Study of Impact of Sleep Problems and Habits of People on Physical Fitness

1. Introduction

This markdown file uses a real data file condensed from a study performed to explore impact of sleep problems on various aspects of people's lives. The dataset used in this study was created based on a questionnaire answered by staff from a University in Melbourne, Australia. Physical fitness and well being of people depends primarily on their sleeping patterns and habits like smoking, drinking etc. It is therefore an important area of study to analyse and find predictive relationships between these factors to improve overall well being of individuals. This analysis is carried out to form and investigate a research question, in an attempt to examine a relationship between factors like Physical Fitness and Total Sleepiness and Associated Sensations. It will determine if these relationships could be considered as a basis for a prediction model in future. Furthermore, this analysis also examines differences in Physical Fitness of participants based on their age group, sleeping patterns and other lifestyle choices.

1.1 Research Question

A person's physical fitness can be affected by various physical factors as well as lifestyle and habits of an individual. This research is carried out to find out whether a person's total sleepiness and associated sensations scale directly affects a person's rate of physical fitness. Moreover, this analysis also examines whether people of different age groups have different rating for Physical Fitness levels. In addition to these factors, we also find out if a person's smoking habits can affect their physical fitness levels. Furthermore, can factors like level of stress experienced by a person be considered to make difference in physical fitness of a person? Physical fitness level may also be different for people who do not sleep quite well and for those who enjoy a good sleep more often. This research also analyses the relationship between quality of sleep experienced by an individual and level of physical fitness.

1.2 Hypotheses

The hypotheses investigated in this study are:

Hypothesis 1:

H0: There is no correlation between a respondent's Rate of Physical Fitness and their level of Total Sleepiness.

HA: There is a correlation between a respondent's Rate of Physical Fitness and their level of Total Sleepiness.

Hypothesis 2:

H0: There is no difference in Rate of Physical Fitness for respondents based on their smoking habits.

HA: There is a difference in Rate of Physical Fitness for respondents based on their smoking habits.

Hypothesis 3:

H0: There is no difference in Rate of Physical Fitness for respondents of different age groups.

HA: There is a difference in Rate of Physical Fitness for respondents of different age groups.

Hypothesis 4:

H0: There is no difference in Rate of Physical Fitness for respondents having different levels of stress.

HA: There is a difference in Rate of Physical Fitness for respondents having different levels of stress.

Hypothesis 5:

H0: There is no difference in Rate of Physical Fitness for respondents based on their quality of sleep.

HA: There is a difference in Rate of Physical Fitness for respondents based on their quality of sleep.

```
library(dplyr) # For data manipulation
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr) # For reading dataset file
library(ggplot2) # For creating histograms with more detail than plot
library(psych) # For descriptive functions
```

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

```
library(pastecs)# For descriptive functions
```

```
##
## Attaching package: 'pastecs'
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
library(semTools) # For skewness and kurtosis
```

```
## Loading required package: lavaan
## This is lavaan 0.6-3
## lavaan is BETA software! Please report any bugs.
##
## Attaching package: 'lavaan'
## The following object is masked from 'package:psych':
##
##     cor2cov
##
## ################################################################################
```

```
## This is semTools 0.5-1
```

```
## All users of R (or SEM) are invited to submit functions or ideas for functions.
```

```
## ################################################################################
```

```
##
## Attaching package: 'semTools'
```

```
## The following object is masked from 'package:psych':
##
##     skew
```

```
library(car) # For Levene's test for homogeneity of variance
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(pander) # For creating a table
library(lsr) # For calculating effect size using etasquare statistic
library(effsize) # For calculating effect size using cohen's d statistic
```

```
##
## Attaching package: 'effsize'
```

```
## The following object is masked from 'package:psych':
##
##     cohen.d
```

```
# Read the dataset.
sleep_dataset<-read_csv('sleep5ED.csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   weight = col_double(),
##   height = col_double(),
##   smokenum = col_double(),
##   alchohol = col_double(),
##   caffeine = col_double(),
##   hourwnit = col_double(),
##   hourwend = col_double(),
##   hourneed = col_double(),
##   stressmo = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

2. METHODOLOGY

2.1 Variables of Interest

Summary of Variables used in this research is shown below.

```
panderOptions('table.split.table', Inf)
variables<- "
Concept                                                    | Variable Name     | Statisti

Rate of Physical Fitness                                   | fitrate           | Scale
Smoking                                                    | smoke             | Nominal
Age Group                                                  | agegp3            | Nominal
Rate of Stress Over Last Month                             | stressmo          | Nominal
Quality of Sleep                                           | qualsleeprec      | Nominal
Total Sleepiness and Associated Sensations                 | totsas            | Scale
"

df <- read.delim(textConnection(variables),header=FALSE,sep="|",strip.white=TRUE,stringsAsFactors=FALSE)
names(df) <- unname(as.list(df[1,])) # put headers on
df <- df[-1,] # remove first row
row.names(df)<-NULL
pander(df, style = 'rmarkdown')
```

Concept	Variable Name	Statistical Type	Possible Values
Rate of Physical Fitness	fitrate	Scale	1 = very poor, 10 = very good
Smoking	smoke	Nominal	1 = yes, 2 = no
Age Group	agegp3	Nominal	1 = <= 37, 2 = 38 - 50 years, 3 = 51+ years
Rate of Stress Over Last Month	stressmo	Nominal	1 = not at all, 10 = extremely
Quality of Sleep	qualsleeprec	Nominal	1 = Very Poor, 2 = Fair, 3 = Good, 4 = Very Good
Total Sleepiness and Associated Sensations	totsas	Scale	5 = Low, 50 = Extreme Sleepiness

Each of the variables included in this research were inspected. The scale variables of interest, those representing total sleepiness and associated sensations and rate of physical fitness, were inspected for normality by creating histograms and QQ plots, inspecting standardised scores for skewness and kurtosis as well as considering the percentage of standardised scores for the variables that fell outside the expected boundaries. Decisions on skewness and kurtosis were guided by the advice of (George and Mallory, 2011) which categorizes a distribution as normal if the relevant standardised scores for skewness and kurtosis fall in the range +/- 2 and the advice of (Field, Miles and Field, 2012) which categorizes a distribution as normal if 95% of the standardised scores for the variable fall within the bounds of +/-3.29 for a dataset larger than 80 cases. For nominal variables, statistical summary was presented indicating their representativeness with respect to each group.

2.2 TOTAL SLEEPINESS AND ASSOCIATED SENSATIONS SCALE

Code for Inspecting the Variable

```
pastecs::stat.desc(sleep_dataset$totsas,basic = F)
```

```
##       median         mean      SE.mean CI.mean.0.95          var
##   27.0000000   26.0438247    0.6640035    1.3077539  110.6660717
##      std.dev     coef.var
##   10.5197943    0.4039266
```

```
# Skew & Kurtosis.
sleep_skew<-semTools::skew(sleep_dataset$totsas)
```

```
## Warning in semTools::skew(sleep_dataset$totsas): Missing observations are
## removed from a vector.
```

```
sleep_kurt<-semTools::kurtosis(sleep_dataset$totsas)
```

```
## Warning in semTools::kurtosis(sleep_dataset$totsas): Missing observations
## are removed from a vector.
```

```
# Get standardised value, divide by standard error.
stdskew<-sleep_skew[1]/sleep_skew[2]
stdkurt<-sleep_kurt[1]/sleep_kurt[2]
stdskew
```

```
##  skew (g1)
## -0.3667199
```

```
stdkurt
```

```
## Excess Kur (g2)
##       -3.654642
```

```
# Filter out rows with standardised value outside the range.
library(dplyr)
outliers<- sleep_dataset %>%
  filter(scale(totsas) >3.29 | scale(totsas) < -3.29)
outliers
```

```
## # A tibble: 0 x 55
## # ... with 55 variables: id <int>, sex <int>, age <int>, marital <int>,
## #   edlevel <int>, weight <dbl>, height <dbl>, healthrate <int>,
## #   fitrate <int>, weightrate <int>, smoke <int>, smokenum <dbl>,
## #   alchohol <dbl>, caffeine <dbl>, hourwnit <dbl>, hourwend <dbl>,
## #   hourneed <dbl>, trubslep <int>, trubstay <int>, wakenite <int>,
## #   niteshft <int>, liteslp <int>, refreshd <int>, satsleep <int>,
## #   qualslp <int>, stressmo <dbl>, medhelp <int>, problem <int>,
## #   impact1 <int>, impact2 <int>, impact3 <int>, impact4 <int>,
## #   impact5 <int>, impact6 <int>, impact7 <int>, stopb <int>,
## #   restlss <int>, drvsleep <int>, drvresul <int>, ess <int>,
## #   anxiety <int>, depress <int>, fatigue <int>, lethargy <int>,
## #   tired <int>, sleepy <int>, energy <int>, stayslprec <int>,
## #   getsleprec <int>, qualsleeprec <int>, totsas <int>, cigsgp3 <int>,
## #   agegp3 <int>, probsleeprec <int>, drvslprec <int>
```

```r
# Count outliers.
numoutliers<-outliers %>%
  summarize(count=n())
fullcount<-sleep_dataset %>%
  summarize(count=n())
numoutliers
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1     0
```

```r
fullcount
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1   271
```

```r
sort(scale(sleep_dataset$totsas))
```

```
##   [1] -2.000402683 -1.905343790 -1.810284897 -1.715226003 -1.715226003
##   [6] -1.715226003 -1.715226003 -1.715226003 -1.525108217
##  [11] -1.525108217 -1.525108217 -1.525108217 -1.525108217
##  [16] -1.525108217 -1.430049324 -1.430049324 -1.430049324 -1.430049324
##  [21] -1.430049324 -1.430049324 -1.430049324 -1.334990431 -1.334990431
##  [26] -1.334990431 -1.334990431 -1.334990431 -1.334990431 -1.334990431
##  [31] -1.239931538 -1.239931538 -1.239931538 -1.239931538 -1.239931538
##  [36] -1.239931538 -1.239931538 -1.239931538 -1.144872645
##  [41] -1.144872645 -1.144872645 -1.144872645 -1.144872645 -1.144872645
##  [46] -1.144872645 -1.144872645 -1.144872645 -1.049813752 -1.049813752
##  [51] -1.049813752 -1.049813752 -1.049813752 -1.049813752 -1.049813752
##  [56] -0.954754859 -0.954754859 -0.954754859 -0.954754859 -0.954754859
##  [61] -0.954754859 -0.954754859 -0.859695965 -0.859695965 -0.859695965
##  [66] -0.859695965 -0.859695965 -0.859695965 -0.859695965 -0.859695965
##  [71] -0.764637072 -0.764637072 -0.764637072 -0.764637072 -0.764637072
##  [76] -0.764637072 -0.764637072 -0.764637072 -0.669578179 -0.669578179
##  [81] -0.669578179 -0.574519286 -0.574519286 -0.574519286 -0.574519286
##  [86] -0.574519286 -0.574519286 -0.574519286 -0.574519286 -0.574519286
##  [91] -0.574519286 -0.479460393 -0.479460393 -0.479460393 -0.479460393
##  [96] -0.384401500 -0.289342607 -0.289342607 -0.289342607 -0.289342607
## [101] -0.289342607 -0.289342607 -0.289342607 -0.289342607 -0.289342607
## [106] -0.194283714 -0.194283714 -0.194283714 -0.194283714 -0.194283714
## [111] -0.194283714 -0.194283714 -0.099224821 -0.099224821 -0.099224821
## [116] -0.099224821 -0.099224821 -0.099224821 -0.099224821 -0.099224821
## [121] -0.004165928 -0.004165928 -0.004165928 -0.004165928  0.090892966
## [126]  0.090892966  0.090892966  0.090892966  0.090892966  0.185951859
## [131]  0.185951859  0.185951859  0.185951859  0.185951859  0.185951859
## [136]  0.185951859  0.281010752  0.281010752  0.281010752  0.281010752
## [141]  0.281010752  0.376069645  0.376069645  0.376069645  0.376069645
## [146]  0.376069645  0.376069645  0.376069645  0.376069645  0.471128538
## [151]  0.471128538  0.471128538  0.471128538  0.471128538  0.471128538
## [156]  0.471128538  0.471128538  0.471128538  0.471128538  0.566187431
## [161]  0.566187431  0.566187431  0.566187431  0.661246324  0.661246324
## [166]  0.661246324  0.661246324  0.661246324  0.661246324  0.661246324
## [171]  0.661246324  0.756305217  0.756305217  0.756305217  0.756305217
```

```
## [176]  0.756305217  0.756305217  0.756305217  0.756305217  0.756305217
## [181]  0.756305217  0.756305217  0.851364110  0.851364110  0.851364110
## [186]  0.851364110  0.851364110  0.851364110  0.851364110  0.851364110
## [191]  0.851364110  0.851364110  0.851364110  0.851364110  0.851364110
## [196]  0.851364110  0.851364110  0.851364110  0.946423003  0.946423003
## [201]  0.946423003  0.946423003  0.946423003  0.946423003  0.946423003
## [206]  0.946423003  1.041481897  1.041481897  1.041481897  1.041481897
## [211]  1.041481897  1.041481897  1.041481897  1.041481897  1.041481897
## [216]  1.136540790  1.136540790  1.136540790  1.231599683  1.231599683
## [221]  1.231599683  1.231599683  1.231599683  1.231599683  1.326658576
## [226]  1.326658576  1.326658576  1.326658576  1.326658576  1.326658576
## [231]  1.326658576  1.326658576  1.326658576  1.326658576  1.326658576
## [236]  1.326658576  1.326658576  1.421717469  1.421717469  1.516776362
## [241]  1.611835255  1.611835255  1.706894148  1.706894148  1.706894148
## [246]  1.706894148  1.706894148  1.801953041  1.801953041  1.992070828
## [251]  2.277247507
```

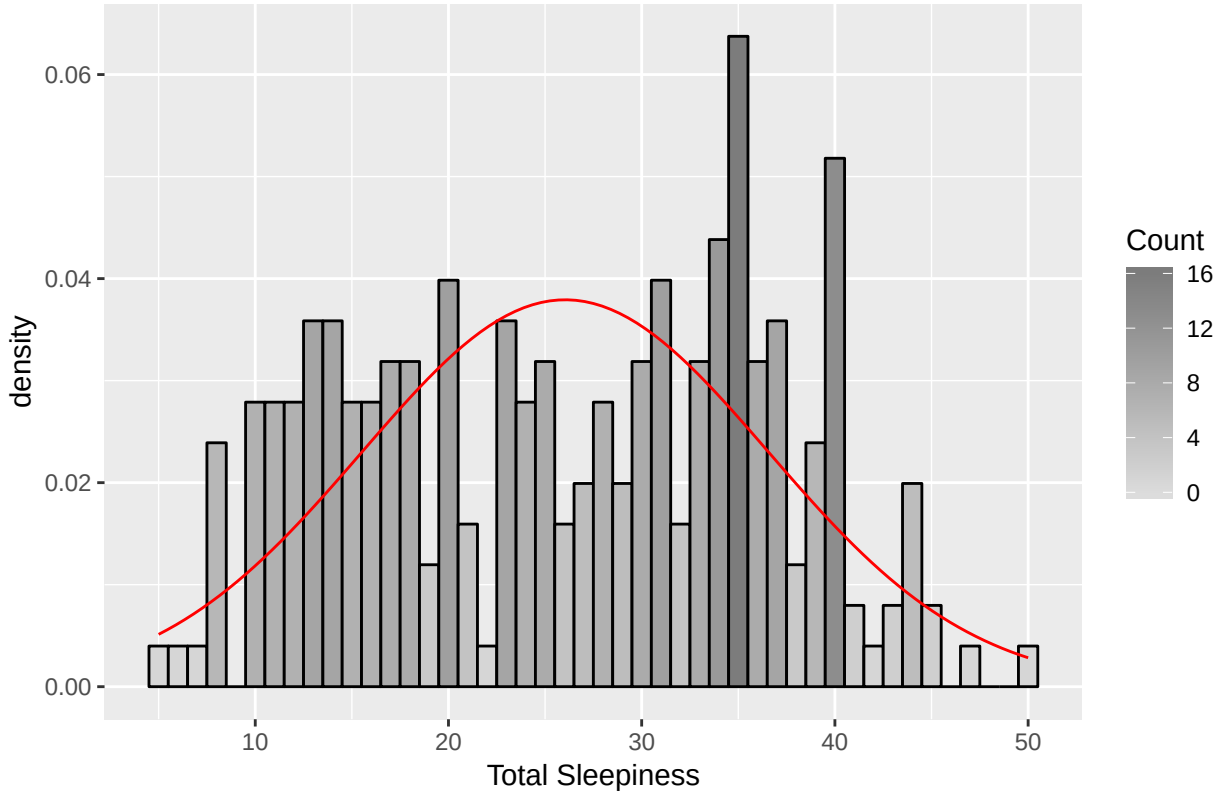
Report of Normality Analysis

Total sleepiness and associated sensations value is represented by a scale variable in the dataset which was calculated as a sum of respondent's rate of fatigue, rate of lethargy, rate of tiredness, rate of sleepiness and rate of lacking in energy. These factors were recorded using a questionnaire. Inspection of the histogram and normality plot (see Figure 1 and Figure 2) shows that the distribution does not conform exactly to a normal distribution. Inspection of standardized normal scores for skewness and kurtosis indicated that while skewness is not outside the acceptable range of +/-2, kurtosis value falls outside the acceptable range of +/-2 (Standardised Skewness: -.36, Standardised Kurtosis: -3.65). Further inspection of the variable using standardized scores showed that 0% standardised scores were outside the acceptable range of +/- 3.29. Total sleepiness and associated sensations can therefore be safely treated as a normal for this research.

```
# Create Histogram.
library(ggplot2)
gg <- ggplot(sleep_dataset, aes(x=sleep_dataset$totsas))
gg <- gg + labs(x="Total Sleepiness")
# Set Binwidth
gg <- gg + geom_histogram(binwidth=1, colour="black", aes(y=..density.., fill=..count..))
gg <- gg + scale_fill_gradient("Count", low="#DCDCDC", high="#7C7C7C")

# stat_function uses mean and standard deviation to compute normalised score for each value.
gg <- gg + stat_function(fun=dnorm, color="red",args=list(mean=mean(sleep_dataset$totsas, na.rm=TRUE), s
gg<-gg+ggtitle('Figure 1: Distribution of Total Sleepiness')
# Display Graph.
gg
```

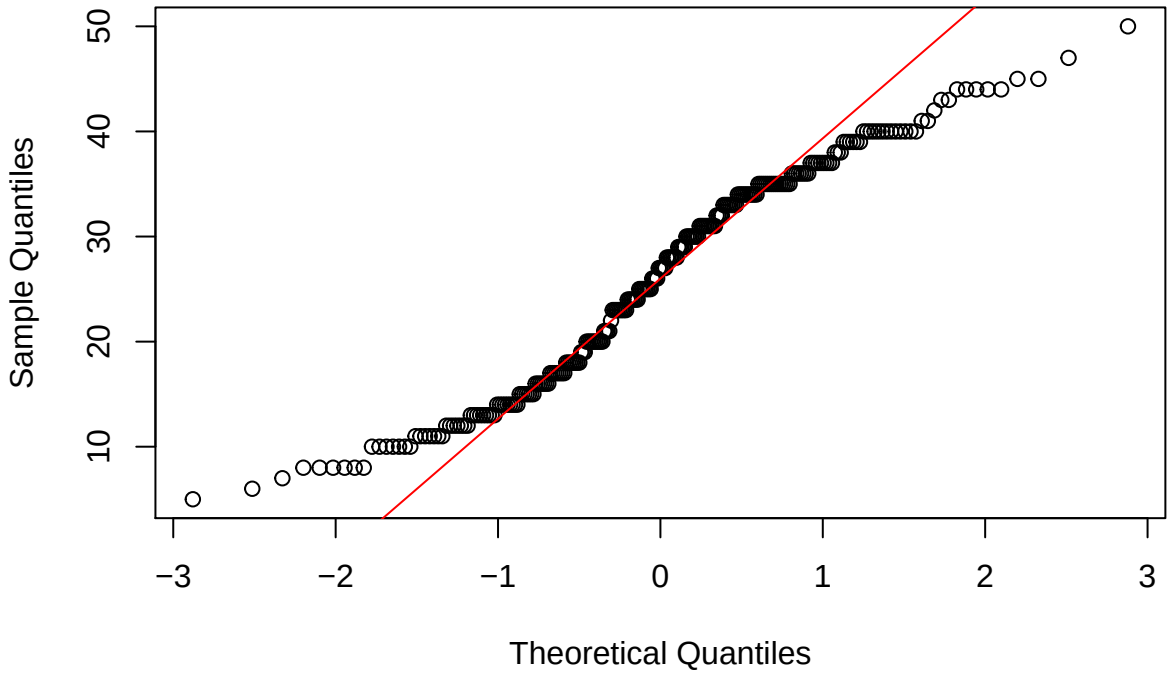
```
## Warning: Removed 20 rows containing non-finite values (stat_bin).
```

Figure 1: Distribution of Total Sleepiness




```
# qqplot
qqnorm(sleep_dataset$totsas, main="Figure 2: QQPlot of Total Sleepiness")
#show a line on theplot
qqline(sleep_dataset$totsas, col=2)
```

Figure 2: QQPlot of Total Sleepiness



2.3 RATE OF PHYSICAL FITNESS

Code for Inspecting the Variable

```
pastecs::stat.desc(sleep_dataset$fitrate,basic = F)
```

```
##       median         mean      SE.mean CI.mean.0.95          var
##    7.0000000    6.4210526    0.1052490    0.2072306    2.9465740
##      std.dev     coef.var
##    1.7165588    0.2673329
```

```
# Skew & Kurtosis.
fitrate_skew<-semTools::skew(sleep_dataset$fitrate)
```

```
## Warning in semTools::skew(sleep_dataset$fitrate): Missing observations are
## removed from a vector.
```

```
fitrate_kurt<-semTools::kurtosis(sleep_dataset$fitrate)
```

```
## Warning in semTools::kurtosis(sleep_dataset$fitrate): Missing observations
## are removed from a vector.
```

```
# Get standardised value, divide by standard error.
stdskew_fitrate<-fitrate_skew[1]/fitrate_skew[2]
stdkurt_fitrate<-fitrate_kurt[1]/fitrate_kurt[2]
stdskew_fitrate
```

```
## skew (g1)
## -3.064094
```

```
stdkurt_fitrate
```

```
## Excess Kur (g2)
##       0.3065625
```

```
# Filter out rows with standardised value outside the range.
library(dplyr)
outliers<- sleep_dataset %>%
  filter(scale(fitrate) >3.29 | scale(fitrate) < -3.29)
outliers
```

```
## # A tibble: 0 x 55
## # ... with 55 variables: id <int>, sex <int>, age <int>, marital <int>,
## #   edlevel <int>, weight <dbl>, height <dbl>, healthrate <int>,
## #   fitrate <int>, weightrate <int>, smoke <int>, smokenum <dbl>,
## #   alchohol <dbl>, caffeine <dbl>, hourwnit <dbl>, hourwend <dbl>,
## #   hourneed <dbl>, trubslep <int>, trubstay <int>, wakenite <int>,
## #   niteshft <int>, liteslp <int>, refreshd <int>, satsleep <int>,
## #   qualslp <int>, stressmo <dbl>, medhelp <int>, problem <int>,
## #   impact1 <int>, impact2 <int>, impact3 <int>, impact4 <int>,
## #   impact5 <int>, impact6 <int>, impact7 <int>, stopb <int>,
## #   restlss <int>, drvsleep <int>, drvresul <int>, ess <int>,
## #   anxiety <int>, depress <int>, fatigue <int>, lethargy <int>,
## #   tired <int>, sleepy <int>, energy <int>, stayslprec <int>,
## #   getsleprec <int>, qualsleeprec <int>, totsas <int>, cigsgp3 <int>,
## #   agegp3 <int>, probsleeprec <int>, drvslprec <int>
```

```
# Count outliers.
numoutliers<-outliers %>%
  summarize(count=n())
fullcount<-sleep_dataset %>%
  summarize(count=n())
numoutliers
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1     0
```

```
fullcount
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1   271
```

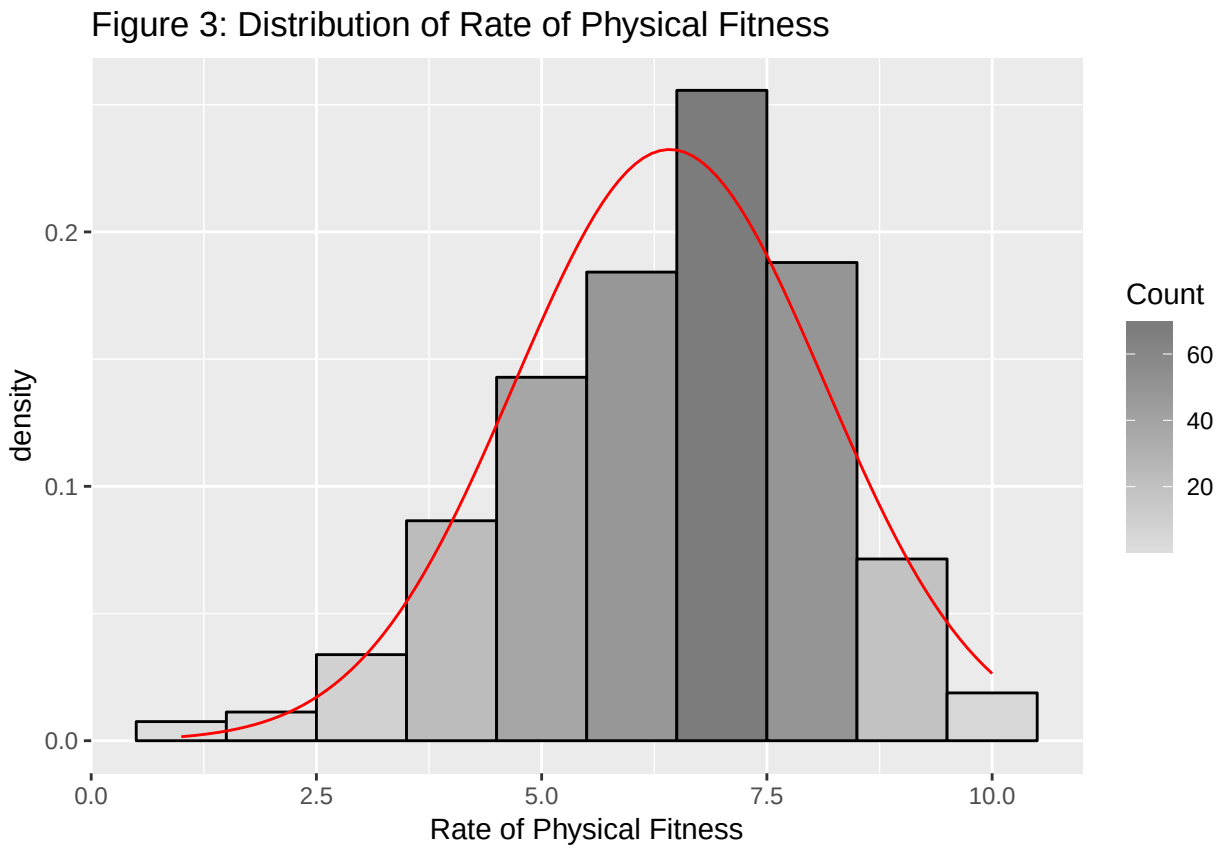
Report of Normality Analysis

Rate of physical fitness is represented by a scale variable in the dataset which is represented by a scale of 1 to 10 where 1 represents very poor physical fitness and 10 represents very good physical fitness. These values were recorded using a questionnaire. Inspection of the histogram and normality plot (see Figure 3 and Figure 4) shows that the distribution does not conform exactly to a normal distribution. Inspection of standardized normal scores for skewness and kurtosis indicated that while kurtosis value falls within the acceptable range of +/-2, skewness value falls outside the acceptable range of +/-2 (Standardised Skewnwss: -3.06, Standardised Kurtosis: .30). Further inspection of the variable using standardized scores showed that

0% standardised scores were outside the acceptable range of +/- 3.29. Rate of physical fitness can therefore be safely treated as a normal within this analysis.

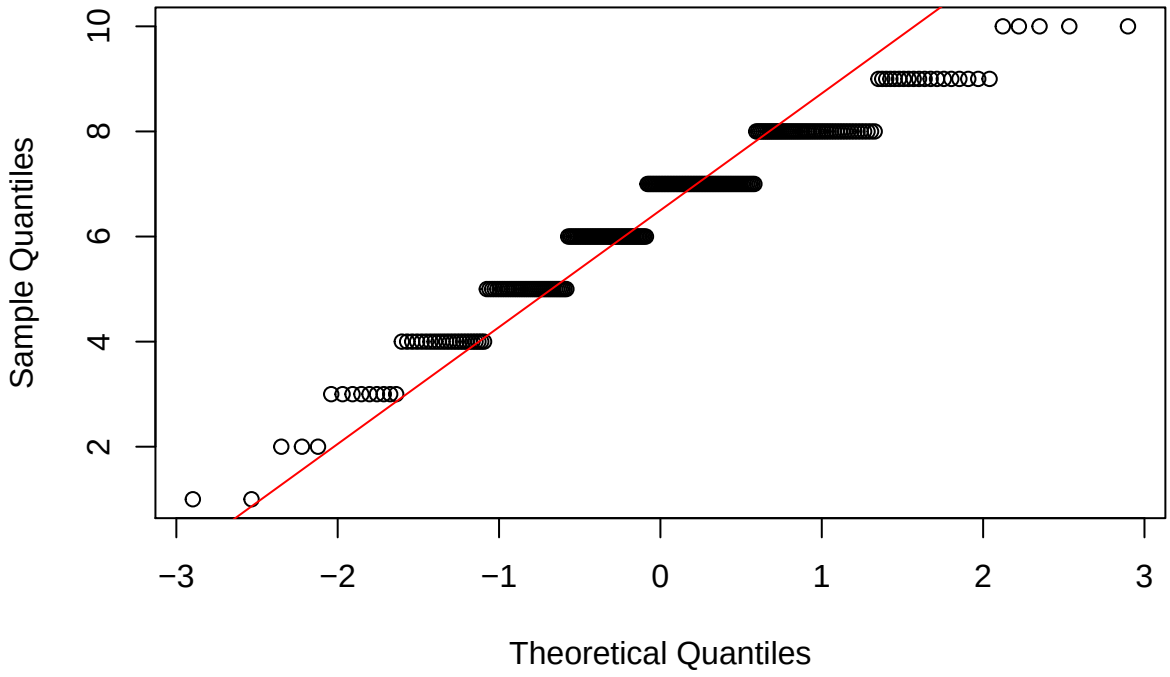
```
# Create Histogram.
library(ggplot2)
gg <- ggplot(sleep_dataset, aes(x=sleep_dataset$fitrate))
gg <- gg + labs(x="Rate of Physical Fitness")
# Set Binwidth
gg <- gg + geom_histogram(binwidth=1, colour="black", aes(y=..density.., fill=..count..))
gg <- gg + scale_fill_gradient("Count", low="#DCDCDC", high="#7C7C7C")

# stat_function uses mean and standard deviation to compute normalised score for each value.
gg <- gg + stat_function(fun=dnorm, color="red",args=list(mean=mean(sleep_dataset$fitrate, na.rm=TRUE),
gg<-gg+ggtitle('Figure 3: Distribution of Rate of Physical Fitness')
# Display Graph.
gg
```

```
## Warning: Removed 5 rows containing non-finite values (stat_bin).
```




```
# qqplot
qqnorm(sleep_dataset$fitrate, main="Figure 4: QQPlot of Rate of Physical Fitness")
#show a line on theplot
qqline(sleep_dataset$fitrate, col=2)
```

Figure 4: QQPlot of Rate of Physical Fitness



2.4 SMOKING

Smoking habit is represented by a categorical variable in the dataset. The dataset contains data from 34 respondents who have smoking habit and 236 respndents who do not smoke. Group 1 represents respondents who have smoking habit, while participants who do not smoke are in Group 2. The distribution of mean rate of physical fitness for each group is shown below using a bar chart (see Figure 5).

```
table(sleep_dataset$smoke)
```

```
##
##   1   2
##  34 236
```

```
# Get statistical description by group.
psych::describeBy(sleep_dataset$fitrate, sleep_dataset$smoke)
```

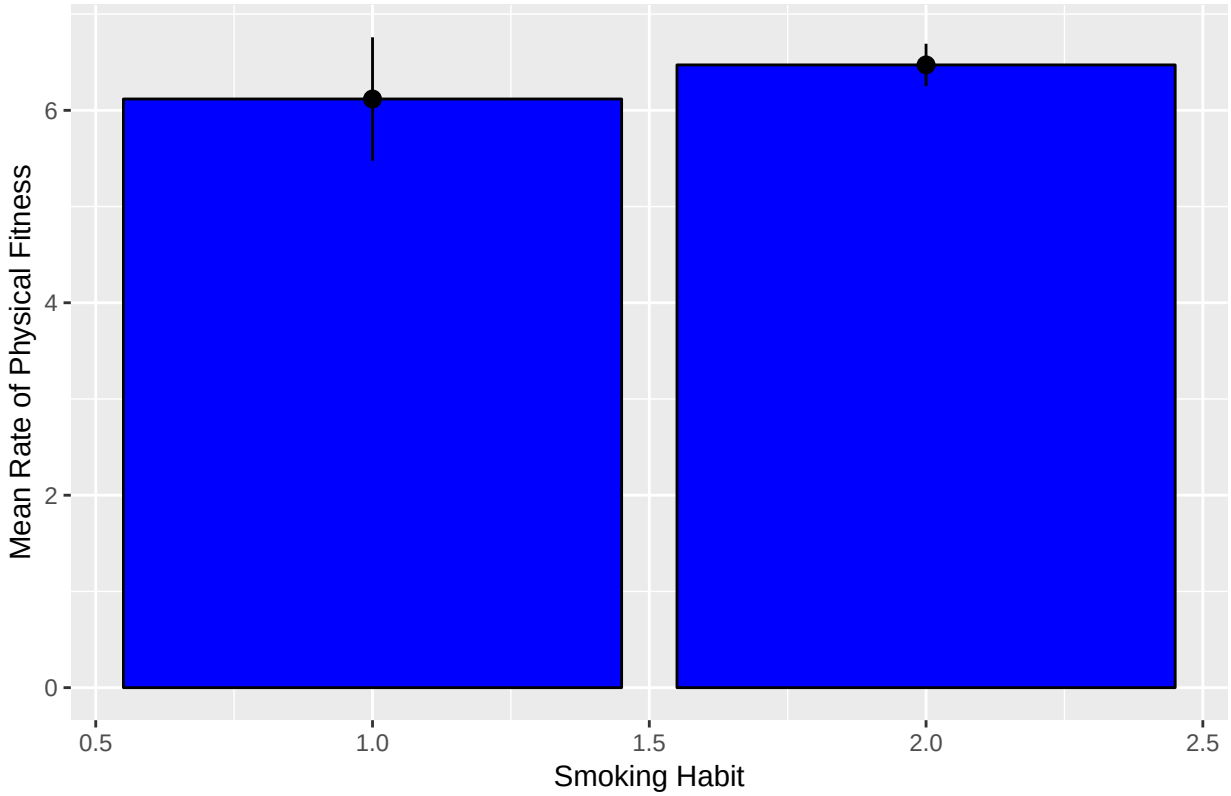
```
##
##  Descriptive statistics by group
## group: 1
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 34 6.12 1.84      7    6.18 1.48   2   9     7 -0.34    -1.04 0.32
## ------------------------------------------------------------
## group: 2
##    vars   n mean  sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 231 6.47 1.7      7    6.54 1.48   1  10     9 -0.47     0.23 0.11
```

```
bar <- ggplot(sleep_dataset, aes(smoke, fitrate))
bar + stat_summary(fun.y = mean, geom = "bar", fill = "Blue", colour = "Black") + stat_summary(fun.data
```

```
## Warning: Removed 6 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 6 rows containing non-finite values (stat_summary).
```

Figure 5: Mean Rate of Physical Fitness by Smoking Habit



2.5 AGE GROUP

Age group is represented by a categorical variable in the dataset. The respondents are divided into 3 groups. Group 1 has people from age 37 and less. Group 2 has people with age between 38 to 50, whereas people whose age 51 years or older are in Group 3. Grpup 1 has 83 respondents, Group 2 has 86 resondents while 79 people belong to Group 3. The distribution of mean rate of physical fitness for each age group is shown below using a bar chart (see Figure 6).

```
table(sleep_dataset$agegp3)
```

```
##
##  1  2  3
## 83 86 79
```

```
# Get statistical description by group.
psych::describeBy(sleep_dataset$fitrate, sleep_dataset$agegp3)
```

```
##
##  Descriptive statistics by group
## group: 1
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 82 6.62 1.53      7     6.7 1.48   2  10     8 -0.46     0.05 0.17
## ------------------------------------------------------------
## group: 2
```

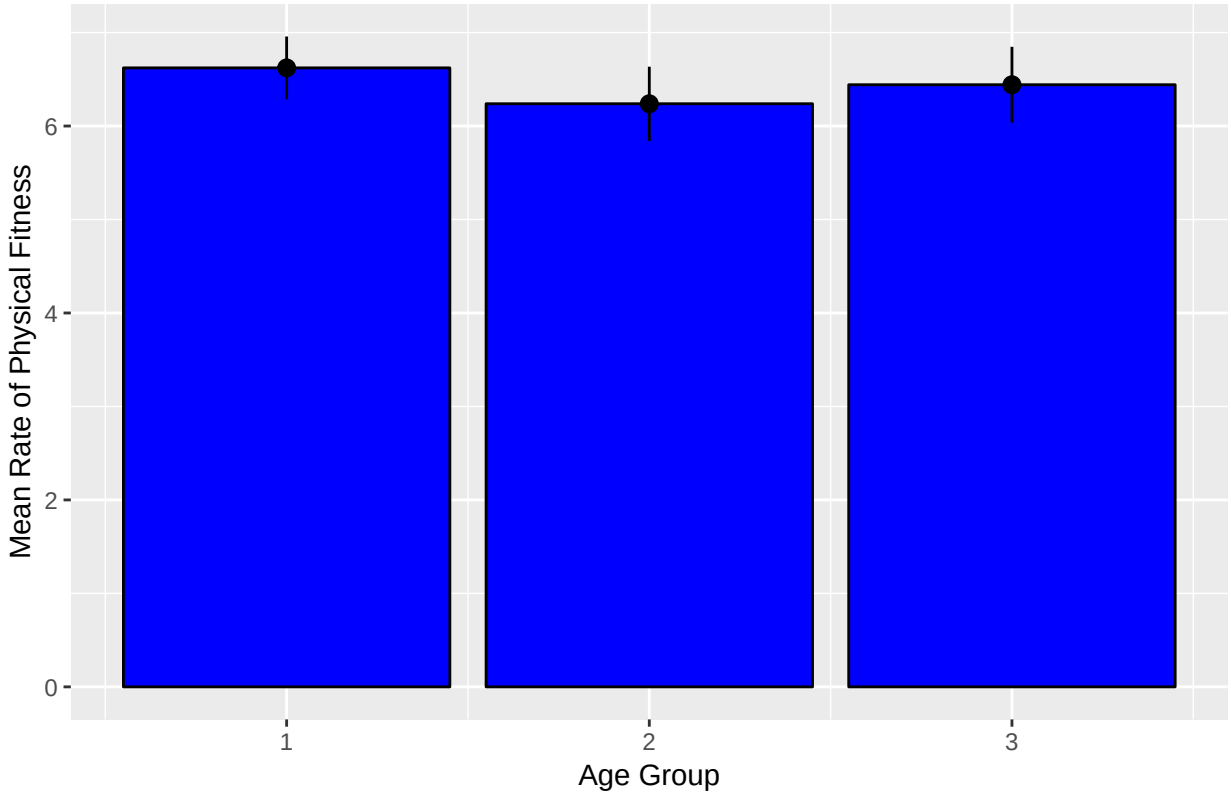
```
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 84 6.24 1.83      6    6.28 1.48   1  10     9 -0.41     0.17 0.2
## ------------------------------------------------------------
## group: 3
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 77 6.44 1.79      7    6.51 1.48   2  10     8 -0.38    -0.48 0.2
```

```
bar <- ggplot(sleep_dataset, aes(agegp3, fitrate))
bar + stat_summary(fun.y = mean, geom = "bar", fill = "Blue", colour = "Black") + stat_summary(fun.data
```

```
## Warning: Removed 28 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 28 rows containing non-finite values (stat_summary).
```

Figure 6: Mean Rate of Physical Fitness by Age Group



2.6 RATING OF STRESS OVER LAST MONTH

Stress level over last month is considered as a categorical variable for this research. The respondents have different levels of stress ranging from 1 to 10, with 1 representing no stress at all and 10 representing a very high stress. The distribution of mean rate of physical fitness for each group is shown below using a bar chart (see Figure 7).

```
table(sleep_dataset$stressmo)
```

```
##
##   1   2   3   4   5   6   7 7.5   8   9  10
##  13  26  26  23  30  25  48   1  52  21   6
```

```r
# Get statistical description by group.
psych::describeBy(sleep_dataset$fitrate, sleep_dataset$stressmo)
```

```
## Warning in min(x, na.rm = na.rm): no non-missing arguments to min;
## returning Inf
```

```
## Warning in max(x, na.rm = na.rm): no non-missing arguments to max;
## returning -Inf
```

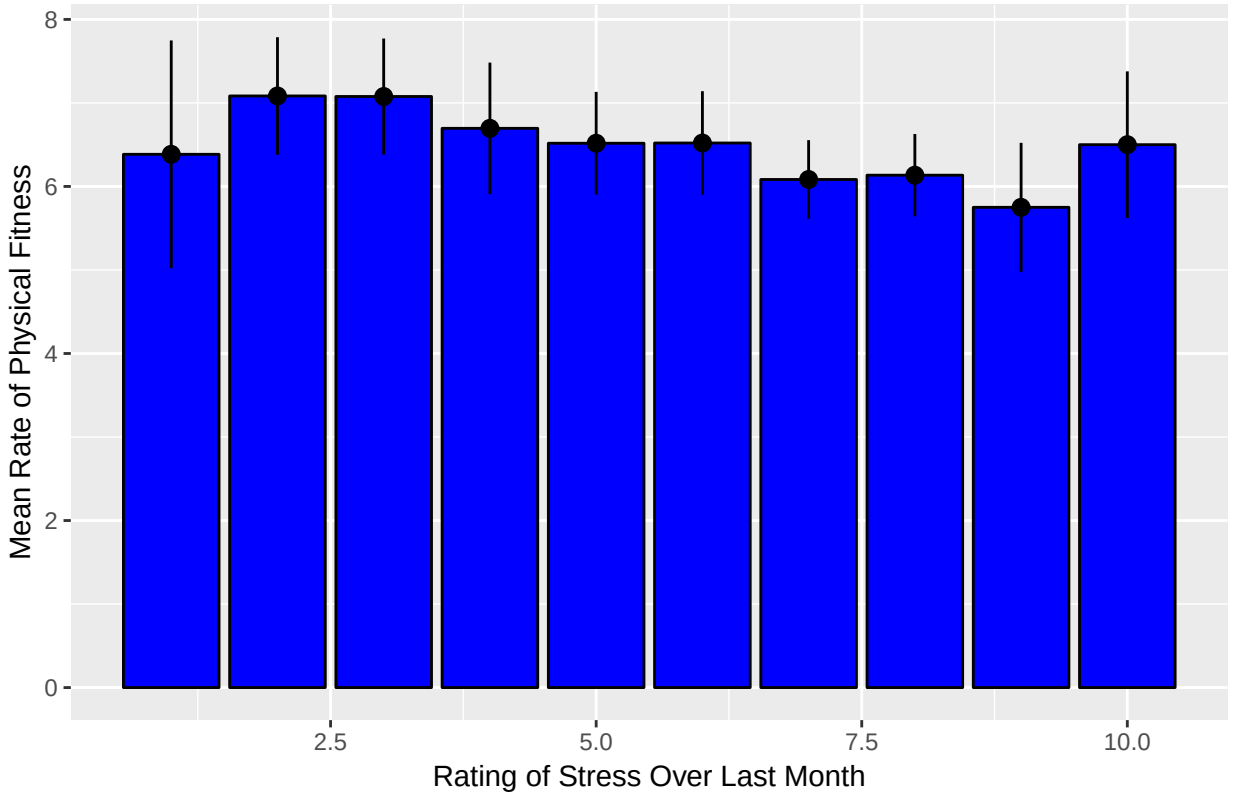
```
##
##  Descriptive statistics by group
## group: 1
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 13 6.38 2.26      6    6.45 2.97   2  10     8 -0.24    -0.99 0.63
## ------------------------------------------------------------
## group: 2
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 24 7.08 1.67    7.5    7.15 1.48   3  10     7 -0.5    -0.43 0.34
## ------------------------------------------------------------
## group: 3
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 26 7.08 1.72      7    7.18 1.48   3  10     7 -0.57    -0.47 0.34
## ------------------------------------------------------------
## group: 4
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 23  6.7 1.82      7    6.84 1.48   3   9     6 -0.69    -0.66 0.38
## ------------------------------------------------------------
## group: 5
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 29 6.52 1.62      7    6.64 1.48   1   9     8 -1.18     2.49 0.3
## ------------------------------------------------------------
## group: 6
##    vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## X1    1 25 6.52 1.5      7    6.52 1.48   4   9     5 0.03    -1.11 0.3
## ------------------------------------------------------------
## group: 7
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 48 6.08 1.62      6    6.17 1.48   1  10     9 -0.63     0.85 0.23
## ------------------------------------------------------------
## group: 7.5
##    vars n mean sd median trimmed mad min  max range skew kurtosis se
## X1    1 0  NaN NA     NA     NaN  NA Inf -Inf  -Inf   NA       NA NA
## ------------------------------------------------------------
## group: 8
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 52 6.13 1.77      6    6.19 1.48   2  10     8 -0.24    -0.63 0.25
## ------------------------------------------------------------
## group: 9
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 20 5.75 1.65      6    5.75 1.48   2   9     7 -0.09    -0.32 0.37
## ------------------------------------------------------------
## group: 10
##    vars n mean   sd median trimmed mad min max range  skew kurtosis   se
## X1    1 6  6.5 0.84      7     6.5   0   5   7     2 -0.85    -1.17 0.34
```

```
bar <- ggplot(sleep_dataset, aes(stressmo, fitrate))
bar + stat_summary(fun.y = mean, geom = "bar", fill = "Blue", colour = "Black") + stat_summary(fun.data
```

```
## Warning: Removed 5 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 5 rows containing non-finite values (stat_summary).
```

Figure 7: Mean Rate of Physical Fitness by Rating of Stress Over Last Month



2.7 QUALITY OF SLEEP

Quality of Sleep is represented by a categorical variable in the dataset. The respondents are divided into 4 groups according to quality of sleep experienced by them. Respondents in Group 1 have very poor quality of sleep. Group 2 has repondents who experience a fair level of sleep. Group 3 contains data from participants who seem to have a good sleep every night. In Group 4, the data is from respondents who enjoy a very good level of sleep. The distribution of mean rate of physical fitness for each age group is shown below using a bar chart (see Figure 8).

```
table(sleep_dataset$qualsleeprec)
```

```
##
##  1  2  3  4
## 36 75 90 67
```

```
# Get statistical description by group.
psych::describeBy(sleep_dataset$fitrate, sleep_dataset$qualsleeprec)
```

```
##
##  Descriptive statistics by group
```

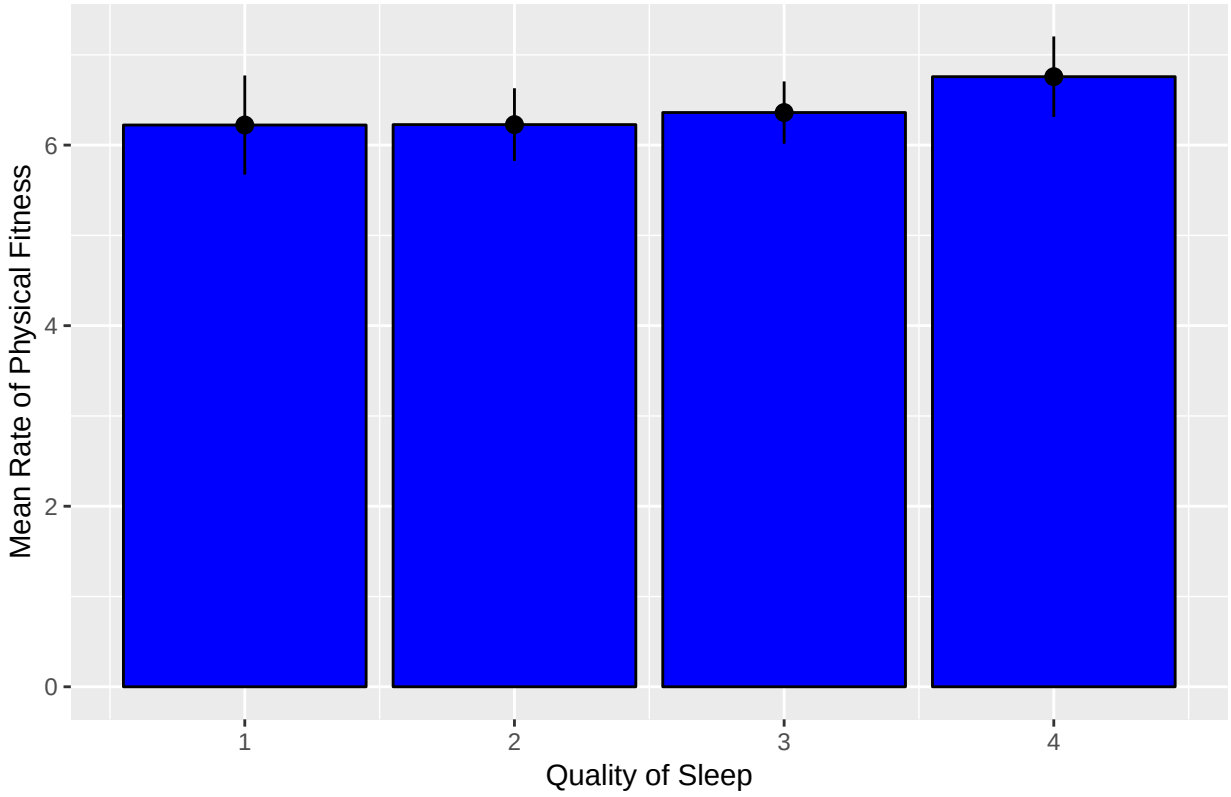
```
## group: 1
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 36 6.22 1.62      6    6.17 1.48   3  10     7 0.23    -0.61 0.27
## ------------------------------------------------------------
## group: 2
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis  se
## X1    1 75 6.23 1.75      7    6.34 1.48   1  10     9 -0.73     0.73 0.2
## ------------------------------------------------------------
## group: 3
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 86 6.36 1.61      7    6.46 1.48   2   9     7 -0.54    -0.35 0.17
## ------------------------------------------------------------
## group: 4
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 66 6.76 1.81      7    6.87 1.48   2  10     8 -0.51    -0.34 0.22
```

```
bar <- ggplot(sleep_dataset, aes(qualsleeprec, fitrate))
bar + stat_summary(fun.y = mean, geom = "bar", fill = "Blue", colour = "Black") + stat_summary(fun.data
```

```
## Warning: Removed 8 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 8 rows containing non-finite values (stat_summary).
```

Figure 8: Mean Rate of Physical Fitness by Quality of Sleep



3. RESULTS

An alpha level 0.05 was adopted and Cohen's conventions on effect size were adopted for all statistical tests.

3.1 Hypothesis 1

H0: There is no correlation between a respondent's Rate of Physical Fitness and their level of Total Sleepiness.

HA: There is a correlation between a respondent's Rate of Physical Fitness and their level of Total Sleepiness.

```
#Pearson Correlation Test.
stats::cor.test(sleep_dataset$fitrate, sleep_dataset$totsas,  method='pearson')
```

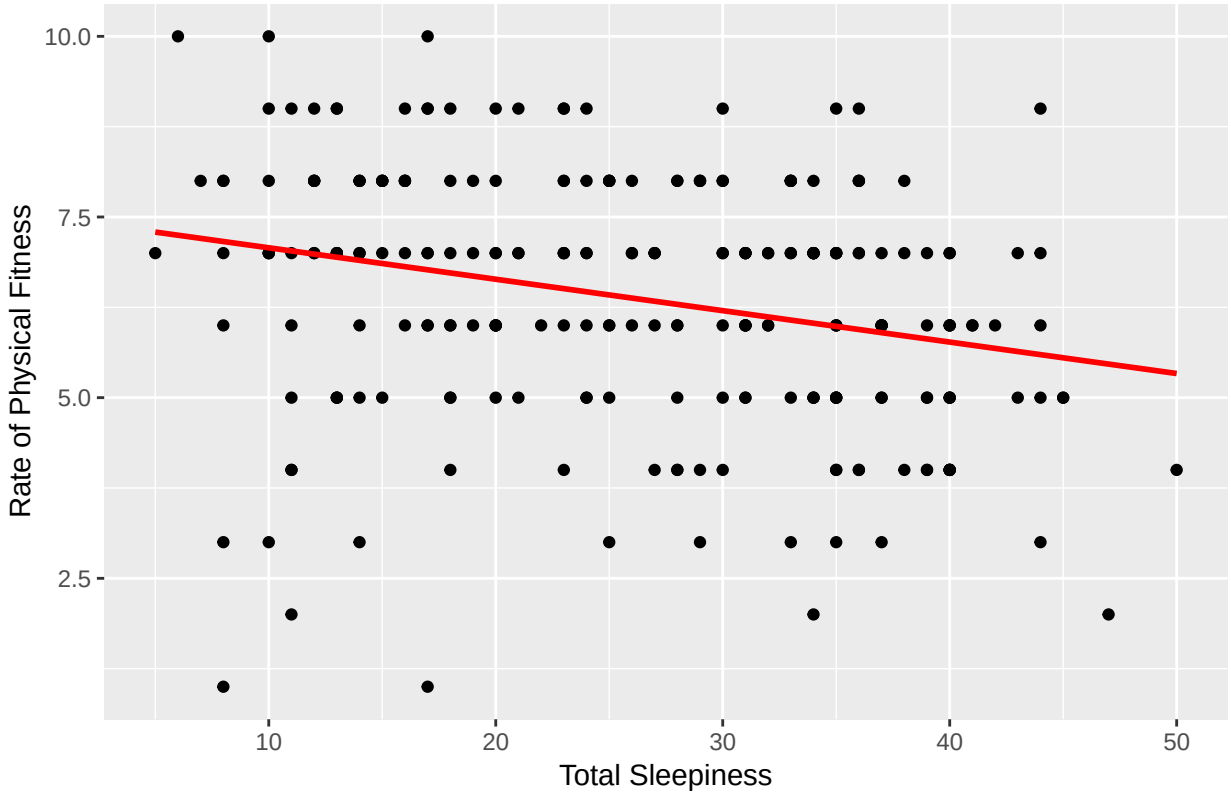
```
##
##  Pearson's product-moment correlation
##
## data:  sleep_dataset$fitrate and sleep_dataset$totsas
## t = -4.3308, df = 245, p-value = 2.168e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3788733 -0.1467290
## sample estimates:
##        cor
## -0.2666646
```

```
#Scatter Plot.
library(ggplot2)
scatter<- ggplot(sleep_dataset,aes(sleep_dataset$totsas,sleep_dataset$fitrate))
scatter<-scatter + geom_point() + geom_smooth(method = "lm", colour = "red", se=F)
scatter<-scatter+labs(x = 'Total Sleepiness', y = 'Rate of Physical Fitness', title = "Figure 9: Scatte
scatter
```

```
## Warning: Removed 24 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 24 rows containing missing values (geom_point).
```

Figure 9: Scatter Plot of Physical Fitness and Total Sleepiness



The relationship between Total sleepiness and associated sensations (calculated by combining factors like rate of lethargy, rate of tiredness, rate of sleepiness, rate of lacking energy, rate of fatigue) and Rate of physical fitness (derived from the sleep problems questionnaire) was investigated using a Pearson correlation. A very weak negative correlation was found (r = -.266, $p < .001$). Scatter plot of physical fitness and total sleepiness also indicated a weak correlation between these factors (see Figure 9). This evidence can be considered to reject the null hypothesis that there is no correlation between a respondent's Rate of Physical Fitness and their level of Total Sleepiness. Coefficient of determination (r^2 = 0.0707) indicates that 7.07% of variations between Rate of Physical Fitness and Total Sleepiness and Associated Sensations scale are common.

3.2 Hypothesis 2

H0: There is no difference in Rate of Physical Fitness for respondents based on their smoking habits.

HA: There is a difference in Rate of Physical Fitness for respondents based on their smoking habits.

Independent t-test

```
# Convert smoke to categorical variable, as Levene's test is not suitable for quantitative variables.
sleep_dataset$smoke<-as.factor(sleep_dataset$smoke)
#Conduct Levene's test for homogeneity of variance in library car
ltest<-car::leveneTest(fitrate ~ smoke, data=sleep_dataset)
ltest
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   1  0.8573 0.3553
##       263
```

```r
# Two sample t-test
stats::t.test(fitrate~smoke,var.equal=TRUE,data=sleep_dataset)
```

```
##
##  Two Sample t-test
##
## data:  fitrate by smoke
## t = -1.1233, df = 263, p-value = 0.2623
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9751245  0.2666957
## sample estimates:
## mean in group 1 mean in group 2
##        6.117647        6.471861
```

```r
# Cohen's d to calculate effect size
cohen.d(sleep_dataset$fitrate, sleep_dataset$smoke,alpha = 0.05,std = TRUE,na.rm = TRUE)
```

```
##
## Cohen's d
##
## d estimate: -0.2063318 (small)
## 95 percent confidence interval:
##        inf        sup
## -0.5684456  0.1557820
```

A Levene's test for homogeneity of variance was conducted and indicated equality of variance for Rate of Physical Fitness scores for respondents who had smoking habit and for those who did not smoke (F = .86, P = .36). An independent-samples t-test was conducted to compare physical fitness scores for respondents having different smoking habits. No significant difference in the scores for Physical Fitness was found (M = 6.12, SD= 1.84 for respondents who smoke, M = 6.47, SD= 1.7 for respondents who do not smoke), ($t(263)$= -1.12, p = .26). The Cohen's d statistic also indicated a small effect size (-.206). The result is not statistically significant which indicates there is no evidence to reject the null hypothesis that there is no difference in physical fitness scores for respondents of having different smoking habits. The very small effect size is a concern and therefore the result should be treated with caution.

3.3 Hypothesis 3

H0: There is no difference in Rate of Physical Fitness for respondents of different age groups.

HA: There is a difference in Rate of Physical Fitness for respondents of different age groups.

ANOVA TEST

```r
# Bartlett's test for homogeneity variance
bartlett.test(sleep_dataset$fitrate, sleep_dataset$agegp3)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  sleep_dataset$fitrate and sleep_dataset$agegp3
## Bartlett's K-squared = 2.9416, df = 2, p-value = 0.2297
```

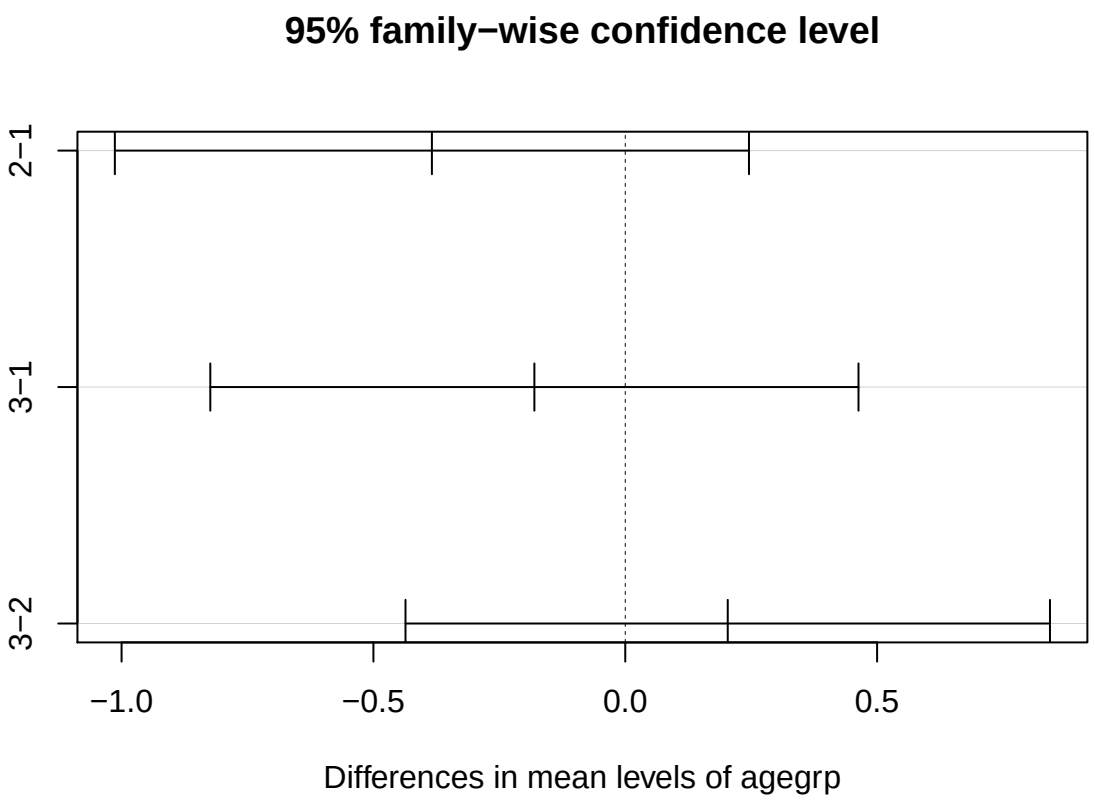
```r
# Compute the analysis of variance
agegrp<-as.factor(sleep_dataset$agegp3)
res.aov <- aov(sleep_dataset$fitrate ~ agegrp, data = sleep_dataset)
# Summary of the analysis
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## agegrp        2    6.1   3.062   1.036  0.357
## Residuals   240  709.5   2.956
## 28 observations deleted due to missingness
```

```r
#Tukey pairwise comparison
tuk_agegp<-TukeyHSD(res.aov)
tuk_agegp
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = sleep_dataset$fitrate ~ agegrp, data = sleep_dataset)
##
## $agegrp
##           diff        lwr       upr     p adj
## 2-1 -0.3838560 -1.0133442 0.2456322 0.3230226
## 3-1 -0.1803928 -0.8238600 0.4630744 0.7862086
## 3-2  0.2034632 -0.4362835 0.8432099 0.7338945
```

```r
plot(tuk_agegp)
```




```r
# Eta square test for calculating effect size.
etaSquared(res.aov,type = 1)
```

```
##             eta.sq eta.sq.part
## agegrp 0.008557546 0.008557546
```

A Bartlett's test was conducted and indicated equality of variance for physical fitness scores for respondents belonging to different age groups (K^2 = 2.94, $P = .23$). A one-way between-groups analysis of variance was conducted to explore physical fitness scores for respondents of different ages. Participants were divided into three groups according to their age (Group 1: 37 yrs or less; Group 2: 38 to 50 yrs; Group 3: 51yrs and above). There was no statistically significant difference level in physical fitness scores for respondents of different age groups. Despite reaching statistical significance, the actual difference in mean scores between groups was quite small. The effect size, calculated using eta squared was .008. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for Group 1 (M = 6.62, SD = 1.53), Group 2 (M = 1.83, SD = 1.83) and Group 3 (M = 34.50, SD = 5.15) was not statistically different. The result is not statistically significant which indicates that there is no evidence to support rejecting the null hypothesis that there is no difference in physical fitness scores for respondents of different age groups.

3.4 Hypothesis 4

H0: There is no difference in Rate of Physical Fitness for respondents having different level of stress.

HA: There is a difference in Rate of Physical Fitness for respondents having different level of stress.

ANOVA TEST

```
# Convert to factor for ANOVA test
stress<-factor(sleep_dataset$stressmo)
# Bartlett's test for homogeneity variance
bartlett.test(sleep_dataset$fitrate, stress)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  sleep_dataset$fitrate and stress
## Bartlett's K-squared = 6.9477, df = 9, p-value = 0.6426
```

```
# Calculate the analysis of variance
res.aov <- aov(sleep_dataset$fitrate ~ stress, data = sleep_dataset)
# Summary of the analysis
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## stress        9   42.8   4.751   1.648  0.102
## Residuals   256  738.1   2.883
## 5 observations deleted due to missingness
```

```
#Tukey pairwise comparison
tuk_wt<-TukeyHSD(res.aov)
plot(tuk_wt)
```


```
# Eta square test for calculating effect size.
etaSquared(res.aov,type = 1)
```

```
##            eta.sq eta.sq.part
## stress 0.05476189  0.05476189
```

A Bartlett's test was conducted and indicated equality of variance for physical fitness scores for respondents within different stress level groups (K^2 = 6.94, $P = .64$). A one-way between-groups analysis of variance was conducted to explore physical fitness scores for respondents within different stress level groups. Participants were divided into 10 groups according to their stress levels over the last month (Group 1 being no stress and Group 10 being highest stress level). There was no statistically significant difference level in physical fitness scores for respondents of different stress level groups. Despite reaching statistical significance, the actual difference in mean scores between groups was quite small. The effect size, calculated using eta squared was .05. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for Group 1 (M = 6.38, SD = 2.26), Group 2 (M = 7.08, SD = 1.67), Group 3 (M = 7.08, SD = 1.72), Group 4 (M = 6.7, SD = 1.82),Group 5 (M = 6.52, SD = 1.62), Group 6 (M = 6.52, SD = 1.5), Group 7 (M = 6.08, SD = 1.62), Group 8 (M = 6.13, SD = 1.77), Group 9 (M = 5.75, SD = 1.65), Group 10 (M = 6.5, SD = 0.84) was not statistically different. The result is not statistically significant which indicates that there is no evidence to support rejecting the null hypothesis that there is no difference in physical fitness scores for respondents of different age groups.

3.5 Hypothesis 5

H0: There is no difference in Rate of Physical Fitness for respondents based on their quality of sleep.

HA: There is a difference in Rate of Physical Fitness for respondents based on their quality of sleep.

ANOVA TEST

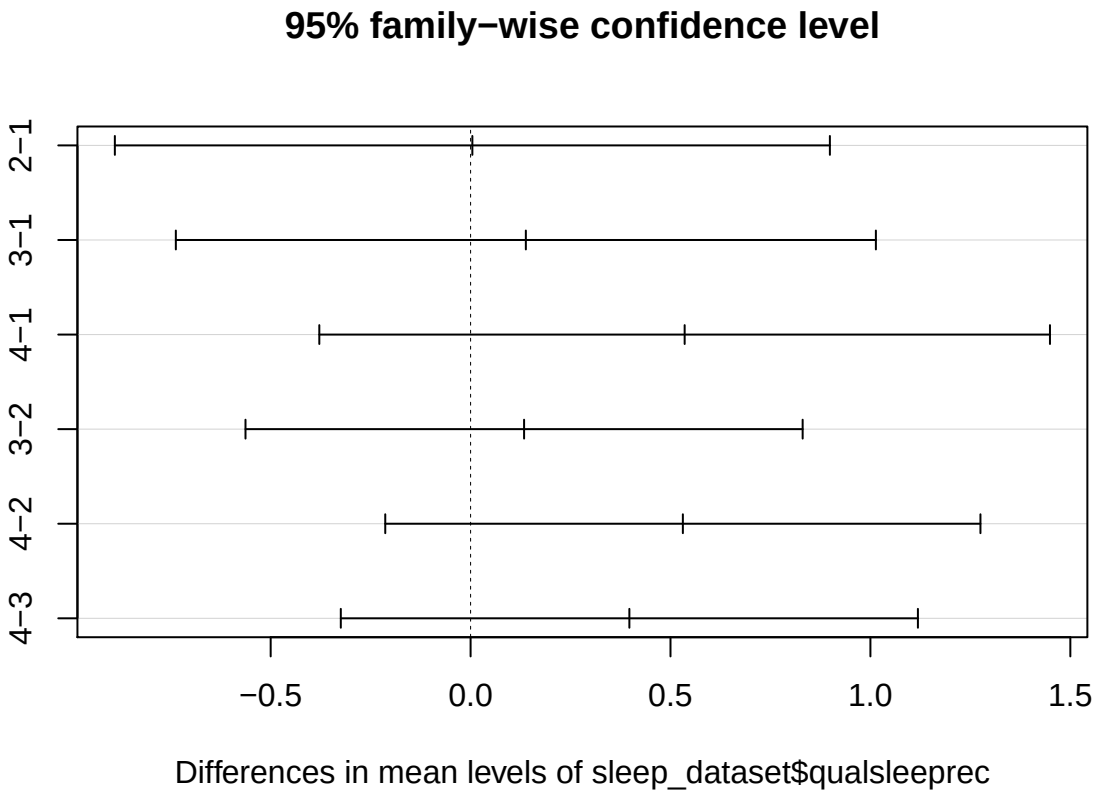
```
# Convert to categorical variable
sleep_dataset$qualsleeprec<-as.factor(sleep_dataset$qualsleeprec)
# Bartlett's test for homogeneity variance
bartlett.test(sleep_dataset$fitrate, sleep_dataset$qualsleeprec)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  sleep_dataset$fitrate and sleep_dataset$qualsleeprec
## Bartlett's K-squared = 1.3544, df = 3, p-value = 0.7163
```

```
# Compute the analysis of variance
res.aov <- aov(sleep_dataset$fitrate ~ sleep_dataset$qualsleeprec, data = sleep_dataset)
# Summary of the analysis
summary(res.aov)
```

```
##                              Df Sum Sq Mean Sq F value Pr(>F)
## sleep_dataset$qualsleeprec   3   12.0   3.987   1.371  0.252
## Residuals                  259  753.3   2.909
## 8 observations deleted due to missingness
```

```
#Tukey pairwise comparison
tuk_qualslp<-TukeyHSD(res.aov)
plot(tuk_qualslp)
```




```
# Eta square test for calculating effect size.
etaSquared(res.aov,type = 1)
```

```
##                                   eta.sq eta.sq.part
## sleep_dataset$qualsleeprec 0.01563078  0.01563078
```

A Bartlett's test was conducted and indicated equality of variance for physical fitness scores for respondents with different rating for sleep quality (K^2 = 1.35, $P = .71$). A one-way between-groups analysis of variance was conducted to explore physical fitness scores for respondents having different sleep quality. Participants were divided into 4 groups according to their quality of sleep (Group 1: Very Poor, Group 2: Fair, Group 3: Good, Group 4: Very Good). There was no statistically significant difference level in physical fitness scores for respondents experiencing different qualities of sleep. Despite reaching statistical significance, the actual difference in mean scores between groups was quite small. The effect size, calculated using eta squared was .01. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for Group 1 (M = 6.22, SD = 1.62), Group 2 (M = 6.23, SD = 1.75), Group 3 (M = 6.36, SD = 1.61), Group 4 (M = 6.76, SD = 1.81),was not statistically different. The result is not statistically significant which indicates that there is no evidence to support rejecting the null hypothesis that there is no difference in physical fitness scores for respondents having different qualities of sleep.

4.0 DISCUSSION

A very weak negative relationship was found between Rate of Physical Fitness and Total Sleepiness and Associated Sensations scale which can be considered to explain 7.07% of their co variance indicating a very weak association between the two factors. No statistically significant difference was found for Physical Fitness scores for respondents having different smoking habits and the effect size found is very small (eta = .004). Moreover, when considering Physical Fitness rating for respondents of different age groups, no statistically significant difference was found between respondents of different age groups. and the effect size is also small (eta = .008). Physical fitness scores for participants who experienced different levels of stress over a period of one month were also not significantly different. Furthermore, physical fitness rate was found to be not significantly different for respondents who did not sleep quite well and those who enjoyed a good sleep.

Concluding this analysis, there is not enough evidence to suggest that Total Sleepiness and Associated Sensations factor can be considered to be a predictor of a person's Rate of Physical Fitness. Also, the small effect sizes found for differences based on age group indicate that there is no significant difference in Physical Fitness ratings based on different age groups. The lack of evidence to support considering smoking habits as a differentiating factor for Physical Fitness also indicates the need for more in depth investigation. Factors like Level of Stress experienced by a person and Quality of Sleep experienced by a person also fail to differentiate the population based on Physical Fitness scores. Hence, further research is required in order to establish a relationship between Physical Fitness and Total Sleepiness and Associated Sensations factor.

5.0 References

Field, A., Miles, J., Field, Z. (2012). *Discovering Statistics Using R. Sage Publications*

George, D. (2011). *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e. Pearson Education India*