BIOS6621-Homework3

Randy

9/24/2019

VAdata3 <- read.csv("C:/Users/Goodgolden5/Desktop/BIOS6621-Gary Grunwald/Week 04/VAdata3.csv")  
str(VAdata3)

## 'data.frame': 26255 obs. of 7 variables:  
## $ hospcode: int 1 1 1 1 1 1 1 1 1 1 ...  
## $ sixmonth: int 34 34 34 34 34 34 34 34 34 34 ...  
## $ proced : int 1 1 1 1 1 1 1 1 1 0 ...  
## $ weight : int 173 162 189 208 140 167 186 172 179 167 ...  
## $ height : int 59 58 57 52 53 60 60 56 59 56 ...  
## $ bmi : int 25 24 29 38 25 23 26 28 26 27 ...  
## $ asa : int 3 4 4 4 3 4 4 4 4 4 ...

summary(VAdata3)

## hospcode sixmonth proced weight   
## Min. : 1.00 Min. :34.00 Min. :0.000 Min. : 41   
## 1st Qu.:11.00 1st Qu.:35.00 1st Qu.:1.000 1st Qu.:158   
## Median :22.00 Median :36.00 Median :1.000 Median :178   
## Mean :22.41 Mean :36.32 Mean :0.804 Mean :174   
## 3rd Qu.:33.00 3rd Qu.:37.00 3rd Qu.:1.000 3rd Qu.:197   
## Max. :44.00 Max. :39.00 Max. :2.000 Max. :288   
## NA's :106 NA's :104   
## height bmi asa   
## Min. :45 Min. : 3.00 Min. :1.000   
## 1st Qu.:56 1st Qu.:24.00 1st Qu.:4.000   
## Median :58 Median :27.00 Median :4.000   
## Mean :58 Mean :27.26 Mean :3.716   
## 3rd Qu.:60 3rd Qu.:30.00 3rd Qu.:4.000   
## Max. :70 Max. :75.00 Max. :5.000   
## NA's :104 NA's :149

VAdata3$proced <- as.character(VAdata3$proced)

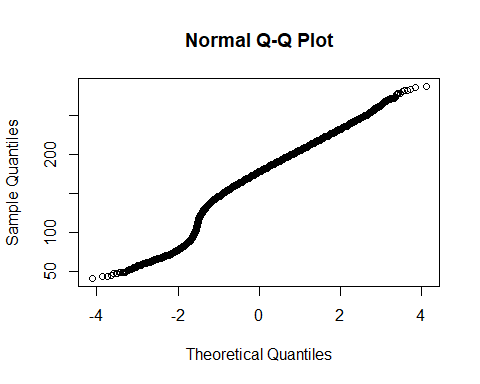
Based on the summary of dataset DVdata, I will assume *hospcode*, *preced*, *sixmonth*, and *asa* are categorical data; and *weight*, *height*, and *bmi* are continueous variables.

I probably will change the data into related data type before further analysis, other than leave the data as integeric form.

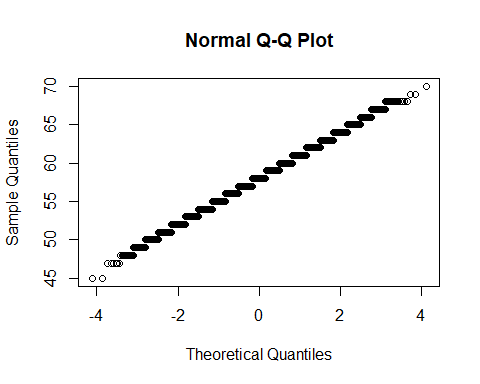
There are few things we should pay more attention:

* 1. For *hospcode*, *preced*, *sixmonth*, and *asa*, those are categorical data with integer-levels, so we do not want to see decimals or other type numiric data type here. There would be more detailed exams for these type of errors.
  2. There are several NA missing data in several columns.
  3. Specifically for *proced*, I would assume the data should be binary varible, however the value of 2 show up. This might be just typo or maybe the data was collected for extra information, for example with both procedures or did it twice.

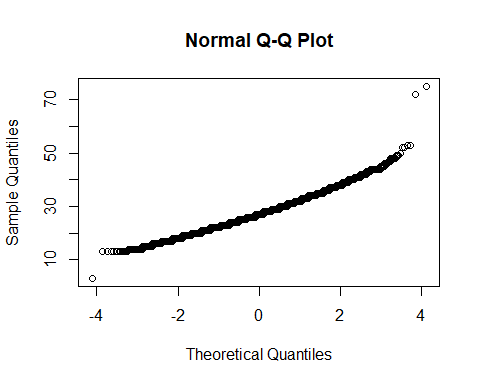
qqnorm(VAdata3$weight)



qqnorm(VAdata3$height)



qqnorm(VAdata3$bmi)



Then we can exam the normality of *height*, and *$bmi*; and there are several extram values.

* 1. The BMI has few valules over 40, the max is 70, which seems impossible for real data. I would guess the mix used metric/imperial unit cause the problem.
  2. I would comfirm the unit levels for the weight and height, then correct the errors.

VAdata3[!complete.cases(VAdata3),]

## hospcode sixmonth proced weight height bmi asa  
## 332 1 37 <NA> 190 60 26 2  
## 381 1 37 <NA> 135 54 24 4  
## 397 1 37 <NA> 142 56 23 2  
## 450 1 37 1 167 59 24 NA  
## 468 1 37 0 183 61 25 NA  
## 483 1 37 <NA> 216 54 38 4  
## 501 1 37 1 197 51 39 NA  
## 626 1 39 <NA> 90 56 32 4  
## 654 1 39 1 111 56 40 NA  
## 960 2 37 <NA> 214 64 26 3  
## 1024 2 37 <NA> 146 55 25 4  
## 1066 2 37 1 140 54 24 NA  
## 1245 2 39 <NA> 83 48 39 4  
## 1284 3 34 1 200 62 27 NA  
## 1425 3 35 1 231 57 36 NA  
## 1512 3 36 0 182 59 26 NA  
## 1898 4 34 <NA> 176 58 26 3  
## 1902 4 34 1 161 55 27 NA  
## 2038 4 35 1 189 59 27 NA  
## 2170 4 37 <NA> 163 59 24 3  
## 2329 4 37 1 145 57 23 NA  
## 2399 4 39 <NA> 86 58 28 2  
## 2513 5 34 0 198 64 24 NA  
## 2605 5 35 1 184 54 32 NA  
## 2701 5 36 <NA> 227 58 34 4  
## 3046 6 34 0 163 61 22 NA  
## 3141 6 35 <NA> 128 57 20 2  
## 3305 6 36 <NA> 215 58 32 3  
## 3384 6 37 1 209 59 30 NA  
## 3536 6 37 <NA> 181 61 25 4  
## 3566 6 39 1 83 62 24 NA  
## 3744 7 35 1 136 59 20 NA  
## 3745 7 35 <NA> 187 60 26 4  
## 3806 7 35 0 201 60 29 NA  
## 3969 7 37 <NA> 186 58 28 4  
## 3972 7 37 1 157 54 27 NA  
## 3991 7 37 0 195 58 30 NA  
## 4114 7 37 1 198 57 30 NA  
## 4184 7 39 1 73 63 20 NA  
## 4203 7 39 1 105 58 34 NA  
## 4236 8 34 1 198 60 28 NA  
## 4253 8 34 <NA> 194 59 28 3  
## 4289 8 34 <NA> 235 57 36 4  
## 4312 8 34 <NA> 222 56 35 4  
## 4440 8 36 0 221 58 33 NA  
## 4506 8 36 1 193 60 27 NA  
## 4862 9 34 <NA> 156 54 27 4  
## 4959 9 35 1 171 60 24 NA  
## 4985 9 35 1 177 60 25 NA  
## 5045 9 35 1 196 53 35 NA  
## 5102 9 36 <NA> 156 56 25 4  
## 5145 9 37 1 196 63 25 NA  
## 5173 9 37 1 172 59 25 NA  
## 5212 9 37 <NA> 161 60 23 3  
## 5315 9 37 0 165 55 27 NA  
## 5424 9 39 1 80 59 26 NA  
## 5471 10 34 1 180 56 29 NA  
## 5575 10 35 1 179 60 25 NA  
## 5921 10 37 <NA> 184 62 24 4  
## 6147 11 35 1 187 60 26 NA  
## 6258 11 36 <NA> 151 58 23 3  
## 6343 11 37 1 190 63 24 NA  
## 6463 11 37 1 157 55 26 NA  
## 6510 11 39 1 87 52 35 NA  
## 6516 11 39 1 86 57 30 NA  
## 6549 11 39 1 89 62 26 NA  
## 6712 12 35 <NA> 195 57 30 4  
## 6720 12 35 <NA> 156 60 22 4  
## 6843 12 36 1 213 56 34 NA  
## 7366 13 35 <NA> 169 62 22 4  
## 7447 13 36 <NA> 210 59 31 4  
## 7463 13 36 <NA> 175 58 26 4  
## 7546 13 37 <NA> 201 57 31 4  
## 7616 13 37 1 234 56 38 NA  
## 7749 13 39 <NA> 75 56 27 4  
## 7785 14 34 <NA> 213 56 35 4  
## 7815 14 34 <NA> 181 57 28 4  
## 7819 14 34 <NA> 164 63 21 4  
## 7866 14 34 <NA> 152 53 27 4  
## 7880 14 35 <NA> 185 55 31 4  
## 8134 14 37 1 193 52 36 NA  
## 8136 14 37 <NA> 178 60 25 3  
## 8245 14 37 1 166 65 20 NA  
## 8256 14 37 <NA> 213 59 31 NA  
## 8395 15 34 <NA> 209 63 27 4  
## 8410 15 34 1 221 59 33 NA  
## 8419 15 34 1 172 59 25 NA  
## 8487 15 35 1 194 59 28 NA  
## 8935 15 39 <NA> 69 57 24 3  
## 9169 16 36 0 221 56 36 NA  
## 9191 16 36 1 145 59 21 NA  
## 9204 16 36 1 163 63 21 NA  
## 9249 16 36 1 171 55 29 NA  
## 9340 16 37 1 156 54 27 NA  
## 9384 16 37 <NA> 175 61 24 4  
## 9492 16 39 1 77 57 27 NA  
## 9509 16 39 0 74 56 26 NA  
## 9549 16 39 0 66 60 20 NA  
## 9610 17 34 <NA> 171 54 30 4  
## 9745 17 35 <NA> 154 58 23 4  
## 9868 17 37 1 227 55 38 NA  
## 9969 17 37 <NA> 185 55 31 3  
## 10115 17 39 <NA> 178 55 29 4  
## 10350 18 35 1 140 62 18 NA  
## 10387 18 36 1 171 59 25 NA  
## 10724 18 39 1 164 53 29 NA  
## 10745 18 39 1 213 60 30 NA  
## 10991 19 36 1 192 55 32 NA  
## 11067 19 36 <NA> 217 61 29 4  
## 11173 19 37 1 195 54 34 NA  
## 11297 19 39 0 176 57 27 NA  
## 11401 20 34 1 132 58 20 NA  
## 11523 20 35 <NA> 167 61 23 4  
## 11627 20 36 1 192 62 25 NA  
## 11630 20 36 <NA> 211 62 28 3  
## 11911 20 39 1 181 58 28 NA  
## 12063 21 35 1 151 59 22 NA  
## 12077 21 35 0 216 57 33 NA  
## 12327 21 37 <NA> 149 59 22 4  
## 12581 22 34 1 213 55 35 NA  
## 12595 22 34 1 181 52 34 NA  
## 12828 22 36 1 141 57 22 NA  
## 12925 22 37 1 167 53 31 NA  
## 12968 22 37 <NA> 167 56 27 4  
## 13006 22 37 <NA> 165 58 25 4  
## 13013 22 37 1 189 59 28 NA  
## 13070 22 39 <NA> 135 55 22 4  
## 13073 22 39 <NA> 170 58 25 4  
## 13437 23 36 1 144 63 19 NA  
## 13442 23 36 1 169 54 30 NA  
## 13447 23 36 <NA> 120 56 19 2  
## 13622 23 37 <NA> 181 61 24 4  
## 13718 23 39 <NA> 197 51 38 4  
## 13759 24 34 1 185 57 29 NA  
## 13786 24 34 0 188 60 27 NA  
## 13943 24 35 <NA> 124 57 19 4  
## 14017 24 36 1 189 60 27 NA  
## 14042 24 36 1 158 58 23 NA  
## 14105 24 37 <NA> 177 57 27 4  
## 14254 24 37 1 171 57 26 NA  
## 14426 25 34 0 170 56 27 NA  
## 14479 25 34 <NA> 182 62 24 4  
## 14582 25 35 <NA> 167 58 25 4  
## 14686 25 36 1 198 61 27 NA  
## 15001 26 34 <NA> 192 53 35 4  
## 15214 26 35 <NA> 185 56 30 4  
## 15230 26 36 <NA> 174 57 27 4  
## 15620 26 39 <NA> 205 59 30 3  
## 15629 26 39 <NA> 173 54 30 4  
## 15718 27 34 1 149 58 23 NA  
## 15801 27 35 1 217 57 34 NA  
## 15971 27 37 1 177 58 27 NA  
## 16021 27 37 1 182 58 28 NA  
## 16042 27 37 <NA> 185 52 34 4  
## 16175 27 39 1 189 58 28 NA  
## 16591 28 37 0 179 56 29 NA  
## 16788 28 39 <NA> 186 59 27 3  
## 16905 29 34 1 158 60 22 NA  
## 17030 29 35 <NA> 145 62 19 2  
## 17059 29 36 1 212 57 33 NA  
## 17126 29 36 1 185 59 27 NA  
## 17308 29 37 1 119 57 18 NA  
## 17319 29 37 <NA> 192 54 34 4  
## 17390 29 39 <NA> 192 58 29 4  
## 17574 30 35 1 165 58 25 NA  
## 17604 30 35 1 213 59 31 NA  
## 17640 30 35 1 164 58 25 NA  
## 17694 30 36 <NA> 196 54 34 3  
## 17752 30 37 1 170 57 26 NA  
## 17789 30 37 <NA> 175 54 30 4  
## 17896 30 39 1 NA 58 NA 3  
## 17897 30 39 1 NA 52 NA 4  
## 17898 30 39 1 NA 56 NA 4  
## 17899 30 39 1 NA 61 NA 4  
## 17900 30 39 1 NA 61 NA 4  
## 17901 30 39 1 NA 57 NA 3  
## 17902 30 39 1 NA 59 NA 4  
## 17903 30 39 1 NA 56 NA 4  
## 17904 30 39 0 NA 61 NA 4  
## 17905 30 39 1 NA 54 NA 4  
## 17906 30 39 0 NA 57 NA 4  
## 17907 30 39 0 NA 62 NA 4  
## 17908 30 39 1 NA 57 NA 4  
## 17909 30 39 0 NA 60 NA 4  
## 17910 30 39 1 NA 58 NA 4  
## 17911 30 39 1 NA 57 NA 4  
## 17912 30 39 1 NA 62 NA 4  
## 17913 30 39 1 NA 55 NA 4  
## 17914 30 39 1 NA 58 NA 4  
## 17915 30 39 1 NA 55 NA NA  
## 17916 30 39 1 NA 60 NA 4  
## 17917 30 39 1 NA 53 NA 2  
## 17918 30 39 0 NA 62 NA 4  
## 17919 30 39 1 NA 58 NA 4  
## 17920 30 39 1 NA 56 NA 4  
## 17921 30 39 1 NA 63 NA 4  
## 17922 30 39 1 NA 59 NA 4  
## 17923 30 39 1 NA 61 NA 3  
## 17924 30 39 1 NA 55 NA 4  
## 17925 30 39 1 NA 57 NA 3  
## 17926 30 39 1 NA 57 NA 3  
## 17927 30 39 1 NA 60 NA 3  
## 17928 30 39 1 NA 57 NA 2  
## 17929 30 39 1 NA 62 NA 4  
## 17930 30 39 1 NA 58 NA 3  
## 17931 30 39 1 NA 58 NA 4  
## 17932 30 39 1 NA 54 NA 4  
## 17933 30 39 0 NA 61 NA 3  
## 17934 30 39 0 NA 55 NA 4  
## 17935 30 39 0 NA 56 NA 4  
## 17936 30 39 1 NA 59 NA 4  
## 17937 30 39 1 NA 59 NA 4  
## 17938 30 39 1 NA 57 NA 4  
## 17939 30 39 1 NA 60 NA 4  
## 17940 30 39 1 NA 58 NA 4  
## 17941 30 39 0 NA 64 NA 4  
## 17942 30 39 0 NA 59 NA 3  
## 17943 30 39 1 NA 62 NA 4  
## 17944 30 39 1 NA 58 NA 4  
## 17945 30 39 0 NA 59 NA 4  
## 17946 30 39 1 NA 62 NA 4  
## 17947 30 39 1 NA 58 NA 4  
## 17948 30 39 1 NA 57 NA 4  
## 17949 30 39 1 NA 57 NA 4  
## 17950 30 39 1 NA 65 NA 4  
## 17951 30 39 1 NA 58 NA 4  
## 17952 30 39 1 NA 61 NA 4  
## 17953 30 39 1 NA 61 NA 3  
## 17954 30 39 0 NA 59 NA 4  
## 17955 30 39 1 NA 58 NA 4  
## 17956 30 39 1 NA 58 NA 4  
## 17957 30 39 0 NA 63 NA 4  
## 17958 30 39 1 NA 57 NA 4  
## 17959 30 39 0 NA 59 NA 4  
## 17960 30 39 0 NA 58 NA 4  
## 17961 30 39 1 NA 57 NA 4  
## 17962 30 39 1 NA 60 NA 4  
## 17963 30 39 1 NA 53 NA 4  
## 17964 30 39 1 NA 54 NA 4  
## 17965 30 39 1 NA 57 NA 4  
## 17966 30 39 1 NA 60 NA 4  
## 17967 30 39 1 NA 56 NA 4  
## 17968 30 39 0 NA 59 NA 2  
## 17969 30 39 1 NA 62 NA 4  
## 17970 30 39 0 NA 54 NA 4  
## 17971 30 39 1 NA 59 NA 4  
## 17972 30 39 1 NA 56 NA 4  
## 17973 30 39 1 NA 60 NA 4  
## 17974 30 39 0 NA 55 NA 3  
## 17975 30 39 1 NA 64 NA 4  
## 17976 30 39 1 NA 62 NA 4  
## 17977 30 39 1 NA 53 NA 4  
## 17978 30 39 1 NA 61 NA 4  
## 17979 30 39 1 NA 60 NA 4  
## 17980 30 39 0 NA 58 NA 4  
## 17981 30 39 1 NA 60 NA 4  
## 17982 30 39 1 NA 55 NA 4  
## 17983 30 39 0 NA 54 NA 4  
## 17984 30 39 1 NA 52 NA 4  
## 17985 30 39 1 NA 62 NA 4  
## 17986 30 39 1 NA 64 NA 4  
## 17987 30 39 0 NA 53 NA 4  
## 17988 30 39 1 NA 54 NA 3  
## 17989 30 39 0 NA 61 NA 4  
## 17990 30 39 1 NA 59 NA 4  
## 17991 30 39 1 NA 59 NA 4  
## 17992 30 39 1 NA 58 NA 4  
## 17993 30 39 1 NA 61 NA 4  
## 17994 30 39 0 NA 59 NA 2  
## 17995 30 39 1 NA 67 NA 4  
## 17996 30 39 1 NA 57 NA 3  
## 17997 30 39 0 NA 54 NA 4  
## 17998 30 39 1 NA 60 NA 3  
## 17999 30 39 1 NA 57 NA 4  
## 18033 31 34 1 140 65 17 NA  
## 18073 31 34 1 219 65 26 NA  
## 18084 31 34 1 189 53 33 NA  
## 18337 31 37 <NA> 172 57 27 4  
## 18486 31 37 1 223 64 28 NA  
## 18527 31 39 1 240 59 35 NA  
## 18598 31 39 <NA> 203 58 30 4  
## 18735 32 35 <NA> 184 56 29 3  
## 18948 32 37 1 144 60 20 NA  
## 19000 32 37 0 200 56 32 NA  
## 19304 33 35 0 213 59 30 NA  
## 19353 33 36 1 196 61 26 NA  
## 19514 33 37 <NA> 179 58 27 4  
## 19540 33 37 1 202 55 34 NA  
## 19625 33 37 0 160 57 25 NA  
## 19633 33 37 <NA> 222 61 30 4  
## 19857 34 35 1 182 58 28 NA  
## 20026 34 36 1 141 55 23 NA  
## 20090 34 37 1 184 55 31 NA  
## 20201 34 37 1 192 52 36 NA  
## 20693 35 37 1 182 64 23 NA  
## 20722 35 37 1 174 56 28 NA  
## 21035 36 35 <NA> 167 61 23 NA  
## 21097 36 35 1 261 56 42 NA  
## 21134 36 36 1 147 54 26 NA  
## 21164 36 36 <NA> 241 58 37 4  
## 21222 36 37 <NA> 230 58 35 4  
## 21324 36 37 1 144 57 23 NA  
## 21468 36 39 <NA> 182 54 31 2  
## 21541 37 34 1 200 58 30 NA  
## 21551 37 34 1 160 56 26 NA  
## 21707 37 35 1 219 60 31 NA  
## 21740 37 36 1 202 62 26 NA  
## 21894 37 37 <NA> 182 57 29 4  
## 21911 37 37 <NA> 196 55 32 4  
## 22131 38 34 1 220 57 34 NA  
## 22200 38 34 1 172 59 25 NA  
## 22433 38 37 0 188 55 32 NA  
## 22485 38 37 <NA> 148 57 23 3  
## 22651 38 39 <NA> 166 53 30 4  
## 22669 38 39 1 169 58 26 NA  
## 22901 39 36 1 210 54 36 NA  
## 22980 39 36 1 142 57 22 NA  
## 22982 39 36 1 231 57 36 NA  
## 23033 39 37 0 162 56 26 NA  
## 23135 39 37 <NA> 177 56 28 4  
## 23198 39 39 <NA> 203 58 31 4  
## 23256 39 39 1 171 57 27 NA  
## 23284 39 39 1 181 57 28 NA  
## 23321 40 34 <NA> 182 53 32 4  
## 23374 40 34 1 142 59 21 NA  
## 23486 40 35 <NA> 213 62 28 4  
## 23552 40 36 <NA> 147 57 23 4  
## 23657 40 37 1 172 55 28 NA  
## 23851 40 39 <NA> 145 59 21 4  
## 23876 40 39 <NA> 197 57 31 4  
## 24298 41 37 <NA> 200 61 27 4  
## 24338 41 37 1 216 54 37 NA  
## 24369 41 37 <NA> 162 61 22 4  
## 24506 41 39 1 188 57 29 NA  
## 24768 42 36 <NA> 181 52 33 4  
## 24801 42 36 1 194 63 24 NA  
## 24854 42 37 <NA> 206 65 25 3  
## 25016 42 39 <NA> 211 55 35 4  
## 25165 43 34 1 169 56 27 NA  
## 25188 43 34 1 139 62 18 NA  
## 25199 43 34 1 210 54 37 NA  
## 25407 43 36 1 172 57 27 NA  
## 25412 43 37 <NA> 154 55 26 4  
## 25415 43 37 0 158 60 22 NA  
## 25452 43 37 <NA> 180 50 36 3  
## 25537 43 37 1 139 59 20 NA  
## 25577 43 37 1 271 63 35 NA  
## 25609 43 39 <NA> 201 62 26 4  
## 25643 43 39 <NA> 100 54 18 4  
## 25656 43 39 <NA> 176 61 24 3  
## 25844 44 35 <NA> 162 51 31 4  
## 25859 44 36 1 152 59 22 NA  
## 25931 44 36 <NA> 189 60 27 3  
## 26054 44 37 <NA> 192 59 28 3  
## 26189 44 39 1 129 51 25 NA  
## 26192 44 39 1 103 62 13 NA

There are a few individuals with missing values. Here I list out all the missing values. I probably will confirm with the researchers or make proper assumptions before the tests, other than delete the those data.