

P8130 Final Report (Project 1)

Huanyu Chen (hc3451)

Xiaoting Tang (xt2288)

Yifei Liu (yl5508)

Longyu Zhang (lz2951)

2023-12-16

Read and Clean Data

```
data =  
read_csv("./data.csv") |>  
janitor::clean_names() |>  
mutate(  
  gender = factor(case_when(  
    gender == "male" ~ 0,  
    gender == "female" ~ 1,  
  )),  
  ethnic_group = factor(case_when(  
    ethnic_group == "group A" ~ 0,  
    ethnic_group == "group B" ~ 1,  
    ethnic_group == "group C" ~ 2,  
    ethnic_group == "group D" ~ 3,  
    ethnic_group == "group E" ~ 4,  
  )),  
  parent_educ = factor(case_when(  
    parent_educ == "some highschool" ~ 0,  
    parent_educ == "some college" ~ 1,  
    parent_educ == "associate's degree" ~ 2,  
    parent_educ == "bachelor's degree" ~ 3,  
    parent_educ == "master's degree" ~ 4,  
  )),  
  lunch_type = factor(case_when(  
    lunch_type == "standard" ~ 0,  
    lunch_type == "free/reduced" ~ 1,  
  )),  
  test_prep = factor(case_when(  
    test_prep == "none" ~ 0,  
    test_prep == "completed" ~ 1,  
  )),  
  parent_marital_status = factor(case_when(  
    parent_marital_status == "married" ~ 0,  
    parent_marital_status == "single" ~ 1,  
    parent_marital_status == "widowed" ~ 2,  
    parent_marital_status == "divorced" ~ 3,  
  )),
```

```

practice_sport = factor(case_when(
  practice_sport == "never" ~ 0,
  practice_sport == "sometimes" ~ 1,
  practice_sport == "regularly" ~ 2,
)),
is_first_child = factor(case_when(
  is_first_child == "no" ~ 0,
  is_first_child == "yes" ~ 1,
)),
transport_means = factor(case_when(
  transport_means == "school_bus" ~ 0,
  transport_means == "private" ~ 1,
)),
wkly_study_hours = factor(case_when(
  wkly_study_hours == "< 5" ~ 0,
  wkly_study_hours == "10-May" ~ 1,
  wkly_study_hours == "> 10" ~ 2,
))
) |>
mutate(nr_siblings = factor(nr_siblings))

```

```

## Rows: 948 Columns: 14
## -- Column specification -----
## Delimiter: ","
## chr (10): Gender, EthnicGroup, ParentEduc, LunchType, TestPrep, ParentMarita...
## dbl (4): NrSiblings, MathScore, ReadingScore, WritingScore
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

```

'
# Deal with NA -- Calculate the column mean (round to integer) and plug it into NA cell
column_means <- round(colMeans(data, na.rm = TRUE), digits = 0)
for (col in names(data)) {
  data[[col]][is.na(data[[col]])] <- column_means[col]
}

head(data)
'

```

```

## [1] "\n# Deal with NA -- Calculate the column mean (round to integer) and plug it into NA cell\ncolumn_means

```

```

# Another data set for EDA
data_long <- data |>
  pivot_longer(cols = c(math_score, reading_score, writing_score),
    names_to = "test", values_to = "score")

```

Summary

```

sum_data_fct =
  data |>
  dplyr::select(1:11) |>
  skimr::skim() |>
  dplyr::select(skim_variable, n_missing, complete_rate, factor.n_unique, factor.top_counts)

colnames(sum_data_fct) = c("Variable", "Missing", "Complete Rate", "Unique", "Top Counts")

knitr::kable(x = sum_data_fct, caption = "Categorical Variables pre-analysis", digits = 1)

```

Table 1: Categorical Variables pre-analysis

Variable	Missing	Complete Rate	Unique	Top Counts
gender	0	1.0	2	1: 488, 0: 460
ethnic_group	59	0.9	5	2: 277, 3: 237, 1: 171, 4: 124
parent_educ	392	0.6	4	1: 199, 2: 198, 3: 104, 4: 55
lunch_type	0	1.0	2	0: 617, 1: 331
test_prep	55	0.9	2	0: 571, 1: 322
parent_marital_status	49	0.9	4	0: 516, 1: 213, 3: 146, 2: 24
practice_sport	16	1.0	3	1: 477, 2: 343, 0: 112
is_first_child	30	1.0	2	1: 604, 0: 314
nr_siblings	46	1.0	8	1: 245, 2: 213, 3: 198, 0: 101
transport_means	102	0.9	2	0: 509, 1: 337
wkly_study_hours	37	1.0	3	1: 508, 0: 253, 2: 150

```

data =
  data |>
  drop_na()

sum_data_score =
  data |>
  dplyr::select(12:14) |>
  skimr::skim() |>
  dplyr::select(skim_variable, numeric.mean, numeric.sd, numeric.p0, numeric.p25, numeric.p50, numeric.p75, numeric.max)

colnames(sum_data_score) = c("Variable", "Mean", "SD", "Min", "Q1", "Median", "Q3", "Max")

knitr::kable(x = sum_data_score, caption = "Continuous Variables pre-analysis", digits = 1)

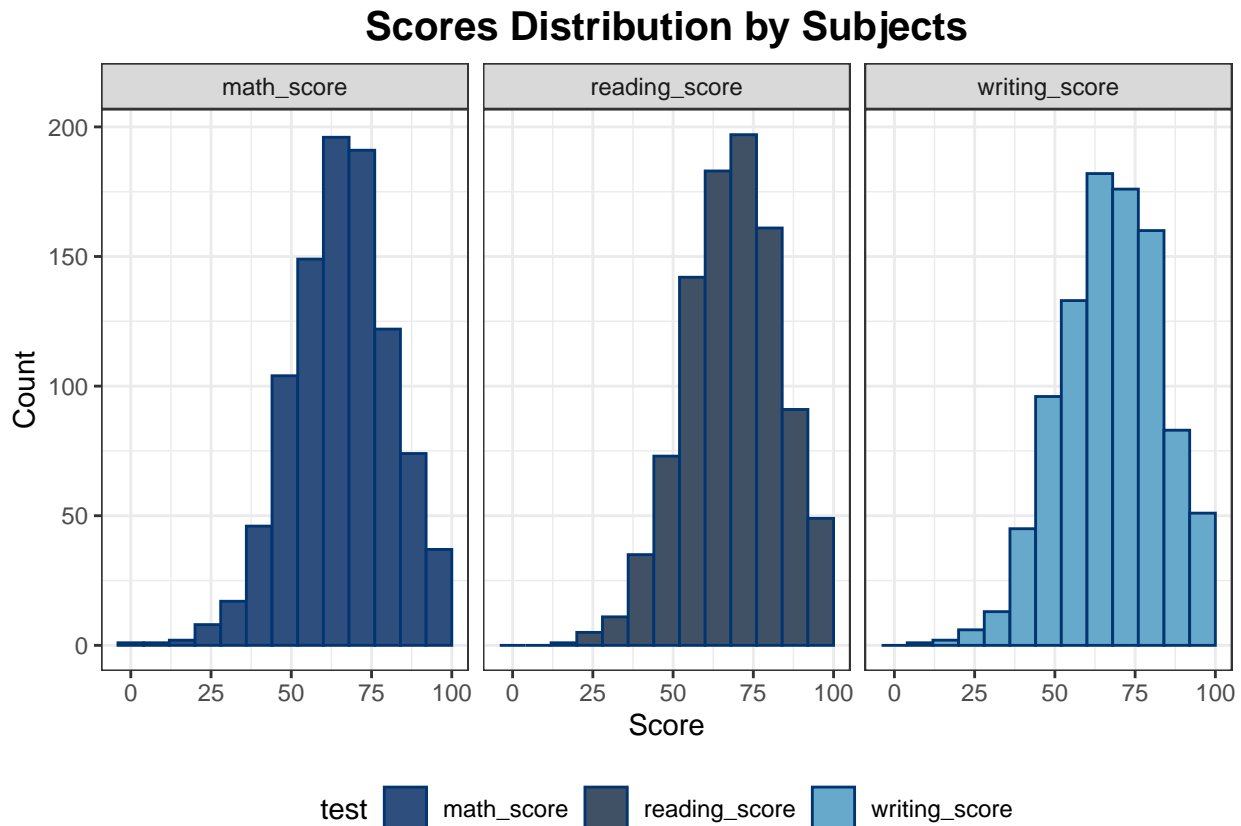
```

Table 2: Continuous Variables pre-analysis

Variable	Mean	SD	Min	Q1	Median	Q3	Max
math_score	68.7	15.9	18	57	69.0	81	100
reading_score	72.3	14.8	23	61	73.0	84	100
writing_score	72.0	15.2	19	62	72.5	84	100

Histograms

```
data_long |>
  ggplot(aes(x = score, fill = test)) +
  geom_histogram(binwidth = 8, color = "#013571") +
  labs(
    title = "Scores Distribution by Subjects",
    x = "Score",
    y = "Count"
  ) +
  scale_fill_manual(values = c("#2E4E7D", "#405165", "#67A9CB")) +
  facet_grid(~ test) +
  theme_bw() +
  theme(legend.position = "bottom") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5))
```



Boxplots

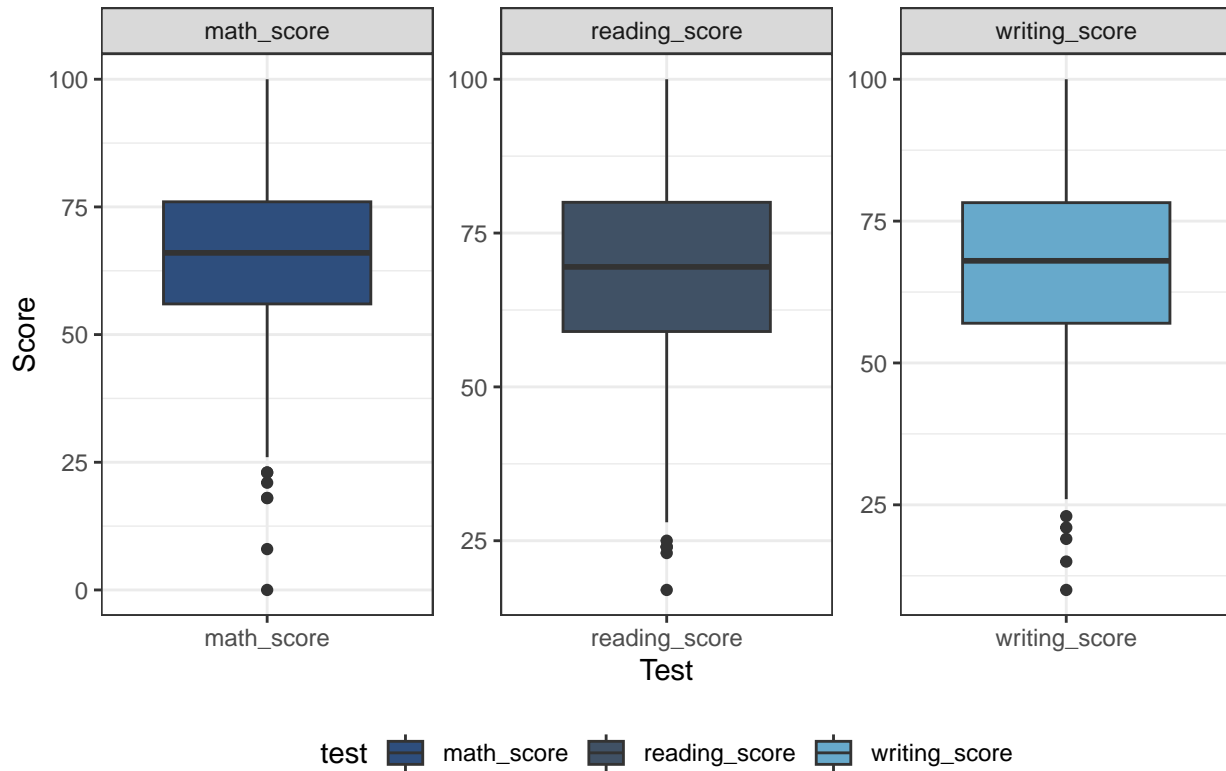
```
data_long |>
  ggplot(aes(x = test, y = score, fill = test)) +
  geom_boxplot() +
  labs(
    title = "Scores Boxplot by Subjects",
```

```

x = "Test",
y = "Score"
) +
facet_wrap(~ test, scales = "free") +
scale_fill_manual(values = c("#2E4E7D", "#405165", "#67A9CB")) +
theme_bw() +
theme(legend.position = "bottom") +
theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5))

```

Scores Boxplot by Subjects



Diagnostics

```

# Math
model_math_full = lm(math_score ~ . - reading_score - writing_score, data = data)
model_math_full |> summary()

```

```

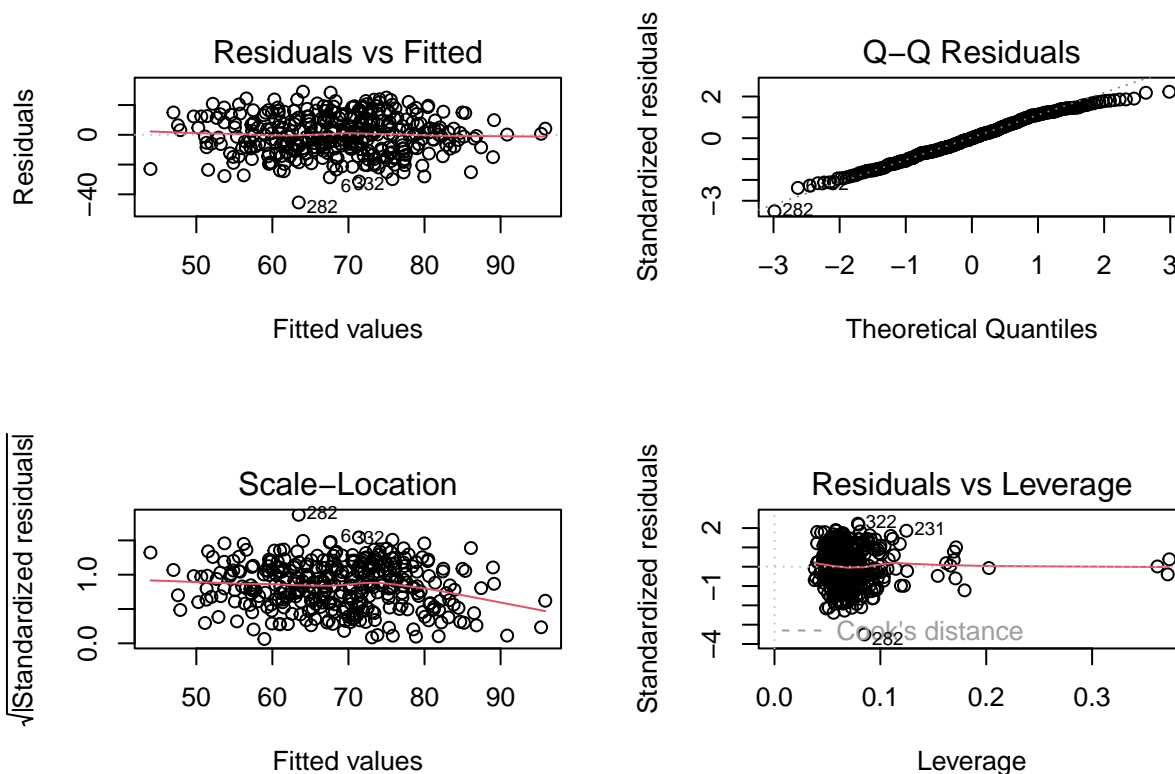
##
## Call:
## lm(formula = math_score ~ . - reading_score - writing_score,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.458  -8.961   0.089   9.800  28.981

```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      62.3523     4.9540  12.586 < 2e-16 ***
## gender1          -3.6522     1.4958  -2.442 0.015150 *
## ethnic_group1      1.8120     3.2790   0.553 0.580912
## ethnic_group2     -1.1247     3.1319  -0.359 0.719748
## ethnic_group3      3.0342     3.1826   0.953 0.341109
## ethnic_group4      8.7423     3.3555   2.605 0.009598 **
## parent_educ2       1.8031     1.7975   1.003 0.316545
## parent_educ3       3.1775     2.0927   1.518 0.129886
## parent_educ4       4.0051     2.5782   1.553 0.121282
## lunch_type1      -12.1275     1.5423  -7.863 5.49e-14 ***
## test_prep1        5.7990     1.5706   3.692 0.000260 ***
## parent_marital_status1 -4.2006     1.8079  -2.323 0.020770 *
## parent_marital_status2  7.0930     4.7226   1.502 0.134083
## parent_marital_status3 -4.8362     2.1726  -2.226 0.026694 *
## practice_sport1     3.0566     2.3818   1.283 0.200295
## practice_sport2     3.2296     2.4896   1.297 0.195466
## is_first_child1    -0.3254     1.6378  -0.199 0.842638
## nr_siblings1       -0.1780     2.7665  -0.064 0.948739
## nr_siblings2       -1.1446     2.8721  -0.399 0.690507
## nr_siblings3        3.1546     2.8049   1.125 0.261548
## nr_siblings4        2.8587     3.3920   0.843 0.399963
## nr_siblings5        2.4937     3.9289   0.635 0.526071
## nr_siblings6       14.5158    13.9723   1.039 0.299617
## nr_siblings7        9.5593     8.3433   1.146 0.252735
## transport_means1    1.0585     1.5640   0.677 0.499003
## wkly_study_hours1   6.4822     1.7525   3.699 0.000254 ***
## wkly_study_hours2   4.2523     2.2536   1.887 0.060065 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.55 on 327 degrees of freedom
## Multiple R-squared:  0.3256, Adjusted R-squared:  0.272
## F-statistic: 6.073 on 26 and 327 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(model_math_full)
```

```
## Warning: not plotting observations with leverage one:
##      186
```



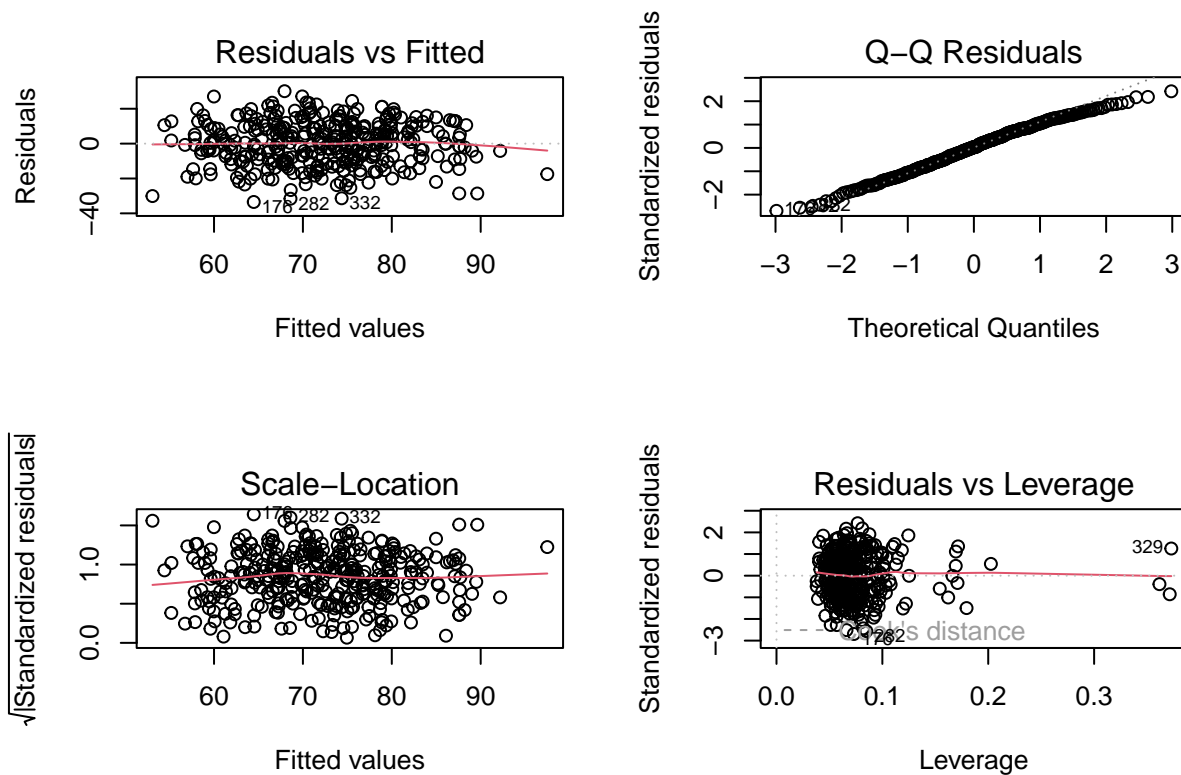
```
# Reading
model_reading_full = lm(reading_score ~ . - math_score - writing_score, data = data)
model_reading_full |> summary()
```

```
##
## Call:
## lm(formula = reading_score ~ . - math_score - writing_score,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.470  -8.942   0.403   9.553  30.063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    59.3627     4.7169  12.585 < 2e-16 ***
## gender1         8.2587     1.4242   5.799 1.57e-08 ***
## ethnic_group1    1.4533     3.1220   0.466 0.64188
## ethnic_group2   -0.5044     2.9819  -0.169 0.86578
## ethnic_group3    2.8080     3.0302   0.927 0.35479
## ethnic_group4    4.7359     3.1949   1.482 0.13921
## parent_educ2     2.6502     1.7114   1.549 0.12246
## parent_educ3     4.5816     1.9925   2.299 0.02211 *
## parent_educ4     6.4240     2.4548   2.617 0.00929 **
## lunch_type1     -7.8783     1.4685  -5.365 1.54e-07 ***
## test_prep1       7.6036     1.4954   5.085 6.21e-07 ***
## parent_marital_status1 -4.6412     1.7214  -2.696 0.00738 **
## parent_marital_status2  4.6364     4.4966   1.031 0.30325
## parent_marital_status3 -4.2660     2.0686  -2.062 0.03997 *
```

```
## practice_sport1      1.9156      2.2678      0.845      0.39890
## practice_sport2      1.2989      2.3705      0.548      0.58408
## is_first_child1      0.6384      1.5594      0.409      0.68252
## nr_siblings1         0.4794      2.6341      0.182      0.85569
## nr_siblings2        -1.4869      2.7347     -0.544      0.58700
## nr_siblings3         1.8958      2.6706      0.710      0.47830
## nr_siblings4         2.3345      3.2296      0.723      0.47028
## nr_siblings5        -1.4797      3.7408     -0.396      0.69269
## nr_siblings6        11.7473     13.3034      0.883      0.37787
## nr_siblings7         7.7275      7.9439      0.973      0.33139
## transport_means1     0.5365      1.4891      0.360      0.71890
## wkly_study_hours1     5.3310      1.6686      3.195      0.00154 **
## wkly_study_hours2     1.1401      2.1458      0.531      0.59557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.9 on 327 degrees of freedom
## Multiple R-squared:  0.2971, Adjusted R-squared:  0.2412
## F-statistic: 5.315 on 26 and 327 DF,  p-value: 6.451e-14
```

```
par(mfrow = c(2,2))
plot(model_reading_full)
```

```
## Warning: not plotting observations with leverage one:
##      186
```



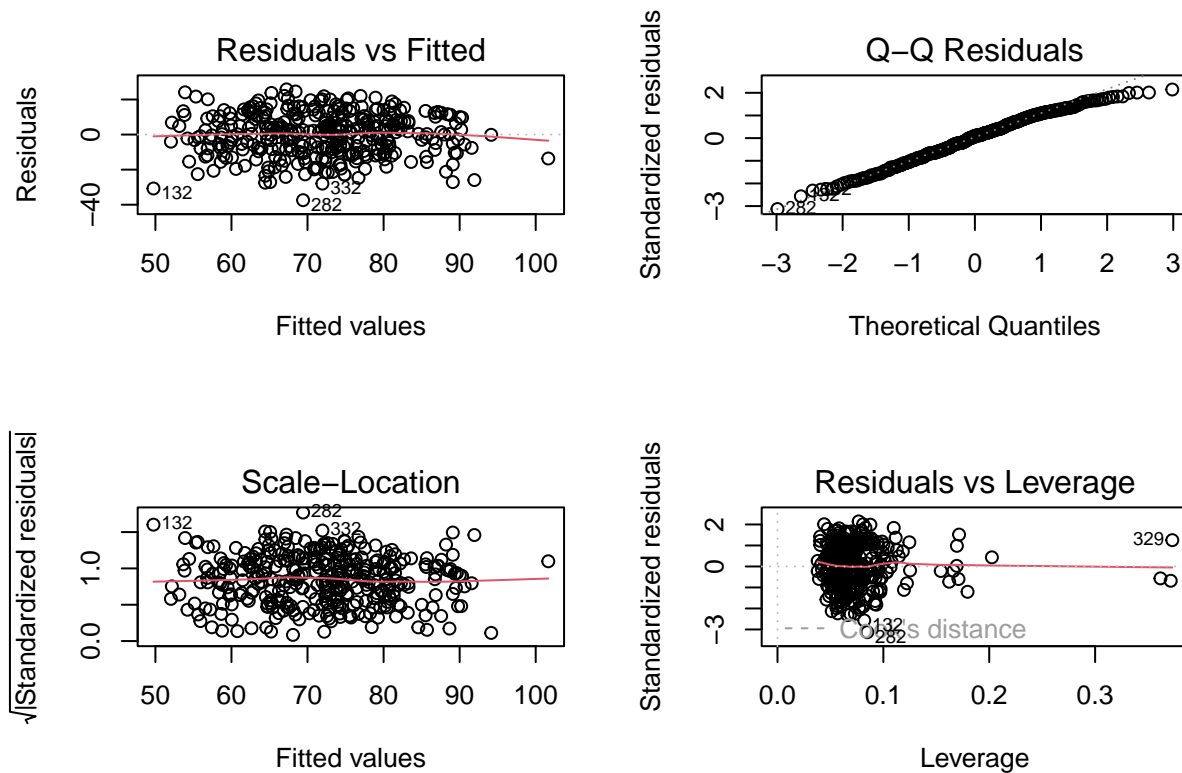

```
# Writing
model_writing_full = lm(writing_score ~ . - reading_score - math_score, data = data)
model_writing_full |> summary()
```

```
##
## Call:
## lm(formula = writing_score ~ . - reading_score - math_score,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37.416  -8.131   1.123   9.165  25.765
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    55.1871     4.5675  12.083 < 2e-16 ***
## gender1        10.0433     1.3791   7.283 2.46e-12 ***
## ethnic_group1    1.7982     3.0232   0.595 0.552382
## ethnic_group2    0.7708     2.8875   0.267 0.789684
## ethnic_group3    5.5577     2.9343   1.894 0.059101 .
## ethnic_group4    5.5666     3.0937   1.799 0.072893 .
## parent_educ2     2.0224     1.6572   1.220 0.223203
## parent_educ3     4.5673     1.9294   2.367 0.018507 *
## parent_educ4     7.5525     2.3771   3.177 0.001629 **
## lunch_type1     -8.9424     1.4220  -6.289 1.03e-09 ***
## test_prep1       9.6428     1.4480   6.659 1.16e-10 ***
## parent_marital_status1 -4.5781     1.6669  -2.747 0.006356 **
## parent_marital_status2  5.2451     4.3542   1.205 0.229221
## parent_marital_status3 -4.4305     2.0031  -2.212 0.027669 *
## practice_sport1    3.3011     2.1960   1.503 0.133746
## practice_sport2    3.0186     2.2954   1.315 0.189415
## is_first_child1   -0.2525     1.5100  -0.167 0.867295
## nr_siblings1       0.3186     2.5507   0.125 0.900665
## nr_siblings2      -1.2993     2.6481  -0.491 0.624008
## nr_siblings3       2.2515     2.5860   0.871 0.384594
## nr_siblings4       2.9536     3.1273   0.944 0.345630
## nr_siblings5      -0.5419     3.6224  -0.150 0.881167
## nr_siblings6      14.3830    12.8821   1.117 0.265024
## nr_siblings7       8.0232     7.6923   1.043 0.297708
## transport_means1   0.9938     1.4420   0.689 0.491208
## wkly_study_hours1  5.4344     1.6157   3.363 0.000861 ***
## wkly_study_hours2  2.0335     2.0778   0.979 0.328454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.5 on 327 degrees of freedom
## Multiple R-squared:  0.3762, Adjusted R-squared:  0.3266
## F-statistic: 7.586 on 26 and 327 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2,2))
plot(model_writing_full)
```

```
## Warning: not plotting observations with leverage one:
```

186



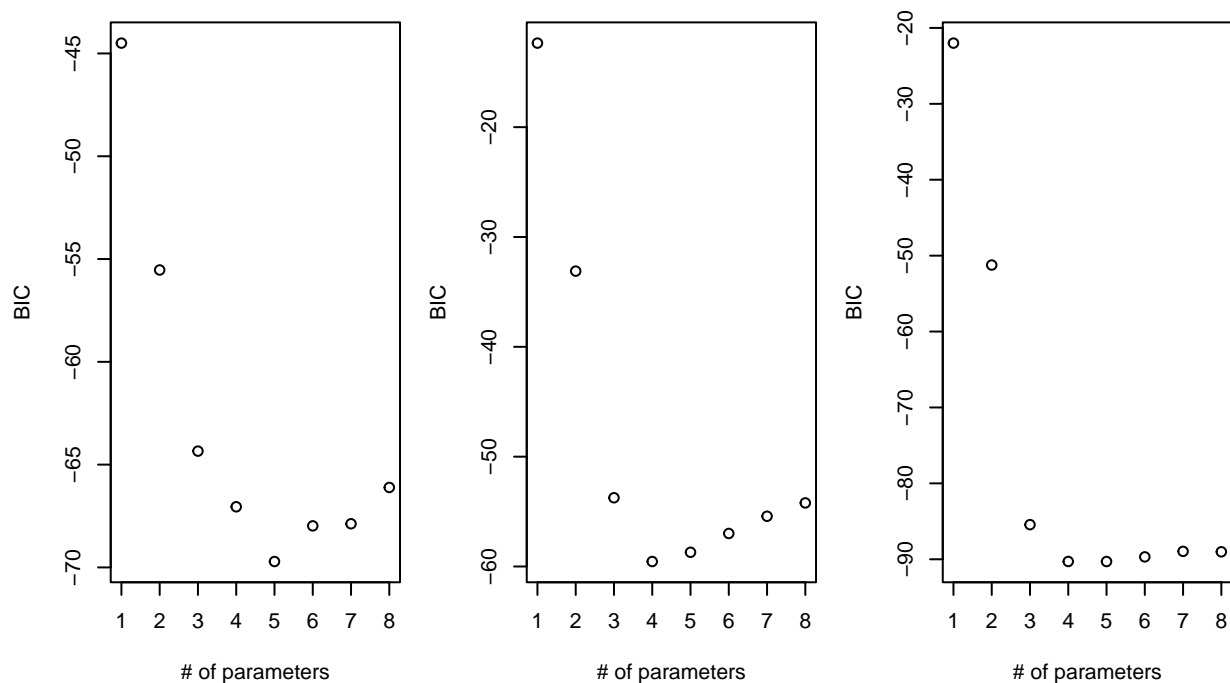
Criterion-based Procedures

```
math_c = regsubsets(math_score ~ . - reading_score - writing_score, data = data)
res_math =
  math_c |>
  summary()

reading_c = regsubsets(reading_score ~ . - math_score - writing_score, data = data)
res_reading =
  reading_c |>
  summary()

writing_c = regsubsets(writing_score ~ . - math_score - reading_score, data = data)
res_writing =
  writing_c |>
  summary()

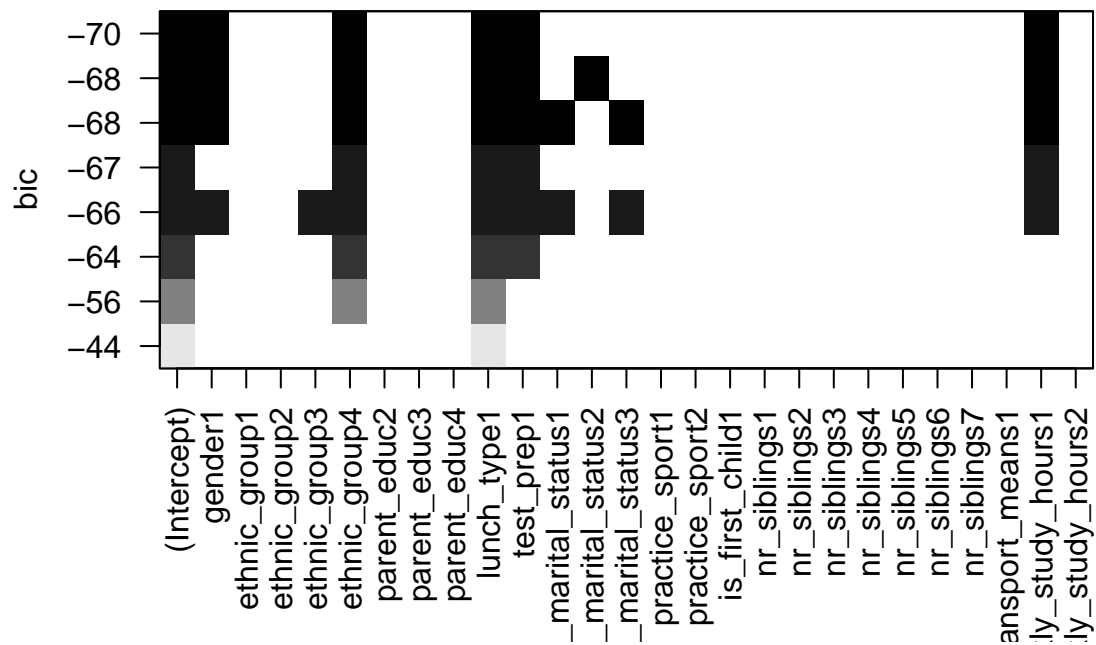
par(mfrow = c(1, 3), mar = c(8, 4, 4, 1))
plot(1:8, res_math$bic, xlab = "# of parameters", ylab = "BIC")
plot(1:8, res_reading$bic, xlab = "# of parameters", ylab = "BIC")
plot(1:8, res_writing$bic, xlab = "# of parameters", ylab = "BIC")
```



```
par(mfrow = c(1, 1))
res_math$outmat[5,]
```

```
##          gender1          ethnic_group1          ethnic_group2
##          "*"           " "           " "
##      ethnic_group3      ethnic_group4      parent_educ2
##          " "           "*"           " "
##      parent_educ3      parent_educ4      lunch_type1
##          " "           " "           "*"
##      test_prep1 parent_marital_status1 parent_marital_status2
##          "*"           " "           " "
## parent_marital_status3      practice_sport1      practice_sport2
##          " "           " "           " "
##      is_first_child1      nr_siblings1      nr_siblings2
##          " "           " "           " "
##      nr_siblings3      nr_siblings4      nr_siblings5
##          " "           " "           " "
##      nr_siblings6      nr_siblings7      transport_means1
##          " "           " "           " "
##      wkly_study_hours1      wkly_study_hours2
##          "*"           " "
```

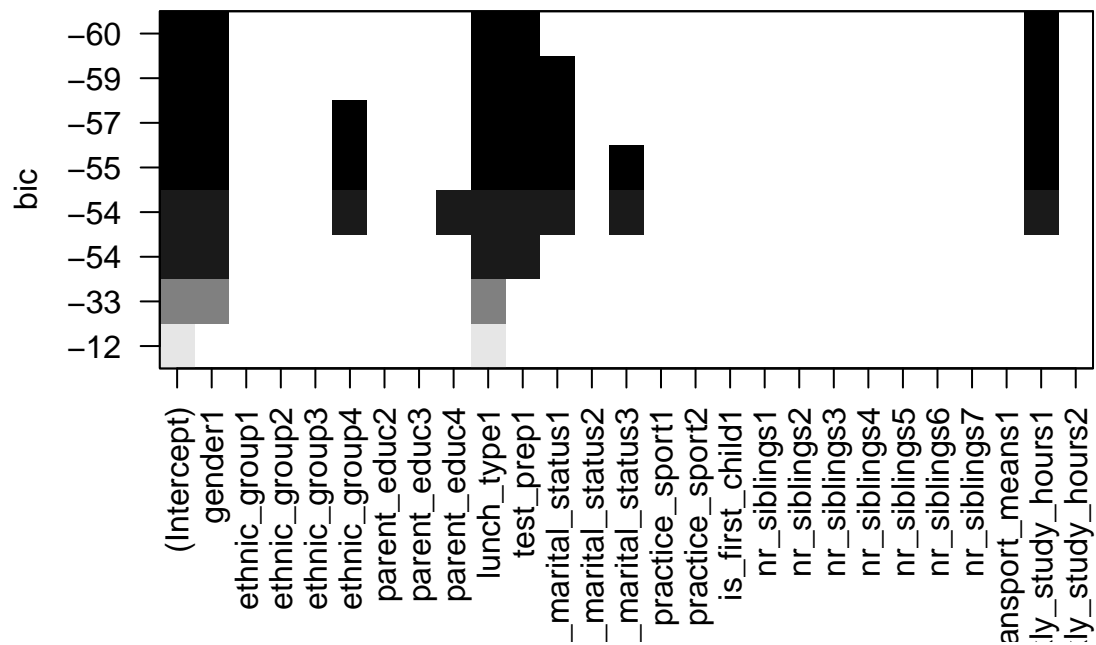
```
plot(math_c, scale = "bic")
```



```
res_reading$outmat[5,]
```

```
##          gender1          ethnic_group1          ethnic_group2
##          "*"          " "          " "
##          ethnic_group3          ethnic_group4          parent_educ2
##          " "          " "          " "
##          parent_educ3          parent_educ4          lunch_type1
##          " "          " "          "*"
##          test_prep1 parent_marital_status1 parent_marital_status2
##          "*"          "*"          " "
## parent_marital_status3          practice_sport1          practice_sport2
##          " "          " "          " "
##          is_first_child1          nr_siblings1          nr_siblings2
##          " "          " "          " "
##          nr_siblings3          nr_siblings4          nr_siblings5
##          " "          " "          " "
##          nr_siblings6          nr_siblings7          transport_means1
##          " "          " "          " "
##          wkly_study_hours1          wkly_study_hours2
##          "*"          " "
```

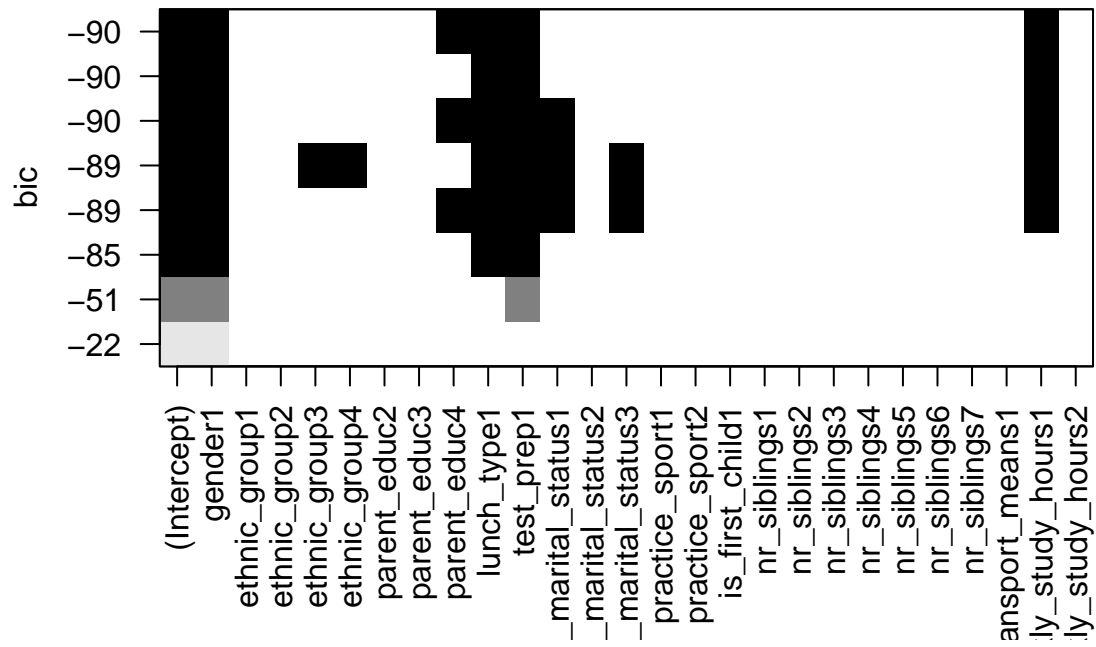
```
plot(reading_c, scale = "bic")
```



```
res_writing$outmat[5,]
```

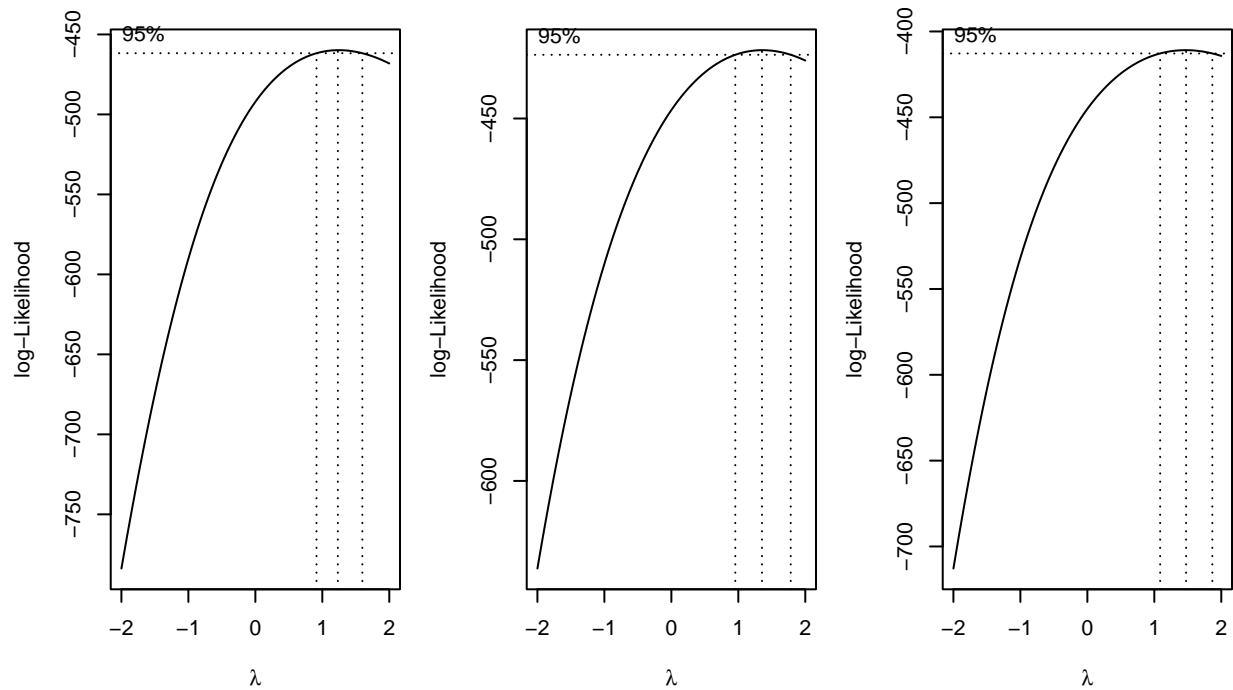
```
##          gender1          ethnic_group1          ethnic_group2
##          "*"            " "            " "
##          ethnic_group3          ethnic_group4          parent_educ2
##          " "            " "            " "
##          parent_educ3          parent_educ4          lunch_type1
##          " "            "*"            "*"
##          test_prep1 parent_marital_status1 parent_marital_status2
##          "*"            " "            " "
## parent_marital_status3          practice_sport1          practice_sport2
##          " "            " "            " "
##          is_first_child1          nr_siblings1          nr_siblings2
##          " "            " "            " "
##          nr_siblings3          nr_siblings4          nr_siblings5
##          " "            " "            " "
##          nr_siblings6          nr_siblings7          transport_means1
##          " "            " "            " "
##          wkly_study_hours1          wkly_study_hours2
##          "*"            " "
```

```
plot(writing_c, scale = "bic")
```



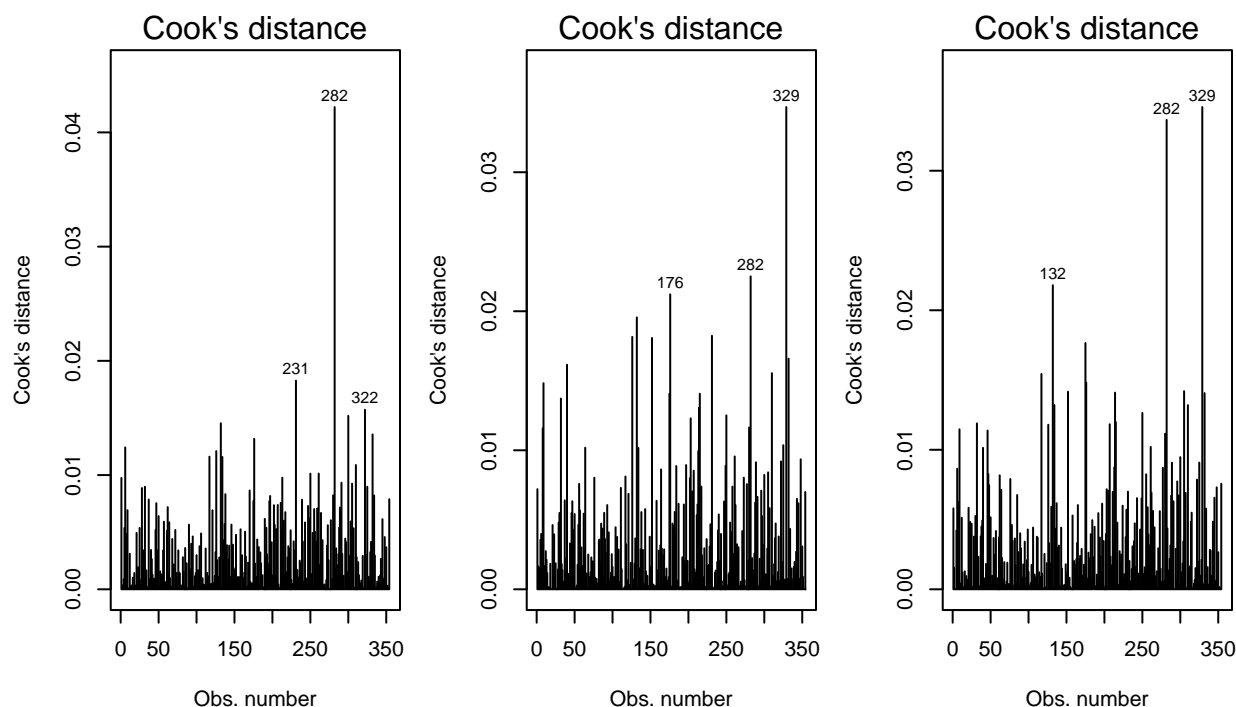
Transformation

```
par(mfrow = c(1, 3), mar = c(8, 4, 4, 1))
boxcox(model_math_full)
boxcox(model_reading_full)
boxcox(model_writing_full)
```



Outlier and influence points

```
par(mfrow = c(1, 3), mar = c(8, 4, 4, 1))
plot(model_math_full, which = 4)
plot(model_reading_full, which = 4)
plot(model_writing_full, which = 4)
```



Multicollinearity

```
# check VIF
vif_math =
  performance::check_collinearity(model_math_full) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_math, caption = "VIF for Math Score", digits = 1)
```

Table 3: VIF for Math Score

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.4]	0.9
ethnic_group	1.2	[1.1, 1.4]	0.8
parent_educ	1.2	[1.1, 1.4]	0.8
lunch_type	1.1	[1, 1.4]	1.0
test_prep	1.1	[1, 1.3]	0.9

Term	VIF	VIF_CI	Tolerance
parent_marital_status	1.2	[1.1, 1.4]	0.9
practice_sport	1.2	[1.1, 1.4]	0.9
is_first_child	1.2	[1.1, 1.3]	0.9
nr_siblings	1.5	[1.4, 1.8]	0.6
transport_means	1.1	[1, 1.3]	0.9
wkly_study_hours	1.1	[1.1, 1.3]	0.9

```
vif_reading =
  performance::check_collinearity(model_reading_full) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_reading, caption = "VIF for Reading Score", digits = 1)
```

Table 4: VIF for Reading Score

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.4]	0.9
ethnic_group	1.2	[1.1, 1.4]	0.8
parent_educ	1.2	[1.1, 1.4]	0.8
lunch_type	1.1	[1, 1.4]	1.0
test_prep	1.1	[1, 1.3]	0.9
parent_marital_status	1.2	[1.1, 1.4]	0.9
practice_sport	1.2	[1.1, 1.4]	0.9
is_first_child	1.2	[1.1, 1.3]	0.9
nr_siblings	1.5	[1.4, 1.8]	0.6
transport_means	1.1	[1, 1.3]	0.9
wkly_study_hours	1.1	[1.1, 1.3]	0.9

```
vif_writing =
  performance::check_collinearity(model_writing_full) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_writing, caption = "VIF for Reading Score", digits = 1)
```

Table 5: VIF for Reading Score

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.4]	0.9
ethnic_group	1.2	[1.1, 1.4]	0.8
parent_educ	1.2	[1.1, 1.4]	0.8
lunch_type	1.1	[1, 1.4]	1.0
test_prep	1.1	[1, 1.3]	0.9
parent_marital_status	1.2	[1.1, 1.4]	0.9
practice_sport	1.2	[1.1, 1.4]	0.9
is_first_child	1.2	[1.1, 1.3]	0.9
nr_siblings	1.5	[1.4, 1.8]	0.6

Term	VIF	VIF_CI	Tolerance
transport_means	1.1	[1, 1.3]	0.9
wkly_study_hours	1.1	[1.1, 1.3]	0.9

Model Selections

Stepwise Regressions

```
# math
math_sr = step(model_math_full, direction = 'both', trace = FALSE)

res_math_sr = math_sr |>
  summary()
res_math_sr

##
## Call:
## lm(formula = math_score ~ gender + ethnic_group + lunch_type +
##     test_prep + parent_marital_status + wkly_study_hours, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.985  -9.397   0.110  10.638  30.842
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      67.3260     3.2947  20.435 < 2e-16 ***
## gender1          -3.7049     1.4716  -2.518 0.012274 *
## ethnic_group1      2.4461     3.2493   0.753 0.452084
## ethnic_group2      0.3026     3.0827   0.098 0.921866
## ethnic_group3      4.1687     3.1287   1.332 0.183624
## ethnic_group4     10.1791     3.3078   3.077 0.002258 **
## lunch_type1     -12.3773     1.5158  -8.166 6.28e-15 ***
## test_prep1        6.0788     1.5224   3.993 8.00e-05 ***
## parent_marital_status1 -4.0821     1.7731  -2.302 0.021925 *
## parent_marital_status2  6.7982     4.6451   1.464 0.144250
## parent_marital_status3 -5.2507     2.1346  -2.460 0.014398 *
## wkly_study_hours1      5.9171     1.7108   3.459 0.000612 ***
## wkly_study_hours2      3.8301     2.2148   1.729 0.084647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.53 on 341 degrees of freedom
## Multiple R-squared:  0.2989, Adjusted R-squared:  0.2742
## F-statistic: 12.11 on 12 and 341 DF, p-value: < 2.2e-16

tb_math_sr = res_math_sr |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
```

```
dplyr::select(term, estimate, p.value)
colnames(tb_math_sr) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_math_sr, caption = "Math Scores Models by Stepwise Regression", digits = 1)
```

Table 6: Math Scores Models by Stepwise Regression

Term	Estimate	P Value
gender1	-3.7	0.0
ethnic_group1	2.4	0.5
ethnic_group2	0.3	0.9
ethnic_group3	4.2	0.2
ethnic_group4	10.2	0.0
lunch_type1	-12.4	0.0
test_prep1	6.1	0.0
parent_marital_status1	-4.1	0.0
parent_marital_status2	6.8	0.1
parent_marital_status3	-5.3	0.0
wkly_study_hours1	5.9	0.0
wkly_study_hours2	3.8	0.1

```
# reading
rea_sr = step(model_reading_full, direction = 'both')

## Start:  AIC=1836.68
## reading_score ~ (gender + ethnic_group + parent_educ + lunch_type +
##   test_prep + parent_marital_status + practice_sport + is_first_child +
##   nr_siblings + transport_means + wkly_study_hours + math_score +
##   writing_score) - math_score - writing_score
##
##           Df Sum of Sq  RSS    AIC
## - nr_siblings      7    887.9 55342 1828.4
## - practice_sport     2    123.8 54578 1833.5
## - transport_means     1     21.6 54476 1834.8
## - is_first_child      1     27.9 54482 1834.9
## - ethnic_group       4    1227.5 55682 1836.6
## <none>                        54454 1836.7
## - parent_educ        3    1558.4 56013 1840.7
## - parent_marital_status 3    1908.7 56363 1842.9
## - wkly_study_hours    2    2004.0 56459 1845.5
## - test_prep           1    4305.6 58760 1861.6
## - lunch_type          1    4793.1 59248 1864.5
## - gender              1    5599.8 60054 1869.3
##
## Step:  AIC=1828.41
## reading_score ~ gender + ethnic_group + parent_educ + lunch_type +
##   test_prep + parent_marital_status + practice_sport + is_first_child +
##   transport_means + wkly_study_hours
##
##           Df Sum of Sq  RSS    AIC
## - practice_sport     2    145.3 55488 1825.3
## - transport_means     1     11.4 55354 1826.5
```

```

## - is_first_child      1      40.1 55382 1826.7
## <none>                  55342 1828.4
## - ethnic_group        4     1318.8 56661 1828.7
## - parent_educ          3     1681.4 57024 1833.0
## - parent_marital_status 3     1924.1 57267 1834.5
## + nr_siblings          7       887.9 54454 1836.7
## - wkly_study_hours     2     1969.7 57312 1836.8
## - test_prep            1     4222.4 59565 1852.4
## - lunch_type           1     5437.8 60780 1859.6
## - gender               1     5693.8 61036 1861.1
##
## Step:  AIC=1825.34
## reading_score ~ gender + ethnic_group + parent_educ + lunch_type +
##      test_prep + parent_marital_status + is_first_child + transport_means +
##      wkly_study_hours
##
##              Df Sum of Sq  RSS    AIC
## - transport_means      1        5.8 55493 1823.4
## - is_first_child       1       40.9 55529 1823.6
## <none>                   55488 1825.3
## - ethnic_group         4     1294.8 56782 1825.5
## + practice_sport       2       145.3 55342 1828.4
## - parent_educ          3     1654.8 57143 1829.7
## - parent_marital_status 3     1902.9 57391 1831.3
## + nr_siblings          7       909.4 54578 1833.5
## - wkly_study_hours     2     1959.0 57447 1833.6
## - test_prep            1     4316.3 59804 1849.8
## - lunch_type           1     5421.7 60909 1856.3
## - gender               1     5678.5 61166 1857.8
##
## Step:  AIC=1823.37
## reading_score ~ gender + ethnic_group + parent_educ + lunch_type +
##      test_prep + parent_marital_status + is_first_child + wkly_study_hours
##
##              Df Sum of Sq  RSS    AIC
## - is_first_child       1       39.4 55533 1821.6
## <none>                   55493 1823.4
## - ethnic_group         4     1295.8 56789 1823.5
## + transport_means      1        5.8 55488 1825.3
## + practice_sport       2       139.7 55354 1826.5
## - parent_educ          3     1649.4 57143 1827.7
## - parent_marital_status 3     1899.1 57393 1829.3
## + nr_siblings          7       901.8 54592 1831.6
## - wkly_study_hours     2     1958.5 57452 1831.7
## - test_prep            1     4422.7 59916 1848.5
## - lunch_type           1     5422.5 60916 1854.4
## - gender               1     5674.9 61168 1855.8
##
## Step:  AIC=1821.62
## reading_score ~ gender + ethnic_group + parent_educ + lunch_type +
##      test_prep + parent_marital_status + wkly_study_hours
##
##              Df Sum of Sq  RSS    AIC
## <none>                   55533 1821.6

```

```
## - ethnic_group      4      1305.9 56839 1821.8
## + is_first_child    1        39.4 55493 1823.4
## + transport_means    1         4.3 55529 1823.6
## + practice_sport     2       140.5 55392 1824.7
## - parent_educ        3      1654.8 57188 1826.0
## - parent_marital_status 3      1899.5 57432 1827.5
## + nr_siblings        7       917.5 54615 1829.7
## - wkly_study_hours   2      1974.9 57508 1830.0
## - test_prep          1     4531.6 60064 1847.4
## - lunch_type         1     5440.2 60973 1852.7
## - gender             1     5644.2 61177 1853.9
```

```
res_rea_sr = rea_sr |>
  summary()
res_rea_sr
```

```
##
## Call:
## lm(formula = reading_score ~ gender + ethnic_group + parent_educ +
##     lunch_type + test_prep + parent_marital_status + wkly_study_hours,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.083  -9.288   0.232   9.530  30.265
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    61.6474     3.2387  19.035 < 2e-16 ***
## gender1         8.1816     1.3959   5.861 1.09e-08 ***
## ethnic_group1    1.8945     3.0850   0.614 0.53956
## ethnic_group2    0.3778     2.9216   0.129 0.89720
## ethnic_group3    3.3789     2.9670   1.139 0.25559
## ethnic_group4    5.6870     3.1395   1.811 0.07096 .
## parent_educ2     2.3964     1.6650   1.439 0.15101
## parent_educ3     4.6728     1.9527   2.393 0.01726 *
## parent_educ4     6.4917     2.3912   2.715 0.00697 **
## lunch_type1     -8.2631     1.4360  -5.754 1.95e-08 ***
## test_prep1       7.6175     1.4505   5.252 2.67e-07 ***
## parent_marital_status1 -4.5976     1.6839  -2.730 0.00666 **
## parent_marital_status2  4.1841     4.4058   0.950 0.34296
## parent_marital_status3 -4.3042     2.0287  -2.122 0.03460 *
## wkly_study_hours1  5.1565     1.6242   3.175 0.00164 **
## wkly_study_hours2  1.0458     2.1132   0.495 0.62102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.82 on 338 degrees of freedom
## Multiple R-squared:  0.2831, Adjusted R-squared:  0.2513
## F-statistic: 8.9 on 15 and 338 DF, p-value: < 2.2e-16
```

```
tb_rea_sr = res_rea_sr |>
  broom::tidy() |>
```

```

filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_rea_sr) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_rea_sr, caption = "Reading Scores Models by Stepwise Regression", digits = 1)

```

Table 7: Reading Scores Models by Stepwise Regression

Term	Estimate	P Value
gender1	8.2	0.0
ethnic_group1	1.9	0.5
ethnic_group2	0.4	0.9
ethnic_group3	3.4	0.3
ethnic_group4	5.7	0.1
parent_educ2	2.4	0.2
parent_educ3	4.7	0.0
parent_educ4	6.5	0.0
lunch_type1	-8.3	0.0
test_prep1	7.6	0.0
parent_marital_status1	-4.6	0.0
parent_marital_status2	4.2	0.3
parent_marital_status3	-4.3	0.0
wkly_study_hours1	5.2	0.0
wkly_study_hours2	1.0	0.6

```

# writing
wri_sr = step(model_writing_full, direction = 'both')

## Start:  AIC=1813.9
## writing_score ~ (gender + ethnic_group + parent_educ + lunch_type +
##   test_prep + parent_marital_status + practice_sport + is_first_child +
##   nr_siblings + transport_means + wkly_study_hours + math_score +
##   reading_score) - reading_score - math_score
##
##           Df Sum of Sq  RSS    AIC
## - nr_siblings      7   1019.1 52079 1806.9
## - is_first_child    1     4.4 51064 1811.9
## - practice_sport    2    361.2 51421 1812.4
## - transport_means   1     74.2 51134 1812.4
## <none>                        51060 1813.9
## - ethnic_group      4    1779.1 52839 1818.0
## - parent_educ       3    1940.3 53000 1821.1
## - parent_marital_status 3    1991.7 53052 1821.4
## - wkly_study_hours  2    1901.4 52961 1822.8
## - lunch_type        1    6175.3 57235 1852.3
## - test_prep         1    6924.6 57985 1856.9
## - gender            1    8281.3 59341 1865.1
##
## Step:  AIC=1806.89
## writing_score ~ gender + ethnic_group + parent_educ + lunch_type +
##   test_prep + parent_marital_status + practice_sport + is_first_child +
##   transport_means + wkly_study_hours

```

```

##
##           Df Sum of Sq  RSS    AIC
## - is_first_child      1      1.2 52080 1804.9
## - transport_means      1     52.4 52132 1805.2
## - practice_sport       2    404.8 52484 1805.6
## <none>                  52079 1806.9
## - ethnic_group         4    1870.2 53949 1811.4
## + nr_siblings          7    1019.1 51060 1813.9
## - parent_marital_status 3    2027.5 54107 1814.4
## - parent_educ          3    2069.1 54148 1814.7
## - wkly_study_hours     2    1830.3 53910 1815.1
## - test_prep            1    6879.5 58959 1848.8
## - lunch_type           1    6955.3 59035 1849.3
## - gender               1    8444.0 60523 1858.1
##
## Step:  AIC=1804.9
## writing_score ~ gender + ethnic_group + parent_educ + lunch_type +
##      test_prep + parent_marital_status + practice_sport + transport_means +
##      wkly_study_hours
##
##           Df Sum of Sq  RSS    AIC
## - transport_means      1     53.0 52133 1803.3
## - practice_sport       2    408.3 52489 1803.7
## <none>                  52080 1804.9
## + is_first_child       1      1.2 52079 1806.9
## - ethnic_group         4    1869.4 53950 1809.4
## + nr_siblings          7    1015.9 51064 1811.9
## - parent_marital_status 3    2028.9 54109 1812.4
## - parent_educ          3    2068.7 54149 1812.7
## - wkly_study_hours     2    1829.2 53910 1813.1
## - test_prep            1    6907.4 58988 1847.0
## - lunch_type           1    6954.4 59035 1847.3
## - gender               1    8463.2 60544 1856.2
##
## Step:  AIC=1803.26
## writing_score ~ gender + ethnic_group + parent_educ + lunch_type +
##      test_prep + parent_marital_status + practice_sport + wkly_study_hours
##
##           Df Sum of Sq  RSS    AIC
## - practice_sport       2    397.6 52531 1802.0
## <none>                  52133 1803.3
## + transport_means      1     53.0 52080 1804.9
## + is_first_child       1      1.8 52132 1805.2
## - ethnic_group         4    1901.9 54035 1808.0
## + nr_siblings          7     992.6 51141 1810.5
## - parent_marital_status 3    1986.8 54120 1810.5
## - parent_educ          3    2041.4 54175 1810.9
## - wkly_study_hours     2    1821.0 53954 1811.4
## - lunch_type           1    6905.0 59038 1845.3
## - test_prep            1    7190.9 59324 1847.0
## - gender               1    8443.2 60577 1854.4
##
## Step:  AIC=1801.95
## writing_score ~ gender + ethnic_group + parent_educ + lunch_type +

```

```
##      test_prep + parent_marital_status + wkly_study_hours
##
##              Df Sum of Sq  RSS    AIC
## <none>                    52531 1802.0
## + practice_sport         2    397.6 52133 1803.3
## + transport_means        1     42.3 52489 1803.7
## + is_first_child         1      6.3 52525 1803.9
## - ethnic_group           4   1950.7 54482 1806.9
## - parent_educ            3   1925.8 54457 1808.7
## + nr_siblings            7   1035.0 51496 1808.9
## - parent_marital_status  3   1962.6 54494 1808.9
## - wkly_study_hours       2   1804.0 54335 1809.9
## - lunch_type             1   6837.1 59368 1843.3
## - test_prep              1   7210.3 59741 1845.5
## - gender                 1   8486.0 61017 1853.0
```

```
res_wri_sr = wri_sr |>
  summary()
res_wri_sr
```

```
##
## Call:
## lm(formula = writing_score ~ gender + ethnic_group + parent_educ +
##      lunch_type + test_prep + parent_marital_status + wkly_study_hours,
##      data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.350  -8.531   0.962   9.592  25.874
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    58.522     3.150  18.579 < 2e-16 ***
## gender1        10.032     1.358   7.389 1.17e-12 ***
## ethnic_group1    2.213     3.000   0.738  0.46124
## ethnic_group2    1.850     2.842   0.651  0.51544
## ethnic_group3    6.338     2.886   2.196  0.02874 *
## ethnic_group4    6.617     3.053   2.167  0.03094 *
## parent_educ      1.789     1.619   1.105  0.27005
## parent_educ3     4.598     1.899   2.421  0.01599 *
## parent_educ4     7.212     2.326   3.101  0.00209 **
## lunch_type1     -9.263     1.397  -6.633 1.31e-10 ***
## test_prep1       9.609     1.411   6.811 4.44e-11 ***
## parent_marital_status1 -4.417     1.638  -2.697  0.00734 **
## parent_marital_status2  4.668     4.285   1.089  0.27678
## parent_marital_status3 -4.644     1.973  -2.353  0.01917 *
## wkly_study_hours1    5.168     1.580   3.271  0.00118 **
## wkly_study_hours2    1.893     2.055   0.921  0.35769
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.47 on 338 degrees of freedom
## Multiple R-squared:  0.3583, Adjusted R-squared:  0.3298
## F-statistic: 12.58 on 15 and 338 DF,  p-value: < 2.2e-16
```

```
tb_wri_sr = res_wri_sr |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_wri_sr) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_wri_sr, caption = "Writing Scores Models by Stepwise Regression", digits = 1)
```

Table 8: Writing Scores Models by Stepwise Regression

Term	Estimate	P Value
gender1	10.0	0.0
ethnic_group1	2.2	0.5
ethnic_group2	1.9	0.5
ethnic_group3	6.3	0.0
ethnic_group4	6.6	0.0
parent_educ2	1.8	0.3
parent_educ3	4.6	0.0
parent_educ4	7.2	0.0
lunch_type1	-9.3	0.0
test_prep1	9.6	0.0
parent_marital_status1	-4.4	0.0
parent_marital_status2	4.7	0.3
parent_marital_status3	-4.6	0.0
wkly_study_hours1	5.2	0.0
wkly_study_hours2	1.9	0.4

Lasso Models

```
lambda_seq = 10 ^ seq(-2, 2, by = .1)

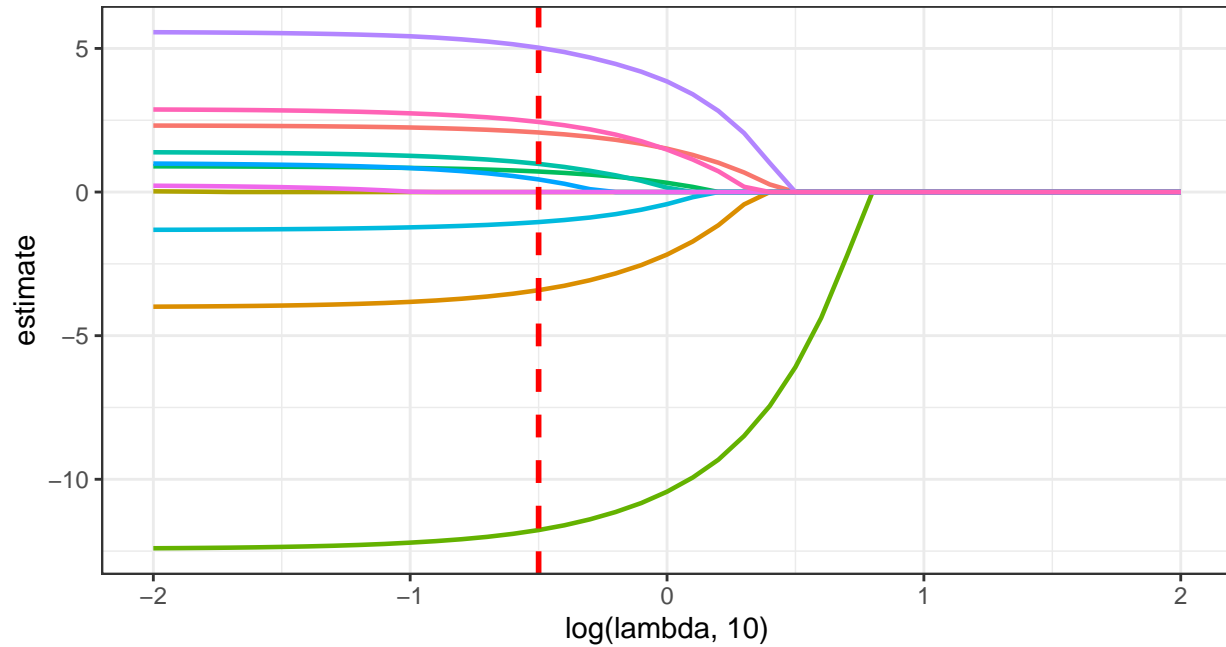
#math
cv_object_math =
  cv.glmnet(as.matrix(data[1:11]), data$math_score,
            lambda = lambda_seq,
            nfolds = 5)

opt_lambda_math = cv_object_math$lambda.min

#variables contraction
glmnet(as.matrix(data[1:11]), data$math_score, lambda = lambda_seq) |>
  broom::tidy() |>
  dplyr::select(term, lambda, estimate) |>
  complete(term, lambda, fill = list(estimate = 0)) |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = log(lambda, 10), y = estimate, group = term, color = term)) +
  geom_path(size = 0.8) +
  geom_vline(xintercept = log(opt_lambda_math, 10), color = "red", linetype = "dashed", size = 1) +
  theme_bw() +
  theme(legend.position = "bottom")
```



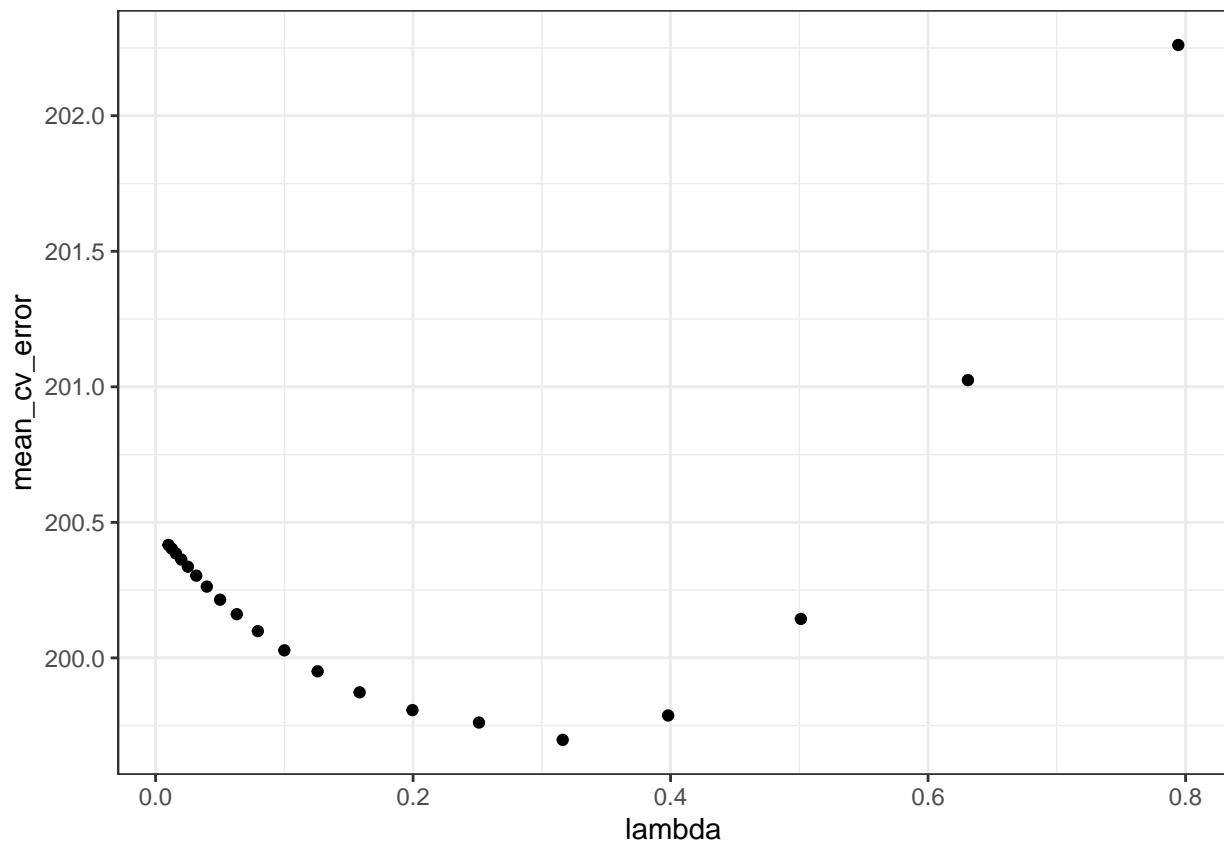
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



	ethnic_group	lunch_type	parent_marital_status	transport_means
term	gender	nr_siblings	practice_sport	wkly_study_hours
	is_first_child	parent_educ	test_prep	

```
tb_la_math = tibble(
  lambda = cv_object_math$lambda,
  mean_cv_error = cv_object_math$cvm) |>
  filter(lambda < 1)

#choosing optimal lambda
tb_la_math |>
  ggplot(aes(x = lambda, y = mean_cv_error)) +
  geom_point() +
  theme_bw()
```



```
#math result
model_math_lasso = glmnet(as.matrix(data[1:11]), data$math_score, lambda = opt_lambda_math, alpha = 1)
coef(model_math_lasso)
```

```
## 12 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept)  62.7158706
## gender       -3.4172517
## ethnic_group  2.0740949
## parent_educ   0.9804808
## lunch_type   -11.7678104
## test_prep     5.0255504
## parent_marital_status -1.0446103
## practice_sport  0.4391390
## is_first_child .
## nr_siblings    0.7146589
## transport_means .
## wkly_study_hours 2.4395500
```

```
model_math_lasso$dev.ratio
```

```
## [1] 0.2622201
```

```
#reading
cv_object_reading =
```

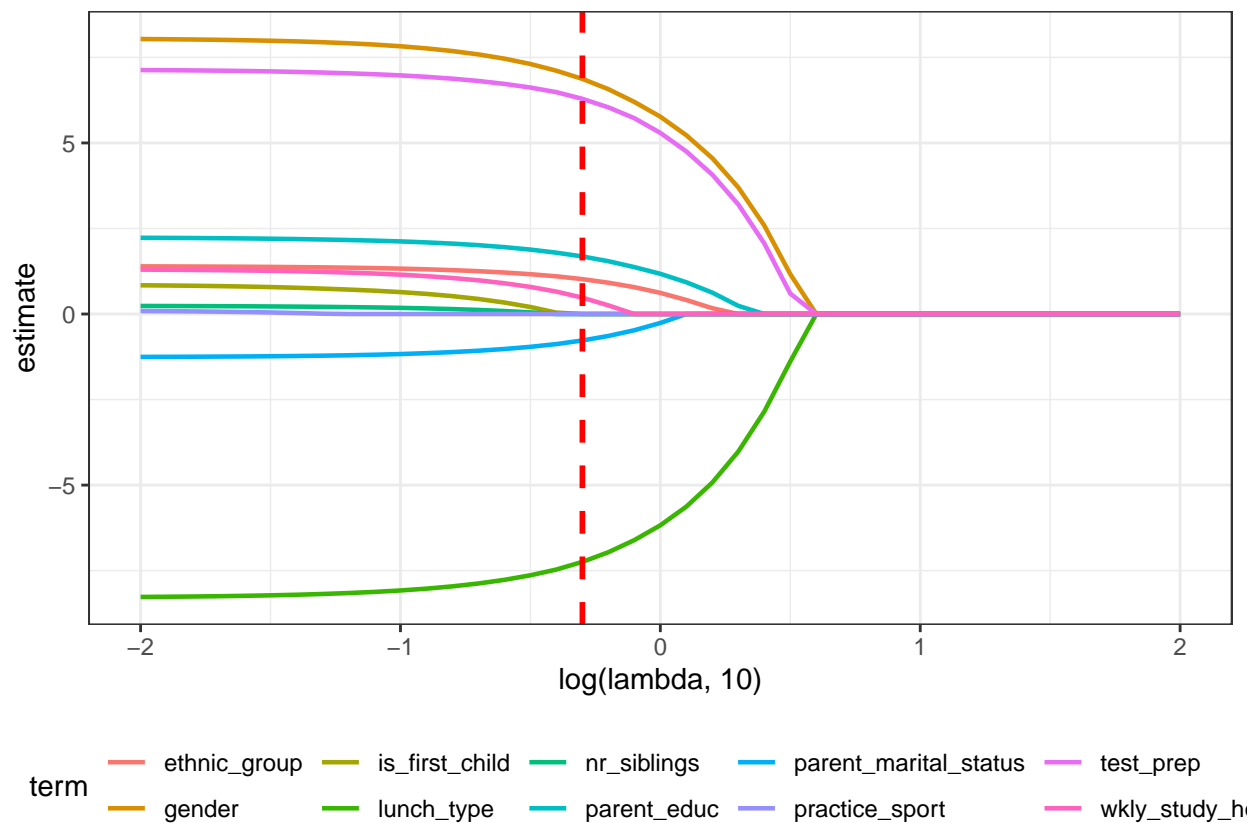
```

cv.glmnet(as.matrix(data[1:11]), data$reading_score,
          lambda = lambda_seq,
          nfolds = 5)

opt_lambda_reading = cv_object_reading$lambda.min

#variables contraction
glmnet(as.matrix(data[1:11]), data$reading_score, lambda = lambda_seq) |>
  broom::tidy() |>
  dplyr::select(term, lambda, estimate) |>
  complete(term, lambda, fill = list(estimate = 0) ) |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = log(lambda, 10), y = estimate, group = term, color = term)) +
  geom_path(size = 0.8) +
  geom_vline(xintercept = log(opt_lambda_reading, 10), color = "red", linetype = "dashed", size = 1) +
  theme_bw() +
  theme(legend.position = "bottom")

```



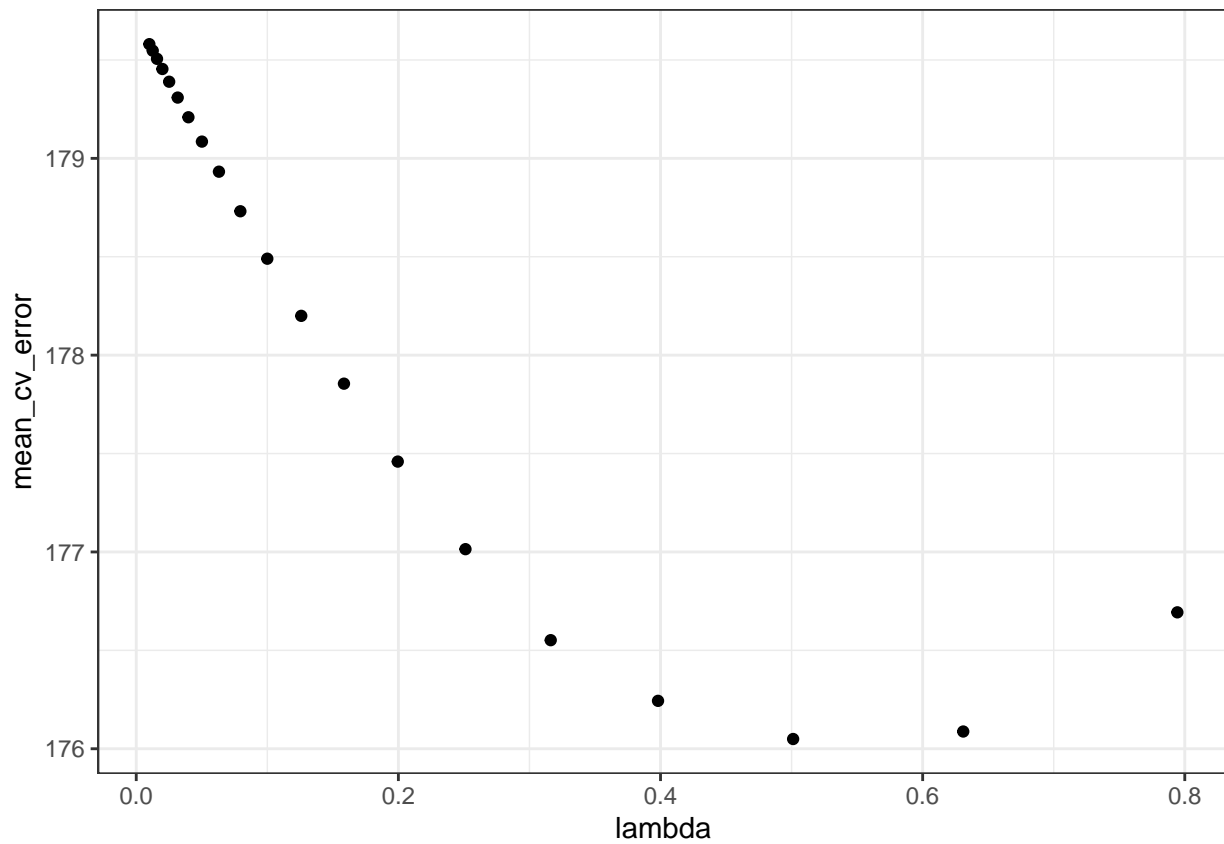
```

tb_la_reading = tibble(
  lambda = cv_object_reading$lambda,
  mean_cv_error = cv_object_reading$cvm) |>
  filter(lambda < 1)

#choosing optimal lambda
tb_la_reading |>
  ggplot(aes(x = lambda, y = mean_cv_error)) +

```

```
geom_point() +  
theme_bw()
```



```
#reading result
```

```
model_reading_lasso = glmnet(as.matrix(data[1:11]), data$reading_score, lambda = opt_lambda_math, alpha  
coef(model_reading_lasso)
```

```
## 12 x 1 sparse Matrix of class "dgCMatrix"  
##              s0  
## (Intercept)  61.64060428  
## gender       7.30735439  
## ethnic_group 1.16352168  
## parent_educ  1.88265624  
## lunch_type   -7.63626242  
## test_prep     6.61888731  
## parent_marital_status -0.95887140  
## practice_sport .  
## is_first_child 0.20510870  
## nr_siblings    0.04774974  
## transport_means .  
## wkly_study_hours 0.79132994
```

```
model_reading_lasso$dev.ratio
```

```
## [1] 0.2357325
```

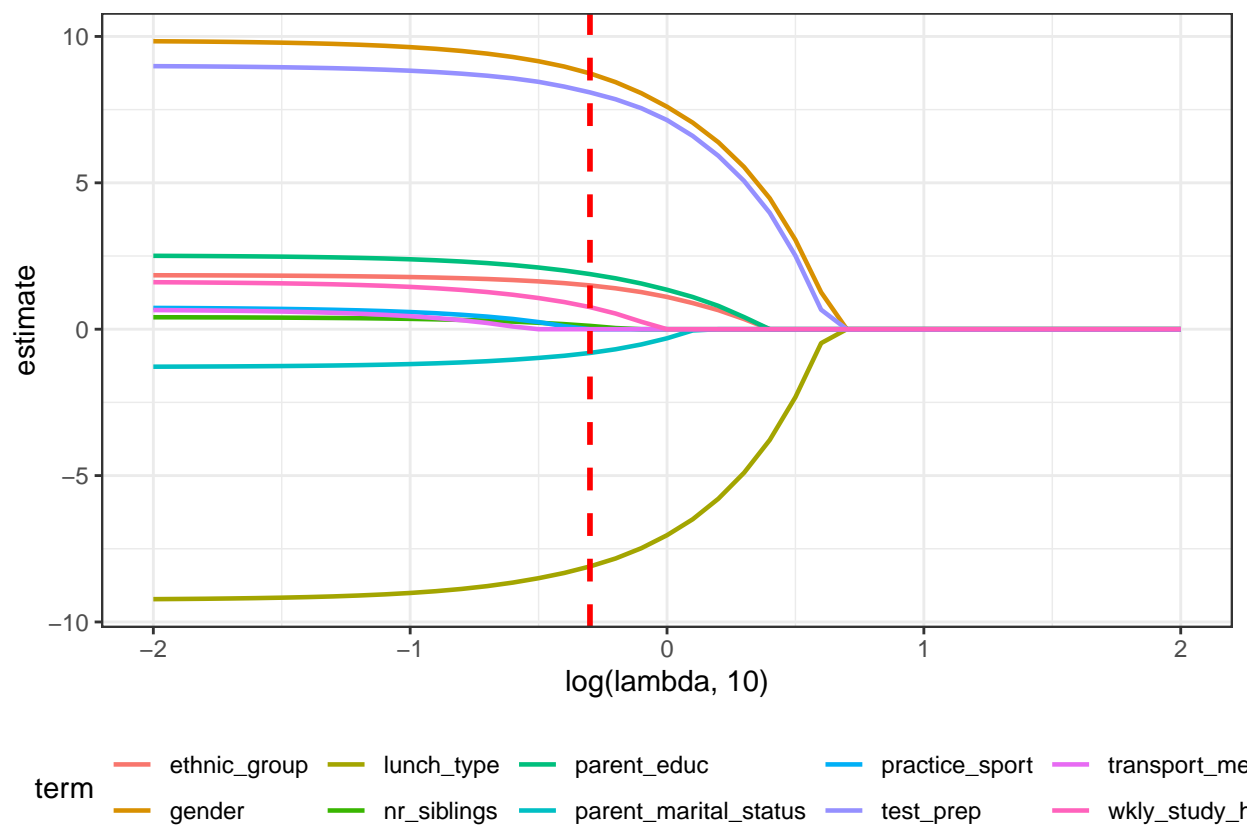
```

#writing
cv_object_writing =
  cv.glmnet(as.matrix(data[1:11]), data$writing_score,
            lambda = lambda_seq,
            nfolds = 5)

opt_lambda_writing = cv_object_writing$lambda.min

#variables contraction
glmnet(as.matrix(data[1:11]), data$writing_score, lambda = lambda_seq) |>
  broom::tidy() |>
  dplyr::select(term, lambda, estimate) |>
  complete(term, lambda, fill = list(estimate = 0)) |>
  filter(term != "(Intercept)") |>
  ggplot(aes(x = log(lambda, 10), y = estimate, group = term, color = term)) +
  geom_path(size = 0.8) +
  geom_vline(xintercept = log(opt_lambda_writing, 10), color = "red", linetype = "dashed", size = 1) +
  theme_bw() +
  theme(legend.position = "bottom")

```



```

tb_la_writing = tibble(
  lambda = cv_object_writing$lambda,
  mean_cv_error = cv_object_writing$cvm) |>
  filter(lambda < 1)

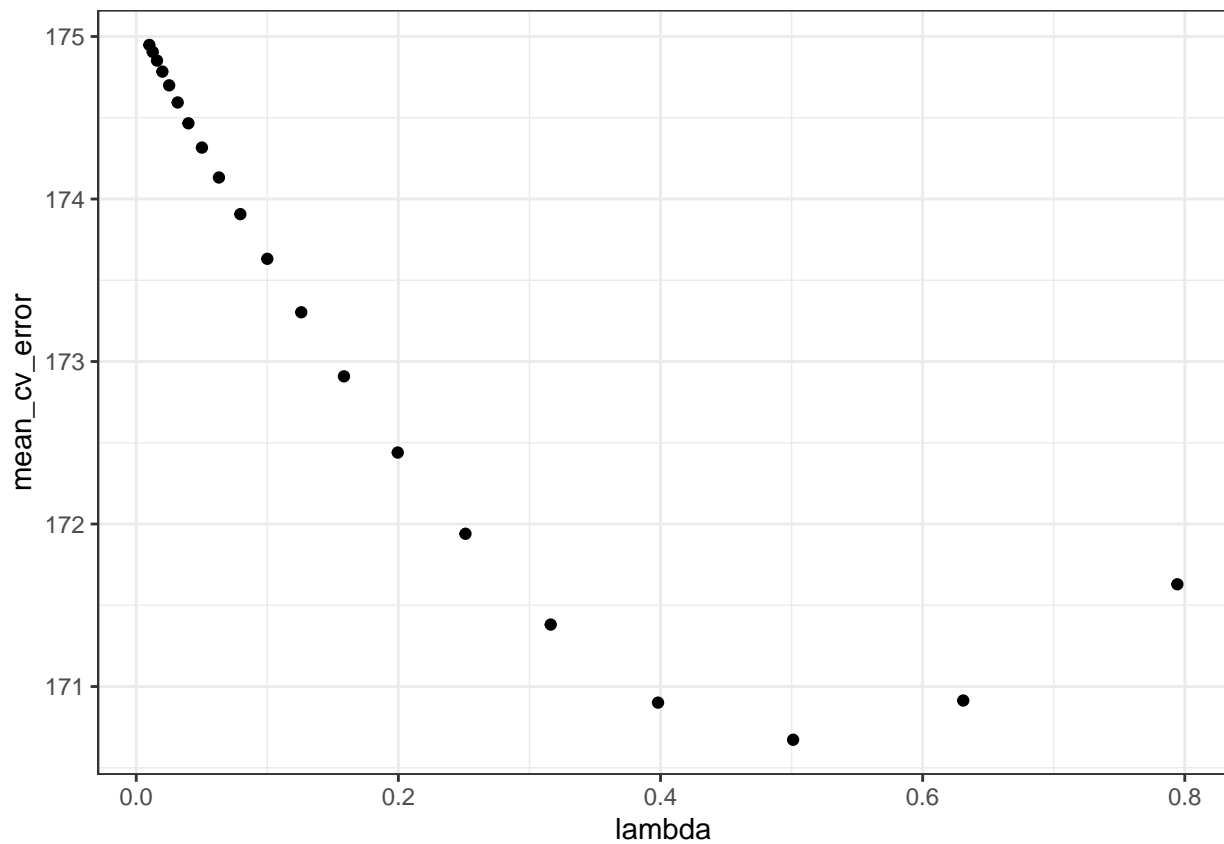
```

```

#choosing optimal lambda
tb_la_writing |>

```

```
ggplot(aes(x = lambda, y = mean_cv_error)) +
  geom_point() +
  theme_bw()
```



```
#writing result
model_writing_lasso = glmnet(as.matrix(data[1:11]), data$writing_score, lambda = opt_lambda_writing, alpha = 0.1)
coef(model_writing_lasso)
```

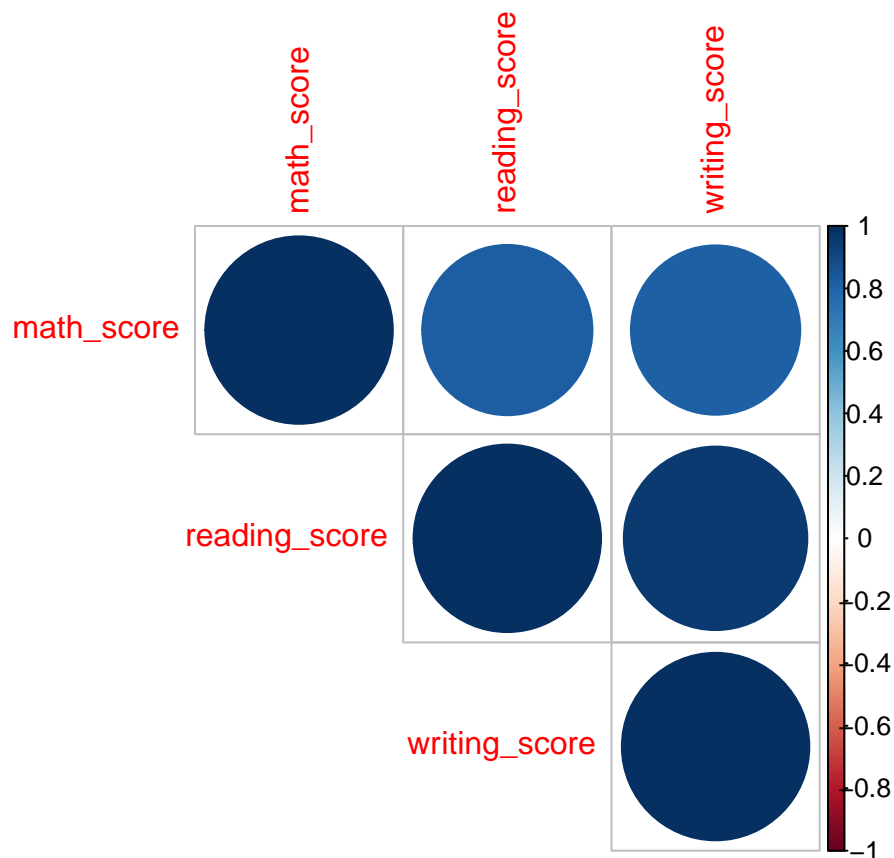
```
## 12 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept)  59.3844759
## gender       8.7384396
## ethnic_group  1.4955961
## parent_educ   1.8826016
## lunch_type   -8.1037819
## test_prep     8.0886240
## parent_marital_status -0.8123378
## practice_sport .
## is_first_child .
## nr_siblings   0.1133873
## transport_means .
## wkly_study_hours 0.7539334
```

```
model_writing_lasso$dev.ratio
```

```
## [1] 0.3119987
```

Adding one score variable

```
#correlation plot for scores  
corrplot::corrplot(cor(data[12:14]), type = "upper")
```



```
# Math  
math_reading = lm(math_score ~ . - writing_score, data = data) |>  
  step(direction = "both", trace = FALSE)  
  
math_reading_sum =  
  math_reading |>  
  summary()  
  
tb_math_reading =  
  math_reading_sum |>  
  broom::tidy() |>  
  filter(term != "(Intercept)") |>  
  dplyr::select(term, estimate, p.value)  
colnames(tb_math_reading) = c("Term", "Estimate", "P Value")  
knitr::kable(x = tb_math_reading, caption = "Math Scores Models Using Reading Score as Additional Prediction")
```

Table 9: Math Scores Models Using Reading Score as Additional Predictor

Term	Estimate	P Value
gender1	-11.5	0.0
ethnic_group1	0.6	0.7
ethnic_group2	-0.2	0.9
ethnic_group3	0.7	0.6
ethnic_group4	4.6	0.0
lunch_type1	-4.7	0.0
test_prep1	-1.2	0.1
practice_sport1	1.4	0.2
practice_sport2	2.4	0.0
wkly_study_hours1	1.4	0.1
wkly_study_hours2	3.4	0.0
reading_score	0.9	0.0

```

math_writing = lm(math_score ~ . - reading_score, data = data) |>
  step(direction = "both", trace = FALSE)

math_writing_sum =
  math_writing |>
  summary()

tb_math_writing =
  math_writing_sum |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_math_writing) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_math_writing, caption = "Math Scores Models Using Writing Score as Additional Predictor")

```

Table 10: Math Scores Models Using Writing Score as Additional Predictor

Term	Estimate	P Value
gender1	-13.7	0.0
ethnic_group1	0.1	0.9
ethnic_group2	-1.6	0.2
ethnic_group3	-2.2	0.1
ethnic_group4	3.5	0.0
parent_educ2	-0.1	0.9
parent_educ3	-1.4	0.1
parent_educ4	-3.8	0.0
lunch_type1	-3.1	0.0
test_prep1	-3.7	0.0
wkly_study_hours1	1.0	0.2
wkly_study_hours2	2.3	0.0
writing_score	1.0	0.0


```

# Reading
reading_math = lm(reading_score ~ . - writing_score, data = data) |>
  step(direction = "both", trace = FALSE)

reading_math_sum =
  reading_math |>
  summary()

tb_reading_math =
  reading_math_sum |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_reading_math) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_reading_math, caption = "Reading Scores Models Using Math Score as Additional Predictor")

```

Table 11: Reading Scores Models Using Math Score as Additional Predictor

Term	Estimate	P Value
gender1	11.3	0.0
ethnic_group1	-0.2	0.9
ethnic_group2	0.2	0.9
ethnic_group3	-0.1	1.0
ethnic_group4	-2.9	0.0
parent_educ2	1.0	0.2
parent_educ3	2.1	0.0
parent_educ4	3.5	0.0
lunch_type1	2.2	0.0
test_prep1	2.5	0.0
is_first_child1	1.1	0.1
wkly_study_hours1	0.0	1.0
wkly_study_hours2	-2.4	0.0
math_score	0.8	0.0

```

reading_writing = lm(reading_score ~ . - math_score, data = data) |>
  step(direction = "both", trace = FALSE)

reading_writing_sum =
  reading_writing |>
  summary()

tb_reading_writing =
  reading_writing_sum |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_reading_writing) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_reading_writing, caption = "Reading Scores Models Using Writing Score as Additional Predictor")

```

Table 12: Reading Scores Models Using Writing Score as Additional Predictor

Term	Estimate	P Value
gender1	-1.6	0.0
ethnic_group1	-0.2	0.8
ethnic_group2	-1.3	0.2
ethnic_group3	-2.7	0.0
ethnic_group4	-0.7	0.5
lunch_type1	0.8	0.1
test_prep1	-1.9	0.0
practice_sport1	-1.2	0.1
practice_sport2	-1.5	0.0
is_first_child1	0.8	0.1
writing_score	1.0	0.0

```
# Writing
writing_math = lm(writing_score ~ . - reading_score, data = data) |>
  step(direction = "both", trace = FALSE)

writing_math_sum =
  writing_math |>
  summary()

tb_writing_math =
  writing_math_sum |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_writing_math) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_writing_math, caption = "Writing Scores Models Using Math Score as Additional Predictor")
```

Table 13: Writing Scores Models Using Math Score as Additional Predictor

Term	Estimate	P Value
gender1	13.1	0.0
ethnic_group1	0.3	0.8
ethnic_group2	1.8	0.1
ethnic_group3	2.9	0.0
ethnic_group4	-1.9	0.1
parent_educ2	0.3	0.6
parent_educ3	2.0	0.0
parent_educ4	4.3	0.0
lunch_type1	1.2	0.1
test_prep1	4.6	0.0
wkly_study_hours1	0.0	0.9
wkly_study_hours2	-1.5	0.1
math_score	0.8	0.0

```
writing_reading = lm(writing_score ~ . - math_score, data = data) |>
  step(direction = "both", trace = FALSE)

writing_reading_sum =
  writing_reading |>
  summary()

tb_writing_reading =
  writing_reading_sum |>
  broom::tidy() |>
  filter(term != "(Intercept)") |>
  dplyr::select(term, estimate, p.value)
colnames(tb_writing_reading) = c("Term", "Estimate", "P Value")
knitr::kable(x = tb_writing_reading, caption = "Writing Scores Models Using Reading Score as Additional Predictor")
```

Table 14: Writing Scores Models Using Reading Score as Additional Predictor

Term	Estimate	P Value
gender1	2.3	0.0
ethnic_group1	0.4	0.6
ethnic_group2	1.4	0.1
ethnic_group3	3.0	0.0
ethnic_group4	1.3	0.2
parent_educ2	-0.5	0.4
parent_educ3	0.3	0.7
parent_educ4	1.3	0.1
lunch_type1	-1.6	0.0
test_prep1	2.7	0.0
practice_sport1	1.5	0.0
practice_sport2	1.8	0.0
is_first_child1	-0.8	0.1
reading_score	0.9	0.0

```
# check VIF
vif_math_reading =
  performance::check_collinearity(math_reading) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_math_reading, caption = "VIF for Math Score (include reading score)", digits = 1)
```

Table 15: VIF for Math Score (include reading score)

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.3]	0.9
ethnic_group	1.1	[1, 1.3]	0.9
lunch_type	1.1	[1, 1.3]	0.9
test_prep	1.1	[1, 1.3]	0.9
practice_sport	1.0	[1, 1.5]	1.0

Term	VIF	VIF_CI	Tolerance
wkly_study_hours	1.1	[1, 1.3]	0.9
reading_score	1.3	[1.2, 1.5]	0.8

```
vif_math_writing =
  performance::check_collinearity(math_writing) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_math_writing, caption = "VIF for Math Score (include writing score)", digits = 1)
```

Table 16: VIF for Math Score (include writing score)

Term	VIF	VIF_CI	Tolerance
gender	1.2	[1.1, 1.4]	0.8
ethnic_group	1.1	[1, 1.3]	0.9
parent_educ	1.1	[1, 1.3]	0.9
lunch_type	1.1	[1.1, 1.4]	0.9
test_prep	1.2	[1.1, 1.4]	0.8
wkly_study_hours	1.1	[1, 1.3]	0.9
writing_score	1.5	[1.3, 1.7]	0.7

```
vif_reading_math =
  performance::check_collinearity(reading_math) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_reading_math, caption = "VIF for Reading Score (include math score)", digits = 1)
```

Table 17: VIF for Reading Score (include math score)

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.4]	1.0
ethnic_group	1.1	[1.1, 1.4]	0.9
parent_educ	1.1	[1, 1.3]	0.9
lunch_type	1.2	[1.1, 1.4]	0.8
test_prep	1.1	[1, 1.3]	0.9
is_first_child	1.0	[1, 3.7]	1.0
wkly_study_hours	1.1	[1, 1.3]	0.9
math_score	1.4	[1.2, 1.6]	0.7

```
vif_reading_writing =
  performance::check_collinearity(reading_writing) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_reading_writing, caption = "VIF for Reading Score (include writing score)", digits = 1)
```

Table 18: VIF for Reading Score (include writing score)

Term	VIF	VIF_CI	Tolerance
gender	1.2	[1.1, 1.4]	0.9
ethnic_group	1.1	[1, 1.3]	0.9
lunch_type	1.1	[1.1, 1.3]	0.9
test_prep	1.2	[1.1, 1.4]	0.9
practice_sport	1.1	[1, 1.4]	0.9
is_first_child	1.0	[1, 1.6]	1.0
writing_score	1.4	[1.3, 1.6]	0.7

```
vif_writing_math =
  performance::check_collinearity(writing_math) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_writing_math, caption = "VIF for Writing Score (include math score)", digits = 1)
```

Table 19: VIF for Writing Score (include math score)

Term	VIF	VIF_CI	Tolerance
gender	1.0	[1, 1.5]	1.0
ethnic_group	1.1	[1.1, 1.3]	0.9
parent_educ	1.1	[1, 1.4]	0.9
lunch_type	1.2	[1.1, 1.4]	0.8
test_prep	1.1	[1, 1.3]	0.9
wkly_study_hours	1.1	[1, 1.3]	0.9
math_score	1.4	[1.2, 1.6]	0.7

```
vif_writing_reading =
  performance::check_collinearity(writing_reading) |>
  as_tibble() |>
  mutate(VIF_CI = str_c("[", round(VIF_CI_low, 1), ", ", round(VIF_CI_high, 1), "]")) |>
  dplyr::select(Term, VIF, VIF_CI, Tolerance)
knitr::kable(x = vif_writing_reading, caption = "VIF for Writing Score (include reading score)", digits = 1)
```

Table 20: VIF for Writing Score (include reading score)

Term	VIF	VIF_CI	Tolerance
gender	1.1	[1, 1.3]	0.9
ethnic_group	1.1	[1, 1.3]	0.9
parent_educ	1.1	[1, 1.3]	0.9
lunch_type	1.1	[1, 1.3]	0.9
test_prep	1.1	[1, 1.3]	0.9
practice_sport	1.1	[1, 1.3]	0.9
is_first_child	1.0	[1, 1.5]	1.0
reading_score	1.3	[1.2, 1.5]	0.8