Applied Biostatistics Assigment 1

Léonard Berney, Mohammad Aquil

Data

The dataset we will be working on contains information on life-cycle savings for the 1960-1970 period in different countries. The data consists of 50 observations on 5 variables:

- sr: personal savings
- pop15: percentage of population under 15
- pop75: percentage of population over 75
- dpi: per-capita disposable income
- ddpi: percentage growth rate of dpi

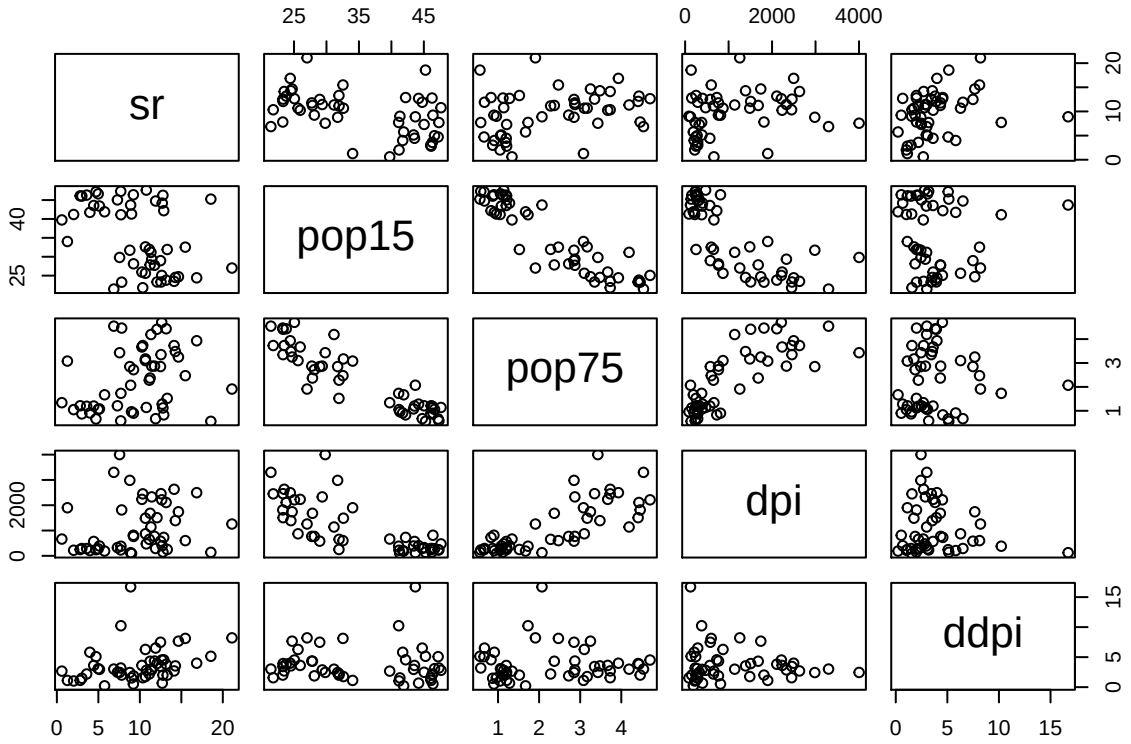

Figure 1: Pairs of variables

Linear Model

The objective is to build a linear model that can predict the personal saving ratio of a country. We will start by fitting a model using every variables and then try to prune it as much as possible, without sacrificing to much accuracy.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	28.5661	7.3545	3.88	0.0003
pop15	-0.4612	0.1446	-3.19	0.0026
pop75	-1.6915	1.0836	-1.56	0.1255
dpi	-0.0003	0.0009	-0.36	0.7192
ddpi	0.4097	0.1962	2.09	0.0425

Table 1: summary of the full model

Looking at the summary from Table 1, pop75 and dpi might not be significant for the model. For the full model, we have an AIC of 282.1961371 which improves to 280.3413931 when omitting the dpi variable.

On Figure 1 we see that the correlation between pop15 and pop75 is rather high (-0.9084787). Calculating the variance inflation factor, pop15 and pop75 both are above 5, which might indicate multicolinearity.

We will now experiment what happens when these variables are present or not. By removing pop15 from the model, we obtain an AIC of 288.4228964 and by removing pop75 we get 281.8861237. These two results are worse than what we obtained previously but the difference when removing pop75 is small enough that we should still consider removing the variable in the final model.

After all the pruning, we obtain the following final model:

```
## (Intercept)       pop15        ddpi
##  15.5995758  -0.2163762   0.4428302
```

Looking at the diagnostic plots on Figure 2, the model seems to preserve homoscedasticity and the residuals are normally distributed.

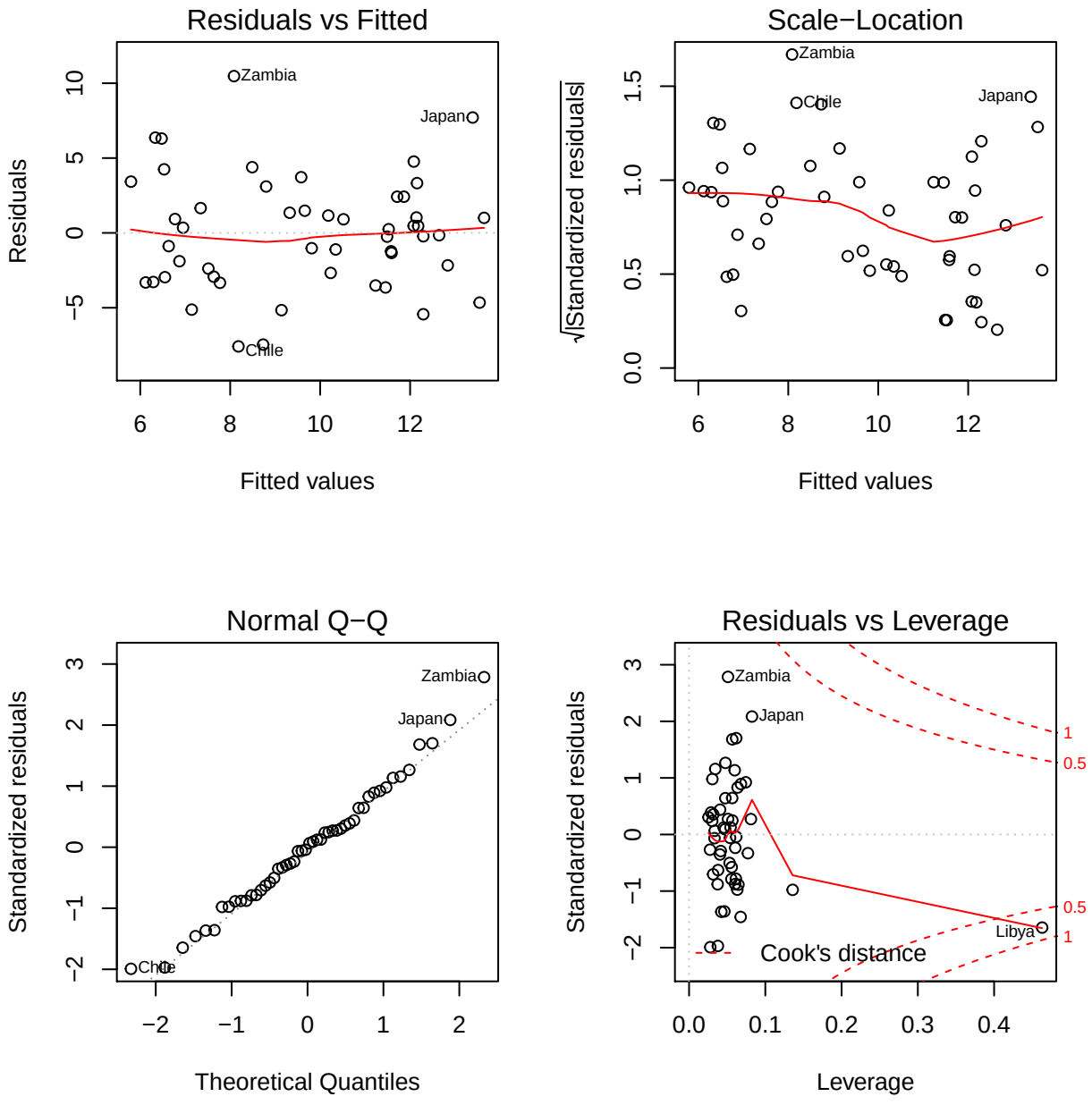

Figure 2: Diagnostic plots for the final model