# LAB 8 – CORRELATED ERRORS

## 1. Model

Suppose that

$$Y = X\beta + \varepsilon,$$

where $\varepsilon$ are mean 0 Normals with covariance matrix $\sigma^2\Sigma$. Our usual theory relies on the assumption that $\varepsilon \sim N(0, \sigma I)$, so does not directly apply in this case.

However, multiplying both sides by $\Sigma^{-1/2}$, we have

$$\underbrace{\Sigma^{-1/2}Y}_{Y'} = \underbrace{\Sigma^{-1/2}X}_{X'}\beta + \underbrace{\Sigma^{-1/2}\varepsilon}_{\varepsilon'}$$

Note that

$$\mathrm{Cov}(\varepsilon') = \sigma^2 I\,,$$

so we can apply out usual theory and methods to the new transformed problem

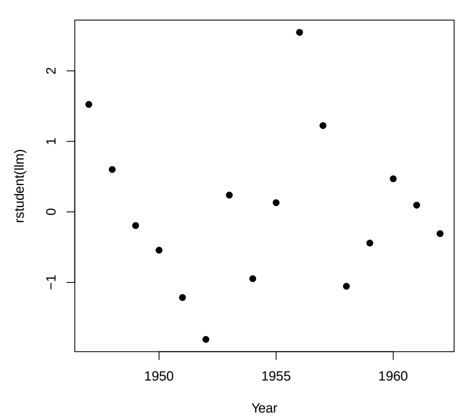
$$Y' = X'\beta + \varepsilon'\,.$$

We consinder the dataset `longley`

```
> data(longley)
> longley[1:4,]
     GNP.deflator     GNP Unemployed Armed.Forces Population Year Employed
1947         83.0 234.289      235.6        159.0    107.608 1947   60.323
1948         88.5 259.426      232.5        145.6    108.632 1948   61.122
1949         88.2 258.054      368.2        161.6    109.773 1949   60.171
1950         89.5 284.599      335.1        165.0    110.929 1950   61.187
```

We consider the response variable `Employed`

```
> llm = lm(Employed~GNP+Population, data=longley)
> plot(rstudent(llm)~Year,data=longley,pch=19)
```



There is perhaps correlation between $\varepsilon_i$ and $\varepsilon_{i-1}$.

We can compute the sample correlation of the residuals with the shifted residuals:

```
> rho = cor(residuals(llm)[-1],residuals(llm)[-16])
> print(rho)
[1] 0.3104092
```

Here we model the residuals as

$$\varepsilon_{i+1} = \rho\varepsilon_i + \delta_i,$$

where $\delta_i \sim N(0, \tau^2)$. In this case,

$$\Sigma_{i,j} = \rho^{|i-j|}$$

We can then directly estimate $\beta$:

```
> x = model.matrix(llm)
> Sigma = diag(16)
> Sigma = rho^abs(row(Sigma)-col(Sigma))
> SigInv = solve(Sigma)
> xtxi = solve(t(x) %*% SigInv %*% x)
> beta = xtxi %*% t(x) %*% SigInv %*% longley$Employed
```

We can check the residuals and standard errors

```
> res = longley$Employed - x%*%beta
> (sig = sqrt((t(res)%*%SigInv%*%res)/llm$df))
         [,1]
[1,] 0.542443
> sqrt(diag(xtxi))*sig
[1] 13.94477226  0.01070339  0.15338547
```

Compare this result to regressing $\Sigma^{-1/2}Y$ on $\Sigma^{-1/2}X$.

```
> sm = chol(Sigma)
> smi = solve(t(sm))
> sx = smi%*%x
> sy = smi%*%longley$Empl
```

Regress `sy` onto `sx` and compare with the results above. Note that `sx` already contains an intercept.

But note we have estimated $\rho$ from the residuals of the initial model. We can improve the estimate of $\rho$ using the residuals of the fitted model. We can then refit using this new $\rho$. Continuing, the estimate of $\rho$ should converge.

Try a few interations.

The package `nlme` will do this interative reweighting:

```
> library(nlme)
> g = gls(Employed~GNP+Population, correlation=corAR1(form=~Year),data=
+          longley)
> summary(g)
Generalized least squares fit by REML
  Model: Employed ~ GNP + Population
  Data: longley
       AIC      BIC    logLik
  44.66377 47.48852 -17.33188
```

```
Correlation Structure: AR(1)
 Formula: ~Year
 Parameter estimate(s):
      Phi
0.6441692

Coefficients:
                Value Std.Error   t-value p-value
(Intercept) 101.85813 14.198932  7.173647  0.0000
GNP           0.07207  0.010606  6.795485  0.0000
Population   -0.54851  0.154130 -3.558778  0.0035

 Correlation:
           (Intr) GNP
GNP         0.943
Population -0.997 -0.966

Standardized residuals:
       Min         Q1        Med         Q3        Max
-1.5924564 -0.5447822 -0.1055401  0.3639202  1.3281898

Residual standard error: 0.689207
Degrees of freedom: 16 total; 13 residual
```

```
> intervals(g)
```

```
Approximate 95% confidence intervals

 Coefficients:
                  lower         est.       upper
(Intercept) 71.18320460 101.85813305 132.5330615
GNP          0.04915865   0.07207088   0.0949831
Population  -0.88149053  -0.54851350  -0.2155365
attr(,"label")
[1] "Coefficients:"

 Correlation structure:
         lower      est.     upper
Phi -0.4432383 0.6441692 0.9645041
attr(,"label")
[1] "Correlation structure:"

 Residual standard error:
    lower      est.     upper
0.2477527 0.6892070 1.9172599
```

## 2. Weighted Least Squares

Suppose that

$$\Sigma = \begin{bmatrix} 1/w_1 & 0 & \cdots & 0 \\ 0 & 1/w_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1/w_n \end{bmatrix}$$

Then

$$\Sigma^{-1/2} = \begin{bmatrix} \sqrt{w_1} & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sqrt{w_n} \end{bmatrix}$$

```
> library(faraway)
> savings[1,]
             sr pop15 pop75     dpi ddpi
Australia 11.43 29.35  2.87 2329.68 2.87
```

Savings data-set: savings rate is regressed on

- percent population under 15,
- percent population over 75,
- per-capita disposable income,
- percent growth rate of per-capita disposable income.

Plot the residuals against `pop15`. What do you see about the variance?

Test the hypothesis that the variances are the same when `pop15` is smaller than 35 and when `pop15` is larger than 35. Use `var.test`.

Figure out weights, and add the argument `weights` to `lm` to fit the model with these weights.

Note that

$$\mathrm{sd}(\sqrt{w_i}(Y_i - EY_i)) = \sigma\,.$$

so should check if the spread of $\sqrt{w_i} \times \hat{\varepsilon}_i$ depends on $i$.