

STAC67 Regression Analyses Blood Pressure data

Group 12

Donglin Que, Yichen Bao, Yifan Cui, Yuxin Zhang

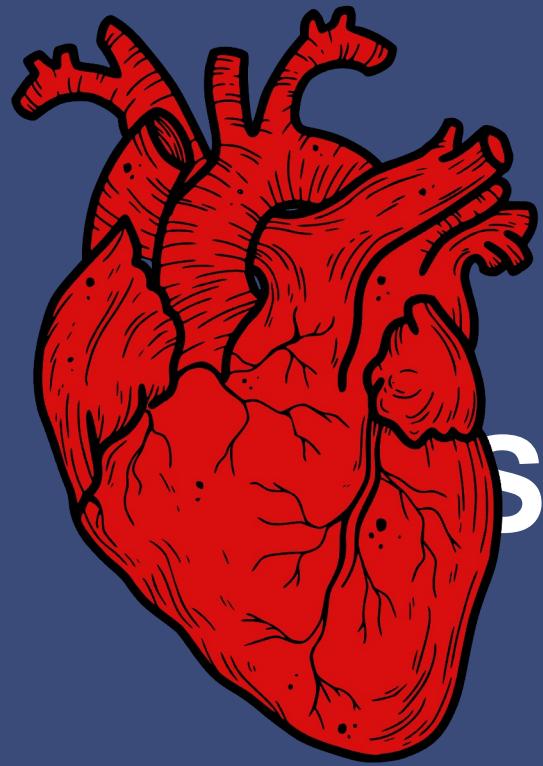
CATALOG

Systolic blood pressure

Data

Model

What is next



01

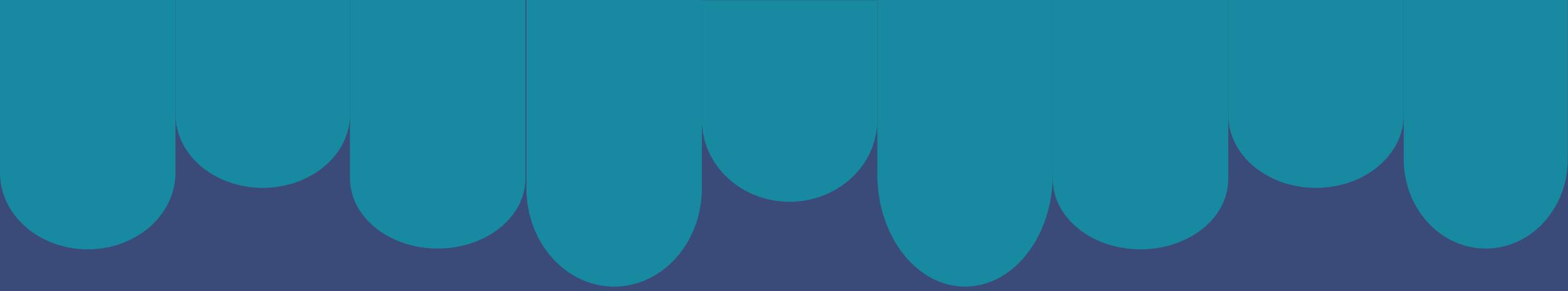
Systolic blood pressure

A normal blood pressure level is less than 120/80 mmHg
Elevated systolic pressure is the most common form of hypertension

Purpose: find out the factors that affect SBP

systolic

diastolic



02

Data

Cleaning Data Set

- 500 observations
- 18 related variables

```
## # A tibble: 500 x 18
##   sbp gender married smoke exercise age weight height overwt race alcohol
##   <dbl> <chr>  <chr>   <chr>    <dbl> <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl>
## 1 133 F     N      N        3     60    159    56    3     1     1     2
## 2 115 M     N      Y        1     55    107    65    1     1     1     2
## 3 140 M     N      Y        1     18    130    59    2     1     1     1
## 4 132 M     Y      N        2     19    230    57    3     2     2     3
## 5 133 M     N      N        2     58    201    74    2     1     1     3
## 6 138 F     N      N        3     55    166    67    2     1     1     1
## 7 133 F     Y      N        1     22    188    66    3     1     1     3
## 8 67  F     Y      N        3     52    123    67    1     1     1     2
## 9 138 M     Y      N        1     46    106    73    1     1     1     3
## 10 130 M    Y      Y        3     38    166    72    1     1     1    1
## # ... with 490 more rows, and 7 more variables: trt <dbl>, bmi <dbl>,
## #   stress <dbl>, salt <dbl>, chldbear <dbl>, income <dbl>, educatn <dbl>
```

Correlation matrix

Test the correlation

Avoid multicollinearity

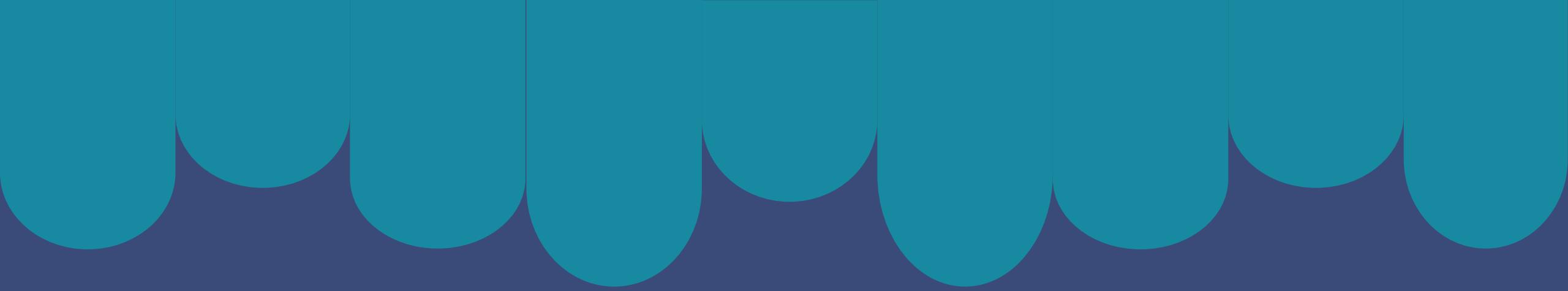
##	sbp	exercise	age	weight	height	overwt	race	alcohol	trt	bmi	stress	salt	chldbear	income	educatn	male	marry	smoker
## sbp	1.000	-0.145	0.037	0.230	-0.117	0.267	-0.008	0.133	-0.126	0.267	0.067	-0.029	0.025	0.046	-0.009	0.002	0.061	0.193
## exercise	-0.145	1.000	0.048	0.025	0.045	-0.008	0.012	-0.012	-0.028	-0.018	-0.017	0.038	0.029	0.084	-0.031	-0.032	-0.036	0.060
## age	0.037	0.048	1.000	-0.002	-0.001	0.050	-0.022	-0.094	0.035	0.002	0.040	0.052	-0.023	0.034	-0.020	0.005	-0.017	-0.004
## weight	0.230	0.025	-0.002	1.000	0.028	0.717	0.032	-0.105	0.121	0.768	0.057	-0.020	-0.207	0.010	-0.006	0.236	-0.081	0.079
## height	-0.117	0.045	-0.001	0.028	1.000	-0.514	0.031	-0.062	0.014	-0.594	0.058	0.079	-0.269	0.039	0.033	0.293	0.017	-0.069
## overwt	0.267	-0.008	0.050	0.717	-0.514	1.000	0.023	-0.084	0.093	0.889	0.034	-0.050	0.003	-0.023	-0.030	0.004	-0.064	0.122
## race	-0.008	0.012	-0.022	0.032	0.031	0.023	1.000	0.075	-0.076	0.003	0.075	-0.024	-0.052	0.054	0.080	0.020	-0.069	-0.040
## alcohol	0.133	-0.012	-0.094	-0.105	-0.062	-0.084	0.075	1.000	0.063	-0.040	-0.029	-0.081	0.093	0.038	-0.039	-0.089	0.073	-0.049
## trt	-0.126	-0.028	0.035	0.121	0.014	0.093	-0.076	0.063	1.000	0.093	0.058	-0.025	-0.035	0.036	-0.011	0.063	-0.043	0.063
## bmi	0.267	-0.018	0.002	0.768	-0.594	0.889	0.003	-0.040	0.093	1.000	0.003	-0.060	0.016	-0.006	-0.018	0.000	-0.077	0.106
## stress	0.067	-0.017	0.040	0.057	0.058	0.034	0.075	-0.029	0.058	0.003	1.000	-0.029	-0.039	0.015	-0.009	0.031	-0.080	0.029
## salt	-0.029	0.038	0.052	-0.020	0.079	-0.050	-0.024	-0.081	-0.025	-0.060	-0.029	1.000	-0.025	0.010	-0.085	0.009	-0.054	-0.053
## chldbear	0.025	0.029	-0.023	-0.207	-0.269	0.003	-0.052	0.093	-0.035	0.016	-0.039	-0.025	1.000	0.002	0.115	-0.895	0.039	0.040
## income	0.046	0.084	0.034	0.010	0.039	-0.023	0.054	0.038	0.036	-0.006	0.015	0.010	0.002	1.000	-0.027	0.044	-0.019	-0.088
## educatn	-0.009	-0.031	-0.020	-0.006	0.033	-0.030	0.080	-0.039	-0.011	-0.018	-0.009	-0.085	0.115	-0.027	1.000	-0.109	0.036	-0.007
## male	0.002	-0.032	0.005	0.236	0.293	0.004	0.020	-0.089	0.063	0.000	0.031	0.009	-0.895	0.044	-0.109	1.000	-0.039	-0.045
## marry	0.061	-0.036	-0.017	-0.081	0.017	-0.064	-0.069	0.073	-0.043	-0.077	-0.080	-0.054	0.039	-0.019	0.036	-0.039	1.000	0.031
## smoker	0.193	0.060	-0.004	0.079	-0.069	0.122	-0.040	-0.049	0.063	0.106	0.029	-0.053	0.040	-0.088	-0.007	-0.045	0.031	1.000

Significance

```
full_model <- lm(sbp ~ factor(exercise) + age + factor(race) + factor(alcohol) +
  factor(trt) + bmi + factor(stress) + factor(salt) +
  factor(chldbear) + factor(income) + factor(educatn) +
  factor(male) + factor(marry) + factor(smoker), data)
summary(full_model)
```

```
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.95249  6.96415 15.214 < 2e-16 ***
## factor(exercise)2 -10.98009  2.90353 -3.782 0.000176 ***
## factor(exercise)3 -10.52344  2.71324 -3.879 0.000120 ***
## age            0.13255  0.08653  1.532 0.126224
## factor(race)2    0.69652  2.93500  0.237 0.812514
## factor(race)3    1.72464  5.36298  0.322 0.747910
## factor(race)4   -6.90089  5.85545 -1.179 0.239168
## factor(alcohol)2   1.41978  2.86598  0.495 0.620554
## factor(alcohol)3   12.18700  2.89115  4.215 2.98e-05 ***
## factor(trt)1     -14.13742  2.90852 -4.861 1.59e-06 ***
## bmi             0.93620  0.13616  6.876 1.94e-11 ***
## factor(stress)2    3.00199  2.86855  1.047 0.295851
## factor(stress)3    5.67193  2.86696  1.978 0.048462 *
## factor(salt)2      2.44908  2.88140  0.850 0.395773
## factor(salt)3      1.67061  2.81687  0.593 0.553414
## factor(chldbear)2   -3.47968  2.75239 -1.264 0.206761
## factor(chldbear)3   0.97013  2.90152  0.334 0.738260
## factor(income)2     1.30556  2.78627  0.469 0.639591
## factor(income)3     4.29918  2.85513  1.506 0.132787
## factor(educatn)2    0.19730  2.82985  0.070 0.944445
## factor(educatn)3    0.07465  2.80788  0.027 0.978801
## factor(male)1        NA       NA      NA      NA
## factor(marry)1      2.82132  2.33366  1.209 0.227274
## factor(smoker)1     10.72304  2.34779  4.567 6.30e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remove not significant variables



03

Model

Correlation matrix

$$\text{BMI} = \text{weight} / \text{height}^2$$

```
##          sbp exercise age weight height overwt race alcohol trt bmi stress salt chldbear income educatn male marry smoker
## sbp      1.000 -0.145 0.037 0.230 -0.117 0.267 -0.008 0.133 -0.126 0.267 0.067 -0.029 0.025 0.046 -0.009 0.002 0.061 0.193
## exercise -0.145  1.000 0.048 0.025 0.045 -0.008 0.012 -0.012 -0.028 -0.018 -0.017 0.038 0.029 0.084 -0.031 -0.032 -0.036 0.060
## age       0.037  0.048 1.000 -0.002 -0.001 0.050 -0.022 -0.094 0.035 0.002 0.040 0.052 -0.023 0.034 -0.020 0.005 -0.017 -0.004
## weight    0.230  0.025 -0.002 1.000 0.028 0.717 0.032 -0.105 0.121 0.768 0.057 -0.020 -0.207 0.010 -0.006 0.236 -0.081 0.079
## height   -0.117  0.045 -0.001 0.028 1.000 -0.514 0.031 -0.062 0.014 -0.594 0.058 0.079 -0.269 0.039 0.033 0.293 0.017 -0.069
## full_model <- lm(sbp ~ factor(exercise) + age + factor(race) + factor(alcohol) +
##                      factor(trt) + bmi + factor(stress) + factor(salt) +
##                      factor(chldbear) + factor(income) + factor(educatn) +
##                      factor(male) + factor(marry) + factor(smoker), data)
##          sbp exercise age weight height overwt race alcohol trt bmi stress salt chldbear income educatn male marry smoker
## chldbear  0.025  0.029 -0.023 -0.207 -0.269  0.003 -0.052  0.093 -0.035  0.016 -0.039 -0.025  1.000  0.002  0.115 -0.895  0.039  0.040
## income    0.046  0.084  0.034  0.010  0.039 -0.023  0.054  0.038  0.036 -0.006  0.015  0.010  0.002  1.000 -0.027  0.044 -0.019 -0.088
## educatn -0.009 -0.031 -0.020 -0.006  0.033 -0.030  0.080 -0.039 -0.011 -0.018 -0.009 -0.085  0.115 -0.027  1.000 -0.109  0.036 -0.007
## male      0.002 -0.032  0.005  0.236  0.293  0.004  0.020 -0.089  0.063  0.000  0.031  0.009 -0.895  0.044 -0.109  1.000 -0.039 -0.045
## marry     0.061 -0.036 -0.017 -0.081  0.017 -0.064 -0.069  0.073 -0.043 -0.077 -0.080 -0.054  0.039 -0.019  0.036 -0.039  1.000  0.031
## smoker    0.193  0.060 -0.004  0.079 -0.069  0.122 -0.040 -0.049  0.063  0.106  0.029 -0.053  0.040 -0.088 -0.007 -0.045  0.031  1.000
```

Model selection

Stepwise selection based on AIC

male	-	-	-	-	-	-	-	-
race		-	-	-	-	-	-	-
educatn			-	-	-	-	-	-
salt				-	-	-	-	-
chldbear					-	-	-	-
income						-	-	-
marry							-	-
stress								-
exercise								
alcohol								
smoker								
treatment								
bmi								
age								

Tricky “age”

```
lm(formula = sbp ~ factor(exercise) + age + factor(alcohol) +  
    factor(trt) + bmi + factor(smoker), data = data)

Residuals:  
    Min      1Q  Median      3Q     Max  
-75.344 -17.733 -1.147  16.480  69.701

Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 113.11227   5.73018 19.740 < 2e-16 ***  
factor(exercise)2 -10.30448   2.85115 -3.614 0.000332 ***  
factor(exercise)3 -10.27216   2.67461 -3.841 0.000139 ***  
age                0.13572   0.08593  1.579 0.114882  
factor(alcohol)2    0.65455   2.81069  0.233 0.815952  
factor(alcohol)3    11.81848   2.81015  4.206 3.09e-05 ***  
factor(trt)1       -13.42877   2.86645 -4.685 3.63e-06 ***  
bmi                 0.91330   0.13486  6.772 3.64e-11 ***  
factor(smoker)1     10.90908   2.29620  4.751 2.66e-06 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.35 on 491 degrees of freedom  
Multiple R-squared:  0.1934,    Adjusted R-squared:  0.1803  
F-statistic: 14.72 on 8 and 491 DF,  p-value: < 2.2e-16
```

~~lm(formula = sbp ~ factor(exercise) + factor(alcohol) +
 factor(trt) + bmi + factor(smoker), data = data)~~

~~Residuals:
 Min 1Q Median 3Q Max
-74.502 -17.748 -1.536 16.518 66.626~~

~~Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.7139 4.5075 26.337 < 2e-16 ***
factor(exercise)2 -10.3364 2.8554 -3.620 0.000325 ***
factor(exercise)3 -10.0605 2.6753 -3.761 0.000190 ***
factor(alcohol)2 0.4343 2.8115 0.154 0.877297
factor(alcohol)3 11.3992 2.8018 4.068 5.51e-05 ***
factor(trt)1 -13.2432 2.8684 -4.617 4.98e-06 ***
bmi 0.9133 0.1351 6.762 3.87e-11 ***
factor(smoker)1 10.8576 2.2995 4.722 3.05e-06 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1~~

~~Residual standard error: 25.38 on 492 degrees of freedom
Multiple R-squared: 0.1893, Adjusted R-squared: 0.1778
F-statistic: 16.41 on 7 and 492 DF, p-value: < 2.2e-16~~

Final model

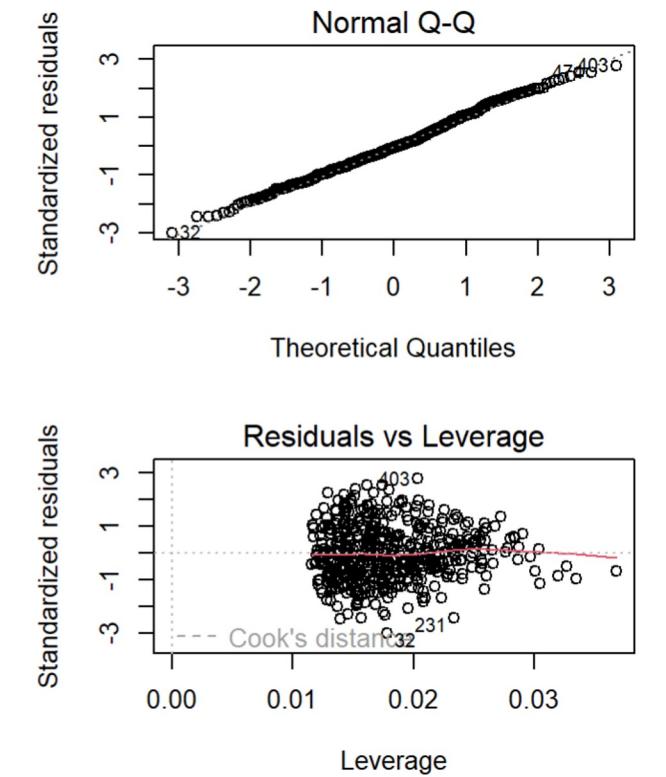
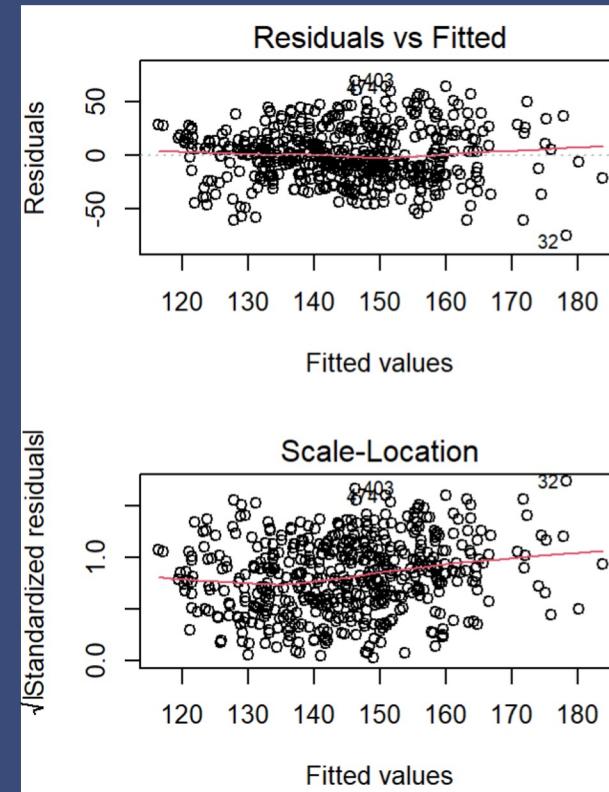
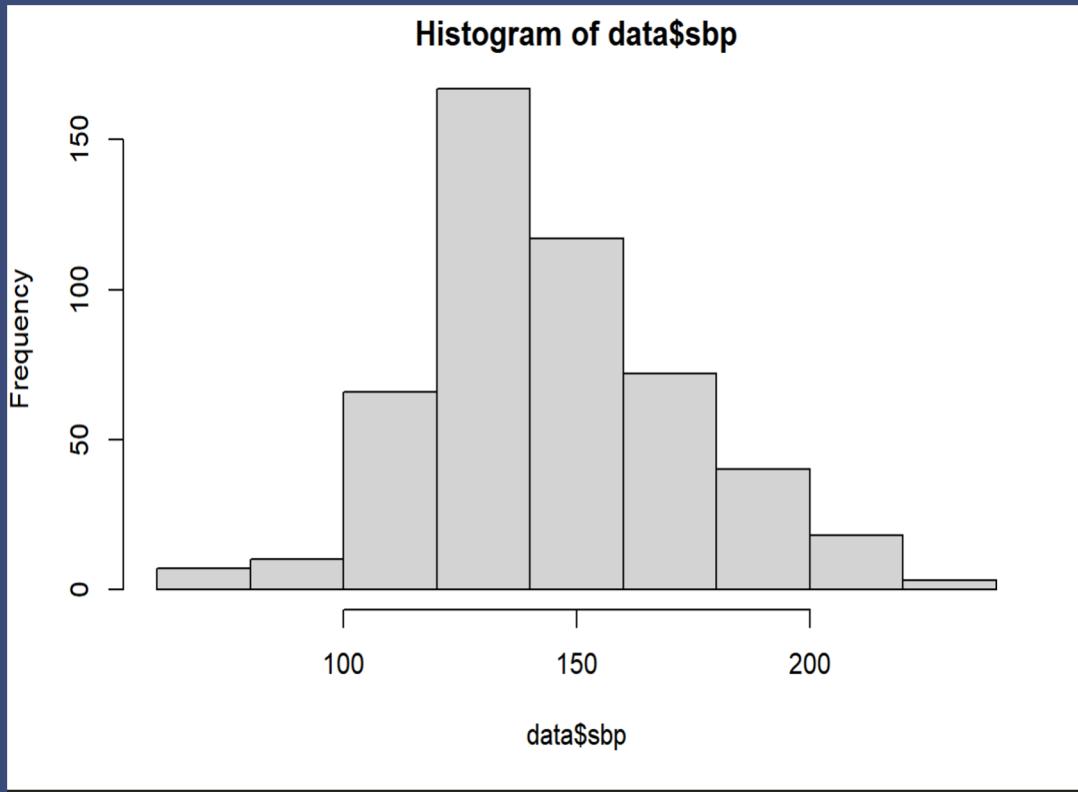
```
lm(formula = sbp ~ factor(exercise) + age + factor(alcohol) +
  factor(trt) + bmi + factor(smoker), data = data)

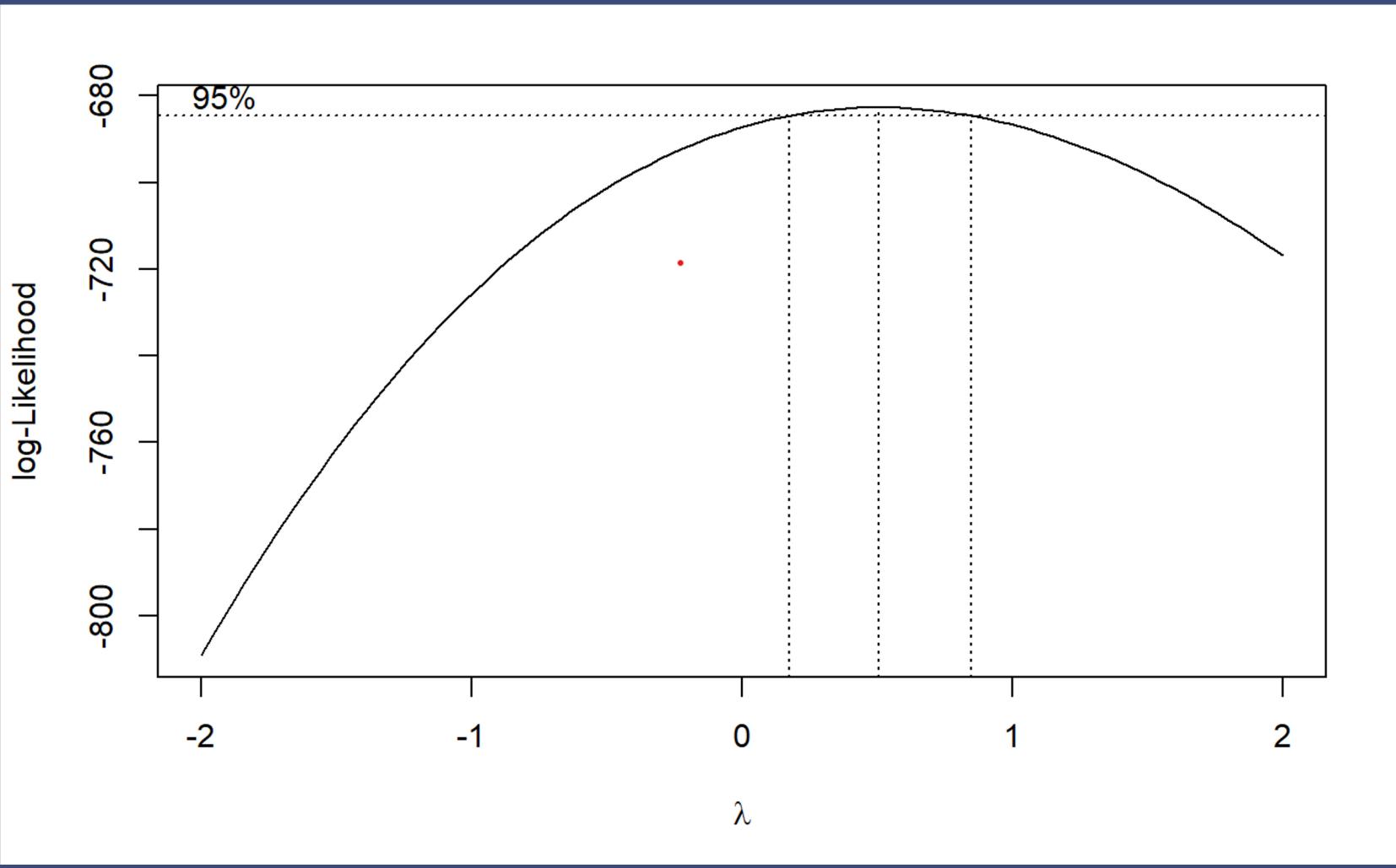
Residuals:
    Min      1Q  Median      3Q     Max 
-75.344 -17.733 -1.147  16.480  69.701 

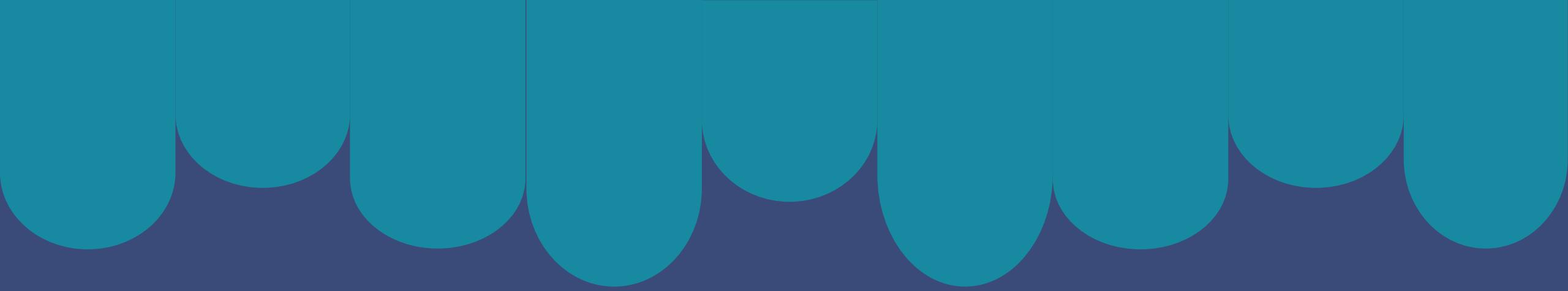
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 113.11227   5.73018 19.740 < 2e-16 ***
factor(exercise)2 -10.30448   2.85115 -3.614 0.000332 ***
factor(exercise)3 -10.27216   2.67461 -3.841 0.000139 *** 
age              0.13572   0.08593  1.579 0.114882  
factor(alcohol)2  0.65455   2.81069  0.233 0.815952  
factor(alcohol)3  11.81848   2.81015  4.206 3.09e-05 ***
factor(trt)1      -13.42877  2.86645 -4.685 3.63e-06 ***
bmi              0.91330   0.13486  6.772 3.64e-11 *** 
factor(smoker)1   10.90908   2.29620  4.751 2.66e-06 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 25.35 on 491 degrees of freedom
Multiple R-squared:  0.1934, Adjusted R-squared:  0.1803 
F-statistic: 14.72 on 8 and 491 DF,  p-value: < 2.2e-16
```

Model validation







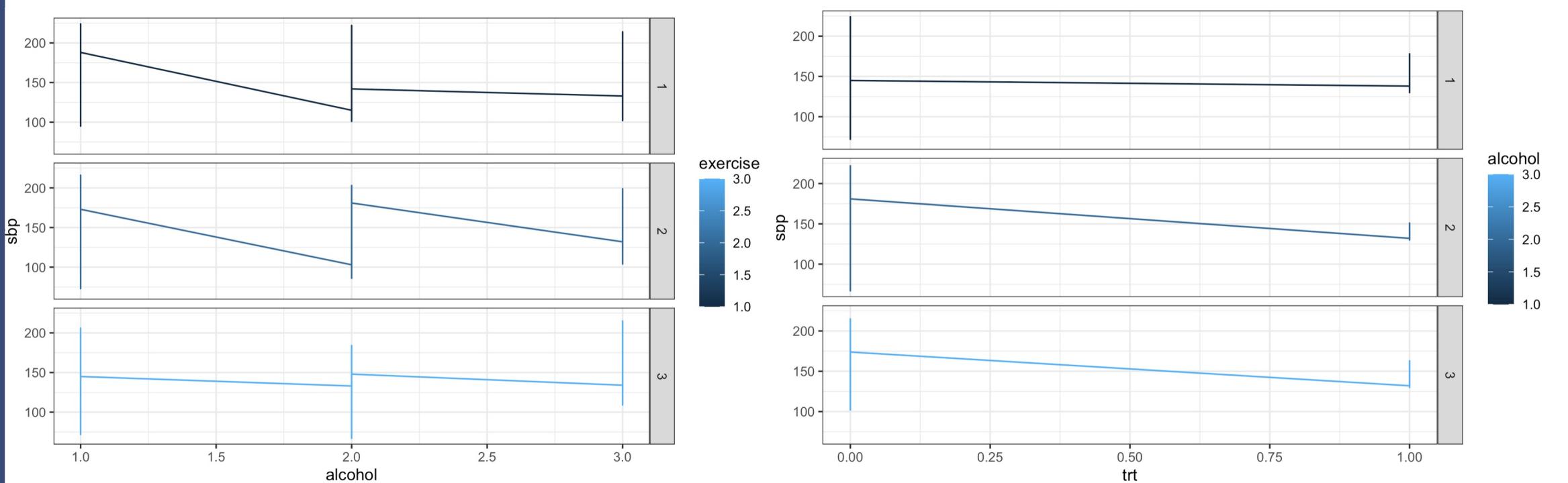
04

What is next

Statistic Conclusion

Basic on the analysis and models, the personal exercise leve, alcohol usage, Body Mass Index(BMI), somking stage and treatment for hypertensdion are all the factors which can effect the Systolic Blood Pressure

What we get



exercise: 1 = Low, 2 = Medium, 3 = High
High

alcohol use: 1 = Low, 2 = Medium, 3 =

Limitation and future direction

We only have blood pressure data for people **aged 18-64**, but high blood pressure is not unique to this age group

Our sample data size is **too small** to find provenance, with an average of only 10 samples per age. With more samples, we can draw more precise conclusions.

In the future, we can **continue to track** these people and observe their blood pressure fluctuations in the future, which will play an important role in the prevention and treatment of hypertension.

For reference, a 2001 study had 5,600 samples and followed them for a long time

References

- CDC. (2021, May 18). High Blood Pressure Symptoms, Causes, and Problems | cdc.gov. Centers for Disease Control and Prevention. <https://www.cdc.gov/bloodpressure/about.htm>
- Strandberg T. E., & Pitkala K. (2003). What is the most important component of blood pressure: Systolic, diastolic or pulse pressure? *Current Opinion in Nephrology and Hypertension*, 12(3), 293.
- Psaty, B. M., Furberg, C. D., Kuller, L. H.(2001). Association between blood pressure level and the risk of myocardial infarction, stroke, and total mortality: the cardiovascular health study. *Archives of internal medicine*, 161(9), 1183-1192.