

HW 21

Nick Huo & Anqi Zang

2022-11-09

Question 1

a. – d.

```
set.seed(567)
x1 <- rnorm(25)
x2 <- rnorm(25)
x3 <- rnorm(25)
x4 <- rnorm(25)
x5 <- x1 + 2*x2 + rnorm(25)

X <- matrix(c(scale(x1),scale(x2),scale(x3),scale(x4),scale(x5)),25,5)
X
```

##		[,1]	[,2]	[,3]	[,4]	[,5]
##	[1,]	1.005610402	0.4272745	1.28338315	-0.87528983	0.2417061
##	[2,]	0.692674894	-0.6545358	0.26970586	0.99724147	-1.0699490
##	[3,]	-0.272935279	1.0622097	-1.24663991	0.02218045	1.4620225
##	[4,]	-1.060591849	1.2221747	-0.09677013	0.59537572	0.6023148
##	[5,]	0.576054934	-1.0548486	1.38577689	0.89918540	-0.9001702
##	[6,]	-0.045304947	-0.1536169	-2.11681431	1.10740043	-0.1276701
##	[7,]	-0.008577994	1.8160744	-1.37269288	-0.50007685	0.6014312
##	[8,]	-1.733885301	-0.1121685	-0.59131893	0.23742197	-0.3028700
##	[9,]	0.417464961	0.5859900	-0.72378806	-1.09865642	0.9769817
##	[10,]	-1.552945960	0.3089871	-0.05932636	1.37293609	-2.0867843
##	[11,]	-0.763074680	-1.5198531	0.48366664	0.48062373	-1.5329439
##	[12,]	-0.386882476	1.7189474	0.70570469	-0.27520232	2.1635218
##	[13,]	-1.758132070	0.3898116	0.82839207	0.26360412	-0.4798485
##	[14,]	1.697973007	-1.7081067	-1.24865721	1.57354307	-0.3849342
##	[15,]	0.092826984	0.4490587	-0.24083251	-0.79527192	0.8134242
##	[16,]	0.445966135	1.1656258	-0.03331354	1.06216613	1.4411231
##	[17,]	-1.542537982	0.5045969	1.77740612	-0.79811908	-0.1657626
##	[18,]	0.920229816	-0.4864580	0.98667362	-0.69255869	-0.1611692
##	[19,]	1.753629143	-0.1430292	0.15190717	-0.89859268	0.5191825
##	[20,]	-0.187148844	-1.4615712	-0.53328551	-2.25158043	-0.1700902
##	[21,]	-0.087212248	0.6195860	1.36816428	0.63190033	0.9148398
##	[22,]	-0.100048019	-0.2209135	0.28103941	-1.97648295	-0.4860733
##	[23,]	0.962104671	-1.3763602	-1.29986968	0.71637659	-0.7147598
##	[24,]	-0.146279677	-0.5149594	0.48770922	0.05037786	-1.2689633
##	[25,]	1.081022379	-0.8639157	-0.44622007	0.15149780	0.1154409

```
cov(X)
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,]  1.00000000 -0.34226166 -0.16807250 -0.02423402  0.16675144
## [2,] -0.34226166  1.00000000  0.06642093 -0.10071817  0.66920677
## [3,] -0.16807250  0.06642093  1.00000000 -0.16803382 -0.09799935
## [4,] -0.02423402 -0.10071817 -0.16803382  1.00000000 -0.25788598
## [5,]  0.16675144  0.66920677 -0.09799935 -0.25788598  1.00000000
```

e.

```
calc_vif <- function(mtx) {
  res <- c()
  for (i in 1:ncol(mtx)) {
    mod <- lm(mtx[,i] ~ mtx[, -i])
    rsq <- summary(mod)$r.squared
    res <- c(res, 1/(1-rsq))
  }
  return(res)
}
```

```
calc_vif(X)
```

```
## [1] 1.689539 3.059697 1.099142 1.149274 2.980510
```

Question 2

a.

```
Xnew <- rbind(c(1, -1.195, 0.147, 1.339), c(1, 0.88, 0.555, 0.039), c(1, -0.368, -0.148, 1.021), c(1, 1.728, -0.134, 0.398), c(1, -0.490, -0.013, 1.332), c(1, 0.290, 0.551, -0.279), c(1, -1.513, 0.168, -1.384), c(1, 0.572, -2.320, -0.638))
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,]      1 -1.195  0.147  1.339
## [2,]      1  0.880  0.555  0.039
## [3,]      1 -0.368 -0.148  1.021
## [4,]      1  1.728 -0.134  0.398
## [5,]      1 -0.490 -0.013  1.332
## [6,]      1  0.290  0.551 -0.279
## [7,]      1 -1.513  0.168 -1.384
## [8,]      1  0.572 -2.320 -0.638
```

```
cor(Xnew[, -1])
```

```
##           [,1]      [,2]      [,3]
## [1,]  1.00000000 -0.1851990 -0.03164395
## [2,] -0.18519897  1.00000000  0.21591824
## [3,] -0.03164395  0.2159182  1.00000000
```

```
t(Xnew) %*% Xnew
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,]  8.000 -0.096000 -1.194000  1.828000
## [2,] -0.096  8.264386 -1.279417 -0.258303
## [3,] -1.194 -1.279417  6.083888  1.090641
## [4,]  1.828 -0.258303  1.090641  7.169852
```

The matrix shows that these covariation are loosely correlated with each other, so that they are orthogonal. And looking at $X^T X$, we see that only the diagonal elements have large values, other elements are very small. So this also means that the matrix is orthogonal

b.

```
Xb <- solve(t(Xnew)%*%Xnew)
Xb
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,]  0.139793078  0.005920606  0.03601343 -0.040906026
## [2,]  0.005920606  0.125325647  0.02773495 -0.001213379
## [3,]  0.036013426  0.027734950  0.18374648 -0.036133237
## [4,] -0.040906026 -0.001213379 -0.03613324  0.155354837
```

The matrix is symmetric. The components of the matrix are very small.

c.

The components of matrix are very small, so the variance of our beta estimators are also small. Thus we will have a more narrow CI around the true value. And it will be easier for us to find the beta values significant.

d.

```
calc_vif(Xnew[, -1])
```

```
## [1] 1.035595 1.085149 1.048980
```

Question 3

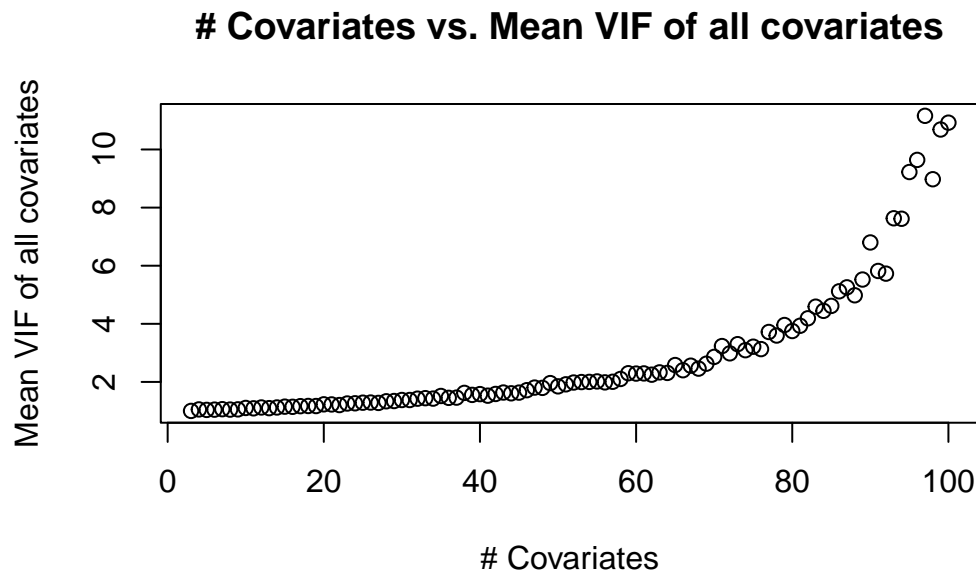
(a)

As we have seen from an earlier homework, as the number of covariates increase, R^2 increases even when there is no relationship between the covariates. Thus, even if there is actually no multicollinearity present in our data, if we have too many predictors, the VIF will be large for all covariates.

(b)

```
mean_vif <- c()
n <- 110
for (i in 3:100) {
  test_mat <- matrix(NA,n,i)
  for (j in 1:i) {
    test_mat[,j] <- rnorm(n)
  }
  this_mean <- mean(calc_vif(test_mat))
  mean_vif <- c(mean_vif,this_mean)
}

plot(3:100, mean_vif, main="# Covariates vs. Mean VIF of all covariates",
     xlab="# Covariates", ylab="Mean VIF of all covariates")
```



All the covariates (columns of the design matrices) are randomly generated $N(0,1)$ vectors, so we know that there is no multicollinearity in the data. However, we see that as we increase the number of covariates, there is this trend of increased mean VIF throughout all covariates.