Heart failure data analysis in R

Jyotishka Datta

March 29, 2019

Data reading

```
rm(list=ls())
setwd("C:/Users/jd033/OneDrive/Documents/R/Nursing")
heart <- read.csv("Heart-failure-project-data.csv",sep=",",header=T)
str(heart)
```

'data.frame': 1350 obs. of 5 variables: $ ï..Patient.ID: int 1 1 1 1 1 1 1 1 1 1 ... $ Question : Factor w/ 25 levels "Act on Sx","Activity",..: 25 25 22 22 21 21 2 2 3 3 ... $ Score : int 1 3 4 4 4 4 3 3 4 4 ... $ Section : Factor w/ 3 levels "A","B","C": 1 1 1 1 1 1 1 1 1 1 ... $ Group : Factor w/ 2 levels "Post","Pre": 2 1 2 1 2 1 2 1 2 1 ...

Before we analyze the data in R, we need to convert the data from a long format to a wide format. We also change the name of the first column from `patient ID` to `subject`.

This is done as follows:

```
library(tidyverse)
library(reshape2)
colnames(heart)[1] <- "subject"
heart.new <- heart[,c(1:3,5)]
heart.wide <-  dcast(heart.new, subject + Group ~ Question, sum, value.var="Score")
dfSummary(heart.wide)
```

Data Frame Summary

heart.wide
Dimensions: 60 x 27
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	subject [integer]	Mean (sd) : 17.3 (19.6) min < med < max: 1 < 15 < 151 IQR (CV) : 14.5 (1.1)	30 distinct values	: : : : : : :	59 (98.33%)	1 (1.67%)
2	Group [factor]	1. Post 2. Pre	30 (50.0%) 30 (50.0%)	IIIIIIIIII IIIIIIIIII	60 (100%)	0 (0%)
3	Act on Sx [integer]	Mean (sd) : 2.6 (1) min < med < max: 0 < 3 < 4 IQR (CV) : 1 (0.4)	0 : 3 (5.0%) 1 : 4 (6.7%) 2 : 17 (28.3%) 3 : 26 (43.3%) 4 : 10 (16.7%)	I I IIIII IIIIIIII III	60 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
4	Activity [integer]	Mean (sd) : 2.5 (1) min < med < max: 0 < 3 < 4 IQR (CV) : 1 (0.4)	0 : 2 (3.3%) 1 : 9 (15.0%) 2 : 18 (30.0%) 3 : 22 (36.7%) 4 : 9 (15.0%)	III IIIIII IIIIIII III	60 (100%)	0 (0%)
5	Appointment [integer]	Mean (sd) : 3.3 (1.2) min < med < max: 0 < 4 < 4 IQR (CV) : 1 (0.4)	0 : 2 (3.3%) 1 : 7 (11.7%) 2 : 4 (6.7%) 3 : 7 (11.7%) 4 : 40 (66.7%)	II I II IIIIIIIIIIII	60 (100%)	0 (0%)
6	Ask for low Salt [integer]	Mean (sd) : 1.4 (0.8) min < med < max: 0 < 1 < 4 IQR (CV) : 1 (0.6)	0 : 2 (3.4%) 1 : 41 (69.5%) 2 : 11 (18.6%) 3 : 2 (3.4%) 4 : 3 (5.1%)	IIIIIIIIIIII III I	59 (98.33%)	1 (1.67%)
7	Contact NP [integer]	Mean (sd) : 2.6 (1.3) min < med < max: 0 < 3 < 4 IQR (CV) : 3 (0.5)	0 : 2 (3.3%) 1 : 16 (26.7%) 2 : 9 (15.0%) 3 : 10 (16.7%) 4 : 23 (38.3%)	IIIII III III IIIIIII	60 (100%)	0 (0%)
8	Eval Sx [integer]	Mean (sd) : 2.6 (1.1) min < med < max: 0 < 3 < 5 IQR (CV) : 1 (0.4)	0 : 4 (6.7%) 1 : 3 (5.0%) 2 : 15 (25.0%) 3 : 27 (45.0%) 4 : 10 (16.7%) 5 : 1 (1.7%)	I I IIIII IIIIIIII III	60 (100%)	0 (0%)
9	Eval Tx [integer]	Mean (sd) : 2.6 (1.2) min < med < max: 0 < 3 < 5 IQR (CV) : 1 (0.4)	0 : 4 (6.7%) 1 : 6 (10.0%) 2 : 13 (21.7%) 3 : 24 (40.0%) 4 : 12 (20.0%) 5 : 1 (1.7%)	I II IIII IIIIIII IIII	60 (100%)	0 (0%)
10	Exercise [integer]	Mean (sd) : 0.1 (0.5) min < med < max: 0 < 0 < 2 IQR (CV) : 0 (3.9)	0 : 56 (93.3%) 1 : 1 (1.7%) 2 : 3 (5.0%)	IIIIIIIIIIIIIIIII I	60 (100%)	0 (0%)
11	Exercise30 [integer]	Mean (sd) : 1.7 (1.1) min < med < max: 0 < 2 < 4 IQR (CV) : 1 (0.6)	0 : 6 (10.0%) 1 : 23 (38.3%) 2 : 18 (30.0%) 3 : 9 (15.0%) 4 : 4 (6.7%)	II IIIIIII IIIIII III I	60 (100%)	0 (0%)
12	Follow Tx [integer]	Mean (sd) : 2.6 (1) min < med < max: 0 < 3 < 5 IQR (CV) : 1 (0.4)	0 : 3 (5.0%) 1 : 2 (3.3%) 2 : 21 (35.0%) 3 : 24 (40.0%) 4 : 9 (15.0%) 5 : 1 (1.7%)	I IIIIIII IIIIIIII III	60 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
13	Forget Meds [integer]	Mean (sd) : 1.2 (0.7) min < med < max: 0 < 1 < 4 IQR (CV) : 0 (0.5)	0 : 2 (3.3%) 1 : 47 (78.3%) 2 : 7 (11.7%) 3 : 3 (5.0%) 4 : 1 (1.7%)	IIIIIIIIIIIIII II I	60 (100%)	0 (0%)
14	Free fromSx [integer]	Mean (sd) : 2.3 (0.8) min < med < max: 0 < 2 < 4 IQR (CV) : 1 (0.4)	0 : 2 (3.3%) 1 : 4 (6.7%) 2 : 31 (51.7%) 3 : 20 (33.3%) 4 : 3 (5.0%)	I IIIIIIIII IIIIII I	60 (100%)	0 (0%)
15	Inc Med [integer]	Mean (sd) : 2.2 (1.2) min < med < max: 0 < 2.5 < 4 IQR (CV) : 2 (0.6)	0 : 2 (3.3%) 1 : 23 (38.3%) 2 : 5 (8.3%) 3 : 20 (33.3%) 4 : 10 (16.7%)	IIIIIII I IIIIII III	60 (100%)	0 (0%)
16	Low Salt Diet [integer]	Mean (sd) : 2.4 (1.1) min < med < max: 0 < 2 < 4 IQR (CV) : 1.2 (0.5)	0 : 2 (3.3%) 1 : 13 (21.7%) 2 : 16 (26.7%) 3 : 19 (31.7%) 4 : 10 (16.7%)	IIII IIIII IIIIII III	60 (100%)	0 (0%)
17	Pill box [integer]	Mean (sd) : 3.3 (1.3) min < med < max: 0 < 4 < 4 IQR (CV) : 0 (0.4)	0 : 2 (3.4%) 1 : 9 (15.2%) 2 : 2 (3.4%) 3 : 1 (1.7%) 4 : 45 (76.3%)	III IIIIIIIIIIIIII	59 (98.33%)	1 (1.67%)
18	Rec change [integer]	Mean (sd) : 2.6 (1.1) min < med < max: 0 < 3 < 5 IQR (CV) : 1 (0.4)	0 : 4 (6.7%) 1 : 4 (6.7%) 2 : 15 (25.0%) 3 : 27 (45.0%) 4 : 9 (15.0%) 5 : 1 (1.7%)	I I IIIII IIIIIIII III	60 (100%)	0 (0%)
19	Recognize [integer]	Mean (sd) : 1.2 (1.3) min < med < max: 0 < 1 < 4 IQR (CV) : 2.2 (1.1)	0 : 27 (48.2%) 1 : 6 (10.7%) 2 : 9 (16.1%) 3 : 13 (23.2%) 4 : 1 (1.8%)	IIIIIIII II III IIII	56 (93.33%)	4 (6.67%)
20	Reduce H2O [integer]	Mean (sd) : 2.3 (1.1) min < med < max: 0 < 3 < 4 IQR (CV) : 2 (0.5)	0 : 2 (3.3%) 1 : 16 (26.7%) 2 : 11 (18.3%) 3 : 24 (40.0%) 4 : 7 (11.7%)	IIIII III IIIIIIII II	60 (100%)	0 (0%)
21	Reduce Na [integer]	Mean (sd) : 2.5 (1.1) min < med < max: 0 < 3 < 4 IQR (CV) : 1 (0.5)	0 : 2 (3.3%) 1 : 12 (20.0%) 2 : 14 (23.3%) 3 : 21 (35.0%) 4 : 11 (18.3%)	IIII IIII IIIIIII III	60 (100%)	0 (0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
22	Remedy Eval [integer]	Mean (sd) : 1.2 (1.4) min < med < max: 0 < 0 < 4 IQR (CV) : 3 (1.1)	0 : 31 (51.7%) 1 : 4 (6.7%) 2 : 7 (11.7%) 3 : 17 (28.3%) 4 : 1 (1.7%)	IIIIIIIII I II IIIII	60 (100%)	0 (0%)
23	Sick [integer]	Mean (sd) : 2.9 (1.1) min < med < max: 0 < 3 < 4 IQR (CV) : 2 (0.4)	0 : 2 (3.3%) 1 : 7 (11.7%) 2 : 10 (16.7%) 3 : 19 (31.7%) 4 : 22 (36.7%)	II III IIIIII IIIIIII	60 (100%)	0 (0%)
24	Swelling [integer]	Mean (sd) : 3.1 (1.2) min < med < max: 0 < 4 < 4 IQR (CV) : 2 (0.4)	0 : 2 (3.3%) 1 : 4 (6.7%) 2 : 12 (20.0%) 3 : 9 (15.0%) 4 : 33 (55.0%)	I IIII III IIIIIIIIII	60 (100%)	0 (0%)
25	Symptoms [integer]	Mean (sd) : 0.7 (0.7) min < med < max: 0 < 1 < 4 IQR (CV) : 1 (1)	0 : 21 (35.0%) 1 : 37 (61.7%) 3 : 1 (1.7%) 4 : 1 (1.7%)	IIIIIII IIIIIIIIII	60 (100%)	0 (0%)
26	Tx Helped [integer]	Mean (sd) : 0.6 (1.1) min < med < max: 0 < 0 < 4 IQR (CV) : 1.2 (1.8)	0 : 44 (73.3%) 1 : 1 (1.7%) 2 : 10 (16.7%) 3 : 3 (5.0%) 4 : 2 (3.3%)	IIIIIIIIIIII III I	60 (100%)	0 (0%)
27	Weigh [integer]	Mean (sd) : 2.6 (1.2) min < med < max: 0 < 3 < 4 IQR (CV) : 2 (0.5)	0 : 2 (3.3%) 1 : 11 (18.3%) 2 : 14 (23.3%) 3 : 12 (20.0%) 4 : 21 (35.0%)	III IIII IIII IIIIIII	60 (100%)	0 (0%)

One of the variable `Forget Meds` is coded in reverse order. We recode that here using the recode function in R.

```
(heart.wide$`Forget Meds`)
```

[1] 1 3 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 3 1 1 1 1 1 2 1 1 1 3 1 [36] 1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 1 1 1 1 1 4 1 1 0 0

```
heart.wide$`Forget Meds`<- recode(heart.wide$`Forget Meds`, `1` = 4, `2` = 3, `3`= 2, `4` = 1, `0` = 0)
(heart.wide$`Forget Meds`)
```

[1] 4 2 4 4 4 4 4 4 4 4 4 3 4 4 4 4 4 4 4 3 4 4 4 2 4 4 4 4 4 4 3 4 4 4 2 4 [36] 4 4 4 4 4 4 3 4 3 4 4 4 3 4 3 4 4 4 4 4 1 4 4 0 0

```
heart.wide[is.na(heart.wide)] <- 0
```

The next step is calculating the normalized score for all section A items after adding them.

```
heart.wide$sec.A.score = with(heart.wide, Weigh+Swelling+Sick+Activity+Appointment+`Low Salt Diet`+
       Exercise+`Forget Meds`+`Ask for  low Salt`+`Pill box`)
Max.sec.A.score = 4*10
```

```
Min.sec.A.score = 0
(heart.wide$normalized.sec.A.score = (heart.wide$sec.A.score - Min.sec.A.score)/(Max.sec.A.score - Min.s
```

[1] 70.0 70.0 82.5 75.0 82.5 65.0 82.5 65.0 80.0 55.0 75.0 40.0 90.0 75.0 [15] 77.5 70.0 80.0 72.5 60.0 27.5 90.0 90.0 75.0 25.0 77.5 70.0 50.0 42.5 [29] 80.0 62.5 72.5 62.5 70.0 42.5 72.5 67.5 72.5 42.5 60.0 32.5 55.0 45.0 [43] 70.0 27.5 75.0 60.0 72.5 47.5 55.0 37.5 70.0 60.0 82.5 67.5 77.5 62.5 [57] 75.0 60.0 0.0 5.0

Now, we can look at whether the normalized score was different between the **pre** and **post** groups:

```
fit <- lm(normalized.sec.A.score ~ Group, data = heart.wide)
summary(fit)
```

Call: lm(formula = normalized.sec.A.score ~ Group, data = heart.wide)

Residuals: Min 1Q Median 3Q Max -71.083 -11.083 3.917 10.979 35.833

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 71.083 3.239 21.949 < 2e-16 ***GroupPre -16.917 4.580 -3.694 0.000491*** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ‘’ 1

Residual standard error: 17.74 on 58 degrees of freedom Multiple R-squared: 0.1904, Adjusted R-squared: 0.1765 F-statistic: 13.64 on 1 and 58 DF, p-value: 0.0004912

We can also look at finer resolution, i.e. if there was a difference between `Ask for low salt` variable between the pre and post groups. This should be done only for specific pre-determined variables. Otherwise there is an issue of multiplicity here.

```
fit <- lm(`Ask for  low Salt` ~ Group, data = heart.wide)
summary(fit)
```

Call: lm(formula = `Ask for  low Salt` ~ Group, data = heart.wide)

Residuals: Min 1Q Median 3Q Max -1.5667 -0.5667 -0.1333 0.4333 2.8667

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.5667 0.1494 10.487 5.17e-15 ** *GroupPre -0.4333 0.2113 -2.051 0.0448*
— Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ‘’ 1

Residual standard error: 0.8183 on 58 degrees of freedom Multiple R-squared: 0.06763, Adjusted R-squared: 0.05155 F-statistic: 4.207 on 1 and 58 DF, p-value: 0.04479