

Multivariate Analysis for the Behavioral Sciences,
Second Edition (Chapman and Hall/CRC, 2019)

**Examples of Chapter 11:
Missing Values**

Kimmo Vehkalahti and Brian S. Everitt

22 November 2018

Contents

Examples	2
Table 11.1: Car Wash Data That Includes Missing Values	2
Table 11.2	3
Table 11.3	4
Table 11.4	5
Beat the Blues Revisited (Again)	6
Table 11.5	8
Table 11.6	9
Table 11.7	10
Table 11.8	12

Examples

Table 11.1: Car Wash Data That Includes Missing Values

```
library("tidyr"); library("dplyr")

##
## 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
WASH2 <- read.table("data/carwash2.txt", sep = ' ', header = TRUE)

glimpse(WASH2)

## Observations: 20
## Variables: 3
## $ Time <int> 46, 79, 63, 42, 44, 59, 27, 40, 30, 61, 50, 65, 70, 20, ...
## $ Age <int> NA, NA, 26, 34, NA, 26, 25, NA, 45, 31, 28, 34, 44, NA, ...
## $ Extro <int> 40, NA, NA, 43, 41, 42, NA, 30, 35, 48, 37, NA, NA, 28, ...

WASH2

##   Time Age Extro
## 1   46  NA   40
## 2   79  NA   NA
## 3   63  26   NA
## 4   42  34   43
## 5   44  NA   41
## 6   59  26   42
## 7   27  25   NA
## 8   40  NA   30
## 9   30  45   35
## 10  61  31   48
## 11  50  28   37
## 12  65  34   NA
## 13  70  44   NA
## 14  20  NA   28
## 15  32  NA   34
## 16  65  28   44
## 17  30  50   33
## 18  48  23   NA
## 19  20  NA   NA
## 20  56  NA   47
```

Table 11.2

```
fit02 <- lm(Time ~ Age + Extro, data = WASH2)
summary(fit02)
```

```
##
## Call:
## lm(formula = Time ~ Age + Extro, data = WASH2)
##
## Residuals:
```

	4	6	9	10	11	16	17
##	-9.3856	0.6487	-2.5691	1.7053	-1.4362	6.6655	4.3713

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	42.8029	38.6640	1.107	0.3304
## Age	-0.9939	0.3979	-2.498	0.0669 .
## Extro	0.9855	0.6849	1.439	0.2236

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.396 on 4 degrees of freedom
## (13 observations deleted due to missingness)
## Multiple R-squared:  0.8708, Adjusted R-squared:  0.8063
## F-statistic: 13.49 on 2 and 4 DF, p-value: 0.01668
```

Table 11.3

```
#install.packages("mice")
library("mice")

## Loading required package: lattice

##
## Attaching package: 'mice'

## The following object is masked from 'package:tidyr':
##
##   complete

## The following objects are masked from 'package:base':
##
##   cbind, rbind

imp1_WASH2 <- mice(WASH2, method = "mean", m = 1, maxit = 1, seed = 2018)

##
## iter imp variable
##   1   1 Age  Extro

fit12 <- with(imp1_WASH2,
              lm(Time ~ Age + Extro))
summary(fit12)

##           term      estimate std.error  statistic    p.value
## 1 (Intercept) -20.84188389  37.9367823  -0.54938460  0.58988618
## 2           Age  -0.05408842   0.5452381  -0.09920148  0.92213870
## 3          Extro   1.81191481   0.7339292   2.46878707  0.02445887

summary(fit12$analyses[[1]]) # single data: get the traditional summary

##
## Call:
## lm(formula = Time ~ Age + Extro)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.350  -7.782  -3.656   7.789  31.650
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.84188   37.93678  -0.549   0.5899
## Age          -0.05409    0.54524  -0.099   0.9221
## Extro         1.81191    0.73393   2.469   0.0245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.36 on 17 degrees of freedom
## Multiple R-squared:  0.2877, Adjusted R-squared:  0.2039
## F-statistic: 3.433 on 2 and 17 DF,  p-value: 0.05593
```

Table 11.4

```
library("mice")

imp2_WASH2 <- mice(WASH2, method = "pmm", m = 5, seed = 2018, print = TRUE)
```

```
##
## iter imp variable
## 1 1 Age Extro
## 1 2 Age Extro
## 1 3 Age Extro
## 1 4 Age Extro
## 1 5 Age Extro
## 2 1 Age Extro
## 2 2 Age Extro
## 2 3 Age Extro
## 2 4 Age Extro
## 2 5 Age Extro
## 3 1 Age Extro
## 3 2 Age Extro
## 3 3 Age Extro
## 3 4 Age Extro
## 3 5 Age Extro
## 4 1 Age Extro
## 4 2 Age Extro
## 4 3 Age Extro
## 4 4 Age Extro
## 4 5 Age Extro
## 5 1 Age Extro
## 5 2 Age Extro
## 5 3 Age Extro
## 5 4 Age Extro
## 5 5 Age Extro
```

```
fit22 <- with(imp2_WASH2,
              lm(Time ~ Age + Extro))
summary(pool(fit22))
```

```
##           estimate std.error statistic      df    p.value
## (Intercept) -13.050709 45.9645422  -0.2839299   4.682013 0.78167580
## Age         -0.303211  0.4478045  -0.6771057  11.131626 0.51216486
## Extro        1.795865  0.9101368   1.9731811   3.587528 0.07381966
```

Beat the Blues Revisited (Again)

Using the Beat the Blues data introduced in **Chapter 9** and revisited in **Chapter 10**:

```
BtB <- read.table("data/BtB.txt", header = TRUE, sep = '\t')

# to make sure that the factor levels are logical (esp. Treatment):
BtB <- within(BtB, {
  Drug <- factor(Drug, levels=c("No", "Yes")) # default
  Length <- factor(Length, levels=c("<6m", ">6m")) # default
  Treatment <- factor(Treatment, levels=c("TAU", "BtheB")) # NOT default!
})

glimpse(BtB); head(BtB); tail(BtB)

## Observations: 100
## Variables: 9
## $ Subject <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ Drug <fct> No, Yes, Yes, No, Yes, Yes, Yes, No, Yes, Yes, No, Y...
## $ Length <fct> >6m, >6m, <6m, >6m, >6m, <6m, <6m, >6m, <6m, >6m, >6...
## $ Treatment <fct> TAU, BtheB, TAU, BtheB, BtheB, BtheB, TAU, TAU, Bthe...
## $ BDIpre <int> 29, 32, 25, 21, 26, 7, 17, 20, 18, 20, 30, 49, 26, 3...
## $ BDI2m <int> 2, 16, 20, 17, 23, 0, 7, 20, 13, 5, 32, 35, 27, 26, ...
## $ BDI4m <int> 2, 24, NA, 16, NA, 0, 7, 21, 14, 5, 24, NA, 23, 36, ...
## $ BDI6m <int> NA, 17, NA, 10, NA, 0, 3, 19, 20, 8, 12, NA, NA, 27,...
## $ BDI8m <int> NA, 20, NA, 9, NA, 0, 7, 13, 11, 12, 2, NA, NA, 22, ...

## Subject Drug Length Treatment BDIpre BDI2m BDI4m BDI6m BDI8m
## 1 1 No >6m TAU 29 2 2 NA NA
## 2 2 Yes >6m BtheB 32 16 24 17 20
## 3 3 Yes <6m TAU 25 20 NA NA NA
## 4 4 No >6m BtheB 21 17 16 10 9
## 5 5 Yes >6m BtheB 26 23 NA NA NA
## 6 6 Yes <6m BtheB 7 0 0 0 0

## Subject Drug Length Treatment BDIpre BDI2m BDI4m BDI6m BDI8m
## 95 95 No >6m BtheB 16 11 4 2 3
## 96 96 Yes >6m BtheB 16 16 10 10 8
## 97 97 Yes <6m TAU 28 NA NA NA NA
## 98 98 No >6m BtheB 11 22 9 11 11
## 99 99 No <6m TAU 13 5 5 0 6
## 100 100 Yes <6m TAU 43 NA NA NA NA

# Convert data to long form for the analyses, adding Time:
BtBL <- gather(BtB, key = Visit, value = BDI, BDI2m, BDI4m, BDI6m, BDI8m) %>%
  mutate(Time = as.integer(substr(Visit, 4, 4)))

glimpse(BtBL); head(BtBL); tail(BtBL)

## Observations: 400
## Variables: 8
## $ Subject <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ Drug <fct> No, Yes, Yes, No, Yes, Yes, Yes, No, Yes, Yes, No, Y...
## $ Length <fct> >6m, >6m, <6m, >6m, >6m, <6m, <6m, >6m, <6m, >6m, >6...
## $ Treatment <fct> TAU, BtheB, TAU, BtheB, BtheB, BtheB, TAU, TAU, Bthe...
## $ BDIpre <int> 29, 32, 25, 21, 26, 7, 17, 20, 18, 20, 30, 49, 26, 3...
```

```
## $ Visit      <chr> "BDI2m", "BDI2m", "BDI2m", "BDI2m", "BDI2m", "BDI2m"...
## $ BDI        <int> 2, 16, 20, 17, 23, 0, 7, 20, 13, 5, 32, 35, 27, 26, ...
## $ Time       <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2...
```

```
## Subject Drug Length Treatment BDIpre Visit BDI Time
## 1      1   No   >6m      TAU      29 BDI2m   2    2
## 2      2  Yes   >6m     BtheB     32 BDI2m  16    2
## 3      3  Yes   <6m      TAU      25 BDI2m  20    2
## 4      4   No   >6m     BtheB     21 BDI2m  17    2
## 5      5  Yes   >6m     BtheB     26 BDI2m  23    2
## 6      6  Yes   <6m     BtheB      7 BDI2m   0    2
```

```
## Subject Drug Length Treatment BDIpre Visit BDI Time
## 395     95   No   >6m     BtheB     16 BDI8m   3    8
## 396     96  Yes   >6m     BtheB     16 BDI8m   8    8
## 397     97  Yes   <6m      TAU      28 BDI8m  NA    8
## 398     98   No   >6m     BtheB     11 BDI8m  11    8
## 399     99   No   <6m      TAU      13 BDI8m   6    8
## 400    100  Yes   <6m      TAU      43 BDI8m  NA    8
```

Table 11.5

```

library("mice")

imp1_BtB <- mice(BtBL, method = "mean", m = 1, maxit = 1, seed = 2019)

##
## iter imp variable
## 1 1 BDI

## Warning: Number of logged events: 1

fitOmean <- with(imp1_BtB,
                 lm(BDI ~ BDIPre + Time + Treatment + Drug + Length))
summary(fitOmean)

##           term      estimate std.error statistic      p.value
## 1 (Intercept) 10.4971960 1.41511372  7.417917 7.372448e-13
## 2      BDIPre  0.3563400 0.03866106  9.217025 1.859872e-18
## 3        Time -0.6810893 0.17778482 -3.830975 1.484779e-04
## 4 TreatmentBtheB -2.6111204 0.83967818 -3.109668 2.009121e-03
## 5      DrugYes -1.8077718 0.87095075 -2.075630 3.857633e-02
## 6    Length>6m  2.3140858 0.81990676  2.822377 5.008494e-03

summary(fitOmean$analyses[[1]]) # single data: get the traditional summary

##
## Call:
## lm(formula = BDI ~ BDIPre + Time + Treatment + Drug + Length)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.1132  -4.6752  -0.2941   4.0395  30.9403
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.49720     1.41511    7.418 7.37e-13 ***
## BDIPre          0.35634     0.03866    9.217 < 2e-16 ***
## Time           -0.68109     0.17778   -3.831 0.000148 ***
## TreatmentBtheB -2.61112     0.83968   -3.110 0.002009 **
## DrugYes        -1.80777     0.87095   -2.076 0.038576 *
## Length>6m       2.31409     0.81991    2.822 0.005008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.951 on 394 degrees of freedom
## Multiple R-squared:  0.269, Adjusted R-squared:  0.2597
## F-statistic:    29 on 5 and 394 DF, p-value: < 2.2e-16

```


Table 11.6

```
imp2_BtB <- mice(BtBL, method = "pmm", m = 5, seed = 2020, print = FALSE)
```

```
## Warning: Number of logged events: 1
```

```
fit0mi <- with(imp2_BtB,
               lm(BDI ~ BDIpre + Time + Treatment + Drug + Length))
summary(pool(fit0mi))
```

##	estimate	std.error	statistic	df	p.value
## (Intercept)	11.237577	2.05663684	5.464055	26.40342	1.236282e-07
## BDIpre	0.342330	0.04662927	7.341525	223.77554	3.877343e-12
## Time	-0.956200	0.21465531	-4.454584	219.49920	1.327789e-05
## TreatmentBtheB	-3.572554	1.52330496	-2.345265	10.57120	1.988848e-02
## DrugYes	-2.322066	1.20899720	-1.920654	36.67885	5.604625e-02
## Length>6m	4.033926	1.07151103	3.764708	65.89928	2.131144e-04

Table 11.7

```

# install.packages("lme4")
library("lme4")

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand

imp1_BtB <- mice(BtBL, method = "mean", m = 1, maxit = 1, seed = 2019)

##
## iter imp variable
##    1    1  BDI

## Warning: Number of logged events: 1

fitlmean <- with(imp1_BtB,
                 lmer(BDI ~ BDIPre + Time + Treatment + Drug + Length + (1 | Subject)))
summary(fitlmean)

##              term      estimate std.error statistic
## 1  (Intercept) 10.4971960 1.86230332  5.636674
## 2      BDIPre  0.3563400 0.06142892  5.800850
## 3      Time -0.6810893 0.12765705 -5.335305
## 4 TreatmentBtheB -2.6111204 1.33417245 -1.957109
## 5      DrugYes -1.8077718 1.38386172 -1.306324
## 6      Length>6m  2.3140858 1.30275745  1.776298

summary(fitlmean$analyses[[1]]) # single data: get the traditional summary

## Linear mixed model fit by REML ['lmerMod']
## Formula: BDI ~ BDIPre + Time + Treatment + Drug + Length + (1 | Subject)
##
## REML criterion at convergence: 2679.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6348 -0.5627 -0.0602  0.4319  4.2923
##
## Random effects:
##  Groups   Name      Variance Std.Dev.
##  Subject (Intercept) 31.75    5.635
##  Residual              32.59    5.709
## Number of obs: 400, groups: Subject, 100
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  10.49720    1.86230    5.637
## BDIPre        0.35634    0.06143    5.801
## Time         -0.68109    0.12766   -5.335
## TreatmentBtheB -2.61112    1.33417   -1.957

```

```

## DrugYes      -1.80777    1.38386   -1.306
## Length>6m    2.31409    1.30276    1.776
##
## Correlation of Fixed Effects:
##              (Intr) BDIPre Time   TrtmBB DrugYs
## BDIPre      -0.671
## Time        -0.343  0.000
## TretmntBthB -0.365  0.145  0.000
## DrugYes     -0.086 -0.246  0.000 -0.311
## Length>6m   -0.243 -0.201  0.000 -0.048  0.181

```

Table 11.8

```
imp2_BtB <- mice(BtBL, method = "pmm", m = 5, seed = 2020, print = FALSE)

## Warning: Number of logged events: 1

fit1mi <- with(imp2_BtB,
               lmer(BDI ~ BDIPre + Time + Treatment + Drug + Length + (1 | Subject)))
summary(pool(fit1mi))

##               estimate std.error statistic      df      p.value
## (Intercept)   11.237577  2.3320020   4.818854  41.55931 2.246766e-06
## BDIPre         0.342330  0.0636655   5.377009  315.61685 1.476294e-07
## Time          -0.956200  0.1828993  -5.228013  159.76439 3.123800e-07
## TreatmentBtheB -3.572554  1.7907569  -1.994997   19.70943 4.690280e-02
## DrugYes        -2.322066  1.5541164  -1.494139   84.84596 1.361384e-01
## Length>6m      4.033926  1.4118211   2.857250  145.24565 4.557036e-03
```