

4 Punto

Se cargan las librerias

```
library(MASS)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(nortest)
library(leaps)
```

carga de data

```
data("swiss")
attach(swiss)
```

Para hacer el calculo del Press se usan las siguientes funciones como referencia

https://fhernanb.github.io/libro_regresion/diag2.html#estad%C3%ADstica-press-y-r2-de-predicci%C3%B3n

```
PRESS <- function(linear.model) {
  # calculate the predictive residuals
  pr <- residuals(linear.model) / (1-lm.influence(linear.model)$hat)
  # calculate the PRESS
  PRESS <- sum(pr^2)
  return(PRESS)
}
```

```
pred_r_squared <- function(linear.model) {
  #' Use anova() to get the sum of squares for the linear model
  lm.anova <- anova(linear.model)
  #' Calculate the total sum of squares
  tss <- sum(lm.anova$'Sum Sq')
  # Calculate the predictive R^2
  pred.r.squared <- 1-PRESS(linear.model)/(tss)
  return(pred.r.squared)
}
```

Se crea el modelo completo para iniciar su interpretacion y su conjunto de covariables regresoras

```
elegir <- regsubsets(Fertility ~ Agriculture + Examination+Education + Catholic + Infant.Mortality, data = swiss)
summary(elegir)
```

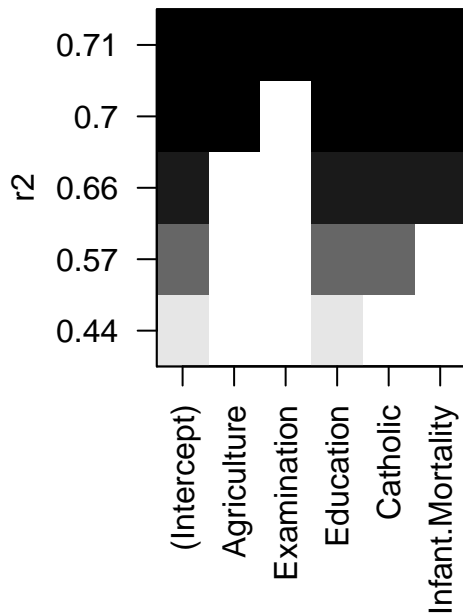
```
## Subset selection object
## Call: regsubsets.formula(Fertility ~ Agriculture + Examination + Education +
##      Catholic + Infant.Mortality, data = swiss)
## 5 Variables (and intercept)
##              Forced in Forced out
## Agriculture      FALSE      FALSE
## Examination      FALSE      FALSE
## Education         FALSE      FALSE
## Catholic          FALSE      FALSE
## Infant.Mortality  FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##      Agriculture Examination Education Catholic Infant.Mortality
## 1  ( 1 ) " "      " "      "*"      " "      " "
## 2  ( 1 ) " "      " "      "*"      "*"      " "
## 3  ( 1 ) " "      " "      "*"      "*"      "*"
## 4  ( 1 ) "*"      " "      "*"      "*"      "*"
## 5  ( 1 ) "*"      "*"      "*"      "*"      "*"

```

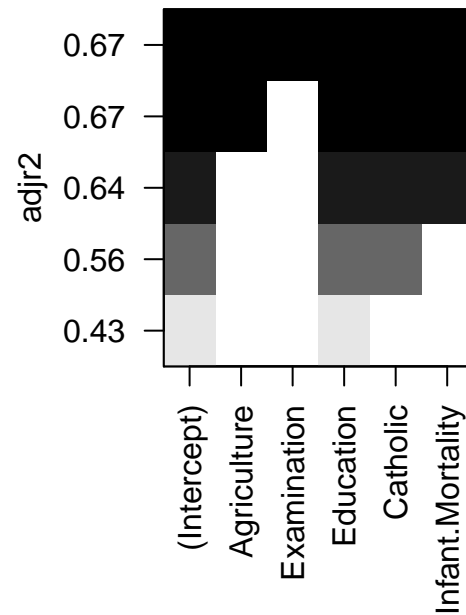
Se grafican cada uno de los modelos entregados por regsubsets

```
par(mfrow=c(1,2))
plot(elegir,scale="r2", main="R^2, Swiss.")
plot(elegir,scale="adjr2", main="R^2.ajust, Swiss.")
```

R², Swiss.

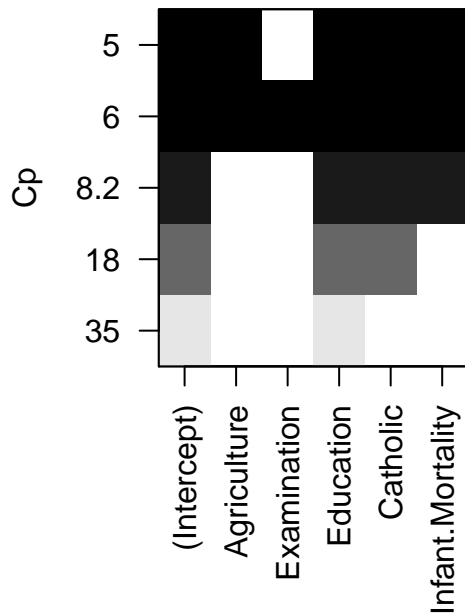


R².ajust, Swiss.

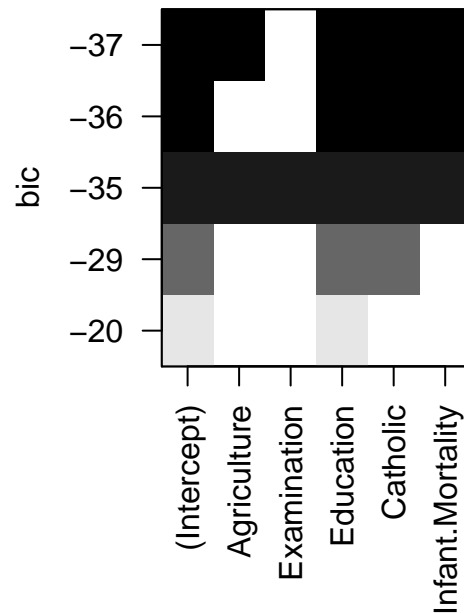


```
par(mfrow=c(1,2))
plot(elegir, scale="Cp", main="Cp, Swiss")
plot(elegir, scale="bic", main="(BIC, Swiss.)")
```

Cp, Swiss



(BIC, Swiss.



Se toma como candidato el modelo a partir del R^2 0.67 ajustado: $\text{Fertility} = B_0 + B_1 * \text{Education} + B_2 * \text{Catholic} + B_3 * \text{Infant.Mortality}$

```
modelo <- lm(Fertility ~ Education + Catholic + Infant.Mortality)
```

```
ValidarSupuestos <- function (respuesta ,modelo,confianza){
```

```
  # Test de normalidad
```

```
  shapiro <- shapiro.test(modelo$residuals)
```

```
  shapiro <- shapiro$p.value
```

```
  lillie <- lillie.test(modelo$residuals)
```

```
  lillie <- lillie$p.value
```

```
  ifelse((shapiro > confianza) & (lillie > confianza), print("Existe normalidad de los errores"), print("No existe normalidad de los errores"))
```

```
  # Test de homocedasticidad
```

```
  baruch <- bptest(modelo)
```

```
  baruch <- baruch$p.value
```

```

golfred <- qqtest(modelo)
golfred <- golfred$p.value

ifelse((baruch > confianza) & (golfred > confianza), print("Existe homocedasticidad"), print("No existe homocedasticidad"))

# Test de independencia

independencia <- dwtest(respuesta ~ modelo$residuals)
independencia <- independencia$p.value

ifelse((independencia > confianza), print("Hay independencia"), print("No hay independencia"))

# Construccion de la Tabla de respuestas

tabla <- rbind(shapiro, lillie, baruch, golfred, independencia)
rownames(tabla) <- c("shapiro", "lillie", "baruch", "golfred", "independencia")
colnames(tabla) <- c("p.value")

print(round(tabla,5))

}

```

Funcion para evaluacion de supuestos se validan supuestos mediante la funcion y se obtiene una conclusion

```
ValidarSupuestos(Fertility, modelo, 0.1)
```

```

## [1] "Existe normalidad de los errores"
## [1] "Existe homocedasticidad"
## [1] "No hay independencia"
##           p.value
## shapiro    0.52297
## lillie     0.82607
## baruch     0.53231
## golfred    0.33480
## independencia 0.00006

```

El modelo propuesto no cumple todas las validaciones de supuestos, por lo tanto existe riesgo al evaluar y no se recomienda estadísticamente:

Fertility = $B_0 + B_1 * \text{Education} + B_2 * \text{Catholic} + B_3 * \text{Infant.Mortality}$

R2 de predicción

```
PRESS(modelo)
```

```
## [1] 2901.555
```

```
pred_r_squared(modelo)
```

```
## [1] 0.5957686
```

De la salida anterior se observa que el R2 de predicción toma un valor de 0.5957686, esto indica que el modelo mod tiene una buena capacidad de predicción para nuevas observaciones.