STATS 201/208 Assignment 3

Chunkai Meng / 857058772

Due Date: 3pm Thursday 29th September 2016

## Loading required package: s20x

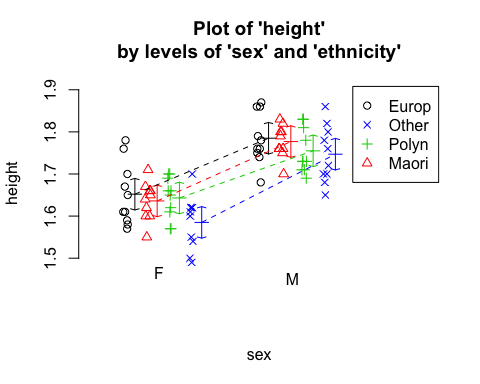
# Question 1

### Question of interest/goal of the study:

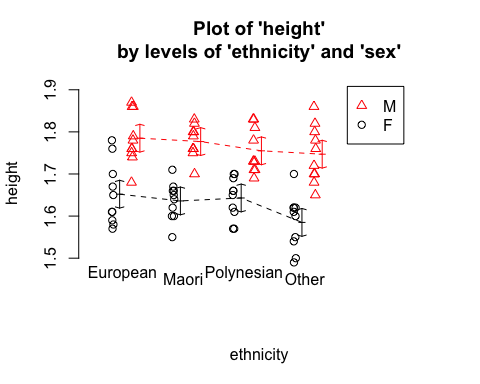
We investigated how people's height depends on their sex and their the ethnicity, and whether these factors influence each other.

### Read in and inspect the data:

height.df=read.table("height.txt",header=T)  
  
# create interaction plot(s)  
interactionPlots(height ~ sex + ethnicity, data=height.df)



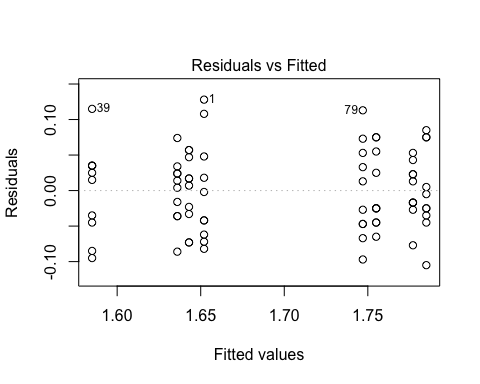
interactionPlots(height ~ ethnicity + sex, data=height.df)



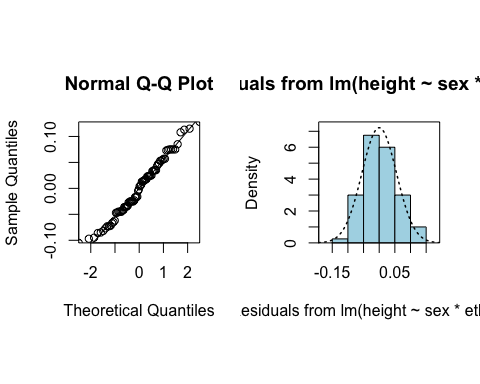
Here we see that the European males seem to be higher than all the other people on average. There are a slight difference among some of the above slops, but these is not so much interaction.  
However, for these type of data (i.e., 2 categorical explanatories) we start by fitting the model with interaction. We will see whether it can be simplifyed by Occam's razor rule.

### Fit model and check assumptions

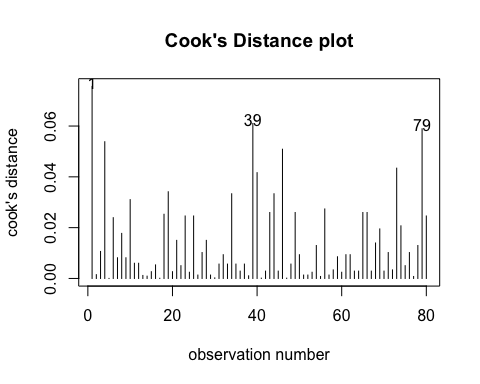
height.fit = lm(height ~ sex \* ethnicity, data = height.df)  
eovcheck(height.fit)



# showed approximately constant and not trend  
normcheck(height.fit)



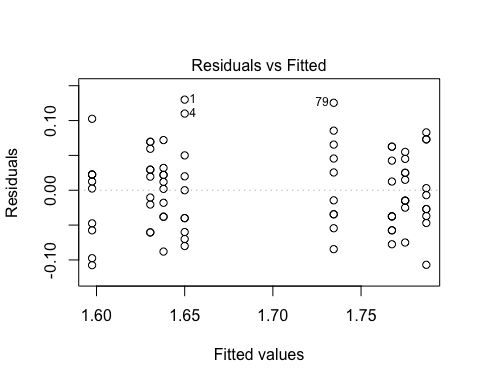
# residuals looks normal  
cooks20x(height.fit)



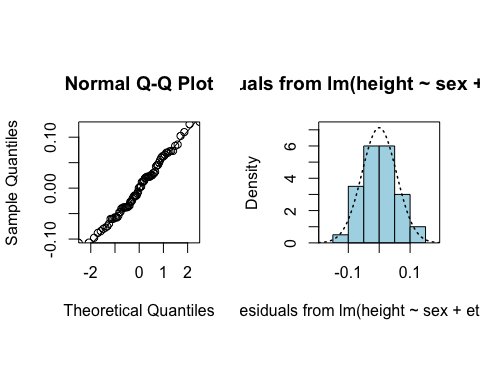
# no unduly influential points  
anova(height.fit)

## Analysis of Variance Table  
##   
## Response: height  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sex 1 0.37538 0.37538 112.2492 2.427e-16 \*\*\*  
## ethnicity 3 0.03033 0.01011 3.0232 0.0351 \*   
## sex:ethnicity 3 0.00641 0.00214 0.6389 0.5924   
## Residuals 72 0.24078 0.00334   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

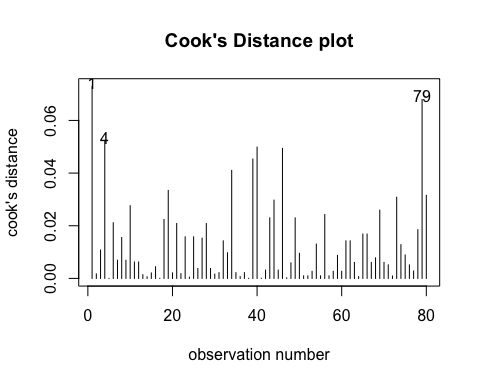
# There is no evidence of an interaction so we can simplify it to an additive model (i.e., no interaction term).  
  
height.fit2 = lm(height ~ sex + ethnicity, data = height.df)  
eovcheck(height.fit2)



# showed approximately constant and not trend  
normcheck(height.fit2)



# residuals looks normal  
cooks20x(height.fit2)



# no unduly influential points  
anova(height.fit2)

## Analysis of Variance Table  
##   
## Response: height  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sex 1 0.37538 0.37538 113.8942 <2e-16 \*\*\*  
## ethnicity 3 0.03033 0.01011 3.0675 0.033 \*   
## Residuals 75 0.24719 0.00330   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# It's ok now, let use summary to get coefficients R^2, and summary2way to get Tukey pairwise comparisons within and between the two factors.  
summary(height.fit2)

##   
## Call:  
## lm(formula = height ~ sex + ethnicity, data = height.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.10750 -0.03850 0.00100 0.03463 0.13000   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.65000 0.01435 114.963 < 2e-16 \*\*\*  
## sexM 0.13700 0.01284 10.672 < 2e-16 \*\*\*  
## ethnicityMaori -0.01200 0.01815 -0.661 0.51064   
## ethnicityOther -0.05250 0.01815 -2.892 0.00501 \*\*   
## ethnicityPolynesian -0.01950 0.01815 -1.074 0.28622   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05741 on 75 degrees of freedom  
## Multiple R-squared: 0.6214, Adjusted R-squared: 0.6012   
## F-statistic: 30.77 on 4 and 75 DF, p-value: 3.695e-15

summary2way(height.fit2, page="nointeraction")

##   
##   
## sex comparisons:  
##   
## Estimate Tukey.L Tukey.U Tukey.p  
## F - M -0.137 -0.1626 -0.1114 0  
##   
##   
## ethnicity comparisons:  
##   
## Estimate Tukey.L Tukey.U Tukey.p  
## European - Maori 0.0120 -0.0357 0.0597 0.9113  
## European - Other 0.0525 0.0048 0.1002 0.0253  
## European - Polynesian 0.0195 -0.0282 0.0672 0.7063  
## Maori - Other 0.0405 -0.0072 0.0882 0.1243  
## Maori - Polynesian 0.0075 -0.0402 0.0552 0.9761  
## Other - Polynesian -0.0330 -0.0807 0.0147 0.2732

# extract the CIs where Tukey adjusted P-values are less than 0.05.  
## Estimate Tukey.L Tukey.U Tukey.p  
## F - M -0.137 -0.1626 -0.1114 0  
## European - Other 0.0525 0.0048 0.1002 0.0253