Blood Pressure

Tarek El-Hajjaoui

2023-03-05

Data loading & pre-processing

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/BloodPressure/Blood
df = read.table(file_path, header=TRUE, sep="", dec=".")
```

Summary of dataset

```
str(df)
```

```
## 'data.frame':    500 obs. of  3 variables:
##  $ SystolicBP: int  133 115 140 132 133 138 133 67 138 130 ...
##  $ Smoke     : int  0 1 1 0 0 0 0 0 0 1 ...
##  $ Overwt    : int  2 0 1 2 1 1 2 0 0 0 ...
```

Transforming categorical columns to factor data type.

```
categorical_cols <- c('Smoke', 'Overwt')
df[categorical_cols] <- lapply(df[categorical_cols], as.factor)
```

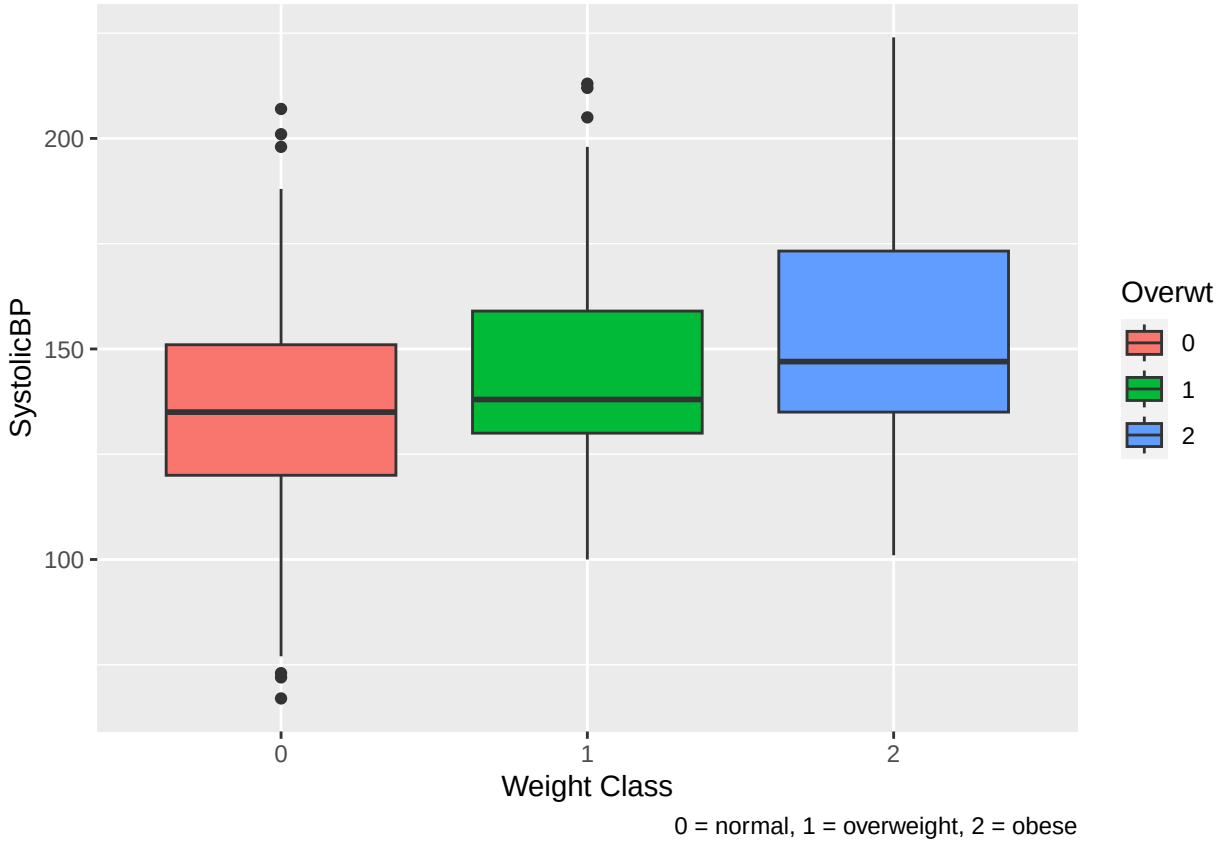
Ensuring column data types are correct now.

```
str(df)
```

```
## 'data.frame':    500 obs. of  3 variables:
##  $ SystolicBP: int  133 115 140 132 133 138 133 67 138 130 ...
##  $ Smoke     : Factor w/ 2 levels "0","1": 1 2 2 1 1 1 1 1 1 2 ...
##  $ Overwt    : Factor w/ 3 levels "0","1","2": 3 1 2 3 2 2 3 1 1 1 ...
```

Side by side boxplots comparing the blood pressure for the three weight groups.

```
ggplot(df, aes(x=Overwt, y=SystolicBP, fill=Overwt)) +
  geom_boxplot() +
  xlab('Weight Class') +
  labs(caption = '0 = normal, 1 = overweight, 2 = obese')
```



The sample means, standard deviations and sample sizes for the three weight groups.

Weight Group 0

```
# class 0: mean, std, sample size
df_0 <- subset(df, Overwt == 0)
str(df_0)
```

```
## 'data.frame':    187 obs. of  3 variables:
##  $ SystolicBP: int  115 67 138 130 134 107 127 117 139 124 ...
##  $ Smoke     : Factor w/ 2 levels "0","1": 2 1 1 2 2 1 2 2 1 1 ...
##  $ Overwt    : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
mean_0 <- mean(df_0$SystolicBP)
sd_0 <- sd(df_0$SystolicBP)
n_0 <- length(df_0$SystolicBP)
stats_0 <- c(mean_0, sd_0, n_0)
print(stats_0)
```

```
## [1] 136.31551  27.26852 187.00000
```

Weight Group 1

```
# class 0: mean, std, sample size
df_1 <- subset(df, Overwt == 1)
str(df_1)
```

```
## 'data.frame':    109 obs. of  3 variables:
##  $ SystolicBP: int  140 133 138 140 131 120 132 135 126 114 ...
##  $ Smoke     : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 2 1 1 ...
##  $ Overwt    : Factor w/ 3 levels "0","1","2": 2 2 2 2 2 2 2 2 2 2 ...
```

```
mean_1 <- mean(df_1$SystolicBP)
sd_1 <- sd(df_1$SystolicBP)
n_1 <- length(df_1$SystolicBP)
stats_1 <- c(mean_1, sd_1, n_1)
print(stats_1)
```

```
## [1] 144.36697  25.07864 109.00000
```

Weight Group 2

```
# class 0: mean, std, sample size
df_2 <- subset(df, Overwt == 2)
str(df_2)
```

```
## 'data.frame':    204 obs. of  3 variables:
##  $ SystolicBP: int  133 132 133 103 137 131 113 131 130 103 ...
##  $ Smoke     : Factor w/ 2 levels "0","1": 1 1 1 2 2 2 1 2 1 2 ...
##  $ Overwt    : Factor w/ 3 levels "0","1","2": 3 3 3 3 3 3 3 3 3 3 ...
```

```
mean_2 <- mean(df_2$SystolicBP)
sd_2 <- sd(df_2$SystolicBP)
n_2 <- length(df_2$SystolicBP)
stats_2 <- c(mean_2, sd_2, n_2)
print(stats_2)
```

```
## [1] 153.18137  27.81397 204.00000
```

Conducting a one-way anova modeling

```
anova_model <- aov(SystolicBP ~ as.factor(Overwt), data=df)
summary(anova_model)
```

```
##                     Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(Overwt)    2  27801   13900   19.02 1.1e-08 ***
## Residuals          497 363274     731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

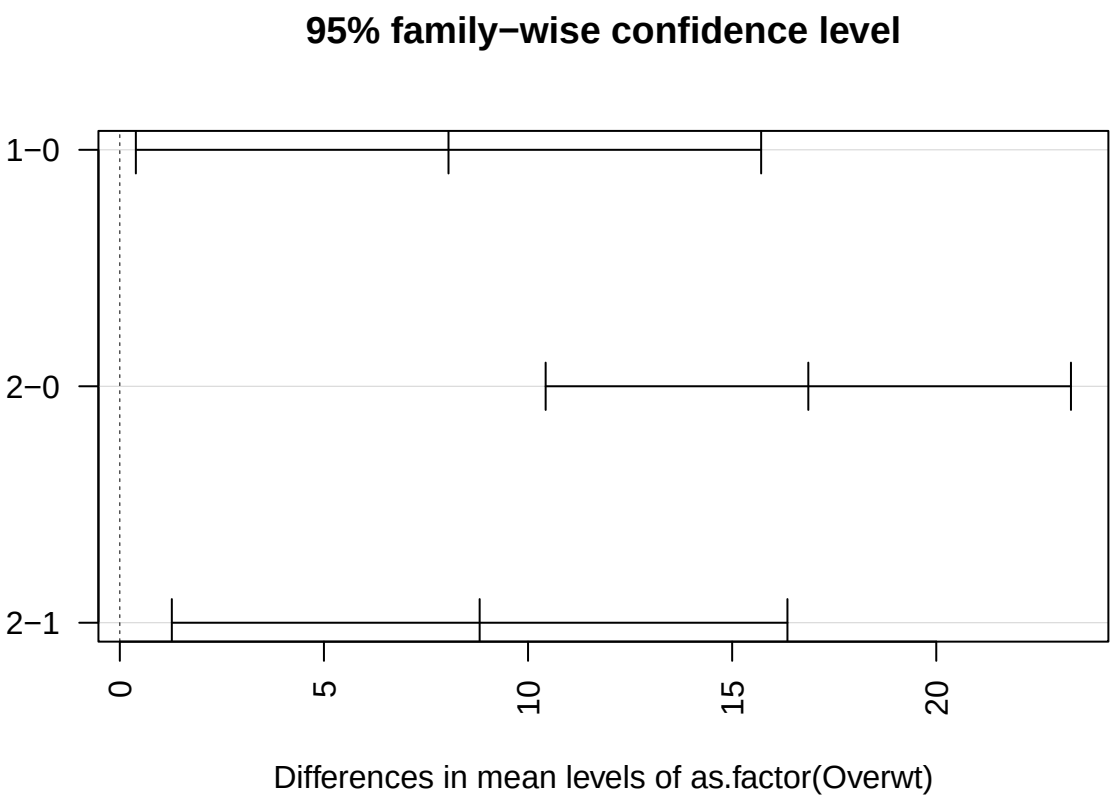
Tukey multiple comparisons procedure to determine which population means differ (if any)

```
TukeyHSD(anova_model, conf.level=.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = SystolicBP ~ as.factor(Overwt), data = df)
##
## $`as.factor(Overwt)`
##          diff        lwr      upr     p adj
## 1-0  8.051464  0.3927115 15.71022 0.0366867
## 2-0 16.865865 10.4316024 23.30013 0.0000000
## 2-1  8.814400  1.2740746 16.35473 0.0170703
```

Tukey graph for one-way anova

```
plot(TukeyHSD(anova_model, conf.level=.95), las = 2)
```



two factor ANOVA model with factors Smoke and Overwt (and response being SystolicBP) with no interaction

```
two_way_anova <- aov(SystolicBP ~ as.factor(Overwt) + as.factor(Smoke), data=df)
summary(two_way_anova)
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(Overwt)    2  27801   13900   19.53 6.84e-09 ***
## as.factor(Smoke)     1  10277   10277   14.44 0.000163 ***
## Residuals          496 352997     712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

two factor ANOVA model with factors Smoke and Overwt (and response being SystolicBP) interaction

```
two_way_anova_i <- aov(SystolicBP ~ as.factor(Overwt) * as.factor(Smoke), data=df)
summary(two_way_anova_i)
```

```
##                                    Df Sum Sq Mean Sq F value   Pr(>F)
## as.factor(Overwt)                   2  27801   13900  19.495 7.09e-09 ***
## as.factor(Smoke)                    1  10277   10277  14.414 0.000165 ***
## as.factor(Overwt):as.factor(Smoke)  2    769     384   0.539 0.583614
## Residuals                         494 352228     713
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```