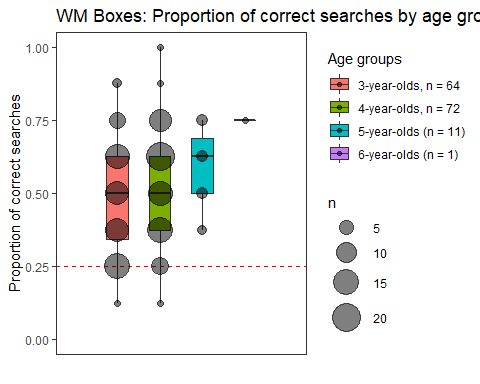
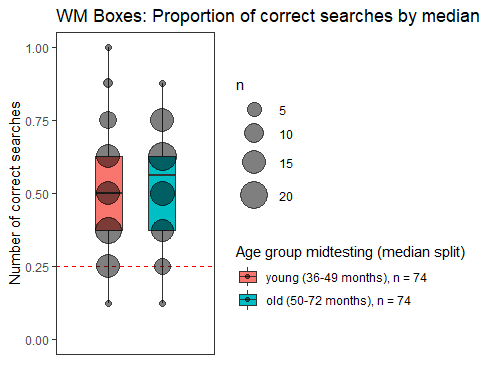
WM\_Boxes

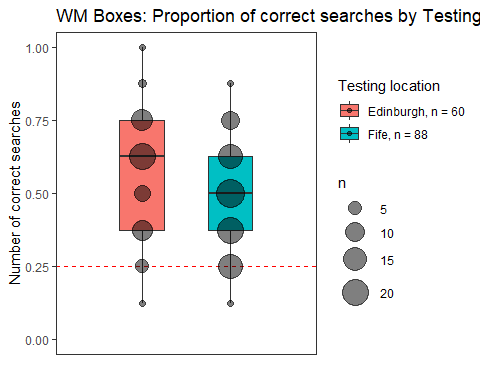
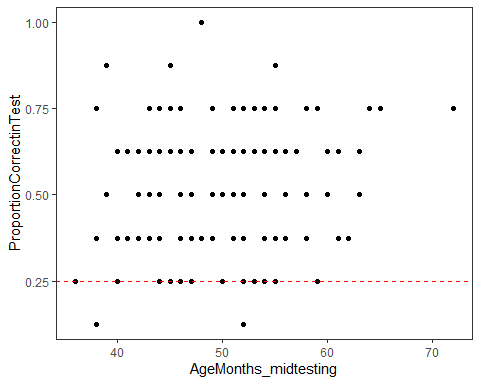
Eva Reindl

31 3 2020

Key findings:

* 148 valid datapoints
* DV “Proportion of correct searches”
  + Not normally distributed
  + Children significantly above chance (entire sample; 3- and 4-year-olds; young and old children)
  + 4-year-olds better than 3-year-olds; older children better than younger children
  + Children in Edinburgh better than children in Fife
  + Age and Testing location have significant effects on Proportion correct

## [1] "R version 3.6.1 (2019-07-05)"

# Dropouts

There is **1 dropout**. This child stopped after 4 test trials. This child will be removed from the analysis.

# Valid data

There are **148 valid datapoints**.

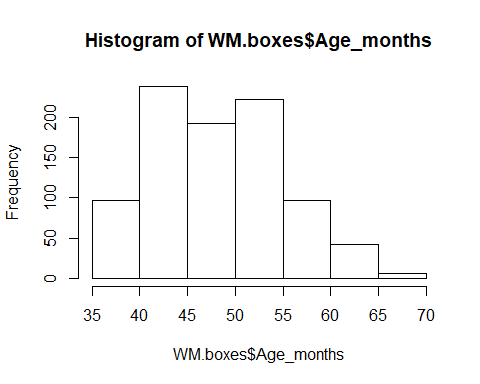
# Description of sample

## Gender distribution

There are 79 girls and 69 boys.

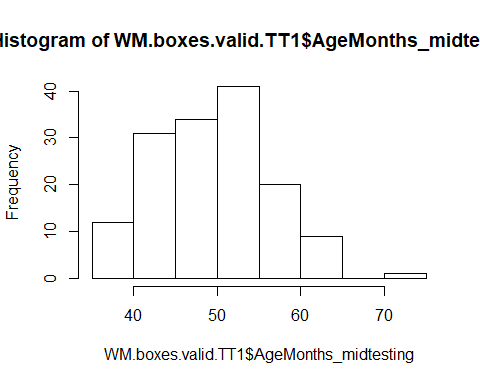
## Age

### Age at beginning of testing



At the beginning of testing, mean age was 48.81 months (SD = 6.87, range 36-70 months). There were 67 3-year-olds, 73 4-year-olds, and 8 5-year-olds.

### Age in the middle of testing



In the middle of testing, mean age was 49.92 months (SD = 6.81, range 36-72 months). There were 64 3-year-olds, 72 4-year-olds, 11 5-year-olds, and 1 6-year-old.

### Age mediansplit by entire sample

Median is 49 months

There were **74 young** and **74 old** children.

# Testing Location

88 children were from the Fife area, 60 children were from Edinburgh.

## Testing location and age

Fife:

* 3y: 39
* 4y: 37
* 5y: 11
* 6y: 1

Edinburgh:

* 3y: 25
* 4y: 35

# Warm-up: Number of correct trials

## Across the sample

Out of the 148 children, **105 children (71%) had both warm-up trials correct**, 42 (28%) had 1 trial correct and 1 child had no warm-up trial correct.

## By age groups (at midtesting)

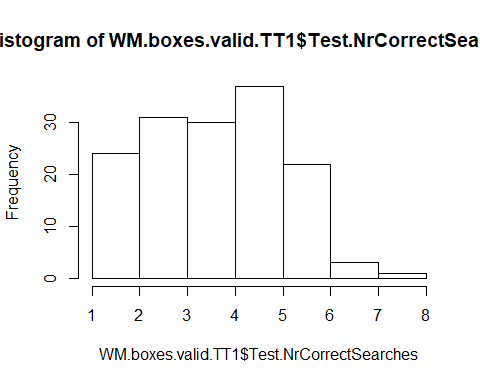
* **3y (n = 64): 41 (64%)** had both trials correct, 23 (36%) had 1 trial correct
* **4y (n = 72): 41 (75%)** had both trials correct, 23 (24%) had 1 trial correct, 1 child had no trial correct
* 5y (n = 11): 41 (82%) had both trials correct, 2 (18%) had 1 trial correct
* 6y (n = 1): had both trials correct

## By mediansplit (entire sample)

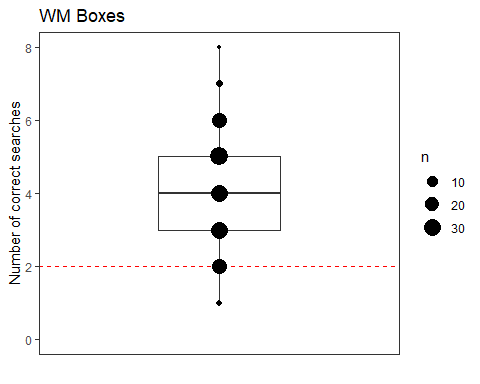
* **young (n = 74): 48 (65%)** had both trials correct, 26 (35%) had 1 trial correct
* **old (n = 74): 57 (77%)** had both trials correct, 16 (22%) had 1 trial correct, 1 child had no trial correct

# Number of correct searches

## Across the sample



The mean number of correct searches in the test trials is 4.09 (SD = 1.44, range 1-8). 50% of the children have 4 searches or less correct. The DV is not normally distributed, W = 0.942, p < .001.



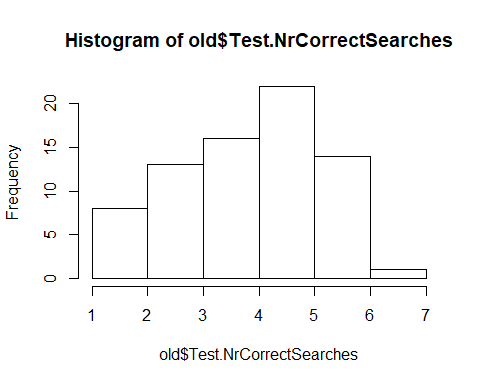
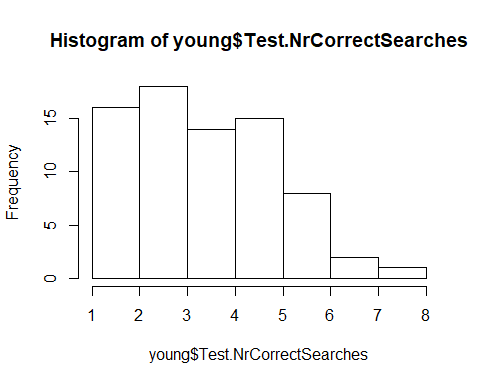
### Does performance deviate from chance?

The probability of getting a correct search on one platform by chance is 0.25. 0.25\*8 = 2, so by chance, one should get 2 trials correct

Children’s mean number of correct searches (4.09, SD = 1.44, range 1-8) is significantly higher than what we would expect by chance (2 trials correct), V = 7967, p < .001.

## By mediansplit (based on entire sample)

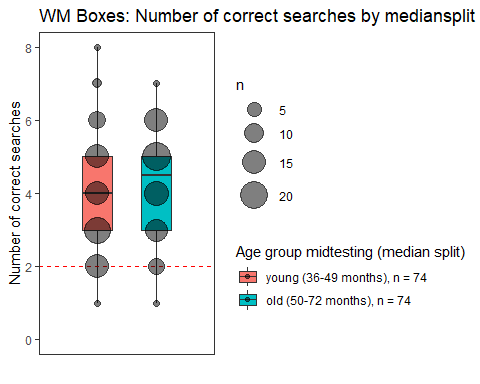
* Young children (n = 74): M = 3.86 (SD = 1.51, range 1-8)
* Old children (n = 74): M = 4.31 (SD = 1.33, range 1-7)

The variable “number of correct searches” is not normally distributed (old: W = 0.929, p < .001; young: W = 0.935, p < .001). Older children perform significantly better than younger children (one-tailed Wilcoxon rank sum test, W = 3272, p = .018).

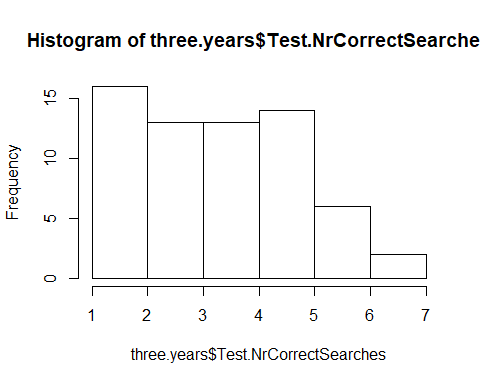
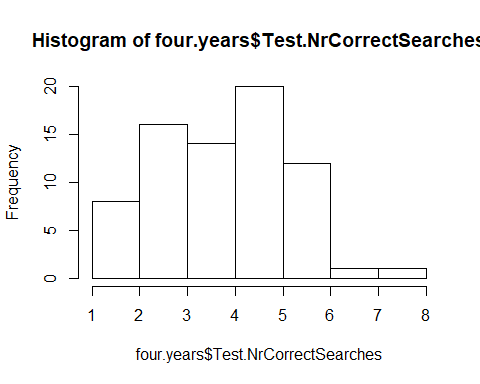
### Does performance deviate from chance?

Performance of young children is significantly better than chance value (2), V = 1760, p < .001. Performance of old children is significantly better than chance value (2), V = 2270.5, p < .001.



## By age groups

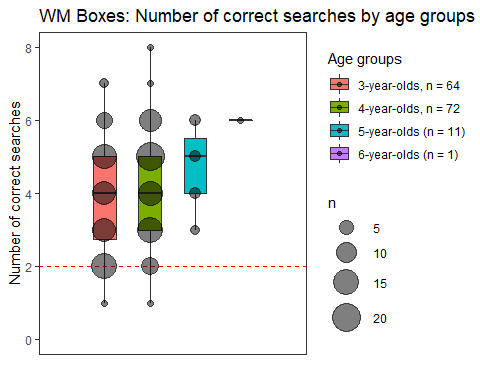
* 3y (n = 64): M = 3.78 (SD = 1.46, range 1-7)
* 4y (n = 72): M = 4.25 (SD = 1.41, range 1-8)
* 5y (n = 11): M = 4.64 (SD = 1.12, range 3-6)
* 6y (n = 1): 6

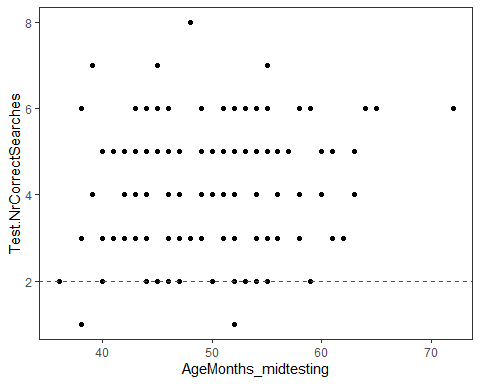
The variable “number of correct searches” is not normally distributed (3y: W = 0.929, p < .001; 4y: W = 0.946, p = .004). 4-year-olds (M = 4.25, SD = 1.41, range 1-8) perform significantly better than 3-year-olds (M = 3.78, SD = 1.46, range 1-7) (one-tailed Wilcoxon rank sum test, W = 1876, p = .028).

### Does performance deviate from chance?

Performance of 3-year-olds is significantly better than chance value (2), V = 1217.5, p < .001. Performance of 4-year-olds is significantly better than chance value (2), V = 2136, p < .001.

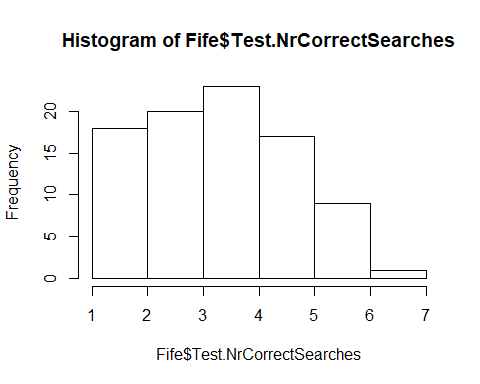
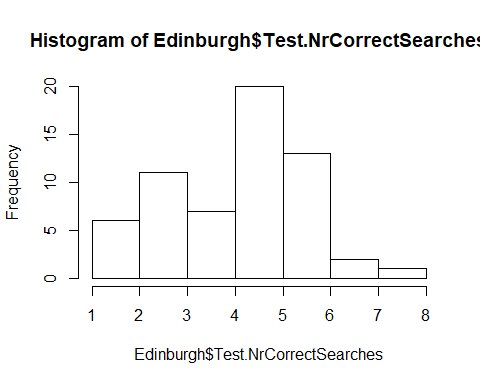


## Plot number of correct searches against age as continuous variable



## By testing location

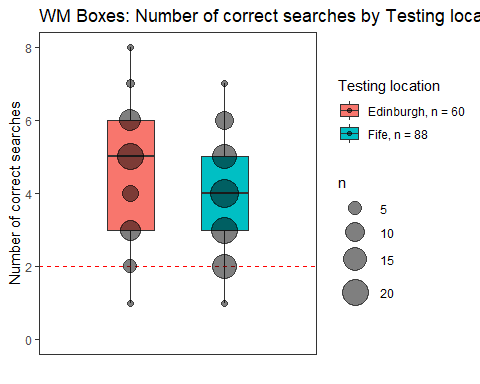
* Fife (n = 88): M = 3.78 (SD = 1.33, range 1-7)
* Edinburgh (n = 60): M = 4.53 (SD = 1.48, range 1-8)



The variable “number of correct searches” is not normally distributed (Fife: W = 0.933, p < .001; Edinburgh: W = 0.936, p = .003). Edinburgh children perform significantly better than Fife children (one-tailed Wilcoxon rank sum test, W = 3428, p = .001).

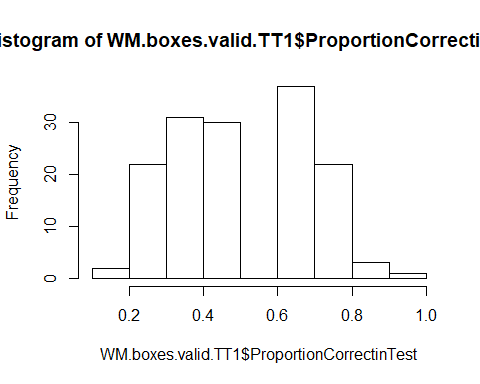
### Does performance deviate from chance?

Performance of Fife children is significantly better than chance value (2), V = 2545, p < .001. Performance of old children is significantly better than chance value (2), V = 1533.5, p < .001.

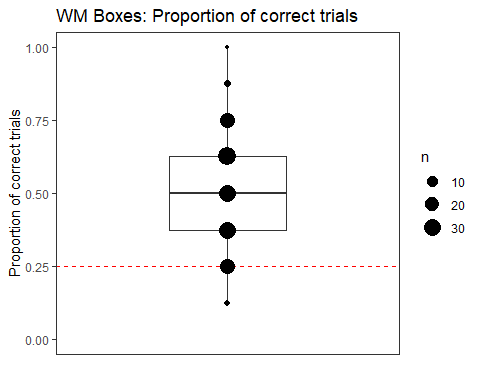


# DV “Proportion correct”

## Across the sample



Children had on average **51.10% (SD = 17.97, range 12.5-100%)** of the trials correct. This DV is **not normally distributed**, W = 0.942, p < .001.



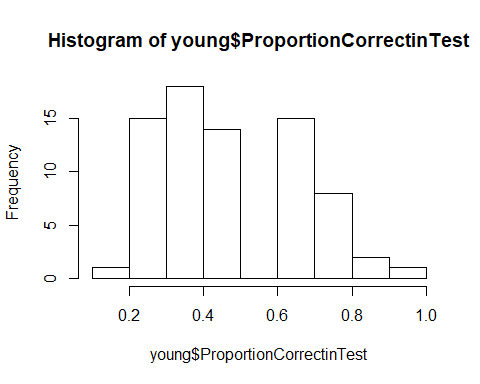
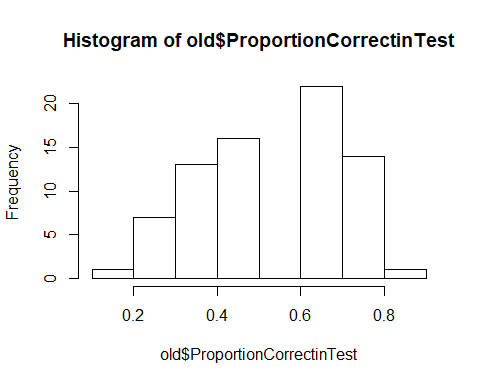
### Does performance deviate from chance?

The percentage of success if choosing completely randomly would be 0.25 (one would get 2 out of 8 trials right, which is 2/8 = 1/4 = 0.25)

Children’s average proportion of trials correct (51.10%, SD = 17.97, range 12.5-100%) is **significantly higher than the proportion we would expect by chance (0.25)**, V = 7967, p < .001.

## By mediansplit (based on entire sample)

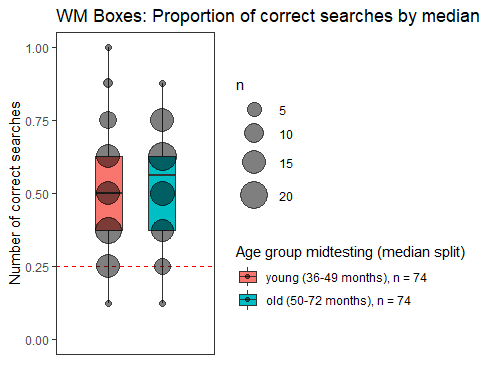
* **Young** children (n = 74): **M = 48.31%** (SD = 18.89, range 12.5-1)
* **Old** children (n = 74): **M = 53.88** (SD = 16.67, range 12.5-87.5)

The variable “proportion of correct searches” is not normally distributed (old: W = 0.929, p < .001; young: W = 0.935, p < .001). **Older children perform significantly better** than younger children (one-tailed Wilcoxon rank sum test, W = 2204, p = .018).

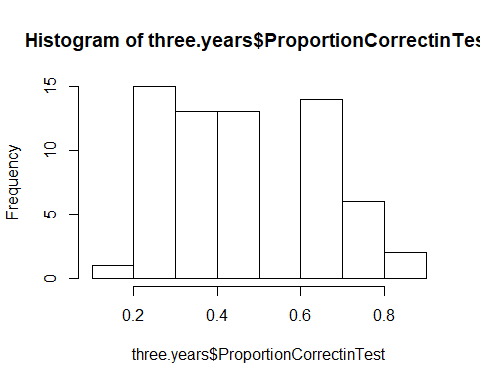
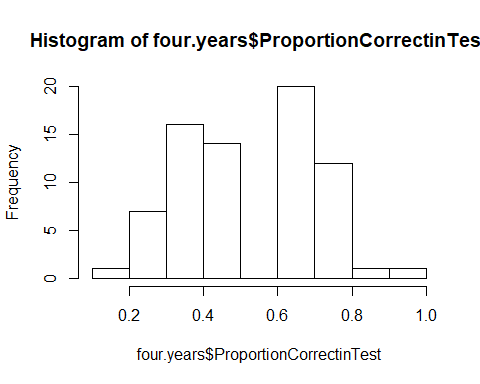
### Does performance deviate from chance?

Performance of young children is significantly better than chance value (2), V = 1760, p < .001. Performance of old children is significantly better than chance value (2), V = 2270.5, p < .001.



## By age groups

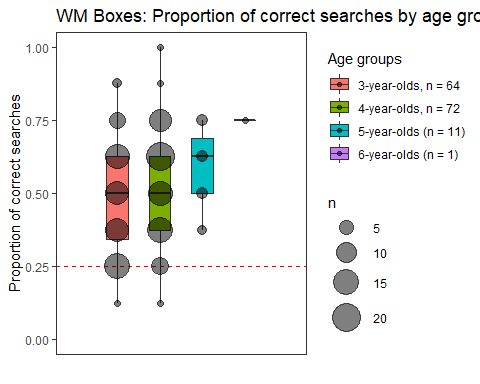
* **3y (n = 64): M = 47.26%** (SD = 18.29, range 12.5-87.5)
* **4y (n = 72): M = 53.12%** (SD = 17.65, range 12.5-1)
* 5y (n = 11): M = 57.95 (SD = 14.00, range 37.5-75.0)
* 6y (n = 1): 75.0

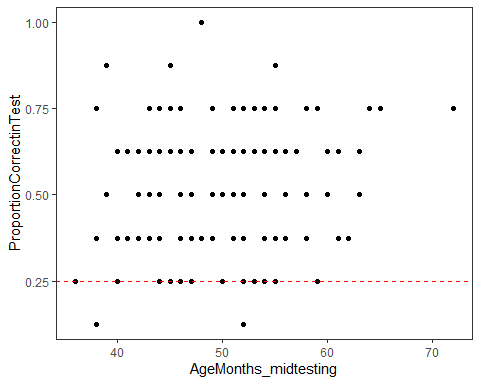
The variable “proportion of correct searches” is not normally distributed (3y: W = 0.929, p = .001; 4y: W = 0.946, p = .004). 4-year-olds (M = 53.12%, SD = 17.65, range 12.5-1) perform significantly better than 3-year-olds (M = 47.26%, SD = 18.29, range 12.5-87.5) (one-tailed Wilcoxon rank sum test, W = 1876, p = .028).

### Does performance deviate from chance?

Performance of 3-year-olds is significantly better than chance value (0.25), V = 1217.5, p < .001. Performance of 4-year-olds is significantly better than chance value (0.25), V = 2136, p < .001.

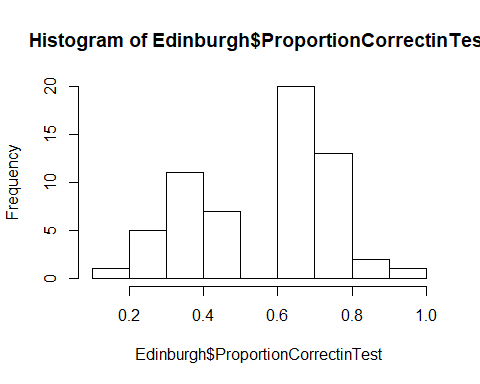
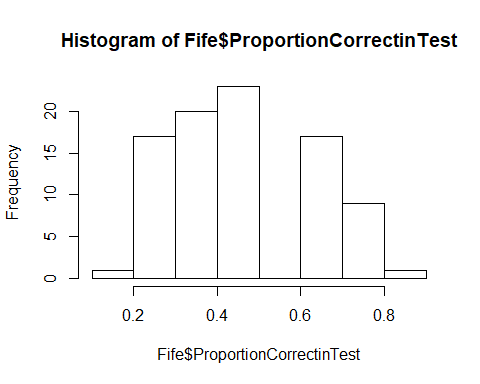


## Plot proportion of correct searches against age as continuous variable



## By testing location

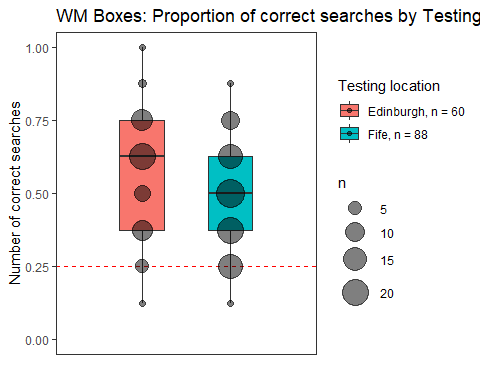
* Fife (n = 88): M = 47.30% (SD = 16.68, range 12.5-87.5)
* Edinburgh (n = 60): M = 56.67% (SD = 18.48, range 12.5-1)



The variable “proportion of correct searches” is not normally distributed (Fife: W = 0.933, p < .001; Edinburgh: W = 0.936, p = .003). Edinburgh children perform significantly better than Fife children (one-tailed Wilcoxon rank sum test, W = 3428, p = .001).

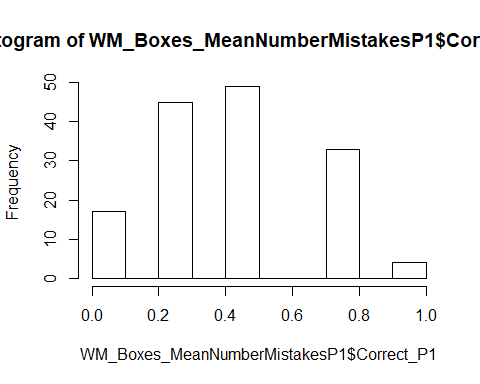
### Does performance deviate from chance?

Performance of Fife children is significantly better than chance value (0.25), V = 2545, p < .001. Performance of old children is significantly better than chance value (0.25), V = 1533.5, p < .001.



# Performance on platform 1

## Mean number of mistakes on platform 1



In the test trials, on platform 1, the mean number of mistakes was 0.43 (SD = 0.25, range 0-1). The variable “mean number of mistakes on platform 1” was not normally distributed, W = 0.905, p < .001.

### Does this performance deviate from chance?

On platform 1, if one would choose completely randomly, the propbability of making an error in each trial is 0.75.

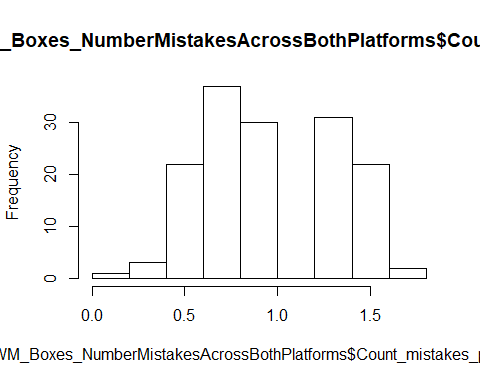
The **mean number of mistakes made on platform 1** **(0.43, SD = 0.25, range 0-1)** is significantly **smaller than what would be expected by chance** (1), V = 108, p < .001.

**Chimps:** Mean number of mistakes on platform 1 vs chance value (0.75): **Mean = 0.61**, 95% CI [0.57, 0.66], t(52)=-6.57, **p=0**

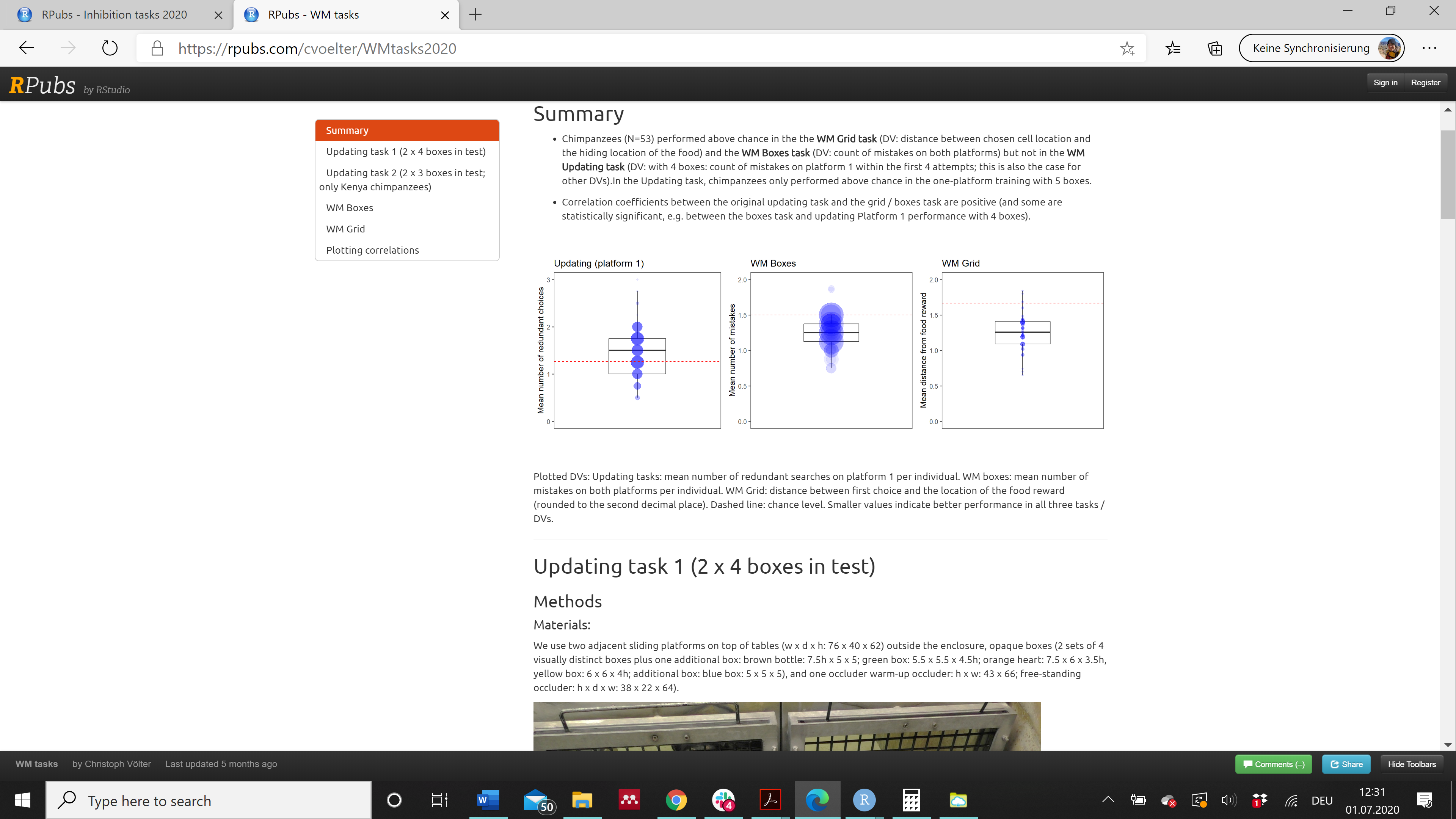
# Performance on both platforms

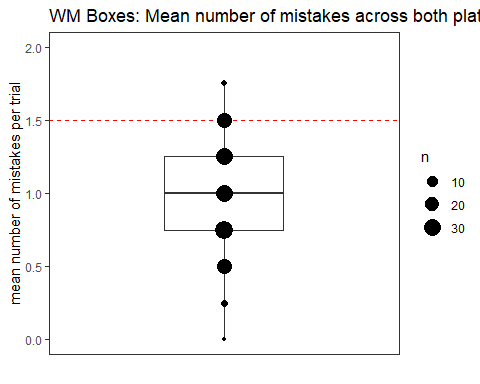
## Mean number of mistakes on both platforms

### Across the sample



Across the test trials, the mean number of mistakes per trial was **0.98 (SD = 0.36, range 0-1.75).** This variable is not normally distributed, W = 0.942, p < .001.   
**Children’s mean number of mistakes** is significantly lower than what would be expected by chance (0.75 + 075 = 1.5), V = 34, p < .001.  
**Chimps:** Mean number of mistakes on both platforms vs chance value (1.5): **Mean = 1.26**, 95% CI [1.2, 1.33], t(52)=-7.32, **p=0**

Children: Chimps:



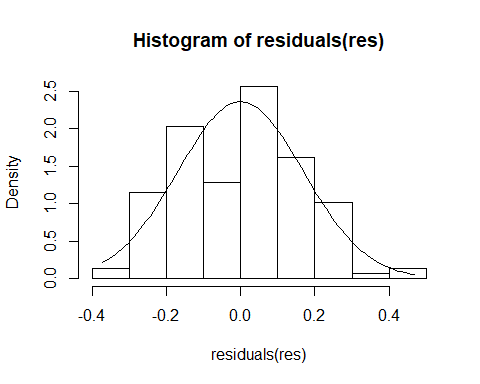
# Can age and testing location predict children’s proportion of correct searches?

res <- lm(ProportionCorrectinTest ~ z.age + TestingLocation + z.age:TestingLocation, data = WM.boxes.valid.TT1)

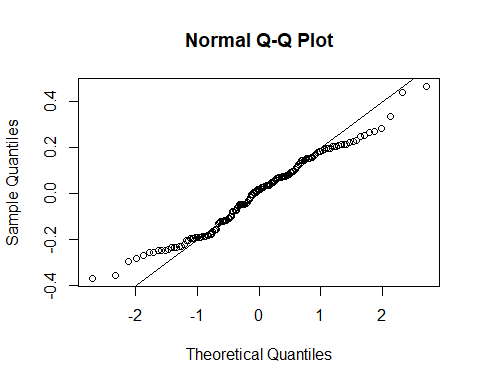
#### Check of assumptions

1. Normality of residuals

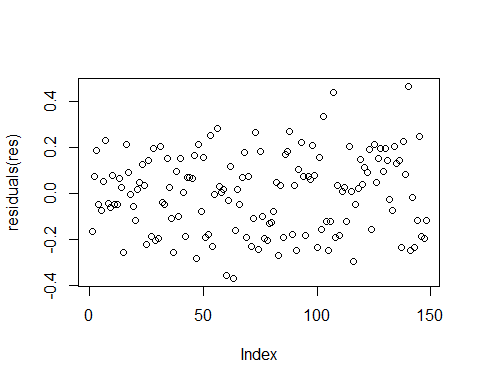
hist(residuals(res), probability=T)  
  
#add line:  
x=seq(from=min(residuals(res)),  
to=max(residuals(res)), length.out=100)  
lines(x=x, y=dnorm(x, mean=0,  
sd=sd(residuals(res))))



#a safer option is to look at the qq plot:  
qqnorm(residuals(res));qqline(residuals(res))



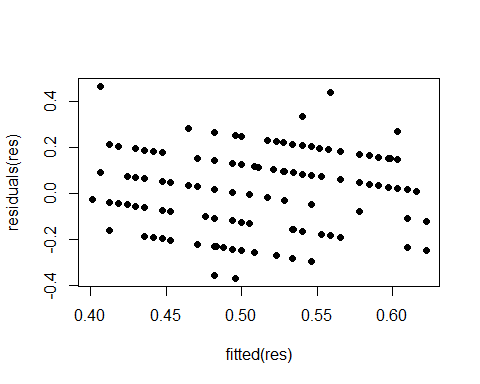
plot(residuals(res))



Histogram looks fairly normal. The qq-plot reveals that small values are too large and large values are too small - see also the histogram.

1. Homogeneity of residuals

plot(fitted(res), residuals(res), pch=19)



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 = -0.35056, df = 146, p-value = 0.7264  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1894575 0.1329659  
## sample estimates:  
## cor   
## -0.02900011

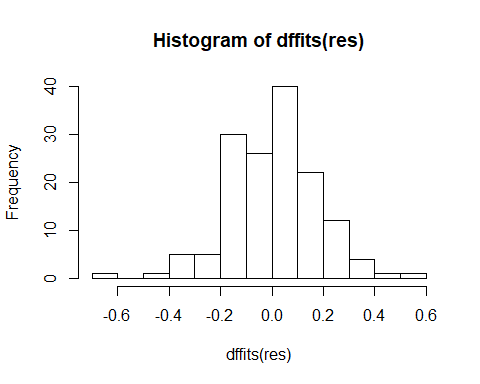
No, so this seems ok.

1. Influence diagnostics DFFits

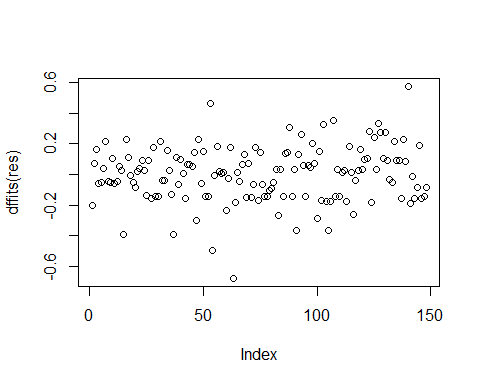
max(abs(dffits(res)))

## [1] 0.6785863

hist(dffits(res))



plot(dffits(res))



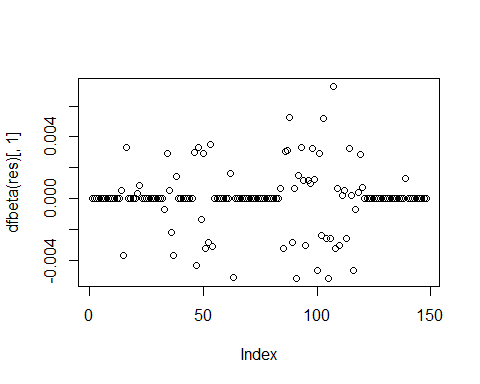
DFFit: Values around 2 are reason to worry, but the maximum value is 0.68, so this is ok.

DFBetas

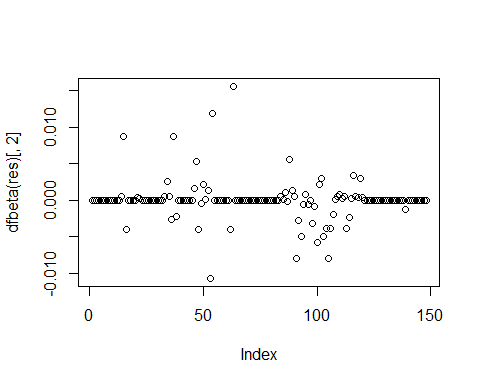
round(cbind(coefficients(res), coefficients(res)+  
t(apply(X=dfbeta(res), MARGIN=2, FUN=range))), 5)

## [,1] [,2] [,3]  
## (Intercept) 0.57114 0.56596 0.57843  
## z.age 0.04298 0.03233 0.05852  
## TestingLocationFife -0.10093 -0.10822 -0.09484  
## z.age:TestingLocationFife -0.00352 -0.01906 0.00712

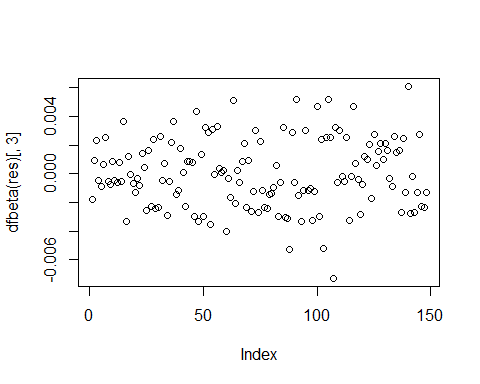
plot(dfbeta(res)[,1])#needs to be done by column



plot(dfbeta(res)[,2])



plot(dfbeta(res)[,3])



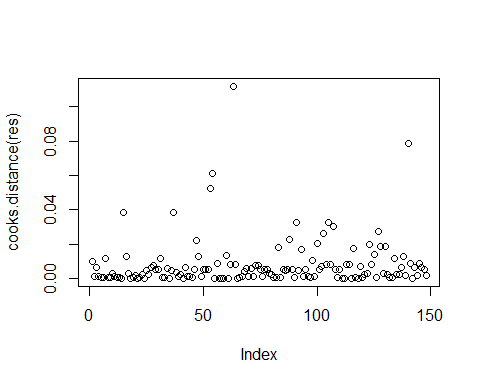
DFBeta: looks good, little variation.

Cook’s distance

max(cooks.distance(res))

## [1] 0.1117928

plot(cooks.distance(res))



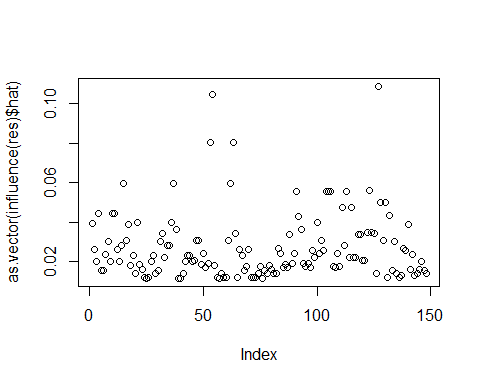
Cook’s distance: values above 1 are reason to worry, but our maximum value is 0.11.

Leverage

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

## [1] 0.1087807

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



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

## [1] 4

2\*(4+1)/148

## [1] 0.06756757

3\*(4+1)/148

## [1] 0.1013514

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 0.07 and 0.10, and our max value is 0.11, so there is evidence of some influential cases.

Together, age, testing location, and the interaction between age and testing location explain the data better than a null model only containing the intercept, X2(3) = 0.55, p < .001.

The interaction term does not increase model fit, X2(1) = 0.00, p = .910, so we exclude it from the model.

res <- lm(ProportionCorrectinTest ~ z.age + TestingLocation, data = WM.boxes.valid.TT1)

Together, age and testing location explain the data better than a null model only containing the intercept, X2(2) = 0.55, p < .001.

summary(res)

##   
## Call:  
## lm(formula = ProportionCorrectinTest ~ z.age + TestingLocation,   
## data = WM.boxes.valid.TT1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.37499 -0.13860 0.01777 0.13361 0.46981   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.57088 0.02201 25.938 < 2e-16 \*\*\*  
## z.age 0.04049 0.01408 2.875 0.004651 \*\*   
## TestingLocationFife -0.10074 0.02859 -3.524 0.000569 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1701 on 145 degrees of freedom  
## Multiple R-squared: 0.1163, Adjusted R-squared: 0.1041   
## F-statistic: 9.54 on 2 and 145 DF, p-value: 0.0001282

drop1(res)

## Single term deletions  
##   
## Model:  
## ProportionCorrectinTest ~ z.age + TestingLocation  
## Df Sum of Sq RSS AIC  
## <none> 4.1957 -521.35  
## z.age 1 0.23915 4.4349 -515.14  
## TestingLocation 1 0.35939 4.5551 -511.18

cbind(coefficients(res), confint(res))

## 2.5 % 97.5 %  
## (Intercept) 0.57088145 0.52738071 0.61438219  
## z.age 0.04048549 0.01265188 0.06831910  
## TestingLocationFife -0.10074381 -0.15724298 -0.04424463

There is a significant effect of age (X2(1) = 0.239, p = .004) and testing location (X2(1) = 0.359, p < .001). With each standard deviation increase in age, the proportion correct increases by 0.04 (95% CI [0.01; 0.01]) percentage points. When moving from Edinburgh to Fife, the expected score for a child of mean age decreases by 0.10 (95% CI [0.04; 0.16]) percentage points.