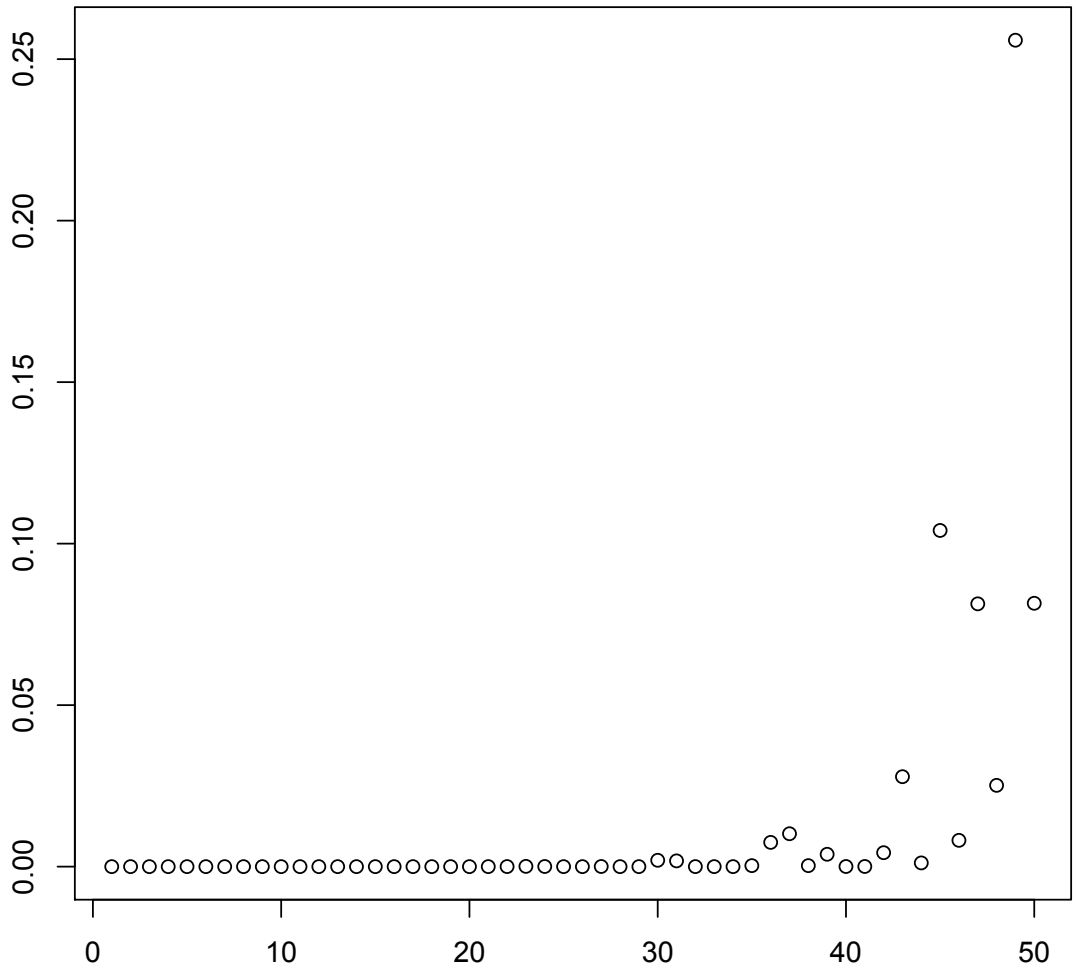
```
### A simulation example.

> n=40;
> x1= rnorm(n);
> x1= x1-mean(x1);
> y=  x1+ rnorm(n);

> summary(lm(y~x1))$coef
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) -0.0578191  0.1677409 -0.344693 7.322257e-01
x1           1.2095332  0.1382422  8.749380 1.219937e-10

> r=seq(0.5, 0.99, by=0.01);
> m=length(r);
> x.pvalue=matrix(0,m,2);

> for(i in 1:m){
+   x2=rnorm(n); x2=x2-mean(x2);
+   x2=sqrt(1-r[i]^2)*x2+r[i]*x1
+   mycoef=summary(lm(y~x1+x2))$coef;
+   x.pvalue[i,]=mycoef[-1,4];
+ }

> x.pvalue=apply(x.pvalue, 1, min);
> plot(x.pvalue, xlab="", ylab="");
```


0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50

```
> library(faraway)

> data(seatpos)
# Car Seat Position Data
# Car drivers like to adjust the seat position for their own comfort.
# Car designers would find it helpful to know how different drivers
# will position the seat depending on their size and age. Researchers
# at the HuMoSim laboratory at the University of Michigan collected
# data on 38 drivers.

# Age:
# Weight:
# HtShoes: height with shoes in cm
# Ht: height without shoes in cm
# Seated: seated height in cm
# Arm: lower arm length in cm
# Thigh: thigh length in cm
# Leg: lower leg length in cm
# hipcenter: horizontal distance of the midpoint of the hips from a
# fixed location in the car in mm
```

```
> g=lm(hipcenter ~ ., seatpos)

> summary(g)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 436.43213  166.57162   2.620   0.0138 *
Age           0.77572    0.57033   1.360   0.1843
Weight        0.02631    0.33097   0.080   0.9372
HtShoes      -2.69241    9.75304  -0.276   0.7845
Ht            0.60134   10.12987   0.059   0.9531
Seated        0.53375    3.76189   0.142   0.8882
Arm          -1.32807    3.90020  -0.341   0.7359
Thigh        -1.14312    2.66002  -0.430   0.6706
Leg          -6.43905    4.71386  -1.366   0.1824

Residual standard error: 37.72 on 29 degrees of freedom
Multiple R-squared: 0.6866,   Adjusted R-squared: 0.6001
F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05
```

```
# check pairwise correlation
> round(cor(seatpos), dig=2)
            Age Weight HtShoes    Ht Seated   Arm Thigh   Leg hipcenter
Age        1.00   0.08   -0.08 -0.09  -0.17  0.36  0.09 -0.04      0.21
Weight     0.08   1.00    0.83  0.83   0.78  0.70  0.57  0.78     -0.64
HtShoes   -0.08   0.83    1.00  1.00   0.93  0.75  0.72  0.91     -0.80
Ht        -0.09   0.83    1.00  1.00   0.93  0.75  0.73  0.91     -0.80
Seated    -0.17   0.78    0.93  0.93   1.00  0.63  0.61  0.81     -0.73
Arm        0.36   0.70    0.75  0.75   0.63  1.00  0.67  0.75     -0.59
Thigh      0.09   0.57    0.72  0.73   0.61  0.67  1.00  0.65     -0.59
Leg       -0.04   0.78    0.91  0.91   0.81  0.75  0.65  1.00     -0.79
hipcenter  0.21  -0.64   -0.80 -0.80  -0.73 -0.59 -0.59 -0.79      1.00
```

```
# condition number
> x = model.matrix(g)[,-1]
> e = eigen(t(x) %*% x)
> e$val
[1] 3.653671e+06 2.147948e+04 9.043225e+03 2.989526e+02 1.483948e+02
[6] 8.117397e+01 5.336194e+01 7.298209e+00
> sqrt(e$val[1]/e$val)
[1]   1.00000  13.04226  20.10032 110.55123 156.91171 212.15650
[7] 261.66698 707.54911

> x = model.matrix(g)[,-1]
> x = x - matrix(apply(x,2, mean), 38,8, byrow=TRUE)
> x = x / matrix(apply(x, 2, sd), 38,8, byrow=TRUE)
> apply(x,2,mean)
          Age        Weight       HtShoes            Ht        Seated
-2.193512e-17  2.810252e-16  9.566280e-16  1.941574e-16 -1.073010e-15
          Arm         Thigh           Leg
-1.070022e-16  8.909895e-17 -9.114182e-17
> apply(x,2,var)
    Age  Weight HtShoes      Ht  Seated     Arm   Thigh     Leg
      1       1       1       1       1       1       1       1

> e = eigen(t(x) %*% x)
> sqrt(e$val[1]/e$val)
[1]  1.000000  2.141737  3.497636  4.852243  5.404643  6.384606
[7] 10.615424 59.766197
```

```
# VIF
> round(vif(x), dig=2)
    Age  Weight HtShoes      Ht  Seated     Arm   Thigh     Leg
   2.00    3.65  307.43  333.14    8.95    4.50    2.76    6.69

> sqrt(307.43)
 [1] 17.53368
# that is, the se for the coef associated with HtShoes is 17.5 times
# larger than it would have been without collinearity.


# remove some variables
> cor(Seated+Thigh, Ht)
[1] 0.9389819
> cor(Seated+Leg, Ht)
[1] 0.965607
> cor(Seated+Arm, Ht)
[1] 0.9465523
> g2 = lm(hipcenter ~ Age + Weight + Ht + Seated, data=seatpos)
> summary(g2)

             Estimate Std. Error t value Pr(>|t|)
(Intercept) 478.65890  159.73362   2.997  0.00515 **
Age           0.58396    0.42573   1.372  0.17943
Weight       -0.01535    0.31640  -0.049  0.96159
Ht           -4.99025    1.64389  -3.036  0.00466 **
Seated        2.04632    3.41283   0.600  0.55287
```

```
Residual standard error: 36.83 on 33 degrees of freedom
Multiple R-squared: 0.6599,   Adjusted R-squared: 0.6186
F-statistic: 16.01 on 4 and 33 DF,  p-value: 2.224e-07

> g3 = lm(hipcenter ~ Ht, data=seatpos)
> summary(g3)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 556.2553    90.6704   6.135 4.59e-07 ***
Ht           -4.2650     0.5351  -7.970 1.83e-09 ***

Residual standard error: 36.37 on 36 degrees of freedom
Multiple R-squared: 0.6383,   Adjusted R-squared: 0.6282
F-statistic: 63.53 on 1 and 36 DF,  p-value: 1.831e-09

> anova(g3, g2)

Model 1: hipcenter ~ Ht
Model 2: hipcenter ~ Age + Weight + Ht + Seated
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1     36 47616
2     33 44774  3    2841.6 0.6981 0.5599
```