

# Test Clean & Explore of PMS '07

Michelle K Jamieson

11/06/2020

```
# read in 2007 data
psychmorb07 <- read_spss("pms07.sav")

# Subsetting
pm07.sub <- psychmorb07[,c("ResSex", "ResAge", "SF1", "Happy",
                           "Health6", "Everwk", "Psycdis")]

# Change Labels
attr(pm07.sub$ResSex, "label") <- "Respondant Sex"
attr(pm07.sub$ResAge, "label") <- "Respondant Age"
attr(pm07.sub$SF1, "label") <- "General Health"
attr(pm07.sub$Happy, "label") <- "How Happy Are You?"
attr(pm07.sub$Health6, "label") <- "Mental Health Condition Mentioned"
attr(pm07.sub$Everwk, "label") <- "Ever Had a Paid Job"
attr(pm07.sub$Psycdis, "label") <- "Definitive Psychosis in Last Year"

colnames(pm07.sub)
```

```
## [1] "ResSex" "ResAge" "SF1" "Happy" "Health6" "Everwk" "Psycdis"
```

```
# rename variables
pm07.sub <- pm07.sub %>% rename(Sex = ResSex, Age = ResAge,
                               GenHeal = SF1, HowHap = Happy,
                               MenHeal = Health6, PaidJob = Everwk,
                               DefPsych = Psycdis)

colnames(pm07.sub)
```

```
## [1] "Sex" "Age" "GenHeal" "HowHap" "MenHeal" "PaidJob" "DefPsych"
```

```
# Make NAs explicit
pm07.sub = pm07.sub %>% mutate_if(is.factor,
                                   fct_explicit_na,
                                   na_level = "No Answer")

# mutate to factors and keep
cleanpm07 <- pm07.sub %>%
  mutate(Sex = as_factor(Sex),
         Age = as_factor(Age),
         GenHeal = as_factor(GenHeal),
         HowHap = as_factor(HowHap),
```

```

MenHeal = as_factor(MenHeal),
PaidJob = as_factor(PaidJob),
DefPsych = as_factor(DefPsych))
head(cleanpm07)

```

```

## # A tibble: 6 x 7
##   Sex    Age  GenHeal   HowHap      MenHeal    PaidJob DefPsych
##   <fct> <fct> <fct>    <fct>      <fct>      <fct>   <fct>
## 1 Female 53   very good Very happy   Mentioned    <NA>    No
## 2 Female 78   fair      Fairly happy Not mentioned Yes      No
## 3 Female 52   good      Fairly happy Mentioned    <NA>    No
## 4 Female 58   or, poor? Fairly happy Not mentioned <NA>    No
## 5 Male   66   good      Fairly happy Not mentioned Yes      No
## 6 Male   86   good      Fairly happy Mentioned    Yes      No

```

```

# look
summary(cleanpm07)

```

```

##              Sex              Age              GenHeal
## No answer/refused : 0 39 : 157 very good:2354
## Don't know        : 0 38 : 146 good :2079
## Missing data      : 0 40 : 144 Excellent:1271
## Proxy             : 0 41 : 144 fair :1183
## Item not applicable: 0 36 : 141 or, poor?: 513
## Male              :3197 37 : 141 (Other) : 0
## Female            :4206 (Other):6530 NA's : 3
##              HowHap              MenHeal              PaidJob
## Fairly happy :3956 Not mentioned :5686 Yes :3073
## Very happy :2746 Mentioned :1714 No : 296
## Not too happy : 697 No answer/refused: 0 No answer/refused: 0
## No answer/refused: 0 Don't know : 0 Don't know : 0
## Don't know : 0 Missing data : 0 Missing data : 0
## (Other) : 0 (Other) : 0 (Other) : 0
## NA's : 4 NA's : 3 NA's :4034
##              DefPsych
## No :7266
## Yes : 23
## No answer/refused: 0
## Don't know : 0
## Missing data : 0
## (Other) : 0
## NA's : 114

```

```

skim(cleanpm07)

```

Table 1: Data summary

Name	cleanpm07
Number of rows	7403
Number of columns	7

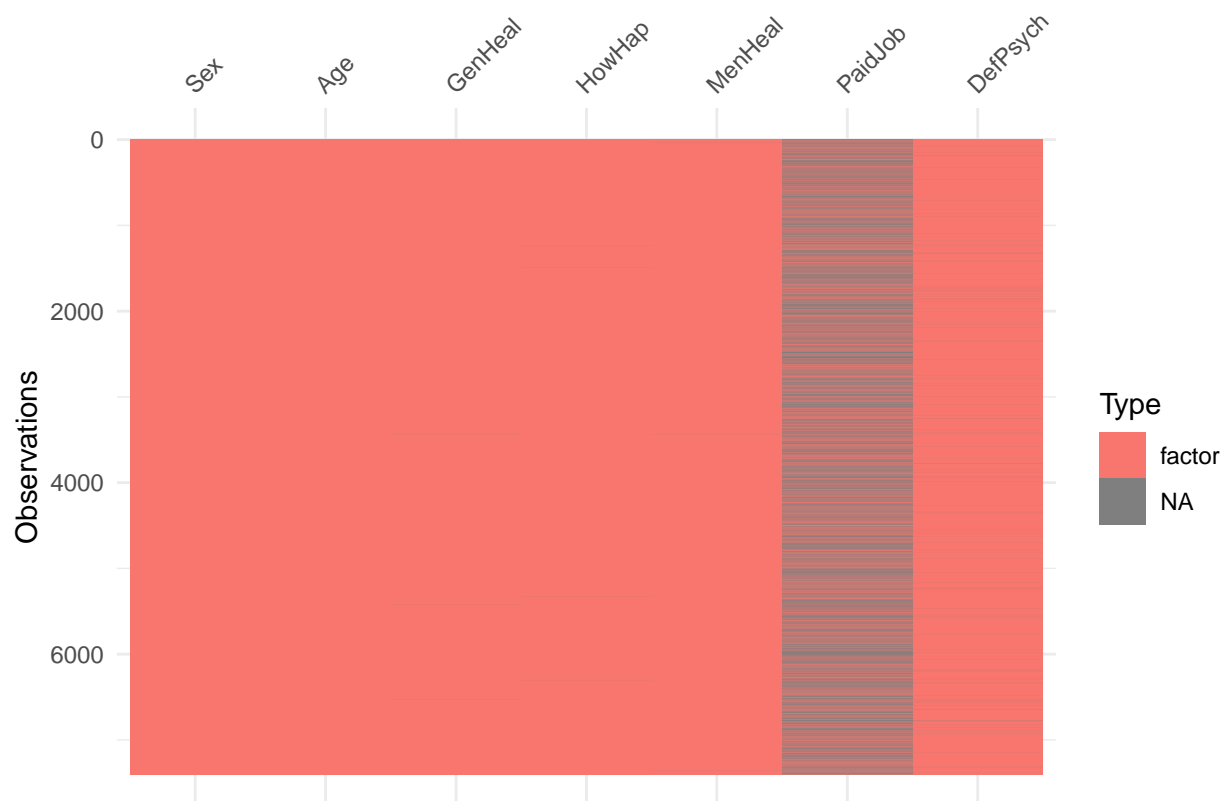
Table 1: Data summary

Column type frequency:	
factor	7
Group variables	None

**Variable type: factor**

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
Sex	0	1.00	FALSE	2	Fem: 4206, Mal: 3197, No : 0, Don: 0
Age	0	1.00	FALSE	80	39: 157, 38: 146, 40: 144, 41: 144
GenHeal	3	1.00	FALSE	5	ver: 2354, goo: 2079, Exc: 1271, fai: 1183
HowHap	4	1.00	FALSE	3	Fai: 3956, Ver: 2746, Not: 697, No : 0
MenHeal	3	1.00	FALSE	2	Not: 5686, Men: 1714, No : 0, Don: 0
PaidJob	4034	0.46	FALSE	2	Yes: 3073, No: 296, No : 0, Don: 0
DefPsych	114	0.98	FALSE	2	No: 7266, Yes: 23, No : 0, Don: 0

```
vis_dat(cleanpm07)
```



```
describe(cleanpm07)
```

```
##      vars    n  mean    sd median trimmed  mad min max range  skew
```

```
## Sex*      1 7403  6.57  0.50      7      6.59  0.00  6      7      1 -0.28
## Age*      2 7403 41.12 18.59     40     40.94 22.24  6     85     79  0.09
## GenHeal*   3 7400  8.64  1.14      9      8.58  1.48  7     11      4  0.33
## HowHap*    4 7399  7.72  0.62      8      7.66  0.00  7      9      2  0.28
## MenHeal*   5 7400  7.23  0.42      7      7.16  0.00  7      8      1  1.27
## PaidJob*   6 3369  7.09  0.28      7      7.00  0.00  7      8      1  2.91
## DefPsych*  7 7289  6.00  0.06      6      6.00  0.00  6      7      1 17.71
##          kurtosis  se
## Sex*      -1.92 0.01
## Age*      -0.94 0.22
## GenHeal*   -0.67 0.01
## HowHap*    -0.65 0.01
## MenHeal*   -0.38 0.00
## PaidJob*    6.47 0.00
## DefPsych* 311.83 0.00
```

```
# descriptives for mental health condition mentioned x
# def had psychosis condition in past year

# total no of people who mentioned having any mental health
# condition by no of people who definetly had psychosis type
# condition in the past year (2006-2007)
cleanpm07 %>%
  group_by(MenHeal, DefPsych)%>%
  summarize(n=n())%>%
  kable()
```

```
## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'
```

```
## Warning: Factor 'DefPsych' contains implicit NA, consider using
## 'forcats::fct_explicit_na'
```

MenHeal	DefPsych	n
Not mentioned	No	5659
Not mentioned	Yes	2
Not mentioned	NA	25
Mentioned	No	1604
Mentioned	Yes	21
Mentioned	NA	89
NA	No	3

```
# crosstab of people who mentioned having any mental health
# condition by no of people who definetly had psychosis type
# condition in the past year (2006-2007)
cleanpm07 %>%
  group_by(MenHeal, DefPsych)%>%
  summarise(n=n())%>%
  spread(DefPsych, n)%>%
  kable()
```

```
## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'

## Warning: Factor 'DefPsych' contains implicit NA, consider using
## 'forcats::fct_explicit_na'

## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'

## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'

## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'

## Warning: Factor 'MenHeal' contains implicit NA, consider using
## 'forcats::fct_explicit_na'
```

MenHeal	No	Yes	
Not mentioned	5659	2	25
Mentioned	1604	21	89
NA	3	NA	NA

```
# quick way to pull together row/column
# frequencies and proportions mental health and
# psychosis conditions vars
table(cleanpm07$MenHeal, cleanpm07$DefPsych)
```

```
##
##               No answer/refused Don't know Missing data Proxy
## No answer/refused                0          0          0      0
## Don't know                      0          0          0      0
## Missing data                    0          0          0      0
## Proxy                          0          0          0      0
## Schedule not applicable          0          0          0      0
## Item not applicable              0          0          0      0
## Not mentioned                   0          0          0      0
## Mentioned                       0          0          0      0
##
##               Item not applicable    No  Yes
## No answer/refused                0    0    0
## Don't know                      0    0    0
## Missing data                    0    0    0
## Proxy                          0    0    0
## Schedule not applicable          0    0    0
## Item not applicable              0    0    0
## Not mentioned                   0 5659    2
## Mentioned                       0 1604   21
```

```
# table frequency can also be called by using the ftable() command.
```

```
cpm07_table<- table(cleanpm07$MenHeal, cleanpm07$DefPsych) #define object w/table parameters for simple
ftable(cpm07_table)
```

```
##
##          No answer/refused Don't know Missing data Proxy Item not applicable No Y
##
## No answer/refused          0          0          0    0          0    0
## Don't know                0          0          0    0          0    0
## Missing data              0          0          0    0          0    0
## Proxy                     0          0          0    0          0    0
## Schedule not applicable    0          0          0    0          0    0
## Item not applicable        0          0          0    0          0    0
## Not mentioned             0          0          0    0          0 5659
## Mentioned                 0          0          0    0          0 1604
```

```
# For row frequencies,
margin.table(cpm07_table, 1)
```

```
##
##          No answer/refused          Don't know          Missing data
##                0                0                0
##          Proxy Schedule not applicable    Item not applicable
##                0                0                0
##          Not mentioned          Mentioned
##          5661                1625
```

```
# For column frequencies
margin.table(cpm07_table, 2)
```

```
##
##          No answer/refused          Don't know          Missing data          Proxy
##                0                0                0                0
##          Item not applicable          No          Yes
##                0                7263                23
```

```
# For proportion of the entire table,
prop.table(cpm07_table) #proportion of entire table
```

```
##
##          No answer/refused Don't know Missing data
##          No answer/refused 0.000000000 0.000000000 0.000000000
##          Don't know        0.000000000 0.000000000 0.000000000
##          Missing data      0.000000000 0.000000000 0.000000000
##          Proxy             0.000000000 0.000000000 0.000000000
##          Schedule not applicable 0.000000000 0.000000000 0.000000000
##          Item not applicable 0.000000000 0.000000000 0.000000000
##          Not mentioned     0.000000000 0.000000000 0.000000000
##          Mentioned         0.000000000 0.000000000 0.000000000
##
##          Proxy Item not applicable          No
```

```
## No answer/refused      0.000000000      0.000000000 0.000000000
## Don't know             0.000000000      0.000000000 0.000000000
## Missing data          0.000000000      0.000000000 0.000000000
## Proxy                  0.000000000      0.000000000 0.000000000
## Schedule not applicable 0.000000000      0.000000000 0.000000000
## Item not applicable     0.000000000      0.000000000 0.000000000
## Not mentioned          0.000000000      0.000000000 0.776695032
## Mentioned              0.000000000      0.000000000 0.220148229
##
##                               Yes
## No answer/refused      0.000000000
## Don't know             0.000000000
## Missing data          0.000000000
## Proxy                  0.000000000
## Schedule not applicable 0.000000000
## Item not applicable     0.000000000
## Not mentioned          0.000274499
## Mentioned              0.002882240
```

```
# For column proportions,
prop.table(cpm07_table, 2) #proportion of entire column
```

```
##
##                               No answer/refused Don't know Missing data Proxy
## No answer/refused
## Don't know
## Missing data
## Proxy
## Schedule not applicable
## Item not applicable
## Not mentioned
## Mentioned
##
##                               Item not applicable      No      Yes
## No answer/refused      0.00000000 0.00000000
## Don't know             0.00000000 0.00000000
## Missing data          0.00000000 0.00000000
## Proxy                  0.00000000 0.00000000
## Schedule not applicable 0.00000000 0.00000000
## Item not applicable     0.00000000 0.00000000
## Not mentioned          0.77915462 0.08695652
## Mentioned              0.22084538 0.91304348
```

```
# produces frequencies, and table,
# row, & column proportions with a single command
CrossTable(cleanpm07$MenHeal, cleanpm07$DefPsych)
```

```
##
##
##      Cell Contents
## |-----|
## |                               N |
## | Chi-square contribution |
```

```
## |          N / Row Total |
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table:  7286
##
##
##              | cleanpm07$DefPsych
## cleanpm07$MenHeal |          No |          Yes | Row Total |
## -----|-----|-----|-----|
##      Not mentioned |         5659 |           2 |        5661 |
##                   |         0.045 |        14.094 |           |
##                   |         1.000 |         0.000 |        0.777 |
##                   |         0.779 |         0.087 |           |
##                   |         0.777 |         0.000 |           |
## -----|-----|-----|-----|
##           Mentioned |         1604 |          21 |        1625 |
##                   |         0.155 |        49.100 |           |
##                   |         0.987 |         0.013 |        0.223 |
##                   |         0.221 |         0.913 |           |
##                   |         0.220 |         0.003 |           |
## -----|-----|-----|-----|
##      Column Total |         7263 |          23 |        7286 |
##                   |         0.997 |         0.003 |           |
## -----|-----|-----|-----|
##
##
```

```
####
# descriptives for def had psychosis in past year x
# ever had a paid job

#cleanpm07 %>%
#   group_by(DefPsych, PaidJob)%>%
#   summarize(n=n())%>%
#   kable()

#cleanpm07 %>%
#   group_by(DefPsych, PaidJob)%>%
#   summarise(n=n())%>%
#   spread(PaidJob, n)%>%
#   kable()

#table(cleanpm07$DefPsych, cleanpm07$PaidJob)

# Creates really cool exploratory report of data set
# DataExplorer::create_report(cleanpm07)
```