Flipped Assignment 13

Group 5

2022/3/31

Input Data

```
setwd('G:/OneDrive - Texas Tech University/IE 5344 Statistical Data Analysis/Flipped Assignment 13')
data <- read.csv('data-heartbeat-3.csv', header = TRUE)
dat <- data %>%
  na_if("") %>%
  na.omit
dat <- dat[,-c(5,6,7,8,9,10)]
colnames(dat) <- c('y', 'x1', 'x2', 'x3', 'x4')
head(dat)
```

```
##     y x1 x2 x3   x4
## 51 72 AS  2  M Some
## 52 68 AS  4  M  Yes
## 53 88 AS  3  M Some
## 54 80 AS  3  M  Yes
## 55 96 AS  3  M  Yes
## 56 68 AF  3  M   No
```

Part a.

Continent is categorical.

```
dat$x1 <- as.factor(dat$x1)
fit1 <- lm(y ~ x1, dat)
summary(fit1)
```

```
##
## Call:
## lm(formula = y ~ x1, data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.667 -5.667 -1.667  2.333 18.333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   71.000      3.842  18.479 3.13e-11 ***
## x1AS           6.667      4.436   1.503    0.155
## x1SA           1.000      8.591   0.116    0.909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.684 on 14 degrees of freedom
## Multiple R-squared:  0.1522, Adjusted R-squared:  0.03107
## F-statistic: 1.257 on 2 and 14 DF,  p-value: 0.3148
```

We don't think continent is significant because p-values of both its coded variables are greater than 0.05. Moreover, the p-value of whole model is also greater than 0.05. For parameters' interpretation, when an observation is from Africa, his/her heartbeat is expected to be 71 bpm. If an observation is from Asia, his/her heartbeat is expected to be 77.667 bpm. If an observation is from South America, his/her heartbeat is expected to be 72 bpm.

Part b.

Exercise frequency is ordinal. Assume that the marginal response is fixed.

```
str(dat)
```

```
## 'data.frame':    17 obs. of  5 variables:
##  $ y : int  72 68 88 80 96 68 80 80 72 76 ...
##  $ x1: Factor w/ 3 levels "AF","AS","SA": 2 2 2 2 2 1 1 2 2 2 ...
##  $ x2: int  2 4 3 3 3 3 2 4 2 1 ...
##  $ x3: chr  "M" "M" "M" "M" ...
##  $ x4: chr  "Some" "Yes" "Some" "Yes" ...
```

```
fit2 <- lm(y ~ x2, dat)
summary(fit2)
```

```
##
## Call:
## lm(formula = y ~ x2, data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.500 -5.000 -2.667  3.000 20.167
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   72.333      6.012  12.032 4.17e-09 ***
## x2             1.167      1.936   0.603    0.556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.967 on 15 degrees of freedom
## Multiple R-squared:  0.02365,    Adjusted R-squared:  -0.04144
## F-statistic: 0.3633 on 1 and 15 DF,  p-value: 0.5557
```

We don't think exercise frequency is significant because its p-value greater than 0.05. Moreover, the p-value of whole model is also greater than 0.05. For parameters' interpretation, when the level of exercise frequency increases by 1, the expected increase in heartbeat is 1.167 bpm.

Part c.

Exercise frequency is ordinal. Assuming that the marginal response varies, it is treated categorical.

```
dat$x2 <- as.factor(dat$x2)
str(dat)
```

```
## 'data.frame':    17 obs. of  5 variables:
##  $ y : int  72 68 88 80 96 68 80 80 72 76 ...
```

```
##  $ x1: Factor w/ 3 levels "AF","AS","SA": 2 2 2 2 2 1 1 2 2 2 ...
##  $ x2: Factor w/ 4 levels "1","2","3","4": 2 4 3 3 3 3 2 4 2 1 ...
##  $ x3: chr  "M" "M" "M" "M" ...
##  $ x4: chr  "Some" "Yes" "Some" "Yes" ...
```

```
fit3 <- lm(y ~ x2, dat)
summary(fit3)
```

```
##
## Call:
## lm(formula = y ~ x2, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.333  -6.000  -2.667   5.333  16.667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   70.000      5.620  12.454 1.34e-08 ***
## x22            4.667      7.256   0.643    0.531
## x23            9.333      6.490   1.438    0.174
## x24            4.667      6.490   0.719    0.485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.949 on 13 degrees of freedom
## Multiple R-squared:  0.1577, Adjusted R-squared:  -0.03673
## F-statistic: 0.8111 on 3 and 13 DF,  p-value: 0.5102
```

We don't think exercise frequency is significant because p-values of all its coded variables are greater than 0.05. Moreover, the p-value of whole model is also greater than 0.05. For parameters' interpretation, when an observation's exercise frequency is 1, his/her heartbeat is expected to be 70 bpm. If an observation's exercise frequency is 2, his/her heartbeat is expected to be 74.667 bpm. If an observation's exercise frequency is 3, his/her heartbeat is expected to be 79.333 bpm. If an observation's exercise frequency is 4, his/her heartbeat is expected to be 74.667 bpm.

Part d.

(1)

We consider the marginal effects of x_2 and x_4 to vary because we have no background knowledge for their effects. Assuming the marginal effects of them gives flexibility to do regression. The full model is

```
dat$x4 <- as.factor(dat$x4)
dat$x3 <- as.factor(dat$x3)
str(dat)
```

```
## 'data.frame':    17 obs. of  5 variables:
##  $ y : int  72 68 88 80 96 68 80 80 72 76 ...
##  $ x1: Factor w/ 3 levels "AF","AS","SA": 2 2 2 2 2 1 1 2 2 2 ...
##  $ x2: Factor w/ 4 levels "1","2","3","4": 2 4 3 3 3 3 2 4 2 1 ...
##  $ x3: Factor w/ 2 levels "F","M": 2 2 2 2 2 2 2 2 1 1 2 ...
##  $ x4: Factor w/ 3 levels "No","Some","Yes": 2 3 2 3 3 1 2 2 1 2 ...
```

```
fit4 <- lm(y ~., dat)
summary(fit4)
```

```
##
## Call:
## lm(formula = y ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.820 -3.564 -1.085  1.277 13.597
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   67.470     10.912   6.183 0.000264 ***
## x1AS           3.573      5.634   0.634 0.543660
## x1SA          -1.349     10.688  -0.126 0.902674
## x22            2.234      8.424   0.265 0.797590
## x23            6.905      7.575   0.912 0.388620
## x24            1.424      8.466   0.168 0.870576
## x3M           -2.384      7.973  -0.299 0.772533
## x4Some         6.256      5.581   1.121 0.294878
## x4Yes          6.839      6.536   1.046 0.325933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.717 on 8 degrees of freedom
## Multiple R-squared:  0.3765, Adjusted R-squared:  -0.2469
## F-statistic: 0.6039 on 8 and 8 DF,  p-value: 0.7542
```

Here we remove the variable that has the largest p-value, which is continent. The model becomes

```r
dat <- dat[,-2]
fit5 <- lm(y ~., dat)
summary(fit5)
```

```
##
## Call:
## lm(formula = y ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.784 -3.154 -1.693  2.370 14.307
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   70.528      9.212   7.656 1.73e-05 ***
## x22            2.071      7.795   0.266    0.796
## x23            8.154      6.816   1.196    0.259
## x24            1.614      7.729   0.209    0.839
## x3M           -4.157      7.016  -0.593    0.567
## x4Some         7.260      4.961   1.463    0.174
## x4Yes          7.169      5.866   1.222    0.250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.069 on 10 degrees of freedom
## Multiple R-squared:  0.3322, Adjusted R-squared:  -0.06846
## F-statistic: 0.8291 on 6 and 10 DF,  p-value: 0.5732
```

Removing exercise frequency, the model becomes

```
dat <- dat[,-2]
fit6 <- lm(y ~., dat)
summary(fit6)
```

```
##
## Call:
## lm(formula = y ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.333 -5.333 -1.333  2.286 18.667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   72.571      6.115  11.868 2.39e-08 ***
## x3M           -1.714      6.219  -0.276    0.787
## x4Some         6.857      4.818   1.423    0.178
## x4Yes          6.476      4.971   1.303    0.215
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.949 on 13 degrees of freedom
## Multiple R-squared:  0.1577, Adjusted R-squared:  -0.03673
## F-statistic: 0.8111 on 3 and 13 DF,  p-value: 0.5102
```

Removing gender, the model becomes

```
dat <- dat[,-2]
fit7 <- lm(y ~., dat)
summary(fit7)
```

```
##
## Call:
## lm(formula = y ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.333 -5.333 -1.333  2.000 18.667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   71.200      3.435  20.725 6.63e-12 ***
## x4Some         6.800      4.652   1.462    0.166
## x4Yes          6.133      4.652   1.319    0.208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.682 on 14 degrees of freedom
## Multiple R-squared:  0.1527, Adjusted R-squared:  0.0317
## F-statistic: 1.262 on 2 and 14 DF,  p-value: 0.3134
```