Assignment II

Group2

2024-12-05

Introduction

Question 1:

Create table 1. Describe all available variables in the data. Show both, the original data and the imputed data.

Overview :

All available variables in the data are as follows:

- *dibpat0f:* Dichotomous behavior pattern factor with levels A and B instead of 1 and 0.
- *agegroup:* Categories for age.
- *smoker:* Binary variable for smoker or not.
- *smokerf:* Smoker factor with levels No and Yes.
- *heightcm:* Convert height from inches to cm.
- *weightkg:* Convert weight from pounds to kg.
- *bmi:* Calculate BMI.
- *bmicat:* Categories for BMI.
- *cholmmol:* Convert cholesterol from mg/dl to mmol/l.
- *sbp10:* Categories of sbp (systolic blood pressure).
- *sbpcat:* Systolic blood pressure factor.
- *chd69f:* Coronary heart disease factor.

Then, we have created tables for the original data. In the next step, we have imputed the data using Multivariate Imputation and created tables for the imputed data.

Code :

```
# Define variables
variables <- c("id", "agegroup", "age0", "cholmmol", "sbp10", "bmi", "smokerf",
               "arcus0", "dibpat0f", "chd69")
categorical <- c("smokerf", "dibpat0f", "chd69")

# Create Table 1 for the original data
table_original <- CreateTableOne(vars = variables, data = wcgs, factorVars = categorical)

# Create Table 1 for the imputed data
table_imputed <- CreateTableOne(vars = variables, data = di, factorVars = categorical)
```

Output :

```
##
##      ### Summary of continuous variables ###
##
## strata: Overall
##             n miss p.miss  mean    sd median  p25   p75  min   max skew kurt
## id       3154    0   0.00 1e+04 6e+03  11406 3741 13115 2001 22101  0.2 -0.7
## age0     3154    0   0.00 5e+01 6e+00     45   42    50   39    59  0.5 -0.8
## cholmmol 3154   13   0.41 6e+00 1e+00      6    5     6    3    11  0.4  0.5
## sbp10    3154    0   0.00 1e+01 2e+00     13   12    14   10    23  1.2  2.8
## bmi      3154    0   0.00 2e+01 3e+00     24   23    26   11    39  0.5  2.0
## arcus0   3154    2   0.06 3e-01 5e-01      0    0     1    0     1  0.9 -1.2
##
## ============================================================================================
##
##      ### Summary of categorical variables ###
##
## strata: Overall
##       var    n miss p.miss   level freq percent cum.percent
##  agegroup 3154    0    0.0 [39,45) 1448    45.9        45.9
##                            [45,55) 1384    43.9        89.8
##                            [55,60]  322    10.2       100.0
##
##   smokerf 3154    0    0.0      No 1652    52.4        52.4
##                                Yes 1502    47.6       100.0
##
##  dibpat0f 3154    0    0.0       B 1565    49.6        49.6
##                                 A 1589    50.4       100.0
##
##     chd69 3154    0    0.0       0 2897    91.9        91.9
##                                 1  257     8.1       100.0
##

##
##      ### Summary of continuous variables ###
##
## strata: Overall
##             n miss p.miss  mean    sd median  p25   p75  min   max skew kurt
## id       3154    0      0 1e+04 6e+03  11406 3741 13115 2001 22101  0.2 -0.7
## age0     3154    0      0 5e+01 6e+00     45   42    50   39    59  0.5 -0.8
## cholmmol 3154    0      0 6e+00 1e+00      6    5     6    3    11  0.4  0.5
## sbp10    3154    0      0 1e+01 2e+00     13   12    14   10    23  1.2  2.8
## bmi      3154    0      0 2e+01 3e+00     24   23    26   11    39  0.5  2.0
## arcus0   3154    0      0 3e-01 5e-01      0    0     1    0     1  0.9 -1.2
##
## ============================================================================================
##
##      ### Summary of categorical variables ###
##
## strata: Overall
##       var    n miss p.miss   level freq percent cum.percent
##  agegroup 3154    0    0.0 [39,45) 1448    45.9        45.9
##                            [45,55) 1384    43.9        89.8
##                            [55,60]  322    10.2       100.0
```

```
##
##   smokerf 3154    0    0.0      No 1652    52.4        52.4
##                                Yes 1502    47.6       100.0
##
##  dibpat0f 3154    0    0.0       B 1565    49.6        49.6
##                                 A 1589    50.4       100.0
##
##     chd69 3154    0    0.0       0 2897    91.9        91.9
##                                 1  257     8.1       100.0
##
```

Conclusion :

The imputed data has been created using Multivariate Imputation where the missing data of cholmmol has been imputed. The imputed data has been created using Predictive Mean Matching (PMM) method.

Question 2

Calculate the overall risk of CHD in the cohort.

Overview :

a. What is the outcome we are interested in?

The outcome we are interested in is Coronary Heart Disease (CHD).

b. What are the known risk factors for our outcome of interest?

The known risk factors for Coronary Heart Disease (CHD) are as follows:

- Behaviour type A/B
- Age
- Cholesterol
- Systolic Blood Pressure
- BMI
- Smoking
- Corneal arcus

c. How many persons are included?

3154 middle-aged men, from 39 to 59 years of age, during the years 1960-1961 are included in this prospective cohort study.

d. What is the overall risk or rate and prevalence of the disease in our cohort?

The overall risk or rate and prevalence of the disease in our cohort is as follows:

```r
# Overall risk or rate
overall_rate <- table(di$chd69)

#calculate risk of CHD
overall_risk <- overall_rate / sum(overall_rate)

# extract the rate and risk into a data frame
chd_frame <- data.frame(
  "CHD Presence" = c("No", "Yes"),
  "Overall Rate" = c(as.matrix(overall_rate)[1], as.matrix(overall_rate)[2]),
  "Overall Risk" = c(as.matrix(overall_risk)[1], as.matrix(overall_risk)[2])
```

```
)
#print overall rate and risk into a table
knitr::kable(chd_frame, col.names = c("CHD Presence", "Overall Rate", "Overall Risk"))
```

CHD Presence	Overall Rate	Overall Risk
No	2897	0.9185162
Yes	257	0.0814838

Analysis :

The overall risk of Coronary Heart Disease (CHD) in the cohort is 0.08, which indicates that prevalence of the disease is 8% in the cohort.

Question 3

Overview :

Code :

Output :

Analysis :

Question 4

Overview :

Code :

Output :

Analysis :

Question 5

Overview :

Code :

Output :

Analysis :

Question 6

Overview :

Code :

Output :

Analysis :

Question 7

Overview :

Code :

Output :

Analysis :

Question 8

Overview :

Code :

Output :

Analysis :

Question 9

Overview :

Code :