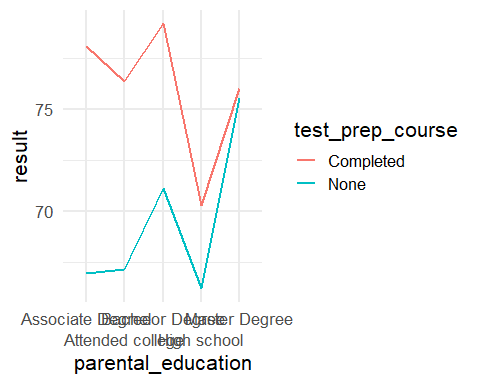
Kovariacinė Analizė

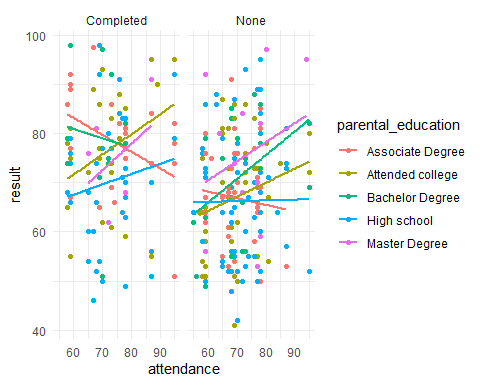
Vainius Gataveckas, Matas Gaulia, Dovydas Martinkus

library(tidyverse)  
library(car)  
library(readxl)  
library(janitor)  
  
x <- readxl::read\_xlsx("HighSchool.xlsx", sheet = 1) %>% clean\_names()

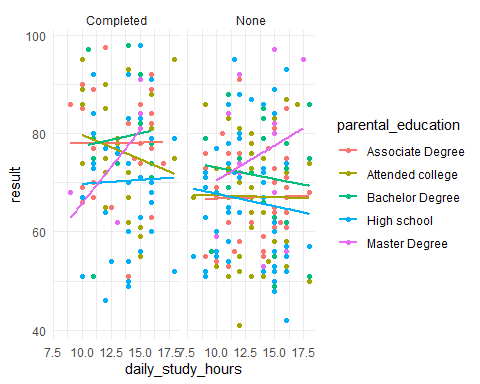
x <- x %>%  
 # sudaromas jungtinis faktorius  
 mutate(combined = factor(paste(parental\_education, test\_prep\_course))) %>%  
 drop\_na()  
  
  
write\_csv(x, "high\_school\_modified.csv")  
  
  
# Tiriamieji grafikai  
  
# Faktorių efektai neatsižvelgiant į kovariantes  
ggplot(x, aes(parental\_education, result, color = test\_prep\_course, group = test\_prep\_course)) +  
 stat\_summary(fun = "mean", geom = "line", size = 1) +  
 theme\_minimal(base\_size = 16) +  
 guides(x = guide\_axis(n.dodge = 2))



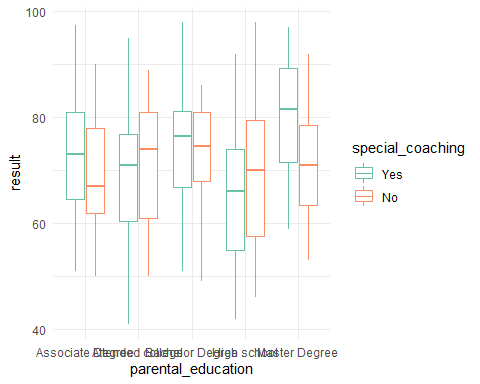
# Hipotezės apie koeficientų lygybę visiems faktorių lygmenims  
ggplot(x, aes(attendance, result, color = parental\_education)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE) +  
 facet\_wrap(vars(test\_prep\_course)) +  
 theme\_minimal()



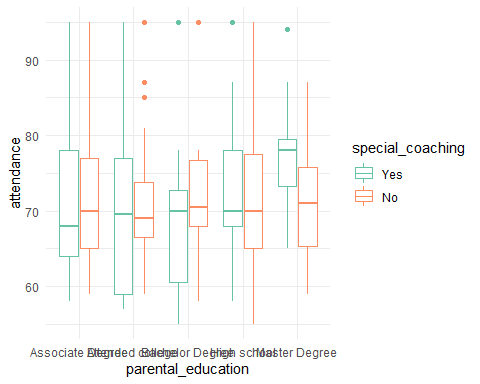
ggplot(x, aes(daily\_study\_hours, result, color = parental\_education)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE) +  
 facet\_wrap(vars(test\_prep\_course)) +  
 theme\_minimal()



# Kovariančių pasiskirstymas pagal faktorių lygmenis  
ggplot(x, aes(x = parental\_education, y = result, color = special\_coaching)) +  
 geom\_boxplot() +  
 theme\_minimal() +  
 scale\_color\_brewer(palette = "Set2")



ggplot(x, aes(x = parental\_education, y = attendance, color = special\_coaching)) +  
 geom\_boxplot() +  
 theme\_minimal() +  
 scale\_color\_brewer(palette = "Set2")



library(rstatix)  
  
# Hipotezė apie koeficientų lygybę neatmetama  
anova\_test(result ~ attendance \* combined + daily\_study\_hours \* combined, data = x, type = 3, detailed = TRUE)

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F p p<.05  
## 1 (Intercept) 6366.834 38499.07 1 264 43.659 2.15e-10 \*  
## 2 attendance 294.059 38499.07 1 264 2.016 1.57e-01   
## 3 combined 1286.732 38499.07 9 264 0.980 4.57e-01   
## 4 daily\_study\_hours 74.310 38499.07 1 264 0.510 4.76e-01   
## 5 attendance:combined 1848.959 38499.07 9 264 1.409 1.84e-01   
## 6 combined:daily\_study\_hours 691.507 38499.07 9 264 0.527 8.55e-01   
## ges  
## 1 0.142  
## 2 0.008  
## 3 0.032  
## 4 0.002  
## 5 0.046  
## 6 0.018

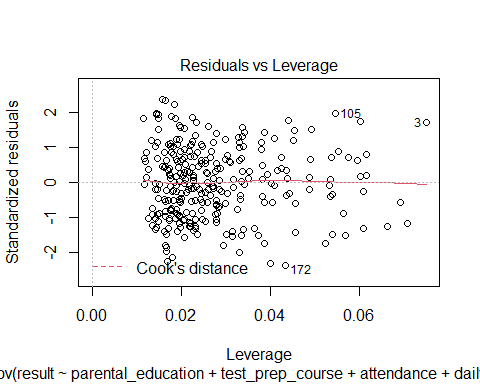
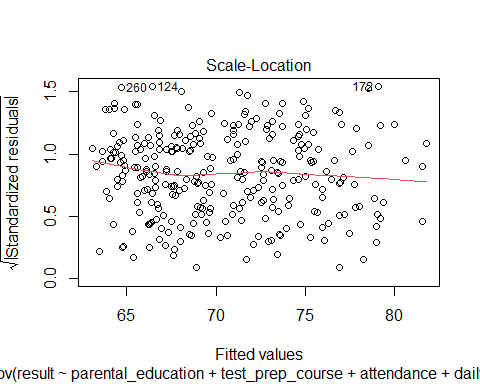
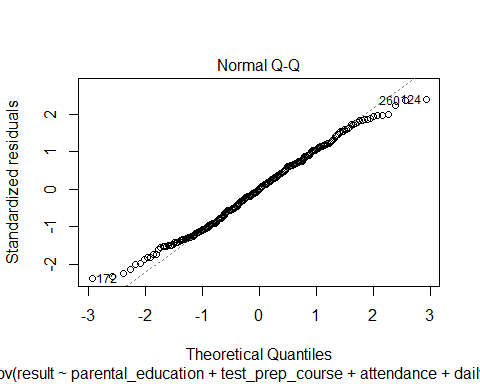
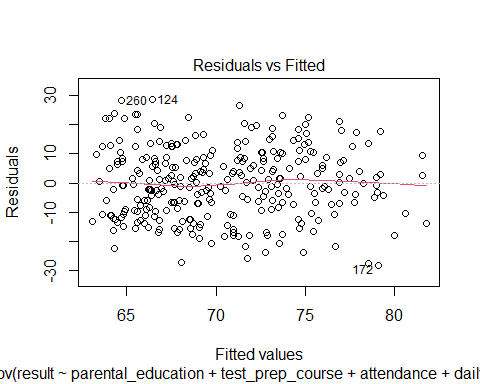
# Hipotezė apie faktorių sąveikos nebuvimą neatmetama  
anova\_test(result ~ attendance + daily\_study\_hours + parental\_education \* test\_prep\_course, data = x, type = 3, detailed = TRUE)

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F  
## 1 (Intercept) 15651.089 41014.03 1 282 107.612  
## 2 attendance 406.149 41014.03 1 282 2.793  
## 3 daily\_study\_hours 63.175 41014.03 1 282 0.434  
## 4 parental\_education 1837.177 41014.03 4 282 3.158  
## 5 test\_prep\_course 1888.974 41014.03 1 282 12.988  
## 6 parental\_education:test\_prep\_course 696.990 41014.03 4 282 1.198  
## p p<.05 ges  
## 1 1.44e-21 \* 0.276  
## 2 9.60e-02 0.010  
## 3 5.10e-01 0.002  
## 4 1.50e-02 \* 0.043  
## 5 3.71e-04 \* 0.044  
## 6 3.12e-01 0.017

model <- anova\_test(result ~ parental\_education + test\_prep\_course + attendance + daily\_study\_hours, data = x, type = 3, detailed = TRUE)  
model

## ANOVA Table (type III tests)  
##   
## Effect SSn SSd DFn DFd F p p<.05 ges  
## 1 (Intercept) 15880.601 41711.02 1 286 108.889 8.20e-22 \* 0.276  
## 2 parental\_education 1832.896 41711.02 4 286 3.142 1.50e-02 \* 0.042  
## 3 test\_prep\_course 3170.732 41711.02 1 286 21.741 4.79e-06 \* 0.071  
## 4 attendance 449.317 41711.02 1 286 3.081 8.00e-02 0.011  
## 5 daily\_study\_hours 89.378 41711.02 1 286 0.613 4.34e-01 0.002

# Modelio prielaidų patikrinimas  
model\_aov <- aov(result ~ parental\_education + test\_prep\_course + attendance + daily\_study\_hours, data = x)  
  
plot(model\_aov)



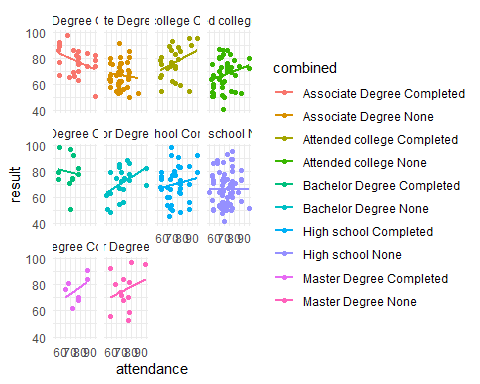
leveneTest(result ~ combined, data = x, center = "mean")

## Levene's Test for Homogeneity of Variance (center = "mean")  
## Df F value Pr(>F)  
## group 9 0.9914 0.4472  
## 284

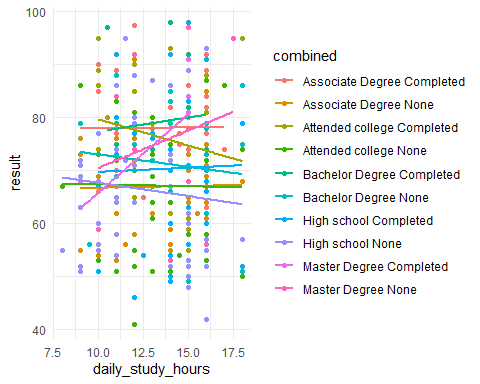
shapiro.test(resid(model\_aov))

##   
## Shapiro-Wilk normality test  
##   
## data: resid(model\_aov)  
## W = 0.99195, p-value = 0.1105

# Tiesinis ryšys tarp kovariančių ir priklausomo kintamojo  
ggplot(x, aes(attendance, result, color = combined)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE) +  
 theme\_minimal() +  
 facet\_wrap(vars(combined))



ggplot(x, aes(daily\_study\_hours, result, color = combined)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", se = FALSE) +  
 theme\_minimal()



library(emmeans)  
  
res <- x %>% emmeans\_test(result ~ parental\_education, model = model\_aov)  
res

## # A tibble: 10 x 9  
## term .y. group1 group2 df statistic p p.adj p.adj.signif  
## \* <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 parental\_education result Assoc~ Atten~ 286 0.248 0.804 1 ns   
## 2 parental\_education result Assoc~ Bache~ 286 -1.25 0.211 1 ns   
## 3 parental\_education result Assoc~ High ~ 286 1.83 0.0678 0.678 ns   
## 4 parental\_education result Assoc~ Maste~ 286 -1.42 0.157 1 ns   
## 5 parental\_education result Atten~ Bache~ 286 -1.46 0.147 1 ns   
## 6 parental\_education result Atten~ High ~ 286 1.58 0.116 1 ns   
## 7 parental\_education result Atten~ Maste~ 286 -1.59 0.113 1 ns   
## 8 parental\_education result Bache~ High ~ 286 2.73 0.00667 0.0667 ns   
## 9 parental\_education result Bache~ Maste~ 286 -0.323 0.747 1 ns   
## 10 parental\_education result High ~ Maste~ 286 -2.65 0.00849 0.0849 ns

get\_emmeans(res)

## # A tibble: 5 x 7  
## parental\_education emmean se df conf.low conf.high method   
## <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 Associate Degree 72.0 1.47 286 69.1 74.9 Emmeans test  
## 2 Attended college 71.5 1.46 286 68.6 74.4 Emmeans test  
## 3 Bachelor Degree 75.3 2.16 286 71.0 79.5 Emmeans test  
## 4 High school 68.6 1.20 286 66.2 70.9 Emmeans test  
## 5 Master Degree 76.4 2.72 286 71.0 81.7 Emmeans test