

# SOK-2011 v.2024: Utfordring 2 - kode

89

2024-03-19

Dette dokumentet inneholder kodene til figurene i utfordring 2, i emnet SOK-2011. Kodene tar utgangspunkt i støtten gitt i oppgaveteksten.

## Forberedelser

```
# Nullstiller minnet.
rm(list = ls())

# Laster inn pakker.
suppressPackageStartupMessages(library(tidyverse))
suppressPackageStartupMessages(library(WDI))
suppressPackageStartupMessages(library(countrycode))
suppressPackageStartupMessages(library(ggrepel))
suppressPackageStartupMessages(library(scales))

# Definerer egendefinert tema.
theme_mod <- function() {
  theme_minimal() +
    theme(
      plot.background = element_rect(fill = "white", color = NA),
      plot.margin = unit(c(5, 10, 5, 10), units = "mm"),
      plot.title = element_text(face = "bold.italic", color = "gray10"),
      axis.title = element_text(face = "bold", color = "gray10"),
      axis.text = element_text(color = "gray10"),
      legend.text = element_text(color = "gray10"),
      legend.title = element_text(face = "bold", color = "gray10"),
      panel.grid = element_line(color = "gray60"),
      axis.text.x = element_text(angle = 0, vjust = 1, hjust = 1)
    )
}
```

```
) }
```

## Utfordring 2.2

```
# Laster inn data.
df_01 <- WDI(
  country = "all",
  indicator = c("gdppc" = "NY.GDP.PCAP.PP.KD"),
  start = 1999,
  end = 2019,
  extra = TRUE)

df_02 <- WDI(
  country = "all",
  indicator = c("savings" = "NY.ADJ.NNAT.GN.ZS", "popgrowth" = "SP.POP.GROW"),
  start = 2015,
  end = 2019,
  extra = TRUE)

df_04 <- WDI(
  country = "all",
  indicator = c("hci_index" = "HD.HCI.OVRL"),
  start = 2015,
  end = 2019,
  extra = TRUE)

df_05 <- WDI(
  country = "all",
  indicator = c("nry" = "NY.ADJ.DRES.GN.ZS"),
  start = 2010,
  end = 2015,
  extra = TRUE)

# Definerer relevant tidsperiode.
start_year_gdp_01 <- 2015
start_year_gdp_02 <- 1999
end_year_gdp_01 <- 2019
end_year_gdp_02 <- 2005
```

## Oppgave B)

### NIVÅ og VEKST i BNP per innbygger mellom 2015 - 2019

```
# NIVÅ og VEKST i BNP per innbygger mellom 2015 - 2019.
# Filtrerer data.
df_01_subset_01 <- df_01[df_01$year %in% start_year_gdp_01:end_year_gdp_01, ]

# Lager et datasett med informasjon om gj.snittlig nivå på, og vekst i BNP per innbygger.
df_01_gdp_01 <- df_01_subset_01 %>%
  filter(iso3c != "", income != "Aggregates") %>%
  mutate(year = as.numeric(year)) %>%
  select(country, region, income, year, gdppc) %>%
  drop_na(gdppc) %>%
  group_by(country) %>%
  arrange(country, year) %>%
  mutate(avg_gdppc2019 = mean(gdppc),
         gdpgrowth = (log(gdppc) - lag(log(gdppc))) * 100,
         gy = mean(gdpgrowth, na.rm = TRUE)) %>%
  arrange(desc(year)) %>%
  slice(1) %>%
  select(-gdppc,
         -gdpgrowth) %>%
  ungroup()
```

### Initielt NIVÅ på BNP per innbygger

```
# Filtrerer data.
df_01_subset_02 <- df_01[df_01$year %in% start_year_gdp_02:end_year_gdp_02, ]

# Lager et datasett med initialt nivå på BNP per innbygger.
df_01_gdp_02 <- df_01_subset_02 %>%
  filter(iso3c != "", income != "Aggregates") %>%
  mutate(year = as.numeric(year)) %>%
  select(country, year, gdppc) %>%
  drop_na(gdppc) %>%
  group_by(country) %>%
  arrange(country, year) %>%
  filter(year == min(year)) %>%
  select(country, gdppc_0 = gdppc, year_0 = year) %>%
  mutate(ln_gdppc_0 = log(gdppc_0)) %>%
```

```
ungroup()
```

## Kombinerer og rydder datasettene

```
# Lager et datasett som kombinerer datasettene ovenfor.
df_gdp <- df_01_gdp_01 %>%
  left_join(df_01_gdp_02, by = "country")

# Rydder i datasettet.
colnames(df_gdp) <- c("country", "Region", "Income", "Year_2019", "Avg_GDP_Per_Cap_2015_20
                    "Avg_GDP_Per_cap_1999_2005", "Year_1999", "LN_GDP_Per_cap_1999_2005")

# Fjerner land som ikke har data fra intervallet 1999-2005.
df_gdp <- df_gdp %>%
  filter(Year_1999 == 1999)
```

## Beregning av variabler knyttet til sparing og befolkningsvekst

```
# Lager et datasett med info. om gj.snittlig nivå på sparing.
df_savings <- df_02 %>%
  filter(iso3c != "", income != "Aggregates") %>%
  mutate(year = as.numeric(year)) %>%
  select(country, year, savings) %>%
  drop_na(savings) %>%
  group_by(country) %>%
  arrange(country, year) %>%
  mutate(AVG_Savings = mean(savings)) %>%
  slice(1) %>%
  select(country, AVG_Savings) %>%
  ungroup()

# Laget er datasett med info. om gj.snittlig nivå på befolkningsvekst.
df_pop <- df_02 %>%
  filter(iso3c != "", income != "Aggregates") %>%
  mutate(year = as.numeric(year)) %>%
  select(country, year, popgrowth) %>%
  drop_na(popgrowth) %>%
  group_by(country) %>%
  arrange(country, year) %>%
```

```
mutate(AVG_popgrowth = mean(popgrowth)) %>%
slice(1) %>%
select(country, AVG_popgrowth) %>%
ungroup()
```

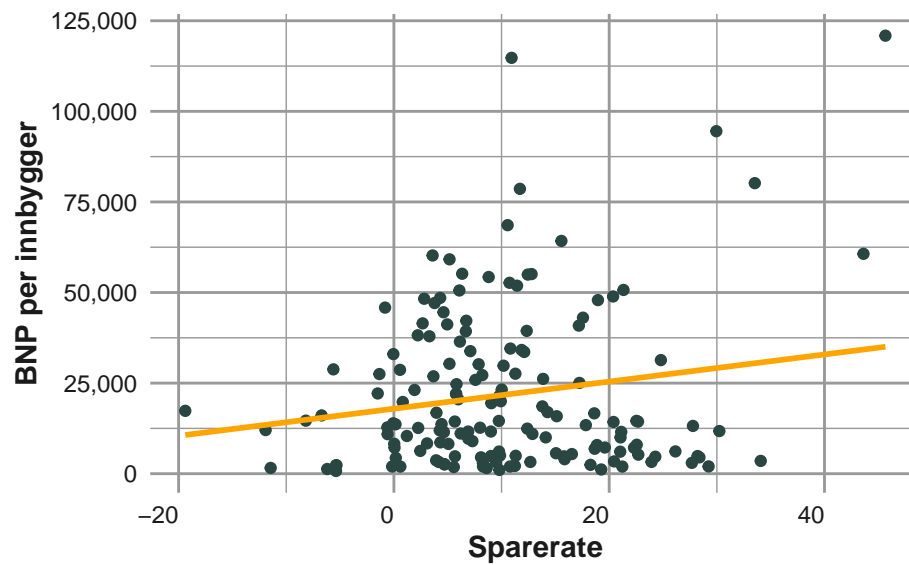
## Kombinerer og rydder datasettene

```
# Kombinerer datasettene.
df_03 <- df_gdp %>%
  left_join(df_savings, by = "country") %>%
  left_join(df_pop, by = "country")
```

## Figur 2.1

```
# Lager figur som viser spareraten og BNP per innbygger.
fig_01 <- ggplot(df_03, aes(x = AVG_Savings, y = Avg_GDP_Per_Cap_2015_2019)) +
  geom_point(color = "#294643") +
  geom_smooth(method = lm, se = FALSE, color = "orange") +
  labs(x = "Sparerate", y = "BNP per innbygger",
       title = " ") +
  theme_mod() +
  scale_y_continuous(labels = comma)

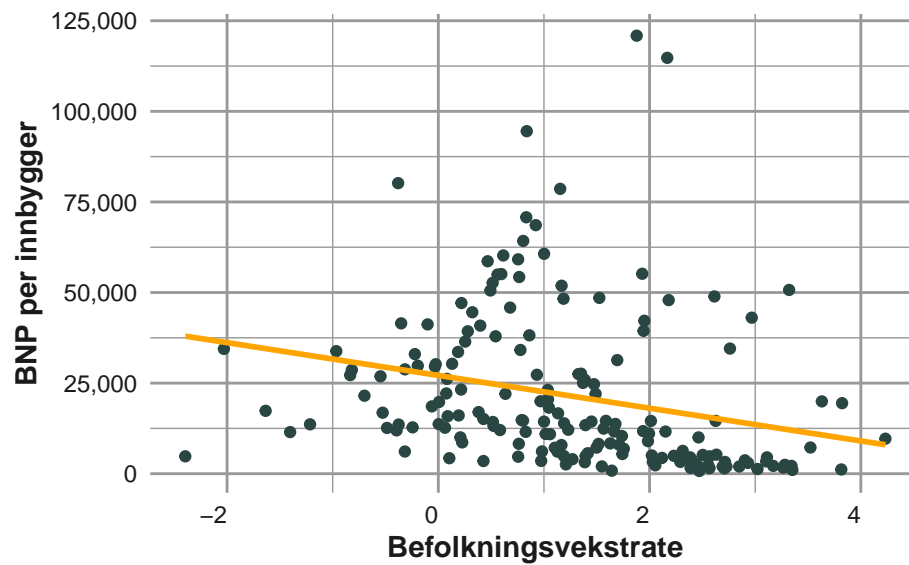
# Viser figuren.
fig_01
```



**Figur 2.2**

```
# Lager figur som viser befolkningsvekstraten og BNP per innbygger.
fig_02 <- ggplot(df_03, aes(x = AVG_popgrowth, y = Avg_GDP_Per_Cap_2015_2019)) +
  geom_point(color = "#294643") +
  geom_smooth(method = lm, se = FALSE, color = "orange") +
  labs(x = "Befolkningsvekstrate", y = "BNP per innbygger",
       title = " ") +
  theme_mod() +
  scale_y_continuous(labels = comma)

# Viser figuren.
fig_02
```



## Regresjonsmodell

```
# Setter opp regresjonsmodellen.
reg_01 <- lm(Avg_GDP_Per_Cap_2015_2019 ~ AVG_Savings + AVG_popgrowth,
             data = df_03)

# Viser resultat.
summary(reg_01)
```

Call:

```
lm(formula = Avg_GDP_Per_Cap_2015_2019 ~ AVG_Savings + AVG_popgrowth,
    data = df_03)
```

Residuals:

Min	1Q	Median	3Q	Max
-34528	-12665	-6163	7499	97510

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	23395.8	2755.0	8.492	1.5e-14	***
AVG_Savings	514.9	168.1	3.063	0.002585	**

```
AVG_popgrowth -5429.8      1408.0  -3.856 0.000168 ***
```

```
---
```

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

```
Residual standard error: 20930 on 156 degrees of freedom  
(19 observations deleted due to missingness)
```

```
Multiple R-squared:  0.114, Adjusted R-squared:  0.1026
```

```
F-statistic: 10.04 on 2 and 156 DF,  p-value: 7.942e-05
```

## Utfordring 2.3

### Beregning av variabler knyttet til kvalitet på arbeid

```
# Lager et datasett med info. om gj.snittlig utdanningsnivå.  
df_educ <- df_04 %>%  
  filter(iso3c != "") %>%  
  mutate(year = as.numeric(year)) %>%  
  select(country, year, hci_index) %>%  
  drop_na(hci_index) %>%  
  group_by(country) %>%  
  arrange(country, year) %>%  
  mutate(hci = mean(hci_index)) %>%  
  slice(1) %>%  
  select(country, hci) %>%  
  ungroup()
```

### Beregning av variabler knyttet til forbruk av naturressurser

```
# Lager et datasett med info. om gj.snittlig naturressurs forbruk.  
df_natural <- df_05 %>%  
  filter(iso3c != "", income != "Aggregates") %>%  
  mutate(year = as.numeric(year)) %>%  
  select(country, year, nry) %>%  
  drop_na(nry) %>%  
  group_by(country) %>%  
  arrange(country, year) %>%  
  mutate(u = mean(nry)) %>%  
  slice(1) %>%  
  select(country, u) %>%
```



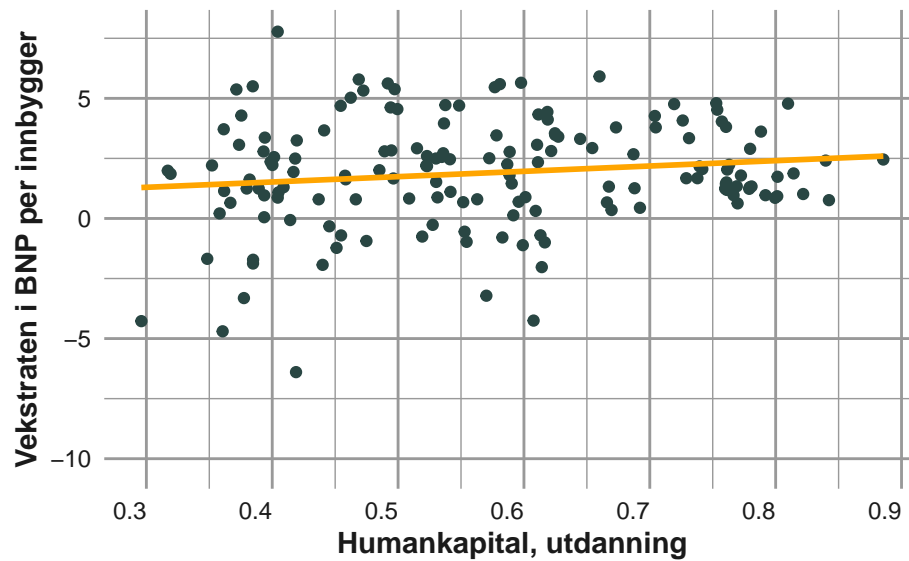
```
ungroup()
```

### Setter sammen datasettene

```
# Setter sammen datasettene.  
df_06 <- df_03 %>%  
  left_join(df_educ, by = "country") %>%  
  left_join(df_natural, by = "country")
```

### Figur 3.1

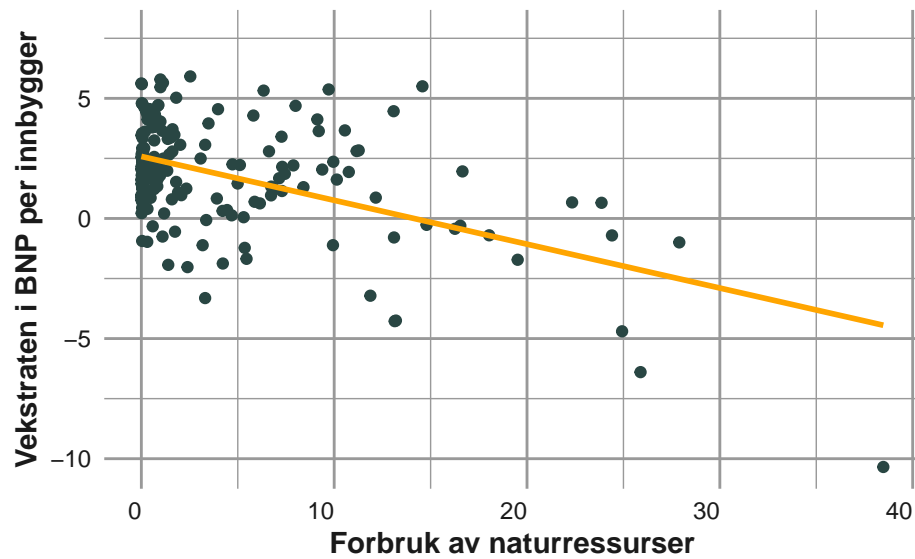
```
# Lager figur som viser befolkningsvekstraten og BNP per innbygger.  
fig_03 <- ggplot(df_06, aes(x = hci, y = Avg_Yearly_Growth)) +  
  geom_point(color = "#294643") +  
  geom_smooth(method = lm, se = FALSE, color = "orange") +  
  labs(x = "Humankapital, utdanning", y = "Vekstraten i BNP per innbygger",  
       title = " ") +  
  theme_mod() +  
  scale_x_continuous(labels = comma)  
  
# Viser figuren.  
fig_03
```



**Figur 3.2**

```
# Lager figur som viser befolkningsvekstraten og BNP per innbygger.
fig_04 <- ggplot(df_06, aes(x = u, y = Avg_Yearly_Growth)) +
  geom_point(color = "#294643") +
  geom_smooth(method = lm, se = FALSE, color = "orange") +
  labs(x = "Forbruk av naturressurser", y = "Vekstraten i BNP per innbygger",
       title = " ") +
  theme_mod() +
  scale_x_continuous(labels = comma)

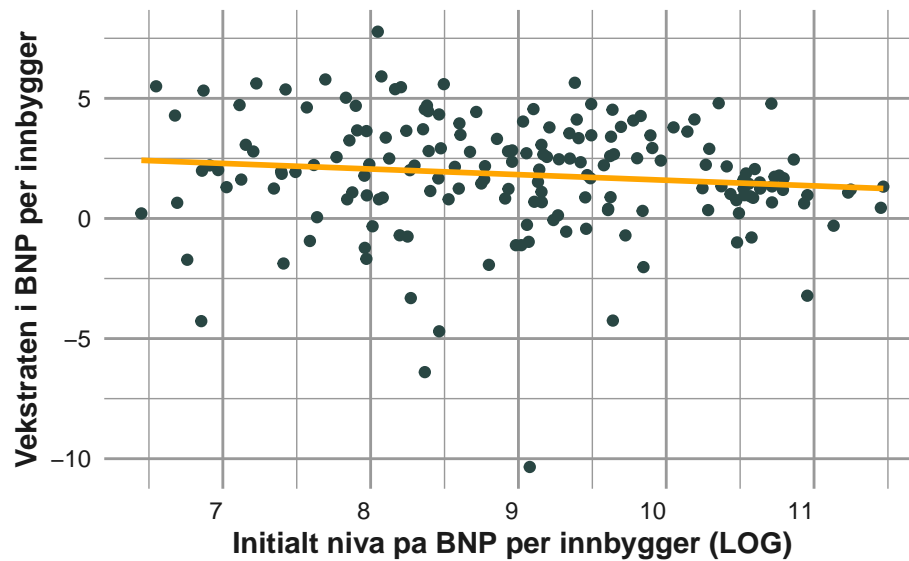
# Viser figuren.
fig_04
```



**Figur 3.3**

```
# Lager figur som viser befolkningsvekstraten og BNP per innbygger.
fig_05 <- ggplot(df_06, aes(x = LN_GDP_Per_cap_1999_2005,
                           y = Avg_Yearly_Growth)) +
  geom_point(color = "#294643") +
  geom_smooth(method = lm, se = FALSE, color = "orange") +
  labs(x = "Initialt niva pa BNP per innbygger (LOG)",
       y = "Vekstraten i BNP per innbygger",
       title = " ") +
  theme_mod() +
  scale_x_continuous(labels = comma)

# Viser figuren.
fig_05
```



## Regresjonsmodell

```
# Setter opp regresjonsmodellen.
reg_02 <- lm(Avg_Yearly_Growth ~ AVG_popgrowth + AVG_Savings + hci + u +
             LN_GDP_Per_cap_1999_2005, data = df_06)

# Viser regresjonsmodellen.
summary(reg_02)
```

Call:

```
lm(formula = Avg_Yearly_Growth ~ AVG_popgrowth + AVG_Savings +
    hci + u + LN_GDP_Per_cap_1999_2005, data = df_06)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.0422	-0.8118	-0.0140	1.0557	4.6624

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	9.98230	1.34684	7.412	1.37e-11 ***
AVG_popgrowth	-0.54154	0.17568	-3.082	0.002504 **

```

AVG_Savings          0.03068    0.01625    1.888 0.061208 .
hci                  6.78760    2.10116    3.230 0.001563 **
u                   -0.10629    0.02835   -3.749 0.000265 ***
LN_GDP_Per_cap_1999_2005 -1.24132    0.22511   -5.514 1.80e-07 ***
---
```

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

Residual standard error: 1.722 on 131 degrees of freedom

(41 observations deleted due to missingness)

Multiple R-squared: 0.4053, Adjusted R-squared: 0.3826

F-statistic: 17.85 on 5 and 131 DF, p-value: 1.804e-13

## Utfordring 2.1

```

# Velger variablene jeg trenger.
table_01 <- df_06 %>%
  select(Avg_GDP_Per_Cap_2015_2019, Avg_GDP_Per_cap_1999_2005, Avg_Yearly_Growth,
         AVG_Savings, AVG_popgrowth, hci, u)

# Bytter ut NA verdier med 0.
table_01[is.na(table_01)] <- 0

# Lager en funksjon som beregner mean, sd, min og max.
Calc_Stats <- function(x) {
  c(mean = mean(x), sd = sd(x), min = min(x), max = max(x))}

# Bruker funksjonen over hele datasettet.
resultat <- apply(table_01, 2, Calc_Stats)

# Gjør til datasett.
table_01 <- t(as.data.frame(resultat))

# Skifter navn på radene.
rownames(table_01) <- c("BNP per person (2015-2019)",
                      "BNP per person (1999-2015)",
                      "BNP vekstrate", "Sparerate",
                      "Befolkningsvekstrate", "Humankapital",
                      "Forbruk av naturressurser")

# Skifter navn på kolonnene.
colnames(table_01) <- c("Gjennomsnitt", "Standardavvik", "Minimum", "Maksimum")

```

```
# Viser tabellen.  
print(table_01)
```

	Gjennomsnitt	Standardavvik	Minimum	Maksimum
BNP per person (2015-2019)	2.145995e+04	2.174827e+04	753.362038	1.208806e+05
BNP per person (1999-2015)	1.580594e+04	1.836906e+04	631.837595	9.568429e+04
BNP vekstrate	1.821326e+00	2.374669e+00	-10.344038	7.775845e+00
Sparerate	9.154200e+00	1.009605e+01	-19.377402	4.565886e+01
Befolkningsvekstrate	1.252046e+00	1.222196e+00	-2.395368	4.232678e+00
Humankapital	4.985574e-01	2.322403e-01	0.000000	8.855418e-01
Forbruk av naturressurser	4.167362e+00	6.461700e+00	0.000000	3.847753e+01