

## Problem 10.2

Beenish Ali, Ruba Khan, Zara Shamim

12/29/22

### Part (a)

Is the data set a balanced panel?

*The dataset is unbalanced because data are available over different years for different countries.*

### Part (b)

i. What are the minimum and maximum values of dem\_ind in the data set? What are the mean and standard deviation of dem\_ind in the data set? What are the 10th, 25th, 50th, 75th, and 90th percentiles of its distribution?

```
summary(dem_ind)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.0000	0.1667	0.5000	0.4991	0.8333	1.0000	103

```
sd(dem_ind, na.rm=TRUE)
```

```
[1] 0.3713367
```

```
quantile(dem_ind, probs=c(0.1, 0.25, 0.5, 0.75, 0.9), na.rm=TRUE)
```

10%	25%	50%	75%	90%
0.0000000	0.1666667	0.5000000	0.8333333	1.0000000

(ii) What is the value of dem\_ind for the United States in 2000? Averaged over all years in the data set?

```
us2000 <- subset(inde, country=="United States" & year=="2000")

us2000$dem_ind # dem_ind for US in 2000
```

```
[1] 1
```

```
us <- subset(inde, country=="United States") # US data 1960-2000

mean(us$dem_ind) # averaged US dem_ind over 1960-2000
```

```
[1] 0.9855556
```

*The value for the U.S. in 2000 is dem\_ind=1.0. The average for the nine years in the sample is 0.99.*

- (iii) What is the value of dem\_ind for Libya in 2000? Averaged over all years in the data set?

```
libya2000 <- subset(inde, country=="Libya" & year=="2000") # Libya data 2000
libya2000$dem_ind # dem_ind for Libya in 2000
```

```
[1] 0
```

```
libya <- subset(inde, country=="Libya") # Libya data 1960-2000

mean(libya$dem_ind) # averaged Libya dem_ind over 1960-2000
```

```
[1] 0.1092593
```

*The value for Libya in 2000 is dem\_ind=0.0. The average for the nine years in the sample is 0.11.*

- (iv) List five countries with an average value of dem\_ind greater than 0.95; less than 0.10; and between 0.3 and 0.7.

```
#(b.iv) Countries with low, mid, and high dem_ind
# (b.iv.1) Remove rows with NA in Column "dem_ind"
```

```

inde.c <- inde[complete.cases(inde[, c('dem_ind')]), ]

# (b.iv.2) Compute average dem_ind for each country
inde.c$ave_dem <- ave(inde.c$dem_ind, inde.c$country)

# (b.iv.3) Countries with high averaged dem_ind: ave_dem>0.95
list.high <- subset(inde.c, ave_dem>0.95)$country
unique(list.high)

```

[1] "Australia"	"Austria"	"Belgium"
[4] "Belize"	"Barbados"	"Canada"
[7] "Switzerland"	"Costa Rica"	"Czech Republic"
[10] "Germany"	"Germany, West"	"Denmark"
[13] "France"	"United Kingdom"	"Ireland"
[16] "Iceland"	"Italy"	"Japan"
[19] "Kiribati"	"St. Kitts and Nevis"	"St. Lucia"
[22] "Lithuania"	"Luxembourg"	"Malta"
[25] "Netherlands"	"Norway"	"New Zealand"
[28] "Slovakia"	"Slovenia"	"Sweden"
[31] "United States"		

```

# (b.iv.4) Countries with low averaged dem_ind: ave_dem<0.1
list.low <- subset(inde.c, ave_dem<0.1)$country
unique(list.low)

```

[1] "Afghanistan"	"Angola"	"Burundi"
[4] "Brunei"	"China"	"Cuba"
[7] "Germany, East"	"Eritrea"	"Equatorial Guinea"
[10] "Iraq"	"Myanmar"	"Korea, Dem. Rep."
[13] "Rwanda"	"Saudi Arabia"	"Turkmenistan"
[16] "Uzbekistan"	"Vietnam"	"Congo, Dem. Rep."

```

#(b.iv.5) Countries with mid averaged dem_ind: 0.3<ave_dem<0.7
list.mid <- subset(inde.c, 0.3<ave_dem & ave_dem<0.7)$country
unique(list.mid)

```

[1] "Argentina"	"Armenia"	"Antigua"
[4] "Bangladesh"	"Bulgaria"	"Bosnia and Herzegovina"
[7] "Bolivia"	"Brazil"	"Chile"

[10] "Comoros"	"Cape Verde"	"Dominican Republic"
[13] "Ecuador"	"Spain"	"Ethiopia 1993-"
[16] "Fiji"	"Georgia"	"Ghana"
[19] "Gambia, The"	"Guinea-Bissau"	"Guatemala"
[22] "Guyana"	"Honduras"	"Hungary"
[25] "Jordan"	"Korea, Rep."	"Kuwait"
[28] "Lebanon"	"Lesotho"	"Morocco"
[31] "Madagascar"	"Maldives"	"Mexico"
[34] "Macedonia, FYR"	"Mozambique"	"Malaysia"
[37] "Nigeria"	"Nicaragua"	"Nepal"
[40] "Pakistan-post-1972"	"Pakistan-pre-1972"	"Panama"
[43] "Peru"	"Philippines"	"Poland"
[46] "Paraguay"	"Russia"	"Senegal"
[49] "Singapore"	"El Salvador"	"Sao Tome and Principe"
[52] "Suriname"	"Seychelles"	"Thailand"
[55] "Tonga"	"Turkey"	"Taiwan"
[58] "Ukraine"	"Yemen"	"Yugoslavia - post 1991"
[61] "South Africa"	"Zambia"	"Zimbabwe"

## Part (c)

Regress `dem_ind` on `log_gdppc`. Use standard errors that are clustered by country.

*To obtain the clustered standard errors, use command `vcovCL`. Note that the default “type” in “`vcovCL`” is “HC1” for `lm` objects and “HC0” otherwise.*

```
# OLS regression with clustered standard errors
fit.c <- lm(dem_ind~log_gdppc, data=inde)
summary(fit.c) # results with unclustered standard errors
```

Call:

```
lm(formula = dem_ind ~ log_gdppc, data = inde)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.72854	-0.19534	0.02586	0.19123	0.72698

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1.354828	0.070919	-19.10	<2e-16 ***
log_gdppc	0.235673	0.008626	27.32	<2e-16 ***

---

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

Residual standard error: 0.2719 on 956 degrees of freedom

(411 observations deleted due to missingness)

Multiple R-squared: 0.4385, Adjusted R-squared: 0.4379

F-statistic: 746.5 on 1 and 956 DF, p-value: < 2.2e-16