HW3

Nicole Baldaccini

2024-11-09

```
setwd("/cloud/project")
Text <- read.csv("TextMessages.csv", header =TRUE)
names(Text)
```

```
## [1] "Group"       "Baseline"    "Six_months"  "Participant"
```

```
is.factor(Text$Six_months)
```

```
## [1] FALSE
```

```
install.packages("pastecs")
library(pastecs)
round( stat.desc(Text$Baseline) , 2)
```

```
##      nbr.val     nbr.null       nbr.na          min          max        range
##        50.00         0.00         0.00        46.00        89.00        43.00
##          sum       median         mean      SE.mean CI.mean.0.95          var
##      3261.00        64.50        65.22         1.51         3.03       113.52
##      std.dev     coef.var
##        10.65         0.16
```

```
round( stat.desc(Text$Six_months) , 2)
```

```
##      nbr.val     nbr.null       nbr.na          min          max        range
##        50.00         0.00         0.00         9.00        79.00        70.00
##          sum       median         mean      SE.mean CI.mean.0.95          var
##      2870.00        60.50        57.40         1.97         3.96       194.12
##      std.dev     coef.var
##        13.93         0.24
```

```
#from this given data set we can determine that their
#are 50 participants. The mean is 57.4. It is lower than the baseline mean at
#65.22, which indicates a decline from baseline to sixmonths. The standard
#deviation is 13.93 which shows that there is variability among the participants
#six month scores. The min value is 9 showing that a least one person scored
#very low.The max score is 79, which is lower than baseline of 89. This shows
#that participants generally scored lower at the six month point versus baseline.
#The first quartile is 53 meaning 25% of people scored below this value. The
#median is the midpoint which is 60.5. Lastly, the 3rd quartile is 63, the point
#of where 75% of the scores fall.
stats_list <- lapply(Text[, c("Baseline", "Six_months")], function(x) stat.desc(x, basic = TRUE, desc =
combined_stats <- do.call(cbind, stats_list)
colnames(combined_stats) <- c("Baseline", "Six_Months")
combined_stats <- round(combined_stats, 2)
print(combined_stats)
```

```
##              Baseline Six_Months
## nbr.val         50.00      50.00
## nbr.null         0.00       0.00
## nbr.na           0.00       0.00
## min             46.00       9.00
## max             89.00      79.00
## range           43.00      70.00
## sum           3261.00    2870.00
## median          64.50      60.50
## mean            65.22      57.40
## SE.mean          1.51       1.97
## CI.mean.0.95     3.03       3.96
## var            113.52     194.12
## std.dev         10.65      13.93
## coef.var         0.16       0.24
## skewness         0.24      -1.28
## skew.2SE         0.35      -1.90
## kurtosis        -0.80       2.54
## kurt.2SE        -0.61       1.92
## normtest.W       0.98       0.88
## normtest.p       0.40       0.00
```

```
#The purpose of the code above is to get a clearer look of the stats side by
#side. We can note some significant differences as stated above.
```