Lab3 YZ Draft

Yulia Zamriy

March 29, 2018

```r
#setwd("/home/yulia/Documents/MIDS/W203/Lab_3/")
crime <- read.csv("crime_v2.csv", stringsAsFactors = FALSE)
crime <- na.omit(crime)
#str(crime)
```

```r
crime$prbconv <- as.numeric(crime$prbconv)
crime$county <- NULL
crime$year <- NULL
```

```r
crime_summary <- data.frame(t(mapply(summary, crime)))
#str(crime_summary)
crime_summary <- crime_summary[,c("Min.","Mean","Max.")]
crime_summary$Min. <- round(crime_summary$Min.,5)
crime_summary$Mean <- round(crime_summary$Mean,4)
crime_summary$Max. <- round(crime_summary$Max.,4)
```

```r
kable(crime_summary, booktabs = TRUE) %>%
  kable_styling(font_size = 7)
```

|	Min.	Mean	Max.
crmrte	0.00553	0.0334	0.0990
prbarr	0.09277	0.2949	1.0909
prbconv	0.06838	0.5513	2.1212
prbpris	0.15000	0.4108	0.6000
avgsen	5.38000	9.6468	20.7000
polpc	0.00075	0.0017	0.0091
density	0.00002	1.4288	8.8277
taxpc	25.69287	38.0551	119.7615
west	0.00000	0.2527	1.0000
central	0.00000	0.3736	1.0000
urban	0.00000	0.0879	1.0000
pctmin80	1.28365	25.4955	64.3482
wcon	193.64316	285.3585	436.7666
wtuc	187.61726	411.6680	613.2261
wtrd	154.20900	211.5529	354.6761
wfir	170.94017	322.0982	509.4655
wser	133.04306	275.5642	2177.0681
wmfg	157.41000	335.5887	646.8500
wfed	326.10001	442.9007	597.9500
wsta	258.32999	357.5220	499.5900
wloc	239.17000	312.6808	388.0900
mix	0.01961	0.1288	0.4651
pctymle	0.06216	0.0840	0.2487

```r
nrow(crime[crime$prbarr >= 1,])
```

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

```r
nrow(crime[crime$prbconv >= 1,])
```

```
## [1] 10
```

```
par(mfrow=c(1,2))
plot(crime$prbarr, crime$crmrte)
plot(crime$prbconv, crime$crmrte)
```
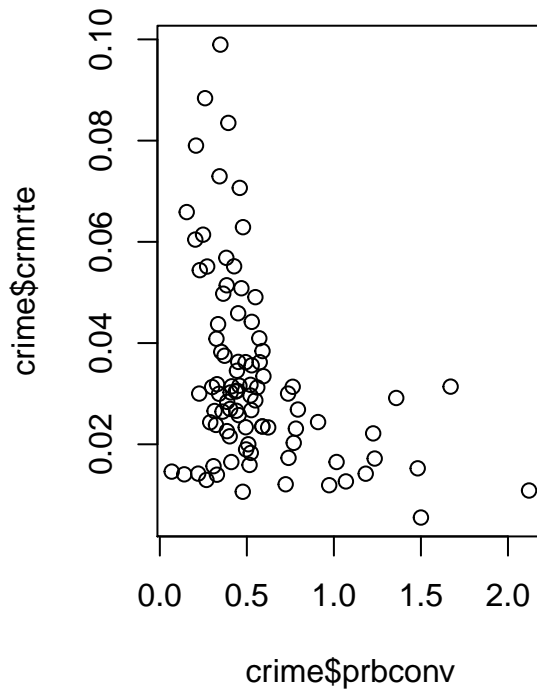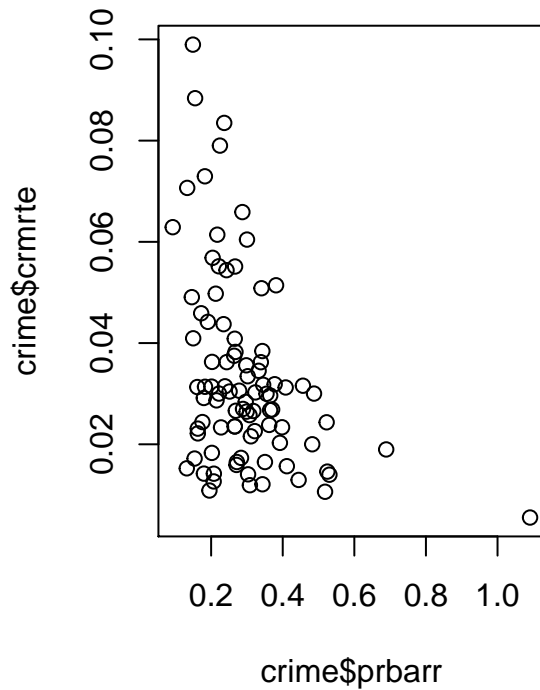


```
crime$prbarr_imp <- ifelse(crime$prbarr > 1, mean(crime$prbarr), crime$prbarr)
summary(crime$prbarr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09277 0.20568 0.27095 0.29492 0.34438 1.09091
```

```
summary(crime$prbarr_imp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09277 0.20568 0.27095 0.28617 0.34323 0.68902
```
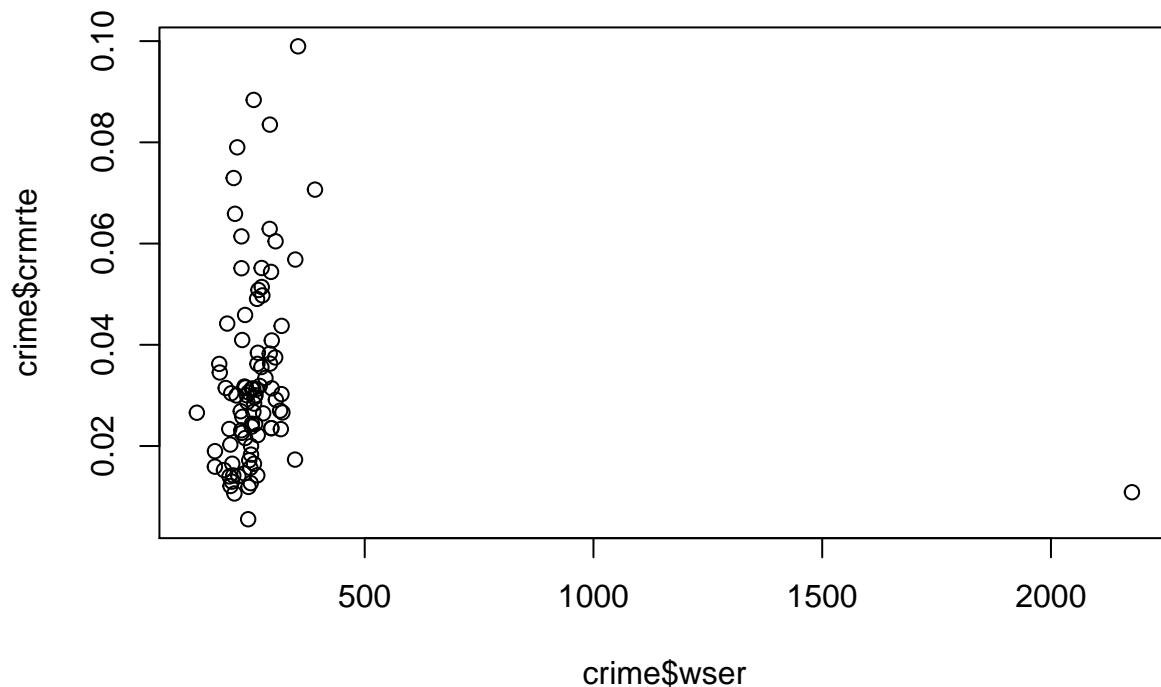
```
crime$prbconv_imp <- ifelse(crime$prbconv > 1, 1, crime$prbconv)
summary(crime$prbconv)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.55128 0.58886 2.12121
```

```
summary(crime$prbconv_imp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.06838 0.34541 0.45283 0.50888 0.58886 1.00000
```

```
plot(crime$wser, crime$crmrte)
```

```r
crime$wser_imp <- ifelse(crime$wser > 2000, mean(crime[crime$wser < 2000,]$wser), crime$wser)
summary(crime$wser)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   133.0   229.7   253.2   275.6   280.5  2177.1
```

```r
summary(crime$wser_imp)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   133.0   229.7   253.2   254.4   277.2   391.3
```

```r
# Prepare a .RData for easier sharing and usage.
ind_variables <- c( 'crmrte',
  'prbarr_imp', 'prbconv_imp', 'prbpris', 'avgsen',
  'polpc', 'density', 'taxpc', 'west', 'central', 'urban', 'pctmin80', 'wcon',
  'wtuc', 'wtrd', 'wfir', 'wser_imp', 'wmfg', 'wfed', 'wsta', 'wloc', 'mix',
  'pctymle'
)
var_labels <- c('crimes committed per person',
  'probability of arrest', 'probability of conviction',
  'probability of prison sentence', 'avg. sentence, days',
  'police per capita', 'people per sq. mile', 'tax revenue per capita',
  '=1 if in western N.C.', '=1 if in central N.C.', '=1 if in SMSA',
  'perc. minority, 1980', 'weekly wage, construction',
  'wkly wge, trns, util, commun', 'wkly wge, whlesle, retail trade',
  'wkly wge, fin, ins, real est', 'wkly wge, service industry',
  'wkly wge, manufacturing', 'wkly wge, fed employees',
  'wkly wge, state employees', 'wkly wge, local gov emps',
```

```r
  'offense mix: face-to-face/other', 'percent young male'
)
impact <- c("Dependent",
  "Negative" , "Negative", "Negative", "Negative",
            "Negative", "Positive", "Negative",
            "Unclear", "Unclear", "Unclear", "Unclear",
            "Negative","Negative","Negative",
            "Negative", "Negative", "Negative", "Negative",
            "Negative", "Negative", "Unclear","Positive")
control <- c("NA","Yes", "Yes", "Yes", "Yes",
             "Yes", "No", "Yes",
             "No", "No", "No","No",
             "Yes", "Yes", "Yes",
             "Yes", "Yes", "Yes", "Yes",
             "Yes", "Yes", "No", "No")
cor_w_crimerate <- round(cor(crime[,ind_variables])[1,],2)
desc <- data.frame(ind_variables, var_labels, impact, cor_w_crimerate, control,
                   row.names = NULL)
colnames(desc) <- c("Explanatory Variables",
                    "Explanation",
                    "Expected Impact on Crime Rate",
                    "Correlation w/ Crime Rate",
                    "Can Gov Impact This?")
```

```r
kable(desc, booktabs = TRUE, align = c("llccc")) %>%
  kable_styling(latex_options = c("scale_down"),
                full_width = FALSE) %>%
  row_spec(0, bold = TRUE) %>%
  column_spec(1, width = "8em") %>%
  column_spec(3, width = "10em") %>%
  column_spec(4, width = "8em") %>%
  column_spec(5, width = "9em")
```
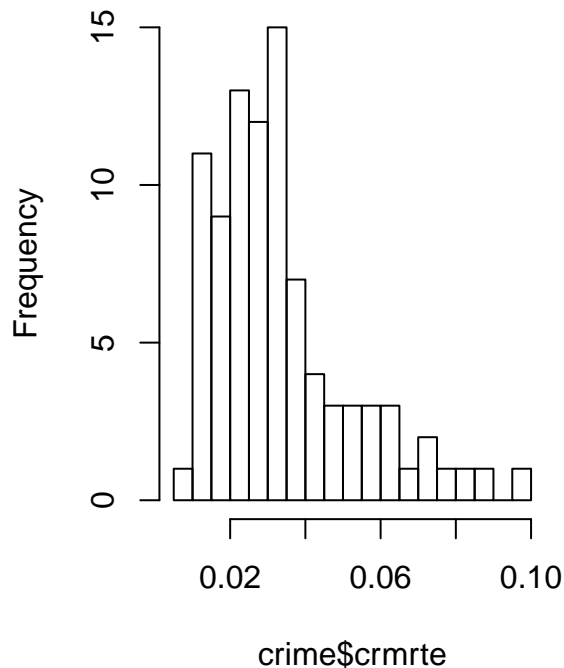
| Explanatory Variables	Explanation	Expected Impact on Crime Rate	Correlation w/ Crime Rate	Can Gov Impact This?
crmrte	crimes committed per person	Dependent	1.00	NA
prbarr_imp	probability of arrest	Negative	-0.38	Yes
prbconv_imp	probability of conviction	Negative	-0.42	Yes
prbpris	probability of prison sentence	Negative	0.05	Yes
avgsen	avg. sentence, days	Negative	0.03	Yes
polpc	police per capita	Negative	0.17	Yes
density	people per sq. mile	Positive	0.73	No
taxpc	tax revenue per capita	Negative	0.45	Yes
west	=1 if in western N.C.	Unclear	-0.35	No
central	=1 if in central N.C.	Unclear	0.17	No
urban	=1 if in SMSA	Unclear	0.62	No
pctmin80	perc. minority, 1980	Unclear	0.19	No
wcon	weekly wage, construction	Negative	0.39	Yes
wtuc	wkly wge, trns, util, commun	Negative	0.23	Yes
wtrd	wkly wge, whlesle, retail trade	Negative	0.41	Yes
wfir	wkly wge, fin, ins, real est	Negative	0.33	Yes
wser_imp	wkly wge, service industry	Negative	0.34	Yes
wmfg	wkly wge, manufacturing	Negative	0.35	Yes
wfed	wkly wge, fed employees	Negative	0.49	Yes
wsta	wkly wge, state employees	Negative	0.20	Yes
wloc	wkly wge, local gov emps	Negative	0.35	Yes
mix	offense mix: face-to-face/other	Unclear	-0.13	No
pctymle	percent young male	Positive	0.29	No

```r
par(mfrow=c(1,2))
hist(crime$crmrte, breaks = 15)
hist(log(crime$crmrte), breaks = 15)
```

Histogram of crime$crmrte   **Histogram of log(crime$crmrte)**
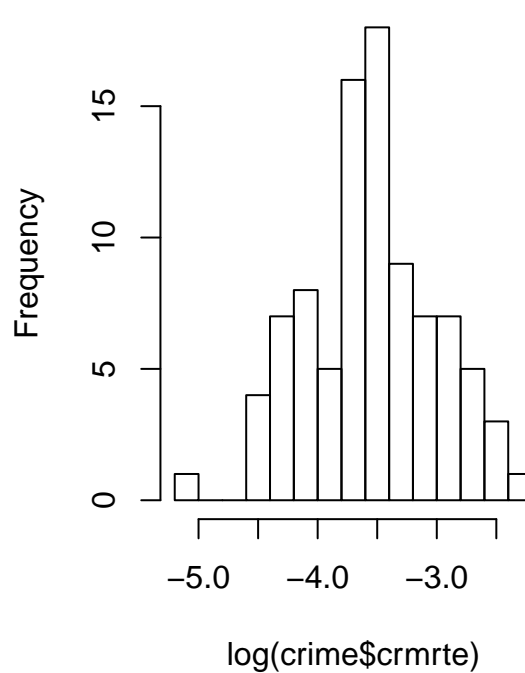


```r
crime$region <- ifelse(crime$west == 1, "west",
                       ifelse(crime$central == 1, "central", "other"))
aggregate(crmrte ~ region, data = crime, mean)
```

```
##    region     crmrte
## 1 central 0.03699627
## 2   other 0.03739491
## 3    west 0.02216183
```

```r
cor_mat <- round(cor(crime[,ind_variables]),2)
get_upper_tri <- function(cor_mat){
    cor_mat[lower.tri(cor_mat)]<- NA
    return(cor_mat)
}
cor_mat_upper <- get_upper_tri(cor_mat)
cor_mat_upper2 <- melt(cor_mat_upper, na.rm = TRUE)
cor_mat_upper2[cor_mat_upper2$value == 1,]$value <- 0
```

```r
ggplot(data = cor_mat_upper2, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
                       midpoint = 0, limit = c(-1,1), space = "Lab",
                       name = "Correlation") +
  theme_minimal() +
  scale_x_discrete(position = "top") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, size = 8, hjust = 0),
        axis.title.x=element_blank(),
```

```
        axis.title.y=element_blank()) +
  coord_fixed()
```
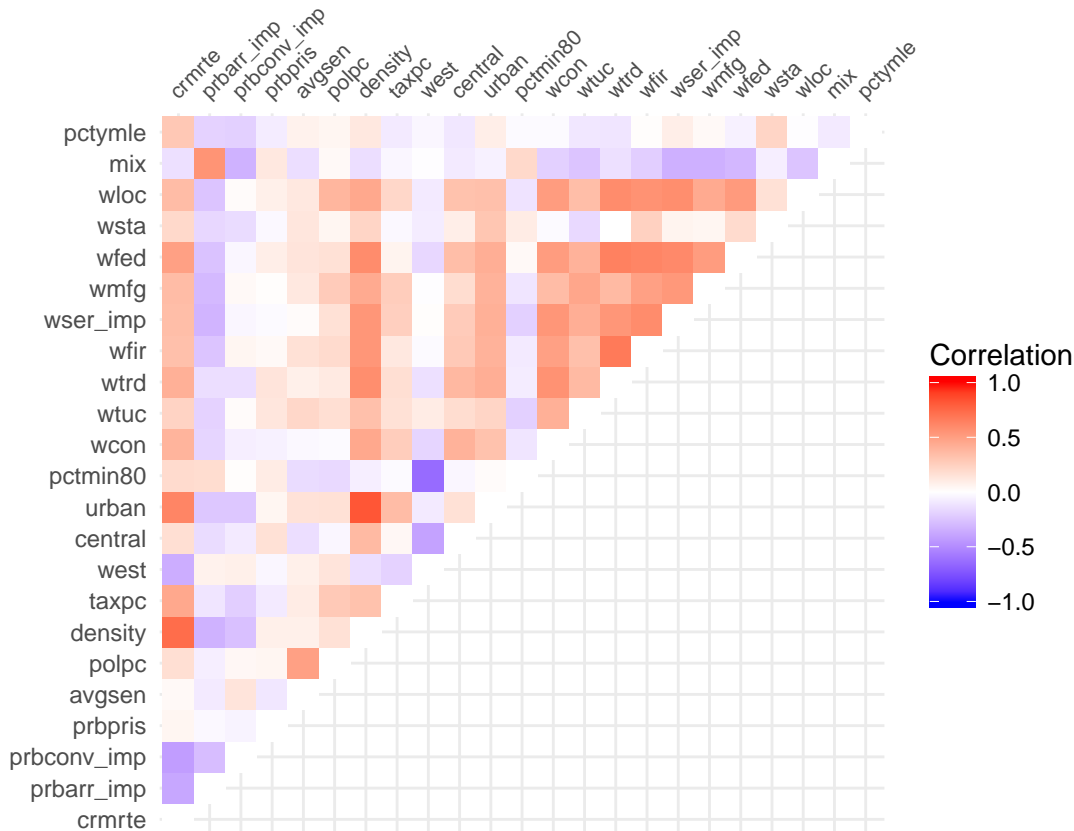


```
ind_vars_transf <- c("wcon", "wtuc", "wtrd", "wfir",
                "wser", "wser_imp", "wmfg", "wfed", "wsta", "wloc",
                "prbarr", "prbarr_imp", "prbconv", "prbconv_imp", "prbpris", "avgsen",
                 "polpc", "density", "taxpc", "pctmin80", "mix", "pctymle")
crime[paste(ind_vars_transf,"ln",sep = ".")] <- log(crime[ind_vars_transf])
crime[paste(ind_vars_transf,"sq",sep = ".")] <- (crime[ind_vars_transf])**2
```

```
ind_vars0 <- c("density", "west", "prbarr_imp", "prbconv")
crmrte_formula0 <- as.formula(paste("log(crmrte) ~", paste(ind_vars0, collapse = "+"), sep = ""))
crmrte_lm0 <- lm(crmrte_formula0, data = crime)
coeftest(crmrte_lm0)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -2.779520   0.158884 -17.4940 < 2.2e-16 ***
## density      0.137050   0.024811   5.5237 3.475e-07 ***
## west        -0.394264   0.076347  -5.1641 1.540e-06 ***
## prbarr_imp  -1.685787   0.352951  -4.7763 7.276e-06 ***
## prbconv     -0.686121   0.104887  -6.5416 4.183e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm0)$adj.r.squared
```

[1] 0.6713567

```
ind_vars1 <- c("density", "prbarr_imp", "prbconv", "taxpc.ln", "pctymle", "pctmin80")
crmrte_formula1 <- as.formula(paste("log(crmrte) ~", paste(ind_vars1, collapse = "+"), sep = ""))
crmrte_lm1 <- lm(crmrte_formula1, data = crime)
coeftest(crmrte_lm1)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -4.0506892  0.5099198 -7.9438 7.859e-12 ***
## density      0.1327965  0.0240191  5.5288 3.561e-07 ***
## prbarr_imp  -2.0565679  0.3439648 -5.9790 5.275e-08 ***
## prbconv     -0.7333852  0.1009028 -7.2682 1.730e-10 ***
## taxpc.ln     0.2279732  0.1253097  1.8193   0.07243 .
## pctymle      2.3122532  1.3903078  1.6631   0.10001
## pctmin80     0.0115781  0.0018476  6.2666 1.515e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm1)$adj.r.squared
```

[1] 0.7186894

```
ind_vars2 <- c("density", "west","prbarr_imp", "prbconv", "polpc.ln", "wfed.ln", "pctmin80",
               "west*polpc.ln")
crmrte_formula2 <- as.formula(paste("log(crmrte) ~", paste(ind_vars2, collapse = "+"), sep = ""))
crmrte_lm2 <- lm(crmrte_formula2, data = crime)
coeftest(crmrte_lm2)
```

```
##
## t test of coefficients:
##
##                Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   -2.1355351  1.6276340 -1.3120  0.193165
## density        0.0693844  0.0223425  3.1055  0.002609 **
## west          -5.0359292  0.9453904 -5.3268 8.580e-07 ***
## prbarr_imp    -1.9552511  0.2794704 -6.9963 6.523e-10 ***
## prbconv       -0.7397929  0.0814656 -9.0810 4.965e-14 ***
## polpc.ln       0.5915122  0.0996189  5.9377 6.684e-08 ***
## wfed.ln        0.5068172  0.2255788  2.2467  0.027344 *
## pctmin80       0.0097429  0.0020442  4.7660 8.026e-06 ***
## west:polpc.ln -0.7555459  0.1458665 -5.1797 1.559e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm2)$adj.r.squared
```

[1] 0.8149695

```
ind_vars_all <- c("prbarr_imp", "prbconv", "prbpris", "avgsen", "polpc.ln", "density", "taxpc.ln",
                  "west", "central", "urban", "pctmin80", "wcon.ln", "wtuc.ln", "wtrd.ln", "wfir.ln",
                  "wser_imp.ln", "wmfg.ln", "wfed.ln", "wsta.ln", "wloc.ln", "mix", "pctymle")
crmrte_formula_all <- as.formula(paste("log(crmrte) ~", paste(ind_vars_all, collapse = "+"), sep = ""))
```

```
crmrte_lm_all <- lm(crmrte_formula_all, data = crime)
coeftest(crmrte_lm_all)
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -5.1144125  3.4802606 -1.4695  0.146296
## prbarr_imp  -1.8479511  0.3519482 -5.2506 1.638e-06 ***
## prbconv     -0.7073694  0.0999706 -7.0758 1.036e-09 ***
## prbpris     -0.2360137  0.3695126 -0.6387  0.525153
## avgsen      -0.0235995  0.0125974 -1.8734  0.065316 .
## polpc.ln     0.2963644  0.1117292  2.6525  0.009935 **
## density      0.1217640  0.0407389  2.9889  0.003894 **
## taxpc.ln     0.1354969  0.1456482  0.9303  0.355506
## west        -0.1771316  0.1149754 -1.5406  0.128054
## central     -0.1583385  0.0793778 -1.9947  0.050081 .
## urban       -0.1712048  0.1890994 -0.9054  0.368466
## pctmin80     0.0086468  0.0027427  3.1527  0.002407 **
## wcon.ln      0.4073996  0.2388575  1.7056  0.092643 .
## wtuc.ln      0.1212886  0.1762608  0.6881  0.493718
## wtrd.ln      0.1522455  0.3127806  0.4867  0.628002
## wfir.ln     -0.1373597  0.2635058 -0.5213  0.603867
## wser_imp.ln -0.4375964  0.2435199 -1.7970  0.076782 .
## wmfg.ln     -0.0359514  0.1663831 -0.2161  0.829576
## wfed.ln      0.9354652  0.3484072  2.6850  0.009104 **
## wsta.ln     -0.3122219  0.2849330 -1.0958  0.277044
## wloc.ln     -0.0899533  0.4489397 -0.2004  0.841791
## mix         -0.3454249  0.4866422 -0.7098  0.480247
## pctymle      2.9288555  1.4126527  2.0733  0.041936 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(crmrte_lm_all)$adj.r.squared
```

```
## [1] 0.7832702
```

```
AIC(crmrte_lm0)
```

```
## [1] 53.74684
```

```
AIC(crmrte_lm2)
```

```
## [1] 5.137531
```

```
AIC(crmrte_lm_all)
```

```
## [1] 30.49119
```

```
par(mfrow=c(1,3))
plot(crmrte_lm0, which = 1)
plot(crmrte_lm2, which = 1)
plot(crmrte_lm_all, which = 1)
```
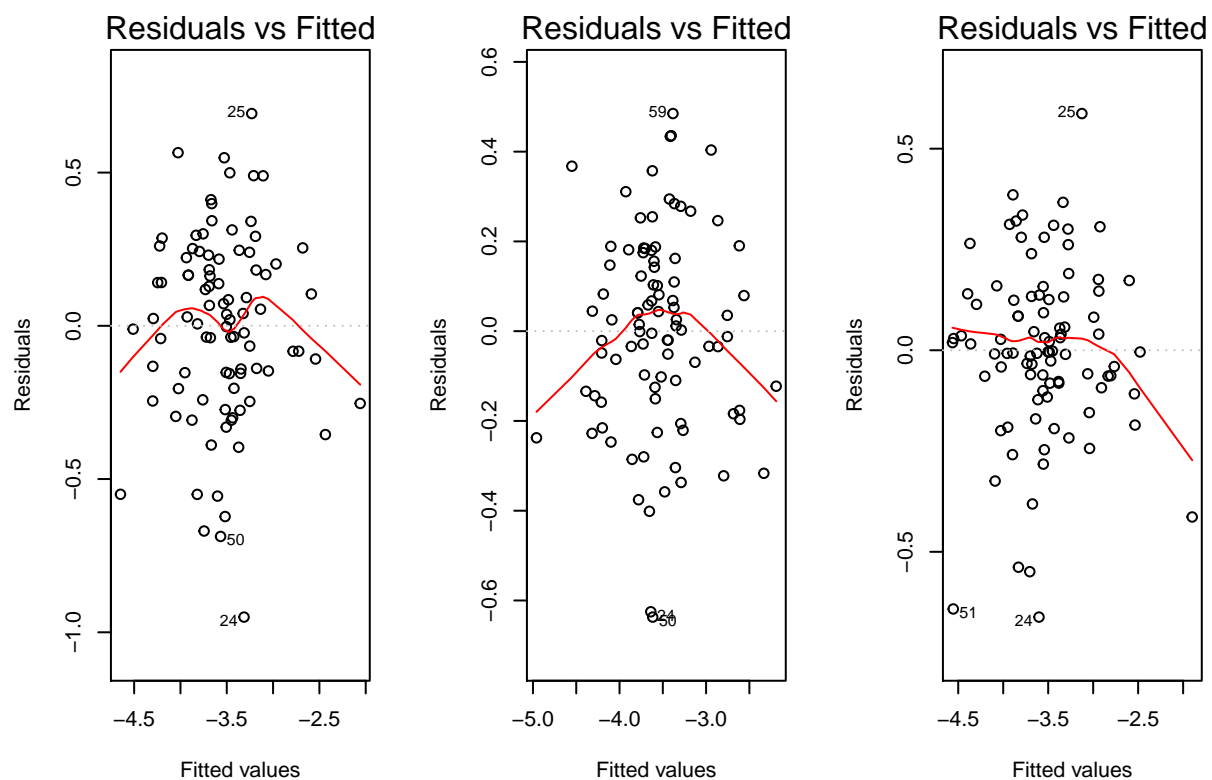
```r
par(mfrow=c(1,3))
plot(crmrte_lm0, which = 2)
plot(crmrte_lm2, which = 2)
plot(crmrte_lm_all, which = 2)
```

Normal Q–Q


```
par(mfrow=c(1,3))
plot(crmrte_lm0, which = 3)
plot(crmrte_lm2, which = 3)
plot(crmrte_lm_all, which = 3)
```
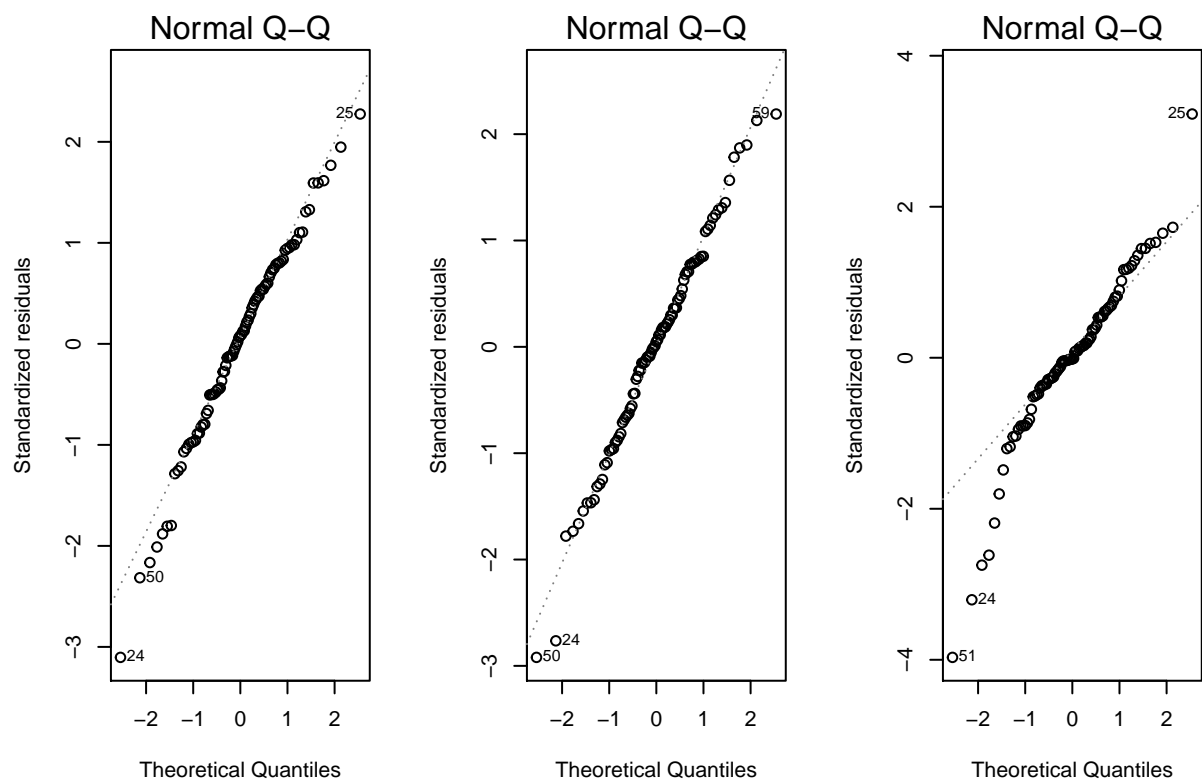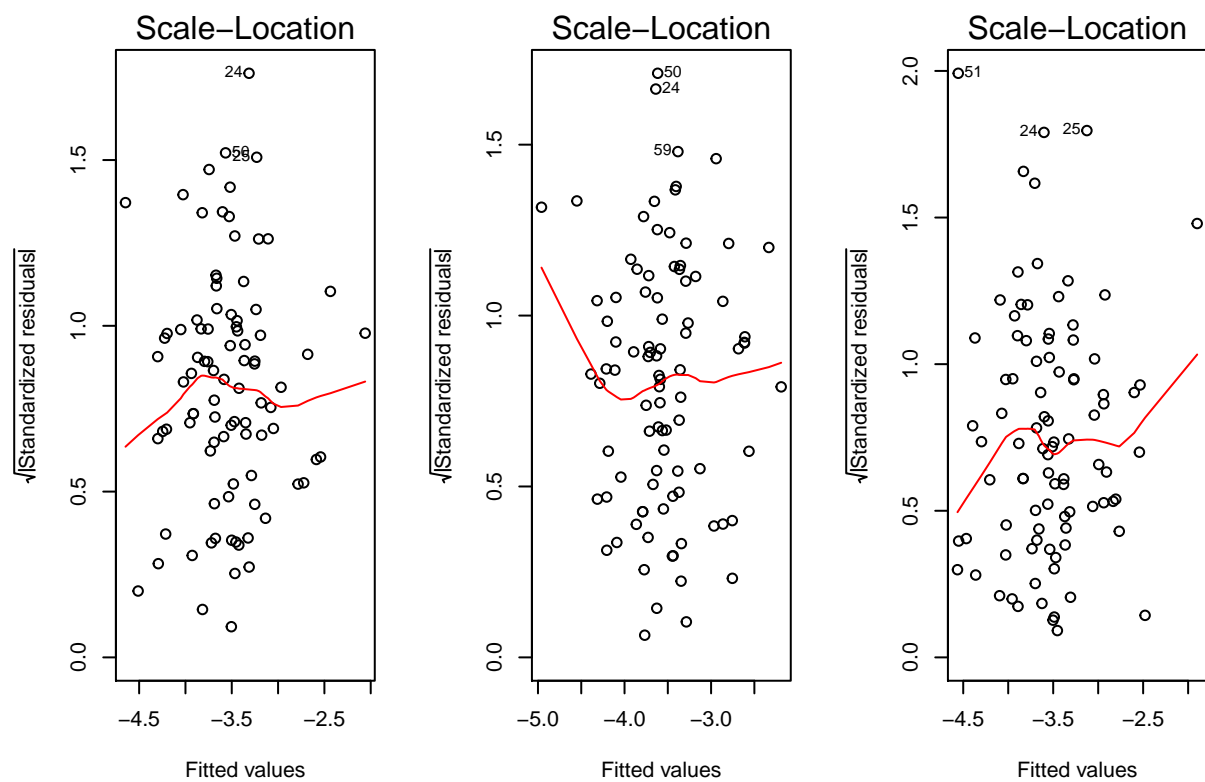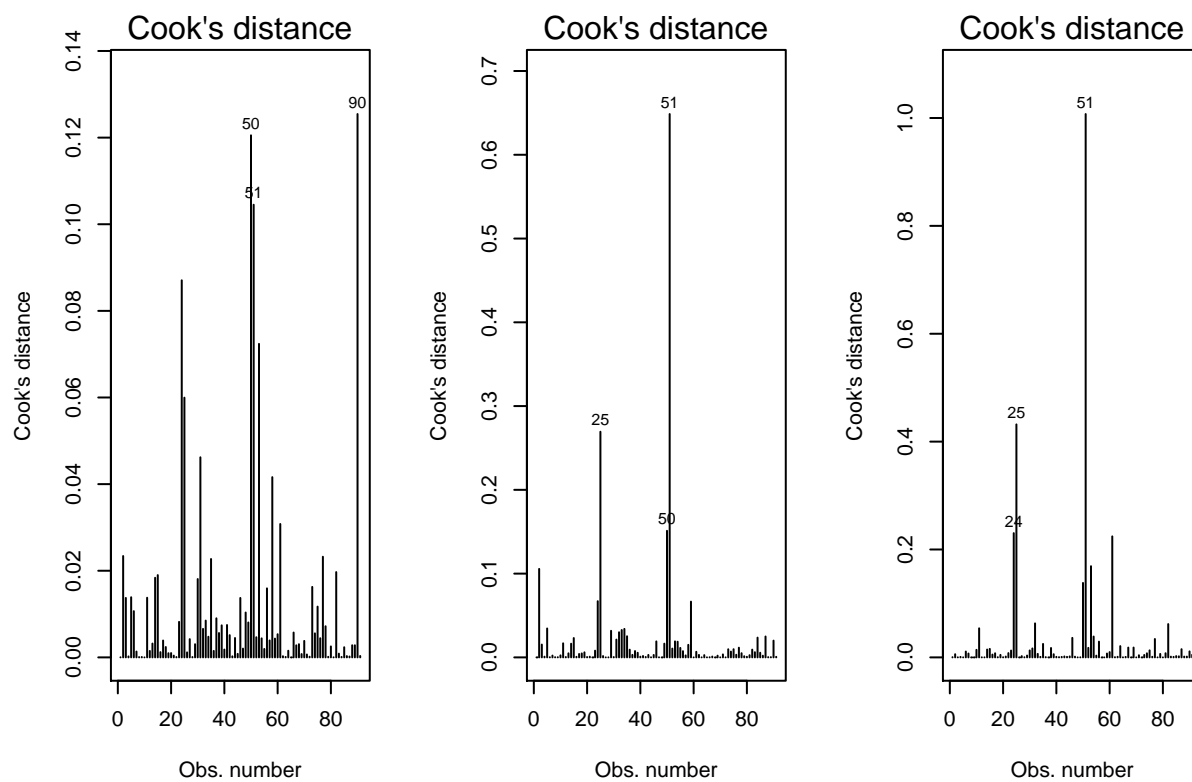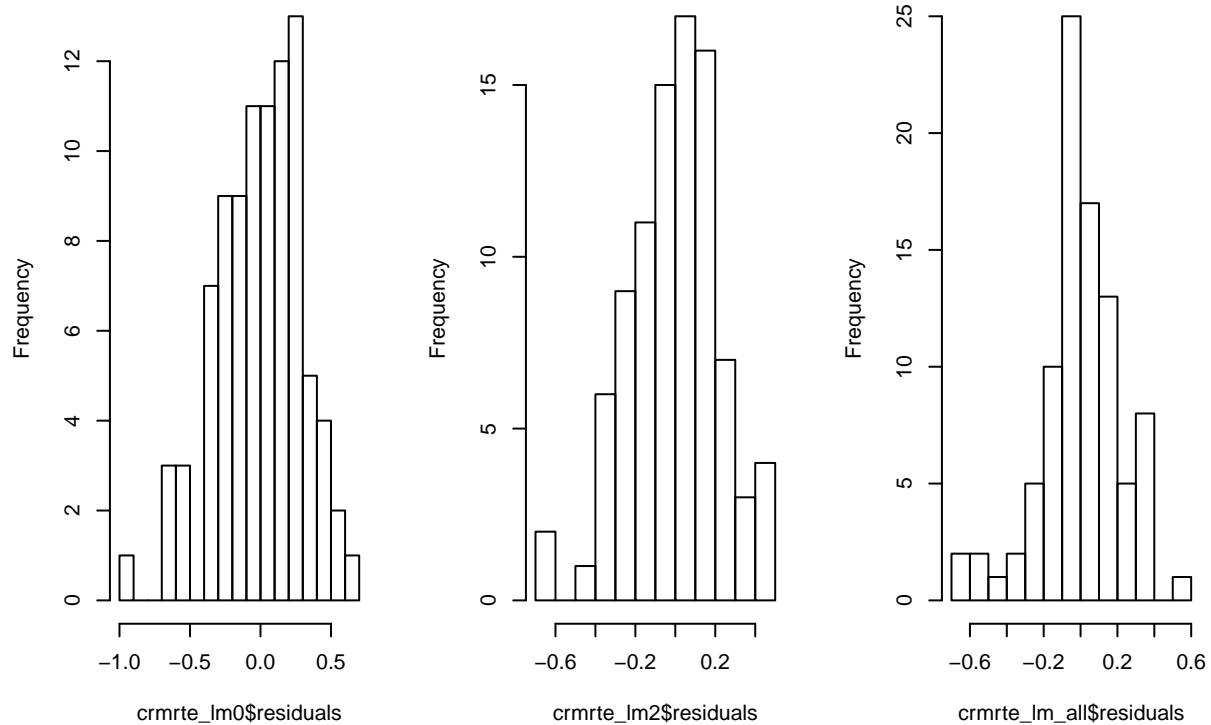
Scale–Location

```
par(mfrow=c(1,3))
plot(crmrte_lm0, which = 4)
plot(crmrte_lm2, which = 4)
plot(crmrte_lm_all, which = 4)
```

```r
par(mfrow=c(1,3))
hist(crmrte_lm0$residuals, breaks = 15)
hist(crmrte_lm2$residuals, breaks = 15)
hist(crmrte_lm_all$residuals, breaks = 15)
```

Histogram of crmrte_lm0$residu Histogram of crmrte_lm2$residu Histogram of crmrte_lm_all$resid


```r
se.crmrte_lm0 <- sqrt(diag(vcovHC(crmrte_lm0)))
se.crmrte_lm2 <- sqrt(diag(vcovHC(crmrte_lm2)))
se.crmrte_lm_all <- sqrt(diag(vcovHC(crmrte_lm_all)))
stargazer(crmrte_lm0, crmrte_lm2,crmrte_lm_all,
          type = "text", omit.stat = "f",
          se = list(se.crmrte_lm0, se.crmrte_lm2, se.crmrte_lm_all),
          star.cutoffs = c(0.05, 0.01, 0.001))
```

```
##
## ======================================================================
##                              Dependent variable:
##              ---------------------------------------------
##                                 log(crmrte)
##                   (1)             (2)             (3)
## ----------------------------------------------------------------------
## density         0.137***        0.069**         0.122*
##                 (0.027)         (0.024)         (0.062)
##
## taxpc.ln                                         0.135
##                                                 (0.342)
##
## west           -0.394***       -5.036**         -0.177
##                 (0.075)         (1.819)         (0.131)
##
## central                                         -0.158
##                                                 (0.086)
```

```
##
## urban                                                    -0.171
##                                                          (0.249)
##
## prbarr_imp            -1.686***       -1.955***        -1.848***
##                        (0.372)         (0.322)          (0.389)
##
## prbconv               -0.686***       -0.740***        -0.707***
##                        (0.144)         (0.104)          (0.158)
##
## prbpris                                                 -0.236
##                                                         (0.422)
##
## avgsen                                                  -0.024
##                                                         (0.016)
##
## polpc.ln                              0.592***           0.296
##                                        (0.161)          (0.318)
##
## wfed.ln                                 0.507            0.935
##                                        (0.337)          (0.546)
##
## wsta.ln                                                 -0.312
##                                                         (0.336)
##
## wloc.ln                                                 -0.090
##                                                         (0.861)
##
## mix                                                     -0.345
##                                                         (0.668)
##
## pctymle                                                 2.929*
##                                                         (1.470)
##
## pctmin80                              0.010***          0.009**
##                                        (0.002)          (0.003)
##
## west:polpc.ln                         -0.756**
##                                        (0.279)
##
## wcon.ln                                                  0.407
##                                                         (0.309)
##
## wtuc.ln                                                  0.121
##                                                         (0.314)
##
## wtrd.ln                                                  0.152
##                                                         (0.382)
##
## wfir.ln                                                 -0.137
##                                                         (0.399)
##
## wser_imp.ln                                             -0.438
##                                                         (0.349)
```

```
##
## wmfg.ln                                              -0.036
##                                                      (0.185)
##
## Constant                  -2.780***     -2.136       -5.114
##                           (0.202)       (2.764)      (5.451)
##
## ----------------------------------------------------------------
## Observations              91            91           91
## R2                        0.686         0.831        0.836
## Adjusted R2               0.671         0.815        0.783
## Residual Std. Error 0.313 (df = 86) 0.235 (df = 82) 0.254 (df = 68)
## ================================================================
## Note:                           *p<0.05; **p<0.01; ***p<0.001
```