Untitled

2024-11-07

undergraduate extra credit (5) (Use R) This problem is required for Grad students (those of you registered in STAT 505) but can be extra credit for undergrad students. A researcher needs your expertise to combine the so called SRH (Health rate) before and after a treatment. A random sample of size n = 2000 individuals were selected. A random sample of 1000 individuals were then subjected to the treatment and another 1000 to the control group. In the first data file called "SRH", their status pre and post experiment are reported in. The weights of the treatment and control group were then record in the files called "weight_treatment" and "weight_control." Combine the 3 data files properly to create an analytic dataset as shown in the sample below: Remember to sort the id in increasing order and sort the time in PRE and POST order. Provide code for how you arrived at your analytic data set.

```
setwd("~/Desktop/Stat_405_R/Stat 405 M10 HW")
SRH <- as.data.frame(read.csv(file = "SRH(1).csv", header = TRUE))
weight_control <- as.data.frame(read.csv(file = "weight_control(1).csv", header = TRUE))
weight_treatment <- as.data.frame(read.csv(file = "weight_treatment(1).csv", header = TRUE))
```

```
(head(SRH,5))
```

```
##   id trt TIME          SRH
## 1  1   1 POST         Poor
## 2  1   1  PRE         Good
## 3  2   1  PRE         Poor
## 4  2   1 POST    Very Poor
## 5  3   1  PRE Satisfactory
```

```
(head(weight_control,5))
```

```
##   obs_id PRE_WEIGHT POST_WEIGHT
## 1   2501   159.7587          NA
## 2   2501         NA     158.692
## 3   2502   176.1611          NA
## 4   2502         NA     174.827
## 5   2503   181.3907          NA
```

```
(head(weight_treatment,5))
```

```
##   Id PRE_WEIGHT POST_WEIGHT
## 1  1   135.2510          NA
## 2  1         NA    125.6678
## 3  2   154.8713          NA
## 4  2         NA    153.9882
## 5  3   128.1951          NA
```

```r
SRH_id_sorted <- SRH[order(SRH[,1], SRH[,3] == "POST"), ]
WC_id_sorted <- weight_control[order(weight_control[,1]),]
WT_id_sorted <- weight_treatment[order(weight_treatment[,1]),]

Pre_WT <- na.omit(WT_id_sorted[,1:2])
Post_WT <- na.omit(WT_id_sorted[,c(1,3)])
Pre_WC <- na.omit(WC_id_sorted[,1:2])
Post_WC <- na.omit(WC_id_sorted[,c(1,3)])

Append <- numeric(nrow(SRH_id_sorted))

Final <- cbind(SRH_id_sorted,Append)

Final[Final[,3] == "PRE",5] <- Pre_WT[,2]
Final[Final[,3] == "POST",5] <- Post_WT[,2]

Pre_WC_id_nums <- Pre_WC[,1]
Post_WC_id_nums <- Post_WC[,1]

Final[which(Final[,1] %in% Pre_WC_id_nums & Final[,3] %in% "PRE"),5] <- Pre_WC[,2]
Final[which(Final[,1] %in% Post_WC_id_nums & Final[,3] %in% "POST"),5] <- Post_WC[,2]

trt <- Final[,2]
time <- Final[,3]
Final[,2] <- time
Final[,3] <- trt

colnames(Final) <- c("id","time","trt","SRH","weight")
print(head(Final,10))
```

```
##    id time trt          SRH   weight
## 2   1  PRE   1         Good 135.2510
## 1   1 POST   1         Poor 125.6678
## 3   2  PRE   1         Poor 154.8713
## 4   2 POST   1    Very Poor 153.9882
## 5   3  PRE   1 Satisfactory 128.1951
## 6   3 POST   1         Good 115.5969
## 7   4  PRE   1         Poor 183.4600
## 8   4 POST   1         Good 177.0187
## 10  5  PRE   1         Poor 166.3726
## 9   5 POST   1         Poor 163.9262
```

```r
write.csv(Final, "analytic_data_set.csv", row.names = FALSE)
```