Outpatient Survey Analysis

Anupam kumar Singh, MD

28 October 2017

The outpatient health survey was conducted to assess satisfaction levels of patients in our jhajjar outreach clinic. The questionaire was adapted from a pre-existing questionaire made of 23 questions with sub-domains evaluating Interpersonal Skills, Physical Environment , Availability, Quality and Accessibility on a Likert scale. One question evaluated overall global satisfaction with physician. Our Goal in this study is to

1. Assess Distribution of Scores across various questions.
2. Assess Distribution of Scores across sub-domains.
3. correlation of of age,education and income with mean score
4. Assess which sub-domain has highest score and conduct an Anova analysis.
5. Evaluate how the various sub-domains affect via multiple linear regression.
6. Assess if five domains distinguish themselves on confirmatory factor analysis.
7. Test Divergent validity, cronbach alpha

so lets get started

Assess Distribution of Scores across various questions

```
jhar %>% dplyr::dplyr::select(one:Twenty.three) %>% gather(key="Question" ,value="Score") %>% group_by(
```

Let us see summary of scores

```
jhar6= read.csv("jhar6.csv")
jharx= read.csv("jharx.csv")


jhar= read.csv("jhar.csv")


jhar %>% dplyr::select(one:Twenty.three) %>% summary()
```

```
##       one             Two            Three           Four      
##  Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.000  
##  1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.00   1st Qu.:4.000  
##  Median :5.000   Median :5.000   Median :5.00   Median :5.000  
##  Mean   :4.687   Mean   :4.565   Mean   :4.56   Mean   :4.644  
##  3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.00   3rd Qu.:5.000  
##  Max.   :5.000   Max.   :5.000   Max.   :5.00   Max.   :5.000  
##       Five           Six             Seven           Eight      
##  Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000  
##  1st Qu.:4.00   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.000  
##  Median :5.00   Median :5.000   Median :4.000   Median :4.000  
##  Mean   :4.46   Mean   :4.393   Mean   :4.119   Mean   :4.266  
##  3rd Qu.:5.00   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000  
##  Max.   :5.00   Max.   :5.000   Max.   :5.000   Max.   :5.000  
##       Nine             Ten            Eleven          Twelve     
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000  
##  1st Qu.:4.000   1st Qu.:4.00   1st Qu.:4.000   1st Qu.:4.000  
##  Median :5.000   Median :5.00   Median :5.000   Median :4.000  
```

```
##  Mean   :4.251   Mean   :4.54   Mean   :4.485   Mean   :4.119
##  3rd Qu.:5.000   3rd Qu.:5.00   3rd Qu.:5.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.00   Max.   :5.000   Max.   :5.000
##     Thirteen        Fourteen        Fifteen         Sixteen
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00
##  1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.00
##  Median :5.000   Median :5.000   Median :4.000   Median :5.00
##  Mean   :4.465   Mean   :4.453   Mean   :4.284   Mean   :4.54
##  3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.00
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :6.00
##    Seventeen        Eighteen       Nineteen        Twenty
##  Min.   :1.000   Min.   :1.00   Min.   :1.00   Min.   :1.00
##  1st Qu.:4.000   1st Qu.:4.00   1st Qu.:4.00   1st Qu.:4.00
##  Median :5.000   Median :5.00   Median :5.00   Median :5.00
##  Mean   :4.537   Mean   :4.53   Mean   :4.58   Mean   :4.49
##  3rd Qu.:5.000   3rd Qu.:5.00   3rd Qu.:5.00   3rd Qu.:5.00
##  Max.   :5.000   Max.   :5.00   Max.   :5.00   Max.   :5.00
##    Twenty.one      twenty.two      Twenty.three
##  Min.   :1.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.00   1st Qu.:4.000   1st Qu.:4.000
##  Median :5.00   Median :5.000   Median :5.000
##  Mean   :4.51   Mean   :4.577   Mean   :4.532
##  3rd Qu.:5.00   3rd Qu.:5.000   3rd Qu.:5.000
##  Max.   :5.00   Max.   :5.000   Max.   :5.000
```

Most of the questions have negative skew and bimodal peaks at 4 and 5 indicating high overall satisfaction. This is indicated by negative skew(Median score is 5 and higher than mean across all score) statistic calculated here .

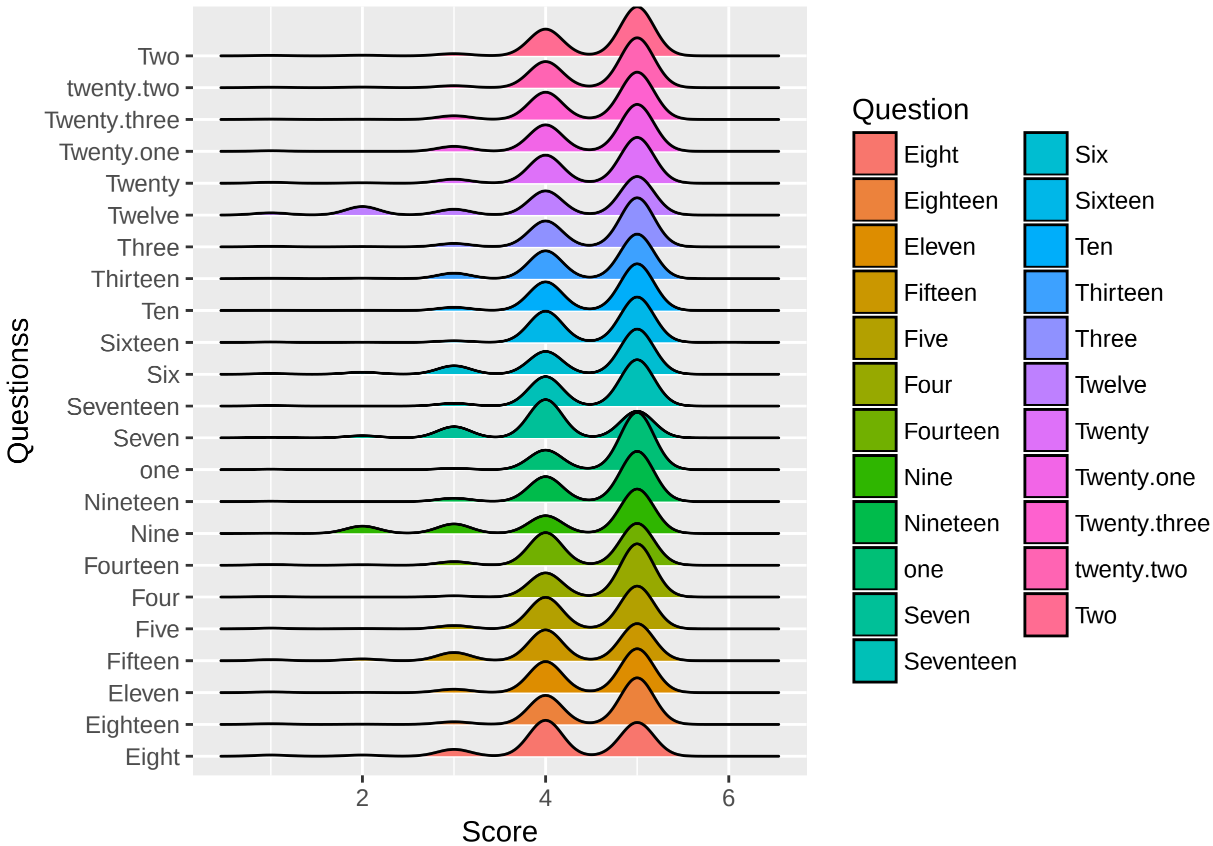
```
jhar %>% dplyr::select(one:Twenty.three) %>%map(~skewness(.))
```

```
## $one
## [1] -2.595851
##
## $Two
## [1] -1.799874
##
## $Three
## [1] -1.662334
##
## $Four
## [1] -1.985509
##
## $Five
## [1] -1.70078
##
## $Six
## [1] -1.405989
##
## $Seven
## [1] -0.8816675
##
## $Eight
## [1] -1.387662
##
```

```
## $Nine
## [1] -1.117693
##
## $Ten
## [1] -1.294795
##
## $Eleven
## [1] -1.582273
##
## $Twelve
## [1] -1.196617
##
## $Thirteen
## [1] -1.403634
##
## $Fourteen
## [1] -1.514912
##
## $Fifteen
## [1] -1.311577
##
## $Sixteen
## [1] -1.485145
##
## $Seventeen
## [1] -1.485481
##
## $Eighteen
## [1] -1.759675
##
## $Nineteen
## [1] -1.686639
##
## $Twenty
## [1] -1.780721
##
## $Twenty.one
## [1] -1.612736
##
## $twenty.two
## [1] -1.858309
##
## $Twenty.three
## [1] -1.54983
```

Let us look at joy plot which indicate bimodal peaks at 4 and 5 indicating Very satisfied(5) or Satisfied patients(4).

```
## Picking joint bandwidth of 0.182
```



Let us look at percent of neutral/dissasatisfied responses across various questions

```
jhar %>% dplyr::select(one:Twenty.three) %>% gather(key="Question" ,value="Score") %>% group_by(Questio
```

```
## # A tibble: 23 x 2
##      Question Neutral_dissatisfied_percent
##         <chr>                        <dbl>
##  1       Nine                   0.21393035
##  2     Twelve                   0.20646766
##  3      Seven                   0.17661692
##  4   Fifteen                    0.13930348
##  5        Six                   0.13930348
##  6      Eight                   0.11691542
##  7   Thirteen                   0.08208955
##  8 Twenty.one                   0.06965174
##  9     Twenty                   0.06467662
## 10       Five                   0.05721393
## # ... with 13 more rows
```

So question 6,7,8,9,12 and 15 have neutral/dissatisfied reponses from greater than 10% of patients and we need to improve on these parameters.

Let us see which questions have the highest percent of very satisfied(5) score.

```
jhar %>% dplyr::select(one:Twenty.three) %>% gather(key="Question" ,value="Score") %>% group_by(Questio
```

```
## # A tibble: 23 x 2
##        Question very_satisfied_percent
##           <chr>                  <dbl>
```

```
##   1          one               0.7263682
##   2         Four               0.6741294
##   3     Nineteen               0.6368159
##   4   twenty.two               0.6318408
##   5        Three               0.6218905
##   6          Two               0.6218905
##   7 Twenty.three               0.5995025
##   8   Twenty.one               0.5945274
##   9          Ten               0.5920398
## 10     Eighteen               0.5895522
## # ... with 13 more rows
```

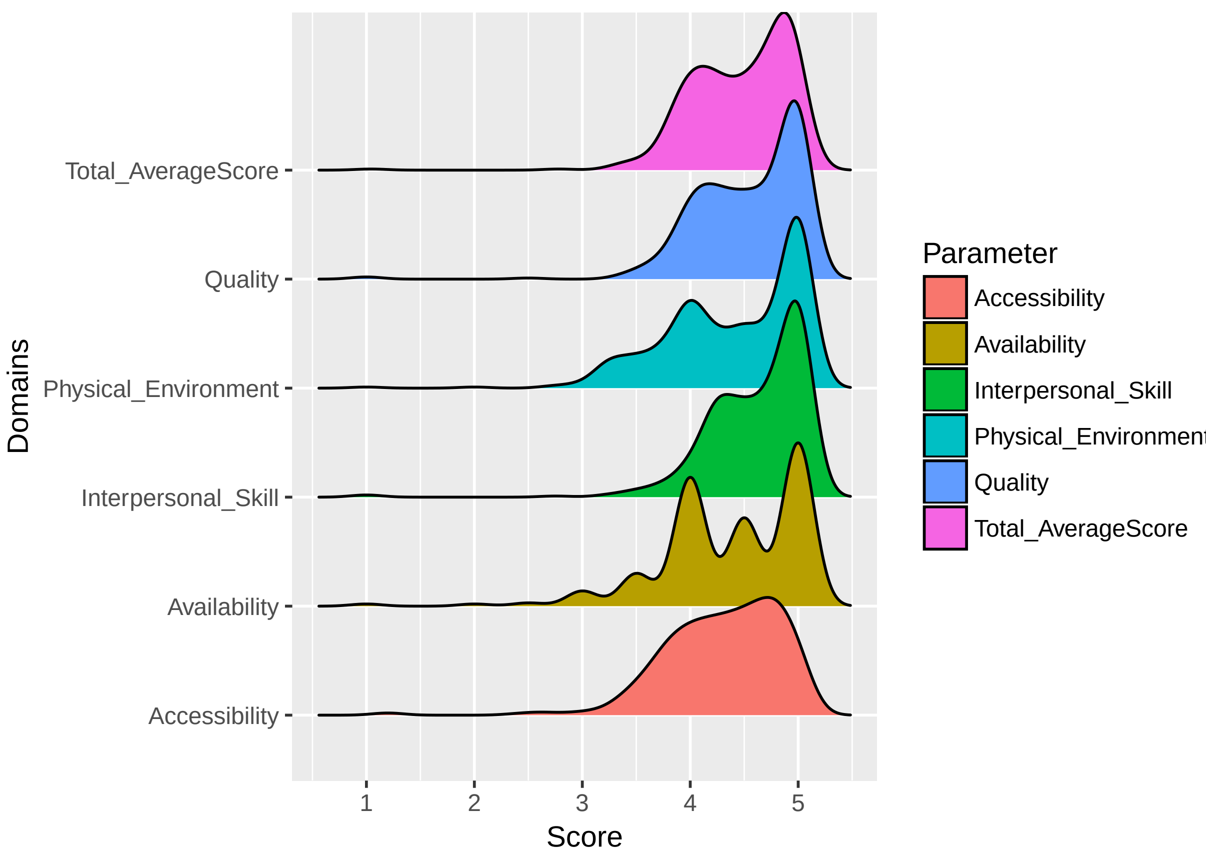
So we are doing well on 1,2,3,4 and 17-22 questions and we need to maintain our performance in these areas.

```
jhar6 %>% dplyr::select(one:Twenty.three) %>% sjp.likert(values = "hide")
```

Let us look at distribution of score across domains . First the visualisation

```
jhar %>% dplyr::select(Accessibility,Quality,Physical_Environment,Interpersonal,Availability,MeanScore)
  labs(x = "Score",y = "Domains")
```

```
## Picking joint bandwidth of 0.147
```



Now statistics part

```
jhar %>% dplyr::select(Accessibility,Quality,Physical_Environment,Interpersonal,Availability,MeanScore)
```

```
##  Accessibility      Quality      Physical_Environment Interpersonal
##  Min.   :1.200   Min.   :1.000   Min.   :1.000        Min.   :1.000
##  1st Qu.:4.000   1st Qu.:4.250   1st Qu.:4.000        1st Qu.:4.250
```

```
##  Median :4.400   Median :4.625   Median :4.500       Median :4.750
##  Mean   :4.298   Mean   :4.537   Mean   :4.402       Mean   :4.614
##  3rd Qu.:4.800   3rd Qu.:5.000   3rd Qu.:5.000       3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000       Max.   :5.000
##   Availability     MeanScore
##  Min.   :1.000   Min.   :1.043
##  1st Qu.:4.000   1st Qu.:4.130
##  Median :4.500   Median :4.543
##  Mean   :4.368   Mean   :4.464
##  3rd Qu.:5.000   3rd Qu.:4.870
##  Max.   :5.000   Max.   :5.043
```

Is there any difference in scores across domains ? Are we scoring better on some domains and lagging on others ?

We need to conduct an ANOVA test to see it..and Tukey's post Hoc correction to see intergroup differences

```
df =jhar %>% dplyr::select(Accessibility,Quality,Physical_Environment,Interpersonal,Availability) %>% ga
fitaov = aov(Score~Domains,data=df)
summary(fitaov)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Domains        4   26.6   6.661   21.05 <2e-16 ***
## Residuals   2005  634.4   0.316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So on ANOVA we see that there is significant difference across Domain Scores. ($p<0.00001$). Now we need to determine post-hoc difference after adjusting for multiple comparison by Tukey's Method

```
TukeyHSD(fitaov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Score ~ Domains, data = df)
##
## $Domains
##                                           diff          lwr         upr
## Availability-Accessibility          0.07014925 -0.038178796  0.17847730
## Interpersonal-Accessibility         0.31579602  0.207467971  0.42412407
## Physical_Environment-Accessibility  0.10435323 -0.003974815  0.21268128
## Quality-Accessibility               0.23899254  0.130664488  0.34732059
## Interpersonal-Availability          0.24564677  0.137318717  0.35397482
## Physical_Environment-Availability   0.03420398 -0.074124069  0.14253203
## Quality-Availability                0.16884328  0.060515234  0.27717133
## Physical_Environment-Interpersonal -0.21144279 -0.319770835 -0.10311474
## Quality-Interpersonal              -0.07680348 -0.185131532  0.03152457
## Quality-Physical_Environment        0.13463930  0.026311254  0.24296735
##                                        p adj
## Availability-Accessibility         0.3926757
## Interpersonal-Accessibility        0.0000000
## Physical_Environment-Accessibility 0.0653889
## Quality-Accessibility              0.0000000
## Interpersonal-Availability         0.0000000
```

```
## Physical_Environment-Availability  0.9106676
## Quality-Availability               0.0002115
## Physical_Environment-Interpersonal 0.0000011
## Quality-Interpersonal              0.2986533
## Quality-Physical_Environment       0.0063212
```

we see Quality and Interpersonal skills domains have significant higher scores than other domains like Availability,Accessibility and Physical Environment though there is not a statistically significant difference between Quality and Interpersonal Skills.

Correlation of Individual Domains with predictor variables like age,gender,Income,Education with individual Domain Scores and inter-Domain Correlation

```r
library(Hmisc)
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
    row = rownames(cormat)[row(cormat)[ut]],
    column = rownames(cormat)[col(cormat)[ut]],
    cor  =(cormat)[ut],
    p = pmat[ut]
    )
}

df2 = jhar %>% dplyr::select(Age,Income,Occupation,Education,Accessibility,Quality,Physical_Environment
res2<-rcorr(as.matrix(df2[,1:9]))
flattenCorrMatrix(res2$r, round(res2$P,3)) %>% arrange(desc(p))
```

```
##                     row               column         cor     p
## 1                   Age               Income -0.01599744 0.749
## 2                Income            Education  0.02827840 0.572
## 3                   Age        Accessibility -0.03505922 0.483
## 4                Income           Occupation -0.05451126 0.276
## 5            Occupation        Accessibility  0.06554248 0.190
## 6            Occupation        Interpersonal  0.06619538 0.185
## 7                   Age Physical_Environment -0.07169959 0.151
## 8                   Age           Occupation -0.08336982 0.095
## 9            Occupation              Quality  0.09402740 0.060
## 10           Occupation         Availability  0.09441951 0.059
## 11                  Age        Interpersonal -0.09851158 0.048
## 12                  Age              Quality -0.11185898 0.025
## 13            Education        Interpersonal  0.11821669 0.018
## 14           Occupation Physical_Environment  0.12015326 0.016
## 15               Income Physical_Environment  0.13437633 0.007
## 16            Education        Accessibility  0.13577312 0.006
## 17               Income              Quality  0.14421451 0.004
## 18               Income        Interpersonal  0.14993410 0.003
## 19                  Age         Availability -0.14545915 0.003
## 20                  Age            Education -0.43327329 0.000
## 21           Occupation            Education  0.38212746 0.000
## 22               Income        Accessibility  0.19889995 0.000
## 23            Education              Quality  0.18045050 0.000
## 24        Accessibility              Quality  0.67797613 0.000
```

```
## 25           Education Physical_Environment  0.19200496 0.000
## 26       Accessibility Physical_Environment  0.70637292 0.000
## 27             Quality Physical_Environment  0.80188841 0.000
## 28       Accessibility        Interpersonal  0.65991789 0.000
## 29             Quality        Interpersonal  0.75494307 0.000
## 30 Physical_Environment       Interpersonal  0.70554692 0.000
## 31              Income         Availability  0.20687886 0.000
## 32           Education         Availability  0.18588264 0.000
## 33       Accessibility         Availability  0.65701002 0.000
## 34             Quality         Availability  0.73149651 0.000
## 35 Physical_Environment        Availability  0.72396451 0.000
## 36       Interpersonal         Availability  0.64695626 0.000
```

We see on an average higher age,Education,Income and Occupation category is linked to higher Domain score

Let us plot a correlogram

```
library(corrplot)
```

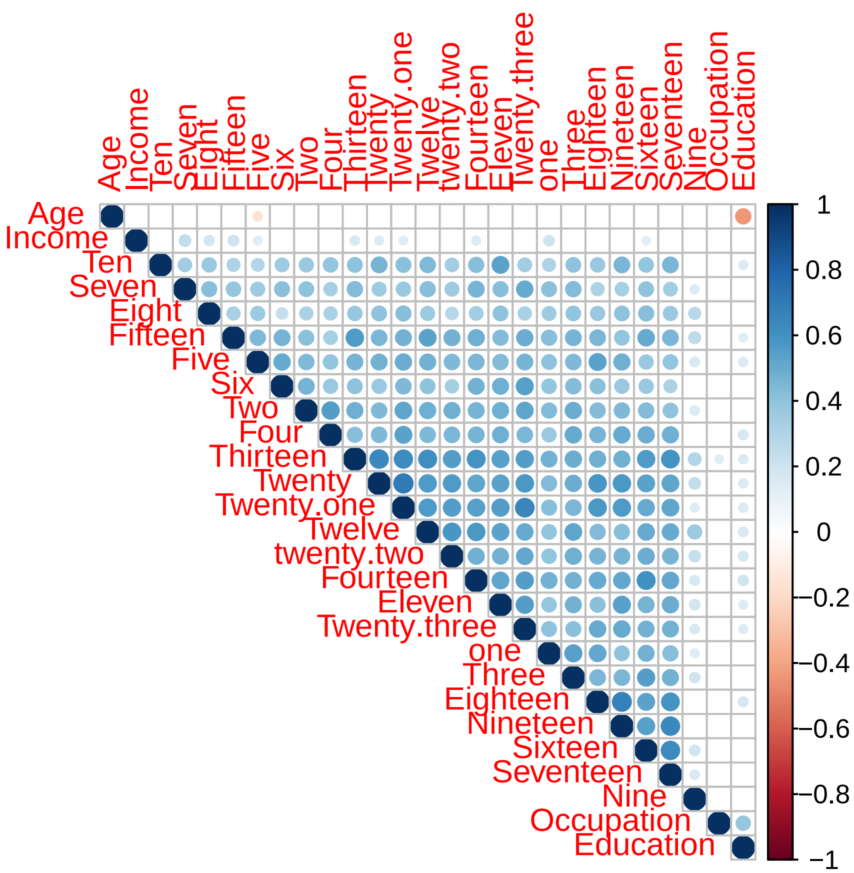
```
## corrplot 0.84 loaded
```

```
df3 = jhar %>% dplyr::select(Age,Income,Occupation,Education,one:Twenty.three)
M<-cor(df3)
head(round(M,2))
```

```
##              Age Income Occupation Education   one   Two Three  Four  Five
## Age         1.00  -0.02      -0.08     -0.43 -0.08 -0.06 -0.08 -0.09 -0.16
## Income     -0.02   1.00      -0.05      0.03  0.21  0.09  0.11  0.07  0.14
## Occupation -0.08  -0.05       1.00      0.38  0.03  0.03  0.06  0.09  0.05
## Education  -0.43   0.03       0.38      1.00  0.07  0.08  0.05  0.18  0.15
## one        -0.08   0.21       0.03      0.07  1.00  0.43  0.55  0.37  0.41
## Two        -0.06   0.09       0.03      0.08  0.43  1.00  0.50  0.56  0.44
##             Six Seven Eight Nine  Ten Eleven Twelve Thirteen Fourteen
## Age        0.00  0.04 -0.02 0.00 0.02  -0.02  -0.10    -0.09    -0.12
## Income     0.11  0.24  0.19 0.01 0.09   0.12   0.07     0.16     0.14
## Occupation 0.04  0.02  0.09 0.02 0.09   0.07   0.09     0.14     0.11
## Education  0.13  0.06  0.05 0.07 0.15   0.14   0.17     0.15     0.19
## one        0.39  0.41  0.36 0.14 0.31   0.39   0.39     0.47     0.47
## Two        0.47  0.40  0.32 0.16 0.37   0.48   0.49     0.48     0.47
##            Fifteen Sixteen Seventeen Eighteen Nineteen Twenty Twenty.one
## Age          -0.13   -0.07     -0.10    -0.12    -0.06  -0.10      -0.12
## Income        0.21    0.13      0.12     0.11     0.08   0.14       0.13
## Occupation    0.06    0.07      0.10     0.05     0.04   0.09       0.12
## Education     0.13    0.12      0.11     0.17     0.09   0.15       0.16
## one           0.43    0.47      0.43     0.52     0.40   0.43       0.42
## Two           0.42    0.43      0.40     0.43     0.44   0.44       0.52
##            twenty.two Twenty.three
## Age             -0.06        -0.07
## Income           0.07         0.12
## Occupation       0.08         0.04
## Education        0.18         0.14
## one              0.39         0.41
## Two              0.48         0.52
```

```
corrplot(M, method="number")
```

```r
cor.mtest <- function(mat, ...) {
    mat <- as.matrix(mat)
    n <- ncol(mat)
    p.mat<- matrix(NA, n, n)
    diag(p.mat) <- 0
    for (i in 1:(n - 1)) {
        for (j in (i + 1):n) {
            tmp <- cor.test(mat[, i], mat[, j], ...)
            p.mat[i, j] <- p.mat[j, i] <- tmp$p.value
        }
    }
  colnames(p.mat) <- rownames(p.mat) <- colnames(mat)
  p.mat
}
# matrix of the p-value of the correlation
p.mat <- cor.mtest(df3)

corrplot(M, type="upper", order="hclust",
         p.mat = p.mat, sig.level = 0.01, insig = "blank")
```



Now having visualised correlation plots and predictors affecting them , let's go to factors affecting mean score

```r
library(arm)
fit=lm(MeanScore~Interpersonal+Accessibility+Physical_Environment+Availability+Quality+Sex+Age+Income+E

summary(fit)
```

```
##
## Call:
## lm(formula = MeanScore ~ Interpersonal + Accessibility + Physical_Environment +
##     Availability + Quality + Sex + Age + Income + Education,
##     data = jhar6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28504 -0.02039  0.00438  0.01799  1.23743
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.350e-01  3.953e-02   8.474 4.86e-16 ***
## Interpersonal         1.230e-01  1.211e-02  10.163  < 2e-16 ***
## Accessibility         1.997e-01  1.011e-02  19.745  < 2e-16 ***
## Physical_Environment  2.232e-01  1.074e-02  20.779  < 2e-16 ***
## Availability          8.339e-02  8.786e-03   9.492  < 2e-16 ***
## Quality               2.946e-01  1.372e-02  21.479  < 2e-16 ***
## SexM                  3.087e-03  7.491e-03   0.412    0.680
## Age                  -1.172e-06  2.517e-04  -0.005    0.996
## Income                3.803e-03  3.543e-03   1.073    0.284
## Education             3.261e-03  2.442e-03   1.336    0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07045 on 392 degrees of freedom
## Multiple R-squared:  0.9768, Adjusted R-squared:  0.9762
## F-statistic:  1830 on 9 and 392 DF,  p-value: < 2.2e-16
```

We see that after controlling for Domain average score, age ,occupation and Income,education dont impact mean score

A cleaner output here

```
display(fit)
```

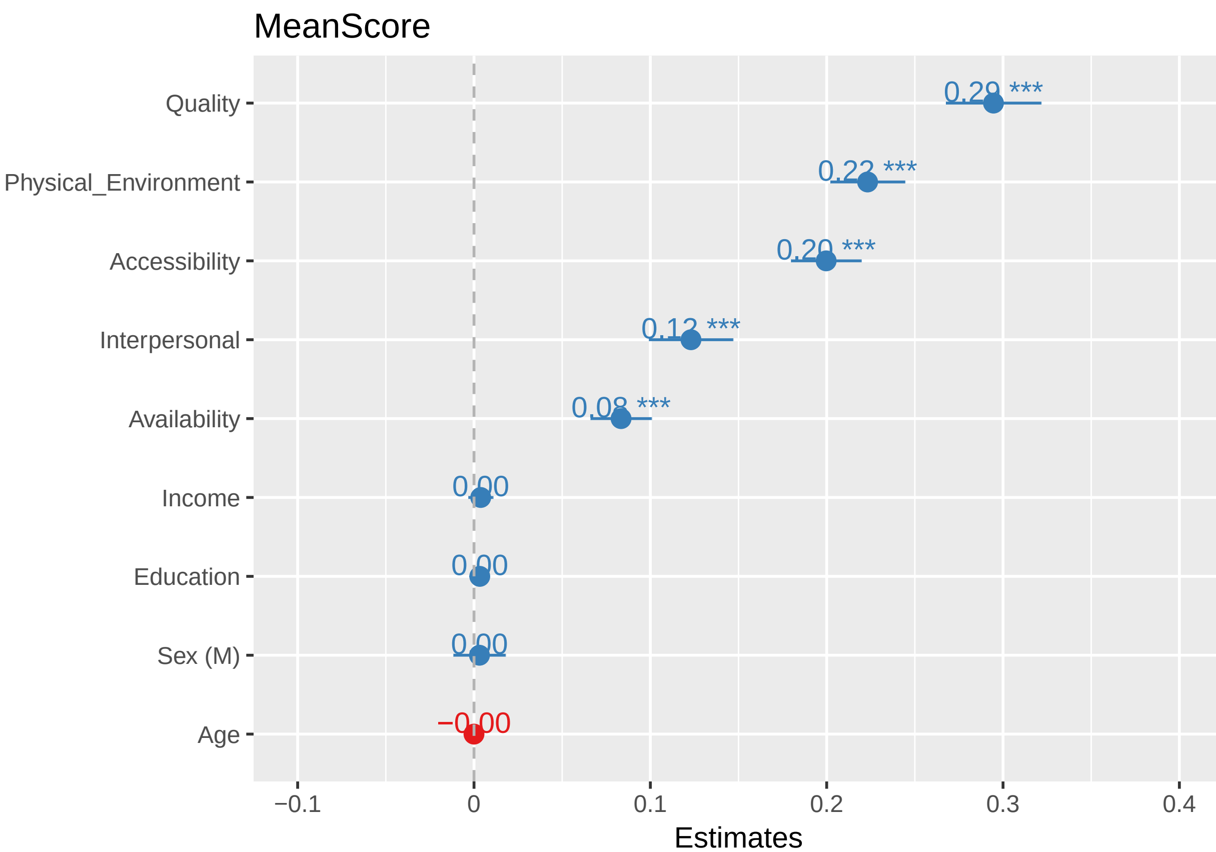
```
## lm(formula = MeanScore ~ Interpersonal + Accessibility + Physical_Environment +
##     Availability + Quality + Sex + Age + Income + Education,
##     data = jhar6)
##                      coef.est coef.se
## (Intercept)          0.33     0.04
## Interpersonal        0.12     0.01
## Accessibility        0.20     0.01
## Physical_Environment 0.22     0.01
## Availability         0.08     0.01
## Quality              0.29     0.01
## SexM                 0.00     0.01
## Age                  0.00     0.00
## Income               0.00     0.00
## Education            0.00     0.00
## ---
## n = 402, k = 10
## residual sd = 0.07, R-Squared = 0.98
```

It implies that one point improvement in Quality will lead to 0.29 point improvement in mean score while controlling for other variables. Corresponding values for other domains are 0.22 for physical environment , 0.20 for Accessibility, 0.08 for Availability and 0.12 for Interpersonal skills other variables are non-significant. It implies Quality and Physical environment play a major role in affecting average score in our study

Let us visualise the linear regression as forest plot to emphasise this impression.

```
sjp.lm(fit)
```

```
## Warning: 'sjstats::get_model_pval' is deprecated.
## Use 'p_value' instead.
## See help("Deprecated")
```



CONFIRMATORY FACTOR ANALYSIS

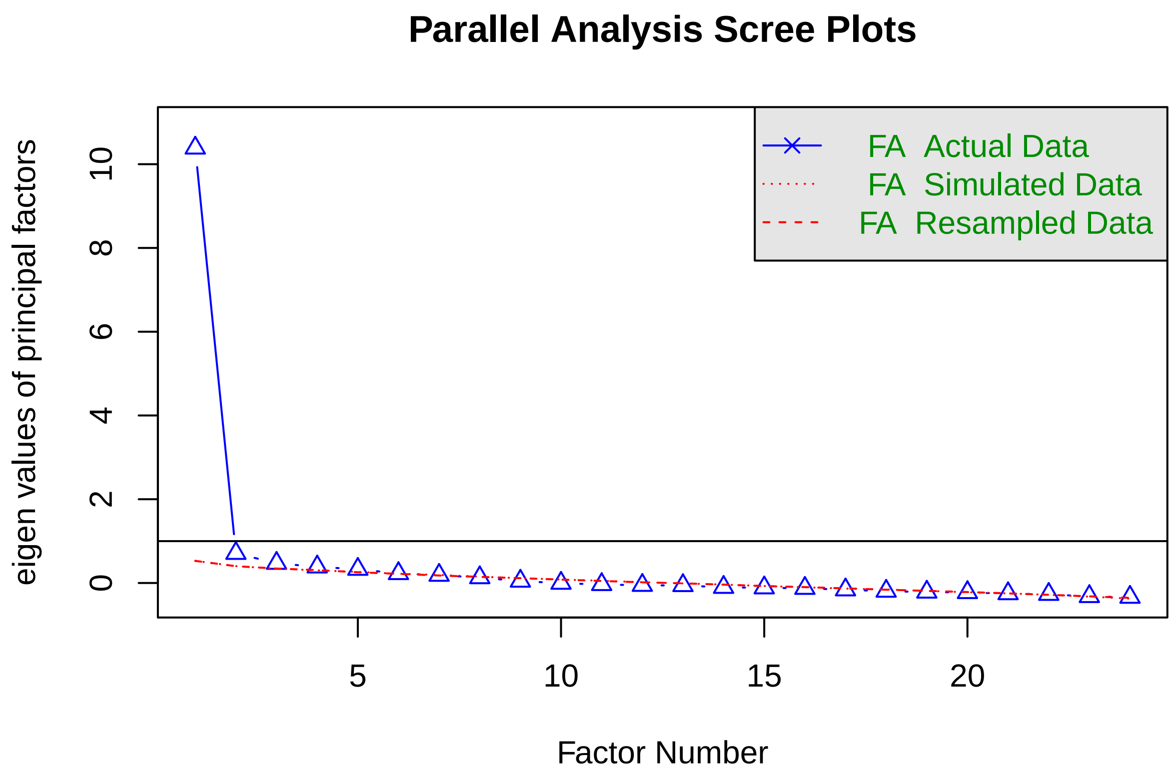
Since this core was adapted from a score used in thailand outpatient set up . One of our goals is to evaluate construct validity and if the domains differentiate between each other to five domains

```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:arm':
##
##     logit, rescale, sim

## The following object is masked from 'package:sjstats':
##
##     phi

## The following object is masked from 'package:Hmisc':
```

```
##
##     describe

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(GPArotation)
parallel <- fa.parallel(jharx, fm = 'minres', fa = 'fa')
```




```
## Parallel analysis suggests that the number of factors =  6  and the number of components =  NA
```

The blue line shows eigenvalues of actual data and the two red lines (placed on top of each other) show simulated and resampled data. Here we look at the large drops in the actual data and spot the point where it levels off to the right. Also we locate the point of inflection – the point where the gap between simulated data and actual data tends to be minimum.

Looking at this plot and parallel analysis, anywhere between 1 to 3 factors factors would be good choice instead of five proposed in original survey.

In this case, we will dplyr::select oblique rotation (rotate = "oblimin") as we believe that there is correlation in the factors. Note that Varimax rotation is used under the assumption that the factors are completely uncorrelated. We will use Ordinary Least Squared/Minres factoring (fm = "minres"), as it is known to provide results similar to Maximum Likelihood without assuming multivariate normal distribution and derives solutions through iterative eigendecomposition like principal axis.

```
fivefactor <- fa(jharx,nfactors = 5,rotate = "oblimin",fm="minres")
print(fivefactor)
```

```
sixfactor <- fa(jharx,nfactors = 6,rotate = "oblimin",fm="minres")
print(sixfactor)
```

```r
print(sixfactor$loadings,cutoff = 0.3)
```

```r
jharm = as.matrix(jharx)
cortest.bartlett(jharx)
```

```
## R was not square, finding R from data

## $chisq
## [1] 5295.678
##
## $p.value
## [1] 0
##
## $df
## [1] 276
```

For these data, Bartlett's test is highly significant,chisquare(253) =5180, p < .00001, and therefore factor analysis is appropriate.

```r
km =kmo(jharx)

list(km$overall,km$report,km$individual)
```

```
## [[1]]
## [1] 0.954155
##
## [[2]]
## [1] "The KMO test yields a degree of common variance marvelous."
##
## [[3]]
##           X         one         Two       Three        Four
##   0.8271621   0.9600921   0.9688859   0.9619139   0.9619071
##        Five         Six       Seven       Eight        Nine
##   0.9590314   0.9383805   0.9439488   0.9319054   0.8323736
##         Ten      Eleven      Twelve    Thirteen    Fourteen
##   0.9399816   0.9629930   0.9597399   0.9632573   0.9713668
##     Fifteen     Sixteen   Seventeen    Eighteen    Nineteen
##   0.9661195   0.9553893   0.9499095   0.9450933   0.9510554
##      Twenty  Twenty.one  twenty.two Twenty.three
##   0.9552838   0.9545005   0.9706909   0.9572043
```

So Both KMO test and barlett test significant.

```r
pc2 <- principal(jharx, nfactors=length(jharx), rotate="none")
pc2
```

```r
pc3 <- principal(jharx, nfactors=5, rotate="oblimin")
pc3
```

```
## Principal Components Analysis
## Call: principal(r = jharx, nfactors = 5, rotate = "oblimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                TC1   TC5   TC3   TC4   TC2   h2   u2 com
## X            -0.04  0.22  0.28  0.12  0.68 0.66 0.34 1.6
## one           0.61  0.03 -0.12  0.37  0.23 0.63 0.37 2.1
## Two           0.09  0.65  0.03  0.08 -0.03 0.55 0.45 1.1
## Three         0.40  0.21  0.06  0.35 -0.01 0.57 0.43 2.6
```

```
## Four          0.32  0.40  0.00  0.04 -0.21 0.50 0.50 2.5
## Five          0.36  0.42 -0.02  0.01  0.09 0.48 0.52 2.1
## Six          -0.02  0.82 -0.22  0.14  0.09 0.67 0.33 1.2
## Seven        -0.04  0.46  0.05  0.56  0.12 0.67 0.33 2.1
## Eight         0.28 -0.10  0.27  0.58 -0.16 0.63 0.37 2.2
## Nine         -0.09 -0.16  0.88  0.09  0.08 0.72 0.28 1.1
## Ten           0.07  0.36  0.11  0.28 -0.55 0.66 0.34 2.4
## Eleven        0.04  0.60  0.19  0.12 -0.36 0.69 0.31 2.0
## Twelve        0.13  0.43  0.50 -0.01 -0.02 0.69 0.31 2.1
## Thirteen      0.39  0.30  0.37 -0.04  0.04 0.66 0.34 2.9
## Fourteen      0.36  0.39  0.10  0.11  0.01 0.57 0.43 2.3
## Fifteen       0.35  0.29  0.23  0.06  0.28 0.56 0.44 3.8
## Sixteen       0.69 -0.02  0.13  0.19 -0.03 0.64 0.36 1.2
## Seventeen     0.78 -0.06  0.11  0.04 -0.13 0.68 0.32 1.1
## Eighteen      0.90  0.01 -0.11 -0.05  0.07 0.74 0.26 1.1
## Nineteen      0.76  0.09 -0.07 -0.01 -0.18 0.69 0.31 1.2
## Twenty        0.49  0.27  0.30 -0.14 -0.06 0.67 0.33 2.6
## Twenty.one    0.40  0.49  0.15 -0.18 -0.04 0.69 0.31 2.5
## twenty.two    0.34  0.37  0.33 -0.22 -0.01 0.60 0.40 3.6
## Twenty.three  0.18  0.69  0.07 -0.08  0.10 0.67 0.33 1.2
##
##                        TC1  TC5  TC3  TC4  TC2
## SS loadings           5.51 4.87 2.21 1.47 1.22
## Proportion Var        0.23 0.20 0.09 0.06 0.05
## Cumulative Var        0.23 0.43 0.52 0.59 0.64
## Proportion Explained  0.36 0.32 0.14 0.10 0.08
## Cumulative Proportion 0.36 0.68 0.82 0.92 1.00
##
##  With component correlations of
##       TC1   TC5  TC3  TC4   TC2
## TC1  1.00  0.60 0.35 0.25 -0.12
## TC5  0.60  1.00 0.31 0.27 -0.01
## TC3  0.35  0.31 1.00 0.18  0.00
## TC4  0.25  0.27 0.18 1.00  0.02
## TC2 -0.12 -0.01 0.00 0.02  1.00
##
## Mean item complexity =  2
## Test of the hypothesis that 5 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  526.17  with prob <  4.7e-39
##
## Fit based upon off diagonal values = 0.99
```

```r
print.psych(pc3, cut = 0.3, sort = TRUE)
```

```
## Principal Components Analysis
## Call: principal(r = jharx, nfactors = 5, rotate = "oblimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##              item   TC1   TC5   TC3   TC4   TC2   h2   u2 com
## Eighteen       19  0.90                         0.74 0.26 1.1
## Seventeen      18  0.78                         0.68 0.32 1.1
## Nineteen       20  0.76                         0.69 0.31 1.2
## Sixteen        17  0.69                         0.64 0.36 1.2
## one             2  0.61              0.37       0.63 0.37 2.1
```

```
## Twenty         21  0.49                            0.67 0.33 2.6
## Three           4  0.40              0.35          0.57 0.43 2.6
## Thirteen       14  0.39  0.30  0.37                0.66 0.34 2.9
## Fifteen        16  0.35                            0.56 0.44 3.8
## Six             7        0.82                      0.67 0.33 1.2
## Twenty.three   24        0.69                      0.67 0.33 1.2
## Two             3        0.65                      0.55 0.45 1.1
## Eleven         12        0.60                -0.36 0.69 0.31 2.0
## Twenty.one     22  0.40  0.49                      0.69 0.31 2.5
## Five            6  0.36  0.42                      0.48 0.52 2.1
## Four            5  0.32  0.40                      0.50 0.50 2.5
## Fourteen       15  0.36  0.39                      0.57 0.43 2.3
## twenty.two     23  0.34  0.37  0.33                0.60 0.40 3.6
## Nine           10              0.88                0.72 0.28 1.1
## Twelve         13        0.43  0.50                0.69 0.31 2.1
## Eight           9                    0.58          0.63 0.37 2.2
## Seven           8        0.46        0.56          0.67 0.33 2.1
## X               1                           0.68 0.66 0.34 1.6
## Ten            11        0.36                -0.55 0.66 0.34 2.4
##
##                         TC1  TC5  TC3  TC4  TC2
## SS loadings           5.51 4.87 2.21 1.47 1.22
## Proportion Var        0.23 0.20 0.09 0.06 0.05
## Cumulative Var        0.23 0.43 0.52 0.59 0.64
## Proportion Explained  0.36 0.32 0.14 0.10 0.08
## Cumulative Proportion 0.36 0.68 0.82 0.92 1.00
##
##  With component correlations of
##       TC1   TC5  TC3  TC4   TC2
## TC1  1.00  0.60 0.35 0.25 -0.12
## TC5  0.60  1.00 0.31 0.27 -0.01
## TC3  0.35  0.31 1.00 0.18  0.00
## TC4  0.25  0.27 0.18 1.00  0.02
## TC2 -0.12 -0.01 0.00 0.02  1.00
##
## Mean item complexity =  2
## Test of the hypothesis that 5 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  526.17  with prob <  4.7e-39
##
## Fit based upon off diagonal values = 0.99
```

A principal components analysis (PCA) was conducted on the 23 items with orthog-onal rotation (varimax). The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis KMO = .93 ('superb' according to Kaiser, 1974), and all KMO values for individual items were > .77, which is well above the acceptablelimit of .5. Bartlett's test of sphericity,chisquare(253)=19,334, p < .001, indicated that correlations between items were sufficiently large for PCA. An initial analysis wasrun to obtain eigenvalues for each component in the data. Four components hadeigenvalues over Kaiser's criterion of 1 and in combination explained 61% of the variance. The scree plot was slightly ambiguous and showed inflexions that wouldjustify retaining both two and four components. Given the large sample size, and the convergence of the scree plot and Kaiser's criterion on four components, five components were retained in the final analysis. Table shows the factor loadings after rotation. The items that cluster on the same components suggest that component 1 represents a fear Quality of Care, Component 2 represents accessibility Component 3 represents environment , other domains are less clearly marked and there is a correlation and cross-talk between questions in domains.

Cronbach alpha

Let us calculate cronbach alpha for each subscale

```
Interpersonal = jhar %>% dplyr::select(one:Four)
Accessibilty = jhar %>% dplyr::select(Five:Nine)
physical_Environment = jhar %>% dplyr::select(Ten:Thirteen)
Availability = jhar %>% dplyr::select(Fourteen:Fifteen)
Quality = jhar %>% dplyr::select(Sixteen:Twenty.three)
```

Now let us run cronbach alpha test

```
keys = c(1, 1, 1, 1, 1, 1, 1)
summary(alpha(Interpersonal))$raw_alpha
```

```
##
## Reliability analysis
##  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
##       0.79      0.79    0.75      0.48 3.8 0.017  4.6 0.48
```

```
## [1] 0.7902073
```

```
summary(alpha(Accessibilty))$raw_alpha
```

```
##
## Reliability analysis
##  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
##       0.65      0.68    0.66       0.3 2.1 0.028  4.3 0.54
```

```
## [1] 0.6545357
```

```
summary(alpha(physical_Environment))$raw_alpha
```

```
##
## Reliability analysis
##  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
##       0.79      0.81    0.77      0.51 4.2 0.016  4.4 0.62
```

```
## [1] 0.7851515
```

```
summary(alpha(Availability))$raw_alpha
```

```
##
## Reliability analysis
##  raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd
##       0.65      0.66    0.49      0.49 1.9 0.034  4.4 0.65
```

```
## [1] 0.6454973
```

```
summary(alpha(Quality))$raw_alpha
```

```
##
## Reliability analysis
##  raw_alpha std.alpha G6(smc) average_r S/N    ase mean  sd
##       0.91      0.91     0.9      0.55 9.8 0.0069  4.5 0.5
```

```
## [1] 0.9077085
```

The cronbach alpha for Interpersonal,Accesibilty,physical Environment , Avilability and Quality sub- scales are 0.79,0.68,0.81,0.66,0.91 respectively. Thus except for accessibility and availability subscales which had lower than 0.7 recommended limit of cronbach alpha,other subscales had nice reliability and correlation implying the accessibilty and availability subscales need to be worded more precisely for better reliability

```
##                Domains test_retest_reliability cronbach_alpha
## 1        Interpersonal                    0.74           0.79
## 2         Accessibilty                    0.60           0.68
## 3 physical_Environment                    0.78           0.81
## 4         Availability                    0.58           0.66
## 5              Quality                    0.82           0.91
```

KEY POINTS

1. Overall Satisfaction levels in Questionnare is high
2. Quality and Interpersonal subscales had high effect on mean score.
3. Age, education,Income were positively correlated with satisfaction
4. Confirmatory factor analysis explained sixty percent of variance , however not all sub-scales were perfectly delineated, in particular accessibility and availability sub-scale question need to be worded well to improve reliability and internal consistency.