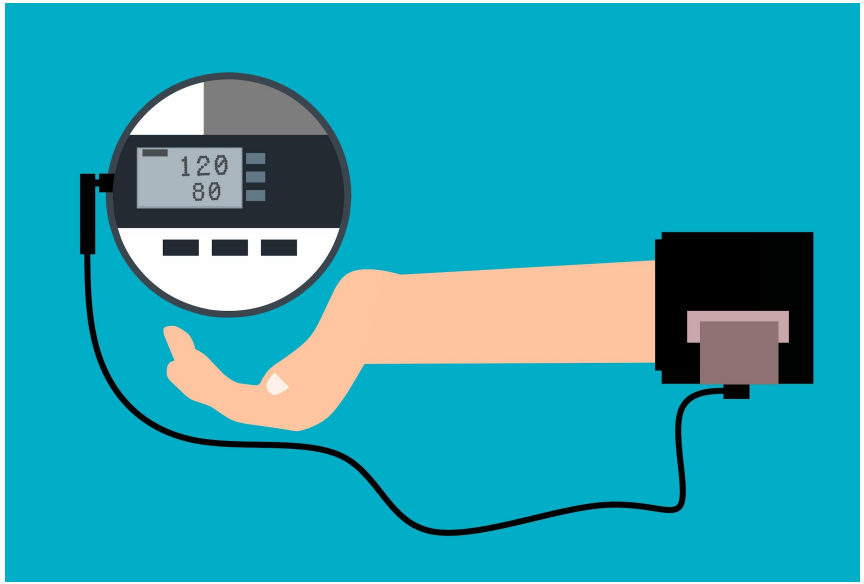
Systolic Blood Pressure

Mingjia Chen
Kuei-Sheng Hou
Raymond Moy
Ruichen Rachel Zhou

Research Interest

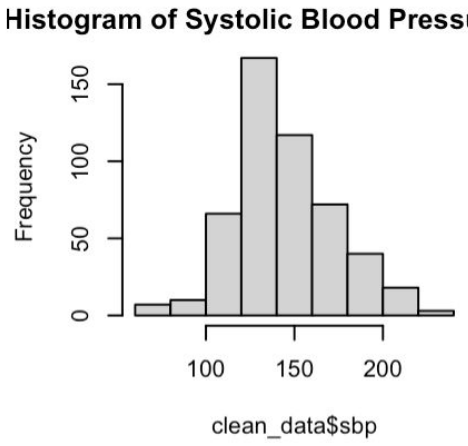
What Factors Play Significant Role of Influencing Systolic Blood Pressure?

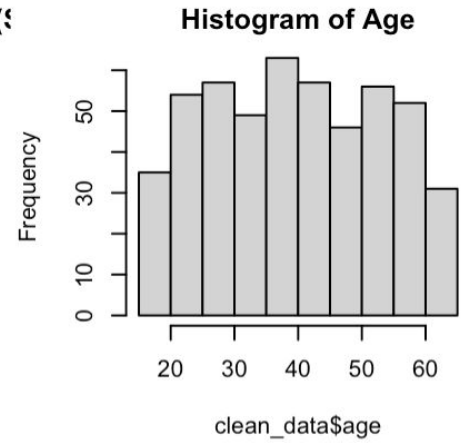


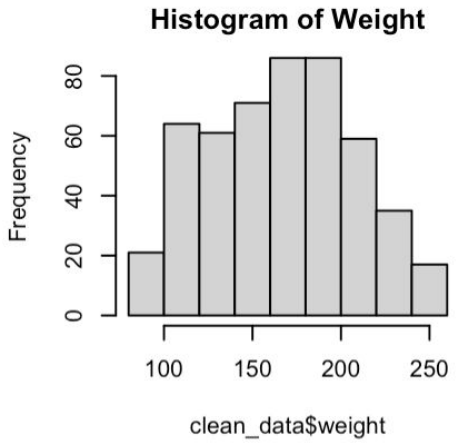
- Gender?
- Age?
- Consumption of alcohol?
- …

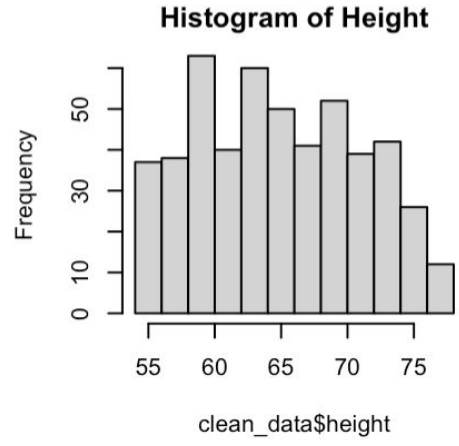
The Data

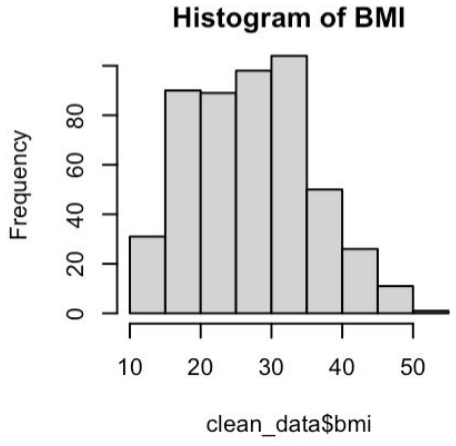
EDA, cleaning, and collinearity check


Histogram of Systolic Blood Pressure(S
Frequency
0
50
100
150
100
150
200
clean_data$sbp


Histogram of Age
Frequency
0
10
30
50
20
30
40
50
60
clean_data$age


Histogram of Weight
Frequency
0
20
40
60
80
100
150
200
250
clean_data$weight


Histogram of Height
Frequency
0
10
30
50
55
60
65
70
75
clean_data$height

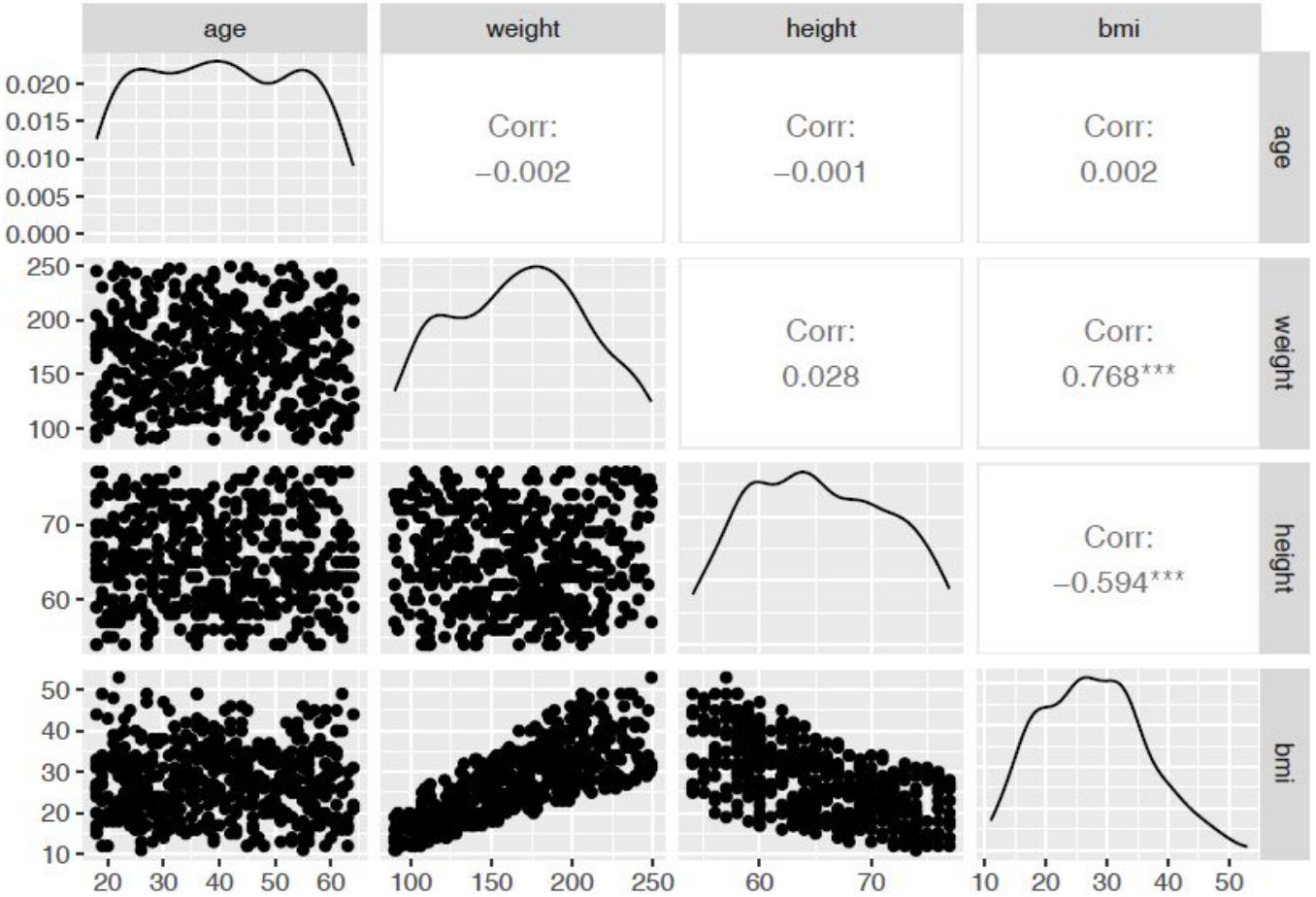

Histogram of BMI
Frequency
0
20
40
60
80
10
20
30
40
50
clean_data$bmi


Historgram of Alcohol
Frequency
0
50
100
150
1.0
1.5
2.0
2.5
3.0
clean_data$alcohol

Data Cleaning

Variables	Descriptions	Min	Q1	Median	Mean	Q3	Max	SD
sbp	Systolic Blood Pressure(SBP)	67	130	140.5	145	162.2	224	28
age	years	18	28	40	40	52	64	13.3
weight	lbs	90	133	168	166.6	198	249	40.9
height	inches	54	60	65	65.33	70	77	6.2
bmi	Body Mass Index(BMI)	11	21	27	27.66	33	53	8.6

Correlation of Continuous Predictors



The Model

Model Building, Model Selection, Model Diagnostics

Model Selection

- Step 1: split the data
- Step 2: run automated model selection (step)
- Step 3: add interaction terms and run automated model selection (step)
- Step 4: Compare the model from step 2, 3 using anova table.

Model Selection

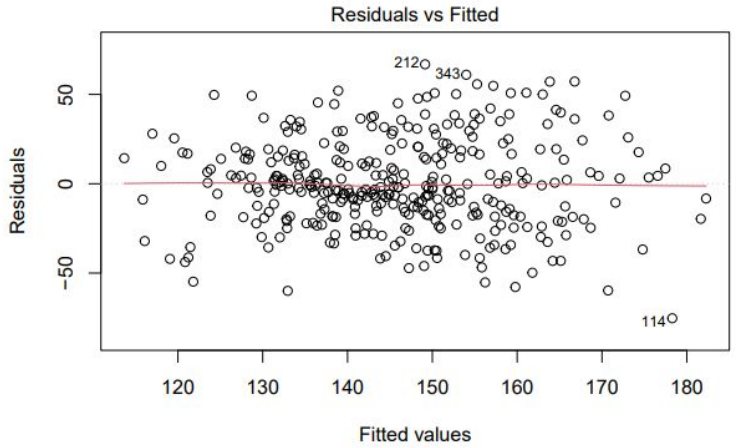
Best model:

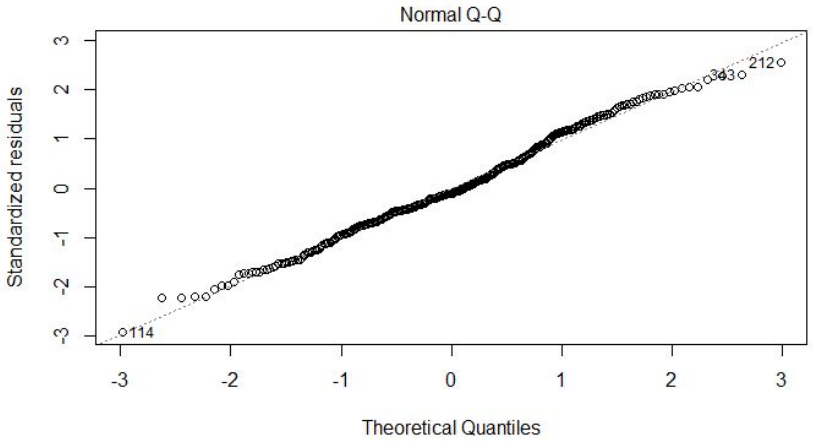
Systolic Blood Pressure ~ Smoking status + exercise level + height + alcohol use + Treatment + bmi + Smoking status:alcohol use + Treatment:bmi

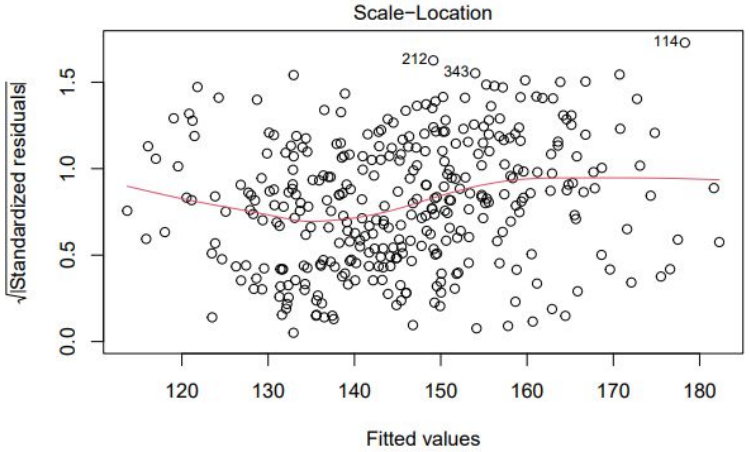
```
Single term deletions

Model:
sbp ~ smoke + exercise + height + alcohol + trt + bmi + smoke:alcohol +
    trt:bmi
             Df Sum of Sq    RSS    AIC Pr(>Chi)
<none>                    218831 2277.3
exercise      2    8289.8 227121 2286.4 0.001493 **
height        1    3556.3 222387 2281.0 0.017532 *
smoke:alcohol 2    5197.1 224028 2281.6 0.016448 *
trt:bmi       1    5857.3 224688 2284.6 0.002361 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

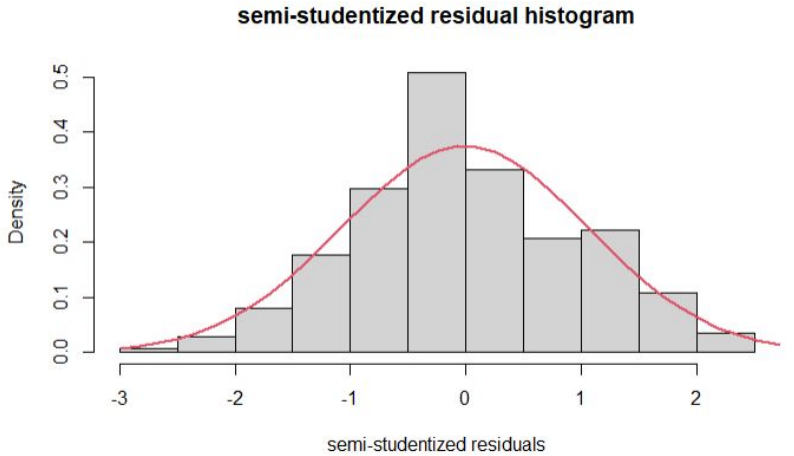
```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        87.4846    25.0817   3.488 0.000551 ***
smokeY              1.0884     4.5716   0.238 0.811971
exerciseLow         9.4380     3.1694   2.978 0.003112 **
exerciseMedium     -1.3733     3.6200  -0.379 0.704652
height              0.6589     0.2811   2.344 0.019671 *
alcoholLow        -19.9067     5.0890  -3.912 0.000111 ***
alcoholMedium     -17.2163     4.7762  -3.605 0.000360 ***
trtUntreated      -19.2780    12.2524  -1.573 0.116561
bmi                 0.2065     0.3978   0.519 0.604011
smokeY:alcoholLow  17.8134     6.8115   2.615 0.009317 **
smokeY:alcoholMedium 14.2997   6.5577   2.181 0.029903 *
trtUntreated:bmi    1.2280     0.4083   3.008 0.002829 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.44 on 338 degrees of freedom
Multiple R-squared:  0.2271,    Adjusted R-squared:  0.2019
F-statistic: 9.028 on 11 and 338 DF,  p-value: 3.763e-14
```

LINE Assumptions

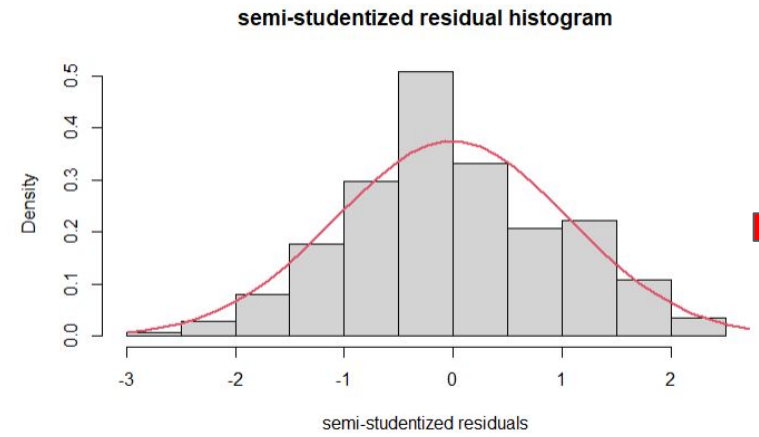


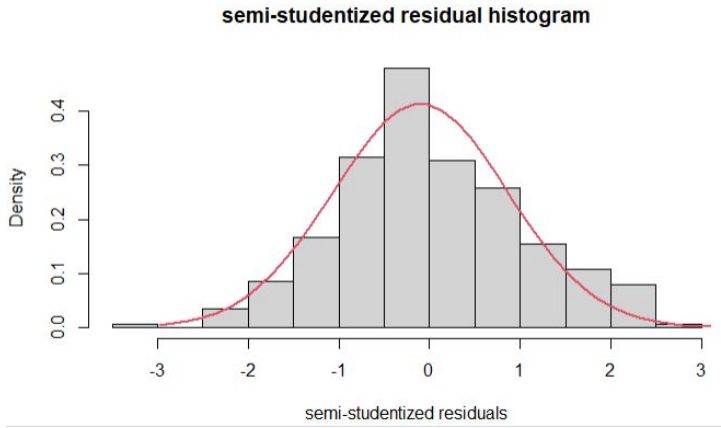


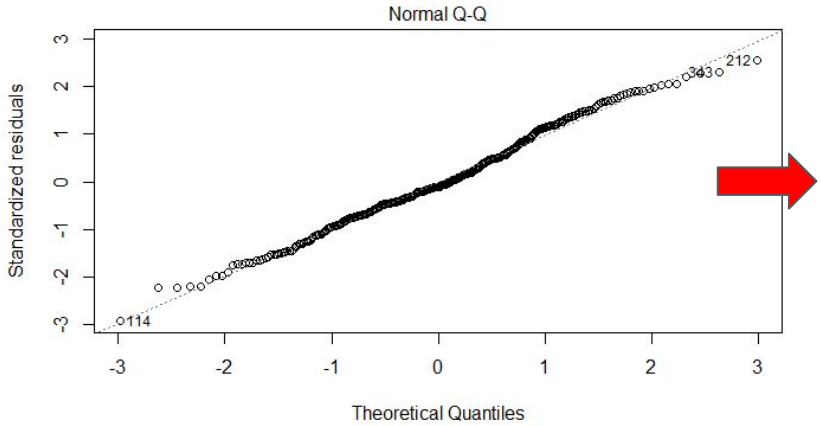


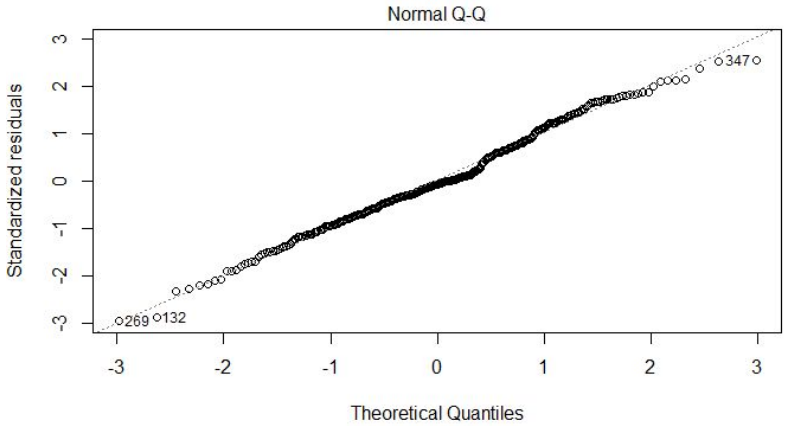


OLS vs WLS


semi-studentized residual histogram
Density
0.0
0.1
0.2
0.3
0.4
0.5
-3
-2
-1
0
1
2
semi-studentized residuals


semi-studentized residual histogram
Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
semi-studentized residuals


Normal Q-Q
Standardized residuals
-3
-2
-1
0
1
2
3
114
343
212
Theoretical Quantiles


Normal Q-Q
Standardized residuals
-3
-2
-1
0
1
2
3
269
132
347
Theoretical Quantiles

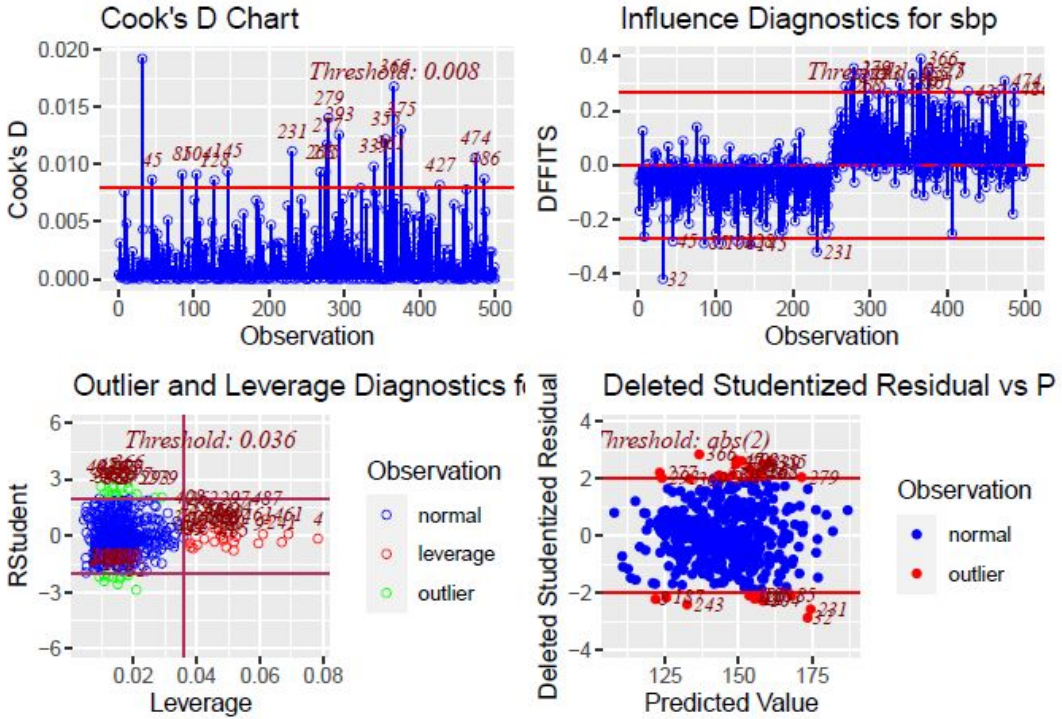
Model Validation / Diagnostics

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 117.0493     8.4547  13.844  < 2e-16 ***
smoking      10.9611     3.3843   3.239 0.001367 **
exercise     -5.8467     1.9355  -3.021 0.002789 **
alcohol       4.7630     2.0279   2.349 0.019639 *
trt         -15.2785     4.1680  -3.666 0.000303 ***
bmi           0.9902     0.2057   4.814  2.6e-06 ***
---
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 114.3237     7.6720  14.901  < 2e-16 ***
smoking      10.6636     3.2109   3.321 0.001034 **
exercise     -4.4703     1.8918  -2.363 0.018913 *
alcohol       6.8356     1.9990   3.419 0.000735 ***
trt          -9.5738     4.0276  -2.377 0.018224 *
bmi           0.7670     0.1809   4.240 3.18e-05 ***
---
```

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	R^2_{adj}
Training	117.05 (8.45)	10.96 (3.38)	-5.85 (1.94)	4.76 (2.03)	-15.28 (4.17)	0.99 (0.21)	0.1657
Validation	114.32 (7.67)	10.66 (3.21)	-4.47 (1.89)	6.84 (2.00)	-9.57 (4.03)	0.77 (0.18)	0.1589

Influential and Outlying Observations



What next?

Impacts, limitations, future directions

Impact

- Dataset: 500 subjects
- Relation between SBP and numerous contributing factors

Limitation

- Outlying observations could be confounding factors.
- Sample representation
- Non-constant variances

Future Directions

- Province, racial specific data
- Replications

Reference

Modi, J. (2022, April 14). *Which Is More Important: Systolic or Diastolic Blood Pressure?* Buzzrx.com. https://www.buzzrx.com/blog/which-is-more-important-systolic-or-diastolic-blood-pressure (for image)

Understanding blood pressure readings. www.heart.org. (2023, February 2). Retrieved April 2, 2023, from https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings