

regression

Bryan Ng, 2427348

2025-05-28

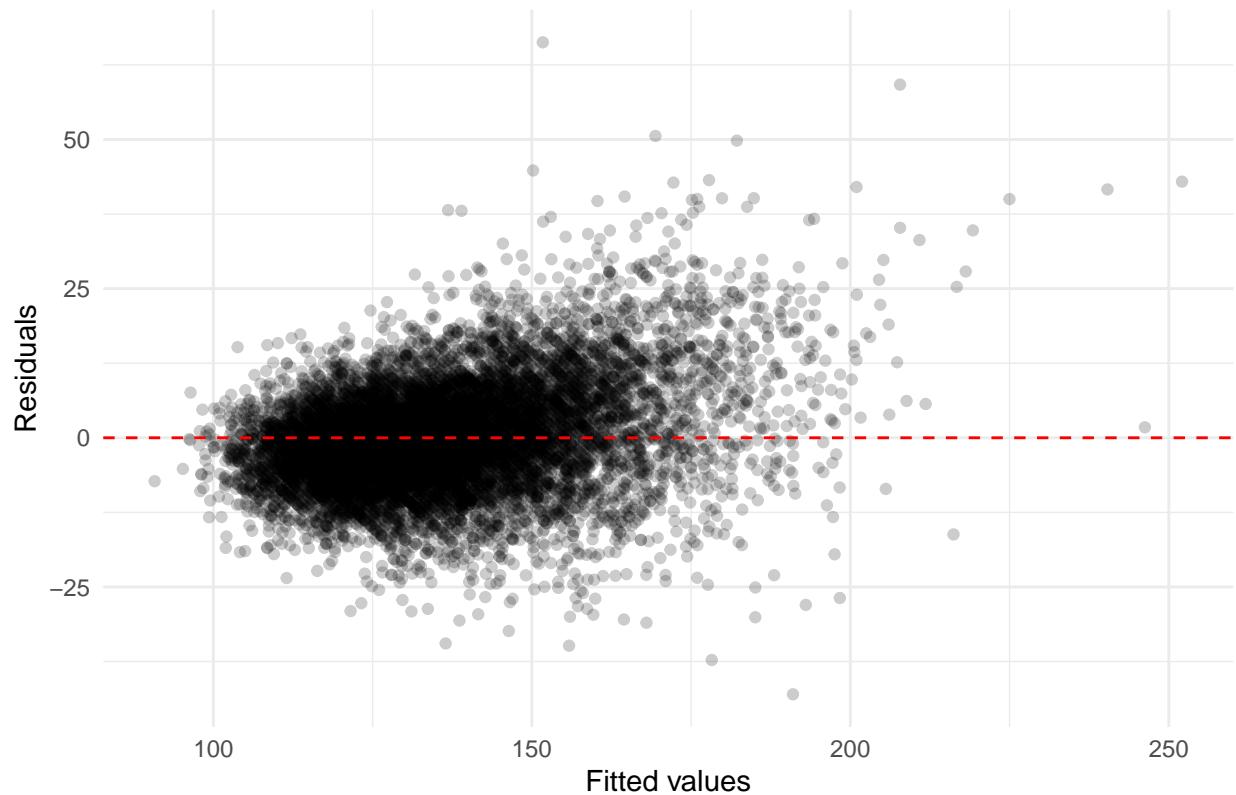
## Question 1

```
vars_q1 <- c("RANDID", "SYSBP", "DIABP", "SEX", "AGE", "BMI", "educ", "TIME", "PERIOD")
keep1 <- complete.cases(df[, vars_q1])
Q1_df <- df[keep1, vars_q1]
```

```
Q1_SYSBP <- lme(
  SYSBP ~ SEX + AGE + BMI + educ + TIME + AGE:BM + SEX:BM,
  random = ~ TIME | RANDID,
  data = Q1_df,
  method = "REML",
  na.action = na.omit
)

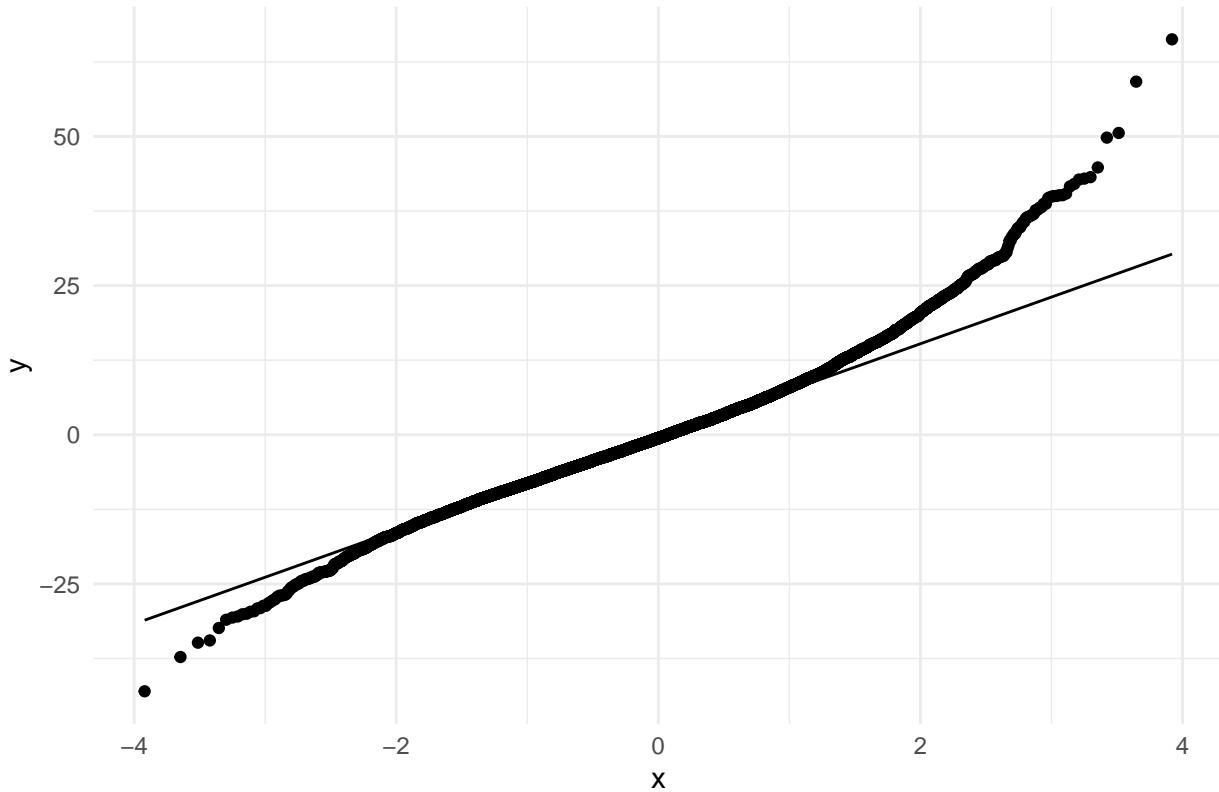
df_diag <- data.frame(
  fitted = fitted(Q1_SYSBP),
  resid = residuals(Q1_SYSBP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()
```

## Residuals vs Fitted



```
ggplot(df_diag, aes(sample = resid)) +  
  stat_qq() +  
  stat_qq_line() +  
  labs(  
    title = "QQ-plot of Residuals"  
) +  
  theme_minimal()
```

## QQ-plot of Residuals



```
summary(Q1_SYSBP)
```

```
## Linear mixed-effects model fit by REML
##   Data: Q1_df
##      AIC      BIC    logLik
##  96051.29 96139.25 -48013.64
##
## Random effects:
##   Formula: ~TIME | RANDID
##   Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev      Corr
## (Intercept) 16.209287681 (Intr)
## TIME        0.002305535 -0.088
## Residual    11.485107279
##
## Fixed effects: SYSBP ~ SEX + AGE + BMI + educ + TIME + AGE:BMI + SEX:BMI
##                  Value Std.Error DF t-value p-value
## (Intercept) 38.83321  9.043915 6970 4.293849 0.0000
## SEX          0.58345  3.306927 4304 0.176432 0.8600
## AGE          1.12579  0.129176 6970 8.715162 0.0000
## BMI          1.68773  0.338162 6970 4.990894 0.0000
## educ         -0.51996  0.272616 4304 -1.907289 0.0565
## TIME         -0.00050  0.000114 6970 -4.413146 0.0000
## AGE:BMI     -0.00669  0.004881 6970 -1.370644 0.1705
## SEX:BMI      0.07903  0.125588 6970  0.629275 0.5292
```

```

## Correlation:
##          (Intr) SEX     AGE     BMI     educ    TIME    AGE:BM
## SEX      -0.605
## AGE      -0.766 -0.012
## BMI      -0.975  0.593  0.747
## educ     -0.072 -0.036  0.024 -0.017
## TIME      0.184 -0.048 -0.220 -0.053 -0.122
## AGE:BM   0.733  0.033 -0.968 -0.761  0.015  0.033
## SEX:BM   0.587 -0.986  0.024 -0.590  0.039  0.050 -0.046
##
## Standardized Within-Group Residuals:
##          Min        Q1        Med        Q3        Max
## -3.74357760 -0.49439363 -0.04982444  0.42494949  5.77087662
##
## Number of Observations: 11282
## Number of Groups: 4307

```

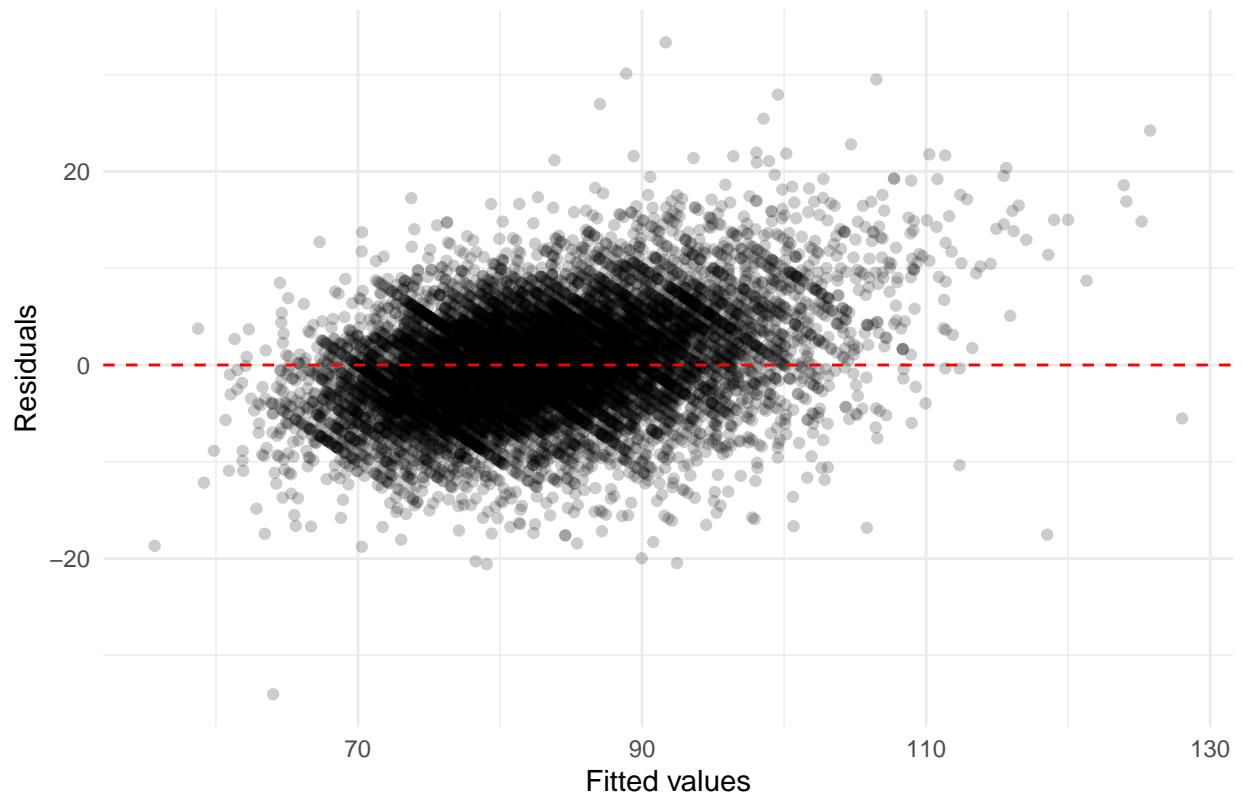
```

Q1_DIABP <- lme(
  DIABP ~ SEX + AGE + BMI + educ + TIME + AGE:BM + SEX:BM,
  random = ~ TIME | RANDID,
  data   = Q1_df,
  method = "REML",
  na.action = na.omit
)

df_diag <- data.frame(
  fitted = fitted(Q1_DIABP),
  resid   = residuals(Q1_DIABP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()

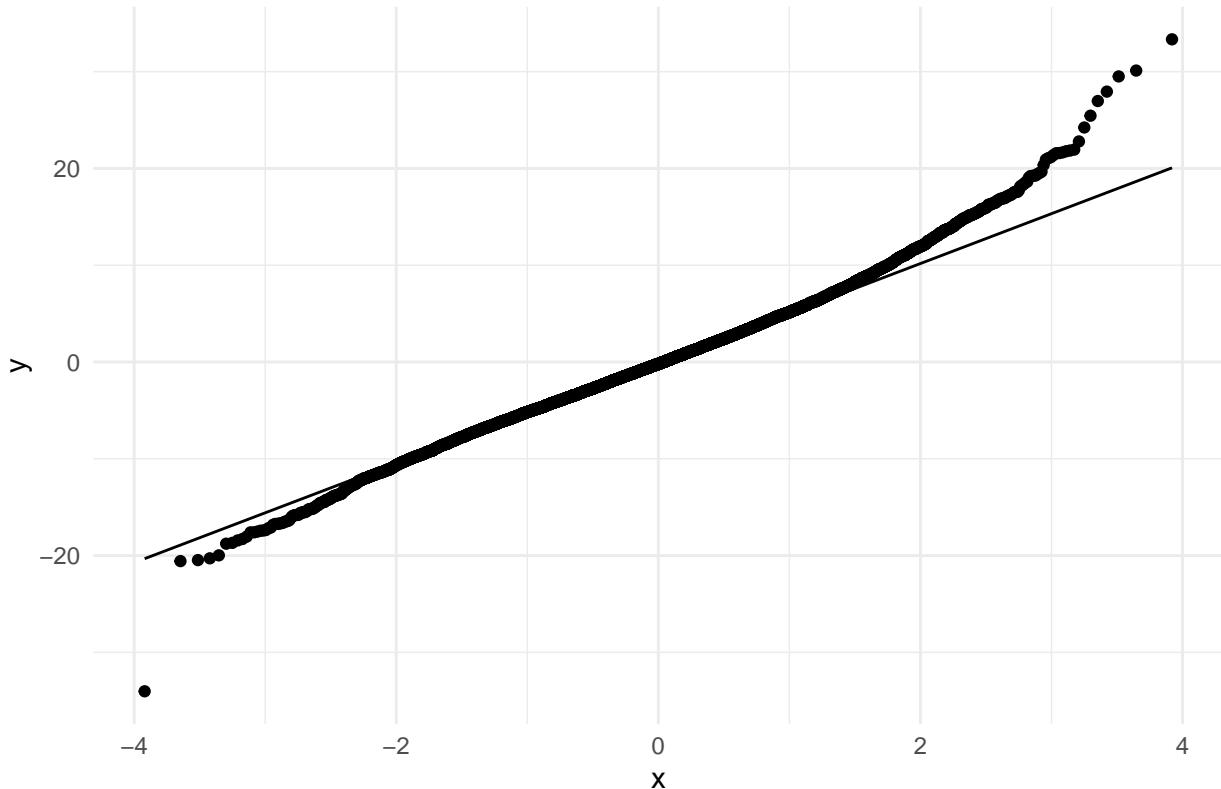
```

## Residuals vs Fitted



```
ggplot(df_diag, aes(sample = resid)) +  
  stat_qq() +  
  stat_qq_line() +  
  labs(  
    title = "QQ-plot of Residuals"  
) +  
  theme_minimal()
```

## QQ-plot of Residuals



```
summary(Q1_DIABP)
```

```
## Linear mixed-effects model fit by REML
##   Data: Q1_df
##      AIC      BIC    logLik
##  83192.34 83280.31 -41584.17
##
## Random effects:
##   Formula: ~TIME | RANDID
##   Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 8.74064290 (Intr)
## TIME        0.00115575 -0.283
## Residual    6.90169010
##
## Fixed effects: DIABP ~ SEX + AGE + BMI + educ + TIME + AGE:BMI + SEX:BMI
##                  Value Std.Error DF t-value p-value
## (Intercept) 23.458955 5.028093 6970 4.665577 0.0000
## SEX          3.007493 1.805179 4304 1.666036 0.0958
## AGE          0.522567 0.072253 6970 7.232490 0.0000
## BMI          2.112417 0.188554 6970 11.203221 0.0000
## educ         0.164745 0.144296 4304 1.141721 0.2536
## TIME         -0.000510 0.000062 6970 -8.212025 0.0000
## AGE:BMI     -0.015624 0.002740 6970 -5.701945 0.0000
## SEX:BMI     -0.133371 0.068600 6970 -1.944185 0.0519
```

```

## Correlation:
##          (Intr) SEX     AGE     BMI     educ    TIME    AGE:BM
## SEX      -0.589
## AGE      -0.776 -0.019
## BMI      -0.978  0.576  0.760
## educ     -0.065 -0.036  0.018 -0.020
## TIME      0.173 -0.050 -0.205 -0.055 -0.119
## AGE:BM   0.745  0.039 -0.971 -0.772  0.019  0.034
## SEX:BM   0.571 -0.987  0.031 -0.572  0.039  0.051 -0.052
##
## Standardized Within-Group Residuals:
##          Min        Q1        Med        Q3        Max
## -4.92973898 -0.52206356 -0.02882177  0.48476286  4.83058534
##
## Number of Observations: 11282
## Number of Groups: 4307

```

## Question 2

```

ind <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ
    + BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BM + BPMEDS:educ,
  id = RANDID,
  data = Q2_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

ar1 <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ
    + BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BM + BPMEDS:educ,
  id = RANDID,
  data = Q2_df,
  waves = TIME,
  corstr = "ar1",
  family = gaussian()
)

exc <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BM + BPMEDS:educ,
  id = RANDID,
  data = Q2_df,
  waves = TIME,
  corstr = "exchangeable",
  family = gaussian()
)
QIC(ind);QIC(ar1);QIC(exc)

```

```

##          QIC       QICu      Quasi Lik       CIC       params
## 3.991590e+06 3.991575e+06 -1.995777e+06 1.857956e+01 1.100000e+01
##          QICC

```

```

##  3.991590e+06

##          QIC          QICu      Quasi Lik          CIC      params
##  3.991590e+06  3.991575e+06 -1.995777e+06  1.857961e+01  1.100000e+01
##          QICC
##  3.991590e+06

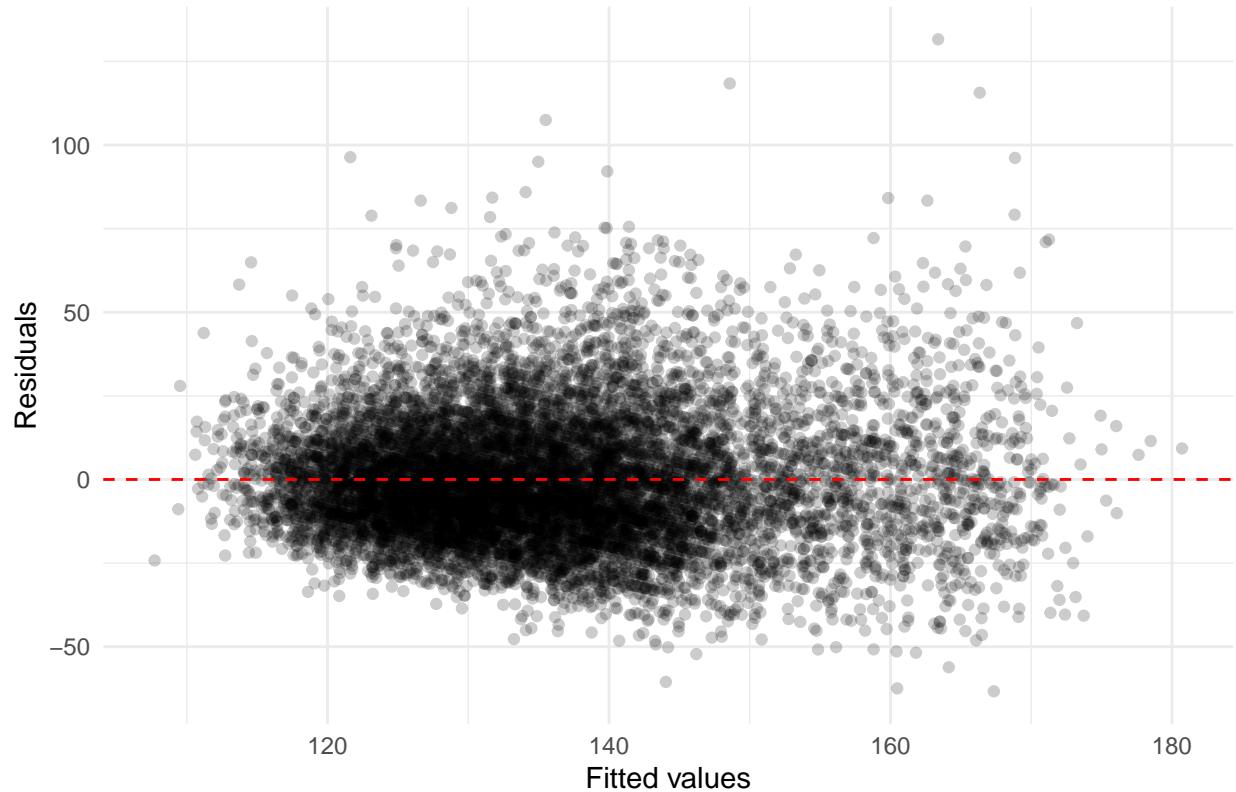
##          QIC          QICu      Quasi Lik          CIC      params
##  4.139917e+06  4.139904e+06 -2.069941e+06  1.794912e+01  1.100000e+01
##          QICC
##  4.139918e+06

Q2_SYSBP <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BMI + BPMEDS:educ,
  id = RANDID,
  data = Q2_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

df_diag <- data.frame(
  fitted = fitted(Q2_SYSBP),
  resid = residuals(Q2_SYSBP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()

```

## Residuals vs Fitted



```
summary(Q2_SYSBP)
```

```
##
## Call:
## geeglm(formula = SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ +
##         BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BMI + BPMEDS:educ, family = gaussian(),
##         data = Q2_df, id = RANDID, waves = TIME, corstr = "independence")
##
## Coefficients:
##             Estimate     Std.err    Wald Pr(>|W|)
## (Intercept) 52.6522763 2.5892916 413.497 < 2e-16 ***
## BPMEDS      57.7942354 9.7693625 34.997 3.30e-09 ***
## AGE          0.8467744 0.0313414 729.960 < 2e-16 ***
## BMI          1.3876543 0.0682577 413.294 < 2e-16 ***
## TIME        -0.0009977 0.0001242 64.537 9.99e-16 ***
## SEX          1.5514618 0.5156168  9.054  0.00262 **
## educ        -0.4316891 0.2472959  3.047  0.08087 .
## BPMEDS:SEX   0.7288104 1.9270658  0.143  0.70528
## BPMEDS:AGE  -0.1824729 0.1019569  3.203  0.07350 .
## BPMEDS:BMI  -0.9877712 0.2028170 23.719 1.11e-06 ***
## BPMEDS:educ -0.6717332 0.9001493  0.557  0.45552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = independence
```

```

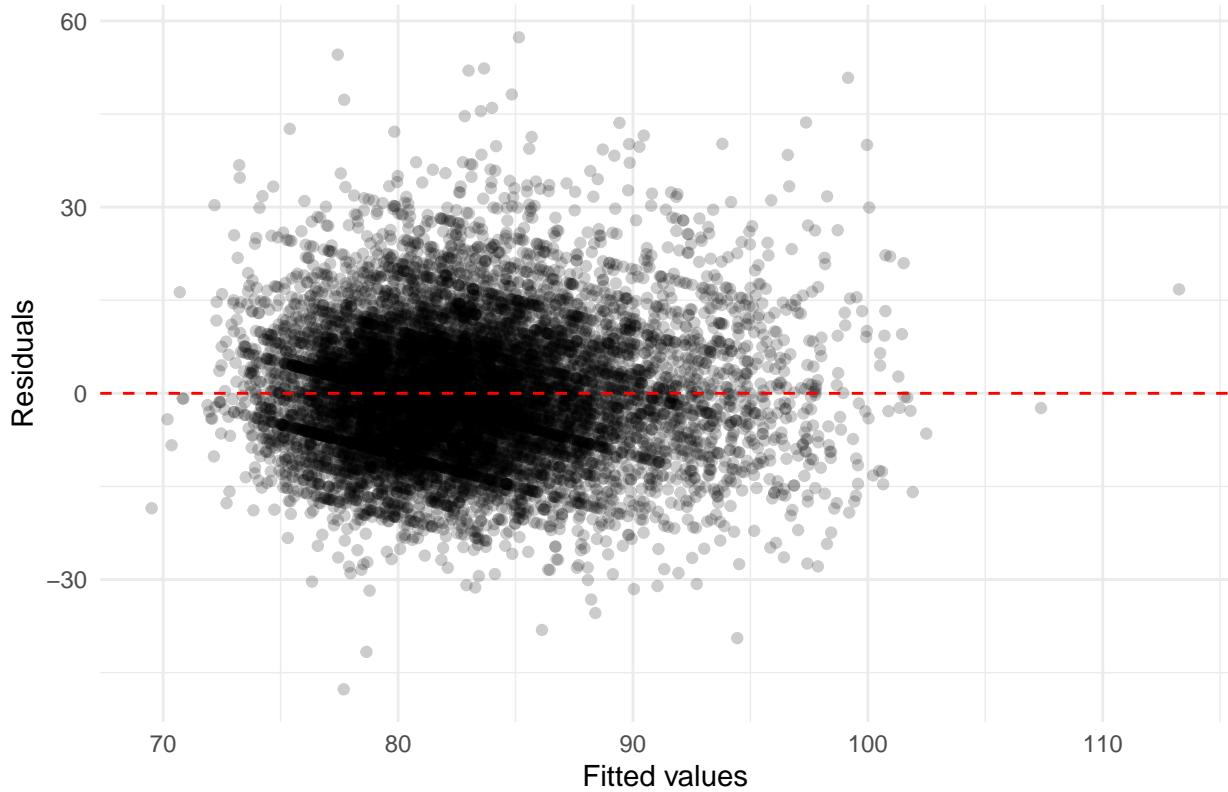
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)    373.1   7.892
## Number of clusters: 4299 Maximum cluster size: 3

Q2_DIABP <- geeglm(
  DIABP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BMI + BPMEDS:educ,
  id = RANDID,
  data = Q2_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

df_diag <- data.frame(
  fitted = fitted(Q2_DIABP),
  resid = residuals(Q2_DIABP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()

```

## Residuals vs Fitted



```
summary(Q2_DIABP)
```

```
##
## Call:
## geeglm(formula = DIABP ~ BPMEDS + AGE + BMI + TIME + SEX + educ +
##         BPMEDS:SEX + BPMEDS:AGE + BPMEDS:BMI + BPMEDS:educ, family = gaussian(),
##         data = Q2_df, id = RANDID, waves = TIME, corstr = "independence")
##
## Coefficients:
##             Estimate Std.err Wald Pr(>|W|)
## (Intercept) 5.49e+01 1.40e+00 1544.65 < 2e-16 ***
## BPMEDS      4.83e+01 4.91e+00   96.58 < 2e-16 ***
## AGE         8.88e-02 1.68e-02   27.82 1.3e-07 ***
## BMI         9.71e-01 3.62e-02   717.25 < 2e-16 ***
## TIME        -8.10e-04 6.65e-05   148.12 < 2e-16 ***
## SEX         -9.15e-01 2.80e-01    10.70  0.0011 **
## educ        1.46e-01 1.33e-01     1.21  0.2705
## BPMEDS:SEX -3.67e-01 9.57e-01     0.15  0.7014
## BPMEDS:AGE -3.59e-01 5.37e-02   44.75 2.2e-11 ***
## BPMEDS:BMI -6.37e-01 1.00e-01   40.45 2.0e-10 ***
## BPMEDS:educ 6.02e-02 4.47e-01     0.02  0.8928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = independence
```

```

## Estimated Scale Parameters:
##
##           Estimate Std.err
## (Intercept)    112     2.11
## Number of clusters:  4299  Maximum cluster size: 3

```

## Question 3

```

df$CIGPDAY <- log(df$CIGPDAY + 1)
vars_q3 <- c("RANDID", "SYSBP", "DIABP",
            "CURSMOKE", "CIGPDAY", "SEX", "AGE", "BMI", "educ", "BPMEDS", "TIME")
keep3 <- complete.cases(df[, vars_q3])
Q3_df <- df[keep3, vars_q3]

ind <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + CIGPDAY + CURSMOKE
    + CURSMOKE:SEX + CIGPDAY:SEX + CURSMOKE:BPMEDS,
  id = RANDID,
  data = Q3_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

ar1 <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + CIGPDAY + CURSMOKE
    + CURSMOKE:SEX + CIGPDAY:SEX + CURSMOKE:BPMEDS,
  id = RANDID,
  data = Q3_df,
  waves = TIME,
  corstr = "ar1",
  family = gaussian()
)

exc <- geeglm(
  SYSBP ~ BPMEDS + AGE + BMI + TIME + SEX + educ + CIGPDAY + CURSMOKE + CURSMOKE:SEX + CIGPDAY:SEX + CIGPDAY:SEX:SEX
  id = RANDID,
  data = Q3_df,
  waves = TIME,
  corstr = "exchangeable",
  family = gaussian()
)
QIC(ind);QIC(ar1);QIC(exc)

```

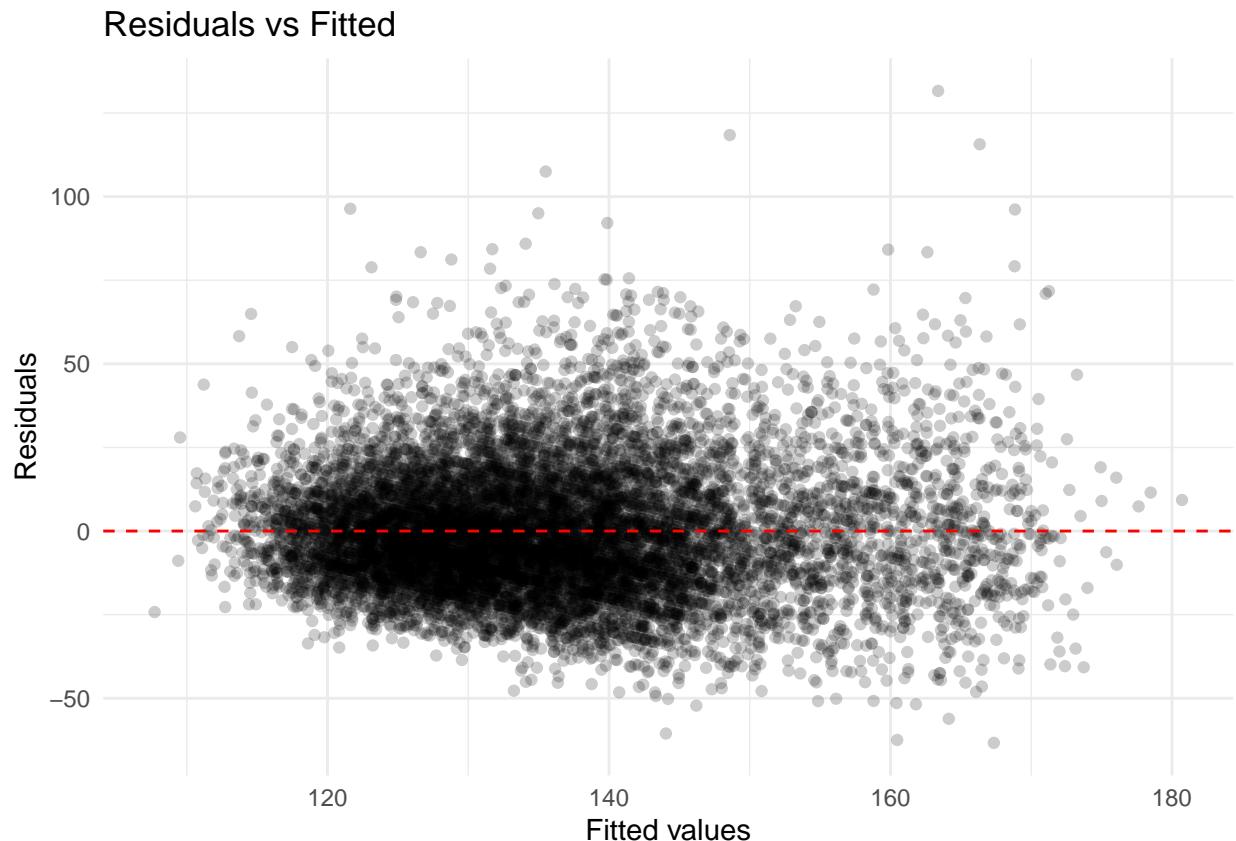
	QIC	QICu	Quasi Lik	CIC	params	QICC
##	3.97e+06	3.97e+06	-1.99e+06	1.93e+01	1.20e+01	3.97e+06
##	4340521	4340287	-2170132	129	12	4340521
##	4.12e+06	4.12e+06	-2.06e+06	1.79e+01	1.20e+01	4.12e+06

```

Q3_SYSBP <- geeglm(
  SYSBP ~ CIGPDAY + CURSMOKE + AGE + BMI + BPMEDS + TIME + SEX + educ + CURSMOKE:SEX + CIGPDAY:SEX + CU
  id = RANDID,
  data = Q3_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

df_diag <- data.frame(
  fitted = fitted(Q2_SYSBP),
  resid   = residuals(Q2_SYSBP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
) +
  theme_minimal()

```



```
summary(Q3_SYSBP)
```

```

## 
## Call:
## geeglm(formula = SYSBP ~ CIGPDAY + CURSMOKE + AGE + BMI + BPMEDS +
##         TIME + SEX + educ + CURSMOKE:SEX + CIGPDAY:SEX + CURSMOKE:BPMEDS,
##         family = gaussian(), data = Q3_df, id = RANDID, waves = TIME,
##         corstr = "independence")
##
## Coefficients:
##              Estimate Std.err Wald Pr(>|W|)
## (Intercept) 54.920508 2.813384 381.08 < 2e-16 ***
## CIGPDAY      2.060501 1.522363   1.83  0.17590
## CURSMOKE    -3.770730 4.697334   0.64  0.42213
## AGE          0.843383 0.031595 712.56 < 2e-16 ***
## BMI          1.261012 0.066319 361.54 < 2e-16 ***
## BPMEDS       18.816024 1.118129 283.19 < 2e-16 ***
## TIME         -0.000993 0.000125  63.13  1.9e-15 ***
## SEX           2.321986 0.678630  11.71  0.00062 ***
## educ          -0.447170 0.243933   3.36  0.06678 .
## CURSMOKE:SEX -0.172996 2.771305   0.00  0.95023
## CIGPDAY:SEX   -0.443417 0.937264   0.22  0.63614
## CURSMOKE:BPMEDS 3.606354 1.980535   3.32  0.06862 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = independence
## Estimated Scale Parameters:
## 
##              Estimate Std.err
## (Intercept)     374     7.96
## Number of clusters: 4298 Maximum cluster size: 3

```

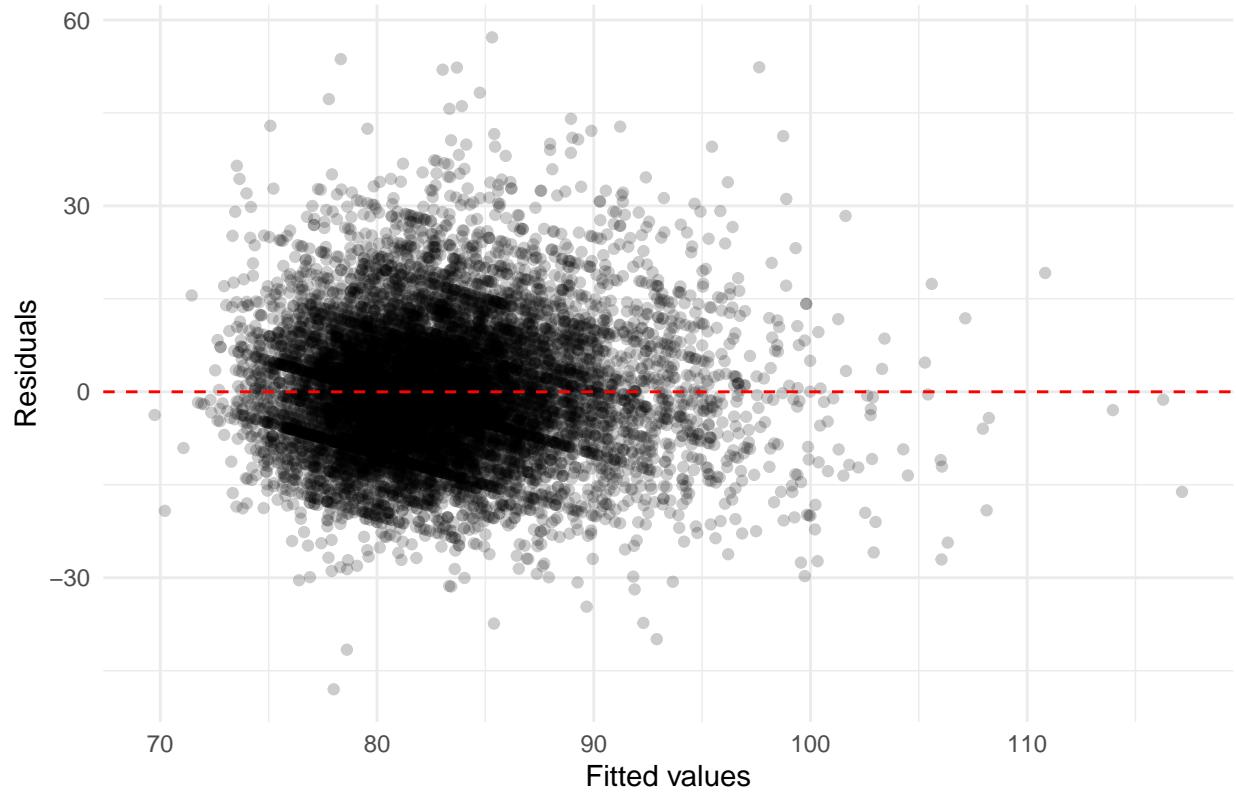
```

Q3_DIABP <- geeglm(
  DIABP ~ CIGPDAY + CURSMOKE + AGE + BMI + BPMEDS + TIME + SEX + educ + CURSMOKE:SEX + CIGPDAY:SEX + CU
  id = RANDID,
  data = Q3_df,
  waves = TIME,
  corstr = "independence",
  family = gaussian()
)

df_diag <- data.frame(
  fitted = fitted(Q3_DIABP),
  resid = residuals(Q3_DIABP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()

```

## Residuals vs Fitted



```
summary(Q3_DIABP)
```

```
##
## Call:
## geeglm(formula = DIABP ~ CIGPDAY + CURSMOKE + AGE + BMI + BPMEDS +
##         TIME + SEX + educ + CURSMOKE:SEX + CIGPDAY:SEX + CURSMOKE:BPMEDS,
##         family = gaussian(), data = Q3_df, id = RANDID, waves = TIME,
##         corstr = "independence")
##
## Coefficients:
##             Estimate Std. err Wald Pr(>|W|)
## (Intercept) 58.555098 1.497452 1529.06 < 2e-16 ***
## CIGPDAY      0.814344 0.833970   0.95  0.32883
## CURSMOKE    -2.199421 2.570547   0.73  0.39221
## AGE          0.061861 0.016914  13.38  0.00025 ***
## BMI          0.886213 0.035156  635.45 < 2e-16 ***
## BPMEDS       7.646517 0.557037  188.43 < 2e-16 ***
## TIME        -0.000795 0.000067  140.48 < 2e-16 ***
## SEX          -0.816581 0.361570   5.10  0.02392 *
## educ         0.170510 0.129591   1.73  0.18826
## CURSMOKE:SEX -0.135694 1.512465   0.01  0.92851
## CIGPDAY:SEX   -0.119362 0.510994   0.05  0.81530
## CURSMOKE:BPMEDS 3.748554 1.050933  12.72  0.00036 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## 
## Correlation structure = independence
## Estimated Scale Parameters:
## 
##           Estimate Std.err
## (Intercept)     113    2.14
## Number of clusters: 4298 Maximum cluster size: 3

```

## Question 4

```

vars_q4 <- c("RANDID", "SYSBP", "DIABP",
           "TOTCHOL", "GLUCOSE",
           "SEX", "AGE", "BMI", "educ", "BPMEDS", "CURSMOKE", "CIGPDAY", "TIME")
keep4   <- complete.cases(df[, vars_q4])
Q4_df   <- df[keep4, vars_q4]

```

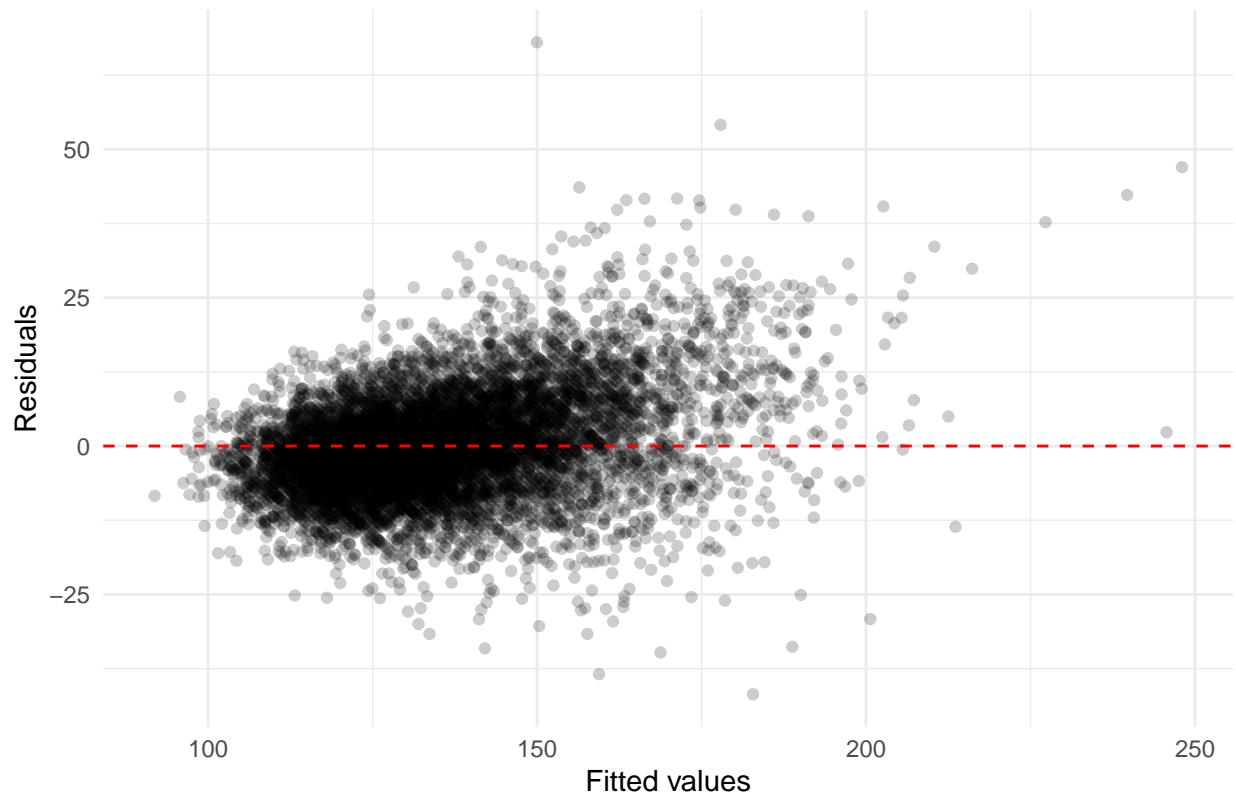
```

Q4_SYSBP <- lme(
  fixed = SYSBP ~ TOTCHOL + GLUCOSE + AGE + BMI + BPMEDS + SEX + educ + CIGPDAY + CURSMOKE + TOTCHOL:SEX,
  random = ~ TIME | RANDID,
  data = Q4_df,
  method = "REML",
  na.action = na.omit
)

df_diag <- data.frame(
  fitted = fitted(Q4_SYSBP),
  resid   = residuals(Q4_SYSBP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
  ) +
  theme_minimal()

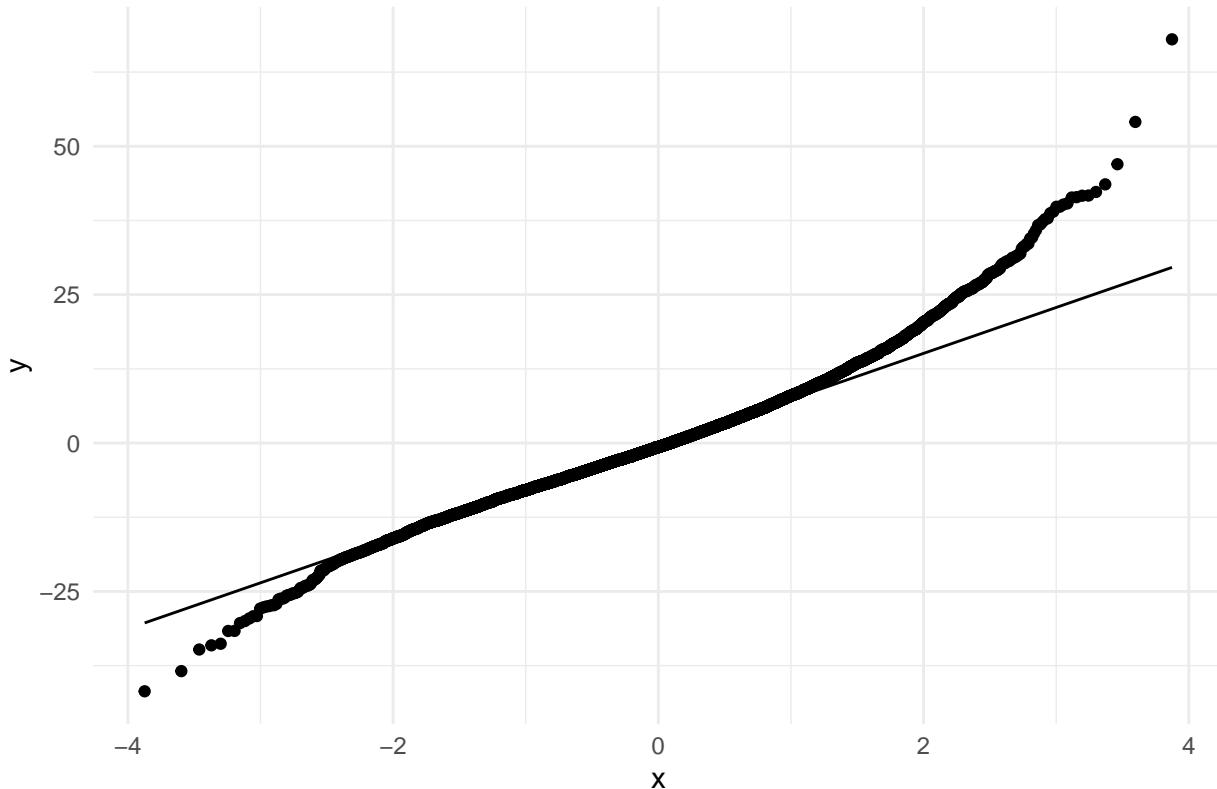
```

## Residuals vs Fitted



```
ggplot(df_diag, aes(sample = resid)) +  
  stat_qq() +  
  stat_qq_line() +  
  labs(  
    title = "QQ-plot of Residuals"  
) +  
  theme_minimal()
```

## QQ-plot of Residuals



```
summary(Q4_SYSBP)
```

```
## Linear mixed-effects model fit by REML
##   Data: Q4_df
##   AIC   BIC logLik
## 79513 79684 -39732
##
## Random effects:
##   Formula: ~TIME | RANDID
##   Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 15.67869 (Intr)
## TIME        0.00211 -0.08
## Residual    11.53320
##
## Fixed effects: SYSBP ~ TOTCHOL + GLUCOSE + AGE + BMI + BPMEDS + SEX + educ +
##                  CIGPDAY + CURSMOK
##             Value Std.Error DF t-value p-value
## (Intercept) 55.7     12.03 5080    4.63  0.0000
## TOTCHOL      0.0      0.04 5080    0.38  0.7017
## GLUCOSE      0.0      0.07 5080   -0.02  0.9820
## AGE          0.8      0.13 5080    6.13  0.0000
## BMI          0.9      0.30 5080    3.00  0.0027
## BPMEDS       4.7      3.84 5080    1.22  0.2229
## SEX          1.5      2.80 4211    0.54  0.5897
## educ         -0.7     0.27 4211   -2.51  0.0121
```

```

## CIGPDAY          0.7    0.40 5080   1.70  0.0890
## CURSMOKE        -3.2   2.83 5080  -1.12  0.2646
## TOTCHOL:SEX      0.0    0.01 5080   0.12  0.9064
## GLUCOSE:SEX      0.0    0.02 5080   0.19  0.8469
## TOTCHOL:AGE      0.0    0.00 5080   0.21  0.8349
## GLUCOSE:AGE      0.0    0.00 5080  -1.44  0.1511
## TOTCHOL:BMI       0.0    0.00 5080   0.29  0.7726
## GLUCOSE:BMI       0.0    0.00 5080   2.77  0.0057
## TOTCHOL:CURSMOKE  0.0    0.01 5080   0.45  0.6535
## GLUCOSE:CURSMOKE  0.0    0.02 5080  -0.03  0.9771
## TOTCHOL:BPMEDS    0.0    0.01 5080   0.16  0.8712
## GLUCOSE:BPMEDS    0.0    0.02 5080   0.32  0.7513
## Correlation:
##              (Intr) TOTCHOL GLUCOSE AGE      BMI      BPMEDS SEX      educ
## TOTCHOL        -0.848
## GLUCOSE        -0.471 -0.016
## AGE            -0.638  0.493   0.421
## BMI            -0.654  0.601   0.209   0.000
## BPMEDS         0.185 -0.134  -0.122  -0.215 -0.087
## SEX             -0.396  0.343   0.132   0.026  0.009 -0.083
## educ            -0.061 -0.013  -0.007  0.019  0.000  0.006  0.010
## CIGPDAY        0.003 -0.035  0.005  -0.014 -0.009  0.007  0.028  0.017
## CURSMOKE       -0.346  0.307   0.160   0.242  0.111 -0.007  0.104  0.011
## TOTCHOL:SEX     0.362 -0.399  -0.004  -0.022 -0.031  0.052 -0.862 -0.007
## GLUCOSE:SEX     0.128  0.009  -0.262   0.025  0.056  0.077 -0.492  0.002
## TOTCHOL:AGE     0.522 -0.609  -0.016  -0.826 -0.010  0.200 -0.006 -0.001
## GLUCOSE:AGE     0.363  0.004  -0.774  -0.549 -0.005  0.066  0.005 -0.001
## TOTCHOL:BMI      0.580 -0.696   0.037   0.009 -0.879  0.036 -0.031  0.020
## GLUCOSE:BMI      0.273  0.019  -0.577  -0.050 -0.410  0.125  0.060  0.018
## TOTCHOL:CURSMOKE 0.330 -0.381  -0.005  -0.195 -0.145  0.008 -0.100  0.001
## GLUCOSE:CURSMOKE 0.166 -0.003  -0.332  -0.180  0.025 -0.002 -0.062 -0.024
## TOTCHOL:BPMEDS   -0.152  0.156   0.026   0.204  0.048 -0.873  0.063 -0.002
## GLUCOSE:BPMEDS   -0.127  0.016  0.228   0.093  0.102 -0.501  0.070 -0.014
## CIGPDA CURSMO TOTCHOL:S GLUCOSE:S TOTCHOL:A GLUCOSE:A
## TOTCHOL
## GLUCOSE
## AGE
## BMI
## BPMEDS
## SEX
## educ
## CIGPDAY
## CURSMOKE       -0.374
## TOTCHOL:SEX      0.015 -0.099
## GLUCOSE:SEX     -0.014 -0.058   0.034
## TOTCHOL:AGE      0.026 -0.198 -0.006  -0.018
## GLUCOSE:AGE      0.002 -0.148   0.005  -0.027   0.021
## TOTCHOL:BMI      0.016 -0.126   0.045  -0.020   0.017  -0.022
## GLUCOSE:BMI     -0.002  0.013  -0.021  -0.108   0.010   0.084
## TOTCHOL:CURSMOKE 0.006 -0.784   0.116   0.005   0.239   0.005
## GLUCOSE:CURSMOKE 0.013 -0.483   0.007   0.118   0.019   0.310
## TOTCHOL:BPMEDS   -0.002  0.015  -0.062  -0.017  -0.234  -0.006
## GLUCOSE:BPMEDS   -0.008 -0.012  -0.006  -0.140  -0.017  -0.139
## TOTCHOL:BM GLUCOSE:BM TOTCHOL:C GLUCOSE:C TOTCHOL:BP

```

```

## TOTCHOL
## GLUCOSE
## AGE
## BMI
## BPMEDS
## SEX
## educ
## CIGPDAY
## CURSMOKE
## TOTCHOL:SEX
## GLUCOSE:SEX
## TOTCHOL:AGE
## GLUCOSE:AGE
## TOTCHOL:BMI
## GLUCOSE:BMI      -0.022
## TOTCHOL:CURSMOKE 0.163     -0.001
## GLUCOSE:CURSMOKE -0.016     -0.030     0.010
## TOTCHOL:BPMEDS   -0.039     -0.030     -0.007    -0.021
## GLUCOSE:BPMEDS   -0.009     -0.225     -0.007    0.043     0.053
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.626 -0.483 -0.058  0.421  5.897
##
## Number of Observations: 9311
## Number of Groups: 4214

```

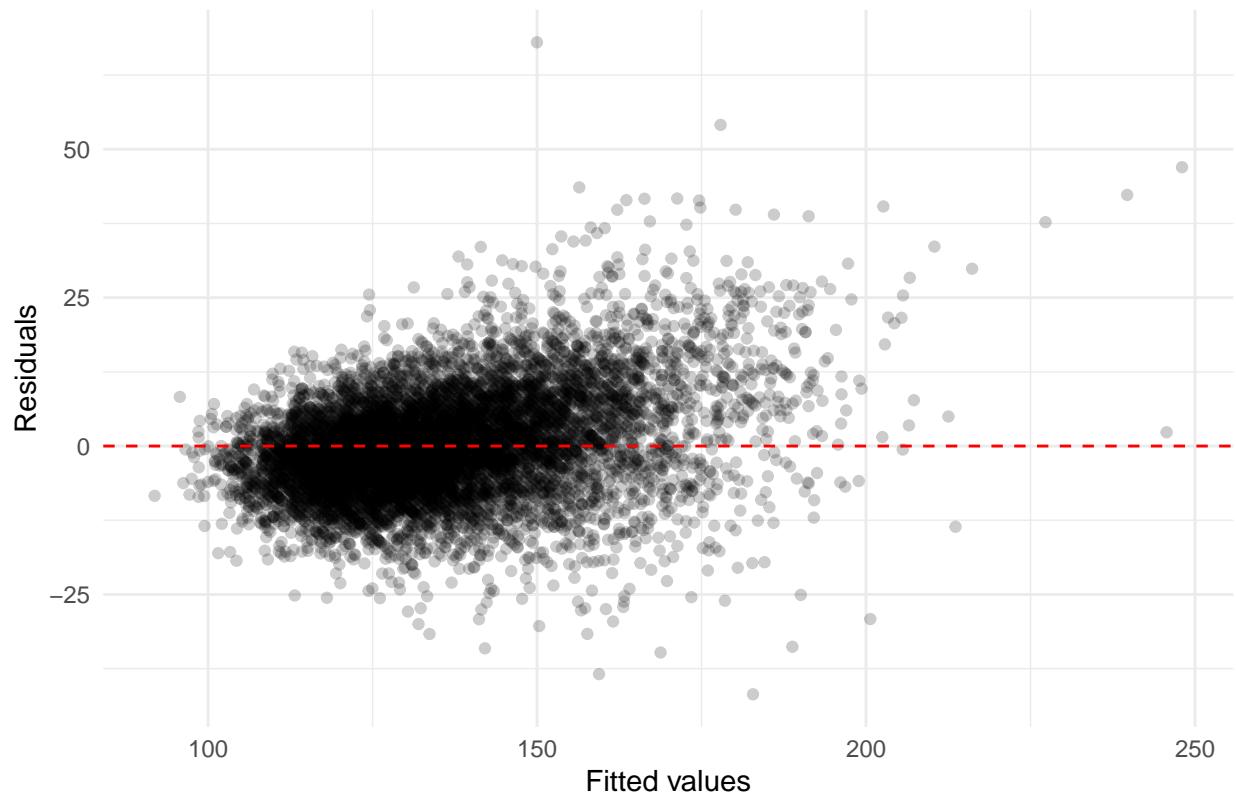
```

Q4_DIABP <- lme(
  fixed = DIABP ~ TOTCHOL + GLUCOSE + AGE + BMI + BPMEDS + CURSMOKE + CIGPDAY + SEX + educ + TOTCHOL:SEX,
  random = ~ TIME | RANDID,
  data = Q4_df,
  method = "REML",
  na.action = na.omit
)

df_diag <- data.frame(
  fitted = fitted(Q4_SYSBP),
  resid   = residuals(Q4_SYSBP)
)
ggplot(df_diag, aes(x = fitted, y = resid)) +
  geom_point(alpha = 0.2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  labs(
    x = "Fitted values",
    y = "Residuals",
    title = "Residuals vs Fitted"
) +
  theme_minimal()

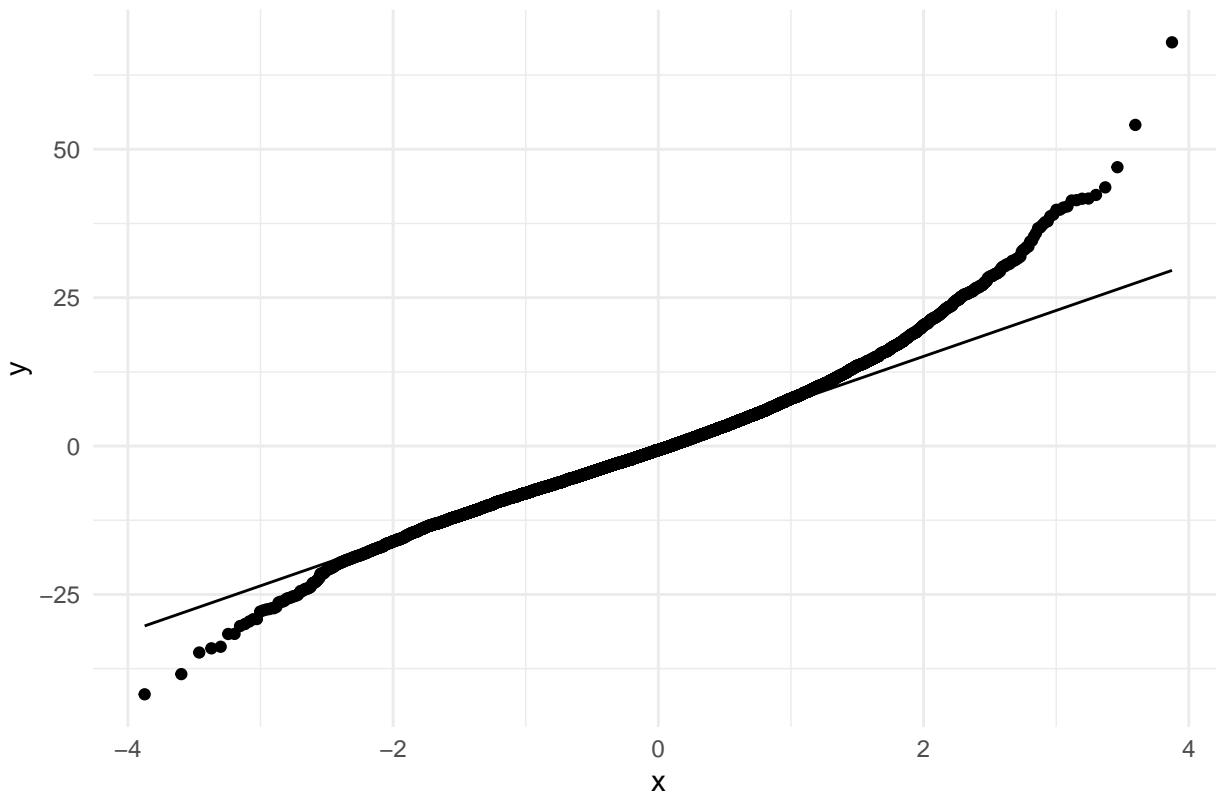
```

## Residuals vs Fitted



```
ggplot(df_diag, aes(sample = resid)) +  
  stat_qq() +  
  stat_qq_line() +  
  labs(  
    title = "QQ-plot of Residuals"  
) +  
  theme_minimal()
```

## QQ-plot of Residuals



```
summary(Q4_DIABP)
```

```
## Linear mixed-effects model fit by REML
##   Data: Q4_df
##   AIC   BIC logLik
## 69037 69208 -34495
##
## Random effects:
##   Formula: ~TIME | RANDID
##   Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 8.79555 (Intr)
## TIME        0.00133 -0.368
## Residual    6.81019
##
## Fixed effects: DIABP ~ TOTCHOL + GLUCOSE + AGE + BMI + BPMEDS + CURSMOKE + CIGPDAY + SEX + educ
##                  Value Std.Error DF t-value p-value
## (Intercept) 35.8     6.92 5080    5.18  0.0000
## TOTCHOL      0.1     0.03 5080    3.98  0.0001
## GLUCOSE      0.0     0.04 5080    0.91  0.3648
## AGE          0.3     0.08 5080    3.98  0.0001
## BMI          0.9     0.17 5080    5.12  0.0000
## BPMEDS       3.4     2.19 5080    1.54  0.1231
## CURSMOKE    -1.2     1.63 5080   -0.72  0.4711
## CIGPDAY       0.4     0.23 5080    1.56  0.1192
```

```

## SEX           -0.7      1.58 4211   -0.42  0.6710
## educ          0.0      0.15 4211    0.21  0.8342
## TOTCHOL:SEX   0.0      0.01 5080   -0.44  0.6609
## GLUCOSE:SEX   0.0      0.01 5080    0.21  0.8341
## TOTCHOL:AGE   0.0      0.00 5080   -2.37  0.0177
## GLUCOSE:AGE   0.0      0.00 5080   -4.62  0.0000
## TOTCHOL:BMI   0.0      0.00 5080   -1.36  0.1735
## GLUCOSE:BMI   0.0      0.00 5080    3.70  0.0002
## TOTCHOL:CURSMOKE 0.0      0.01 5080   -0.02  0.9828
## GLUCOSE:CURSMOKE 0.0      0.01 5080   -0.37  0.7112
## TOTCHOL:BPMEDS 0.0      0.01 5080   -0.93  0.3506
## GLUCOSE:BPMEDS 0.0      0.01 5080    0.89  0.3744
## Correlation:
##              (Intr) TOTCHOL GLUCOSE AGE      BMI      BPMEDS CURSMO CIGPDA
## TOTCHOL        -0.847
## GLUCOSE       -0.478 -0.013
## AGE            -0.648  0.502  0.423
## BMI            -0.653  0.598  0.218  0.007
## BPMEDS         0.188 -0.141 -0.117 -0.211 -0.097
## CURSMOKE      -0.347  0.308  0.161  0.241  0.115 -0.004
## CIGPDAY        0.003 -0.035  0.006 -0.012 -0.010  0.005 -0.380
## SEX            -0.392  0.339  0.135  0.027  0.010 -0.087  0.104  0.029
## educ           -0.056 -0.013 -0.009  0.017 -0.003  0.010  0.011  0.018
## TOTCHOL:SEX    0.356 -0.393 -0.006 -0.021 -0.032  0.056 -0.098  0.013
## GLUCOSE:SEX    0.130  0.008 -0.262  0.022  0.055  0.078 -0.059 -0.017
## TOTCHOL:AGE    0.528 -0.618 -0.015 -0.827 -0.013  0.199 -0.197  0.025
## GLUCOSE:AGE    0.370  0.002 -0.778 -0.549 -0.011  0.058 -0.150  0.001
## TOTCHOL:BMI    0.578 -0.693  0.031  0.004 -0.878  0.045 -0.131  0.017
## GLUCOSE:BMI    0.279  0.015 -0.576 -0.051 -0.420  0.129  0.013 -0.001
## TOTCHOL:CURSMOKE 0.332 -0.384 -0.005 -0.196 -0.149  0.008 -0.782  0.010
## GLUCOSE:CURSMOKE 0.169 -0.004 -0.335 -0.181  0.024 -0.006 -0.486  0.013
## TOTCHOL:BPMEDS -0.156  0.163  0.025  0.203  0.055 -0.872  0.013 -0.001
## GLUCOSE:BPMEDS -0.125  0.017  0.218  0.081  0.111 -0.499 -0.015 -0.006
## SEX             0.011
## educ           -0.860 -0.008
## TOTCHOL:SEX    -0.498  0.003  0.033
## GLUCOSE:SEX    -0.006  0.000 -0.007   -0.016
## TOTCHOL:AGE    0.002  0.002  0.006   -0.023   0.019
## GLUCOSE:AGE    -0.031  0.021  0.045   -0.019   0.020   -0.019
## TOTCHOL:BMI    0.058  0.021 -0.018   -0.107   0.010   0.085
## GLUCOSE:BMI    -0.100  0.002  0.114    0.007   0.241   0.004
## TOTCHOL:CURSMOKE -0.065 -0.025  0.008    0.119   0.016   0.316
## GLUCOSE:CURSMOKE -0.065 -0.004 -0.066   -0.015   -0.233   -0.007
## TOTCHOL:BPMEDS  0.073 -0.016 -0.006   -0.146   -0.015   -0.116
## GLUCOSE:BPMEDS
## TOTCHOL:BM GLUCOSE:BM TOTCHOL:C GLUCOSE:C TOTCHOL:BP

```

```

## TOTCHOL
## GLUCOSE
## AGE
## BMI
## BPMEDS
## CURSMOKE
## CIGPDAY
## SEX
## educ
## TOTCHOL:SEX
## GLUCOSE:SEX
## TOTCHOL:AGE
## GLUCOSE:AGE
## TOTCHOL:BMI
## GLUCOSE:BMI      -0.016
## TOTCHOL:CURSMOKE 0.168     -0.001
## GLUCOSE:CURSMOKE -0.013     -0.033     0.011
## TOTCHOL:BPMEDS    -0.048     -0.028     -0.006    -0.020
## GLUCOSE:BPMEDS    -0.012     -0.238     -0.009     0.051     0.048
##
## Standardized Within-Group Residuals:
##      Min       Q1       Med       Q3       Max
## -4.5135 -0.4941 -0.0345  0.4526  5.0312
##
## Number of Observations: 9311
## Number of Groups: 4214

models <- list(
  Q1_SYSBP = Q1_SYSBP,
  Q1_DIABP = Q1_DIABP,
  Q2_SYSBP = Q2_SYSBP,
  Q2_DIABP = Q2_DIABP,
  Q3_SYSBP = Q3_SYSBP,
  Q3_DIABP = Q3_DIABP,
  Q4_SYSBP = Q4_SYSBP,
  Q4_DIABP = Q4_DIABP
)

extract_coefs <- function(fit) {
  if (inherits(fit, "geeglm")) {

    mat <- summary(fit)$coefficients
    df  <- as.data.frame(mat, check.names = FALSE)
    df$term <- rownames(df)
    df %>%
      rename(
        estimate = Estimate,
        std.error = `Std.error`,
        statistic = Wald,
        p.value   = `Pr(>|W|)`
      ) %>%
      select(term, estimate, std.error, statistic, p.value)
  } else if (inherits(fit, "lme")) {

```

```

mat <- summary(fit)$tTable
df <- as.data.frame(mat, check.names = FALSE)
df$term <- rownames(df)
df %>%
  rename(
    estimate = Value,
    std.error = `Std.Error`,
    statistic = `t-value`,
    p.value = `p-value`
  ) %>%
  select(term, estimate, std.error, statistic, p.value)
} else {
  stop("Extractor only knows geeglm and lme")
}
}

coef_tables <- map(models, extract_coefs)

walk2(
  coef_tables,
  names(coef_tables),
  ~{
    cat("## Model:", .y, "\n\n")
    print(
      kable(
        .x,
        caption = paste("Coefficients for", .y),
        digits = 3,
        col.names = c("Term", "Estimate", "Std.Error", "Stat", "p.value")
      )
    )
    cat("\n\n")
  }
)

```

```

## ## Model: Q1_SYSBP
##
##
##
## Table: Coefficients for Q1_SYSBP
##
## |           |Term          | Estimate| Std.Error|   Stat| p.value|
## |:-----|:-----|-----:|-----:|-----:|-----:|
## |(Intercept)| (Intercept) | 38.833| 9.044| 4.294| 0.000|
## |SEX         |SEX          | 0.583| 3.307| 0.176| 0.860|
## |AGE         |AGE          | 1.126| 0.129| 8.715| 0.000|
## |BMI         |BMI          | 1.688| 0.338| 4.991| 0.000|
## |educ        |educ         | -0.520| 0.273| -1.907| 0.057|
## |TIME        |TIME         | -0.001| 0.000| -4.413| 0.000|
## |AGE:BMI    |AGE:BMI     | -0.007| 0.005| -1.371| 0.171|
## |SEX:BMI    |SEX:BMI     | 0.079| 0.126| 0.629| 0.529|
##
```

```

## ## Model: Q1_DIABP
##
##
##
## Table: Coefficients for Q1_DIABP
##
## | Term | Estimate | Std.Error | Stat | p.value |
## | :----- | :----- | -----: | -----: | -----: |
## |(Intercept) |(Intercept) | 23.459 | 5.028 | 4.67 | 0.000 |
## |SEX |SEX | 3.007 | 1.805 | 1.67 | 0.096 |
## |AGE |AGE | 0.523 | 0.072 | 7.23 | 0.000 |
## |BMI |BMI | 2.112 | 0.189 | 11.20 | 0.000 |
## |educ |educ | 0.165 | 0.144 | 1.14 | 0.254 |
## |TIME |TIME | -0.001 | 0.000 | -8.21 | 0.000 |
## |AGE:BMI |AGE:BMI | -0.016 | 0.003 | -5.70 | 0.000 |
## |SEX:BMI |SEX:BMI | -0.133 | 0.069 | -1.94 | 0.052 |
##
##
## ## Model: Q2_SYSBP
##
##
##
## Table: Coefficients for Q2_SYSBP
##
## | Term | Estimate | Std.Error | Stat | p.value |
## | :----- | :----- | -----: | -----: | -----: |
## |(Intercept) |(Intercept) | 52.652 | 2.589 | 413.497 | 0.000 |
## |BPMEDS |BPMEDS | 57.794 | 9.769 | 34.997 | 0.000 |
## |AGE |AGE | 0.847 | 0.031 | 729.960 | 0.000 |
## |BMI |BMI | 1.388 | 0.068 | 413.294 | 0.000 |
## |TIME |TIME | -0.001 | 0.000 | 64.537 | 0.000 |
## |SEX |SEX | 1.551 | 0.516 | 9.054 | 0.003 |
## |educ |educ | -0.432 | 0.247 | 3.047 | 0.081 |
## |BPMEDS:SEX |BPMEDS:SEX | 0.729 | 1.927 | 0.143 | 0.705 |
## |BPMEDS:AGE |BPMEDS:AGE | -0.182 | 0.102 | 3.203 | 0.074 |
## |BPMEDS:BMI |BPMEDS:BMI | -0.988 | 0.203 | 23.719 | 0.000 |
## |BPMEDS:educ |BPMEDS:educ | -0.672 | 0.900 | 0.557 | 0.456 |
##
##
## ## Model: Q2_DIABP
##
##
##
## Table: Coefficients for Q2_DIABP
##
## | Term | Estimate | Std.Error | Stat | p.value |
## | :----- | :----- | -----: | -----: | -----: |
## |(Intercept) |(Intercept) | 54.907 | 1.397 | 1544.655 | 0.000 |
## |BPMEDS |BPMEDS | 48.276 | 4.912 | 96.583 | 0.000 |
## |AGE |AGE | 0.089 | 0.017 | 27.824 | 0.000 |
## |BMI |BMI | 0.971 | 0.036 | 717.246 | 0.000 |
## |TIME |TIME | -0.001 | 0.000 | 148.117 | 0.000 |
## |SEX |SEX | -0.915 | 0.280 | 10.705 | 0.001 |
## |educ |educ | 0.146 | 0.133 | 1.214 | 0.271 |

```

```

## |BPMEDS:SEX |BPMEDS:SEX | -0.367| 0.957| 0.147| 0.701|
## |BPMEDS:AGE |BPMEDS:AGE | -0.359| 0.054| 44.746| 0.000|
## |BPMEDS:BMI |BPMEDS:BMI | -0.637| 0.100| 40.451| 0.000|
## |BPMEDS:educ |BPMEDS:educ | 0.060| 0.447| 0.018| 0.893|
##
##
## ## Model: Q3_SYSBP
##
##
## Table: Coefficients for Q3_SYSBP
##
## |Term | Estimate| Std.Error| Stat| p.value|
## |:-----|:-----:|-----:|-----:|-----:|
## |(Intercept) |(Intercept)| 54.921| 2.813| 381.076| 0.000|
## |CIGPDAY |CIGPDAY| 2.061| 1.522| 1.832| 0.176|
## |CURSMOKE |CURSMOKE| -3.771| 4.697| 0.644| 0.422|
## |AGE |AGE| 0.843| 0.032| 712.557| 0.000|
## |BMI |BMI| 1.261| 0.066| 361.541| 0.000|
## |BPMEDS |BPMEDS| 18.816| 1.118| 283.186| 0.000|
## |TIME |TIME| -0.001| 0.000| 63.129| 0.000|
## |SEX |SEX| 2.322| 0.679| 11.707| 0.001|
## |educ |educ| -0.447| 0.244| 3.361| 0.067|
## |CURSMOKE:SEX |CURSMOKE:SEX| -0.173| 2.771| 0.004| 0.950|
## |CIGPDAY:SEX |CIGPDAY:SEX| -0.443| 0.937| 0.224| 0.636|
## |CURSMOKE:BPMEDS |CURSMOKE:BPMEDS| 3.606| 1.981| 3.316| 0.069|
##
##
## ## Model: Q3_DIABP
##
##
## Table: Coefficients for Q3_DIABP
##
## |Term | Estimate| Std.Error| Stat| p.value|
## |:-----|:-----:|-----:|-----:|-----:|
## |(Intercept) |(Intercept)| 58.555| 1.497| 1529.057| 0.000|
## |CIGPDAY |CIGPDAY| 0.814| 0.834| 0.953| 0.329|
## |CURSMOKE |CURSMOKE| -2.199| 2.571| 0.732| 0.392|
## |AGE |AGE| 0.062| 0.017| 13.376| 0.000|
## |BMI |BMI| 0.886| 0.035| 635.446| 0.000|
## |BPMEDS |BPMEDS| 7.647| 0.557| 188.434| 0.000|
## |TIME |TIME| -0.001| 0.000| 140.480| 0.000|
## |SEX |SEX| -0.817| 0.362| 5.101| 0.024|
## |educ |educ| 0.171| 0.130| 1.731| 0.188|
## |CURSMOKE:SEX |CURSMOKE:SEX| -0.136| 1.512| 0.008| 0.929|
## |CIGPDAY:SEX |CIGPDAY:SEX| -0.119| 0.511| 0.055| 0.815|
## |CURSMOKE:BPMEDS |CURSMOKE:BPMEDS| 3.749| 1.051| 12.723| 0.000|
##
##
## ## Model: Q4_SYSBP
##
##

```

```

## Table: Coefficients for Q4_SYSBP
##
## | Term | Estimate | Std.Error | Stat | p.value |
## | :--- | :--- | :--- | :--- | :--- |
## | (Intercept) | (Intercept) | 55.706 | 12.025 | 4.632 | 0.000 |
## | TOTCHOL | TOTCHOL | 0.017 | 0.044 | 0.383 | 0.702 |
## | GLUCOSE | GLUCOSE | -0.002 | 0.070 | -0.023 | 0.982 |
## | AGE | AGE | 0.826 | 0.135 | 6.128 | 0.000 |
## | BMI | BMI | 0.898 | 0.299 | 3.001 | 0.003 |
## | BPMEDS | BPMEDS | 4.681 | 3.840 | 1.219 | 0.223 |
## | SEX | SEX | 1.508 | 2.795 | 0.539 | 0.590 |
## | educ | educ | -0.681 | 0.272 | -2.509 | 0.012 |
## | CIGPDAY | CIGPDAY | 0.681 | 0.400 | 1.701 | 0.089 |
## | CURSMOKE | CURSMOKE | -3.163 | 2.835 | -1.116 | 0.265 |
## | TOTCHOL:SEX | TOTCHOL:SEX | 0.001 | 0.010 | 0.118 | 0.906 |
## | GLUCOSE:SEX | GLUCOSE:SEX | 0.003 | 0.016 | 0.193 | 0.847 |
## | TOTCHOL:AGE | TOTCHOL:AGE | 0.000 | 0.000 | 0.208 | 0.835 |
## | GLUCOSE:AGE | GLUCOSE:AGE | -0.001 | 0.001 | -1.436 | 0.151 |
## | TOTCHOL:BMI | TOTCHOL:BMI | 0.000 | 0.001 | 0.289 | 0.773 |
## | GLUCOSE:BMI | GLUCOSE:BMI | 0.004 | 0.001 | 2.769 | 0.006 |
## | TOTCHOL:CURSMOKE | TOTCHOL:CURSMOKE | 0.004 | 0.009 | 0.449 | 0.654 |
## | GLUCOSE:CURSMOKE | GLUCOSE:CURSMOKE | 0.000 | 0.016 | -0.029 | 0.977 |
## | TOTCHOL:BPMEDS | TOTCHOL:BPMEDS | 0.002 | 0.013 | 0.162 | 0.871 |
## | GLUCOSE:BPMEDS | GLUCOSE:BPMEDS | 0.006 | 0.020 | 0.317 | 0.751 |
##
## |
## |
## ## Model: Q4_DIABP
##
## |
## |
## |
## Table: Coefficients for Q4_DIABP
##
## | Term | Estimate | Std.Error | Stat | p.value |
## | :--- | :--- | :--- | :--- | :--- |
## | (Intercept) | (Intercept) | 35.834 | 6.919 | 5.179 | 0.000 |
## | TOTCHOL | TOTCHOL | 0.100 | 0.025 | 3.978 | 0.000 |
## | GLUCOSE | GLUCOSE | 0.037 | 0.041 | 0.906 | 0.365 |
## | AGE | AGE | 0.309 | 0.078 | 3.976 | 0.000 |
## | BMI | BMI | 0.874 | 0.171 | 5.120 | 0.000 |
## | BPMEDS | BPMEDS | 3.383 | 2.194 | 1.542 | 0.123 |
## | CURSMOKE | CURSMOKE | -1.176 | 1.632 | -0.721 | 0.471 |
## | CIGPDAY | CIGPDAY | 0.357 | 0.229 | 1.558 | 0.119 |
## | SEX | SEX | -0.673 | 1.584 | -0.425 | 0.671 |
## | educ | educ | 0.031 | 0.146 | 0.209 | 0.834 |
## | TOTCHOL:SEX | TOTCHOL:SEX | -0.002 | 0.006 | -0.439 | 0.661 |
## | GLUCOSE:SEX | GLUCOSE:SEX | 0.002 | 0.009 | 0.210 | 0.834 |
## | TOTCHOL:AGE | TOTCHOL:AGE | -0.001 | 0.000 | -2.372 | 0.018 |
## | GLUCOSE:AGE | GLUCOSE:AGE | -0.002 | 0.000 | -4.623 | 0.000 |
## | TOTCHOL:BMI | TOTCHOL:BMI | -0.001 | 0.001 | -1.361 | 0.173 |
## | GLUCOSE:BMI | GLUCOSE:BMI | 0.003 | 0.001 | 3.705 | 0.000 |
## | TOTCHOL:CURSMOKE | TOTCHOL:CURSMOKE | 0.000 | 0.005 | -0.022 | 0.983 |
## | GLUCOSE:CURSMOKE | GLUCOSE:CURSMOKE | -0.003 | 0.009 | -0.370 | 0.711 |
## | TOTCHOL:BPMEDS | TOTCHOL:BPMEDS | -0.007 | 0.007 | -0.933 | 0.351 |
## | GLUCOSE:BPMEDS | GLUCOSE:BPMEDS | 0.010 | 0.011 | 0.888 | 0.374 |

```