**Estatística Aplicada II**

**Segunda Lista de Exercícios**

**SAMUEL KOJICOVSKI**

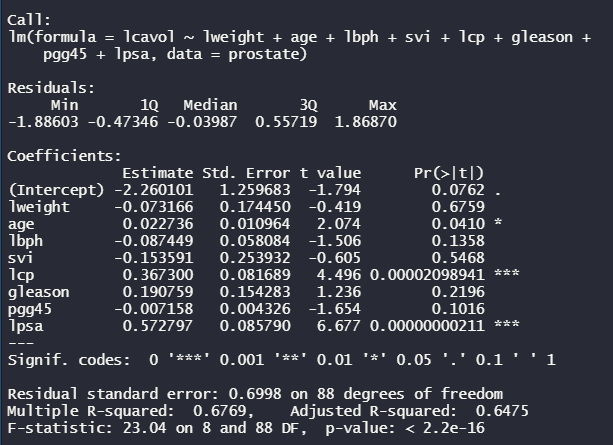
Utilizar a base de dados “prostate” que encontra-se na aba arquivos do TEAMS, explicações do que se trata esta base de dados pode ser encontrado em: <https://rafalab.github.io/pages/649/prostate.html>

Com a base de dados “prostate” obter os seguintes resultados com o auxílio do “R”

**a) Elaborar a regressão linear preliminar com as seguintes variáveis:**

**Variável dependente: lcavol;**

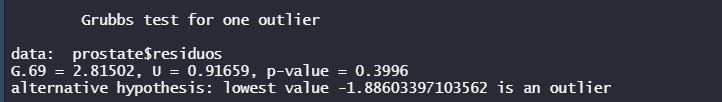
**Variáveis explicativas: lweight, age, lbph, svi, lcp, gleason, pgg45, lpsa**



**b) Testar outliers pelas técnicas de Grubbs, qui-quadrado, Bonferroni e da Distância de Cook e eliminar observações se necessário.**

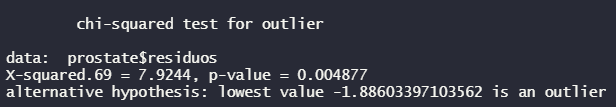
**Teste de Grubbs**

Como p-value é maior que 0.05, não há necessidade de remoção de outlier.

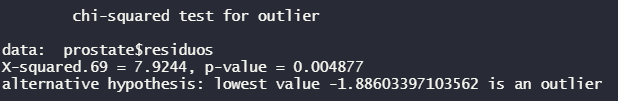


**Teste qui-quadrado**

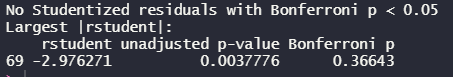
Como p-value é menor que 0.05, existe outlier e o seu valor é -1.88603397103562



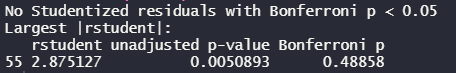
Nova rodada com qui-quadrado, com p-value agora menor que 0,05.



**Teste Bonferroni**

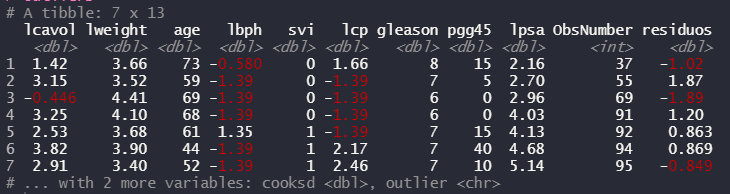


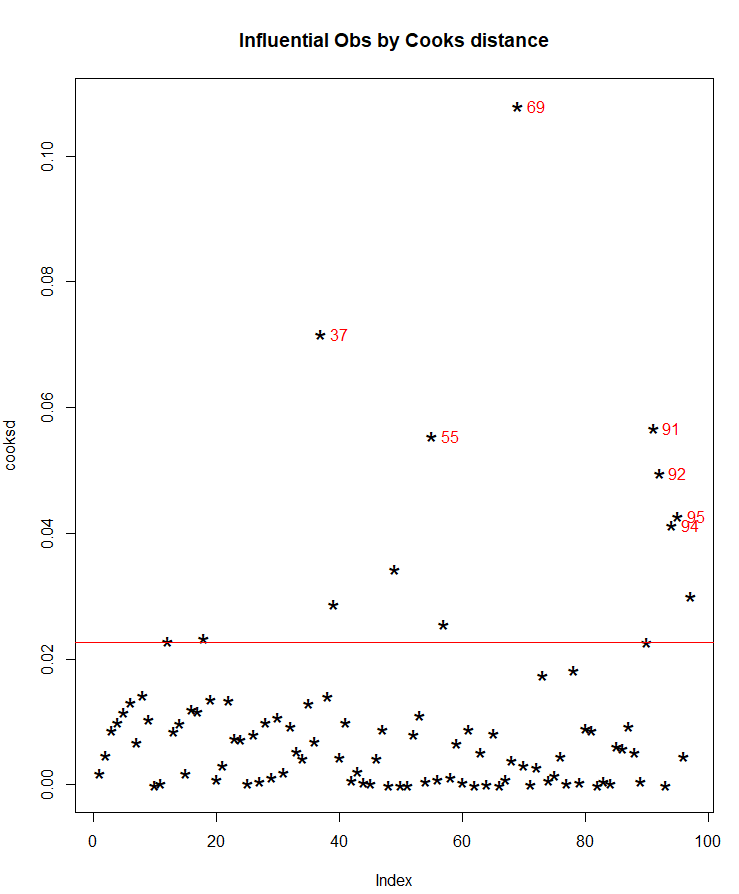
Nova rodada com teste Bonferroni, agora com p-value maior que 0,05, portanto, não necessário remoção de outlier.



**Distância de cook**

Outliers encontrados no teste

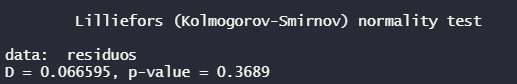




**c) Testar normalidade por Kolmogorov-Smirnov e Shapiro-Wilk.**

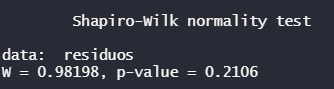
**Kolmogorov-Smirnov**

Para n=96 🡪 D= 1/raiz(96) = 0,102, como D calculado é menor que tabelado, aceita-se H0. Portanto, como p-value é maior que 0.05, a distribuição é normal, aceita-se a normalidade.



**Shapiro-Wilk**

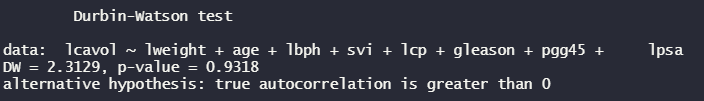
Valor de “W” calculado menor que “W” tabelado, portanto, aceita-se H0. Com um p-value maior que 0.05, distribuição normal.



**d) Testar autocorrelação e corrigir com HAC se for o caso.**

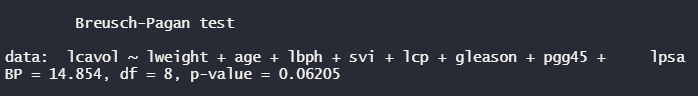
**Dwtest**

Não há autocorrelação, dado que p-value é maior que 0,05.



**e) Testar heterocedasticidade e corrigir se for o caso, com regressão robusta ou HC1.**

**Bptest**

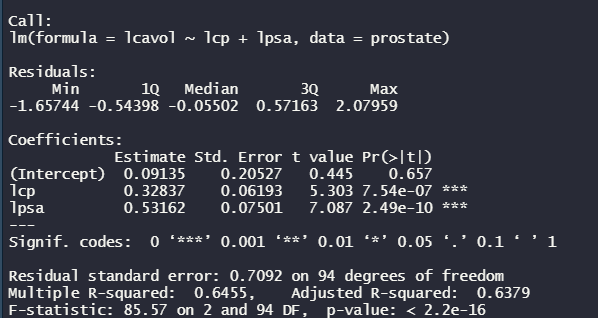




Como o resultado do teste BP (14.854) é menor que o tabelado (15.50731), aceita-se a hipótese de homoscedasticidade.

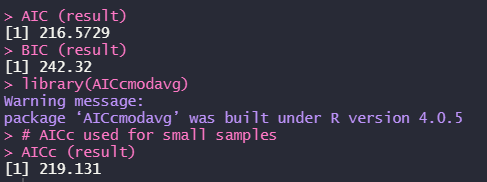
**f) Fazer regressão stepwise e obter a melhor regressão, apresentar o resultado do modelo.**

Melhor modelo stepwise 🡪 result\_stepwise



**g) Obter o AIC, BIC e AICc do melhor modelo selecionado.**

O modelo selecionado para os testes AIC, BIC e AICc é o modelo utilizando todas as variáveis, pois, o melhor modelo stepwise teve um rendimento menor que o modelo preliminar.



**h) Estimar os Intervalos de confiança dos parâmetros.**

Teste de performance e intervalos de confiança, comparando melhor modelo stepwise com modelo preliminar utilizando todas as variáveis.

