QUIZ2 ADEI CURS 20-21Q1

Júlia Gasull

December, 18th 2020

Table of Contents

Network KPI

```
if(!is.null(dev.list())) dev.off(); rm(list = ls())
```

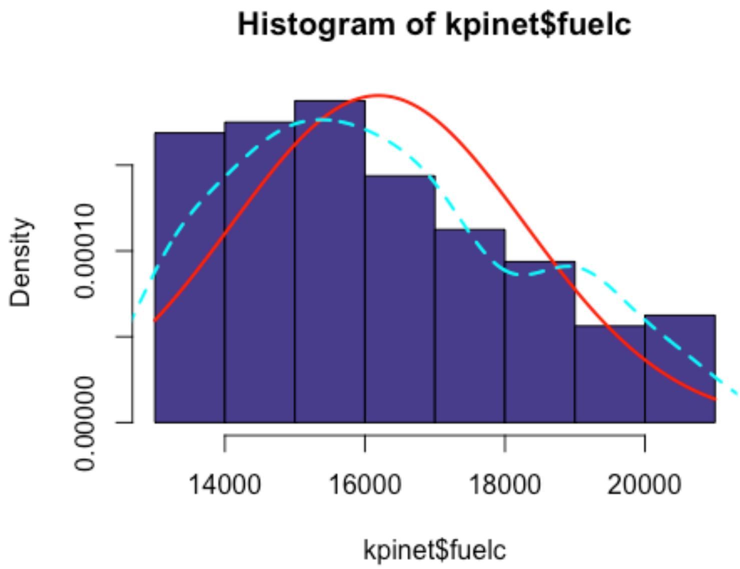
Quiz 2 - Questions

Question 1

1. Variable fuelc contains the total fuel consumption in liters of the selected configuration. Summarize numerically and graphically the response variable. Make an interpretation of the results. Do you think that fuelc may be considered normally distributed?

- És evident que les dades no es distribueixen normalment, ja que veiem que, primerament, no són simètriques.
- El 50% de les mostres es troben entre els 14589 i els 17475 litres, ambdós quantils Q1 i Q3.
- La mitjana i la mediana no s'acosten, per tant declarem que hi ha asimeteia.
- No hi ha valors atípics, ja que els valors mínims i màxims no són valors extrems.
- La prova de Shapiro rebutja la nul·la hipòtesi de normalitat.

```
summary(kpinet$fuelc)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13016   14589   15837   16193   17475   20926

hist(kpinet$fuelc,freq=F,10,col="darkslateblue")
mm<-mean(kpinet$fuelc);dd<-sd(kpinet$fuelc);mm;dd

## [1] 16193.08

## [1] 2093.585

curve(dnorm(x,mm,dd),add=T,col="red",lwd=2)
lines(density(kpinet$fuelc),add=T,col="cyan",lwd=2,lty=2)
```




```
shapiro.test(kpinet$fuelc)

##
##  Shapiro-Wilk normality test
##
## data:  kpinet$fuelc
## W = 0.95701, p-value = 7.594e-05
```

Question 2

2. Which are the variables statistically associated with the target (fuelc)? Indicate the suitable measure of association and/or tests that support your answer.

Hem de respondre la pregunta mitjançant el mètode condes().

```
vars <- names(kpinet)[c(1:6,8:17,25)]; vars

##  [1] "TD"           "GP"           "DP"           "PVD"          "TW"
##  [6] "NS"           "density"      "mflow"        "ttt.h"        "mtt.s.km"
## [11] "mdelay.s.km" "mspeed"       "ttdis"        "fuelc"        "co2"
## [16] "nox"          "thrputrate"

res.con <- condes(kpinet[,vars],which("fuelc"==colnames(kpinet[,vars])))
```

Les variables numèriques globals i relacionades positivament amb fuelc són co2, nox, ttt.h, ttdis, mflow, densitat, mtt.s.km i mdelay.s.km. mspeed està relacionada inversament.

```
res.con$quanti

##              correlation      p.value
## co2            0.9574843 3.312107e-87
## nox            0.9378166 1.688136e-74
## ttt.h          0.8980833 3.063657e-58
## ttdis          0.8695197 2.860851e-50
## mflow          0.7823336 2.633173e-34
## density        0.5482947 6.148990e-14
## mtt.s.km       0.1892277 1.655307e-02
## mdelay.s.km    0.1841043 1.978221e-02
## mspeed        -0.6013352 4.171215e-17
```

Les variables factorials que s'associen globalment a combustibles de significació alta a baixa són: penetració de la guia (GP), patró de demanda (DP), configuració del controlador (TD) i factor d'interval de finestra de temps (TW).

```
res.con$quali

##            R2      p.value
## GP 0.66951548 1.178435e-32
## DP 0.55134607 5.279693e-27
## TD 0.34626062 2.382023e-14
## TW 0.07297049 2.611225e-03
```

Les categories que mostren un efecte additiu positiu sobre la gran mitjana de combustible són:

- GP de 80, 90 i 100
- TW de 1.5 min
- DP de 30%
- TD de 60-20-20.

En el costat negatiu:

- GP de 0, 10 i 50
- TW de 3 min
- TD de 40-50-10
- DP de 0%

```
res.con$category

##                  Estimate      p.value
## DP=DP-30        1996.2021 1.959939e-14
## GP=GP-100       2493.6503 1.058729e-07
```

```
## GP=GP-90         2332.4238 6.959762e-07
## TD=TD-60-20-20    884.6062 7.497169e-04
## GP=GP-80         1238.3998 1.501337e-03
## TW=TW-1.5         696.4612 1.832814e-03
## TD=TD-40-40-20    486.5391 2.080387e-03
## TD=TD-20-70-10    576.5114 2.888558e-02
## DP=DP-10         -436.8683 1.305695e-02
## GP=GP-30        -1543.7091 5.868600e-03
## TW=TW-3          -644.9636 2.595581e-03
## GP=GP-0         -1822.8303 6.610523e-06
## GP=GP-50        -2521.5017 1.172978e-08
## TD=TD-40-50-10 -1947.6567 4.730290e-16
## DP=DP-0         -2342.3820 4.023510e-17
```

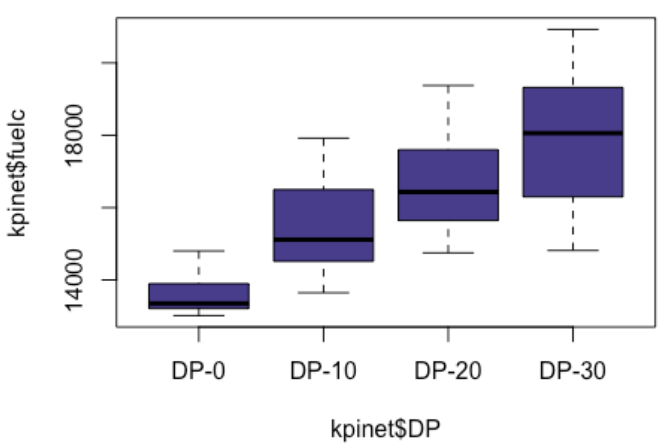
Question 3

3. The average fuelc can be argued to be the same for all Demand Pattern levels (DP)? Which are the groups that show a significant greater average fuelc than the others?.

Segons la prova no paramètrica de Kruskal sobre els mitjans de combustible segons els nivells de demanda (DP), es pot rebutjar clarament la hipòtesi nul·la de la mitjana uniforme.

Es mostra en un boxplot per fuelc segons els nivells de demanda (DP) i, a mesura que augmenta la demanda (number of trips), també augmenta el consum total de fuel.

```
Boxplot(kpinet$fuelc~kpinet$DP,col="darkslateblue")
```



```
kruskal.test(kpinet$fuelc~kpinet$DP)

##
##  Kruskal-Wallis rank sum test
##
## data:  kpinet$fuelc by kpinet$DP
## Kruskal-Wallis chi-squared = 94.008, df = 3, p-value < 2.2e-16
```

Question 4

4. Let us perform a one-way analysis of variance on TRB2018 dataset for target fuelc on Demand Pattern (DP) factor using the standard general linear model (lm() method). Assess the explicability of the model and interpret the model for prediction purposes writing the equation for each GP level.

El model que conté l'efecte brut del Demand Pattern (quantity of trips) explica el 55.13% de la variabilitat total del consum de combustible.

El consum de fuel (litres) a la demanda base es preveu que serà de 13579,5 litres, ja que la demanda augmenta en viatges de llarga distància un 10, 20 i 30%, el consum mitjà es preveu com (13579.5 + 1905.5 = 15485.0) litres, (13579.5+ 3125.4 = 16704.9) litres i (13579-5 + 4338-6 = 17918.1) litres, respectivament.

```
model_1 <- lm(
  fuelc~
    DP
  ,data=kpinet
); summary(model_1)

##
## Call:
## lm(formula = fuelc ~ DP, data = kpinet)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3103.0  -963.6  -229.6  1019.3  3007.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13579.5      258.5  52.537  < 2e-16 ***
## DPDP-10       1905.5      341.9   5.573 1.08e-07 ***
## DPDP-20       3125.4      341.9   9.140 3.16e-16 ***
## DPDP-30       4338.6      327.0  13.270  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1416 on 156 degrees of freedom
```

```
## Multiple R-squared:  0.5513, Adjusted R-squared:  0.5427
## F-statistic:  63.9 on 3 and 156 DF,  p-value: < 2.2e-16

tapply(kpinet$fuelc,kpinet$DP,mean)

##     DP-0    DP-10    DP-20    DP-30
## 13579.54 15485.05 16704.97 17918.12
```

Question 5

5. Use an inferential method to quantify the probability of the null hypothesis of a negligible gross effect for demand pattern (DP) factor.

Clarament, una prova de Fisher és adequada per provar per calcular el valor de la hipòtesi. El valor P és molt petit (l'estadística de Fisher és de 63.9, molt gran) rebutjant així la hipòtesi nul·la: el patró de demanda (quantity of trips) afecta significativament el consum total de combustible de la network.

```
model_2 <- lm(
  fuelc~
    1
  ,data=kpinet
)
model_3 <- lm(
  fuelc~
    DP
  ,data=kpinet
)

anova(model_2,model_3)

## Analysis of Variance Table
##
## Model 1: fuelc ~ 1
## Model 2: fuelc ~ DP
##   Res.Df       RSS Df Sum of Sq      F    Pr(>F)
## 1    159 696912403
## 2    156 312672489  3 384239914 63.902 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 6

6. Let us perform a two-way analysis of variance on TRB2018 dataset for target fuelc on Guidance Penetration (GP) and Demand Pattern (DP) factors using the standard general linear model (lm()

method). Are interactions needed for GP and DP to explain the total fuel consumption in the network? Are net effects for both factors significant?

Segons la producció d'Anova(model_2), els efectes nets d'ambdós factors són significatius.

De totes maneres, donat que els models additius GP+DP i GP*DP complets estan nested, es pot utilitzar una prova de Fisher per comprovar la necessitat d'interaccions: ja que que el valor de p és inferior a 0,05 llindar normal, ambdós models no són equivalents, per tant les interaccions (el poques que es puguin estimar) contribueixen a explicar el consum total de fuel.

```
model_4 <- lm(
  fuelc~
    GP+DP
  ,data=kpinet
)
model_5 <- lm(
  fuelc~
    GP*DP
  ,data=kpinet
)

anova(model_4,model_5)

## Analysis of Variance Table
##
## Model 1: fuelc ~ GP + DP
## Model 2: fuelc ~ GP * DP
##   Res.Df      RSS Df Sum of Sq     F    Pr(>F)
## 1    148 37025155
## 2    136 18597690 12  18427465 11.23 2.288e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova(model_4)

## Anova Table (Type II tests)
##
## Response: fuelc
##              Sum Sq  Df F value    Pr(>F)
## GP        275647334   8  137.73 < 2.2e-16 ***
## DP        193293608   3  257.55 < 2.2e-16 ***
## Residuals  37025155 148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 7

7. Assess the explicability of the BEST model selected in question 6 and calculate the point and 95% confidence prediction interval for fuel consumption for an scenario with a guidance penetration of 80% and demand level of 30%. Try to explain the reason of warning messages appearing for some methods depending on the selected mode.

El disseny experimental no permet estimar les interaccions entre els factors GP i DP, de manera que no es poden determinar moltes de les variables fictícies. Per als nivells particulars de 80% GP i 20% DP, la predicció puntual és: 13145.04 + 1490.20 + 2529.35 + 1964.15 = 19128.74 litres.

L'interval de confiança del 95% per al consum total és de [18361.75 19895.72], en litres.

```
summary(model_5)

##
## Call:
## lm(formula = fuelc ~ GP * DP, data = kpinet)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1489.79  -177.73    13.64   174.16  1109.10
##
## Coefficients: (12 not defined because of singularities)
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13145.04     116.94 112.409  < 2e-16 ***
## GPGP-10            478.00     202.54   2.360 0.019698 *
## GPGP-20           1189.67     165.38   7.194 3.83e-11 ***
## GPGP-30            987.19     233.88   4.221 4.42e-05 ***
## GPGP-50            372.26     150.97   2.466 0.014914 *
## GPGP-70           3172.85     202.54  15.665  < 2e-16 ***
## GPGP-80           1490.20     202.54   7.357 1.60e-11 ***
## GPGP-90           4633.45     202.54  22.876  < 2e-16 ***
## GPGP-100          4864.97     202.54  24.019  < 2e-16 ***
## DPDP-10           -346.56     286.44  -1.210 0.228428
## DPDP-20           1748.55     202.54   8.633 1.41e-14 ***
## DPDP-30           2529.35     165.38  15.294  < 2e-16 ***
## GPGP-10:DPDP-10   1286.17     330.75   3.889 0.000157 ***
## GPGP-20:DPDP-10        NA         NA      NA       NA
## GPGP-30:DPDP-10    452.35     369.79   1.223 0.223352
## GPGP-50:DPDP-10   1508.62     344.26   4.382 2.33e-05 ***
## GPGP-70:DPDP-10    -63.27     309.39  -0.205 0.838267
## GPGP-80:DPDP-10        NA         NA      NA       NA
## GPGP-90:DPDP-10   -749.18     330.75  -2.265 0.025089 *
## GPGP-100:DPDP-10       NA         NA      NA       NA
## GPGP-10:DPDP-20    293.63     309.39   0.949 0.344269
## GPGP-20:DPDP-20   -452.32     286.44  -1.579 0.116639
## GPGP-30:DPDP-20        NA         NA      NA       NA
```

```
## GPGP-50:DPDP-20         NA         NA      NA       NA
## GPGP-70:DPDP-20  -1136.12     309.39  -3.672 0.000345 ***
## GPGP-80:DPDP-20    994.39     309.39   3.214 0.001635 **
## GPGP-90:DPDP-20   -532.41     309.39  -1.721 0.087557 .
## GPGP-100:DPDP-20 -1492.47     309.39  -4.824 3.72e-06 ***
## GPGP-10:DPDP-30         NA         NA      NA       NA
## GPGP-20:DPDP-30         NA         NA      NA       NA
## GPGP-30:DPDP-30         NA         NA      NA       NA
## GPGP-50:DPDP-30         NA         NA      NA       NA
## GPGP-70:DPDP-30         NA         NA      NA       NA
## GPGP-80:DPDP-30   1964.15     261.48   7.512 6.99e-12 ***
## GPGP-90:DPDP-30         NA         NA      NA       NA
## GPGP-100:DPDP-30        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 369.8 on 136 degrees of freedom
## Multiple R-squared:  0.9733, Adjusted R-squared:  0.9688
## F-statistic: 215.7 on 23 and 136 DF,  p-value: < 2.2e-16
```

```
predict(model_5,newdata=data.frame(GP="GP-80",DP="DP-30"))
```

```
##        1
## 19128.73
```

```
predict(model_4,newdata=data.frame(GP="GP-80",DP="DP-30"))
```

```
##        1
## 18772.07
```

```
predict(model_5,newdata=data.frame(GP="GP-80",DP="DP-30"),interval="prediction")
```

```
##        fit      lwr      upr
## 1 19128.73 18361.75 19895.72
```

```
predict(model_4,newdata=data.frame(GP="GP-80",DP="DP-30"),interval="prediction")
```

```
##        fit      lwr      upr
## 1 18772.07 17754.11 19790.02
```

Question 8

8. Consider a general linear modeling for total fuel consumption when all numeric variables included in the list of KPIs are taken into account. Assess explicability, try to reduce the number of explanatory variables without loosing too much explicability and discuss colinearity among explanatory variables.

Les variables explicatives es correlacionen entre elles.

L'explicabilitat segons el coeficient de determinació és del 99,93%, de manera que explica el 99,93% de la variabilitat total en combustible, però el model no és adequat com es pot veure amb vif().

El mètode step(model_6) no ens ajuda a definir un model adequat, ja que no s'elimina cap variable.

S'ha de proposar un model que contingui 2 o 3 variables no correlacionades.

```
model_6 <- lm(
  fuelc~
    co2+
    nox+
    ttt.h+
    ttdis+
    mflow+
    density+
    mtt.s.km+
    mdelay.s.km+
    mspeed
  ,data=kpinet
); summary(model_6)

##
## Call:
## lm(formula = fuelc ~ co2 + nox + ttt.h + ttdis + mflow + density +
##     mtt.s.km + mdelay.s.km + mspeed, data = kpinet)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -161.269  -37.072    0.229   29.150  132.714
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.093e+04  1.919e+03   5.694 6.34e-08 ***
## co2          1.994e-04  3.838e-05   5.195 6.56e-07 ***
## nox          1.994e-02  1.600e-02   1.246  0.21454
## ttt.h        1.614e+00  7.433e-02  21.721  < 2e-16 ***
## ttdis        1.117e-01  9.255e-03  12.068  < 2e-16 ***
## mflow       -1.562e-01  1.349e-02 -11.583  < 2e-16 ***
## density     -3.061e+02  3.203e+01  -9.555  < 2e-16 ***
## mtt.s.km    -1.035e+02  1.979e+01  -5.231 5.58e-07 ***
## mdelay.s.km  9.930e+01  1.961e+01   5.065 1.18e-06 ***
## mspeed      -9.357e+01  3.532e+01  -2.649  0.00893 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.56 on 150 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9993
## F-statistic: 2.419e+04 on 9 and 150 DF,  p-value: < 2.2e-16

vif(model_6)
```

```
##         co2         nox       ttt.h       ttdis       mflow     density
##  1362.66468  1419.05991   100.13455   216.58627    36.40855   154.51691
##    mtt.s.km mdelay.s.km      mspeed
## 12746.72939 12705.61565    96.00476
```

```
step(model_6)
```

```
## Start:  AIC=1300.94
## fuelc ~ co2 + nox + ttt.h + ttdis + mflow + density + mtt.s.km +
##     mdelay.s.km + mspeed
##
##               Df Sum of Sq     RSS    AIC
## - nox          1      4969  484743 1300.6
## <none>                      479774 1300.9
## - mspeed       1     22447  502221 1306.3
## - mdelay.s.km  1     82048  561822 1324.2
## - co2          1     86335  566108 1325.4
## - mtt.s.km     1     87521  567295 1325.8
## - density      1    291986  771759 1375.0
## - mflow        1    429113  908887 1401.2
## - ttdis        1    465835  945608 1407.5
## - ttt.h        1   1509042 1988816 1526.5
##
## Step:  AIC=1300.59
## fuelc ~ co2 + ttt.h + ttdis + mflow + density + mtt.s.km + mdelay.s.km +
##     mspeed
##
##               Df Sum of Sq     RSS    AIC
## <none>                      484743 1300.6
## - mspeed       1     20803  505546 1305.3
## - mdelay.s.km  1    103909  588652 1329.7
## - mtt.s.km     1    110042  594785 1331.3
## - density      1    340603  825346 1383.7
## - mflow        1    491665  976408 1410.6
## - co2          1    509966  994709 1413.6
## - ttdis        1    519488 1004231 1415.1
## - ttt.h        1   1510949 1995692 1525.0
```

```
##
## Call:
## lm(formula = fuelc ~ co2 + ttt.h + ttdis + mflow + density +
##     mtt.s.km + mdelay.s.km + mspeed, data = kpinet)
##
## Coefficients:
## (Intercept)          co2        ttt.h        ttdis        mflow      density
##   1.145e+04    2.409e-04    1.615e+00    1.145e-01   -1.608e-01   -2.861e+02
##    mtt.s.km  mdelay.s.km       mspeed
##  -1.109e+02    1.066e+02   -8.973e+01
```

```
names(kpinet)
```

```
##  [1] "TD"          "GP"          "DP"          "PVD"         "TW"
##  [6] "NS"          "replica"     "density"     "mflow"       "ttt.h"
## [11] "mtt.s.km"    "mdelay.s.km" "mspeed"      "ttdis"       "fuelc"
## [16] "co2"         "nox"         "tvehin"      "tvehout"     "virwait"
## [21] "virlostin"   "virlostout"  "inputveh"    "demand"      "thrputrate"
## [26] "waitrate"
```

```
round(cor(kpinet[,c(15,16,17,10,14,9,8,11,12,13)]),dig=2)
```

```
##             fuelc   co2   nox ttt.h ttdis mflow density mtt.s.km mdelay.s.km
## fuelc        1.00  0.96  0.94  0.90  0.87  0.78    0.55     0.19        0.18
## co2          0.96  1.00  1.00  0.96  0.71  0.65    0.76     0.42        0.42
## nox          0.94  1.00  1.00  0.96  0.67  0.62    0.80     0.47        0.47
## ttt.h        0.90  0.96  0.96  1.00  0.59  0.62    0.82     0.59        0.59
## ttdis        0.87  0.71  0.67  0.59  1.00  0.91    0.10    -0.29       -0.29
## mflow        0.78  0.65  0.62  0.62  0.91  1.00    0.16    -0.13       -0.13
## density      0.55  0.76  0.80  0.82  0.10  0.16    1.00     0.89        0.88
## mtt.s.km     0.19  0.42  0.47  0.59 -0.29 -0.13    0.89     1.00        1.00
## mdelay.s.km  0.18  0.42  0.47  0.59 -0.29 -0.13    0.88     1.00        1.00
## mspeed      -0.60 -0.78 -0.81 -0.86 -0.14 -0.18   -0.96    -0.88       -0.88
##             mspeed
## fuelc        -0.60
## co2          -0.78
## nox          -0.81
## ttt.h        -0.86
## ttdis        -0.14
## mflow        -0.18
## density      -0.96
## mtt.s.km     -0.88
## mdelay.s.km  -0.88
## mspeed        1.00
```

Question 9

9. Consider linear models including as explanatory numeric variables total travel distance (ttdis) and mean speed (mspeed) and the main effect for guidance penetration factor. Discuss the model, while justifying the choice (nor residuals or influent data analysis has to be considered).

El model especificat explica el 99,5% de la variabilitat del consum de fuel, per tant és aparentment un model molt bo. Tots els efectes nets són significatius i el model no es pot reduir mitjançant el mètode step(model_7). Els vifs són inferiors a 3, de manera que no hi ha problemes de colinearitat.

```
model_7 <- lm(
  fuelc~
    ttdis+
    mspeed+
    GP
```

```
  ,data=kpinet
); summary(model_7)

##
## Call:
## lm(formula = fuelc ~ ttdis + mspeed + GP, data = kpinet)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -383.2 -104.3   -7.4  103.6  483.8
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.697e+04  4.458e+02  38.070  < 2e-16 ***
## ttdis        2.128e-01  3.431e-03  62.015  < 2e-16 ***
## mspeed      -8.553e+02  1.574e+01 -54.338  < 2e-16 ***
## GPGP-10      7.685e+01  5.390e+01   1.426    0.156
## GPGP-20      5.783e+01  5.966e+01   0.969    0.334
## GPGP-30      7.589e+01  7.122e+01   1.066    0.288
## GPGP-50      3.949e+02  6.776e+01   5.827 3.35e-08 ***
## GPGP-70      4.833e+02  7.349e+01   6.577 7.64e-10 ***
## GPGP-80      3.608e+02  7.697e+01   4.688 6.17e-06 ***
## GPGP-90      5.993e+02  7.975e+01   7.514 4.92e-12 ***
## GPGP-100     5.918e+02  7.415e+01   7.981 3.57e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 153.3 on 149 degrees of freedom
## Multiple R-squared:  0.995,  Adjusted R-squared:  0.9946
## F-statistic:  2952 on 10 and 149 DF,  p-value: < 2.2e-16

vif(model_7)

##            GVIF Df GVIF^(1/(2*Df))
## ttdis  4.052151  1        2.012996
## mspeed 2.596127  1        1.611250
## GP     9.395873  8        1.150293

step(model_7)

## Start:  AIC=1620.89
## fuelc ~ ttdis + mspeed + GP
##
##          Df Sum of Sq      RSS    AIC
## <none>                 3499920 1620.9
## - GP      8   4190700  7690620 1730.8
## - mspeed  1  69354224 72854144 2104.6
## - ttdis   1  90337813 93837733 2145.1

##
## Call:
```

```
## lm(formula = fuelc ~ ttdis + mspeed + GP, data = kpinet)
##
## Coefficients:
## (Intercept)        ttdis       mspeed      GPGP-10      GPGP-20      GPGP-30
##  16970.0053       0.2128    -855.2664      76.8523      57.8301      75.8852
##     GPGP-50      GPGP-70      GPGP-80      GPGP-90     GPGP-100
##    394.8798     483.2836     360.8376     599.2620     591.8391

Anova(model_7)

## Anova Table (Type II tests)
##
## Response: fuelc
##             Sum Sq  Df  F value    Pr(>F)
## ttdis     90337813   1 3845.897 < 2.2e-16 ***
## mspeed    69354224   1 2952.575 < 2.2e-16 ***
## GP         4190700   8   22.301 < 2.2e-16 ***
## Residuals  3499920 149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 10

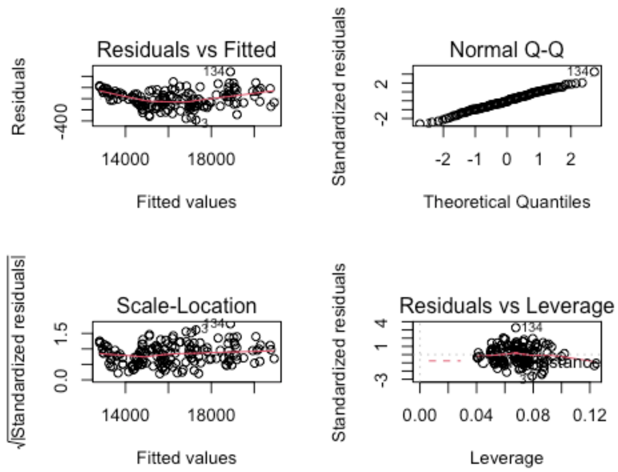
10. Consider the model for fuel consumption proposed in the previous question. Does the model fullfil the properties for linear models? Justify your answer using statistical arguments. Indicate the presence of lack of fit observations and influent data.

Sembla que es compleix la normalitat i homocedasticitat, però sembla que hi ha un patró sistemàtic en els residus (Residuals vs Fitted) que indica la necessitat d'una transformació no lineal o l'addició d'una nova variable (com es pot veure utilitzant MarginalModelPlots(model_7), els perfils per a configuracions de velocitat baixa indiquen dos grups d'observacions).

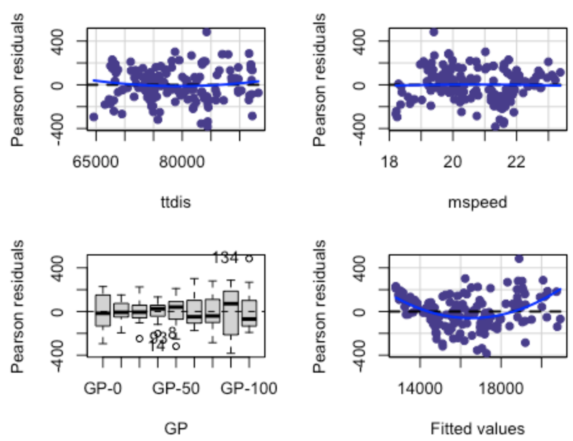
La falta d'ajust és present a l'observació 134 (student residual > 3). El tall 4/(n-p) per a la distància de Cooks és adequat a causa de la petita mida de la mostra: l'observació 134 és la més influent, seguida de 3, 1, 5 i 120.

Nota: he tret els warnings que em donava perquè sigui més llegible.

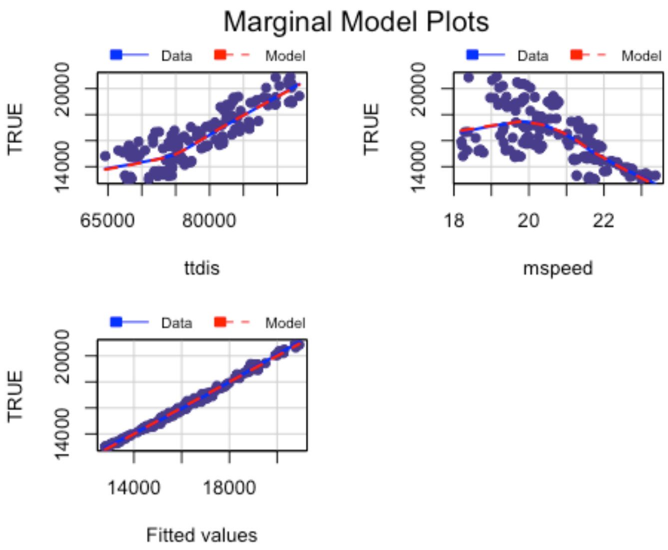
```
par(mfrow=c(2,2))
plot(model_7)
```



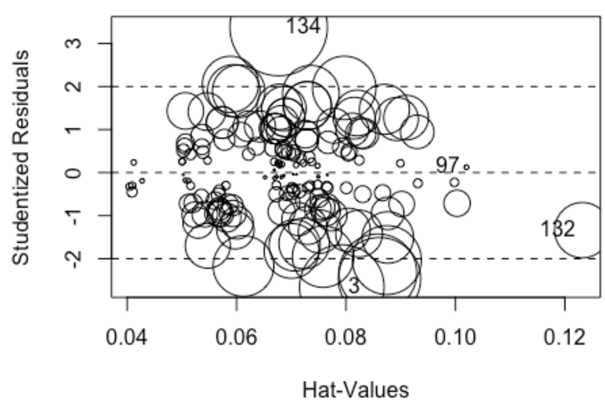

```
par(mfrow=c(1,1))

residualPlots(model_7,id.n=5,id.method=cooks.distance(model_7),col="darkslateblue",pch
=19)
```



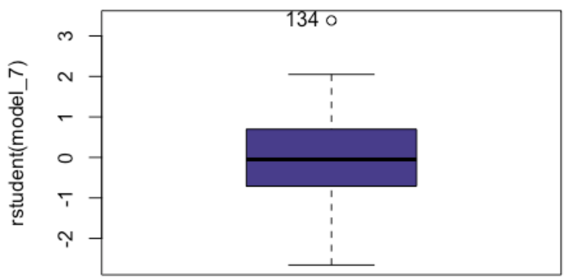

```
##              Test stat Pr(>|Test stat|)
## ttdis           1.6111           0.1093
## mspeed         -0.2785           0.7810
## GP
## Tukey test      8.3676           <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

marginalModelPlots(model_7,id.n=5,id.method=cooks.distance(model_7),col="darkslateblue
",pch=19)
```




```
influencePlot(model_7,id.n=5)
```




```
##         StudRes        Hat        CookD
## 3   -2.6579649 0.07922927 0.0531022681
## 97   0.1246918 0.10201438 0.0001616419
## 132 -1.3376359 0.12317385 0.0227297005
## 134  3.3814196 0.06768643 0.0705261775

Boxplot(rstudent(model_7),col="darkslateblue")  # Lack of fit: outlier in residuals
```




```
## [1] 134

kpinet[134,]

##                TD      GP     DP     PVD    TW     NS replica density mflow    ttt.h
## 134 TD-20-70-10 GP-100 DP-20 PVD-20 TW-6 NS-LI       R3  10.882 44819 5648.025
##      mtt.s.km mdelay.s.km  mspeed    ttdis    fuelc      co2      nox tvehin
## 134 236.9667    163.4125 19.4004 84238.56 19375.19 41347221 101333.7   6297
##      tvehout virwait virlostin virlostout inputveh demand thrputrate waitrate
## 134   44819     963         2        236    45670  46633      96.11     2.07

Boxplot(cooks.distance(model_7),col="darkslateblue")

##  [1] 134   3   1   5 120  90  28  14 135  59

abline(h=4/(nrow(kpinet)-length(coef(model_7))),col="red",lwd=2,lty=2)
```