BPVS

Eva Reindl

5 5 2020

"R version 3.6.1 (2019-07-05)"

There are 156 children in the dataset.

# Dropouts

There are **7 dropouts**:

* for 5 children, no basal set could be established. For two of those children, it was suspected that the children answered wrongly on purpose.
* 1 child stopped the game
* for 1 child, the raw score was so low that the standardized score could not be calculated.

# Valid data

There are **149 children** with valid data.

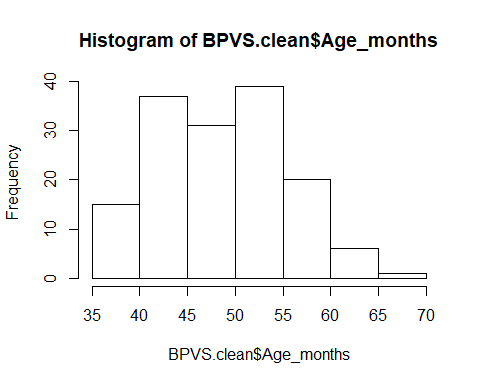
# Sample description

## Gender distribution

There are 79 girls and 70 boys.

## Age

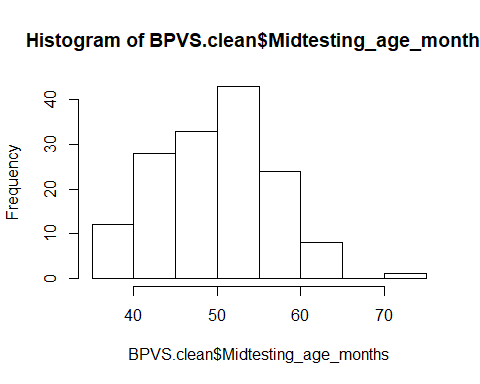
### Age at beginning of testing (which was basis for calculation of scores)



The children who contributed data to the BPVS were on average 49.15 months (SD = 6.82, range 35-70) at the beginning of testing. There were

* 1 2-year-old (scores caluclated as if child was 3)
* 62 3-year-olds
* 79 4-year-olds
* 7 5-year-olds

### Age in the middle of testing



The children who contributed data to the BPVS were on average **50.21 months (SD = 6.77, range 36-72)** in the middle of testing. There were

* 60 3-year-olds
* 77 4-year-olds
* 10 5-year-olds
* 1 6-year-old

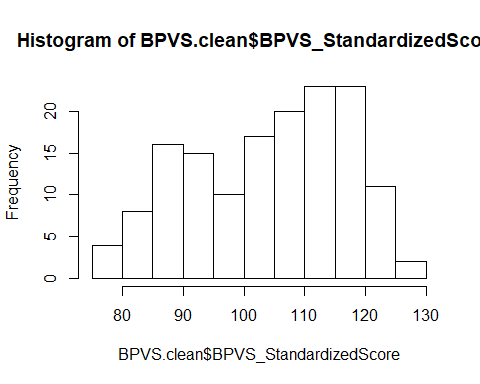
### Mediansplit young/old based on entire sample of test battery children

median is 49 months, “young” children are <= 49 months, “old” children are >49 months

There were **70 young** children and **79 old** children.

# BPVS score

## Across the entire sample



Across the entire sample, the average BPVS standardized score was **104.33 (SD = 12.61, range 77-126)**. The distribution of the variable was bimodal and is not normally distributed, W = 0.959, p < .001.

## By gender

* Girls (n = 79): M = 103.71 (SD = 11.98, range 82-126)
* Boys (n = 70): M = 105.03 (SD = 13.34, range 77-126)

## By age groups at midtesting

* 3y (n = 60): M = 100.03 (SD = 11.79, range 80-122)
* 4y (n = 77): M = 107.66 (SD = 12.21, range 77-126)
* 5y (n = 10): M = 107.10 (SD = 13.26, range 77-122)
* 6y (n = 1): 94

## By mediansplit age

* young (n = 70): M = 100.80 (SD = 12.46, range 80-123)
* old (n = 79): M = 107.45 (SD = 11.96, range 77-126)

## By testing location (Edinburgh vs Fife):

Edinburgh (n = 67): M = 107.24 (SD = 12.56, range 77-126)

Fife (n = 82): M = 101.95 (SD = 12.21, range 77-125)

### We then look at the age distribution of Edinburgh and Fife.

* Edinburgh: 25 3y, 41 4y. M = 49.83 months (SD = 5.83, range 36-58)
* Fife: 35 3y, 36 4y 10 5y, 1 6y. M = 50.52 months (SD = 7.48, range 36-72)

### Can the BPVS score be predicted by testing location, above age?

library(lme4)

## Loading required package: Matrix

BPVS.clean$z.age=as.vector(scale(BPVS.clean$Midtesting\_age\_months))#transform age to mean of 0 and SD of 1  
mean(BPVS.clean$z.age)#check whether it has worked

## [1] -4.707897e-16

sd(BPVS.clean$z.age)

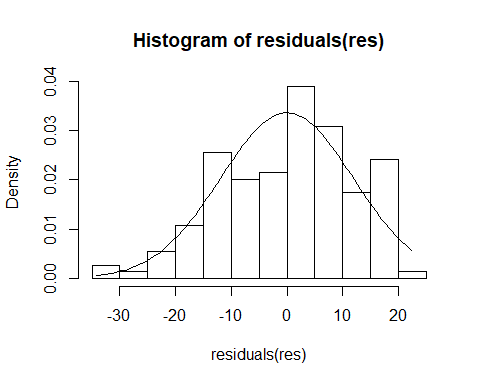
## [1] 1

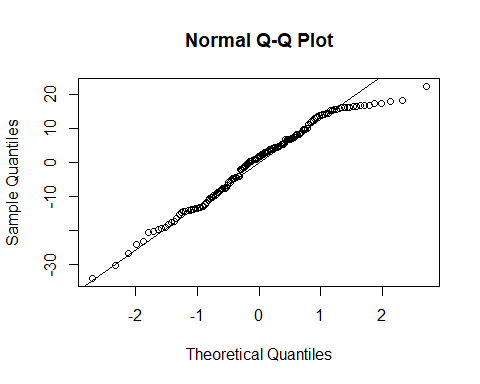
res<-lm(BPVS\_StandardizedScore ~ z.age + TestingLocation, data = BPVS.clean)

#### 

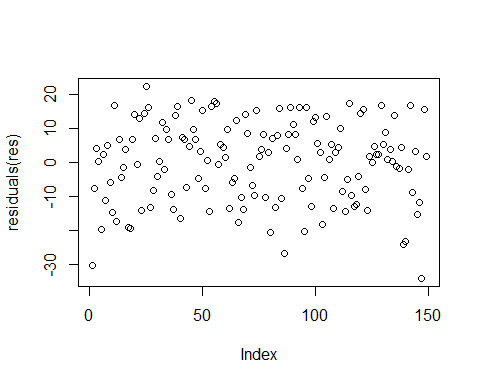
#### Check of assumptions

1. Normality of residuals



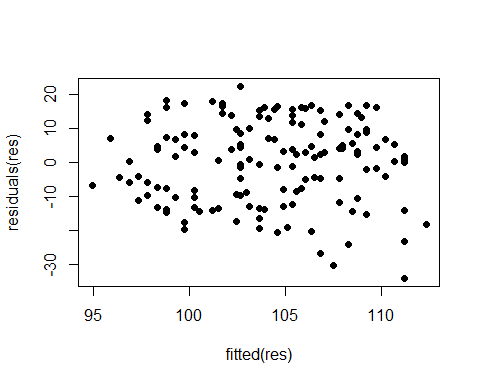


plot(residuals(res))



Not really normal, but only slightly skewed, so probably ok? The qq-plot reveals that large values are too large - see also the histogram.

1. Homogeneity of residuals



A pattern is visible after all - this might be problematic.

1. Is there a correlation between absolute residuals and fitted values?

cor.test(fitted(res), abs(residuals(res)))

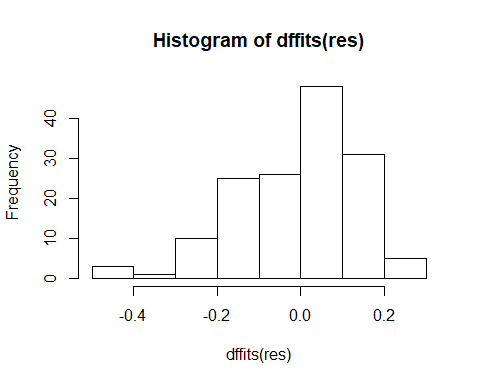
##   
## Pearson's product-moment correlation  
##   
## data: fitted(res) and abs(residuals(res))  
## t = 0.20154, df = 147, p-value = 0.8406  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1445658 0.1769476  
## sample estimates:  
## cor   
## 0.01662053

No, so this seems ok.

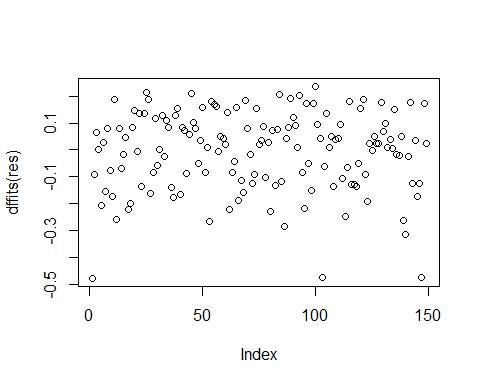
1. Influence diagnostics DFFits

max(abs(dffits(res)))

## [1] 0.4777614



plot(dffits(res))



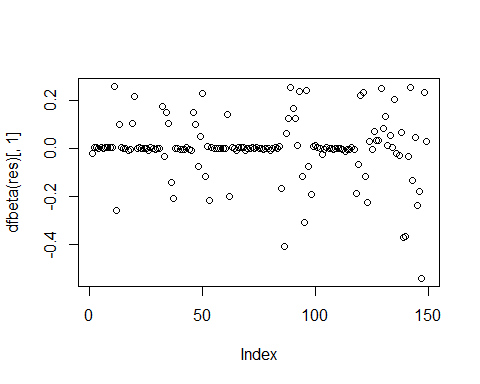
DFFit: Values around 2 are reason to worry, but all our values are below 0.5, so this is ok.

DFBetas

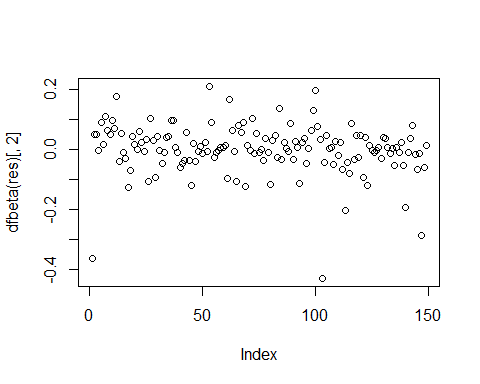
round(cbind(coefficients(res), coefficients(res)+  
t(apply(X=dfbeta(res), MARGIN=2, FUN=range))), 5)

## [,1] [,2] [,3]  
## (Intercept) 107.42240 106.8831 107.68072  
## z.age 3.28289 2.8544 3.49302  
## TestingLocationFife -5.62119 -5.9687 -5.06879

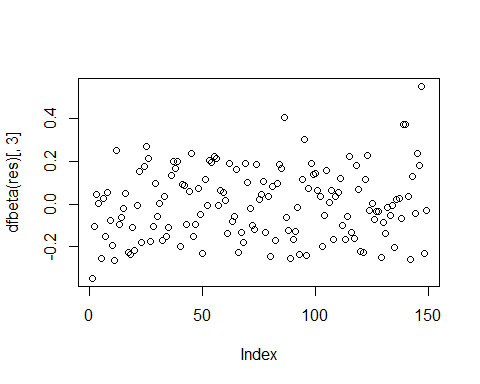
plot(dfbeta(res)[,1])#needs to be done by column



plot(dfbeta(res)[,2])



plot(dfbeta(res)[,3])



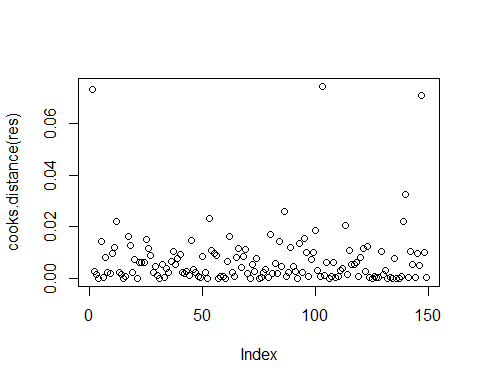
DFBeta: looks good, little variation.

Cook’s distance

max(cooks.distance(res))

## [1] 0.07440391

plot(cooks.distance(res))



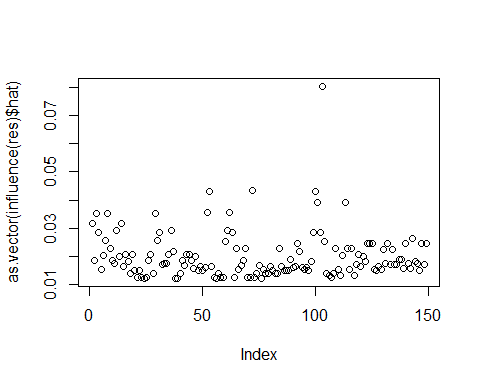
Cook’s distance: values above 1 are reason to worry, but our maximum value is 0.07.

Leverage

max(as.vector(influence(res)$hat))

## [1] 0.08024262

plot(as.vector(influence(res)$hat))



length(coefficients(res))# this is calculating k, which is the number of predictors

## [1] 3

2\*(3+1)/149

## [1] 0.05369128

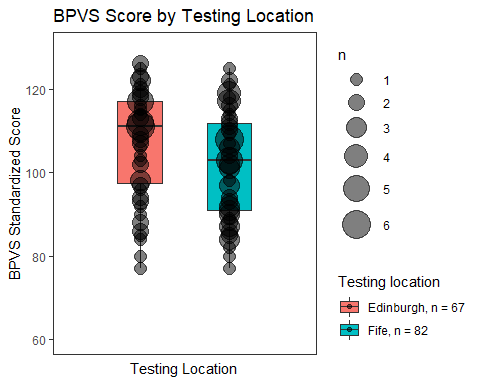
3\*(3+1)/149

## [1] 0.08053691

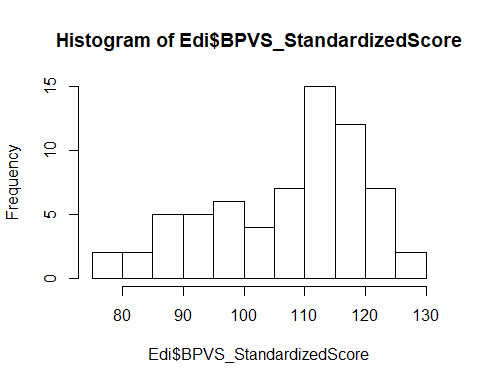
Leverage: values > 2*(k+1)/n or > 3*(k+1)/n are a reason to worry (k = number of predictors); in our case the thresholds are 0.05 and 0.08, and our max value is 0.08, so this is just ok.

Together, age in months and testing location can explain the BPVS scores better than a model just containing the intercept, X2(2) = 2621.8, p < .001. There is a significant effect of age (X2(1) = 1590.9, p = .001) and testing location (X2(1) = 1162.1, p = .005). With each standard deviation increase in age, the BPVS standardized score increases by 3.28 (95% CI [1.34; 5.23])units. When moving from Edinburgh to Fife, the expected score for a child of mean age decreases by 5.62 (95% CI [1.72; 9.52])units.

### Plot the effect of testing location



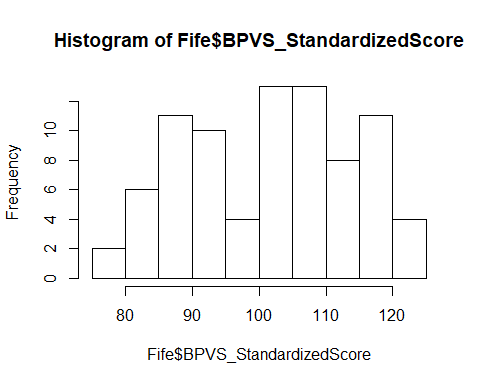
## We further examine the distribution of the standardised score in each testing location subset.



shapiro.test(Edi$BPVS\_StandardizedScore)#not normally distributed

##   
## Shapiro-Wilk normality test  
##   
## data: Edi$BPVS\_StandardizedScore  
## W = 0.9425, p-value = 0.003854

hist(Fife$BPVS\_StandardizedScore)



shapiro.test(Fife$BPVS\_StandardizedScore)#not normally distributed

##   
## Shapiro-Wilk normality test  
##   
## data: Fife$BPVS\_StandardizedScore  
## W = 0.96246, p-value = 0.01695

wilcox.test(BPVS.clean$BPVS\_StandardizedScore ~ BPVS.clean$TestingLocation, alternative = "greater")

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: BPVS.clean$BPVS\_StandardizedScore by BPVS.clean$TestingLocation  
## W = 3457, p-value = 0.003375  
## alternative hypothesis: true location shift is greater than 0

For the Fife children, the distribution is bimodal, which is not the case for the Edinburgh children.

## Plot age as continuous variable against BPVS score

