Applied Regression Analysis: HW2

NCCU, Department of Statistics
統計碩一 葉佐晨
112354016

2023-12-10

Variance Decomposition Proportions

Packages required

```
library(car)
library(magrittr)
library(mctest)
library(matlib)
library(knitr)
library(generics)
```

Import data

Data is from **TABLE 9.4 Unstandardized Regressor and Response Variables from Webster, Gunst, and Mason [1974].**

```
data <- read.table("tb9.4 Webster, Gunst, and Mason.txt",
                   header = TRUE,
                   row.names = "Observation")
```

Center regressors

```
X <- as.matrix(data[-1])
X_cen <- scale(X, scale = FALSE)
```

Find eigenvectors

Show that **TABLE 9.6 Eigenvectors for the Webster, Gunst, and Mason Data**.

```
X_corxz <- scale(X, scale = FALSE) %>%
  apply(2, function(x) x/sqrt(sum(x^2)))

tb_eigvt <- eigen(t(X_corxz) %*% X_corxz) %>%
  .[["vectors"]] %>%
  round(5) %>%
  as.data.frame
colnames(tb_eigvt) <- paste0("t", 1:6)

kable(tb_eigvt, longtable = FALSE,
      caption = "TABLE 9.6 Eigenvectors for the Webster, Gunst, and Mason Data")
```

表 1: TABLE 9.6 Eigenvectors for the Webster, Gunst, and Mason Data

t1	t2	t3	t4	t5	t6
-0.39072	0.33968	0.67980	-0.07990	0.25104	-0.44768
-0.45560	0.05392	-0.70013	-0.05769	0.34447	-0.42114
0.48264	0.45333	-0.16078	-0.19103	-0.45364	-0.54169
0.18766	-0.73547	0.13587	0.27645	-0.01521	-0.57337
-0.49773	0.09714	-0.03185	0.56356	-0.65128	-0.00605
0.35195	0.35476	-0.04864	0.74818	0.43375	-0.00217

Eigenvalues and Variance Decomposition Proportion

Show that **TABLE 9.7 Variance Decomposition Proportions for the Webster, Gunst, and Mason[1974] Data**.

There are two methods can produce table 9.7, first is using package `mctest`, which is writed to do eigenvalues and variance decomposition proportion, second is caculating mathematically by hand.

method 1, with `mctest`

```
md_A <- lm(data$y ~ 0 + X_cen)
md_B <- lm(data$y ~ X)
```

```r
structure(eigprop(md_A))
```

A. Regressors Centered

```
Call:
eigprop(mod = md_A)

  Eigenvalues      CI X_cenx1 X_cenx2 X_cenx3 X_cenx4 X_cenx5 X_cenx6
1      2.4288  1.0000  0.0003  0.0005  0.0004  0.0000  0.0531  0.0350
2      1.5462  1.2533  0.0004  0.0000  0.0005  0.0012  0.0032  0.0559
3      0.9221  1.6230  0.0028  0.0033  0.0001  0.0001  0.0006  0.0018
4      0.7940  1.7490  0.0000  0.0000  0.0002  0.0003  0.2083  0.4845
5      0.3079  2.8086  0.0011  0.0024  0.0025  0.0000  0.7175  0.4199
6      0.0011 46.8605  0.9953  0.9937  0.9964  0.9984  0.0172  0.0029

===============================
Row 6==> X_cenx2, proportion 0.993750 >= 0.50
Row 6==> X_cenx3, proportion 0.996353 >= 0.50
Row 6==> X_cenx4, proportion 0.998383 >= 0.50
Row 5==> X_cenx5, proportion 0.717533 >= 0.50
```

```r
structure(eigprop(md_B))
```

B. Regressors Not Centered

```
Call:
eigprop(mod = md_B)

  Eigenvalues      CI (Intercept)    Xx1    Xx2    Xx3    Xx4    Xx5    Xx6
1      2.6329  1.0000      0.0001 0.0003 0.0003 0.0001 0.0001 0.0217 0.0043
2      1.8206  1.2025      0.0000 0.0001 0.0002 0.0005 0.0000 0.0523 0.0949
3      1.0333  1.5962      0.0000 0.0002 0.0000 0.0002 0.0013 0.0256 0.1010
4      0.6583  1.9999      0.0000 0.0005 0.0000 0.0005 0.0003 0.1906 0.3958
5      0.6057  2.0849      0.0000 0.0025 0.0035 0.0001 0.0001 0.0011 0.0002
```

```
6      0.2488  3.2528      0.0000 0.0012 0.0023 0.0028 0.0000 0.6909 0.4003
7      0.0003 92.2534      0.9999 0.9953 0.9936 0.9959 0.9983 0.0178 0.0034

===============================
Row 7==> Xx1, proportion 0.995292 >= 0.50
Row 7==> Xx2, proportion 0.993595 >= 0.50
Row 7==> Xx3, proportion 0.995885 >= 0.50
Row 7==> Xx4, proportion 0.998256 >= 0.50
Row 6==> Xx5, proportion 0.690893 >= 0.50
```

method 2, without `mctest`

```r
eigval_A <- eigen(t(X_corxz) %*% X_corxz)$values %>% round(5)

D_A <- svd(X_corxz)$d
eta_A <- (max(D_A)/D_A) %>%
  round(5)

vif_A <- data.frame(y = data$y, X_corxz) %>%
  lm(y ~ ., data = .) %>%
  vif
t_A <- svd(X_corxz)$v

vdp_A <- t(t_A^2 %*% diag(1/D_A)^2) %*% diag(1/vif_A) %>% round(4)
dimnames(vdp_A) <- list(NULL,colnames(X))

tb_A <- data.frame(Number = 1:6,
                   Eigenvalue = eigval_A,
                   `Condition Indices` = eta_A,
                   vdp_A)
kable(tb_A, longtable = FALSE, caption = "TABLE 9.7: A. Regressors Centered")
```

A. Regressors Centered

表 2: TABLE 9.7: A. Regressors Centered

Number	Eigenvalue	Condition.Indices	x1	x2	x3	x4	x5	x6
1	2.42879	1.00000	0.0003	0.0005	0.0004	0.0000	0.0531	0.0350
2	1.54615	1.25334	0.0004	0.0000	0.0005	0.0012	0.0032	0.0559
3	0.92208	1.62297	0.0028	0.0033	0.0001	0.0001	0.0006	0.0018
4	0.79398	1.74900	0.0000	0.0000	0.0002	0.0003	0.2083	0.4845
5	0.30789	2.80864	0.0011	0.0024	0.0025	0.0000	0.7175	0.4199
6	0.00111	46.86052	0.9953	0.9937	0.9964	0.9984	0.0172	0.0029

```r
X_corxz0 <- cbind(x0=1, X) %>%
  apply(2, function(x) x/sqrt(sum(x^2)))
eigval_B <- (eigen(t(X_corxz0) %*% X_corxz0)$values)%>%
  round(5)

D_B <- svd(X_corxz0)$d
eta_B <- (max(D_B)/D_B) %>% round(5)

t_B <- svd(X_corxz0)$v

vif_B <- data.frame(y = data$y, cbind(x0=1, X)) %>%
  lm(y ~ 0 + ., data = .) %>%
  vif
t <- svd(X_corxz)$v
vdp_B <- t(t_B^2 %*% diag(1/D_B)^2) %*% diag(1/vif_B) %>% round(4)
dimnames(vdp_B) <- list(NULL, colnames(X_corxz0))

tb_B <- data.frame(Number = 1:7,
                   Eigenvalue = eigval_B,
                   `Condition Indices` = eta_B,
                   vdp_B)

kable(tb_B, longtable = FALSE, caption = "TABLE 9.7: B. Regressors Not Centered ")
```

B. Regressors Not Centered

表 3: TABLE 9.7: B. Regressors Not Centered

Number	Eigenvalue	Condition.Indices	x0	x1	x2	x3	x4	x5	x6
1	2.63287	1.00000	0.0001	0.0003	0.0003	0.0001	0.0001	0.0217	0.0043
2	1.82065	1.20255	0.0000	0.0001	0.0002	0.0005	0.0000	0.0523	0.0949
3	1.03335	1.59622	0.0000	0.0002	0.0000	0.0002	0.0013	0.0256	0.1010
4	0.65826	1.99994	0.0000	0.0005	0.0000	0.0005	0.0003	0.1906	0.3958
5	0.60573	2.08485	0.0000	0.0025	0.0035	0.0001	0.0001	0.0011	0.0002
6	0.24884	3.25280	0.0000	0.0012	0.0023	0.0028	0.0000	0.6909	0.4003
7	0.00031	92.25341	0.9999	0.9953	0.9936	0.9959	0.9983	0.0178	0.0034