1. A data set is provided to predict the divorce rates in the US (# divorce per 1,000 women age 16 years or more). Divorce rate is provided for years 1920-1996. The predictor variables for the same years are:

   – year: the year from 1920-1996
   – unemployed: unemployment rate
   – femlab: percent female participation in labor force aged 16+
   – marriages: marriages per 1000 unmarried women aged 16+
   – birth: births per 1000 women aged 15-44
   – femnolab: Calculated as 100-femlab, to estimate the percent female age 16+ non participating in labor force.

   The proposed regression model presents one of the following problems:

   ( ) Errors are not independent
   ( ) Errors are not normally distributed
   (X) There is exact collinearity
   ( ) Residual variance is not constant

   Justification: The variable femnolab is a linear combination of femlab.

2. The data set divusa from the faraway library is used to predict divorce rate as a function of predictors year, unemployed, femlab, marriage, birth and military. The model output is given in the following figure:

```
Call:
lm(formula = divorce ~ ., data = divusa)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9087 -0.9212 -0.0935  0.7447  3.4689

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 380.14761   99.20371   3.832 0.000274 ***
year         -0.20312    0.05333  -3.809 0.000297 ***
unemployed   -0.04933    0.05378  -0.917 0.362171
femlab        0.80793    0.11487   7.033 1.09e-09 ***
marriage      0.14977    0.02382   6.287 2.42e-08 ***
birth        -0.11695    0.01470  -7.957 2.19e-11 ***
military     -0.04276    0.01372  -3.117 0.002652 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.513 on 70 degrees of freedom
Multiple R-squared:  0.9344,    Adjusted R-squared:  0.9288
F-statistic: 166.2 on 6 and 70 DF,  p-value: < 2.2e-16
```

   If the Variance Inflation Factor for variable year is: 47.27, the R^2 for the regression between year as the response, and the rest of the predictors (year $\sim$ unemployed+ femlab,+marriage+ birth + military) is: (0.98)

   Justification: Recall that the definition of VIF is $VIF = \frac{1}{1-R_k^2}$.

3. A Multiple Linear Regression model has been fitted to the response variable divorce as a function of a set of predictors: year, unemployed, femlab, marriage, birth, military.

   After standardizing the columns of the design matrix X (excluding the first column), the eigenvalues of

$$(X^T X)$$

are: 233.8132582 127.9095015 53.4749991 23.1290289 16.9672685 0.7059438

(a) According to these results the matrix condition number is (18.2)

Justification:

$$\kappa = \sqrt{\frac{\text{largest eigenvalue}}{\text{smallest eigenvalue}}}$$

(b) According to the condition number rule of thumb, can we declare a collinearity problem?
( ) Yes (X) No

Justification: We say that we have a collinearity problem when the condition number is greater than 30, which is not the case here.