Team8_Report

XXX XXXX
XXX

Reports XXXX
XXX

Hao Li
XXX

Report for
ETC5513

2 June 2021

Economics and business

(03) 9905 2478
questions@company.com

ABN: 12 377 614 630

```
AUS <- read_csv("Data/AUS.csv") %>%
  dplyr::filter(year >= 1990 & year <= 2019) %>%
  janitor::clean_names()
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   year = col_double(),
##   GDP_growth = col_double(),
##   Population = col_double(),
##   `Unemployment with advanced education` = col_double(),
##   `Unemployment with basic education` = col_double(),
##   `Unemployment female` = col_double(),
##   `Unemployment male` = col_double()
## )
```

Country XX1 and YY1"

Country XX2 and YY2

Q3 The influence of different factors on GDP.

Because there are missing values in some years of some variables in the original data, which will affect the estimation of the regression model, it is necessary to interpolate the missing values. Here I use the mean to interpolate. At the same time, I deal with the variable `population` in millions of units.And I made some single liner regression in figure 1 to judge whether they have a connection.

```
AUS_lm <- AUS %>%
  mutate(across(everything(),na_mean)) %>%
  dplyr::select(-year) %>%
  mutate(population = population/100000)
AUS_lm %>%
  pivot_longer(-gdp_growth,names_to = "vars",values_to = "value") %>%
  ggplot(aes(value,gdp_growth,color = vars)) +
  geom_point() +
```

```
  facet_wrap(~ vars,scales = "free")+
  geom_smooth(method = lm) +
  guides(color = F) +
  labs(x = "variables",y = "percentage of GDP change")+
  theme(strip.text.x = element_text(size = 8))
```

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

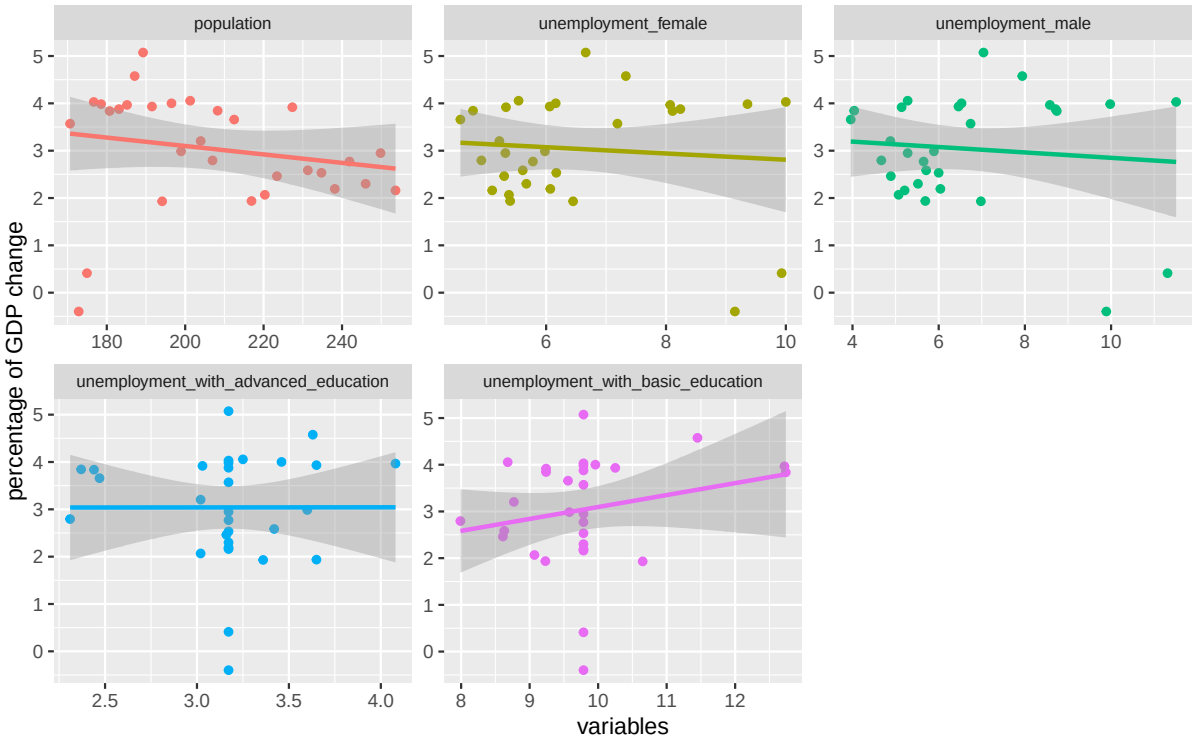

Figure 1: *liner model*

```
lm_model <- lm(gdp_growth ~ .,data = AUS_lm)
lm_model %>%
  tidy() %>%
  mutate(across(where(is.numeric),.fns = function(x) round(x,4))) %>%
  regulartable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	6.5313	4.8638	1.3428	0.1919
population	-0.0234	0.0124	-1.8932	0.0705
unemployment_with_advanced_education	0.0316	1.6093	0.0518	0.9591
unemployment_with_basic_education	0.3809	0.2391	1.5937	0.1241
unemployment_female	-0.1105	1.0815	-0.1022	0.9195
unemployment_male	-0.2652	0.8168	-0.3247	0.7482

It can be found that in table most variables are not significant when using the whole model for regression. I think this is due to the existence of multicollinearity in the data. Let's judge whether there is serious multicollinearity.

```
lm_model %>%
  car::vif() %>%
  as.data.frame() %>%
  setNames("Variance Inflation Factors") %>%
  rownames_to_column("Variables") %>%
  regulartable()
```

Variables	Variance Inflation Factors
population	2.156211
unemployment_with_advanced_education	23.47385
unemployment_with_basic_education	3.484657
unemployment_female	64.27438
unemployment_male	62.09696

By calculating the variance expansion factor in table we can find that the model has serious multi-collinearity, so we need to delete the variables.

```
lm_steped <- step(lm_model,trace = 0)
gglm(lm_steped)
```

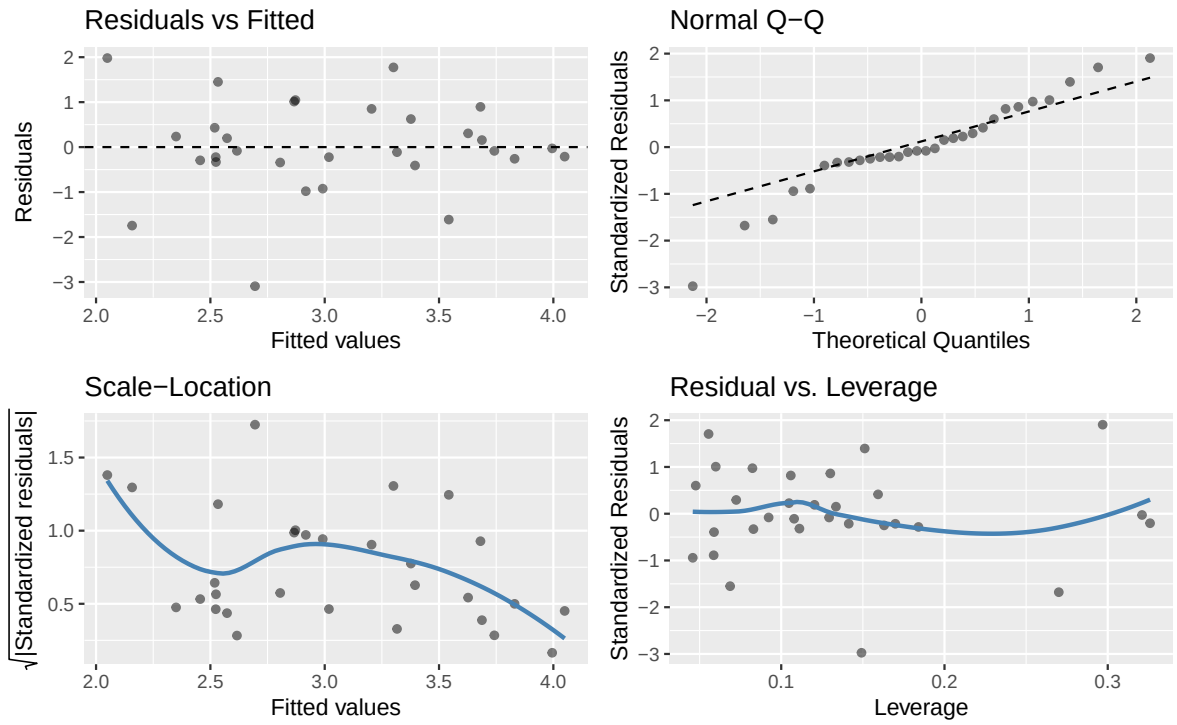

Figure 2: *residuals*

```
lm_steped %>%
  tidy() %>%
  mutate(across(where(is.numeric),.fns = function(x) round(x,4))) %>%
  regulartable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	6.2908	3.5992	1.7478	0.0923
population	-0.0229	0.0113	-2.0351	0.0522
unemployment_with_basic_education	0.3855	0.2245	1.7175	0.0978
unemployment_male	-0.3445	0.1475	-2.3356	0.0275

The regression model with stepwise regression method excluding highly correlated variables shows that variables `population`,`unemployment_with_basic_education`,`unemployment_male` have a significant impact on GDP growth at the significance level of 10%. To be specific, every 1 million increase in `population` will reduce GDP by 0.0229%. This may be because the Australian government is unable to provide sufficient employment opportunities and the demographic dividend has not been fully released. From the perspective of variable `unemployment_with_basic_education`, every 1% increase in the employment rate of people with basic education will increase the GDP growth by 0.38% on average. This main reason may be the industrial structure in Australia is undergoing adjustment, and some industries are eliminating people with only basic education to reduce the labor cost. Finally, every 1% increase in male unemployment rate will reduce GDP growth by an average of 0.34%, which shows that male workers still dominate all walks of life in Australia. Through the above model analysis of variables. We can know that the variables “population”, “unemployment rate of basic education” and “unemployment rate of men” have a certain impact on Australia’s GDP.