project 2

Dizhou Wu

2023-11-22

Project 2

```
# Load library
library(countreg)
```

```
## Loading required package: MASS

## Loading required package: Formula
```

```
library(ROCR)
library(ggplot2)
library(foreign)
library(statmod)
```

```
##
## Attaching package: 'statmod'

## The following object is masked from 'package:countreg':
##
##     qresiduals
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(caret)
```

```
## Loading required package: lattice
```

```r
library(xlsx)
suppressMessages(library(car))
```

```r
# Load data
athletes <- read.xlsx("/home/wud18/Rstudio/STA712/project2/athletes_data.xlsx", sheetIndex = 1)
# Remove NA
athletes <- na.omit(athletes)
# Races under 10 seconds minus 1
athletes$Races.under.10.seconds <- athletes$Races.under.10.seconds-1
athletes
```

```
##     Date.of.birth                    Name   PB    ID Height.cm. X.Weight.kg.
## 1      1946-09-10               Jim Hines 9.95  20720        183           81
## 2      1955-09-20          Silvio Leonard 9.98  40973        173           64
## 3      1961-07-01              Carl Lewis 9.86   6473        188           90
## 4      1961-01-08            Calvin Smith 9.93   6367        178           69
## 5      1959-08-10             Mel Lattany 9.96  30160        175           78
## 7      1960-04-02        Linford Christie 9.87  26791        189           94
## 8      1965-03-18         Raymond Stewart 9.96  36637        178           73
## 9      1967-02-21           Leroy Burrell 9.85  26399        183           82
## 10     1966-02-20         Dennis Mitchell 9.91  10489        174           69
## 11     1967-10-02      Frankie Fredericks 9.86  13845        180           73
## 12     1969-01-20             Andre Cason 9.92   2654        170           70
## 13     1969-08-19        Olapade Adeniken 9.95  33434        186           78
## 14     1967-08-04           Michael Marsh 9.93  30428        178           75
## 15     1971-11-22         Davidson Ezinwa 9.94  10167        184           82
## 16     1972-06-17          Daniel Effiong 9.98   9052        187           80
## 17     1968-09-09            Jon Drummond 9.92  21961        175           75
## 18     1967-12-16          Donovan Bailey 9.84  11422        183           82
## 19     1967-07-12             Bruny Surin 9.84   6164        180           81
## 20     1973-12-30              Ato Boldon 9.86   4186        176           75
## 21     1974-07-23          Maurice Greene 9.79  29969        175           75
## 22     1973-03-30 Kareem Streete-Thompson 9.96  23495        183           84
## 23     1975-01-28          Tim Montgomery 9.92  43767        178           73
## 24     1975-02-24        Percival Spencer 9.98  35013        182           68
## 25     1977-12-28           Seun Ogunkoya 9.92  40352        180           86
## 26     1972-10-20       Vincent Henderson 9.95  45952        175           66
## 27     1976-03-30        Obadele Thompson 9.87  33349        175           67
## 28     1973-05-09     Leonard Myles-Mills 9.98  26307        172           66
## 29     1978-04-05          Dwain Chambers 9.97  11723        180           83
## 30     1975-09-17          Jason Gardener 9.98  19545        178           70
## 31     1974-01-27              Tim Harden 9.92  43738        178           79
## 32     1976-10-19             Coby Miller 9.98   8252        168           68
## 33     1978-01-19        Bernard Williams 9.94   4930        181           73
## 34     1978-11-22        Francis Obikwelu 9.86  13713        195           80
## 35     1978-01-14          Shawn Crawford 9.88  40509        181           80
## 36     1976-05-10       Joshua J. Johnson 9.95  20907        191           91
## 37     1974-12-05             Brian Lewis 9.99   5946        170           72
## 38     1976-04-05              Kim Collins 9.93  24788        175           64
## 39     1972-09-26          Patrick Johnson 9.93  34303        177           73
## 40     1975-11-22                Deji Aliu 9.95  10325        187           75
## 41     1978-10-27               John Capel 9.95  21377        180           82
## 42     1982-02-10            Justin Gatlin 9.74  23308        185           79
## 43     1976-10-10            Mickey Grimes 9.99  30705        185           84
## 44     1976-09-17          Uchenna Emedolu 9.97  45192        183           79
## 45     1982-11-23             Asafa Powell 9.72   4109        190           88
## 46     1976-09-02              Aziz Zakari 9.99   4377        177           73
## 47     1983-01-07               Marc Burns 9.96  28099        185           88
## 48     1984-10-11             Darrel Brown 9.99   9427        184           79
## 49     1982-11-16            Ronald Pognon 9.99  38218        184           70
## 50     1980-01-19            Leonard Scott 9.91  26327        181           84
## 51     1984-07-09           Olusoji Fasuba 9.85  51282        170           78
## 52     1982-08-09                Tyson Gay 9.69  45173        180           73
## 53     1978-04-24           Marcus Brunson 9.99  28377        185           77
## 54     1984-01-05           Derrick Atkins 9.91  10613        185           88
## 55     1986-01-31               Walter Dix 9.88  52892        178           84
## 56     1987-03-27           Samuel Francis 9.99  61273        190           80
## 57     1984-12-24         Wallace Spearmon 9.96  54727        190           80
## 58     1986-08-21               Usain Bolt 9.58  45032        196           88
## 59     1986-12-13           Travis Padgett 9.89  56297        174           80
## 60     1985-06-07         Richard Thompson 9.82  61016        187           79
## 61     1982-12-22            Rodney Martin 9.95  37847        175           79
## 62     1984-04-10               Mark Jelks 9.99  50988        170           66
## 63     1977-12-04            Darvis Patton 9.89   9546        183           75
## 64     1985-05-02           Ivory Williams 9.94  18479        173           77
## 65     1985-10-11             Nesta Carter 9.78  59574        178           70
## 66     1984-07-03         Churandy Martina 9.91  40902        178           75
## 67     1982-10-06           Michael Frater 9.88  30332        170           67
## 68     1986-09-09            Daniel Bailey 9.91  55393        173           70
## 69     1985-04-24             Mike Rodgers 9.85  58527        178           73
## 70     1989-12-26              Yohan Blake 9.69  69837        181           79
## 71     1981-06-12            Lerone Clarke 9.99  26397        174           66
## 72     1990-06-11      Christophe Lemaitre 9.92  84220        189           74
## 73     1985-07-13            Trell Kimmons 9.95  58466        178           77
## 74     1989-04-13              Ryan Bailey 9.88  88407        193           98
## 75     1985-10-30           Mario Forsythe 9.95  59594        173           68
## 76     1982-11-29           Steve Mullings 9.80  41985        173           68
## 77     1987-03-11      Ngonidzashe Makusha 9.89  83912        178           73
## 78     1990-04-07          Nickel Ashmeade 9.90  79234        184           87
## 79     1988-03-08           Keston Bledman 9.86  66064        183           75
## 80     1990-04-05            Mookie Salaam 9.97 105851        180           73
## 81     1984-01-01      Jaysuma Saidy Ndure 9.99  51645        192           72
## 82     1989-11-27              Harry Adams 9.96  72677        182           81
## 83     1989-10-11              Kemar Hyman 9.95  70904        178           74
## 84     1992-01-10        Kemar Bailey-Cole 9.92 104003        195           86
## 85     1990-01-05              Isiah Young 9.89 111759        183           75
## 86     1989-12-12          Dentarius Locke 9.96 105447        170           68
## 87     1988-02-23         Gabriel Mvumvure 9.98  83908        172           75
## 88     1991-07-04           Charles Silmon 9.98  93272        175           72
## 89     1987-09-05            James Dasaolu 9.91  82420        180           75
## 90     1992-02-27             Jimmy Vicaut 9.86 103087        188           83
## 91     1985-05-25            Simon Magakwe 9.98  90520        177           73
## 92     1992-07-20           Kemarley Brown 9.93 128631        180           72
## 93     1994-03-05            Chijindu Ujah 9.96 126069        180           75
## 94     1995-07-10          Trayvon Bromell 9.76 136552        175           71
## 95     1991-05-15             Femi Ogunode 9.91  96247        183           79
## 96     1992-05-15           Clayton Vaughn 9.93 114678        173           77
## 97     1994-11-10          Andre De Grasse 9.89 140003        180           73
## 98     1993-11-13           Bryce Robinson 9.99 135937        178           75
## 99     1993-12-15             Marvin Bracy 9.85 123692        175           74
## 100    1989-08-29              Su Bingtian 9.83  84128        172           64
## 101    1993-10-06              Adam Gemili 9.97 125508        178           73
## 102    1992-07-13           Diondre Batson 9.94 117751        188           75
## 103    1993-03-05               Beejay Lee 9.99 150807        168           72
## 104    1992-09-18           Quentin Butler 9.96 129036        175           70
## 105    1993-09-21            Akani Simbine 9.84 135967        174           67
## 106    1993-07-16      Henricho Bruintjies 9.97 139896        178           70
## 107    1991-12-15            Andrew Fisher 9.94 115854        168           64
## 108    1992-07-15        Wayde Van Niekerk 9.94 122672        183           73
## 109    1994-04-25              Omar McLeod 9.99 127853        180           73
## 110    1991-03-19               Ameer Webb 9.94 117296        175           75
## 111    1986-11-11        Ben Youssef Meïté 9.96  67421        179           70
## 112    1993-12-30         Senoj-Jay Givans 9.96 136582        178           73
## 113    1992-05-27              Aaron Brown 9.96 115675        185           79
## 114    1989-05-04          Jak Ali Harvey  9.92  91407        182           73
## 115    1986-01-21          Rondel Sorrillo 9.99  80875        178           62
## 116    1996-03-06        Christian Coleman 9.76 163672        175           73
## 117    1988-10-11              Joel Fearon 9.96  95593        175           77
## 118    1995-09-26              Thando Roto 9.95 140947        173           68
## 119    1993-10-15             Ronnie Baker 9.83 140764        178           73
## 120    1994-08-28              Odean Skeen 9.98 115857        181           75
## 121    1994-04-02 Nethaneel Mitchell-Blake 9.99 128908        186           75
## 122    1994-09-11          Cameron Burrell 9.93 130614        173           68
## 123    1994-01-29      Christopher Belcher 9.93 140854        175           75
## 124    1993-01-07             Julian Forte 9.91 104013        186           73
## 125    1990-05-29            Ramil Guliyev 9.97  81841        187           73
## 126    1995-12-15          Yoshihide Kiryū 9.98 133239        175           69
## 127    1995-09-23          Kendal Williams 9.99 129731        180           73
## 128    1996-08-05             Jaylen Bacon 9.97 164555        183           75
## 130    1995-07-13           Zharnel Hughes 9.83 136196        190           79
## 131    1997-07-18               Noah Lyles 9.83 147365        180           70
## 133    1993-08-17               Xie Zhenye 9.97 118459        185           80
## 134    1998-06-15            Filippo Tortu 9.99 163520        182           68
## 135    1988-06-15        Barakat Al Harthi 9.97  84020        172           64
## 136    1993-06-10          Tyquendo Tracey 9.96 135524        179           75
## 137    1996-02-29            Reece Prescod 9.93 141072        184           68
## 138    1991-11-12           Roberto Skyers 9.98  90880        187           83
## 139    1996-10-07           Divine Oduduru 9.86 141457        175           70
## 140    1999-03-06   Abdul Hakim Sani Brown 9.97 169716        187           73
## 141    1996-07-31         Cravon Gillespie 9.93 162131        175           66
## 142    1997-03-18              Mario Burke 9.98 138213        175           64
## 143    1995-05-13               Yuki Koike 9.98 133238        178           72
## 145    1997-12-03           Michael Norman 9.86 163895        183           73
## 146    1998-01-14   Benjamin Azamati-Kwaku 9.90 330232        170           60
## 147    1994-07-09               Kyree King 9.97 137530        181           68
## 148    1999-07-15         Jo’Vaughn Martin 9.94 334230        178           74
## 149    1995-05-07              Fred Kerley 9.76 151446        191           93
## 150    1994-09-26           Marcell Jacobs 9.80 141067        184           73
## 151    1998-05-12        Tlotliso Leotlela 9.94 142499        180           70
## 152    1992-06-10           Ryota Yamagata 9.95 114877        176           70
## 154    2001-11-12           Micah Williams 9.86 294929        173           80
## 155    2000-03-03            Enoch Adegoke 9.98 211224        170           71
## 156    1996-01-02         Ferdinand Omurwa 9.77 217943        175           83
## 159    1997-01-12        Joseph Paul Amoah 9.94 225091        180           68
## 160    2003-06-07           Letsile Tebogo 9.91 343821        184           72
## 164    1994-08-22              Elijah Hall 9.90 125641        174           68
## 165    1991-04-15          Emmanuel Matadi 9.98 117969        183           85
## 168    1996-11-21             Reynier Mena 9.99 138833        174           79
## 169    1994-05-19        Méba-Mickaël Zeze 9.99 128539        175           65
## 173    2003-01-29         Udodi Onwuzurike 9.91 369021        178           70
## 177    2001-02-01        Shaun Maswanganyi 9.91 201147        185           77
## 180    1993-08-17           Emmanuel Eseme 9.96 326821        184           83
##     Races.under.10.seconds Total.number.of.races no..Date.of.birth   NA.
## 1                        0                    20             17055     1
## 2                        0                    66             20352     2
## 3                       14                   160             22463     3
## 4                        3                   138             22289     4
## 5                        0                    99             21772     5
## 7                        8                   194             22008     7
## 8                        1                   144             23819     8
## 9                        8                   111             24524     9
## 10                      11                   259             24158    10
## 11                      26                   220             24747    11
## 12                       4                    93             25223    12
## 13                       2                   122             25434    13
## 14                       2                    94             24688    14
## 15                       3                   108             26259    15
## 16                       1                    66             26467    16
## 17                       9                   185             25090    17
## 18                      14                   149             24822    18
## 19                       6                   189             24665    19
## 20                      27                   150             27028    20
## 21                      51                   160             27233    21
## 22                       0                    92             26753    22
## 23                       5                   160             27422    23
## 24                       0                    34             27449    24
## 25                       3                    69             28487    25
## 26                       0                    74             26592    26
## 27                       2                   140             27849    27
## 28                       1                   103             26793    28
## 29                       3                   275             28585    29
## 30                       0                   141             27654    30
## 31                       0                    99             27056    31
## 32                       2                   134             28052    32
## 33                       4                   196             28509    33
## 34                       7                   308             28816    34
## 35                       8                   153             28504    35
## 36                       0                    81             27890    36
## 37                       0                   175             27368    37
## 38                       9                   358             27855    38
## 39                       0                   230             26568    39
## 40                       1                   248             27720    40
## 41                       1                   100             28790    41
## 42                      63                   216             29992    42
## 43                       0                   143             28043    43
## 44                       0                   185             28020    44
## 45                      96                   257             30278    45
## 46                       1                   225             28005    46
## 47                       2                   219             30323    47
## 48                       0                   137             30966    48
## 49                       0                   207             30271    49
## 50                       2                    98             29239    50
## 51                       1                   131             30872    51
## 52                      37                   162             30172    52
## 53                       0                   104             28604    53
## 54                       2                   116             30686    54
## 55                       9                   143             31443    55
## 56                       0                   116             31863    56
## 57                       0                    65             31040    57
## 58                      52                    86             31645    58
## 59                       2                   118             31759    59
## 60                      12                   187             31205    60
## 61                       0                    72             30307    61
## 62                       0                   106             30782    62
## 63                       7                   138             28463    63
## 64                       2                   104             31169    64
## 65                      28                   228             31331    65
## 66                       5                   279             30866    66
## 67                       6                   292             30230    67
## 68                       6                   154             31664    68
## 69                      45                   403             31161    69
## 70                      45                   214             32868    70
## 71                       1                   168             29749    71
## 72                       6                   151             33035    72
## 73                       0                   151             31241    73
## 74                       8                    82             32611    74
## 75                       1                   120             31350    75
## 76                       3                   113             30284 76[7]
## 77                       1                    20             31847    77
## 78                      10                   109             32970    78
## 79                       4                   240             32210    79
## 80                       1                    76             32968    80
## 81                       0                   172             30682    81
## 82                       0                    55             32839    82
## 83                       0                   191             32792    83
## 84                      10                   130             33613    84
## 85                       9                   159             32878    85
## 86                       1                   118             32854    86
## 87                       0                    68             32196    87
## 88                       0                   112             33423    88
## 89                       1                   120             32025    89
## 90                      21                   169             33661    90
## 91                       0                   174             31192    91
## 92                       0                    41             33805    92
## 93                       3                   148             34398    93
## 94                      29                   108             34890 94[8]
## 95                       6                    89             33373    95
## 96                       0                    58             33739    96
## 97                      18                   107             34648 97[9]
## 98                       0                    67             34286    98
## 99                      21                   116             34318    99
## 100                      9                   151             32749   100
## 101                      0                   101             34248   101
## 102                      0                    68             33798   102
## 103                      0                    75             34033   103
## 104                      0                    59             33865   104
## 105                     38                   220             34233   105
## 106                      0                   147             34166   106
## 107                      0                   114             33587   107
## 108                      1                    21             33800   108
## 109                      0                     5             34449   109
## 110                      0                    82             33316   110
## 111                      7                   110             31727   111
## 112                      0                    65             34333   112
## 113                      1                   153             33751   113
## 114                      1                   186             32632   114
## 115                      0                    86             31433   115
## 116                     30                    84             35130   116
## 117                      0                    94             32427   117
## 118                      0                   112             34968   118
## 119                     21                    89             34257   119
## 120                      0                    47             34574   120
## 121                      0                    46             34426   121
## 122                      1                    92             34588   122
## 123                      0                    80             34363   123
## 124                      3                   104             33976   124
## 125                      0                   123             33022   125
## 126                      0                   153             35048   126
## 127                      1                   133             34965   127
## 128                      0                    66             35282   128
## 130                     13                    84             34893   130
## 131                     17                    77             35629   131
## 133                      1                   118             34198   133
## 134                      0                    63             35961   134
## 135                      0                   158             32309   135
## 136                      1                   145             34130   136
## 137                      5                    58             35124   137
## 138                      0                    92             33554   138
## 139                      2                    61             35345   139
## 140                      3                    90             36225   140
## 141                      3                    77             35277   141
## 142                      0                   122             35507   142
## 143                      0                   135             34832   143
## 145                      2                     9             35767   145
## 146                      3                    39             35809   146
## 147                      3                   108             34524   147
## 148                      0                    27             36356   148
## 149                     25                    40             34826   149
## 150                      5                    75             34603   150
## 151                      0                    77             35927   151
## 152                      0                   156             33765   152
## 154                      5                    47             37207   154
## 155                      0                    94             36588   155
## 156                     12                    92             35066   156
## 159                      0                    50             35442   159
## 160                      5                    29             37779   160
## 164                      1                    39             34568   164
## 165                      2                   118             33343   165
## 168                      0                   166             35390   168
## 169                      0                   109             34473   169
## 173                      1                    38             37650   173
## 177                      1                    67             36923   177
## 180                      1                    59             34198   180
```

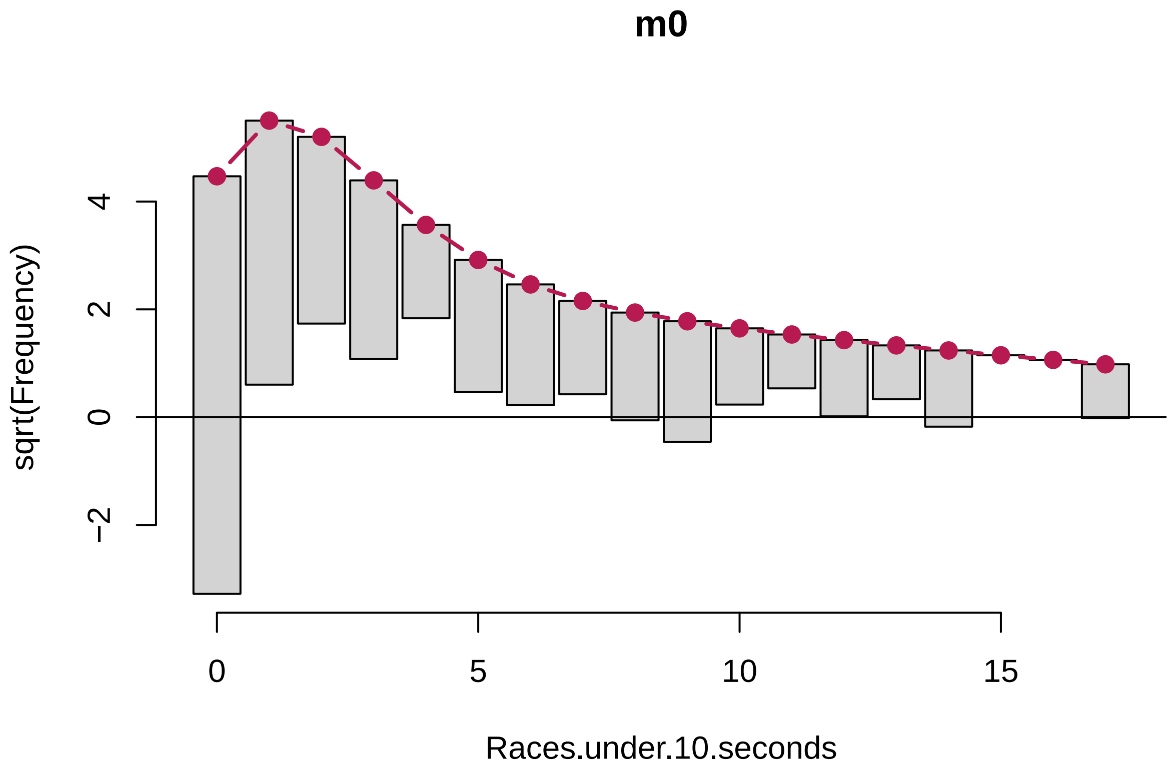
```
# Load necessary library
library(dplyr)

# Calculate detailed summary statistics for all variables
detailed_stats <- athletes %>%
  summarise(
    Mean_PB = mean(PB, na.rm = TRUE),
    Median_PB = median(PB, na.rm = TRUE),
    SD_PB = sd(PB, na.rm = TRUE),
    Min_PB = min(PB, na.rm = TRUE),
    Max_PB = max(PB, na.rm = TRUE),
    IQR_PB = IQR(PB, na.rm = TRUE),

    Mean_Height = mean(Height.cm., na.rm = TRUE),
    Median_Height = median(Height.cm., na.rm = TRUE),
    SD_Height = sd(Height.cm., na.rm = TRUE),
    Min_Height = min(Height.cm., na.rm = TRUE),
    Max_Height = max(Height.cm., na.rm = TRUE),
    IQR_Height = IQR(Height.cm., na.rm = TRUE),

    Mean_Weight = mean(X.Weight.kg., na.rm = TRUE),
    Median_Weight = median(X.Weight.kg., na.rm = TRUE),
    SD_Weight = sd(X.Weight.kg., na.rm = TRUE),
    Min_Weight = min(X.Weight.kg., na.rm = TRUE),
    Max_Weight = max(X.Weight.kg., na.rm = TRUE),
    IQR_Weight = IQR(X.Weight.kg., na.rm = TRUE),

    Mean_RacesUnder10 = mean(Races.under.10.seconds, na.rm = TRUE),
    Median_RacesUnder10 = median(Races.under.10.seconds, na.rm = TRUE),
    SD_RacesUnder10 = sd(Races.under.10.seconds, na.rm = TRUE),
    Min_RacesUnder10 = min(Races.under.10.seconds, na.rm = TRUE),
    Max_RacesUnder10 = max(Races.under.10.seconds, na.rm = TRUE),
    IQR_RacesUnder10 = IQR(Races.under.10.seconds, na.rm = TRUE),

    Mean_TotalRaces = mean(Total.number.of.races, na.rm = TRUE),
    Median_TotalRaces = median(Total.number.of.races, na.rm = TRUE),
    SD_TotalRaces = sd(Total.number.of.races, na.rm = TRUE),
    Min_TotalRaces = min(Total.number.of.races, na.rm = TRUE),
    Max_TotalRaces = max(Total.number.of.races, na.rm = TRUE),
    IQR_TotalRaces = IQR(Total.number.of.races, na.rm = TRUE),

    Mean_DOB = mean(no..Date.of.birth, na.rm = TRUE),
    Median_DOB = median(no..Date.of.birth, na.rm = TRUE),
    SD_DOB = sd(no..Date.of.birth, na.rm = TRUE),
    Min_DOB = min(no..Date.of.birth, na.rm = TRUE),
    Max_DOB = max(no..Date.of.birth, na.rm = TRUE),
    IQR_DOB = IQR(no..Date.of.birth, na.rm = TRUE)
  )

# Print the detailed statistics
print(detailed_stats)
```

```
##   Mean_PB Median_PB      SD_PB Min_PB Max_PB IQR_PB Mean_Height Median_Height
## 1   9.918      9.94 0.07126706   9.58   9.99 0.0825    179.6562         179.5
##   SD_Height Min_Height Max_Height IQR_Height Mean_Weight Median_Weight
## 1  5.979048        168        196          9    74.94375            74
##   SD_Weight Min_Weight Max_Weight IQR_Weight Mean_RacesUnder10
## 1  6.761198         60         98          9           6.48125
##   Median_RacesUnder10 SD_RacesUnder10 Min_RacesUnder10 Max_RacesUnder10
## 1                   1        13.18708                0               96
##   IQR_RacesUnder10 Mean_TotalRaces Median_TotalRaces SD_TotalRaces
## 1                6        123.8688               112      67.09086
##   Min_TotalRaces Max_TotalRaces IQR_TotalRaces Mean_DOB Median_DOB   SD_DOB
## 1              5            403           77.5 31339.29    32259.5 3886.477
##   Min_DOB Max_DOB IQR_DOB
## 1   17055   37779    5814
```

```
# Fit a poisson model
m0 <- glm(Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races + X.Weight.kg. + Height.cm
summary(m0)
```

```
##
## Call:
## glm(formula = Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races +
##     X.Weight.kg. + Height.cm. + PB, family = poisson, data = athletes)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            1.167e+02  3.958e+00  29.489  < 2e-16 ***
## no..Date.of.birth      7.542e-06  1.090e-05   0.692 0.489003
## Total.number.of.races  6.630e-03  3.870e-04  17.133  < 2e-16 ***
## X.Weight.kg.           2.318e-02  7.014e-03   3.305 0.000948 ***
## Height.cm.            -2.958e-02  7.367e-03  -4.016 5.93e-05 ***
## PB                    -1.140e+01  3.328e-01 -34.251  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2475.71  on 159  degrees of freedom
## Residual deviance:  656.96  on 154  degrees of freedom
## AIC: 1013.5
##
## Number of Fisher Scoring iterations: 5
```

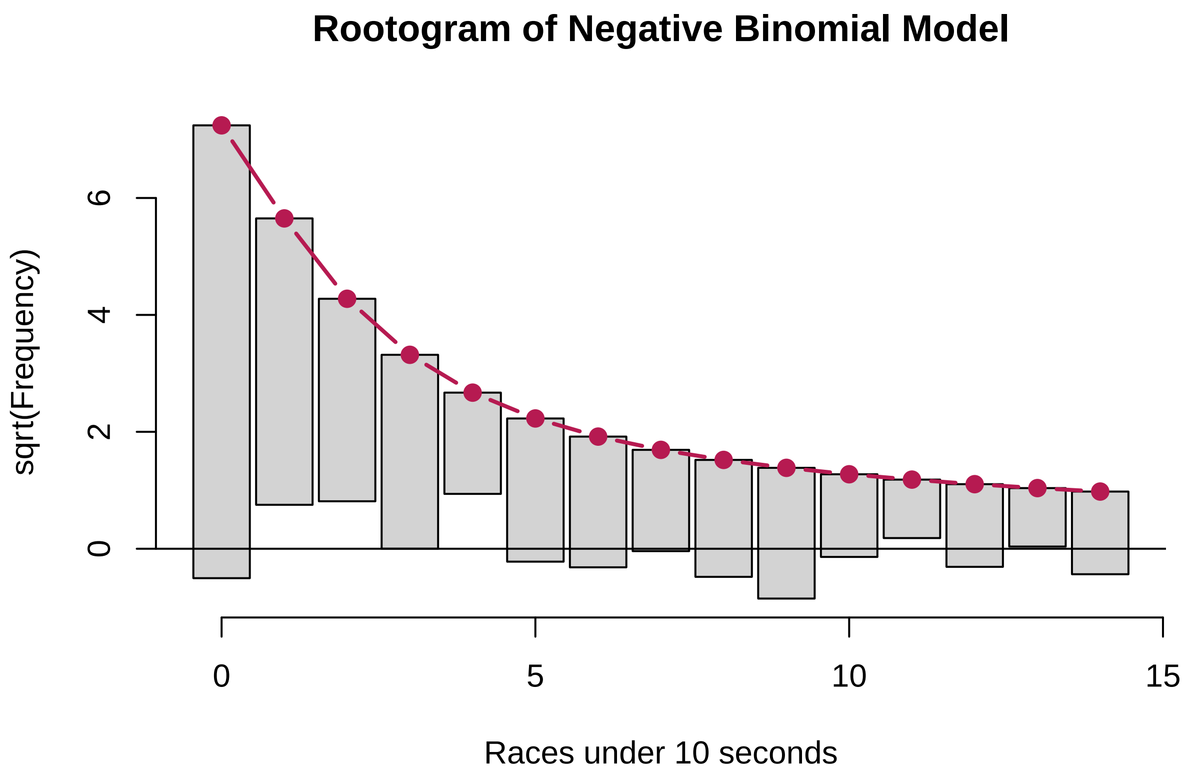
```
rootogram(m0)
```



```
# Fit a negative binomial model
m1 <- glm.nb(Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races + X.Weight.kg. + Height
summary(m1)
```

```
##
## Call:
## glm.nb(formula = Races.under.10.seconds ~ no..Date.of.birth +
##     Total.number.of.races + X.Weight.kg. + Height.cm. + PB, data = athletes,
##     init.theta = 1.481785106, link = log)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            1.773e+02  1.219e+01  14.539  < 2e-16 ***
## no..Date.of.birth     -2.109e-06  2.564e-05  -0.082    0.934
## Total.number.of.races  5.918e-03  1.314e-03   4.503  6.7e-06 ***
## X.Weight.kg.          -3.480e-03  1.741e-02  -0.200    0.842
## Height.cm.             2.170e-02  1.901e-02   1.142    0.254
## PB                    -1.822e+01  1.153e+00 -15.807  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.4818) family taken to be 1)
##
##     Null deviance: 520.88  on 159  degrees of freedom
## Residual deviance: 163.61  on 154  degrees of freedom
## AIC: 685.25
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  1.482
##           Std. Err.:  0.277
##
##  2 x log-likelihood:  -671.246
```

```
rootogram(m1, main = "Rootogram of Negative Binomial Model", xlab = "Races under 10 seconds")
```



```
# 95% Confidence Interval
conf_int <- confint(m1, level = 0.95)
```

```
## Waiting for profiling to be done...
```

```
print(conf_int)
```

```
##                               2.5 %        97.5 %
## (Intercept)            1.492405e+02  2.070541e+02
## no..Date.of.birth     -5.088661e-05  4.650843e-05
## Total.number.of.races  3.273284e-03  8.754190e-03
## X.Weight.kg.          -3.738078e-02  3.087630e-02
## Height.cm.            -1.786242e-02  6.152526e-02
## PB                    -2.118165e+01 -1.544397e+01
```

```
library(pscl)
```

```
## Registered S3 methods overwritten by 'pscl':
##   method                 from
##   print.zeroinfl         countreg
##   print.summary.zeroinfl countreg
##   summary.zeroinfl       countreg
##   coef.zeroinfl          countreg
##   vcov.zeroinfl          countreg
##   logLik.zeroinfl        countreg
##   predict.zeroinfl       countreg
##   residuals.zeroinfl     countreg
##   fitted.zeroinfl        countreg
##   terms.zeroinfl         countreg
##   model.matrix.zeroinfl  countreg
##   extractAIC.zeroinfl    countreg
##   print.hurdle           countreg
##   print.summary.hurdle   countreg
##   summary.hurdle         countreg
##   coef.hurdle            countreg
##   vcov.hurdle            countreg
##   logLik.hurdle          countreg
##   predict.hurdle         countreg
##   residuals.hurdle       countreg
##   fitted.hurdle          countreg
##   terms.hurdle           countreg
##   model.matrix.hurdle    countreg
##   extractAIC.hurdle      countreg
```

```
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
```

```
##
## Attaching package: 'pscl'
```

```
## The following objects are masked from 'package:countreg':
##
##     hurdle, hurdle.control, hurdletest, zeroinfl, zeroinfl.control
```

```
m2 <- hurdle(Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races + X.Weight.kg. + Height
```

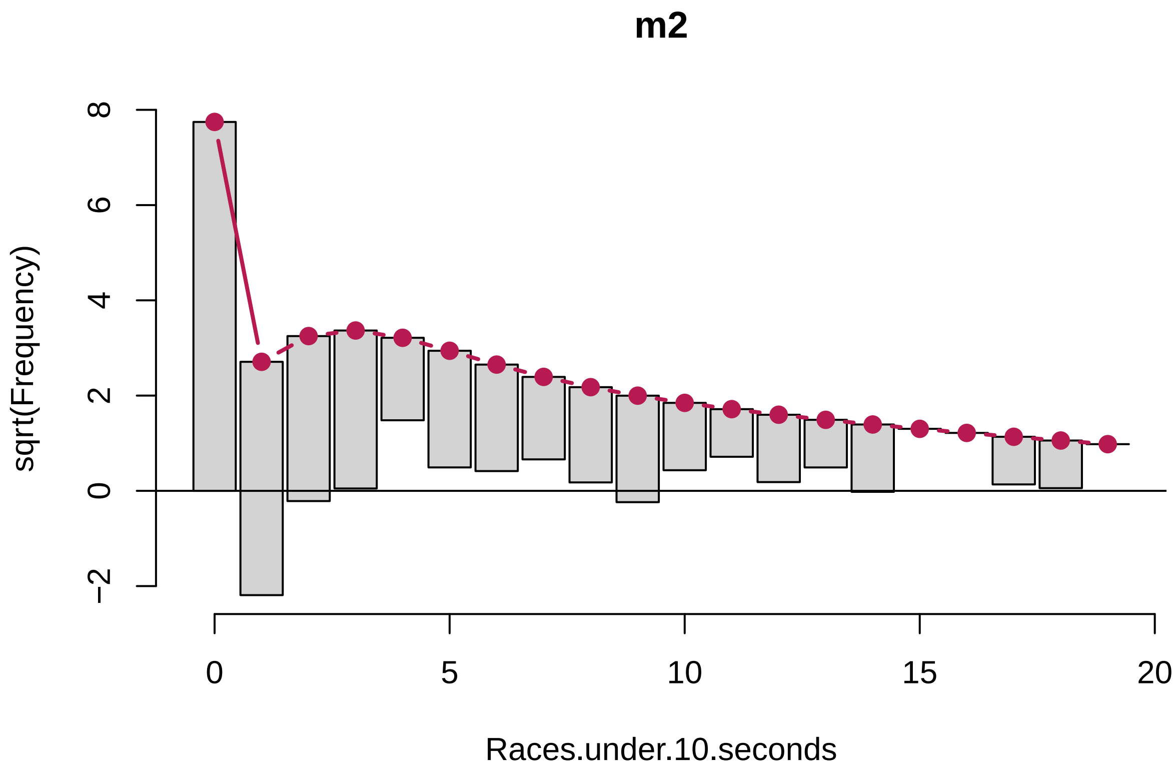
```
## Warning in value[[3L]](cond): system is computationally singular: reciprocal
## condition number = 5.5158e-27FALSE
```

```
summary(m2)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## hurdle(formula = Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races +
##     X.Weight.kg. + Height.cm. + PB, data = athletes, dist = "poisson",
##     zero.dist = "binomial")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -3.7330 -0.8296 -0.3776  0.3952  6.9322
##
## Count model coefficients (truncated poisson with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            9.919e+01         NA      NA       NA
## no..Date.of.birth      1.554e-05         NA      NA       NA
## Total.number.of.races  5.691e-03         NA      NA       NA
## X.Weight.kg.           1.926e-02         NA      NA       NA
## Height.cm.            -2.231e-02         NA      NA       NA
## PB                    -9.711e+00         NA      NA       NA
## Zero hurdle model coefficients (binomial with logit link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            5.114e+02  9.061e+01   5.644 1.66e-08 ***
## no..Date.of.birth     -1.794e-05        NaN     NaN      NaN
## Total.number.of.races  1.335e-02  4.768e-03   2.799  0.00512 **
## X.Weight.kg.          -1.863e-02  4.549e-02  -0.410  0.68213
## Height.cm.             6.079e-02  5.075e-02   1.198  0.23096
## PB                    -5.243e+01  9.208e+00  -5.694 1.24e-08 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Number of iterations in BFGS optimization: 15
## Log-likelihood: -430.6 on 12 Df
```

```
rootogram(m2)
```



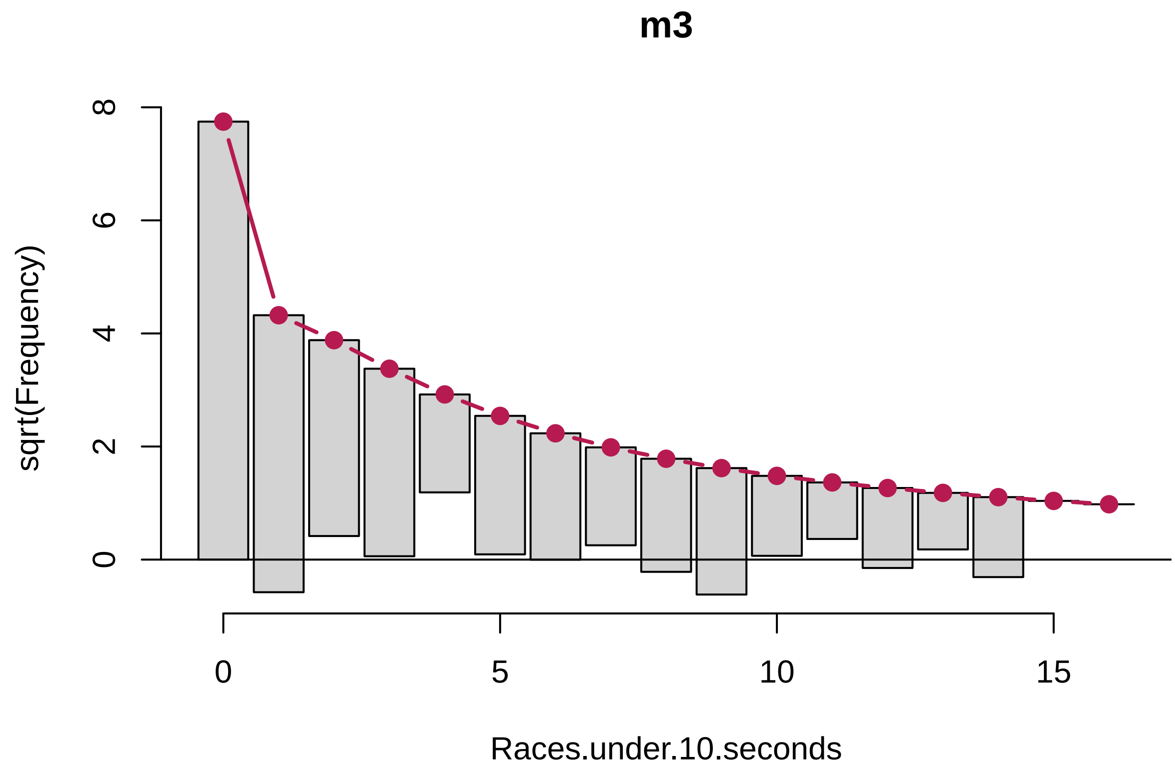
```
m3 <- hurdle(Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races + X.Weight.kg. + Height
summary(m3)
```

```
## Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
##
## Call:
## hurdle(formula = Races.under.10.seconds ~ no..Date.of.birth + Total.number.of.races +
##     X.Weight.kg. + Height.cm. + PB, data = athletes, dist = "negbin",
##     zero.dist = "binomial")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.2363 -0.6148 -0.2895  0.3656  3.8141
##
## Count model coefficients (truncated negbin with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            1.376e+02  1.434e+01   9.600  < 2e-16 ***
## no..Date.of.birth      6.585e-06        NaN     NaN      NaN
## Total.number.of.races  5.216e-03  1.201e-03   4.342 1.41e-05 ***
## X.Weight.kg.          -6.201e-04  1.717e-02  -0.036    0.971
## Height.cm.             9.817e-03  2.082e-02   0.471    0.637
## PB                    -1.402e+01  1.422e+00  -9.860  < 2e-16 ***
## Log(theta)             7.026e-01  1.081e-01   6.497 8.17e-11 ***
## Zero hurdle model coefficients (binomial with logit link):
```

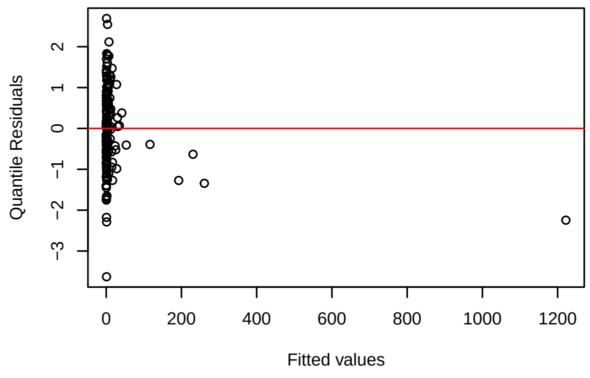
```
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            5.114e+02  9.061e+01   5.644 1.66e-08 ***
## no..Date.of.birth     -1.794e-05        NaN     NaN      NaN
## Total.number.of.races  1.335e-02  4.768e-03   2.799  0.00512 **
## X.Weight.kg.          -1.863e-02  4.549e-02  -0.410  0.68213
## Height.cm.             6.079e-02  5.075e-02   1.198  0.23096
## PB                    -5.243e+01  9.208e+00  -5.694 1.24e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta: count = 2.019
## Number of iterations in BFGS optimization: 32
## Log-likelihood: -322.4 on 13 Df
```

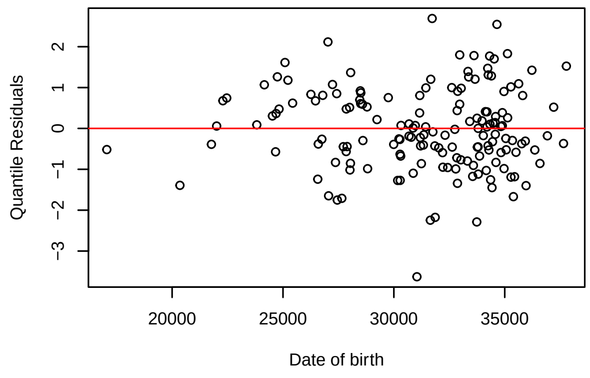
```
rootogram(m3)
```

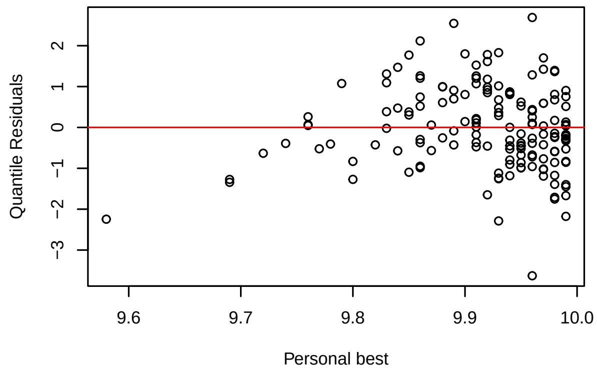


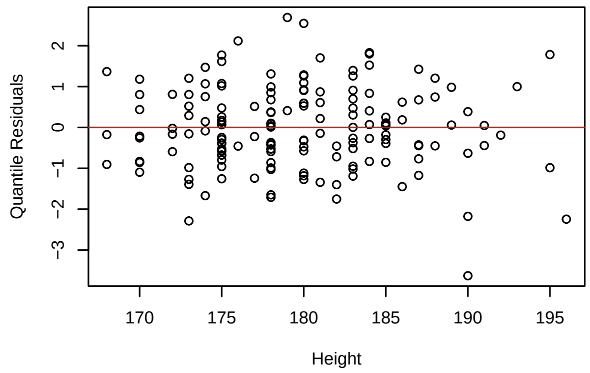
```
# create a new plotting window and set the plotting area into a 2*2 array
par(mfrow=c(3,2))

# Quantile residual plot
qresiduals <- qresid(m1)
plot(m1$fitted.values, qresiduals, xlab="Fitted values", ylab="Quantile Residuals")
abline(h = 0, col = "red")
plot(athletes$no..Date.of.birth, qresiduals, xlab="Date of birth", ylab="Quantile Residuals")
abline(h = 0, col = "red")
```

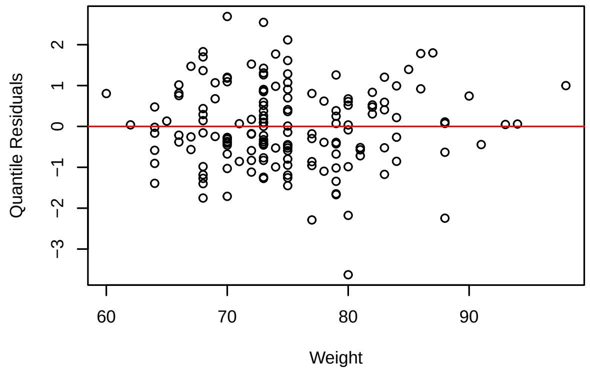
```
plot(athletes$PB, qresiduals, xlab="Personal best", ylab="Quantile Residuals")
abline(h = 0, col = "red")
plot(athletes$Height.cm., qresiduals, xlab="Height", ylab="Quantile Residuals")
abline(h = 0, col = "red")
plot(athletes$X.Weight.kg., qresiduals, xlab="Weight", ylab="Quantile Residuals")
abline(h = 0, col = "red")
plot(athletes$Total.number.of.races, qresiduals, xlab="Total number of races", ylab="Quantile Residuals
abline(h = 0, col = "red")
```

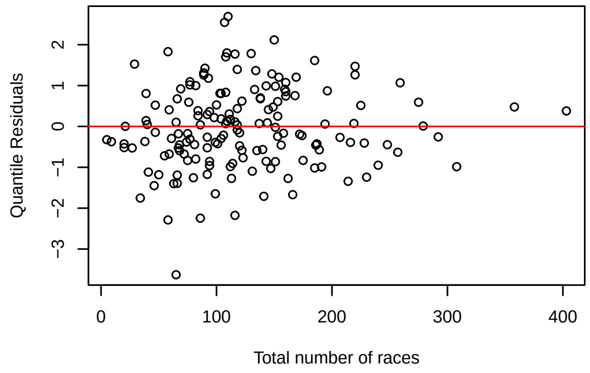












```
max_position = which.max(qresiduals);max_position
```

```
## [1] 110
```

```
max_value = max(qresiduals);max_value
```

```
## [1] 2.690111
```

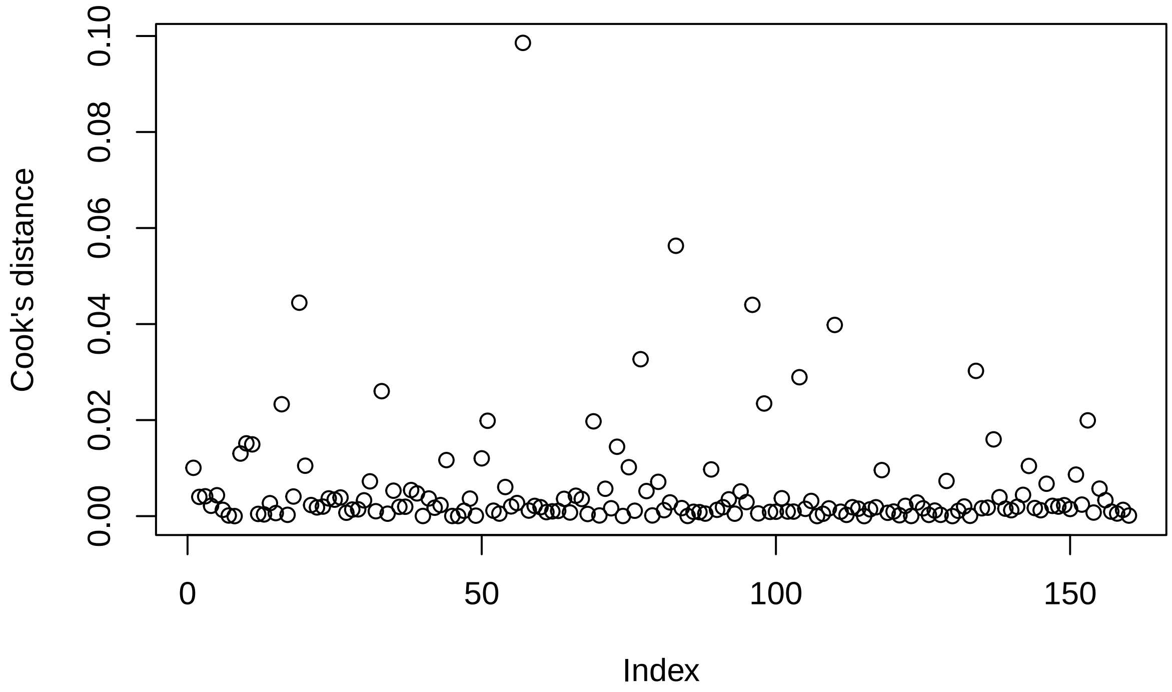
```
athletes[110,]
```

```
##     Date.of.birth              Name   PB    ID Height.cm. X.Weight.kg.
## 111    1986-11-11 Ben Youssef Meïté 9.96 67421        179           70
##     Races.under.10.seconds Total.number.of.races no..Date.of.birth NA.
## 111                      7                   110             31727 111
```

```
# Assuming m1 is your fitted model
predicted_value <- predict(m1, newdata = athletes[19, ], type = "response")
print(predicted_value)
```

```
##       20
## 7.428351
```

```
# Cook's distance
cooks_d <- cooks.distance(m1)
plot(cooks_d, ylab="Cook's distance", xlab="Index")
```



```
#abline(h=c(0.5, 1), col="red")
```

```
# Variance inflation factor
vif(m1)
```

```
##     no..Date.of.birth Total.number.of.races          X.Weight.kg.
##              1.276440              1.223739              1.981531
##            Height.cm.                    PB
##              1.802255              1.112980
```