Question 2 (Amalina)

Ahmad Firdaus Mohamed

2024-02-10

2. Multiple linear regression ; Variables: BMI, fbs as covariates.  Hba1c as outcome  (AMALINA) (?dataset [https://github.com/drkamarul/multivar\_data\_analysis/blob/main/data/metabolic\_syndrome.dta)](https://github.com/drkamarul/multivar_data_analysis/blob/main/data/metabolic_syndrome.dta))

## a. Write regression model and regression equation (10m)

Model\_A <- lm (hba1c ~ bmi + fbs, data = Data1)  
summary(Model\_A)

##   
## Call:  
## lm(formula = hba1c ~ bmi + fbs, data = Data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.9879 -0.4595 -0.1070 0.2929 8.6813   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.057126 0.086672 35.273 < 2e-16 \*\*\*  
## bmi 0.020619 0.003157 6.532 7.31e-11 \*\*\*  
## fbs 0.389686 0.006497 59.979 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.048 on 4065 degrees of freedom  
## (272 observations deleted due to missingness)  
## Multiple R-squared: 0.485, Adjusted R-squared: 0.4847   
## F-statistic: 1914 on 2 and 4065 DF, p-value: < 2.2e-16

tidy(Model\_A, conf.int=T)

## # A tibble: 3 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 3.06 0.0867 35.3 5.18e-238 2.89 3.23   
## 2 bmi 0.0206 0.00316 6.53 7.31e- 11 0.0144 0.0268  
## 3 fbs 0.390 0.00650 60.0 0 0.377 0.402

Y = B0 + B1X1 + B2X2 + Error

hba1c = 3.0571 + 0.0206(bmi) + 0.3896 (fbs) + Error

Interpretation

1. With every increase of 1 unit bmi, the hba1c changes by 0.0206 (95% CI = 0.0144, 0.026808) when fbs is adjusted
2. With every increase of 1 unit fbs, the hba1c changes by 0.3896 ( 95% CI = 0.3769, 0.4024) when bmi is adjusted

## b. Add interaction. Present in table (10m)

Model\_B <- lm (hba1c ~ bmi + fbs + bmi:fbs , data = Data1)  
summary(Model\_B)

##   
## Call:  
## lm(formula = hba1c ~ bmi + fbs + bmi:fbs, data = Data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.5719 -0.4567 -0.1055 0.2914 8.6677   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.637628 0.205549 17.697 < 2e-16 \*\*\*  
## bmi -0.001409 0.007745 -0.182 0.85567   
## fbs 0.287713 0.033385 8.618 < 2e-16 \*\*\*  
## bmi:fbs 0.003821 0.001227 3.114 0.00186 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.047 on 4064 degrees of freedom  
## (272 observations deleted due to missingness)  
## Multiple R-squared: 0.4862, Adjusted R-squared: 0.4858   
## F-statistic: 1282 on 3 and 4064 DF, p-value: < 2.2e-16

tidy(Model\_B)

## # A tibble: 4 × 5  
## term estimate std.error statistic p.value  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 3.64 0.206 17.7 1.43e-67  
## 2 bmi -0.00141 0.00775 -0.182 8.56e- 1  
## 3 fbs 0.288 0.0334 8.62 9.62e-18  
## 4 bmi:fbs 0.00382 0.00123 3.11 1.86e- 3

Table.Model\_B <- tbl\_regression(Model\_B)  
Table.Model\_B

## Table printed with {flextable}, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

| **Characteristic** | **Beta** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| bmi | 0.00 | -0.02, 0.01 | 0.9 |
| fbs | 0.29 | 0.22, 0.35 | <0.001 |
| bmi \* fbs | 0.00 | 0.00, 0.01 | 0.002 |
| 1CI = Confidence Interval | | | |

## c. Write equation with interaction. (10m)

tidy(Model\_B, conf.int = T)

## # A tibble: 4 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 3.64 0.206 17.7 1.43e-67 3.23 4.04   
## 2 bmi -0.00141 0.00775 -0.182 8.56e- 1 -0.0166 0.0138   
## 3 fbs 0.288 0.0334 8.62 9.62e-18 0.222 0.353   
## 4 bmi:fbs 0.00382 0.00123 3.11 1.86e- 3 0.00142 0.00623

Y = B0 + B1X1 + B2X2 + B3X3 Error

hba1c = 3.6376 - 0.00140(bmi) + 0.2877 (fbs) + 0.0038(fbs\*bmi) + Error

1. with increase of 1 unit bmi, the hba1c change by -0.0014 (95% CI = -0.01659, 0.01377) when fbs and fbs\*bmi are adjusted
2. with increase of 1 unit fbs, the hba1c changes by 0.2877 (95% CI = 0.2222, 0.3531) when bmi and fbs\* bmi are adjusted

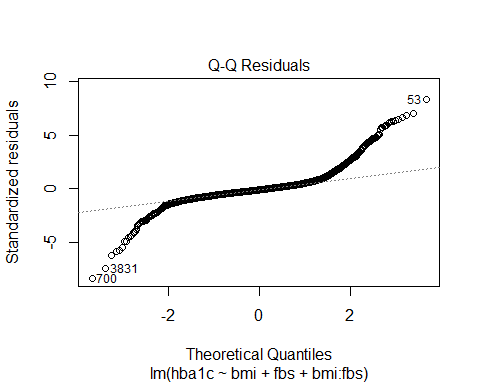
***3) Interaction Interpretation susah, buat next time la***

## d. Assumptions for fitness analysis for this model. (10m)

\*\*comment on adjusted R-squared

### Linearity

plot(Model\_B, which=2)

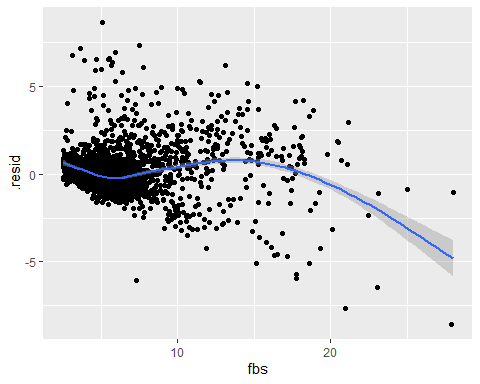


Q-Q plot shows the linearity of residual assumption might be violated as the head and is curved

To compare covariate linearity with residuals

augment(Model\_B) %>%  
 ggplot(aes(x = fbs, y = .resid))+  
 geom\_point() +  
 geom\_smooth()

## `geom\_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'



### Independent

This is based on study design

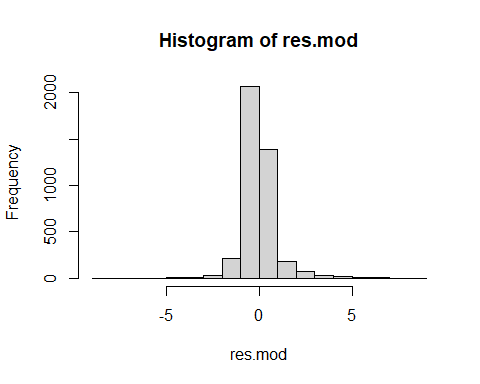
### Normality

shapiro.test(Model\_B$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: Model\_B$residuals  
## W = 0.79717, p-value < 2.2e-16

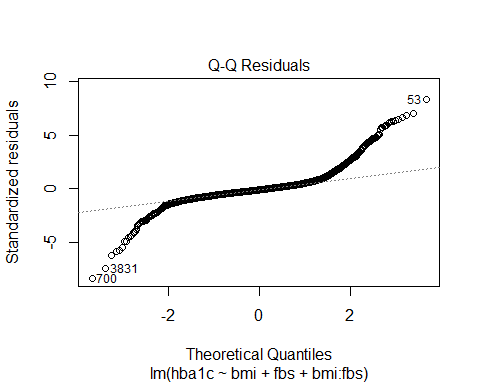
p-value <0.05, normality of residuals assumption violated

res.mod <- residuals(Model\_B)  
hist(res.mod)



however, histogram of residuals seemed normality distributed by eye-balling

plot(Model\_B, which=2)



but the Q-Q plot showed there is no normality as the plot is curved at head and tail

### Equal Variance

ncvTest(Model\_B)

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 3676.72, Df = 1, p = < 2.22e-16

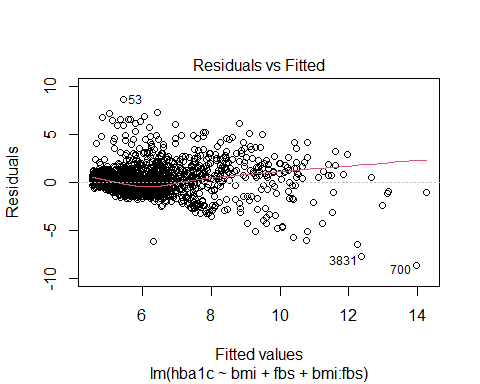
p-value <0.05, homoscedacity assumption is violated

bptest(Model\_B)

##   
## studentized Breusch-Pagan test  
##   
## data: Model\_B  
## BP = 524.12, df = 3, p-value < 2.2e-16

p-value ,0.005, homoscedacity assumption is violated

plot(Model\_B, which = 1 )



plot(Model\_B)

