* (A). **30** countries in the data are OECD members.
* (B). **161** countries in the data are not OECD members.
* (C). About **16%** of the countries in the data are OECD members.
* (D). About **84%** of the countries in the data are not OECD members.

## Q4. Bar Graph

ggplot(data = ft.oecd, aes(x = OECD, y = Freq)) +  
 geom\_bar(stat = "identity", width = 0.5) +  
 labs(x = "OECD membership",  
 y = "Number of countries")



## Q5. per capita GDP Summary

summary(world.data$gdp\_10\_thou)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
 0.0090 0.0503 0.1897 0.6018 0.6320 4.7354 14

sd(world.data$gdp\_10\_thou, na.rm = TRUE)

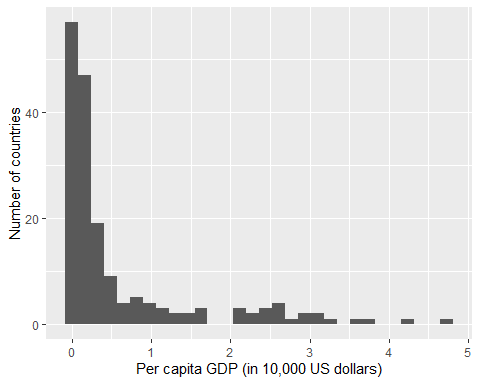
[1] 0.9433982

## Q6. Skewness

* (B). Positively skewed (skewed to the right).

## Q7. describe per capita GDP graphically

ggplot(data = world.data, aes(gdp\_10\_thou)) +  
 geom\_histogram() +  
 labs(x = "Per capita GDP (in 10,000 US dollars)",  
 y = "Number of countries")



## Q8. The standard error of mean

sd(world.data$gdp\_10\_thou, na.rm = TRUE)/sqrt(177)

[1] 0.07091015

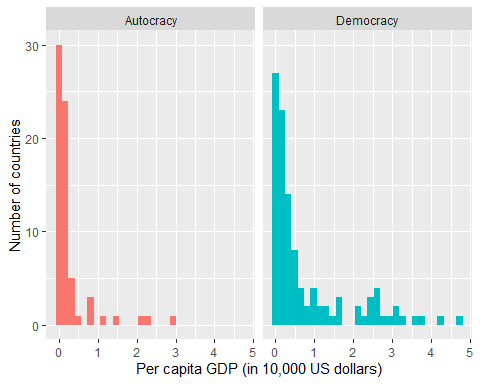
## Q9. 95% CI

lower <- mean(world.data$gdp\_10\_thou, na.rm = TRUE) - 1.96\*sd(world.data$gdp\_10\_thou, na.rm = TRUE)/sqrt(177)  
  
upper <- mean(world.data$gdp\_10\_thou, na.rm = TRUE) + 1.96\*sd(world.data$gdp\_10\_thou, na.rm = TRUE)/sqrt(177)  
  
list("Lower CI" = lower,  
 "Upper CI" = upper)

$`Lower CI`  
[1] 0.4628347  
  
$`Upper CI`  
[1] 0.7408025

## Q10. Histogram for democracy and non-democracy per capita GDP

dem.gdp <- world.data %>%   
 filter(!is.na(gdp\_10\_thou)) %>%   
 mutate(dem.dum = if\_else(democ\_regime == "Yes", "Democracy", "Autocracy"))  
  
  
ggplot(data = dem.gdp, aes(x = gdp\_10\_thou, fill = dem.dum), color = "white") +  
 geom\_histogram(show.legend = FALSE) +  
 facet\_wrap(dem.dum ~ .) +  
 labs(x = "Per capita GDP (in 10,000 US dollars)",  
 y = "Number of countries")



## Q11. mean and CI for democracy

dem.gdp %>%   
 filter(dem.dum == "Democracy") %>%   
 group\_by(dem.dum) %>%   
 summarise(Freq = n(),  
 Mean = mean(gdp\_10\_thou),  
 stdev = sd(gdp\_10\_thou)) %>%   
 mutate(lower.95.CI = Mean - 1.96\*stdev/sqrt(Freq),  
 upper.95.CI = Mean + 1.96\*stdev/sqrt(Freq)) %>%   
 select(dem.dum, Mean, lower.95.CI, upper.95.CI)

# A tibble: 1 × 4  
 dem.dum Mean lower.95.CI upper.95.CI  
 <chr> <dbl> <dbl> <dbl>  
1 Democracy 0.801 0.598 1.00

## Q12. mean nd CI for non-democracy

dem.gdp %>%   
 filter(dem.dum == "Autocracy") %>%   
 group\_by(dem.dum) %>%   
 summarise(Freq = n(),  
 Mean = mean(gdp\_10\_thou),  
 stdev = sd(gdp\_10\_thou)) %>%   
 mutate(lower.95.CI = Mean - 1.96\*stdev/sqrt(Freq),  
 upper.95.CI = Mean + 1.96\*stdev/sqrt(Freq)) %>%   
 select(dem.dum, Mean, lower.95.CI, upper.95.CI)

# A tibble: 1 × 4  
 dem.dum Mean lower.95.CI upper.95.CI  
 <chr> <dbl> <dbl> <dbl>  
1 Autocracy 0.282 0.155 0.409