Exam 1 Extra Practice

Kathleen Wendt

3/1/2020

Question 1: Body Fat

1A: VIF

```
##  triceps    thigh   midarm
## 708.8429 564.3434 104.6060
```

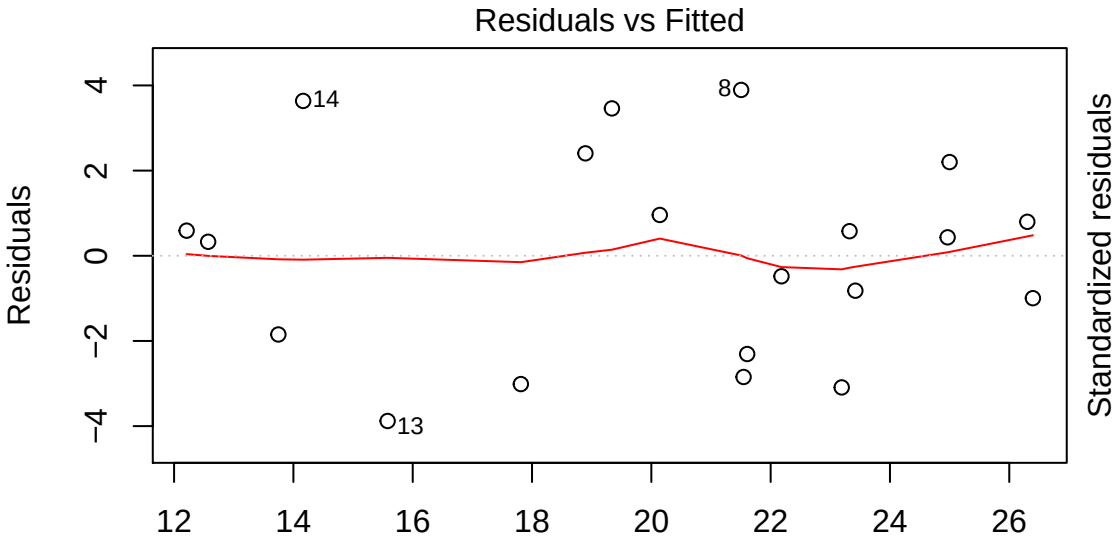
```
## [1] 714.2857
```

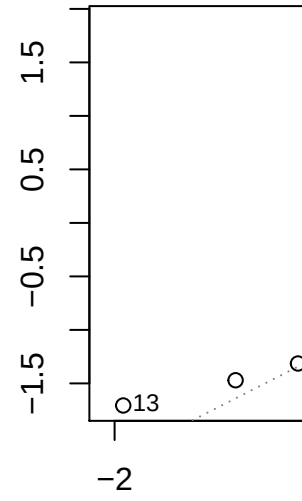
I don't know how to calculate VIF by hand or where to find the information in the slides.

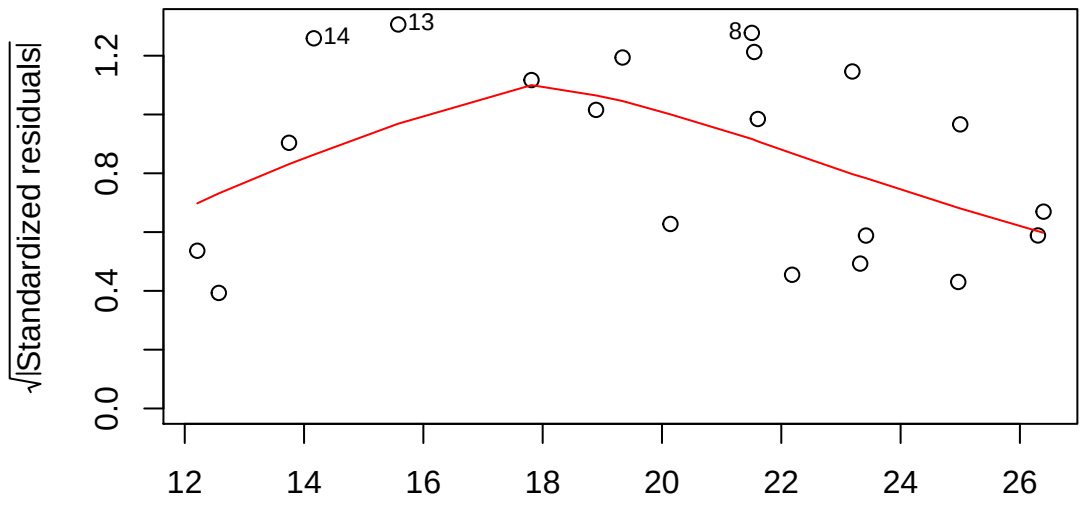
1B: Refit

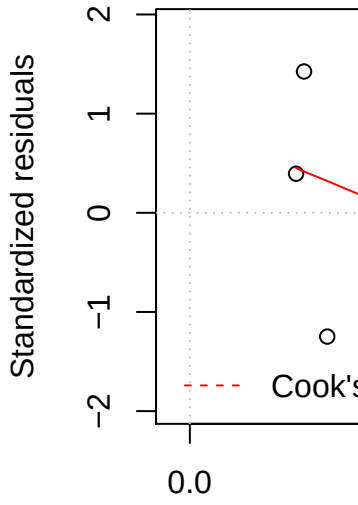
```
##  triceps   midarm
## 1.265118 1.265118
```

1C: Influence










```
## Influence measures of
##   lm(formula = body_fat ~ triceps + midarm, data = fat_data) :
##
##      dfb.1_  dfb.trcp dfb.mdrm   dffit cov.r  cook.d    hat inf
## 1  -0.01417  0.308728 -0.21515 -0.3768 1.295 0.04831 0.1785
## 2   0.00575 -0.075532  0.08371  0.3514 0.865 0.03851 0.0538
## 3   1.05631  0.052521 -1.05722 -1.2439 1.327 0.47826 0.3988   *
## 4   0.30021 -0.168675 -0.20053 -0.4774 0.979 0.07251 0.1119
## 5  -0.01333 -0.074759  0.06649  0.0919 1.643 0.00299 0.2732   *
## 6  -0.05045 -0.030637  0.06128 -0.0780 1.370 0.00215 0.1310
```

```
## 7  -0.01524  0.114639 -0.05289  0.1408 1.379 0.00697 0.1484
## 8  -0.29157  0.066364  0.27028  0.5306 0.787 0.08408 0.0865
## 9   0.33347 -0.023774 -0.26269  0.3956 1.133 0.05195 0.1278
## 10 -0.17910 -0.102926  0.20727 -0.3076 1.113 0.03165 0.0917
## 11 -0.03196  0.042322  0.00616  0.0668 1.356 0.00158 0.1211
## 12 -0.06659  0.236274 -0.07068  0.3259 1.150 0.03568 0.1093
## 13 -0.70148  0.378629  0.34248 -0.8242 0.823 0.19936 0.1704
## 14  0.07310 -0.579789  0.36482  0.7163 0.879 0.15486 0.1560
## 15  0.16720 -0.117851 -0.06671  0.1953 1.757 0.01344 0.3272   *
## 16 -0.03801  0.033493  0.01903  0.0754 1.308 0.00201 0.0926
## 17 -0.03646 -0.065715  0.06818 -0.1116 1.303 0.00438 0.0989
## 18 -0.06520 -0.171702  0.16163 -0.2242 1.462 0.01760 0.2081
## 19 -0.11269  0.153656 -0.03214 -0.3343 0.961 0.03595 0.0648
## 20  0.01396 -0.000594 -0.00237  0.0881 1.228 0.00273 0.0501
```

```
## [1] 0.4472136
```

```
## [1] 0.7745967
```

Observation 3 is influential based on DFBETAS and DFFITS but not according to Cook's distance.

Question 2: Mortality

2A: Correlations

```
## # A tibble: 4 x 5
##   rowname   mortality hc_log nox_log so2_log
##   <chr>         <dbl>  <dbl>   <dbl>   <dbl>
## 1 mortality    NA      0.151   0.292   0.403
## 2 hc_log        0.151 NA       0.947   0.641
## 3 nox_log       0.292  0.947  NA       0.733
## 4 so2_log       0.403  0.641   0.733  NA
```

There is a high correlation between hydrocarbon and nitrogen oxide pollution.

2B: Model selection

The model selected includes `density`, `educ`, `house`, `jantemp`, `julytemp`, `nonwhite`, and `precip`.

2C: Add pollution predictors

2D: VIF

```
##    precip   jantemp  julytemp     house      educ   density  nonwhite    hc_log
##  2.350081  1.980619  3.420447  2.112059  1.871509  1.480251  3.929744 16.835589
##   nox_log
## 12.660836
```

There is collinearity between `hc_log` and `nox_log`.

2E: Interpretation

```
##
## Call:
## lm(formula = mortality ~ precip + jantemp + julytemp + house +
##     educ + density + nonwhite + hc_log + nox_log, data = death_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.989 -19.161   2.703  19.733  76.349
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.353e+03  2.451e+02   5.521 1.20e-06 ***
## precip       1.766e+00  6.436e-01   2.744  0.00841 **
## jantemp     -1.980e+00  5.802e-01  -3.412  0.00128 **
## julytemp    -2.173e+00  1.628e+00  -1.335  0.18793
## house       -5.503e+01  4.504e+01  -1.222  0.22757
## educ        -1.376e+01  6.784e+00  -2.028  0.04792 *
## density      3.128e-03  3.506e-03   0.892  0.37655
## nonwhite     5.032e+00  9.315e-01   5.402 1.83e-06 ***
## hc_log      -2.848e+01  1.463e+01  -1.947  0.05713 .
## nox_log      3.765e+01  1.259e+01   2.989  0.00433 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32.2 on 50 degrees of freedom
## Multiple R-squared:  0.7729, Adjusted R-squared:  0.732
## F-statistic: 18.91 on 9 and 50 DF,  p-value: 3.078e-13
```

A one-unit increase in `hc_log` is associated with a 28.48 unit *decrease* in predicted mortality, holding other variables in the model constant.

A one-unit increase in `nox_log` is associated with a 37.65 unit *increase* in predicted mortality, holding other variables in the model constant.