PSY 653 Module 1: Missing Data

Jan 29, 2020

Neil Yetz & Gemma Wallace

Part 1: In class Demo

Load Libraries

```
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3

## Warning: package 'tibble' was built under R version 3.6.3

## Warning: package 'tidyr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.3

## Warning: package 'forcats' was built under R version 3.6.3
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.6.3
```

```
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 3.6.3
```

Read in data

```
mice_data1 <- read_csv("mice_data1.csv")
```

Simple Linear regression model X1 ~ X2

Using pairwise deletion for missing data by default

```
mod1 <- lm(X1 ~ X2, data = mice_data1)
ols_regress(mod1)
```

```
##                         Model Summary
## --------------------------------------------------------------
## R                       0.026       RMSE              0.714
## R-Squared               0.001       Coef. Var        39.353
## Adj. R-Squared         -0.002       MSE               0.510
## Pred R-Squared         -0.009       MAE               0.592
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
```

```
##                          ANOVA
## ----------------------------------------------------------------------
##                 Sum of
##                Squares        DF     Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression       0.137         1          0.137      0.27      0.6039
## Residual       204.939       402          0.510
## Total          205.077       403
## ----------------------------------------------------------------------
##
##
##                              Parameter Estimates
## --------------------------------------------------------------------------------------
##         model      Beta    Std. Error    Std. Beta       t         Sig     lower    upper
## --------------------------------------------------------------------------------------
## (Intercept)       1.828        0.044                  41.152      0.000    1.741    1.916
##          X2      -0.011        0.022       -0.026     -0.519      0.604   -0.055    0.032
## --------------------------------------------------------------------------------------
```

Impute the dataset 5 times (using mice)

```
imputed_data <- mice(mice_data1, m=5, maxit = 50, method = 'pmm', seed = 500)
```

```
##
##  iter imp variable
##   1   1  X1
##   1   2  X1
##   1   3  X1
##   1   4  X1
##   1   5  X1
##   2   1  X1
##   2   2  X1
##   2   3  X1
##   2   4  X1
##   2   5  X1
##   3   1  X1
##   3   2  X1
##   3   3  X1
##   3   4  X1
##   3   5  X1
##   4   1  X1
##   4   2  X1
##   4   3  X1
##   4   4  X1
##   4   5  X1
##   5   1  X1
##   5   2  X1
##   5   3  X1
##   5   4  X1
##   5   5  X1
##   6   1  X1
##   6   2  X1
##   6   3  X1
##   6   4  X1
##   6   5  X1
```

```
##    7    1    X1
##    7    2    X1
##    7    3    X1
##    7    4    X1
##    7    5    X1
##    8    1    X1
##    8    2    X1
##    8    3    X1
##    8    4    X1
##    8    5    X1
##    9    1    X1
##    9    2    X1
##    9    3    X1
##    9    4    X1
##    9    5    X1
##   10    1    X1
##   10    2    X1
##   10    3    X1
##   10    4    X1
##   10    5    X1
##   11    1    X1
##   11    2    X1
##   11    3    X1
##   11    4    X1
##   11    5    X1
##   12    1    X1
##   12    2    X1
##   12    3    X1
##   12    4    X1
##   12    5    X1
##   13    1    X1
##   13    2    X1
##   13    3    X1
##   13    4    X1
##   13    5    X1
##   14    1    X1
##   14    2    X1
##   14    3    X1
##   14    4    X1
##   14    5    X1
##   15    1    X1
##   15    2    X1
##   15    3    X1
##   15    4    X1
##   15    5    X1
##   16    1    X1
##   16    2    X1
##   16    3    X1
##   16    4    X1
##   16    5    X1
##   17    1    X1
##   17    2    X1
##   17    3    X1
##   17    4    X1
```

```
##   17   5   X1
##   18   1   X1
##   18   2   X1
##   18   3   X1
##   18   4   X1
##   18   5   X1
##   19   1   X1
##   19   2   X1
##   19   3   X1
##   19   4   X1
##   19   5   X1
##   20   1   X1
##   20   2   X1
##   20   3   X1
##   20   4   X1
##   20   5   X1
##   21   1   X1
##   21   2   X1
##   21   3   X1
##   21   4   X1
##   21   5   X1
##   22   1   X1
##   22   2   X1
##   22   3   X1
##   22   4   X1
##   22   5   X1
##   23   1   X1
##   23   2   X1
##   23   3   X1
##   23   4   X1
##   23   5   X1
##   24   1   X1
##   24   2   X1
##   24   3   X1
##   24   4   X1
##   24   5   X1
##   25   1   X1
##   25   2   X1
##   25   3   X1
##   25   4   X1
##   25   5   X1
##   26   1   X1
##   26   2   X1
##   26   3   X1
##   26   4   X1
##   26   5   X1
##   27   1   X1
##   27   2   X1
##   27   3   X1
##   27   4   X1
##   27   5   X1
##   28   1   X1
##   28   2   X1
##   28   3   X1
```

```
## 28  4  X1
## 28  5  X1
## 29  1  X1
## 29  2  X1
## 29  3  X1
## 29  4  X1
## 29  5  X1
## 30  1  X1
## 30  2  X1
## 30  3  X1
## 30  4  X1
## 30  5  X1
## 31  1  X1
## 31  2  X1
## 31  3  X1
## 31  4  X1
## 31  5  X1
## 32  1  X1
## 32  2  X1
## 32  3  X1
## 32  4  X1
## 32  5  X1
## 33  1  X1
## 33  2  X1
## 33  3  X1
## 33  4  X1
## 33  5  X1
## 34  1  X1
## 34  2  X1
## 34  3  X1
## 34  4  X1
## 34  5  X1
## 35  1  X1
## 35  2  X1
## 35  3  X1
## 35  4  X1
## 35  5  X1
## 36  1  X1
## 36  2  X1
## 36  3  X1
## 36  4  X1
## 36  5  X1
## 37  1  X1
## 37  2  X1
## 37  3  X1
## 37  4  X1
## 37  5  X1
## 38  1  X1
## 38  2  X1
## 38  3  X1
## 38  4  X1
## 38  5  X1
## 39  1  X1
## 39  2  X1
```

```
## 39   3   X1
## 39   4   X1
## 39   5   X1
## 40   1   X1
## 40   2   X1
## 40   3   X1
## 40   4   X1
## 40   5   X1
## 41   1   X1
## 41   2   X1
## 41   3   X1
## 41   4   X1
## 41   5   X1
## 42   1   X1
## 42   2   X1
## 42   3   X1
## 42   4   X1
## 42   5   X1
## 43   1   X1
## 43   2   X1
## 43   3   X1
## 43   4   X1
## 43   5   X1
## 44   1   X1
## 44   2   X1
## 44   3   X1
## 44   4   X1
## 44   5   X1
## 45   1   X1
## 45   2   X1
## 45   3   X1
## 45   4   X1
## 45   5   X1
## 46   1   X1
## 46   2   X1
## 46   3   X1
## 46   4   X1
## 46   5   X1
## 47   1   X1
## 47   2   X1
## 47   3   X1
## 47   4   X1
## 47   5   X1
## 48   1   X1
## 48   2   X1
## 48   3   X1
## 48   4   X1
## 48   5   X1
## 49   1   X1
## 49   2   X1
## 49   3   X1
## 49   4   X1
## 49   5   X1
## 50   1   X1
```

```
##    50    2  X1
##    50    3  X1
##    50    4  X1
##    50    5  X1
```

Regress X1 on X2 on imputed dataset using the "with" function

```
fit.mi <- with(imputed_data, exp= lm(X1 ~ X2))
summary(fit.mi)
```

```
## # A tibble: 10 x 5
##    term         estimate std.error statistic   p.value
##    <chr>           <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)   1.84      0.0425   43.3      3.13e-160
##  2 X2           -0.0106     0.0207   -0.513    6.08e-  1
##  3 (Intercept)   1.82      0.0423   43.1      9.48e-160
##  4 X2           -0.00106    0.0205   -0.0516   9.59e-  1
##  5 (Intercept)   1.80      0.0429   42.0      1.46e-155
##  6 X2           -0.00443    0.0208   -0.213    8.32e-  1
##  7 (Intercept)   1.81      0.0419   43.3      3.27e-160
##  8 X2           -0.00242    0.0204   -0.119    9.05e-  1
##  9 (Intercept)   1.84      0.0419   43.9      2.47e-162
## 10 X2           -0.0119     0.0204   -0.584    5.60e-  1
```

Pool model estimates across imputed versions of the dataset

```
combimp <- pool(fit.mi)
summary(combimp)
```

```
##          term      estimate  std.error   statistic      df  p.value
## 1 (Intercept)  1.822457999  0.04641326  39.2658866 100.5742  0.00000
## 2          X2 -0.006076888  0.02123498  -0.2861734 289.4252  0.77495
```

Part 2: Try it yourself

Read in data

```
mice_data2 <- read_csv("mice_data2.csv")
```

```
## Warning: Duplicated column names deduplicated: 'X4' => 'X4_1' [5]
```

Multiple Linear regression model X1 ~ X2 + X3 + X4

Using pairwise deletion for missing data by default

```
mod2 <- lm(X1 ~ X2 + X3 + X4, data = mice_data2)
ols_regress(mod2)
```

```
##                       Model Summary
## -----------------------------------------------------------------
## R                      0.103       RMSE              0.712
## R-Squared              0.011       Coef. Var        39.254
## Adj. R-Squared         0.003       MSE               0.507
## Pred R-Squared        -0.008       MAE               0.588
```

```
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## -----------------------------------------------------------------
##               Sum of
##               Squares       DF     Mean Square      F        Sig.
## -----------------------------------------------------------------
## Regression     2.179        3         0.726       1.432     0.2330
## Residual     202.898      400         0.507
## Total        205.077      403
## -----------------------------------------------------------------
##
##                          Parameter Estimates
## --------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
## --------------------------------------------------------------------------
## (Intercept)     1.681      0.089                    18.832    0.000     1.505    1.856
##          X2    -0.017      0.022        -0.040      -0.789    0.431    -0.061    0.026
##          X3     0.079      0.040         0.101       1.990    0.047     0.001    0.157
##          X4    -0.001      0.024        -0.002      -0.036    0.971    -0.048    0.046
## --------------------------------------------------------------------------
```

Impute the dataset 5 times (using mice)

```
imputed_data <- mice(mice_data2, m=5, maxit = 50, method = 'pmm', seed = 500)
```

```
##
##  iter imp variable
##   1   1  X1
##   1   2  X1
##   1   3  X1
##   1   4  X1
##   1   5  X1
##   2   1  X1
##   2   2  X1
##   2   3  X1
##   2   4  X1
##   2   5  X1
##   3   1  X1
##   3   2  X1
##   3   3  X1
##   3   4  X1
##   3   5  X1
##   4   1  X1
##   4   2  X1
##   4   3  X1
##   4   4  X1
##   4   5  X1
##   5   1  X1
##   5   2  X1
##   5   3  X1
```

```
##    5    4    X1
##    5    5    X1
##    6    1    X1
##    6    2    X1
##    6    3    X1
##    6    4    X1
##    6    5    X1
##    7    1    X1
##    7    2    X1
##    7    3    X1
##    7    4    X1
##    7    5    X1
##    8    1    X1
##    8    2    X1
##    8    3    X1
##    8    4    X1
##    8    5    X1
##    9    1    X1
##    9    2    X1
##    9    3    X1
##    9    4    X1
##    9    5    X1
##   10    1    X1
##   10    2    X1
##   10    3    X1
##   10    4    X1
##   10    5    X1
##   11    1    X1
##   11    2    X1
##   11    3    X1
##   11    4    X1
##   11    5    X1
##   12    1    X1
##   12    2    X1
##   12    3    X1
##   12    4    X1
##   12    5    X1
##   13    1    X1
##   13    2    X1
##   13    3    X1
##   13    4    X1
##   13    5    X1
##   14    1    X1
##   14    2    X1
##   14    3    X1
##   14    4    X1
##   14    5    X1
##   15    1    X1
##   15    2    X1
##   15    3    X1
##   15    4    X1
##   15    5    X1
##   16    1    X1
##   16    2    X1
```

```
## 16   3  X1
## 16   4  X1
## 16   5  X1
## 17   1  X1
## 17   2  X1
## 17   3  X1
## 17   4  X1
## 17   5  X1
## 18   1  X1
## 18   2  X1
## 18   3  X1
## 18   4  X1
## 18   5  X1
## 19   1  X1
## 19   2  X1
## 19   3  X1
## 19   4  X1
## 19   5  X1
## 20   1  X1
## 20   2  X1
## 20   3  X1
## 20   4  X1
## 20   5  X1
## 21   1  X1
## 21   2  X1
## 21   3  X1
## 21   4  X1
## 21   5  X1
## 22   1  X1
## 22   2  X1
## 22   3  X1
## 22   4  X1
## 22   5  X1
## 23   1  X1
## 23   2  X1
## 23   3  X1
## 23   4  X1
## 23   5  X1
## 24   1  X1
## 24   2  X1
## 24   3  X1
## 24   4  X1
## 24   5  X1
## 25   1  X1
## 25   2  X1
## 25   3  X1
## 25   4  X1
## 25   5  X1
## 26   1  X1
## 26   2  X1
## 26   3  X1
## 26   4  X1
## 26   5  X1
## 27   1  X1
```

```
## 27   2   X1
## 27   3   X1
## 27   4   X1
## 27   5   X1
## 28   1   X1
## 28   2   X1
## 28   3   X1
## 28   4   X1
## 28   5   X1
## 29   1   X1
## 29   2   X1
## 29   3   X1
## 29   4   X1
## 29   5   X1
## 30   1   X1
## 30   2   X1
## 30   3   X1
## 30   4   X1
## 30   5   X1
## 31   1   X1
## 31   2   X1
## 31   3   X1
## 31   4   X1
## 31   5   X1
## 32   1   X1
## 32   2   X1
## 32   3   X1
## 32   4   X1
## 32   5   X1
## 33   1   X1
## 33   2   X1
## 33   3   X1
## 33   4   X1
## 33   5   X1
## 34   1   X1
## 34   2   X1
## 34   3   X1
## 34   4   X1
## 34   5   X1
## 35   1   X1
## 35   2   X1
## 35   3   X1
## 35   4   X1
## 35   5   X1
## 36   1   X1
## 36   2   X1
## 36   3   X1
## 36   4   X1
## 36   5   X1
## 37   1   X1
## 37   2   X1
## 37   3   X1
## 37   4   X1
## 37   5   X1
```

```
##   38   1   X1
##   38   2   X1
##   38   3   X1
##   38   4   X1
##   38   5   X1
##   39   1   X1
##   39   2   X1
##   39   3   X1
##   39   4   X1
##   39   5   X1
##   40   1   X1
##   40   2   X1
##   40   3   X1
##   40   4   X1
##   40   5   X1
##   41   1   X1
##   41   2   X1
##   41   3   X1
##   41   4   X1
##   41   5   X1
##   42   1   X1
##   42   2   X1
##   42   3   X1
##   42   4   X1
##   42   5   X1
##   43   1   X1
##   43   2   X1
##   43   3   X1
##   43   4   X1
##   43   5   X1
##   44   1   X1
##   44   2   X1
##   44   3   X1
##   44   4   X1
##   44   5   X1
##   45   1   X1
##   45   2   X1
##   45   3   X1
##   45   4   X1
##   45   5   X1
##   46   1   X1
##   46   2   X1
##   46   3   X1
##   46   4   X1
##   46   5   X1
##   47   1   X1
##   47   2   X1
##   47   3   X1
##   47   4   X1
##   47   5   X1
##   48   1   X1
##   48   2   X1
##   48   3   X1
##   48   4   X1
```

```
##   48   5  X1
##   49   1  X1
##   49   2  X1
##   49   3  X1
##   49   4  X1
##   49   5  X1
##   50   1  X1
##   50   2  X1
##   50   3  X1
##   50   4  X1
##   50   5  X1
```

Regress X1 on X2 on imputed dataset using the "with" function

```
fit.mi2 <- with(imputed_data, exp= lm(X1 ~ X2 + X3 + X4))
summary(fit.mi2)
```

```
## # A tibble: 20 x 5
##    term          estimate std.error statistic   p.value
##    <chr>            <dbl>     <dbl>     <dbl>     <dbl>
##  1 (Intercept)   1.68       0.0874  19.2       4.08e-60
##  2 X2           -0.0263     0.0211  -1.25       2.13e- 1
##  3 X3            0.0871     0.0392   2.22       2.67e- 2
##  4 X4            0.000103   0.0238   0.00435    9.97e- 1
##  5 (Intercept)   1.70       0.0862  19.7       3.09e-62
##  6 X2           -0.00724    0.0208  -0.348      7.28e- 1
##  7 X3            0.0709     0.0386   1.84       6.71e- 2
##  8 X4           -0.000183   0.0234  -0.00783    9.94e- 1
##  9 (Intercept)   1.68       0.0849  19.8       1.31e-62
## 10 X2           -0.0261     0.0205  -1.28       2.03e- 1
## 11 X3            0.0838     0.0380   2.20       2.80e- 2
## 12 X4            0.000285   0.0231   0.0124     9.90e- 1
## 13 (Intercept)   1.69       0.0867  19.5       2.79e-61
## 14 X2           -0.0268     0.0209  -1.28       2.00e- 1
## 15 X3            0.0791     0.0389   2.03       4.25e- 2
## 16 X4           -0.00353    0.0236  -0.150      8.81e- 1
## 17 (Intercept)   1.64       0.0867  18.9       1.59e-58
## 18 X2            0.00604    0.0209   0.288      7.73e- 1
## 19 X3            0.0795     0.0389   2.05       4.14e- 2
## 20 X4            0.00378    0.0236   0.160      8.73e- 1
```

Pool model estimates across imputed versions of the dataset

```
combimp2 <- pool(fit.mi2)
summary(combimp2)
```

```
##          term      estimate  std.error    statistic       df   p.value
## 1 (Intercept)  1.676699e+00 0.09002441 18.624934170 245.66880 0.0000000
## 2          X2 -1.610209e-02 0.02649093 -0.607834073  25.07851 0.5487629
## 3          X3  8.010165e-02 0.03929906  2.038258926 387.12469 0.0422032
## 4          X4  9.198331e-05 0.02364814  0.003889664 418.47772 0.9968984
```