library(readr)  
library(alr4)  
library(dplyr)  
day <- read\_csv("~/126 Regression/Bike-Sharing-Dataset/day.csv")

# Preliminary observations

Welcome to the capital bike share data set. Today we will be exploring the relationship between different variables and how they impact total bike count in future settings..  
Before we dive into exploratory data analysis, we will vet the data for any irregularites and remove them as fit.

## Warning in summary.lm(test0): essentially perfect fit: summary may be  
## unreliable

##   
## Call:  
## lm(formula = cnt ~ . - cnt, data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.541e-11 -2.540e-13 -3.000e-15 2.151e-13 2.653e-11   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.216e-13 4.601e-13 -1.351e+00 0.17710   
## instant 1.442e-15 6.651e-15 2.170e-01 0.82848   
## dteday NA NA NA NA   
## season -9.284e-13 1.060e-13 -8.761e+00 < 2e-16 \*\*\*  
## yr -6.468e-13 2.455e-12 -2.630e-01 0.79229   
## mnth 2.040e-13 2.051e-13 9.940e-01 0.32043   
## holiday -1.644e-13 3.642e-13 -4.510e-01 0.65190   
## weekday 3.328e-14 2.971e-14 1.120e+00 0.26307   
## workingday -3.965e-13 2.203e-13 -1.799e+00 0.07238 .   
## weathersit 5.849e-14 1.482e-13 3.950e-01 0.69324   
## temp -2.146e-12 2.535e-12 -8.460e-01 0.39766   
## atemp 8.693e-12 2.873e-12 3.026e+00 0.00256 \*\*   
## hum 7.270e-13 5.709e-13 1.273e+00 0.20333   
## windspeed 7.140e-13 8.390e-13 8.510e-01 0.39505   
## casual 1.000e+00 1.590e-16 6.288e+15 < 2e-16 \*\*\*  
## registered 1.000e+00 9.141e-17 1.094e+16 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.569e-12 on 716 degrees of freedom  
## Multiple R-squared: 1, Adjusted R-squared: 1   
## F-statistic: 7.944e+31 on 14 and 716 DF, p-value: < 2.2e-16

This tells us we have other responses we must remove. The other responses are: casual and registered. In addtion to removing those, we will get rid of non-variables. Our motivation is to predict count for future years, thus we must remove the categorical variable, year.

day <- select(day, c(- casual, - registered, - instant, - dteday, -yr) )   
y <- day$cnt  
test0 <- lm( y ~ . - cnt , data = day )  
summary(test0)

##   
## Call:  
## lm(formula = y ~ . - cnt, data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4278.2 -973.0 -190.1 1056.6 4006.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3033.31 360.95 8.404 2.30e-16 \*\*\*  
## season 482.12 84.11 5.732 1.46e-08 \*\*\*  
## mnth -29.27 26.23 -1.116 0.2649   
## holiday -479.95 308.83 -1.554 0.1206   
## weekday 61.35 25.04 2.450 0.0145 \*   
## workingday 112.39 110.62 1.016 0.3099   
## weathersit -494.64 120.24 -4.114 4.35e-05 \*\*\*  
## temp 2381.14 2156.21 1.104 0.2698   
## atemp 3634.09 2441.58 1.488 0.1371   
## hum -2234.15 478.65 -4.668 3.63e-06 \*\*\*  
## windspeed -3145.35 700.33 -4.491 8.25e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1340 on 720 degrees of freedom  
## Multiple R-squared: 0.528, Adjusted R-squared: 0.5214   
## F-statistic: 80.53 on 10 and 720 DF, p-value: < 2.2e-16

From the summary everything appears to be fine, but futher we are going to check for redundant variables using VIF. If there are two varaibles with a high correlation, then one must be removed, preferably the one with the higher number.

vif(test0)

## season mnth holiday weekday workingday weathersit   
## 3.547489 3.332573 1.083084 1.023842 1.076379 1.744807   
## temp atemp hum windspeed   
## 63.317230 64.343265 1.888988 1.197228

Since atemp has the highest co-variances, we will remove it.

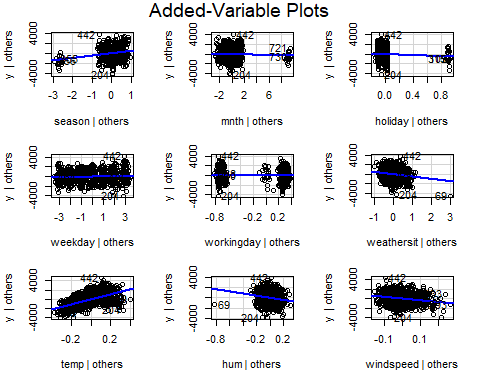
day <- select(day, c(- atemp, - cnt ))  
test0 <- lm( y ~ . , data = day )  
vif(test0)

## season mnth holiday weekday workingday weathersit   
## 3.540512 3.331034 1.081396 1.021227 1.076325 1.737620   
## temp hum windspeed   
## 1.210200 1.875768 1.163155

Now that the data has been cleared of any non-variables, other potiental responses, and redundacies, we can continue to exploratory data analysis.

## Exploratory Data Analysis

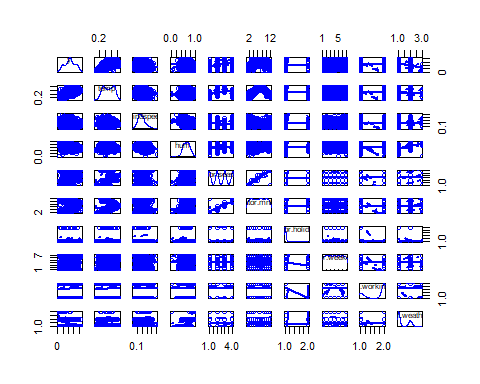
avPlots(test0)



Here we view that the most impactful vairbale is tempature. The other variables do have correlations, but not as much as temp.

scatterplotMatrix(~ y + temp + windspeed + hum + factor(season) + factor(mnth) + factor(holiday) + factor(weekday) + factor(workingday) + factor(weathersit) , data =day)

## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread  
  
## Warning in smoother(x[subs], y[subs], col = smoother.args$col[i], log.x =  
## FALSE, : could not fit negative part of the spread



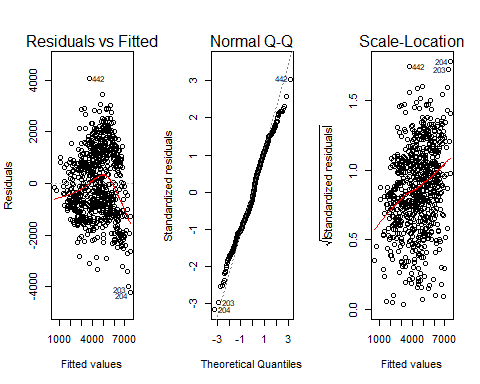
This is ugly but as we see there must be transformations in order to attain a correlation between predictors and the response.

### Diagnostics

test1 <- test0  
summary(test1)

##   
## Call:  
## lm(formula = y ~ ., data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4209.1 -987.7 -170.3 1043.1 4053.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3193.23 344.87 9.259 < 2e-16 \*\*\*  
## season 487.67 84.09 5.799 9.97e-09 \*\*\*  
## mnth -30.11 26.25 -1.147 0.2517   
## holiday -498.10 308.85 -1.613 0.1072   
## weekday 59.46 25.02 2.376 0.0177 \*   
## workingday 111.22 110.71 1.005 0.3154   
## weathersit -506.12 120.10 -4.214 2.82e-05 \*\*\*  
## temp 5559.67 298.35 18.635 < 2e-16 \*\*\*  
## hum -2174.55 477.37 -4.555 6.14e-06 \*\*\*  
## windspeed -3321.20 690.87 -4.807 1.86e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1341 on 721 degrees of freedom  
## Multiple R-squared: 0.5265, Adjusted R-squared: 0.5206   
## F-statistic: 89.08 on 9 and 721 DF, p-value: < 2.2e-16

par(mfrow = c( 1, 3))  
for( i in 1:3){  
 plot( test1 , which = i)}



par(mfrow = c( 1, 1))

The residuals vs fitted and scale-location plot needs work. To this we will use MLR transformation. First checking if we need to add a constant to any numeric variable(need non-zero numbers for transformation).

min(day$temp)

## [1] 0.0591304

min(day$windspeed)

## [1] 0.0223917

min(day$hum)

## [1] 0

Since humidity has value of zero we must add a constant before we proceed.

day$hum1 <- with(day, (hum + .001 ))   
diag1 <- powerTransform(cbind(hum1, windspeed , temp) ~ 1, day)  
summary(diag1)

## bcPower Transformations to Multinormality   
## Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd  
## hum1 0.9971 1.0 0.8058 1.1885  
## windspeed 0.4357 0.5 0.2993 0.5721  
## temp 0.8385 1.0 0.6580 1.0190  
##   
## Likelihood ratio test that transformation parameters are equal to 0  
## (all log transformations)  
## LRT df pval  
## LR test, lambda = (0 0 0) 677.8528 3 < 2.22e-16  
##   
## Likelihood ratio test that no transformations are needed  
## LRT df pval  
## LR test, lambda = (1 1 1) 64.12643 3 7.7161e-14

Adding a small constant that will have a small impact on the data we find that windspeed, should have a square root transformation.

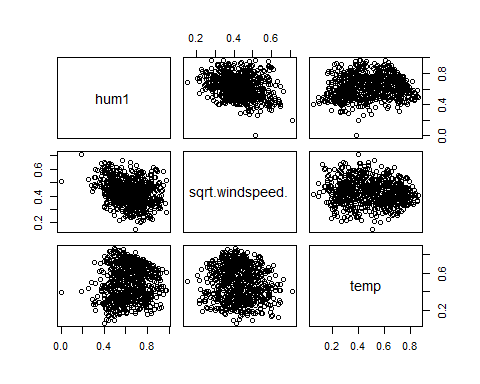
testTransform(diag1, lambda = c(1, .5, 1))

## LRT df pval  
## LR test, lambda = (1 0.5 1) 3.818622 3 0.28173

Here we get a small LRT, which is very good! Despite this we have a p-value of .2

Following with the proceedures from the MLS pdf we have:

day\_trsf <- with(day, data.frame( hum1 , sqrt(windspeed), temp ))  
pairs(day\_trsf)



windspeed.sqrt <- sqrt(day$windspeed)  
day <- cbind( day, windspeed.sqrt )  
day <- select(day , -c(hum, windspeed) )

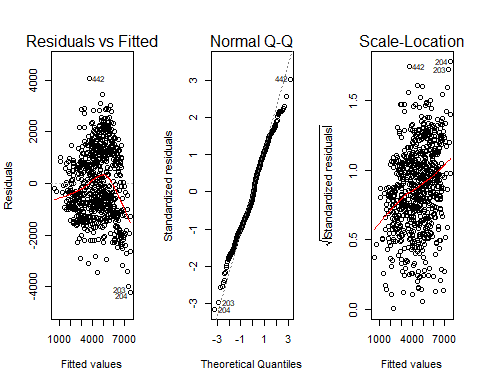
In the plots we can see a grouping between the predictors at hand, because of this it is closer to a linear relationship.

Now that we have transformed our predictors, our data could follow linearity assumptions.

test2 <- lm(y ~. , data = day)  
summary(test2)

##   
## Call:  
## lm(formula = y ~ ., data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4208.8 -979.5 -187.7 1047.6 4043.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3759.62 432.35 8.696 < 2e-16 \*\*\*  
## season 487.95 84.16 5.798 1.00e-08 \*\*\*  
## mnth -31.32 26.27 -1.192 0.2336   
## holiday -502.09 309.07 -1.625 0.1047   
## weekday 59.84 25.04 2.390 0.0171 \*   
## workingday 110.44 110.79 0.997 0.3192   
## weathersit -507.78 120.23 -4.223 2.72e-05 \*\*\*  
## temp 5592.32 298.43 18.739 < 2e-16 \*\*\*  
## hum1 -2163.07 477.94 -4.526 7.04e-06 \*\*\*  
## windspeed.sqrt -2834.46 603.85 -4.694 3.21e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1342 on 721 degrees of freedom  
## Multiple R-squared: 0.5258, Adjusted R-squared: 0.5199   
## F-statistic: 88.83 on 9 and 721 DF, p-value: < 2.2e-16

par(mfrow = c( 1, 3))  
for( i in 1:3){  
 plot( test2 , which = i)  
   
}

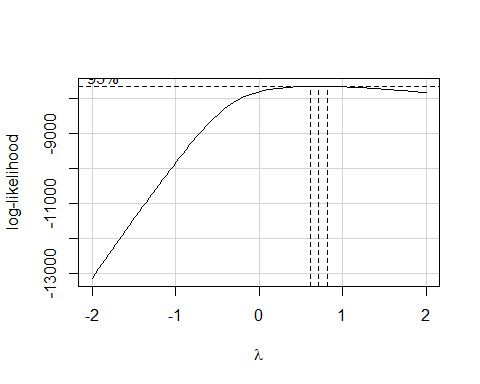


par(mfrow = c( 1, 1))

There area still issues with all 3 of our plots, despite this since our data set is of length the qq-plots do not have to be perfect.

In order to fix the violations in linearity-assumptions, we will consider a response transformation:

bc <- boxCox(test2)

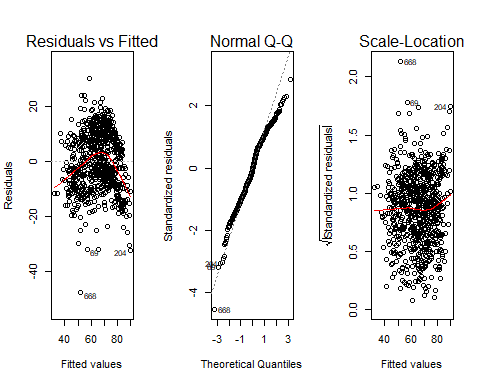


lambda.opt <- bc$x[which.max(bc$y)]   
lambda.opt

## [1] 0.7070707

Now the lambda is closest to .5, so we should transform our response with a square root.

y.sqrt <- sqrt(y)  
test2 <- lm(y.sqrt ~ . , data = day)   
  
par(mfrow = c( 1, 3))  
for( i in 1:3){  
 plot( test2 , which = i)  
   
}



par(mfrow = c( 1, 1))

Here we see that our plots are close to following linear assumptions. With nothing else to do we can continue to the AIC/BIC methods.

### AIC/BIC Model selection.

Doing the AIC/ BIC Methods, we have:

#Inital model  
m0 = lm(y.sqrt ~ temp , day)  
  
#Full model  
f = ~ season + mnth + holiday + weekday + workingday + weathersit + temp + hum1 + windspeed.sqrt  
#AIC method  
m.forward = step(m0 , f ,direction = 'forward' )

## Start: AIC=3649.95  
## y.sqrt ~ temp  
##   
## Df Sum of Sq RSS AIC  
## + weathersit 1 10029.5 97125 3580.1  
## + season 1 8980.4 98174 3588.0  
## + hum1 1 5948.3 101206 3610.2  
## + mnth 1 4355.4 102799 3621.6  
## + windspeed.sqrt 1 3660.3 103494 3626.5  
## + weekday 1 610.4 106544 3647.8  
## + holiday 1 555.6 106599 3648.1  
## <none> 107155 3649.9  
## + workingday 1 251.7 106903 3650.2  
##   
## Step: AIC=3580.11  
## y.sqrt ~ temp + weathersit  
##   
## Df Sum of Sq RSS AIC  
## + season 1 10270.3 86855 3500.4  
## + mnth 1 5392.9 91732 3540.4  
## + windspeed.sqrt 1 3415.5 93710 3555.9  
## + weekday 1 775.9 96349 3576.2  
## + holiday 1 752.6 96373 3576.4  
## + workingday 1 517.2 96608 3578.2  
## + hum1 1 384.7 96740 3579.2  
## <none> 97125 3580.1  
##   
## Step: AIC=3500.41  
## y.sqrt ~ temp + weathersit + season  
##   
## Df Sum of Sq RSS AIC  
## + windspeed.sqrt 1 1472.47 85382 3489.9  
## + hum1 1 1413.54 85441 3490.4  
## + weekday 1 805.60 86049 3495.6  
## + holiday 1 760.41 86095 3496.0  
## + workingday 1 563.55 86291 3497.7  
## + mnth 1 312.23 86543 3499.8  
## <none> 86855 3500.4  
##   
## Step: AIC=3489.91  
## y.sqrt ~ temp + weathersit + season + windspeed.sqrt  
##   
## Df Sum of Sq RSS AIC  
## + hum1 1 2601.53 82781 3469.3  
## + weekday 1 840.97 84541 3484.7  
## + holiday 1 760.56 84622 3485.4  
## + workingday 1 536.42 84846 3487.3  
## + mnth 1 389.16 84993 3488.6  
## <none> 85382 3489.9  
##   
## Step: AIC=3469.29  
## y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1  
##   
## Df Sum of Sq RSS AIC  
## + holiday 1 719.56 82061 3464.9  
## + weekday 1 598.04 82183 3466.0  
## + workingday 1 456.53 82324 3467.3  
## + mnth 1 229.53 82551 3469.3  
## <none> 82781 3469.3  
##   
## Step: AIC=3464.91  
## y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1 +   
## holiday  
##   
## Df Sum of Sq RSS AIC  
## + weekday 1 478.81 81583 3462.6  
## + workingday 1 229.40 81832 3464.9  
## <none> 82061 3464.9  
## + mnth 1 192.59 81869 3465.2  
##   
## Step: AIC=3462.63  
## y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1 +   
## holiday + weekday  
##   
## Df Sum of Sq RSS AIC  
## + workingday 1 225.33 81357 3462.6  
## <none> 81583 3462.6  
## + mnth 1 214.90 81368 3462.7  
##   
## Step: AIC=3462.61  
## y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1 +   
## holiday + weekday + workingday  
##   
## Df Sum of Sq RSS AIC  
## <none> 81357 3462.6  
## + mnth 1 209.39 81148 3462.7

This gives us 8 variables to use.

Now with the BIC method

n=731  
step(m0, f, direction = 'forward', k = log(n), trace = 0)

##   
## Call:  
## lm(formula = y.sqrt ~ temp + weathersit + season + windspeed.sqrt +   
## hum1, data = day)  
##   
## Coefficients:  
## (Intercept) temp weathersit season   
## 60.641 46.560 -4.313 3.517   
## windspeed.sqrt hum1   
## -23.188 -17.998

The BIC method leaves us with 5 variables . In addition to this I calculated goodness of fit by hand and it was 16.6.

Copying and pasting the results from both:

BIC <- lm(formula = y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1, data = day)  
AIC <- lm(y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1 + holiday + weekday + workingday , data = day)  
anova(BIC, AIC)

## Analysis of Variance Table  
##   
## Model 1: y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1  
## Model 2: y.sqrt ~ temp + weathersit + season + windspeed.sqrt + hum1 +   
## holiday + weekday + workingday  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 725 82781   
## 2 722 81357 3 1423.7 4.2115 0.005761 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

From the anova table, we accept the BIC model.

Must do BIC MLR diagnostics, then make the model plots, if still not good, make a non-parallel model.

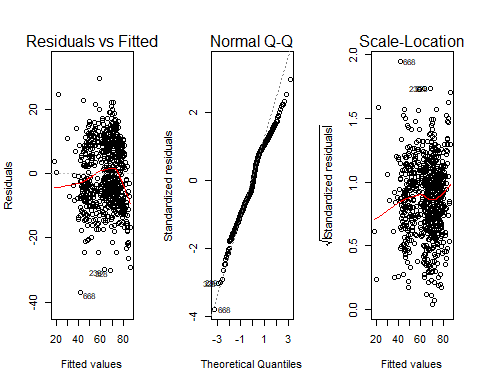
In this chunk I will be usign the linear model BIC and factoring all categorical variables.

BIC.factor <- lm(formula = y.sqrt ~ temp + factor(weathersit) + factor(season) + windspeed.sqrt + hum1 , data = day)  
summary(BIC.factor)

##   
## Call:  
## lm(formula = y.sqrt ~ temp + factor(weathersit) + factor(season) +   
## windspeed.sqrt + hum1, data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.979 -7.049 -1.275 8.710 29.617   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 57.1083 3.3357 17.120 < 2e-16 \*\*\*  
## temp 51.4157 3.7249 13.803 < 2e-16 \*\*\*  
## factor(weathersit)2 -1.6286 0.9891 -1.647 0.1001   
## factor(weathersit)3 -19.6138 2.5300 -7.753 3.06e-14 \*\*\*  
## factor(season)2 8.1372 1.3854 5.874 6.50e-09 \*\*\*  
## factor(season)3 4.0548 1.8258 2.221 0.0267 \*   
## factor(season)4 13.0634 1.1839 11.035 < 2e-16 \*\*\*  
## windspeed.sqrt -22.3666 4.5660 -4.898 1.19e-06 \*\*\*  
## hum1 -20.5732 3.5761 -5.753 1.30e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10.06 on 722 degrees of freedom  
## Multiple R-squared: 0.5971, Adjusted R-squared: 0.5926   
## F-statistic: 133.7 on 8 and 722 DF, p-value: < 2.2e-16

From the summary we see that our increased and Residual standard error decreased. This is good. In the chart we have an issue with the p-value for factor(weekday)[1]. This is not grounds to remove it, it just implies that the slope for factor(weekday)[1] is possibly 0.

par(mfrow = c( 1, 3))  
for( i in 1:3){  
 plot( BIC.factor , which = i)  
   
}



par(mfrow = c( 1,1))

Running a second diagnostic to see if we can transform and make residuals linear.

diag2 <- powerTransform(cbind(hum1, windspeed.sqrt , temp) ~ 1, day)  
summary(diag2)

## bcPower Transformations to Multinormality   
## Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd  
## hum1 0.9971 1 0.8058 1.1885  
## windspeed.sqrt 0.8714 1 0.5986 1.1442  
## temp 0.8385 1 0.6580 1.0190  
##   
## Likelihood ratio test that transformation parameters are equal to 0  
## (all log transformations)  
## LRT df pval  
## LR test, lambda = (0 0 0) 677.8528 3 < 2.22e-16  
##   
## Likelihood ratio test that no transformations are needed  
## LRT df pval  
## LR test, lambda = (1 1 1) 3.818622 3 0.28173

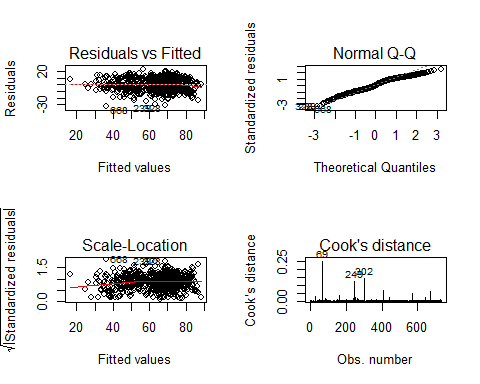
No futher transfermations are possible.

Thus we must consider a non-parallel model. After a lot of testing

test3 <- lm(y.sqrt ~ temp\* factor(weathersit)\*factor(season) + hum1 + windspeed.sqrt, data = day )  
summary(test3)

##   
## Call:  
## lm(formula = y.sqrt ~ temp \* factor(weathersit) \* factor(season) +   
## hum1 + windspeed.sqrt, data = day)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -31.829 -6.702 -0.208 7.925 22.185   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 49.6957 3.8600 12.874  
## temp 92.4099 8.0006 11.550  
## factor(weathersit)2 -0.6694 4.7210 -0.142  
## factor(weathersit)3 57.1681 19.2802 2.965  
## factor(season)2 25.9781 5.0637 5.130  
## factor(season)3 91.2291 9.1684 9.950  
## factor(season)4 18.0056 4.5489 3.958  
## hum1 -27.4089 3.6001 -7.613  
## windspeed.sqrt -25.8184 4.3060 -5.996  
## temp:factor(weathersit)2 2.9755 15.1669 0.196  
## temp:factor(weathersit)3 -269.7100 69.0138 -3.908  
## temp:factor(season)2 -48.5630 11.0007 -4.415  
## temp:factor(season)3 -145.2174 14.5518 -9.979  
## temp:factor(season)4 -20.4688 12.0006 -1.706  
## factor(weathersit)2:factor(season)2 -9.6724 8.1597 -1.185  
## factor(weathersit)3:factor(season)2 -37.0773 34.3558 -1.079  
## factor(weathersit)2:factor(season)3 -56.3957 16.7933 -3.358  
## factor(weathersit)3:factor(season)3 -40.1278 84.4093 -0.475  
## factor(weathersit)2:factor(season)4 13.3726 7.6570 1.746  
## factor(weathersit)3:factor(season)4 -96.2050 24.6115 -3.909  
## temp:factor(weathersit)2:factor(season)2 11.2554 19.4106 0.580  
## temp:factor(weathersit)3:factor(season)2 160.8610 103.0197 1.561  
## temp:factor(weathersit)2:factor(season)3 77.1379 27.8527 2.769  
## temp:factor(weathersit)3:factor(season)3 204.3366 154.6519 1.321  
## temp:factor(weathersit)2:factor(season)4 -35.5306 20.3789 -1.743  
## temp:factor(weathersit)3:factor(season)4 305.6192 76.6710 3.986  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## temp < 2e-16 \*\*\*  
## factor(weathersit)2 0.887282   
## factor(weathersit)3 0.003128 \*\*   
## factor(season)2 3.74e-07 \*\*\*  
## factor(season)3 < 2e-16 \*\*\*  
## factor(season)4 8.32e-05 \*\*\*  
## hum1 8.59e-14 \*\*\*  
## windspeed.sqrt 3.23e-09 \*\*\*  
## temp:factor(weathersit)2 0.844522   
## temp:factor(weathersit)3 0.000102 \*\*\*  
## temp:factor(season)2 1.17e-05 \*\*\*  
## temp:factor(season)3 < 2e-16 \*\*\*  
## temp:factor(season)4 0.088515 .   
## factor(weathersit)2:factor(season)2 0.236262   
## factor(weathersit)3:factor(season)2 0.280861   
## factor(weathersit)2:factor(season)3 0.000827 \*\*\*  
## factor(weathersit)3:factor(season)3 0.634652   
## factor(weathersit)2:factor(season)4 0.081165 .   
## factor(weathersit)3:factor(season)4 0.000102 \*\*\*  
## temp:factor(weathersit)2:factor(season)2 0.562195   
## temp:factor(weathersit)3:factor(season)2 0.118864   
## temp:factor(weathersit)2:factor(season)3 0.005762 \*\*   
## temp:factor(weathersit)3:factor(season)3 0.186841   
## temp:factor(weathersit)2:factor(season)4 0.081682 .   
## temp:factor(weathersit)3:factor(season)4 7.42e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.261 on 705 degrees of freedom  
## Multiple R-squared: 0.6664, Adjusted R-squared: 0.6546   
## F-statistic: 56.33 on 25 and 705 DF, p-value: < 2.2e-16

par(mfrow = c(2,2))   
for( i in 1:4){  
 plot(test3 , which = i )  
}



par(mfrow = c(1,1))

In addition, we will verify it is the best

anova(BIC.factor, test3 )

## Analysis of Variance Table  
##   
## Model 1: y.sqrt ~ temp + factor(weathersit) + factor(season) + windspeed.sqrt +   
## hum1  
## Model 2: y.sqrt ~ temp \* factor(weathersit) \* factor(season) + hum1 +   
## windspeed.sqrt  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 722 73031   
## 2 705 60466 17 12565 8.618 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Since we have such a small p-value for model2, we reject that the slope of non-parallel interactions are 0, arriving at our final model.