Stats 506 Group Project

Jingyan Lu

12/5/2019

Load the four datasets

```
demographics = read.xport("./DEMO_I.XPT")
blood_pressure = read.xport("./BPX_I.XPT")
nutrients_1day = read.xport("./DR1TOT_I.XPT")
BMI = read.xport("./BMX_I.XPT")
```

Transform and merge the data frames

```
blood_pressure = blood_pressure %>%
  select(SEQN, BPXSY1, BPXSY2, BPXSY3, BPXDI1, BPXDI2, BPXDI3) %>%
  rowwise() %>%
  mutate(DI = mean(c(BPXDI1, BPXDI2, BPXDI3), na.rm=TRUE),
         SY = mean(c(BPXSY1, BPXSY2, BPXSY3), na.rm = TRUE)) %>%
  select(SEQN, DI, SY)

demographics = demographics %>%
  select(SEQN, RIDAGEYR)

nutrients_1day = nutrients_1day %>%
  select(SEQN, DBD100)

BMI = BMI %>%
  select(SEQN, BMXWAIST)

data = demographics %>%
  left_join(BMI, by = "SEQN") %>%
  left_join(nutrients_1day, by = "SEQN") %>%
  left_join(blood_pressure, by = "SEQN")


data = data[complete.cases(data), ]
```

Center means to reduce multicolinearity

```
data[c("RIDAGEYR", "BMXWAIST", "DBD100", "DI", "SY")] =
  lapply(data[c("RIDAGEYR", "BMXWAIST", "DBD100", "DI", "SY")],
         function(x) scale(x, center=TRUE, scale=FALSE))
```

Analyse the relationsip between salt intake and two kinds of blood pressure

For diastole blood pressure

```
model_DI = lm(DI ~ DBD100, data)
summary(model_DI)
```

```
##
## Call:
## lm(formula = DI ~ DBD100, data = data)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -67.247  -7.689   0.311   8.311  60.645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.991e-15  1.950e-01   0.000 1.000000
## DBD100      8.918e-01  2.296e-01   3.884 0.000104 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.34 on 4680 degrees of freedom
## Multiple R-squared:  0.003213,   Adjusted R-squared:  0.003
## F-statistic: 15.08 on 1 and 4680 DF,  p-value: 0.0001042
```

For systole blood pressure

```
model_SY = lm(SY ~ DBD100, data)
summary(model_SY)
```

```
##
## Call:
## lm(formula = SY ~ DBD100, data = data)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -47.011 -12.074  -2.740   8.322  86.593
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.744e-15  2.526e-01   0.000 1.000000
## DBD100      1.135e+00  2.975e-01   3.816 0.000137 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.29 on 4680 degrees of freedom
## Multiple R-squared:  0.003102,   Adjusted R-squared:  0.002889
## F-statistic: 14.56 on 1 and 4680 DF,  p-value: 0.0001373
```

Both p-values are less than 0.001, which shows strong evidence that salt intake have siginificant influence on both kinds of blood pressure.

Moderation part

We perform moderation to test if the relationship is dependent on the waist size choose three levels of a moderator(mean, one standard deviation above the mean and one standard deviation below the mean).

Moderation at mean for diastole blood pressure

```
moderation_DI = lm(DI ~ DBD100 + BMXWAIST + DBD100 * BMXWAIST, data)
summary(moderation_DI)
```

```
##
## Call:
## lm(formula = DI ~ DBD100 + BMXWAIST + DBD100 * BMXWAIST, data = data)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -74.261  -7.147   0.498   7.943  53.281
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.010034   0.186699   0.054  0.95714
## DBD100           0.596273   0.219951   2.711  0.00673 **
## BMXWAIST         0.207160   0.009824  21.086  < 2e-16 ***
## DBD100:BMXWAIST -0.009955   0.012081  -0.824  0.41000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.75 on 4678 degrees of freedom
## Multiple R-squared:  0.09015,    Adjusted R-squared:  0.08957
## F-statistic: 154.5 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Moderation at one standard deviation above mean for diastole blood pressure

```
data$BMXWAIST_high = data$BMXWAIST + sd(data$BMXWAIST)
moderation_DI_high = lm(DI ~ DBD100 + BMXWAIST_high + DBD100 * BMXWAIST_high, data)
summary(moderation_DI_high)
```

```
##
## Call:
## lm(formula = DI ~ DBD100 + BMXWAIST_high + DBD100 * BMXWAIST_high,
##     data = data)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -74.261  -7.147   0.498   7.943  53.281
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.928731   0.264392 -14.860   <2e-16 ***
## DBD100            0.785543   0.313193   2.508   0.0122 *
```

```
## BMXWAIST_high        0.207160    0.009824   21.086    <2e-16 ***
## DBD100:BMXWAIST_high -0.009955    0.012081   -0.824    0.4100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.75 on 4678 degrees of freedom
## Multiple R-squared:  0.09015,    Adjusted R-squared:  0.08957
## F-statistic: 154.5 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Modetation at one standard deviation below mean for diastole blood pressure

```
data$BMXWAIST_low = data$BMXWAIST - sd(data$BMXWAIST)
moderation_DI_low = lm(DI ~ DBD100 + BMXWAIST_low + DBD100 * BMXWAIST_low, data)
summary(moderation_DI_low)
```

```
##
## Call:
## lm(formula = DI ~ DBD100 + BMXWAIST_low + DBD100 * BMXWAIST_low,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.261  -7.147   0.498   7.943  53.281
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.948800   0.263806  14.969   <2e-16 ***
## DBD100             0.407004   0.322794   1.261    0.207
## BMXWAIST_low       0.207160   0.009824  21.086   <2e-16 ***
## DBD100:BMXWAIST_low -0.009955   0.012081  -0.824    0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.75 on 4678 degrees of freedom
## Multiple R-squared:  0.09015,    Adjusted R-squared:  0.08957
## F-statistic: 154.5 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Since the regression coefficient for the interation term is not significant with p value 0.41, there does not exist a significant moderation effect. the effect of salt intake on diastole blood pressure may not depends on waist size.

Moderation at mean for systole blood pressure

```
moderation_SY = lm(SY ~ DBD100 + BMXWAIST + DBD100 * BMXWAIST, data)
summary(moderation_SY)
```

```
##
## Call:
## lm(formula = SY ~ DBD100 + BMXWAIST + DBD100 * BMXWAIST, data = data)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q     Max
## -44.138 -10.182  -2.532   6.798  92.036
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.009924   0.230167  -0.043   0.9656
## DBD100            0.610168   0.271162   2.250   0.0245 *
## BMXWAIST          0.379788   0.012112  31.357   <2e-16 ***
## DBD100:BMXWAIST   0.009845   0.014894   0.661   0.5086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.72 on 4678 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1757
## F-statistic: 333.7 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Moderation at one standard deviation above mean for systole blood pressure

```
data$BMXWAIST_high = data$BMXWAIST + sd(data$BMXWAIST)
moderation_SY_high = lm(SY ~ DBD100 + BMXWAIST_high + DBD100 * BMXWAIST_high, data)
summary(moderation_SY_high)
```

```
##
## Call:
## lm(formula = SY ~ DBD100 + BMXWAIST_high + DBD100 * BMXWAIST_high,
##     data = data)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -44.138 -10.182  -2.532   6.798  92.036
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -7.230880   0.325950 -22.184   <2e-16 ***
## DBD100                 0.422982   0.386113   1.095    0.273
## BMXWAIST_high          0.379788   0.012112  31.357   <2e-16 ***
## DBD100:BMXWAIST_high   0.009845   0.014894   0.661    0.509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.72 on 4678 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1757
## F-statistic: 333.7 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Modetation at one standard deviation below mean for diastole blood pressure

```
data$BMXWAIST_low = data$BMXWAIST - sd(data$BMXWAIST)
moderation_SY_low = lm(SY ~ DBD100 + BMXWAIST_low + DBD100 * BMXWAIST_low, data)
summary(moderation_SY_low)
```

```
##
## Call:
```

```
## lm(formula = SY ~ DBD100 + BMXWAIST_low + DBD100 * BMXWAIST_low,
##      data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.138 -10.182  -2.532   6.798  92.036
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        7.211033   0.325227  22.172   <2e-16 ***
## DBD100             0.797353   0.397949   2.004   0.0452 *
## BMXWAIST_low       0.379788   0.012112  31.357   <2e-16 ***
## DBD100:BMXWAIST_low 0.009845  0.014894   0.661   0.5086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.72 on 4678 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1757
## F-statistic: 333.7 on 3 and 4678 DF,  p-value: < 2.2e-16
```

Since the regression coefficient for the interation term is not significant with p value 0.51, there does not exist a significant moderation effect. the effect of salt intake on systole blood pressure may not depends on waist size as well.

Mediation Part

We perform mediation to test if the relationship between salt intake and blood pressure mediated by age.

First, test if there is relationship between age and salt intake, since mediation makes sense only if they have relationship

```
age_salt_DI = lm(RIDAGEYR ~ DBD100, data)
summary(age_salt_DI)
```

```
##
## Call:
## lm(formula = RIDAGEYR ~ DBD100, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.005 -19.872  -3.024  17.823  43.433
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.015e-16  3.145e-01   0.000        1
## DBD100       2.305e+00  3.704e-01   6.223 5.31e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.52 on 4680 degrees of freedom
```

```
## Multiple R-squared:  0.008206,   Adjusted R-squared:  0.007994
## F-statistic: 38.72 on 1 and 4680 DF,  p-value: 5.309e-10
```

The p_value is **5.31e-10. They have strong relationship.**

Second, perform the mediation

For diastole blood pressure

```
mediation_DI = lm(DI ~ DBD100 + RIDAGEYR, data)
summary(mediation_DI)
```

```
##
## Call:
## lm(formula = DI ~ DBD100 + RIDAGEYR, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.075  -7.197   0.551   7.921  60.624
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.089e-15  1.881e-01   0.000   1.0000
## DBD100      5.152e-01  2.224e-01   2.316   0.0206 *
## RIDAGEYR    1.634e-01  8.743e-03  18.687   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.87 on 4679 degrees of freedom
## Multiple R-squared:  0.07244,    Adjusted R-squared:  0.07204
## F-statistic: 182.7 on 2 and 4679 DF,  p-value: < 2.2e-16
```

The effect of salt intake on diastole blood pressure still exists(p-value is 0.021), but in a smaller magnitude.Age partially mediates between salt intake and diastole blood pressure.

For systole blood pressure

```
mediation_SY = lm(SY ~ DBD100 + RIDAGEYR, data)
summary(mediation_SY)
```

```
##
## Call:
## lm(formula = SY ~ DBD100 + RIDAGEYR, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.567  -8.973  -1.109   7.366  73.387
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.019e-15  2.077e-01   0.000    1.000
## DBD100      8.128e-02  2.456e-01   0.331    0.741
## RIDAGEYR    4.574e-01  9.655e-03  47.375   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.21 on 4679 degrees of freedom
## Multiple R-squared:  0.3263, Adjusted R-squared:  0.326
## F-statistic:  1133 on 2 and 4679 DF,  p-value: < 2.2e-16
```

The effect of salt intake on systole blood pressure disappear (p-value is 0,741), age fully mediates salt intake and systole blood pressure.

Third, perform the mediation through bootstrapping

For diastole blood pressure

```
## [1] "Bootstrap resampling has begun. This process may take a considerable amount of time if the numb
```

mediation_boot_DI1

```
##                                           Estimate CI.Lower_Percentile
## Indirect.Effect                         0.3765500632          0.2505787649
## Indirect.Effect.Partially.Standardized  0.0281847931          0.0188093474
## Index.of.Mediation                      0.0239330262          0.0161318694
## R2_4.5                                  0.0021491333          0.0007676457
## R2_4.6                                  0.0005699203          0.0002598030
## R2_4.7                                  0.0078676552          0.0038091572
## Ratio.of.Indirect.to.Total.Effect       0.4222366268          0.2722860156
## Ratio.of.Indirect.to.Direct.Effect      0.7308123818          0.3589352505
## Success.of.Surrogate.Endpoint           0.3869472836          0.2073195740
## Residual.Based_Gamma                    0.0043638728          0.0020185913
## Residual.Based.Standardized_gamma       0.0039579315          0.0018605497
## SOS                                     0.6689288496          0.4724133709
##                                        CI.Upper_Percentile
## Indirect.Effect                                0.507481979
## Indirect.Effect.Partially.Standardized         0.037967238
## Index.of.Mediation                             0.031631566
## R2_4.5                                         0.003742520
## R2_4.6                                         0.000992203
## R2_4.7                                         0.012788460
## Ratio.of.Indirect.to.Total.Effect             0.809033277
## Ratio.of.Indirect.to.Direct.Effect            3.116990645
## Success.of.Surrogate.Endpoint                  0.589696767
## Residual.Based_Gamma                           0.007490006
## Residual.Based.Standardized_gamma              0.006799977
## SOS                                            0.949258888
```

The indirect effect (0.37655) and its confidence interval is different from zero. Age is a mediator between salt intake and diastole blood pressure.

For systole blood pressure

[1] "Bootstrap resampling has begun. This process may take a considerable amount of time if the numb

```
mediation_boot_SY1
```

```
##                                         Estimate CI.Lower_Percentile
## Indirect.Effect                        1.054137647          0.708602105
## Indirect.Effect.Partially.Standardized 0.060897024          0.041186404
## Index.of.Mediation                     0.051710512          0.035397580
## R2_4.5                                 0.003086452          0.000351890
## R2_4.6                                 0.002660286          0.001249634
## R2_4.7                                 0.008153615          0.003859541
## Ratio.of.Indirect.to.Total.Effect      0.928416123          0.621891063
## Ratio.of.Indirect.to.Direct.Effect    12.969626170        -67.404480564
## Success.of.Surrogate.Endpoint          0.492651297          0.266898230
## Residual.Based_Gamma                   0.003966790          0.001564294
## Residual.Based.Standardized_gamma      0.003752141          0.001397527
## SOS                                    0.994917800          0.485932934
##                                        CI.Upper_Percentile
## Indirect.Effect                                1.403703892
## Indirect.Effect.Partially.Standardized         0.080191587
## Index.of.Mediation                             0.067416654
## R2_4.5                                         0.006884625
## R2_4.6                                         0.004504418
## R2_4.7                                         0.013678985
## Ratio.of.Indirect.to.Total.Effect             1.719289633
## Ratio.of.Indirect.to.Direct.Effect           85.562803262
## Success.of.Surrogate.Endpoint                  0.733954514
## Residual.Based_Gamma                           0.007410994
## Residual.Based.Standardized_gamma              0.007175315
## SOS                                            0.999962753
```

The estimate indirect effect(1.054) and its confidence interval is different from zero. Age is a mediator between salt intake and systole blood pressure.