

Assignment 5: GAM

Víctor Villegas, Roger Llorenç, Luis Sierra

2024-03-26

```
# Clears plots
while (dev.cur() != 1) {
  dev.off()
}
# Clears global environment
rm(list = ls())

set.seed(1234)
```

Initial remark: double-blind trials

(Taken from this website)

In double-blind trials, the treatment patients have is unknown to both patients and doctors until after the study is concluded. This differs from other types of trials, such as simple blind trials where only the patients are unaware of the treatment they are receiving, whereas the doctors know.

Double-blind trials are a form of randomized trials and can be ‘upgraded’ to triple-blind trials, in which the statisticians or data clean-up personnel are also blind to treatments.

To be effective, it is generally recommended that double-blind trials include around 100-300 people. If treatments are highly effective, smaller numbers can be used but if only 30 or so patients are enrolled the study is unlikely to be beneficial.

The assignment of patients into treatments is typically done by computers, where the computer assigns each patient a code number and treatment group. The doctor and patients only know the code number to avoid bias, hence allowing the study to be double-blind.

GAMs for hirsutism data

In this report we’ll explore several GAM models (including semi-parametric models) explaining **FGm12** as a function of the variables that were measured at the beginning of the clinical trial and Treatment (treated as factor, which can be used as value of parameter by in function **s()**).

One of the advantages of GAM models is that they can overcome the curse of dimensionality.

Then we use functions **summary**, **plot**, **vis.gam** and **gam.check** to get an insight into the fitted models.

To do this, we will be using the following variables:

- **Treatment**, with values 0 to 3.
- **FGm0**, always greater than 15. Indicates the baseline hirsutism level at the randomization
- **SysPres**, baseline systolic blood pressure.
- **DiaPres**, baseline diastolic blood pressure.
- **weight**
- **height**

Note: The term “baseline“ means that these variables were measured at the beginning of the clinical trial

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.9-1. For overview type 'help("mgcv-package")'.
```

```
library(ggplot2)
```

```
library(car)
```

```
## Loading required package: carData
```

```
# data separated by tab, fill in missing values
```

```
hirsutism <- read.table("hirsutism.dat", header=T, sep="\t", fill=TRUE)
```

```
hirsutism <- na.omit(hirsutism)
```

```
# we're using the variable Treatment as a categorical predictor
```

```
hirsutism$Treatment <- as.factor(hirsutism$Treatment)
```

```
attach(hirsutism)
```

```
summary(hirsutism)
```

```
## Treatment      FGm0      FGm3      FGm6      FGm12
## 0:22      Min.   :14.57      Min.   : 4.381      Min.   : 1.786      Min.   : -1.163
## 1:22      1st Qu.:16.40      1st Qu.: 9.587      1st Qu.: 7.539      1st Qu.: 5.566
## 2:22      Median :17.70      Median :12.643      Median :10.286      Median : 8.069
## 3:25      Mean    :18.67      Mean    :13.129      Mean    :10.940      Mean    : 9.053
##           3rd Qu.:20.27      3rd Qu.:16.219      3rd Qu.:14.285      3rd Qu.:12.402
##           Max.    :28.36      Max.    :25.637      Max.    :23.411      Max.    :22.759
## SysPres      DiaPres      weight      height
## Min.   : 88.0      Min.   :46.00      Min.   : 41.00      Min.   :1.480
## 1st Qu.:110.0      1st Qu.:65.00      1st Qu.: 57.00      1st Qu.:1.580
## Median :115.0      Median :70.00      Median : 64.00      Median :1.610
## Mean    :115.9      Mean    :70.04      Mean    : 68.06      Mean    :1.613
## 3rd Qu.:120.0      3rd Qu.:75.00      3rd Qu.: 74.50      3rd Qu.:1.650
## Max.    :162.0      Max.    :95.00      Max.    :113.00      Max.    :1.800
```

0. Linear additive model

```
gam_0 <- gam(FGm12 ~
  (FGm0) + Treatment + (SysPres) +
  (DiaPres) + (weight) + (height),
  data = hirsutism)
summary(gam_0)
```

```
##
```

```
## Family: gaussian
```

```
## Link function: identity
```

```
##
```

```
## Formula:
```

```
## FGm12 ~ (FGm0) + Treatment + (SysPres) + (DiaPres) + (weight) +
```

```
## (height)
```

```
##
```

```
## Parametric coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 19.49686    14.85245    1.313 0.192945
```

```
## FGm0         0.59983     0.16862    3.557 0.000626 ***
```

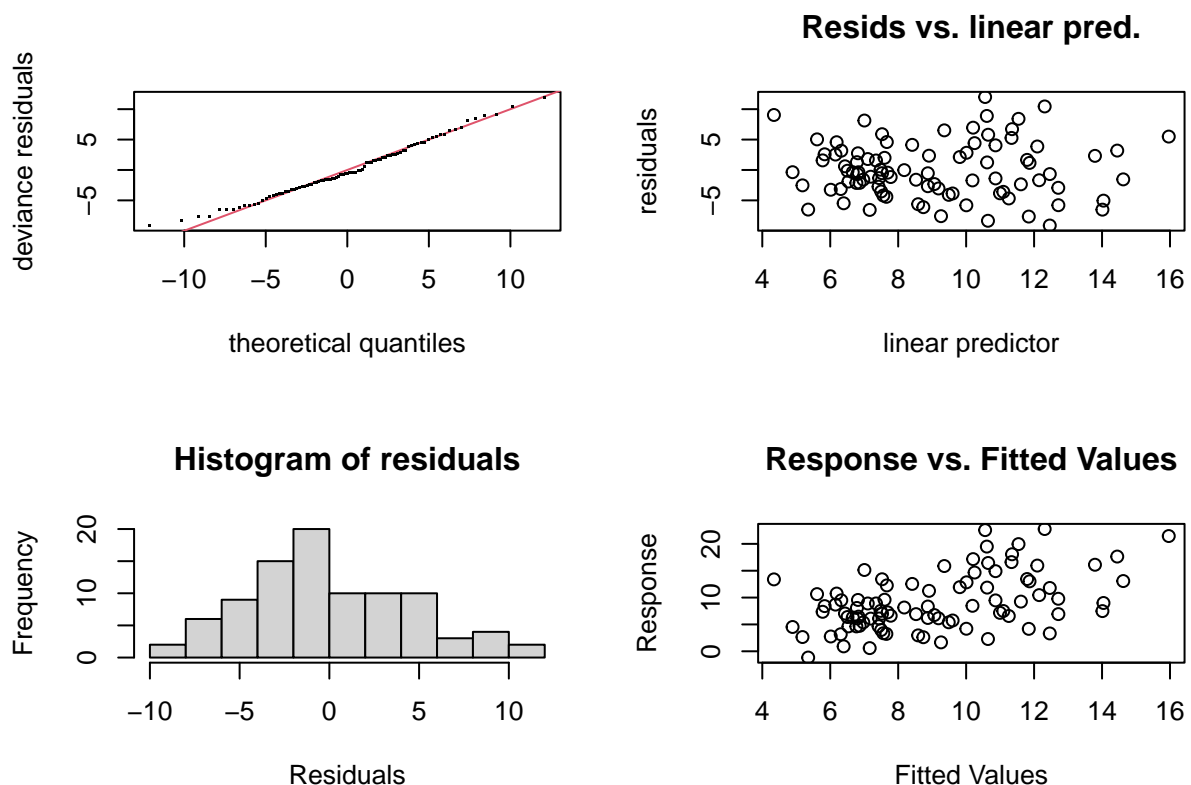
```
## Treatment1 -4.33022 1.48110 -2.924 0.004471 **
## Treatment2 -4.31441 1.49589 -2.884 0.005012 **
## Treatment3 -3.94666 1.44364 -2.734 0.007668 **
## SysPres -0.07570 0.05194 -1.458 0.148787
## DiaPres 0.03525 0.07115 0.495 0.621652
## weight 0.02768 0.04425 0.626 0.533308
## height -8.71024 9.08570 -0.959 0.340540
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) = 0.17 Deviance explained = 24.4%
## GCV = 25.139 Scale est. = 22.653 n = 91
```

- **p-value:** None of the following are statistically significant since their p-value is too high: SysPres, weight, height.
- **R-squared (adjusted):** The model explains around 17% of the variance in FGm12.
- **Deviance explained:** 24.4%.

The previous two values lead us to conclude that the explanatory potential of this model is limited but not null.

- **Number of observations:** notice that the gam functions removes the 8 observations with missing values. So n=91.
- **Visualization:**

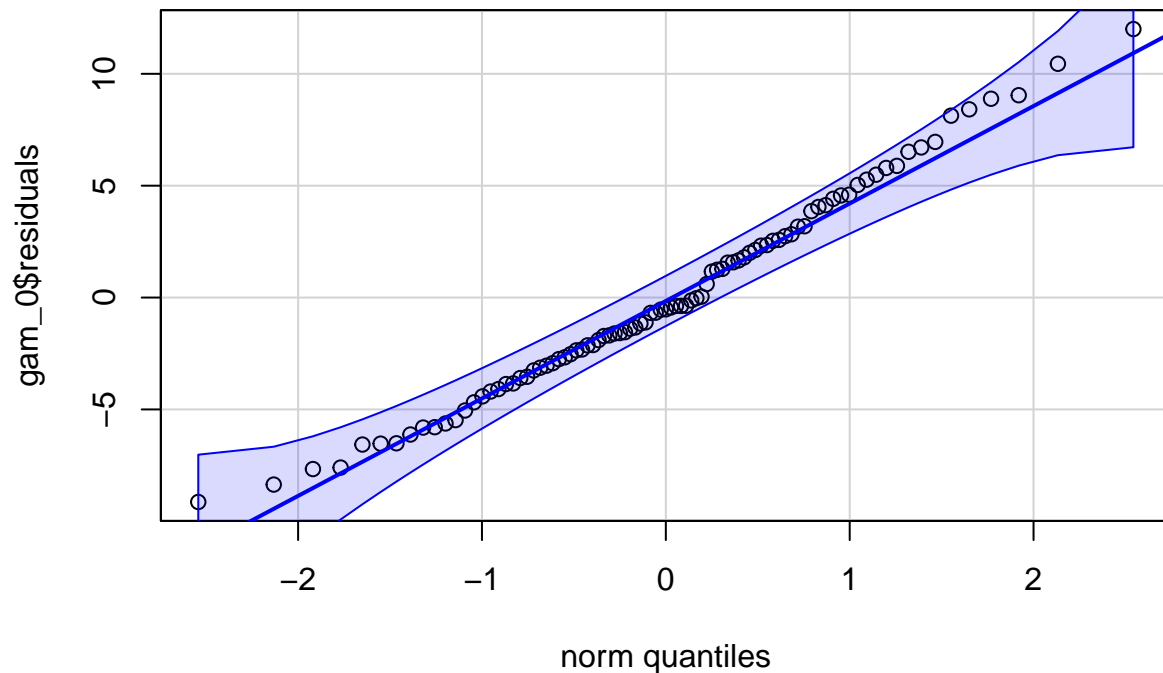
```
gam.check(gam_0)
```



```
##
## Method: GCV Optimizer: magic
```

```
## Model required no smoothing parameter selectionModel rank = 9 / 9
```

```
qqPlot(gam_0$residuals, id=FALSE)
```



1. Smoothing

- Variable Treatment is a factor, so it doesn't have to be smoothed.

```
gam_1 <- gam(FGm12 ~
  s(FGm0) + Treatment + s(SysPres) +
  s(DiaPres) + s(weight) + s(height),
  data = hirsutism)
summary(gam_1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0) + Treatment + s(SysPres) + s(DiaPres) + s(weight) +
##       s(height)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.477      0.985   12.667 < 2e-16 ***
## Treatment1    -4.868      1.405   -3.465 0.000876 ***
## Treatment2    -4.576      1.418   -3.227 0.001847 **
## Treatment3    -4.154      1.383   -3.005 0.003601 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(FGm0)       6.114  7.238  3.798 0.00143 **
```

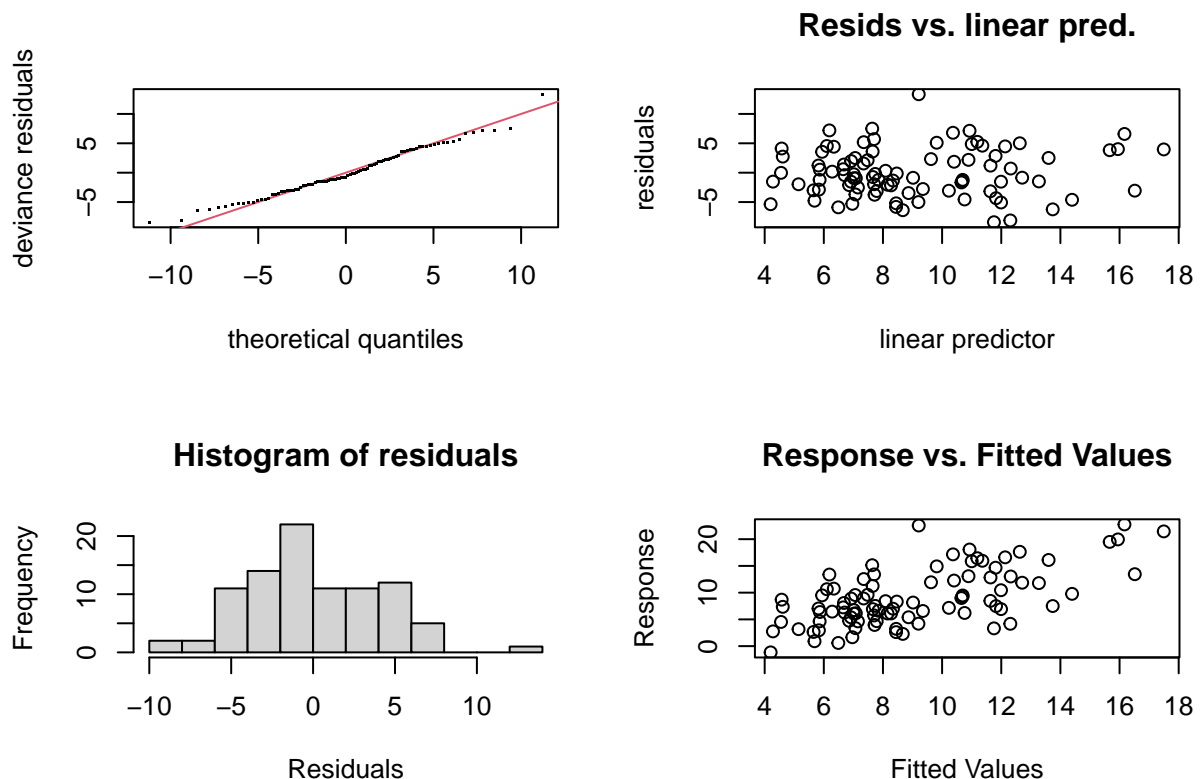
```
## s(SysPres) 1.000 1.000 3.015 0.08653 .
## s(DiaPres) 1.956 2.464 1.191 0.40264
## s(weight) 1.000 1.000 1.146 0.28770
## s(height) 1.000 1.000 1.454 0.23159
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.288 Deviance explained = 39.9%
## GCV = 23.31 Scale est. = 19.45 n = 91
```

- **p-value:** the meaning of the relevant values don't change.
- **R-squared (adjusted):** The model explains around 28.8% of the variance in FGm12.
- **Deviance explained:** 39.9%.

The explanatory potential of the model has improved a considerable amount with respect to the previous one.

- **Number of observations:** notice that the `gam` functions removes the 8 observations with missing values. So $n=91$.
- **Visualization:**

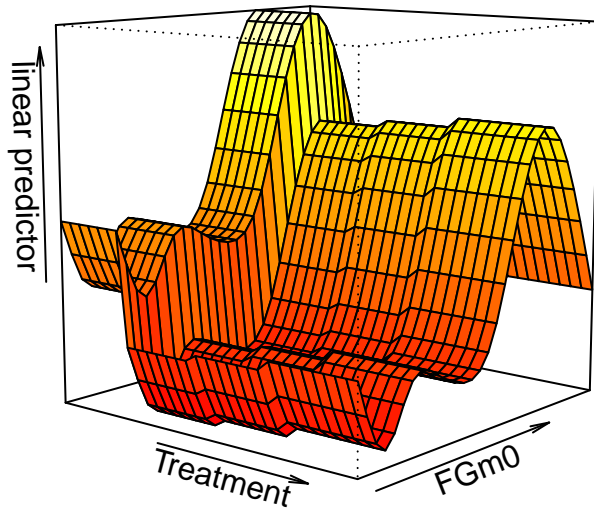
```
gam.check(gam_1)
```



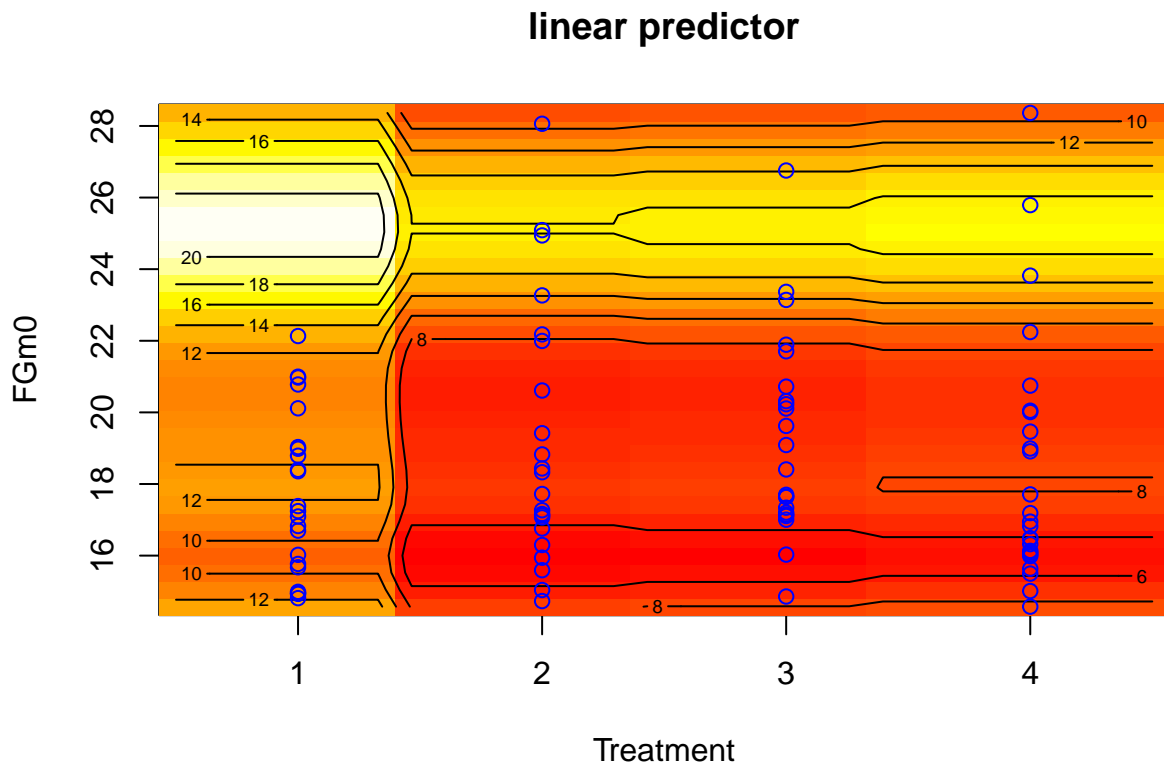
```
##
## Method: GCV Optimizer: magic
## Smoothing parameter selection converged after 22 iterations.
## The RMS GCV score gradient at convergence was 8.714651e-07 .
## The Hessian was positive definite.
## Model rank = 49 / 49
##
```

```
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##           k'   edf k-index p-value
## s(FGm0)    9.00 6.11   1.12  0.860
## s(SysPres) 9.00 1.00   1.09  0.760
## s(DiaPres) 9.00 1.96   1.06  0.690
## s(weight)  9.00 1.00   1.10  0.810
## s(height)  9.00 1.00   0.87  0.085 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

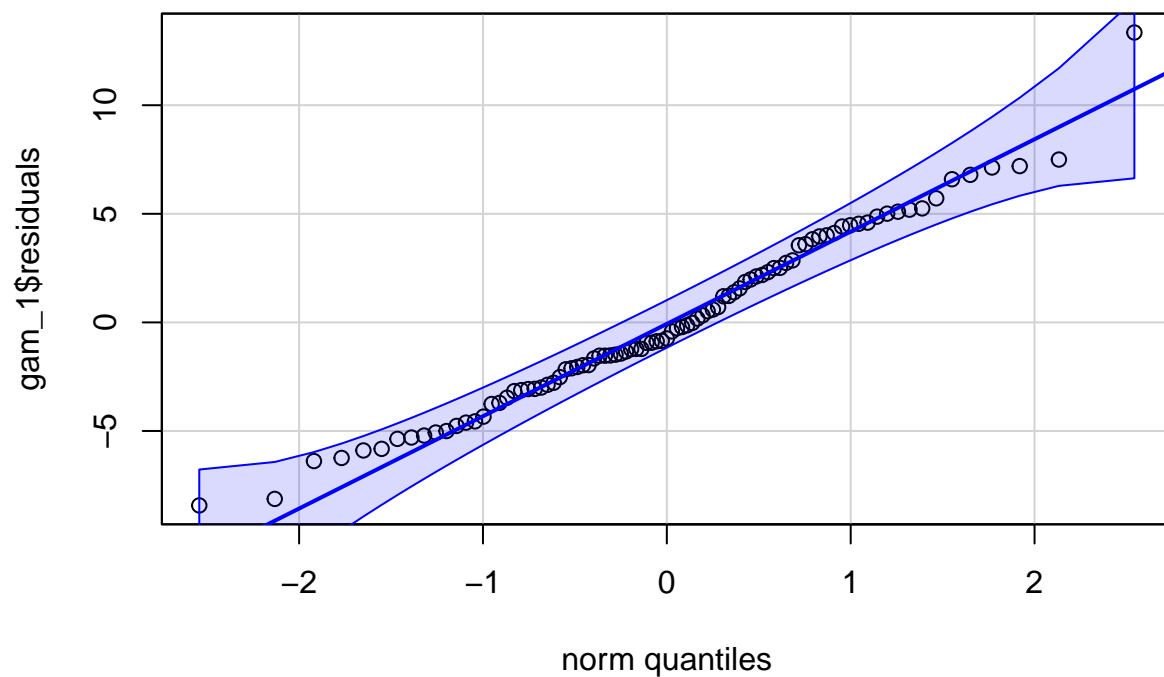
```
vis.gam(gam_1,se=0,theta =40, phi = 10, d=4,nticks=3)
```



```
vis.gam(gam_1,se=0,plot.type="contour",contour.col=1)
points(Treatment,FGm0,col="blue")
```



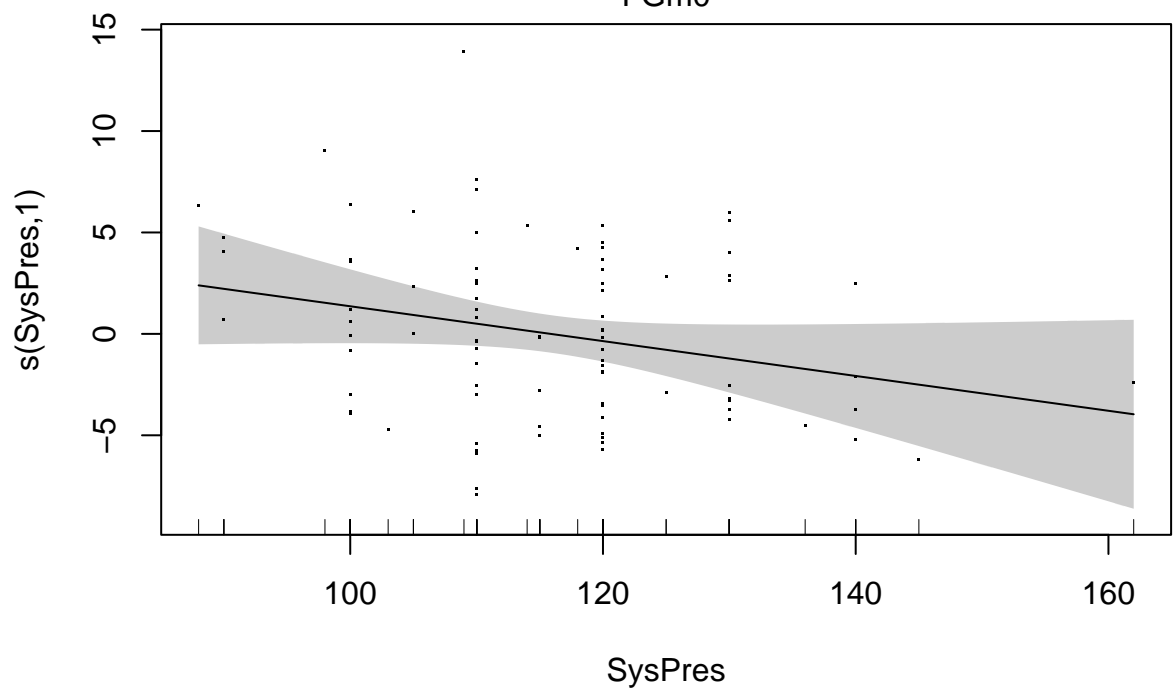
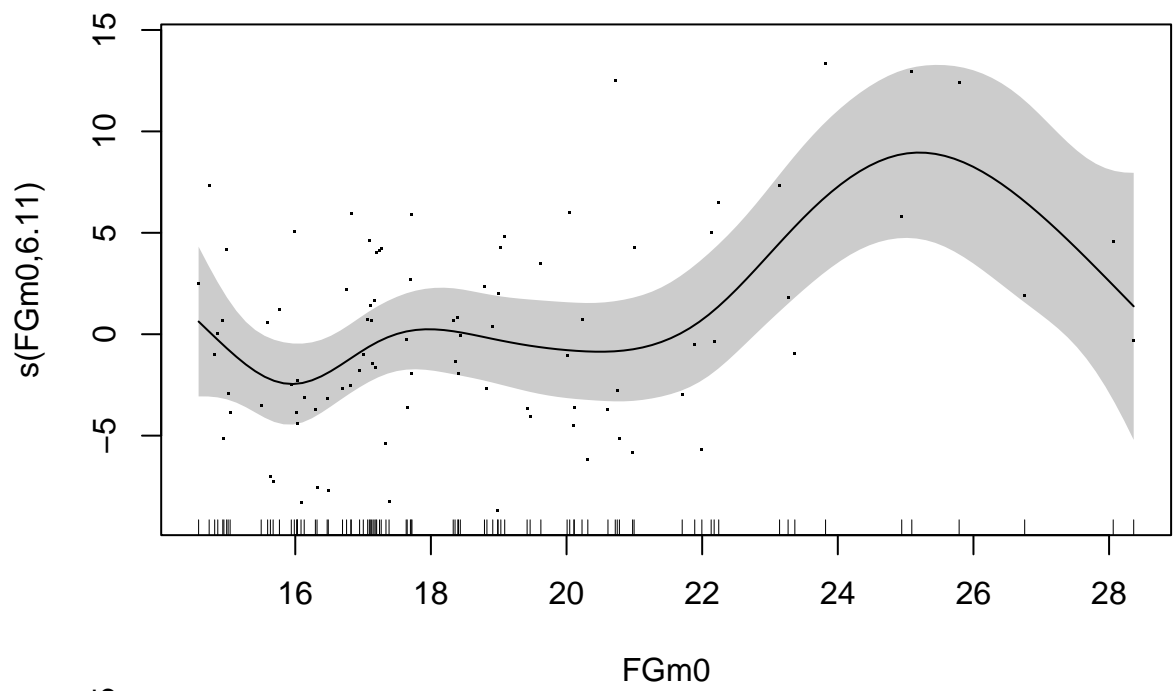
```
qqPlot(gam_1$residuals, id=FALSE)
```

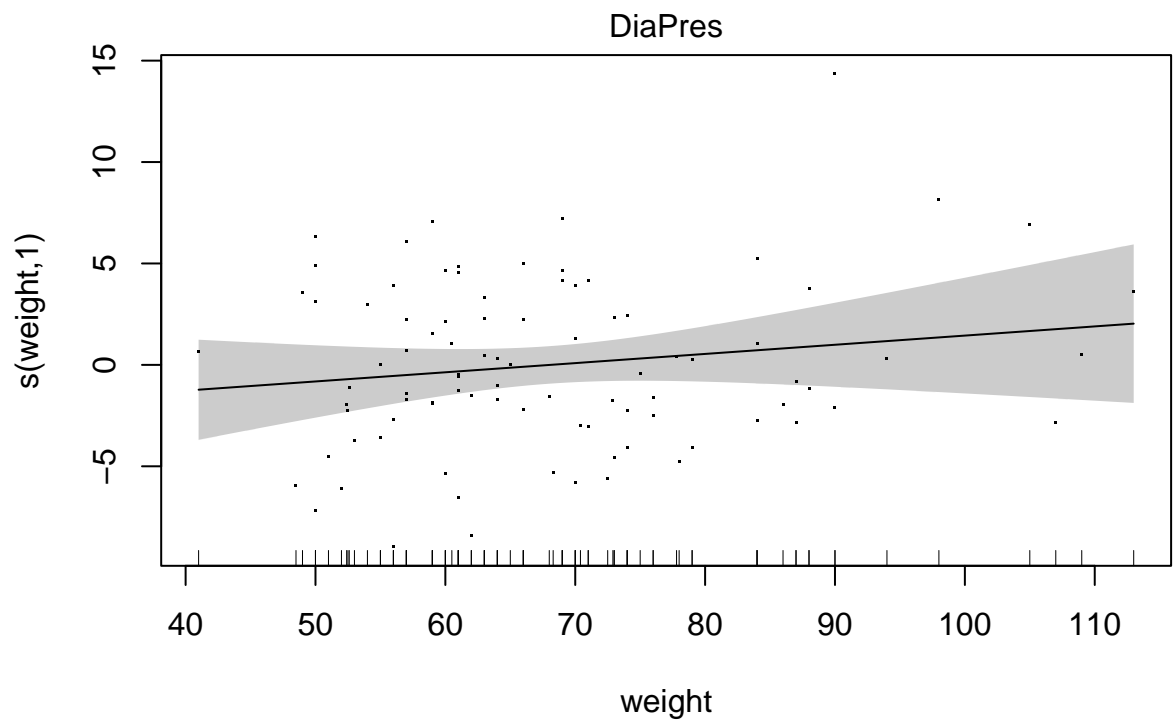
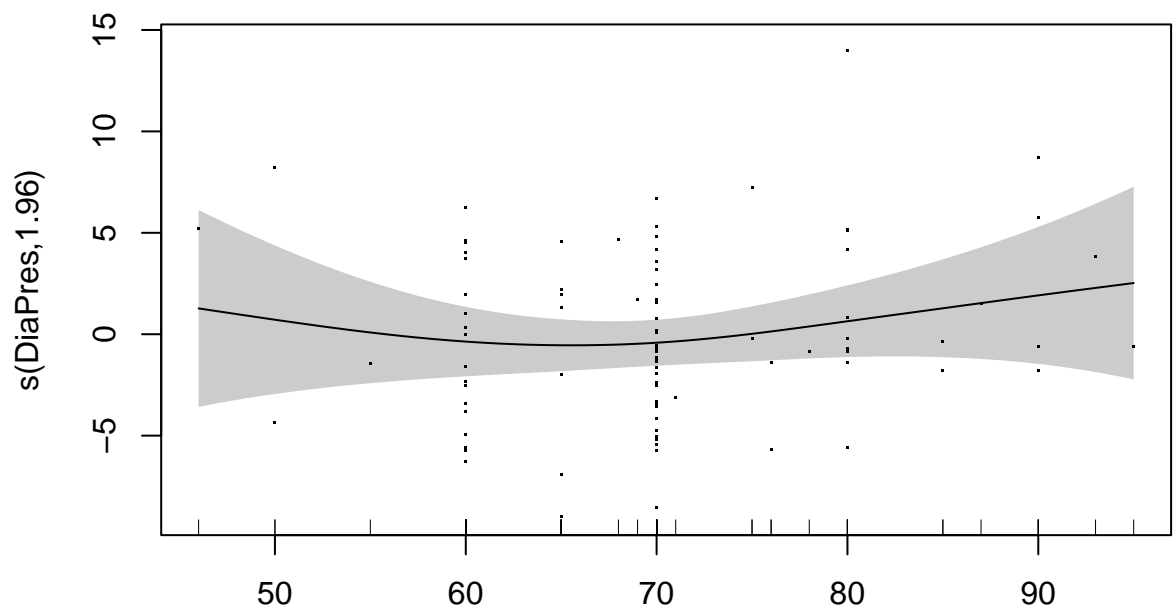


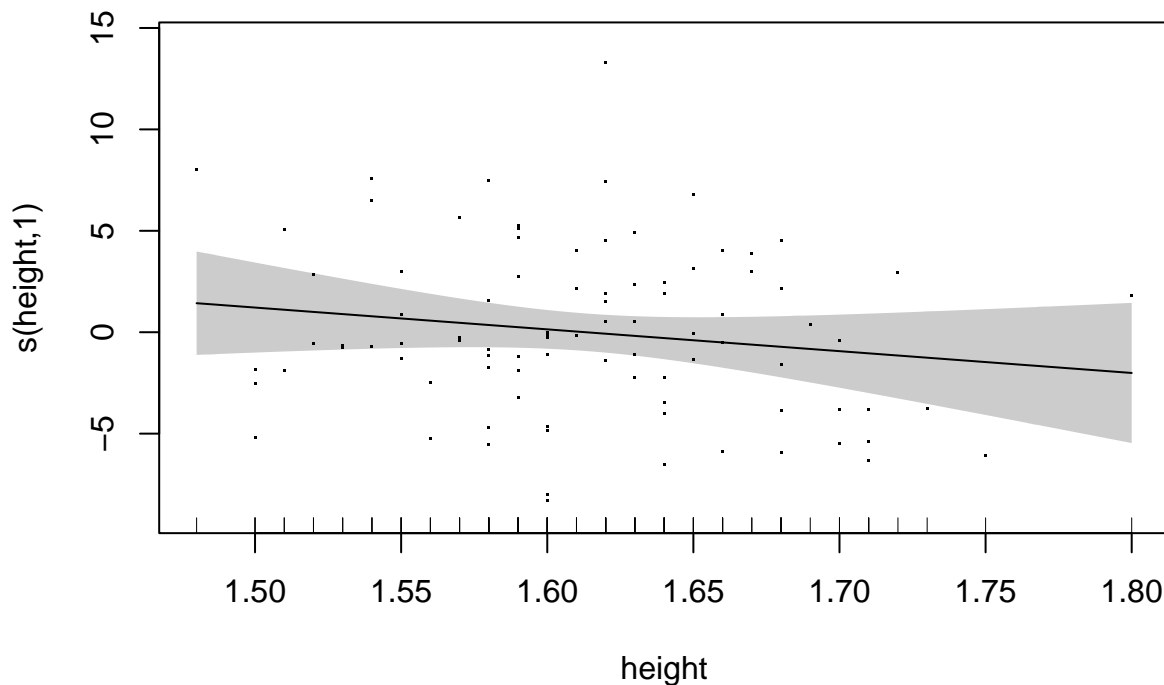
From the quantile-quantile plot, we may observe some evidence for heavy tails, which may indicate that the assumption of normality is too strong.

Individual effects of each explanatory variable:

```
plot(gam_1, residuals = TRUE, shade=TRUE, seWithMean=TRUE)
```







2. Joining variables

- We fit jointly weight and height by taking the tensor product as `te(weight,height)` in order to capture the physical effect.
- We fit jointly systolic and diastolic blood pressure by `te(SysPres,DiaPres)` in order to capture the physical effect.

```
gam_2 <- gam(FGm12 ~
  s(FGm0) + Treatment
  + te(SysPres,DiaPres) + te(weight,height),
  data = hirsutism)
summary(gam_2)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres) + te(weight,
##   height)
##
## Parametric coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.5280    0.9721  12.888 < 2e-16 ***
## Treatment1   -5.0660    1.3952  -3.631 0.000521 ***
## Treatment2   -4.4030    1.4012  -3.142 0.002422 **
## Treatment3   -4.3167    1.3777  -3.133 0.002488 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##               edf Ref.df    F  p-value
## s(FGm0)         6.357  7.428 4.007 0.000699 ***
```

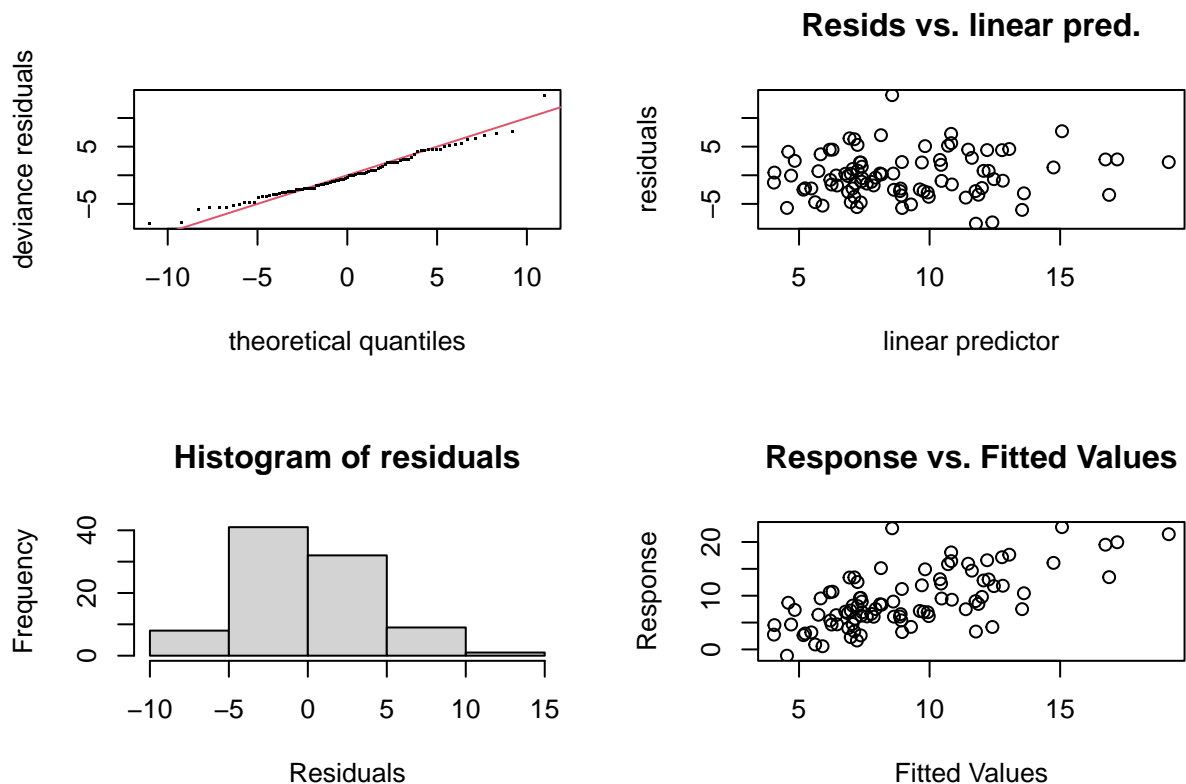
```
## te(SysPres,DiaPres) 4.031 4.660 1.481 0.230636
## te(weight,height) 3.570 3.987 1.013 0.403700
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.316 Deviance explained = 44.5%
## GCV = 23.279 Scale est. = 18.685 n = 91
```

- **p-value:** There hasn't been an improvement, the non-significant variables remain so.
- **R-squared (adjusted):** 31.6%
- **Deviance explained:** It has increased to 44.5%.

The increase of these two last values means an improvement in the explanatory potential of the model, that is, this model is better than the previous one.

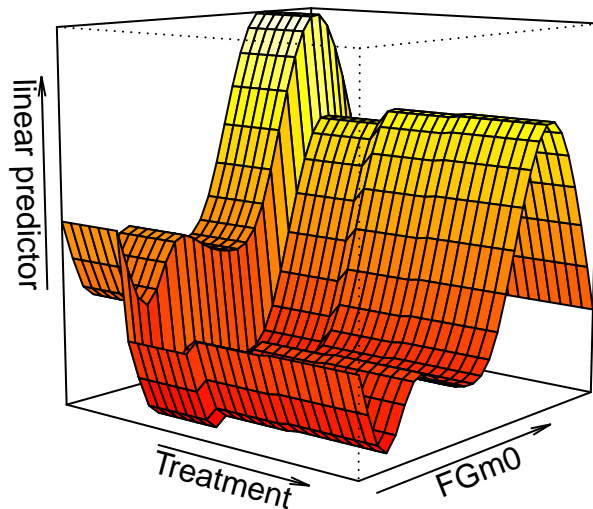
- **Visualization:**

```
gam.check(gam_2)
```

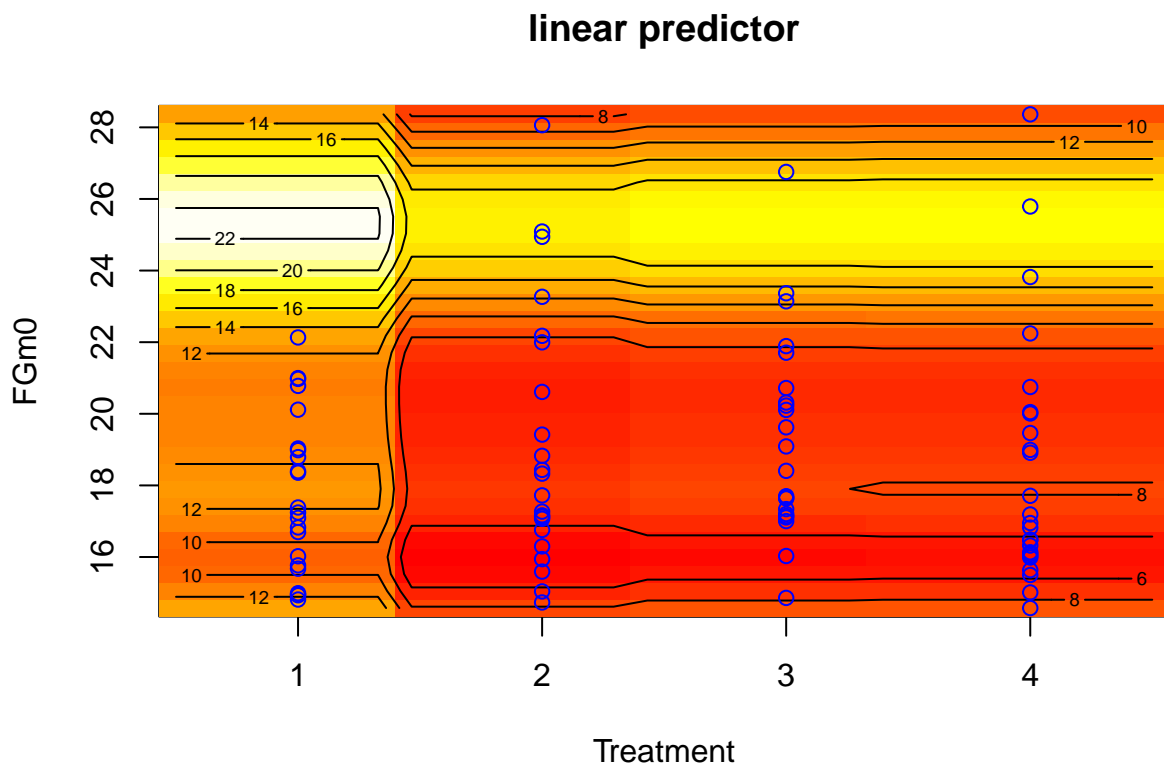


```
##
## Method: GCV Optimizer: magic
## Smoothing parameter selection converged after 30 iterations.
## The RMS GCV score gradient at convergence was 1.10955e-06 .
## The Hessian was positive definite.
## Model rank = 61 / 61
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
## k' edf k-index p-value
```

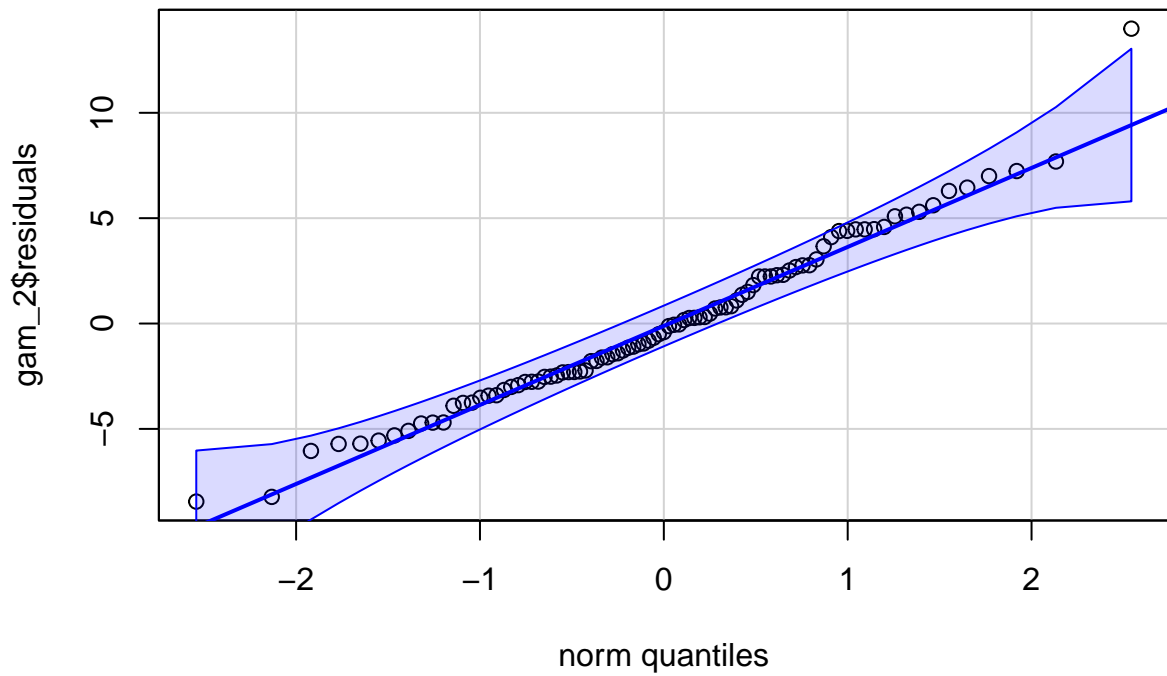
```
## s(FGm0)          9.00  6.36    1.1    0.69
## te(SysPres,DiaPres) 24.00  4.03    1.2    0.95
## te(weight,height)  24.00  3.57    0.9    0.14
vis.gam(gam_2,se=0,theta =40, phi = 10, d=4,nticks=3)
```



```
vis.gam(gam_2,se=0,plot.type="contour",contour.col=1)
points(Treatment,FGm0,col="blue")
```

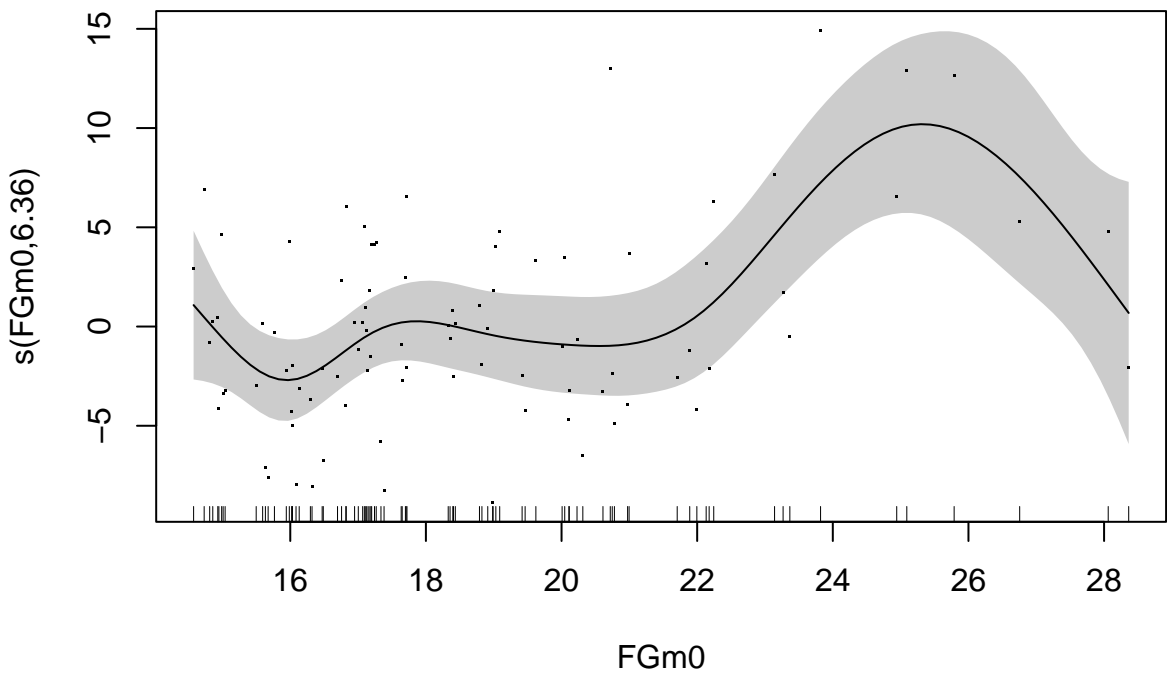


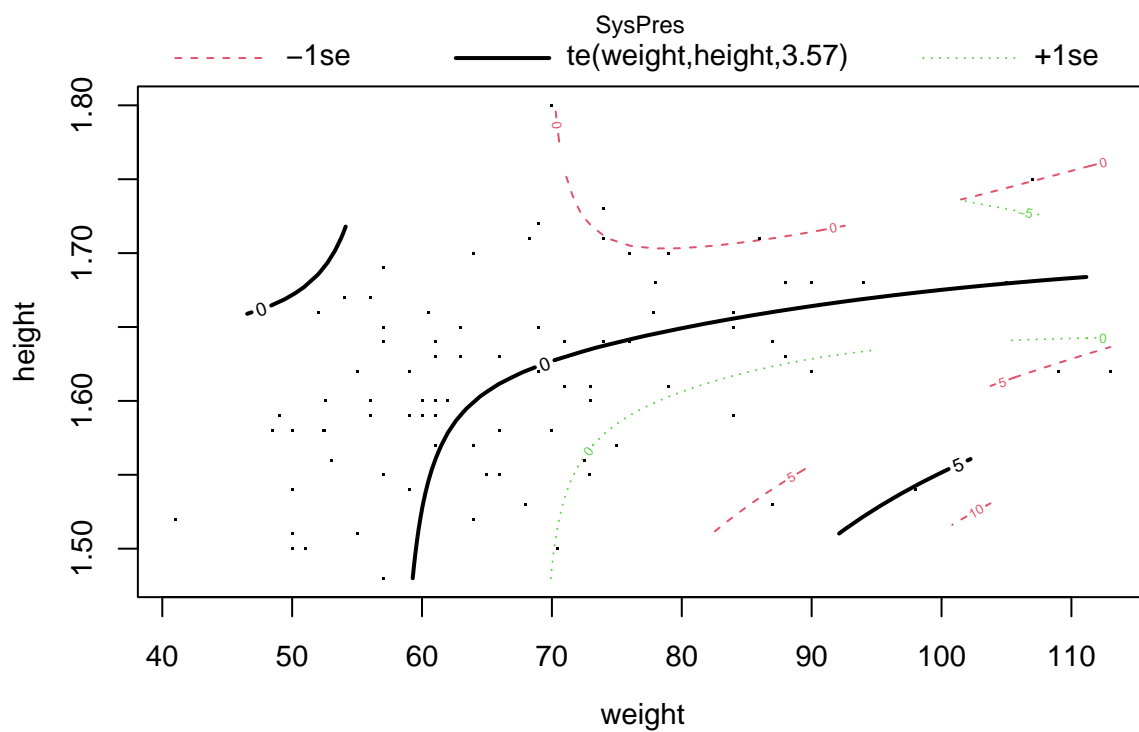
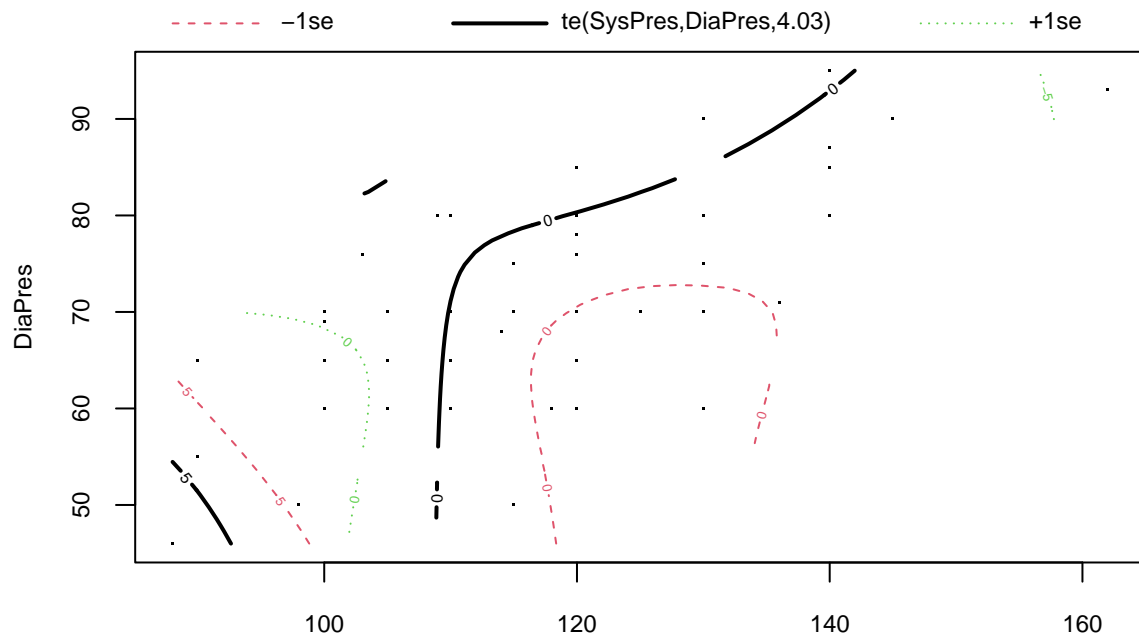
```
qqPlot(gam_2$residuals, id=FALSE)
```



Individual effects of each explanatory variable:

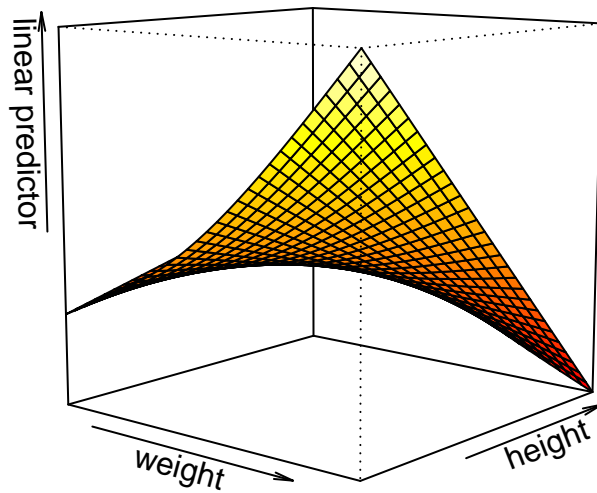
```
plot(gam_2, residuals = TRUE, shade=TRUE, seWithMean=TRUE)
```





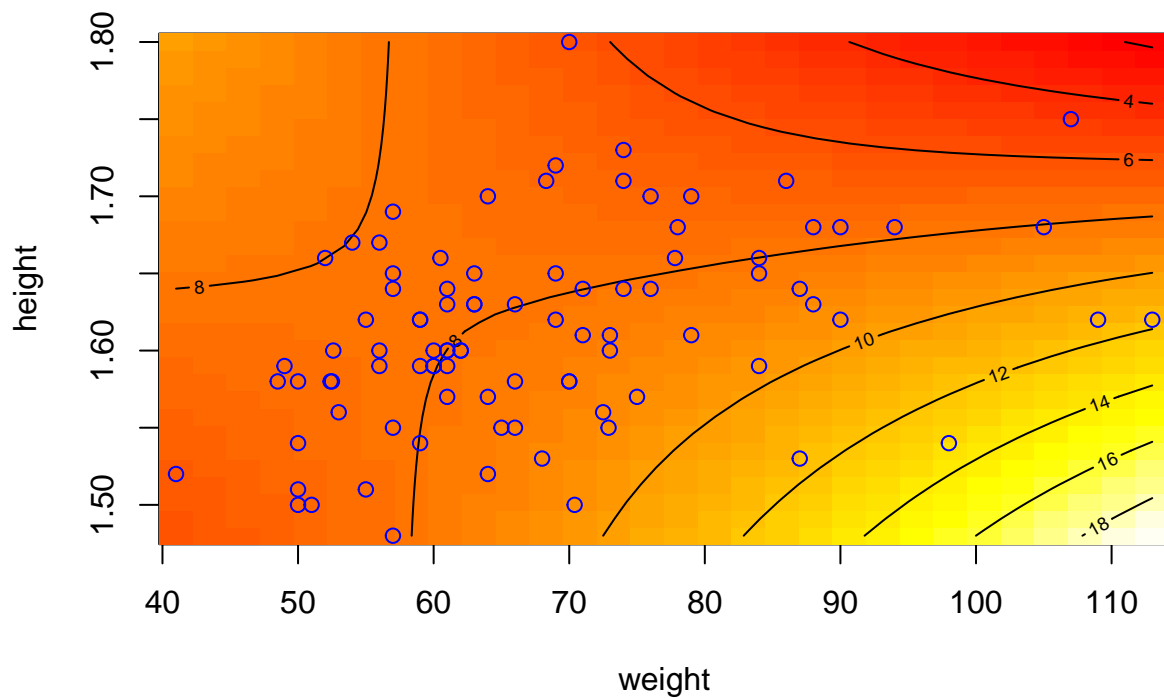
Joint effects of a pair of variables

```
vis.gam(gam_2, view=c("weight", "height"), plot.type = "persp", theta=40, phi=10, d=4, nticks=3)
```

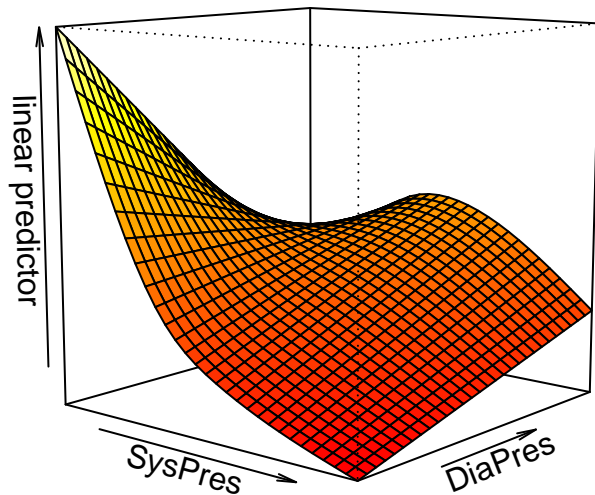


```
vis.gam(gam_2, view=c("weight","height"), plot.type = "contour",contour.col=1)
points(weight,height,col="blue")
```

linear predictor

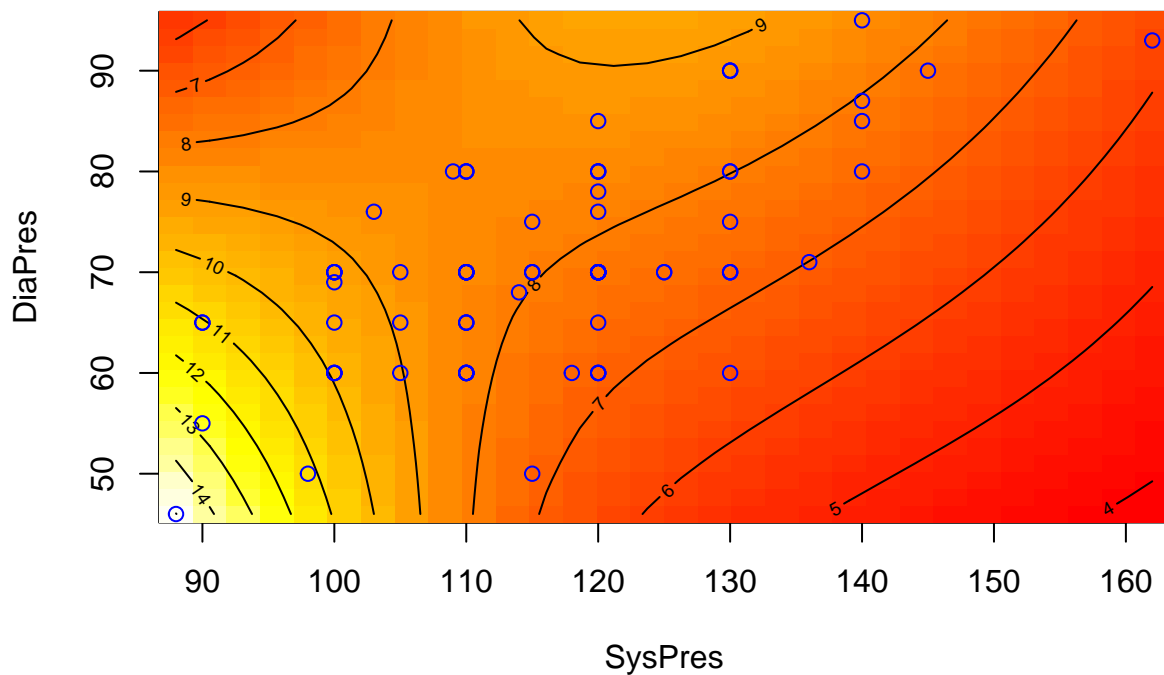


```
vis.gam(gam_2, view=c("SysPres","DiaPres"), plot.type = "persp", theta=40, phi=10,d=4,nticks=3)
```



```
vis.gam(gam_2, view=c("SysPres","DiaPres"), plot.type = "contour", contour.col=1)
points(SysPres,DiaPres,col="blue")
```

linear predictor



Refinements that can be done are:

- Certain variables could be removed from the model because the corresponding p-values are large.

3. Refining (2)

Now we remove non-significant variables from the model one at a time, in decreasing order of p-values, until all variables are significant.

```
gam_3a <- gam(FGm12 ~
  s(FGm0) + Treatment
  + te(SysPres,DiaPres),
```



```

      data = hirsutism)
summary(gam_3a)

##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres)
##
## Parametric coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3097     0.9688  12.706 < 2e-16 ***
## Treatment1   -5.0527     1.3938  -3.625 0.000518 ***
## Treatment2   -4.2882     1.3969  -3.070 0.002962 **
## Treatment3   -3.6348     1.3221  -2.749 0.007445 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df      F p-value
## s(FGm0)      6.372  7.485  3.816 0.00134 **
## te(SysPres,DiaPres) 4.007  4.650  1.505 0.18845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.305   Deviance explained = 40.8%
## GCV = 22.546   Scale est. = 18.984      n = 91
gam_3 <- gam(FGm12 ~
             s(FGm0) + (Treatment),
             data = hirsutism)
summary(gam_3)

```

```

##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0) + (Treatment)
##
## Parametric coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3681     0.9808  12.610 < 2e-16 ***
## Treatment1   -5.0794     1.3986  -3.632 0.000492 ***
## Treatment2   -4.5832     1.3969  -3.281 0.001526 **
## Treatment3   -3.5641     1.3483  -2.643 0.009847 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df      F p-value
## s(FGm0)      5.763  6.892  3.999 0.00102 **
## ---

```

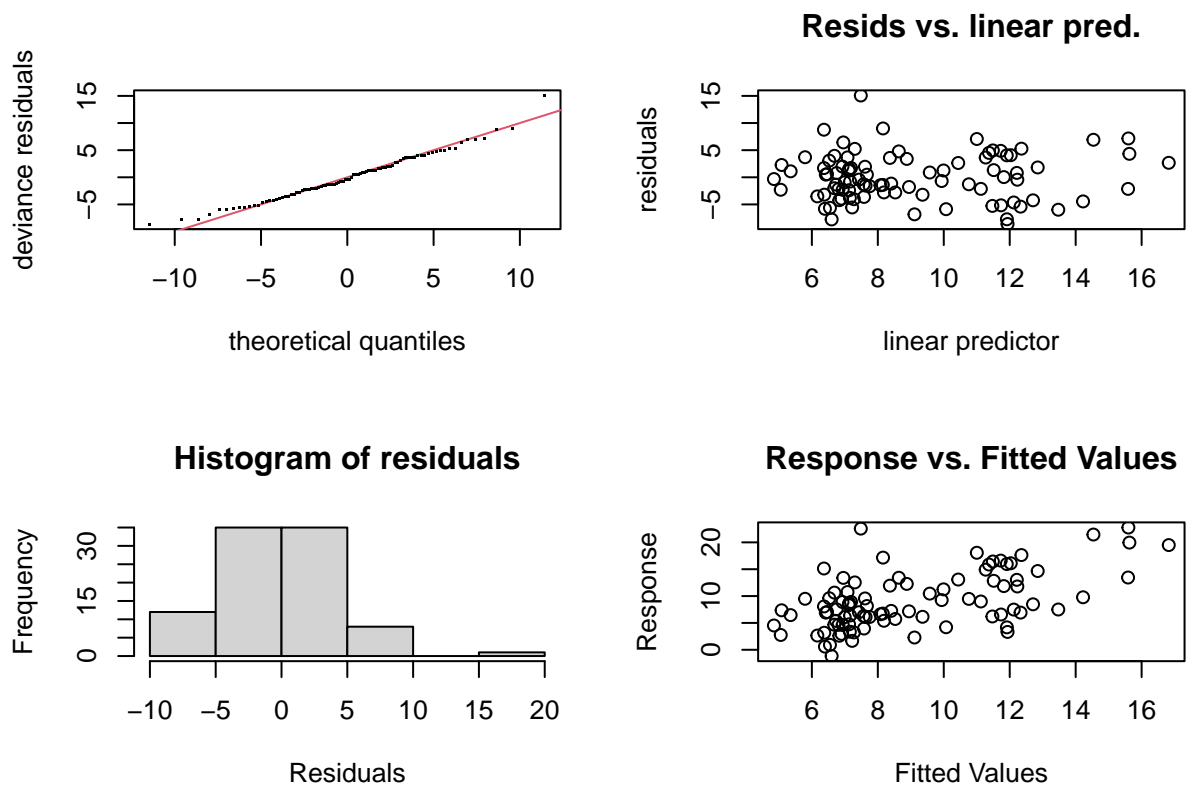
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.259   Deviance explained = 33.1%
## GCV = 22.667   Scale est. = 20.235    n = 91
```

- **p-value:** the only variable left is FGm0.
- **R-squared (adjusted):** 27,7%
- **Deviance explained:** 34,1%.

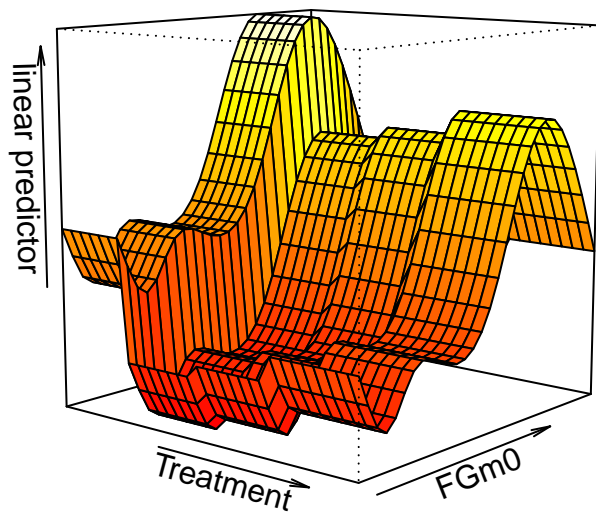
The increase of these two last values means an improvement in the explanatory potential of the model, that is, this model looks better than the previous one.

- **Visualization:**

```
gam.check(gam_3)
```

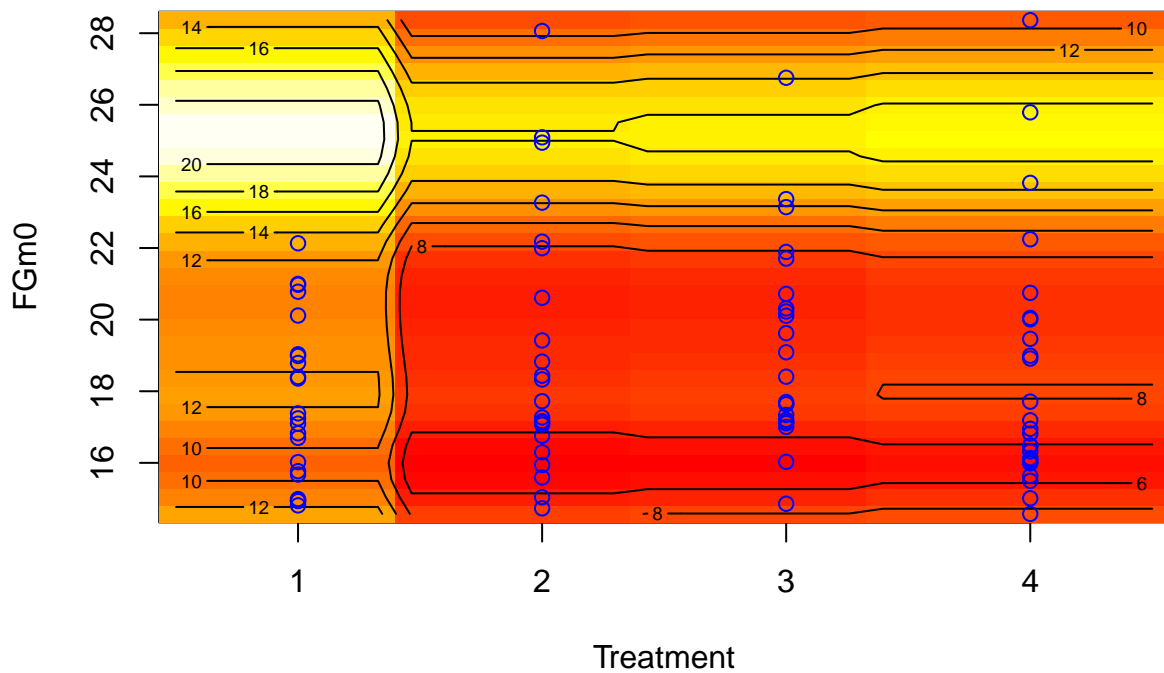


```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 6 iterations.
## The RMS GCV score gradient at convergence was 0.0001388711 .
## The Hessian was positive definite.
## Model rank =  13 / 13
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##           k'   edf k-index p-value
## s(FGm0)  9.00  5.76    1.06   0.65
vis.gam(gam_3, se=0, theta = 40, phi = 10, d=4,nticks=3)
```

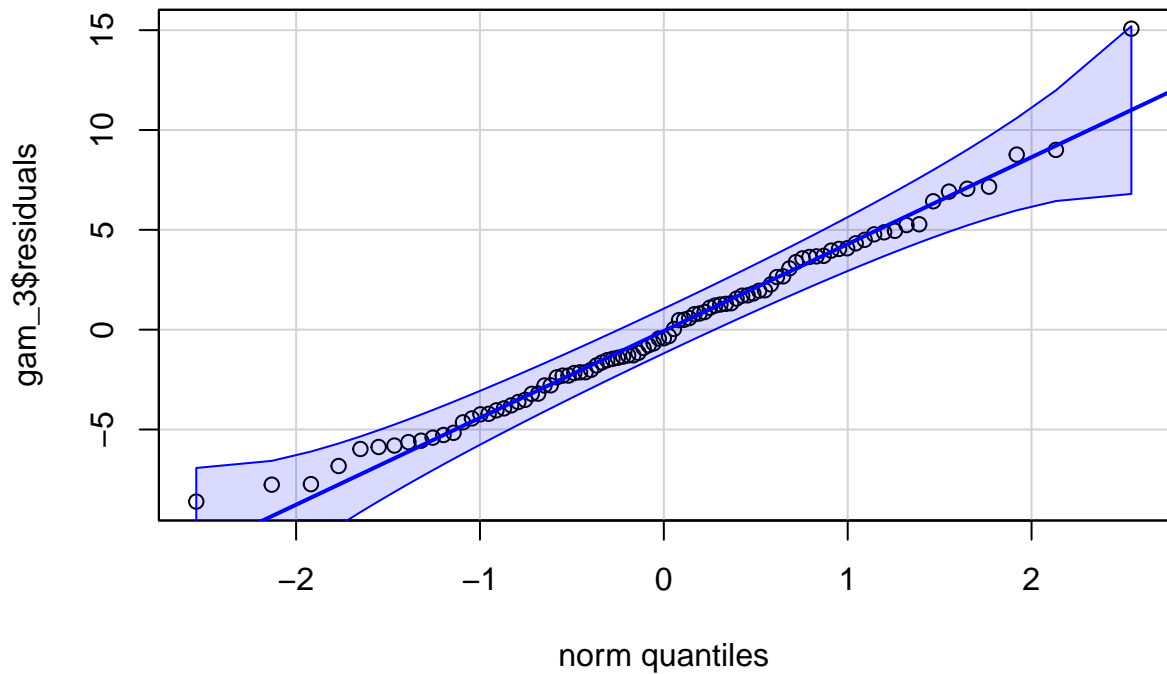


```
vis.gam(gam_1, se=0, plot.type="contour", contour.col=1)
points(Treatment, FGm0, col="blue")
```

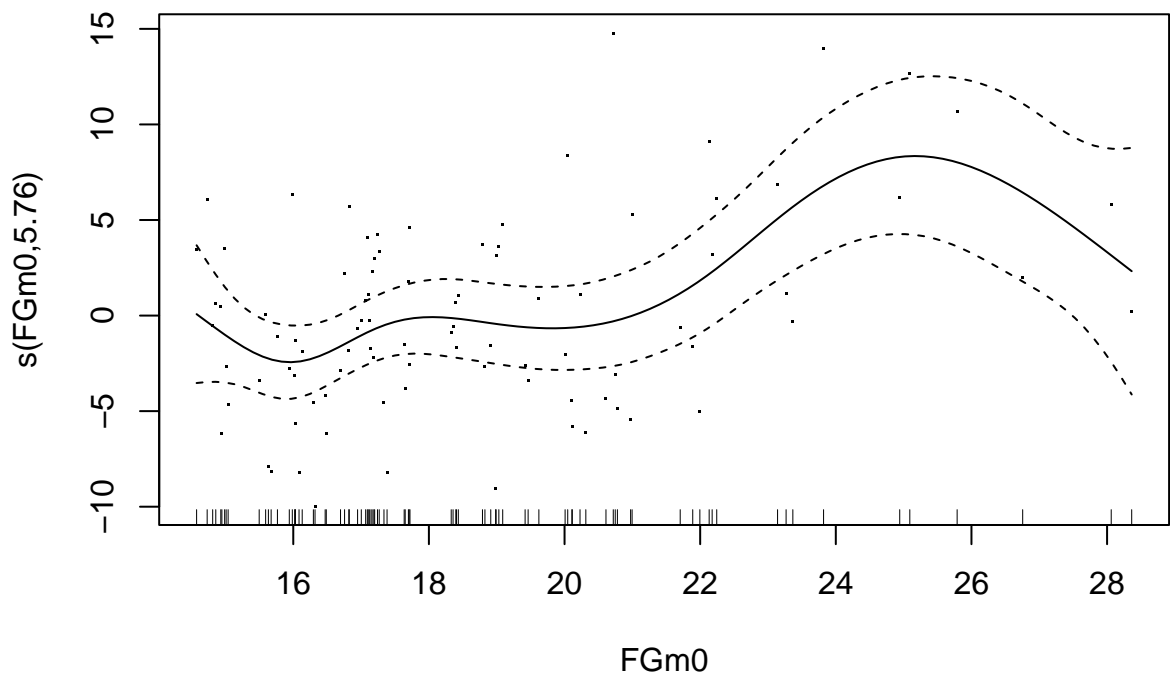
linear predictor



```
qqPlot(gam_3$residuals, id=FALSE)
```



```
plot(gam_3, residuals = TRUE, seWithMean=TRUE)
```



4. Treatment as a parameter of `s()`

Now we separate the continuous variables according to the treatment group.

```
gam_1b <- gam(FGm12 ~
  s(FGm0, by=Treatment) + s(SysPres, by=Treatment) +
  s(DiaPres, by=Treatment) + s(weight, by=Treatment) + s(height, by=Treatment),
  data = hirsutism)
summary(gam_1b)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0, by = Treatment) + s(SysPres, by = Treatment) +
##       s(DiaPres, by = Treatment) + s(weight, by = Treatment) +
##       s(height, by = Treatment)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.5764      0.3142   24.11 1.32e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df      F  p-value
## s(FGm0):Treatment0  7.948  7.962  25.788 < 2e-16 ***
## s(FGm0):Treatment1  8.694  8.889  32.609 < 2e-16 ***
## s(FGm0):Treatment2  1.000  1.000   0.564 0.463070
## s(FGm0):Treatment3  1.000  1.000 260.291 < 2e-16 ***
## s(SysPres):Treatment0 1.000  1.000   0.107 0.747020
## s(SysPres):Treatment1 1.000  1.000  19.500 0.000377 ***
## s(SysPres):Treatment2 4.458  4.929  18.004 6.52e-06 ***
## s(SysPres):Treatment3 7.835  7.918  16.471 2.39e-06 ***
## s(DiaPres):Treatment0 4.372  4.458   4.258 0.088535 .
## s(DiaPres):Treatment1 4.790  5.442   3.164 0.025008 *
## s(DiaPres):Treatment2 1.092  1.144  14.564 0.003574 **
## s(DiaPres):Treatment3 5.816  6.017  26.818 < 2e-16 ***
## s(weight):Treatment0  1.000  1.000   1.670 0.213520
## s(weight):Treatment1  1.000  1.000  29.318 4.52e-05 ***
## s(weight):Treatment2  8.545  8.742  21.293 < 2e-16 ***
## s(weight):Treatment3  1.000  1.000  48.935 1.04e-06 ***
## s(height):Treatment0  6.349  6.558  27.461 8.08e-07 ***
## s(height):Treatment1  1.342  1.544   5.964 0.020313 *
## s(height):Treatment2  3.726  4.209   2.007 0.157563
## s(height):Treatment3  1.000  1.000 198.946 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Rank: 180/181
## R-sq.(adj) =  0.971   Deviance explained = 99.5%
## GCV = 4.1831   Scale est. = 0.78303   n = 91
```

- **p-value:** almost all variables all relevant.
- **R-squared (adjusted):** 97.1%
- **Deviance explained:** 99.5%.

```
gam_2b <- gam(FGm12 ~
  s(FGm0, by=Treatment)
  + te(SysPres,DiaPres, by=Treatment) + te(weight,height, by=Treatment),
  data = hirsutism)
summary(gam_2b)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## FGm12 ~ s(FGm0, by = Treatment) + te(SysPres, DiaPres, by = Treatment) +
##      te(weight, height, by = Treatment)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.9658     0.2963   30.26   0.998
##
## Approximate significance of smooth terms:
##              edf Ref.df      F p-value
## s(FGm0):Treatment0    3.567  3.567    11.149 0.21916
## s(FGm0):Treatment1    8.706  8.706  7841.847 0.00984 **
## s(FGm0):Treatment2    1.000  1.000    1.984 0.39308
## s(FGm0):Treatment3    7.260  7.260   166.868 0.06220 .
## te(SysPres,DiaPres):Treatment0  8.438  8.439    56.856 0.10096
## te(SysPres,DiaPres):Treatment1  5.144  5.144    13.593 0.20020
## te(SysPres,DiaPres):Treatment2 11.724 11.724 44961.654 0.00410 **
## te(SysPres,DiaPres):Treatment3  9.884  9.884   152.141 0.05854 .
## te(weight,height):Treatment0    9.991  9.991    19.420 0.17577
## te(weight,height):Treatment1    7.763  7.763    51.454 0.10797
## te(weight,height):Treatment2    8.870  8.870 55157.761 0.00355 **
## te(weight,height):Treatment3    7.653  7.653   132.998 0.06670 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =      1    Deviance explained = 100%
## GCV = 2.6393  Scale est. = 7.4463e-06  n = 91
```

- **p-value:** almost all variables all relevant.
- **R-squared (adjusted):** 100%
- **Deviance explained:** 100%.

ANOVA type tests

Now we test the null hypothesis that states one model is correct against the alternative that another model is better. To do so, we will compare the previous models among themselves. Recall `gam_3.2` was our simplest model.

```
anova(gam_3, gam_3a, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + (Treatment)
## Model 2: FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1     80.108     1643.8
## 2     74.866     1454.6 5.242   189.29 1.9022 0.1008
```

```
# accept null and gam_3 = gam_3a
```

```
anova(gam_3a, gam_2, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres)
## Model 2: FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres) + te(weight,
## height)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1    74.866    1454.6
## 2    70.925    1364.8 3.9411   89.772 1.2191 0.3105
```

```
# accept null gam_3a = gam_2
```

```
anova(gam_3, gam_2, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + (Treatment)
## Model 2: FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres) + te(weight,
## height)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1    80.108    1643.8
## 2    70.925    1364.8 9.1831   279.06 1.6264 0.1228
```

```
# accept null gam_3 = gam_2
```

```
anova(gam_1, gam_2, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + Treatment + s(SysPres) + s(DiaPres) + s(weight) +
## s(height)
## Model 2: FGm12 ~ s(FGm0) + Treatment + te(SysPres, DiaPres) + te(weight,
## height)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1    74.297    1476.9
## 2    70.925    1364.8 3.3728   112.08 1.7784 0.1528
```

```
anova(gam_1, gam_3, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + Treatment + s(SysPres) + s(DiaPres) + s(weight) +
## s(height)
## Model 2: FGm12 ~ s(FGm0) + (Treatment)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1    74.297    1476.9
## 2    80.108    1643.8 -5.8104  -166.99 1.4776 0.1994
```

```
anova(gam_3, gam_1b, test="F")
```

```
## Analysis of Deviance Table
##
## Model 1: FGm12 ~ s(FGm0) + (Treatment)
## Model 2: FGm12 ~ s(FGm0, by = Treatment) + s(SysPres, by = Treatment) +
## s(DiaPres, by = Treatment) + s(weight, by = Treatment) +
## s(height, by = Treatment)
```

```
##   Resid. Df Resid. Dev    Df Deviance      F      Pr(>F)
## 1    80.108    1643.85
## 2    14.187      13.34 65.92   1630.5 31.588 5.999e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(gam_1b, gam_2b, test="F")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: FGm12 ~ s(FGm0, by = Treatment) + s(SysPres, by = Treatment) +
```

```
##   s(DiaPres, by = Treatment) + s(weight, by = Treatment) +
```

```
##   s(height, by = Treatment)
```

```
## Model 2: FGm12 ~ s(FGm0, by = Treatment) + te(SysPres, DiaPres, by = Treatment) +
```

```
##   te(weight, height, by = Treatment)
```

```
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
```

```
## 1 1.4187e+01      13.338
```

```
## 2 1.5325e-07      0.000 14.187   13.338 126256      1
```

So model gam_1b would be our preferred option.