Applied Statistical Methods - Solution 9

AUTHOR
Peter von Rohr

PUBLISHED
May 6, 2024

WEBR STATUS
Ready!

Problem 1: Repeated Observations in Predictors

Use the dataset on blood presure and pulse frequency and fit linear regression models with response variables `SYS`, `DIA` and `PUL` on the mean of the manually measured pulse frequencies. Check the distribution of the residuals with a dot-plot of the resiudals of the regression model on the fitted values. For which of the responses (`SYS`, `DIA` and `PUL`), the highest absolute regression coefficient can be found? The dataset is available at

```
[1] "https://charlotte-ngs.github.io/asmasss2024/data/20240429_bp_data.csv"
```

Tasks

- Read the data

Run Code

```
1  # read data to a data.frame
2  s_tbl_ex09_p01_path <- "https://charlotte-ngs.github.io/asmasss2024/
3  df_bp <- read.table(s_tbl_ex09_p01_path, header = T, sep = ",")
4  df_bp
```

	Nr	SYS	DIA	PUL	MP_Rep	ManualPulseFreq
1	1	150	95	92	MP1	104
2	1	150	95	92	MP2	92
3	1	150	95	92	MP3	84
4	2	135	89	70	MP1	72
5	2	135	89	70	MP2	70
6	2	135	89	70	MP3	76
7	3	121	85	69	MP1	66
8	3	121	85	69	MP2	72
9	3	121	85	69	MP3	70
10	4	122	82	65	MP1	62
11	4	122	82	65	MP2	68
12	4	122	82	65	MP3	66
13	5	113	79	76	MP1	85
14	5	113	79	76	MP2	84
15	5	113	79	76	MP3	91
16	6	139	83	64	MP1	56
17	6	139	83	64	MP2	72
18	6	139	83	64	MP3	64
19	7	116	75	72	MP1	60
20	7	116	75	72	MP2	64
21	7	116	75	72	MP3	68
22	8	137	85	93	MP1	76
23	8	137	85	93	MP2	84
24	8	137	85	93	MP3	80
25	9	133	83	90	MP1	64
26	9	133	83	90	MP2	68
27	9	133	83	90	MP3	68
28	10	129	83	65	MP1	60
29	10	129	83	65	MP2	68
30	10	129	83	65	MP3	68
31	11	105	72	64	MP1	72

```
32 11 105  72  64    MP2             80
33 11 105  72  64    MP3             68
```

- Compute mean of manual pulse frequencies

Run Code

```r
# compute mean values accross individuals
library(dplyr)
tbl_bp_mpf_mean <- df_bp %>%
  group_by(Nr) %>%
  summarise(mean_sys = mean(SYS),
            mean_dia = mean(DIA),
            mean_pul = mean(PUL),
            mean_mpf = mean(ManualPulseFreq))
tbl_bp_mpf_mean
```

```
# A tibble: 11 × 5
      Nr mean_sys mean_dia mean_pul mean_mpf
   <int>    <dbl>    <dbl>    <dbl>    <dbl>
 1     1      150       95       92     93.3
 2     2      135       89       70     72.7
 3     3      121       85       69     69.3
 4     4      122       82       65     65.3
 5     5      113       79       76     86.7
 6     6      139       83       64     64
 7     7      116       75       72     64
 8     8      137       85       93     80
 9     9      133       83       90     66.7
10    10      129       83       65     65.3
11    11      105       72       64     73.3
```

- Fit a regression model of `SYS` on mean manual pulse frequency

Run Code

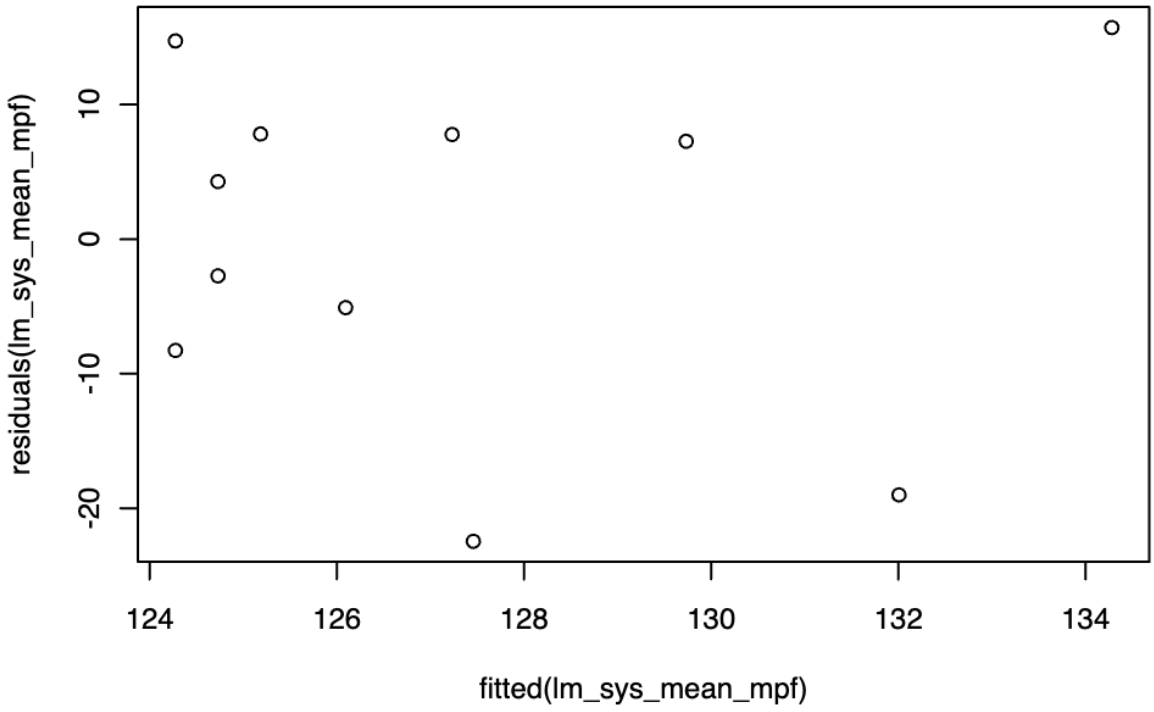
```r
# fit regression model of SYS on mean_mpf
lm_sys_mean_mpf <- lm(mean_sys ~ mean_mpf, data = tbl_bp_mpf_mean)
(smry_sys_mean_mpf <- summary(lm_sys_mean_mpf))
```

```
Call:
lm(formula = mean_sys ~ mean_mpf, data = tbl_bp_mpf_mean)

Residuals:
    Min      1Q  Median      3Q     Max
-22.459  -6.685   4.270   7.792  15.720

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 102.4456    31.4616   3.256   0.0099 **
mean_mpf      0.3411     0.4286   0.796   0.4467
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.43 on 9 degrees of freedom
Multiple R-squared:  0.06573,   Adjusted R-squared:  -0.03808
F-statistic: 0.6332 on 1 and 9 DF,  p-value: 0.4467
```

- Plot residuals

Run Code

```
# plot fitted values versus residuals
plot(fitted(lm_sys_mean_mpf), residuals(lm_sys_mean_mpf))
```



- Fit a regression model of `DIA` on mean manual pulse frequency

Run Code

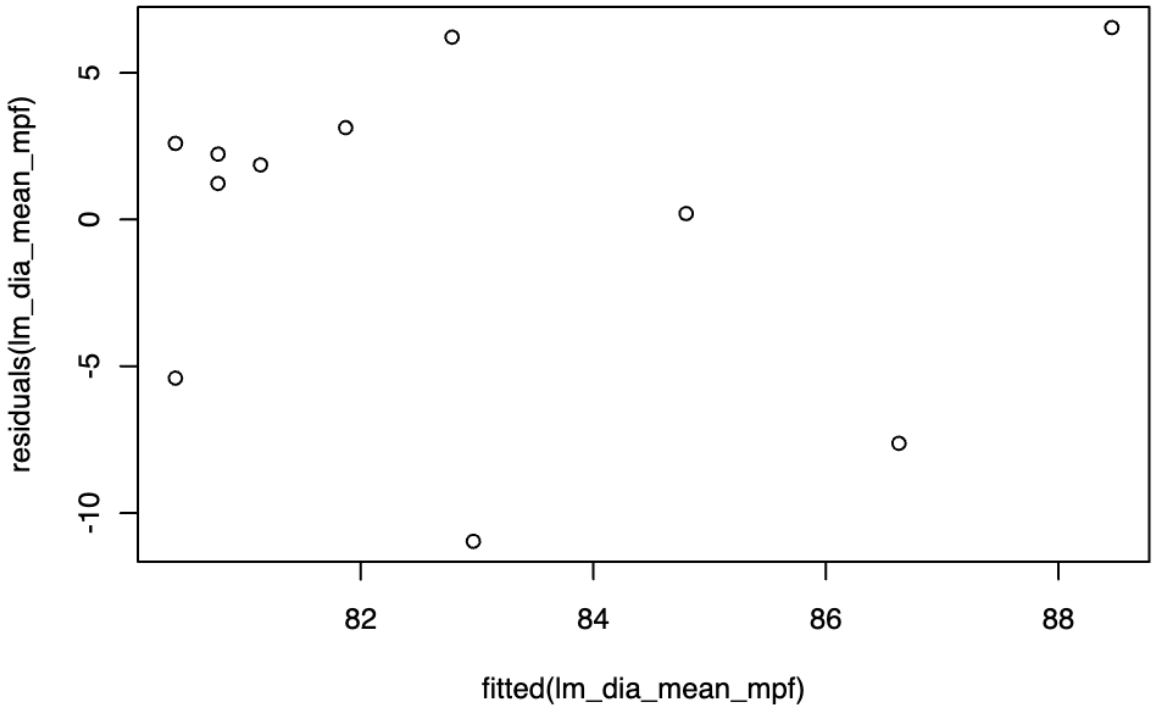
```
# fit regression model of DIA on mean_mpf
lm_dia_mean_mpf <- lm(mean_dia ~ mean_mpf, data = tbl_bp_mpf_mean)
(smry_dia_mean_mpf <- summary(lm_dia_mean_mpf))
```

```
Call:
lm(formula = mean_dia ~ mean_mpf, data = tbl_bp_mpf_mean)

Residuals:
    Min      1Q  Median      3Q     Max
-10.968  -2.602   1.862   2.862   6.542

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  62.8379    13.8705   4.530  0.00143 **
mean_mpf      0.2745     0.1890   1.453  0.18030
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.922 on 9 degrees of freedom
Multiple R-squared:  0.1899,    Adjusted R-squared:  0.09991
F-statistic:  2.11 on 1 and 9 DF,  p-value: 0.1803
```

- Plot residuals

Run Code

```
# plot fitted values versus residuals
plot(fitted(lm_dia_mean_mpf), residuals(lm_dia_mean_mpf))
```



- Fit a regression model of `PUL` on mean manual pulse frequency

Run Code

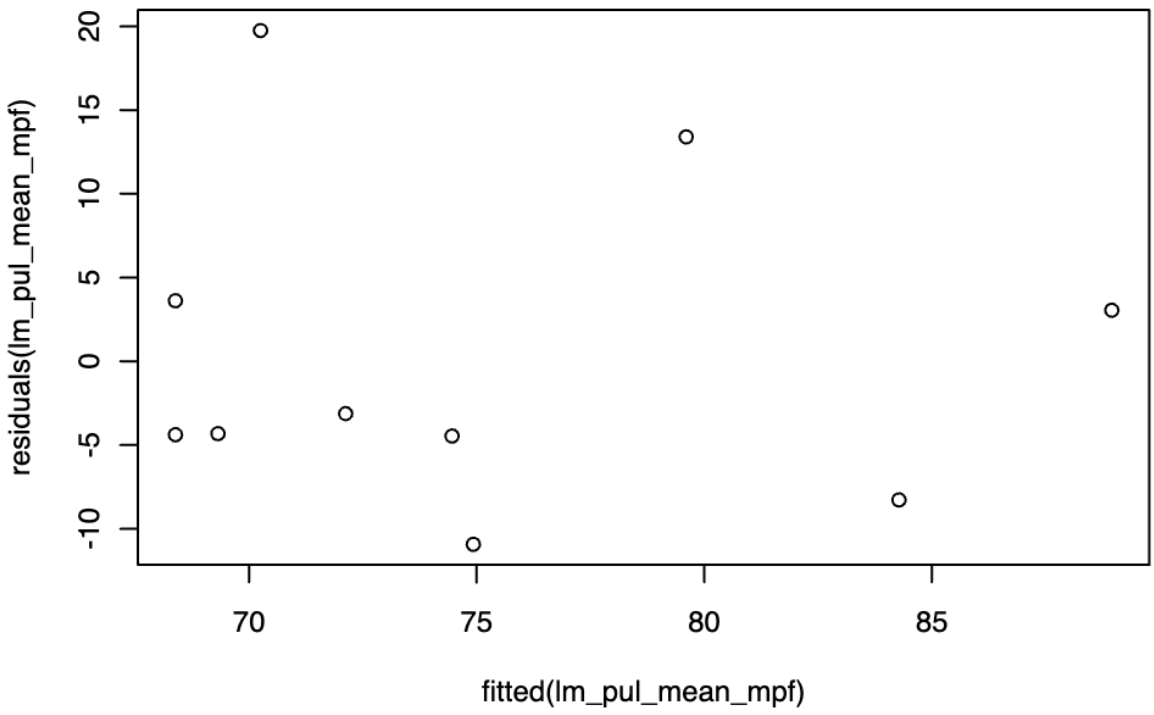
```
# fit regression model of PUL on mean_mpf
lm_pul_mean_mpf <- lm(mean_pul ~ mean_mpf, data = tbl_bp_mpf_mean)
(smry_pul_mean_mpf <- summary(lm_pul_mean_mpf))
```

```
Call:
lm(formula = mean_pul ~ mean_mpf, data = tbl_bp_mpf_mean)

Residuals:
    Min      1Q  Median      3Q     Max
-10.928  -4.422  -4.318   3.332  19.747

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.5077    22.9735   1.023   0.3329
mean_mpf      0.7012     0.3130   2.240   0.0518 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.809 on 9 degrees of freedom
Multiple R-squared:  0.358,     Adjusted R-squared:  0.2867
F-statistic: 5.019 on 1 and 9 DF,  p-value: 0.05182
```

- Plot residuals

Run Code

```
# plot fitted values versus residuals
plot(fitted(lm_pul_mean_mpf), residuals(lm_pul_mean_mpf))
```



- Ranking of the regression coefficients

Run Code

```
# rank estimated regression coefficient
order(c(smry_sys_mean_mpf$coefficients["mean_mpf", "Estimate"],
        smry_dia_mean_mpf$coefficients["mean_mpf", "Estimate"],
        smry_pul_mean_mpf$coefficients["mean_mpf", "Estimate"]))
```

```
[1] 2 1 3
```

Problem 2: Median Predictor Variables

Fit the same type of regression models, but use the median of the predictor variables instead of the mean.

Tasks

- Read the data

Run Code

```
# read data to a data.frame
s_tbl_ex09_p02_path <- "https://charlotte-ngs.github.io/asmasss2024/
df_bp <- read.table(s_tbl_ex09_p02_path, header = T, sep = ",")
df_bp
```

```
   Nr SYS DIA PUL MP_Rep ManualPulseFreq
1   1 150  95  92    MP1             104
2   1 150  95  92    MP2              92
3   1 150  95  92    MP3              84
4   2 135  89  70    MP1              72
5   2 135  89  70    MP2              70
6   2 135  89  70    MP3              76
7   3 121  85  69    MP1              66
8   3 121  85  69    MP2              72
9   3 121  85  69    MP3              70
10  4 122  82  65    MP1              62
11  4 122  82  65    MP2              68
12  4 122  82  65    MP3              66
13  5 113  79  76    MP1              85
14  5 113  79  76    MP2              84
15  5 113  79  76    MP3              91
16  6 139  83  64    MP1              56
17  6 139  83  64    MP2              72
18  6 139  83  64    MP3              64
19  7 116  75  72    MP1              60
20  7 116  75  72    MP2              64
21  7 116  75  72    MP3              68
22  8 137  85  93    MP1              76
23  8 137  85  93    MP2              84
24  8 137  85  93    MP3              80
25  9 133  83  90    MP1              64
26  9 133  83  90    MP2              68
27  9 133  83  90    MP3              68
28 10 129  83  65    MP1              60
29 10 129  83  65    MP2              68
30 10 129  83  65    MP3              68
31 11 105  72  64    MP1              72
32 11 105  72  64    MP2              80
33 11 105  72  64    MP3              68
```

- Compute median of manual pulse frequencies

Run Code

```
# compute median values accross individuals
library(dplyr)
tbl_bp_mpf_median <- df_bp %>%
  group_by(Nr) %>%
  summarise(median_sys = median(SYS),
            median_dia = median(DIA),
            median_pul = median(PUL),
            median_mpf = median(ManualPulseFreq))
tbl_bp_mpf_median
```

```
# A tibble: 11 × 5
      Nr median_sys median_dia median_pul median_mpf
   <int>      <int>      <int>      <int>      <int>
 1     1        150         95         92         92
 2     2        135         89         70         72
 3     3        121         85         69         70
 4     4        122         82         65         66
 5     5        113         79         76         85
 6     6        139         83         64         64
 7     7        116         75         72         64
 8     8        137         85         93         80
 9     9        133         83         90         68
```

```
10    10        129         83         65         68
11    11        105         72         64         72
```

- Fit a regression model of `SYS` on median manual pulse frequency

Run Code

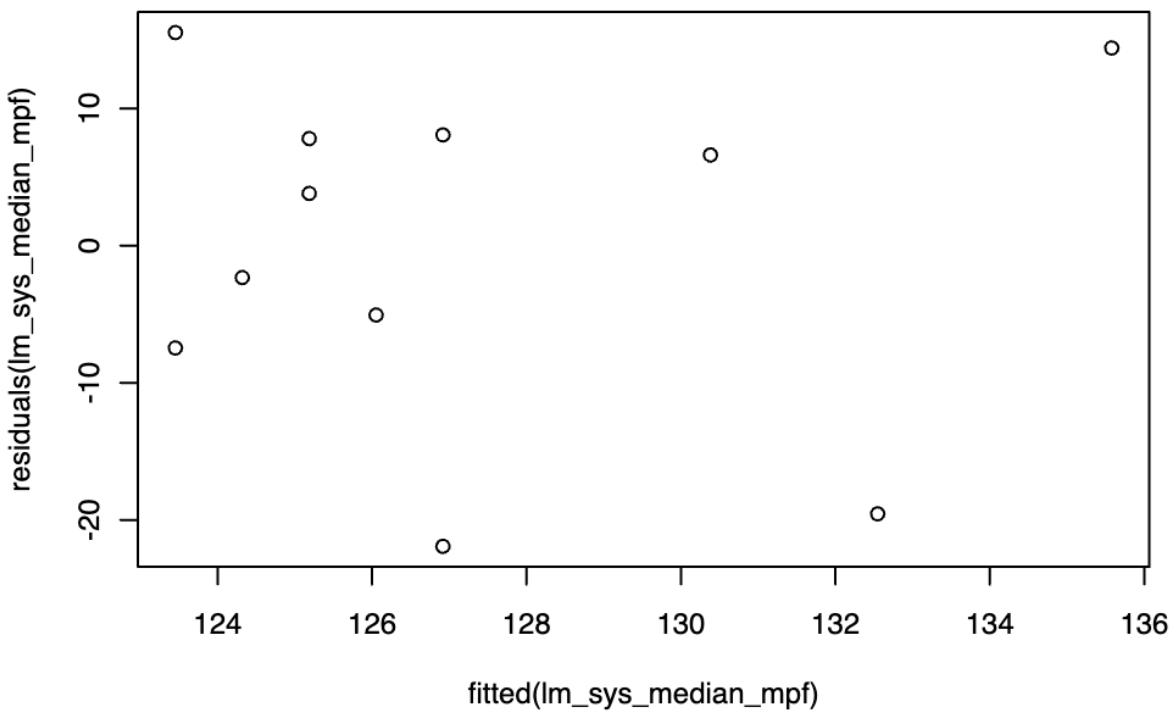
```
1  # fit regression model of SYS on median_mpf
2  lm_sys_median_mpf <- lm(median_sys ~ median_mpf, data = tbl_bp_mpf_m
3  (smry_sys_median_mpf <- summary(lm_sys_median_mpf))
```

```
Call:
lm(formula = median_sys ~ median_mpf, data = tbl_bp_mpf_median)

Residuals:
    Min      1Q  Median      3Q     Max
-21.918  -6.253   3.814   7.948  15.546

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  95.7385    33.8504   2.828   0.0198 *
median_mpf    0.4331     0.4616   0.938   0.3727
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.26 on 9 degrees of freedom
Multiple R-squared:  0.08908,   Adjusted R-squared:  -0.01213
F-statistic: 0.8801 on 1 and 9 DF,  p-value: 0.3727
```

- Plot residuals

Run Code

```
1  # plot fitted values versus residuals
2  plot(fitted(lm_sys_median_mpf), residuals(lm_sys_median_mpf))
```



- Fit a regression model of `DIA` on median manual pulse frequency

Run Code

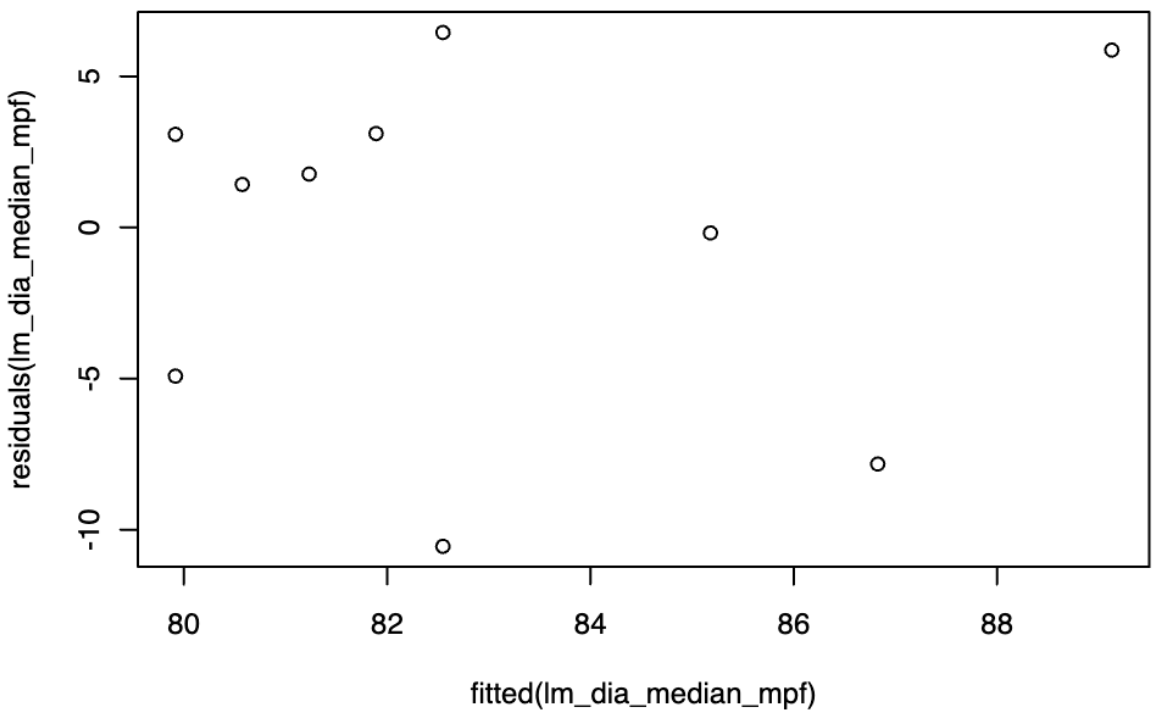
```
# fit regression model of DIA on median_mpf
lm_dia_median_mpf <- lm(median_dia ~ median_mpf, data = tbl_bp_mpf_m
(smry_dia_median_mpf <- summary(lm_dia_median_mpf))
```

```
Call:
lm(formula = median_dia ~ median_mpf, data = tbl_bp_mpf_median)

Residuals:
    Min      1Q  Median      3Q     Max
-10.549  -2.549   1.767   3.096   6.451

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   58.861     14.741   3.993  0.00314 **
median_mpf     0.329      0.201   1.637  0.13613
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.776 on 9 degrees of freedom
Multiple R-squared:  0.2294,    Adjusted R-squared:  0.1437
F-statistic: 2.679 on 1 and 9 DF,  p-value: 0.1361
```

- Plot residuals

Run Code

```
# plot fitted values versus residuals
plot(fitted(lm_dia_median_mpf), residuals(lm_dia_median_mpf))
```



- Fit a regression model of `PUL` on median manual pulse frequency

Run Code

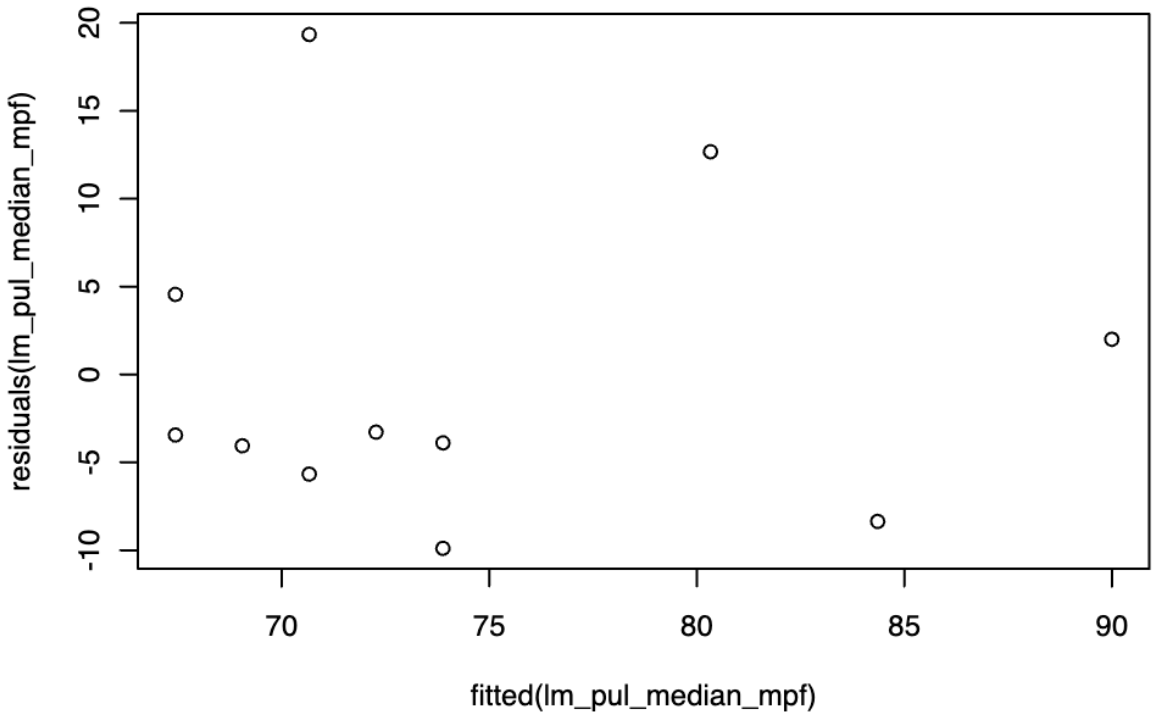
```
# fit regression model of PUL on median_mpf
lm_pul_median_mpf <- lm(median_pul ~ median_mpf, data = tbl_bp_mpf_m
(smry_pul_median_mpf <- summary(lm_pul_median_mpf))
```

```
Call:
lm(formula = median_pul ~ median_mpf, data = tbl_bp_mpf_median)

Residuals:
   Min     1Q Median     3Q    Max
-9.886 -4.859 -3.442  3.280 19.336

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.8868    24.2556   0.655   0.5289
median_mpf    0.8055     0.3308   2.435   0.0376 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.504 on 9 degrees of freedom
Multiple R-squared:  0.3972,    Adjusted R-squared:  0.3303
F-statistic: 5.931 on 1 and 9 DF,  p-value: 0.03765
```

- Plot residuals

Run Code

```
# plot fitted values versus residuals
plot(fitted(lm_pul_median_mpf), residuals(lm_pul_median_mpf))
```



- Ranking of the regression coefficients

Run Code

```
# rank estimated regression coefficient
order(c(smry_sys_median_mpf$coefficients["median_mpf", "Estimate"],
        smry_dia_median_mpf$coefficients["median_mpf", "Estimate"],
        smry_pul_median_mpf$coefficients["median_mpf", "Estimate"]))
```

```
[1] 2 1 3
```