MA3505 Multivariate Statistics Project 1

April 25, 2016

1 Introduction and exploratory data analysis for the variables.

2 Analysis to answer each research question

2.1 Question 1

2.2 Question 2

2.3 Question 3

2.3.1 Cleveland

From running variance inflation factor we get the following

age	sex	cp	trestbps	htn	chol	cigs	years
2.070591	2.379469	1.683710	2.935706	1.734144	1.326342	2.346224	2.315459
fbs	famhist	restecg	ekgmo	ekgday	ekgyr	dig	prop
1.281244	1.291443	1.338021	14.903816	3.357399	78.992867	1.296383	1.679766
nitr	pro	diuretic	thaldur	thaltime	met	thalach	thalrest
1.546570	1.415979	1.480903	9.549788	1.422540	10.328475	2.868773	1.713892
tpeakbps	tpeakbpd	trestbpd	exang	xhypo	oldpeak	slope	rldv5e
2.829387	2.173463	2.785971	1.734917	1.870852	2.831028	2.291928	1.557587
ca	thal	cmo	cday	cyr	lmt	ladprox	laddist
1.841289	2.051953	15.389866	3.413846	80.511913	1.401270	1.496650	1.526869
cxmain	om1	rcaprox	rcadist				
1.543251	1.789705	1.764053	1.835745				

Here we see the variables, ekgmo, ekgyr, cmo and cyr are collinear with other variables in the model.

From PCA we have the following importance of variables:

```
Importance of components:
                          Comp.1     Comp.2     Comp.3     Comp.4     Comp.5
Standard deviation     2.3133041 1.80044630 1.68730136 1.60245102 1.51366282
Proportion of Variance 0.1216222 0.07367288 0.06470422 0.05836021 0.05207216
Cumulative Proportion  0.1216222 0.19529506 0.25999929 0.31835950 0.37043166
                          Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
Standard deviation     1.34915434 1.29424733 1.26446407 1.19770134 1.15778216
Proportion of Variance 0.04136858 0.03806991 0.03633794 0.03260201 0.03046499
Cumulative Proportion  0.41180024 0.44987015 0.48620809 0.51881010 0.54927509
                         Comp.11    Comp.12    Comp.13    Comp.14    Comp.15
Standard deviation     1.14128616 1.12936676 1.09814804 1.05554851 1.0455419
Proportion of Variance 0.02960305 0.02898794 0.02740748 0.02532233 0.0248445
Cumulative Proportion  0.57887814 0.60786608 0.63527356 0.66059589 0.6854404
                         Comp.16    Comp.17    Comp.18    Comp.19   Comp.20
Standard deviation     1.0311984 0.95973900 0.92955111 0.91309717 0.8878647
Proportion of Variance 0.0241675 0.02093407 0.01963785 0.01894878 0.0179160
```

```
Cumulative Proportion  0.7096079 0.73054196 0.75017981 0.76912859 0.7870446
                          Comp.21    Comp.22     Comp.23     Comp.24     Comp.25
Standard deviation     0.87569299 0.8660316 0.84281890 0.82420668 0.80847278
Proportion of Variance 0.01742814 0.0170457 0.01614417 0.01543901 0.01485519
Cumulative Proportion  0.80447273 0.8215184 0.83766260 0.85310161 0.86795680
                          Comp.26     Comp.27    Comp.28     Comp.29     Comp.30
Standard deviation     0.77145114 0.75097291 0.7317500 0.69473103 0.68409574
Proportion of Variance 0.01352584 0.01281728 0.0121695 0.01096935 0.01063607
Cumulative Proportion  0.88148264 0.89429992 0.9064694 0.91743877 0.92807483
                          Comp.31      Comp.32      Comp.33      Comp.34
Standard deviation     0.645436007 0.630410565 0.601172308 0.576105461
Proportion of Variance 0.009467901 0.009032215 0.008213821 0.007543125
Cumulative Proportion  0.937542734 0.946574949 0.954788771 0.962331896
                          Comp.35      Comp.36      Comp.37      Comp.38
Standard deviation     0.549458611 0.529360490 0.511247676 0.466306978
Proportion of Variance 0.006861472 0.006368694 0.005940322 0.004941868
Cumulative Proportion  0.969193368 0.975562062 0.981502384 0.986444252
                          Comp.39      Comp.40      Comp.41      Comp.42
Standard deviation     0.454444269 0.396139455 0.368040283 0.232178628
Proportion of Variance 0.004693627 0.003566511 0.003078492 0.001225157
Cumulative Proportion  0.991137879 0.994704390 0.997782882 0.999008039
                          Comp.43      Comp.44
Standard deviation     0.1936567249 0.0783795601
Proportion of Variance 0.0008523393 0.0001396217
Cumulative Proportion  0.9998603783 1.0000000000
```

Here we see that it is need for first 21 components in order to keep 80% of the variance.

2.3.2 Hungary

2.3.3 Longbeach

2.3.4 Switzerland

2.4 Question 4

3 Summary