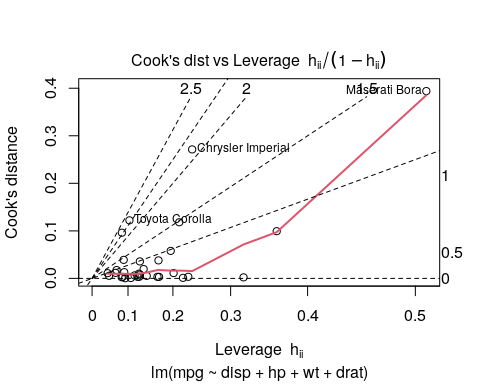
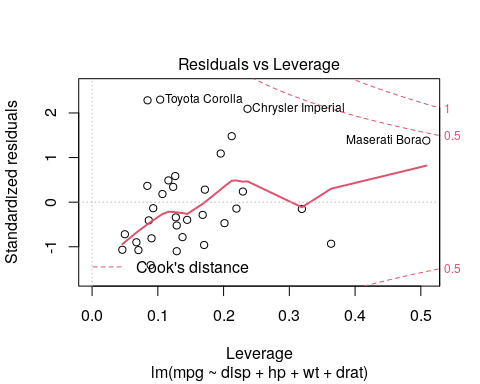
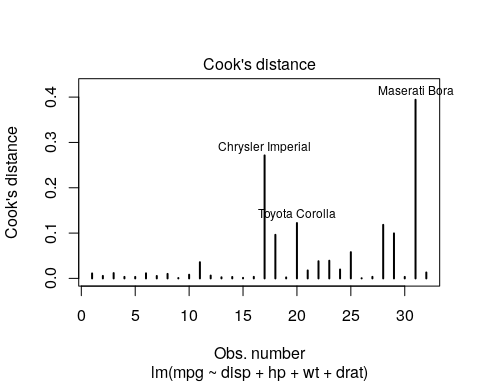
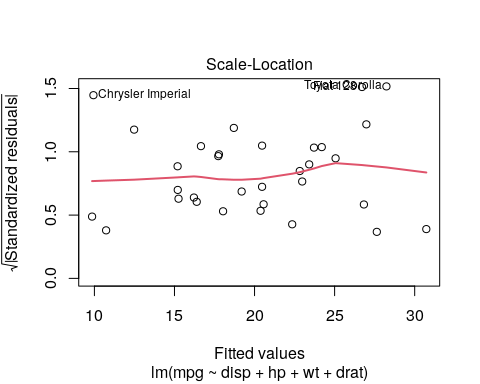
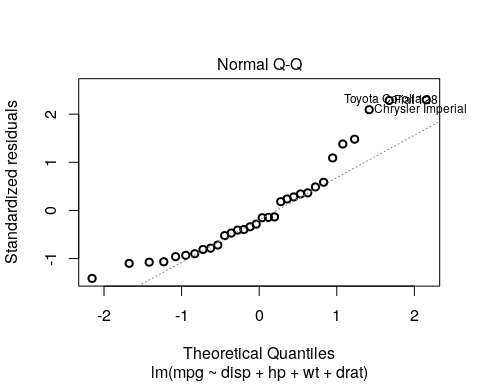
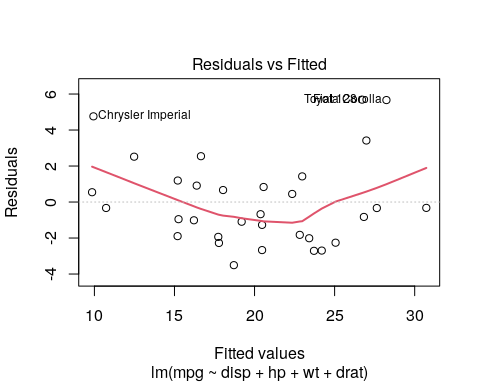
After fitting the model fit <- lm(mpg ∼ disp + hp + wt + drat, data=mtcars), the following outputs are obtained by the R commands summary(fit) and plot(fit), respectively. Describe how to interpret these results, and then suggest how to proceed with further analyses.

##   
## Call:  
## lm(formula = mpg ~ disp + hp + wt + drat, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5077 -1.9052 -0.5057 0.9821 5.6883   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 29.148738 6.293588 4.631 8.2e-05 \*\*\*  
## disp 0.003815 0.010805 0.353 0.72675   
## hp -0.034784 0.011597 -2.999 0.00576 \*\*   
## wt -3.479668 1.078371 -3.227 0.00327 \*\*   
## drat 1.768049 1.319779 1.340 0.19153   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.602 on 27 degrees of freedom  
## Multiple R-squared: 0.8376, Adjusted R-squared: 0.8136   
## F-statistic: 34.82 on 4 and 27 DF, p-value: 2.704e-10

#### Osservazioni

* Modello con 4 regressori continui e variabile risposta continua
* L’intercetta è elevata, ma sembra comunque essere coerente con il contesto
* Alcuni regressori (disp e drat) hanno p-value elevati (maggiori del 5%), quindi probabilmente non sono significativi
* I p-value di hp e wt sono sufficientemente bassi da rifiutare l’ipotesi nulla,
* hp e disp hanno coefficiente stimato piuttosto basso, ma ciò potrebbe essere causato dalla diversa scala delle variabili.
* I coefficienti di determinazione e corretto assumono un buon valore (poco oltre l’80%), ciò significa che il modello copre una parte significativa di variabilità del dataset.
* Il test F suggerisce una buona significatività di tutte le variabili, in relazione ai p-value dei regressori, probabilmente c’è qualche collinearità tra le variabili.

plot(fit, which = c(1:6), lwd=2)



#### Osservazioni

* La curva non è nè lineare, nè parallela alle ascisse, quindi l’ipotesi di linearità è violata (c’è qualche regressore che ha effetti non lineari sulla risposta),
* Sulle code il qq-plot presenta un deciso scostamento rispetto ai quantili teorici, l’ipotesi di normalità dei residui è violata
* Dal grafico dei residui standardizzati sembra essere rispettata l’ipotesi di omogeneità della varianza dei residui (la curva è abbastanza orizzontale)
* Dagli ultimi tre grafici risulta che nessun punto ha effetti leva o distanza di Cook particolarmente elevati.

library(DAAG)

## Loading required package: lattice

vif(fit)

## disp hp wt drat   
## 8.2094 2.8944 5.0966 2.2795

L’analisi VIF conferma la presenza di collinearità tra le variabili esplicative (per disp e wt il fattore VIF è maggiore di 5).

### Suggerimenti per ulteriori analisi

* utilizzare metodi stepforward e backforward (eliminando o aggiungendo regressori) per la scelta di un modello tale che elimini le collinearità (verificabili con la funzione vif) e riduca i valori dei p-value.
* trovato un modello con le caratteristiche di cui sopra, valutare l’utilizzo di qualche trasformata per ridurre o eliminare gli effetti non lineari.