Homework 2

Alex Ojemann

2024-09-27

Problem 1

Part a

The following regression adjustments estimate the causal effect of quitting smoking on weight change, SBP, and DBP using sex, race, age, education, smoking intensity per day, number of years having smoked, exercise level, activity level, weight at the beginning of the study period, and whether the subject had high blood pressure at the beginning of the study period.

```
## Weight Change (1971-1982) :
##  - Estimated effect of quitting smoking:  3.51
##  - Standard Error:  0.44
##  - 95% Confidence Interval: [ 2.64 ,  4.37 ]
##
```

```
## Systolic Blood Pressure :
##  - Estimated effect of quitting smoking:  2.15
##  - Standard Error:  1.03
##  - 95% Confidence Interval: [ 0.14 ,  4.16 ]
##
```

```
## Diastolic Blood Pressure :
##  - Estimated effect of quitting smoking:  1.73
##  - Standard Error:  0.62
##  - 95% Confidence Interval: [ 0.52 ,  2.95 ]
##
```

Part b

The following regression adjustments estimate the causal effect of quitting smoking on weight change, SBP, and DBP using the same covariates as those in part a but with squared and logarithmic terms included for the continuous numeric covariates.

```
## Weight Change (1971-1982) :
##  - Estimated effect of quitting smoking:  3.57
##  - Standard Error:  0.44
##  - 95% Confidence Interval: [ 2.71 ,  4.43 ]
##
```

```
## Systolic Blood Pressure :
##  - Estimated effect of quitting smoking:  2.21
##  - Standard Error:  1.03
##  - 95% Confidence Interval: [ 0.19 ,  4.23 ]
##


## Diastolic Blood Pressure :
##  - Estimated effect of quitting smoking:  1.72
##  - Standard Error:  0.62
##  - 95% Confidence Interval: [ 0.51 ,  2.93 ]
##
```

Part c

The following regression adjustments estimate the causal effect of quitting smoking on weight change, SBP, and DBP using the same covariates as those in part a but with the product of the quit smoking covariate with each of the other covariates as additional features.

```
## Weight Change (1971-1982):
## - Estimated effect of quitting smoking: 1.86
## - Standard Error: 2.52
## - 95% CI: [-3.07, 6.8]


## Systolic Blood Pressure:
## - Estimated effect of quitting smoking: -3.23
## - Standard Error: 5.79
## - 95% CI: [-14.59, 8.12]


## Diastolic Blood Pressure:
## - Estimated effect of quitting smoking: 0.88
## - Standard Error: 3.53
## - 95% CI: [-6.05, 7.8]
```

Part d

The following regression adjustments estimate the causal effect of quitting smoking on weight change, SBP, and DBP using the same covariates as those in parts a, b, and c but with the final models having features selected using stepwise feature selection with BIC as the information criterion.

Note: when initially using stepwise feature selection, the model dropped the quit smoking variable in the SBP model. It was added back in manually after stepwise feature selection because it is the variable for which we are trying to evaluate causal relationships.

```
## Weight Change (1971-1982) (Selected Model):
## - Estimated effect of quitting smoking: 3.31
## - Standard Error: 0.43
## - 95% CI: [2.47, 4.15]


## Systolic Blood Pressure (Selected Model):
## - Estimated effect of quitting smoking: 1.95
## - Standard Error: 1.02
## - 95% CI: [-0.04, 3.94]
```

```
## Diastolic Blood Pressure (Selected Model):
## - Estimated effect of quitting smoking: 1.77
## - Standard Error: 0.61
## - 95% CI: [0.58, 2.96]
```

Below is the summary for the model used to predict weight change:

```
summary(final_model_wt82_71)
```

```
##
## Call:
## lm(formula = wt82_71 ~ qsmk + age + age_squared + wt71_squared,
##     data = nhefs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -41.923  -4.164  -0.242   4.012  45.210
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.8971158  2.4657930  -1.175     0.24
## qsmk1         3.3087995  0.4298223   7.698 2.44e-14 ***
## age           0.5408850  0.1132695   4.775 1.96e-06 ***
## age_squared  -0.0078068  0.0012435  -6.278 4.43e-10 ***
## wt71_squared -0.0005576  0.0000783  -7.122 1.62e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.366 on 1561 degrees of freedom
##   (63 observations deleted due to missingness)
## Multiple R-squared:  0.1284, Adjusted R-squared:  0.1262
## F-statistic: 57.48 on 4 and 1561 DF,  p-value: < 2.2e-16
```

Below is the summary for the model used to predict SBP:

```
summary(final_model_sbp)
```

```
##
## Call:
## lm(formula = sbp ~ age + race + age_squared + wt71 + hbp + qsmk,
##     data = nhefs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.426 -11.423  -1.775   9.077  89.369
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 77.744153   6.004574  12.947  < 2e-16 ***
## age          1.314334   0.264084   4.977 7.18e-07 ***
## race1        5.398950   1.300258   4.152 3.47e-05 ***
## age_squared -0.008209   0.002897  -2.834  0.00466 **
```

```
## wt71          0.123107   0.028679    4.293 1.88e-05 ***
## hbp1          7.575918   1.737206    4.361 1.38e-05 ***
## hbp2         -0.193995   0.911387   -0.213  0.83147
## qsmk1         1.947573   1.015203    1.918  0.05524 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.16 on 1544 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.1927, Adjusted R-squared:  0.189
## F-statistic: 52.64 on 7 and 1544 DF,  p-value: < 2.2e-16
```

Below is the summary for the model used to predict DBP:

```
summary(final_model_dbp)
```

```
##
## Call:
## lm(formula = dbp ~ qsmk + age + race + log_age + log_wt71, data = nhefs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.559  -7.289  -0.579   6.517  47.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -76.4063    18.5592  -4.117 4.04e-05 ***
## qsmk1         1.7685     0.6062   2.917  0.00358 **
## age          -1.1072     0.1564  -7.080 2.18e-12 ***
## race1         3.6813     0.7747   4.752 2.20e-06 ***
## log_age      45.3679     6.6416   6.831 1.21e-11 ***
## log_wt71      7.5481     1.2422   6.077 1.54e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.23 on 1542 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.07708,    Adjusted R-squared:  0.07408
## F-statistic: 25.76 on 5 and 1542 DF,  p-value: < 2.2e-16
```