MP_handendess_reanalysis

G Sala

2022-05-03

```
library(gamlss)
library(mgcv)
library(tidymv)
library(dplyr)
library(ggplot2)

# loading the dataset stored in github
base_address <- "https://raw.githubusercontent.com/SG540/data_science_portfolio"
specific_address <- "/main/GAM_handedness/data_handedness.csv"
data <- read.csv(paste0(base_address, specific_address))

# setting seed
set.seed(1892)

# establishing data types
data$gender <- as.factor(data$gender)
data$school <- as.factor(data$school)
data$h <- as.numeric(data$h)
data$spat <- as.numeric(data$spat)
data$math <- as.numeric(data$math)


data <- data[sample(nrow(data)), ] # shuffling rows

# creating groups for descriptive statistics
data$h_group <- cut(data$h, c(-1, -.95, -.30, .30, .95, 1), include.lowest = TRUE)

# subsetting the datasets by study
dat_1 <- subset(data, Exp == "Exp1") %>% dplyr::select(-c(spat))
dat_2 <- subset(data, Exp == "Exp2") %>% dplyr::select(-c(spat))
dat_3 <- subset(data, Exp == "Exp3") %>% dplyr::select(-c(school))
dat_4 <- subset(data, Exp == "Exp4")
dat_5 <- subset(data, Exp == "Exp5")


# inverting math scores in Studies 3 and 4
dat_3$math_inv <- abs(dat_3$math - max(dat_3$math))
dat_4$math_inv <- abs(dat_4$math - max(dat_4$math))
```

```
# percentage of left-handers
length(data$h[data$h < 0]) / length(data$h)
```

```
## [1] 0.1032844
```

```r
length(dat_1$h[dat_1$h < 0]) / length(dat_1$h)
```

```
## [1] 0.1089588
```

```r
length(dat_2$h[dat_2$h < 0]) / length(dat_2$h)
```

```
## [1] 0.1233333
```

```r
length(dat_3$h[dat_3$h < 0]) / length(dat_3$h)
```

```
## [1] 0.09259259
```

```r
length(dat_4$h[dat_4$h < 0]) / length(dat_4$h)
```

```
## [1] 0.1102757
```

```r
length(dat_5$h[dat_5$h < 0]) / length(dat_5$h)
```

```
## [1] 0.08424337
```

```r
# prevalence sorted by handedness groups and study
sapply(group_split(data %>%
  group_by(h_group, Exp)), function(x) nrow(x))
```

```
##  [1]   6   8   4  25   7  30  24   8  49  29  37  21  13  58  50 245 148 103 372
## [20] 404  95  99  34 294 151
```

```r
# percentage of left-handers sorted by gender

perc_h <- function(x) {aggregate(h ~ gender, data = subset(x, h < 0),
                                 FUN = length)[,2] /
                        aggregate(h ~ gender, data = x, FUN = length)[,2]}

perc_h(data)
```

```
## [1] 0.08346457 0.12739464
```

```r
perc_h(dat_1)
```

```
## [1] 0.08849558 0.13368984
```

```r
perc_h(dat_2)
```

```
## [1] 0.09395973 0.15231788
```

```r
perc_h(dat_3)
```

```
## [1] 0.08333333 0.10256410
```

```r
perc_h(dat_4)
```

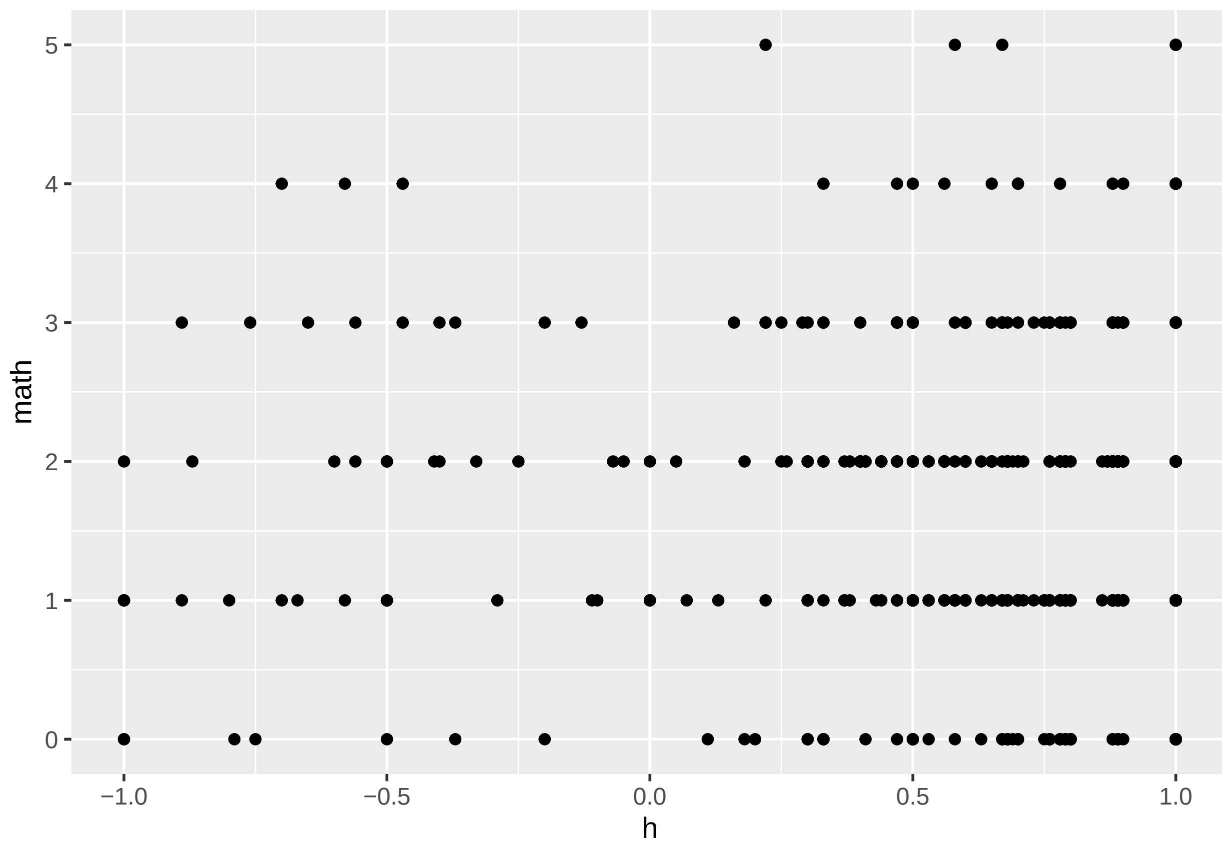
```
## [1] 0.09186352 0.12709832
```

```r
perc_h(dat_5)
```

```
## [1] 0.06976744 0.11374408
```

#Study 1

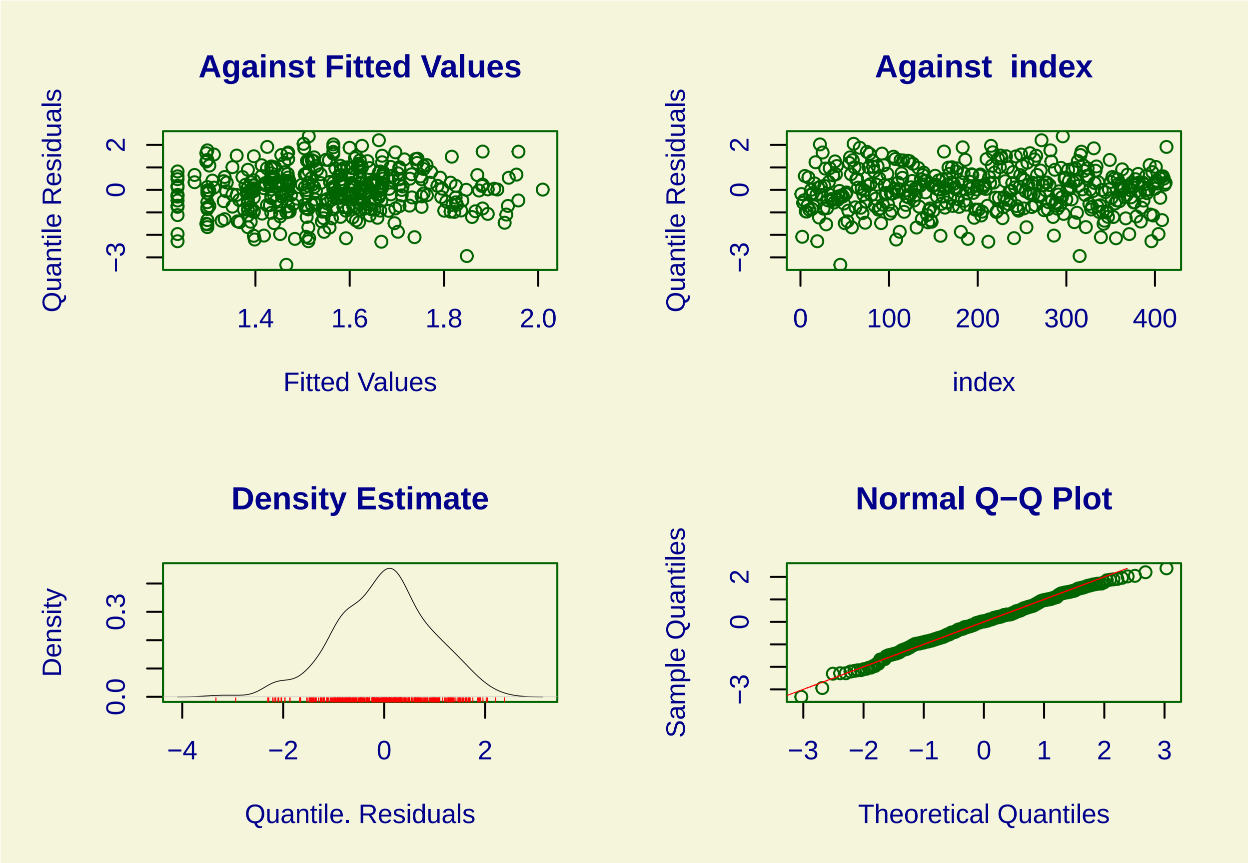
```r
ggplot(dat_1, aes(x = h, y = math)) + geom_point()
```




```
gamlss_1 <- gamlss(math ~ pb(h) + gender + random(school),
                   sigma.formula =~ 1,
                   nu.formula =~ 1,
                   c.crit = 0.001,
                   data = dat_1, n.cyc = 100, family = NBI())
```

```
## GAMLSS-RS iteration 1: Global Deviance = 1240.623
## GAMLSS-RS iteration 2: Global Deviance = 1240.702
## GAMLSS-RS iteration 3: Global Deviance = 1240.702
```

```
plot(gamlss_1)
```




```
## ******************************************************************
##   Summary of the Randomised Quantile Residuals
##                            mean   =  -0.01319988
##                        variance   =  0.8822983
##                coef. of skewness  =  -0.2237538
##                coef. of kurtosis  =  3.123681
## Filliben correlation coefficient  =  0.9972048
## ******************************************************************
```

Figure S1. From the upper-left to the bottom-right: the residuals against the fitted values (*mu* parameter); the residuals against participants' index; the residuals' Kernel density estimate; and the QQ-normal plot comparing estimated and theoretical residuals. The upper plots show no relationship between residuals and other variables (e.g., heteroscedasticity). The bottom plots indicate that the distribution of the residuals is approximately normal.

```
dropterm(gamlss_1, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math ~ pb(h) + gender + random(school)
##                    Df    AIC    LRT Pr(Chi)
## <none>                1255.9
## pb(h)          1.9301 1256.0 3.9217 0.13288
## gender         1.2537 1256.3 2.9516 0.11757
## random(school) 2.7550 1257.7 7.3214 0.05149 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

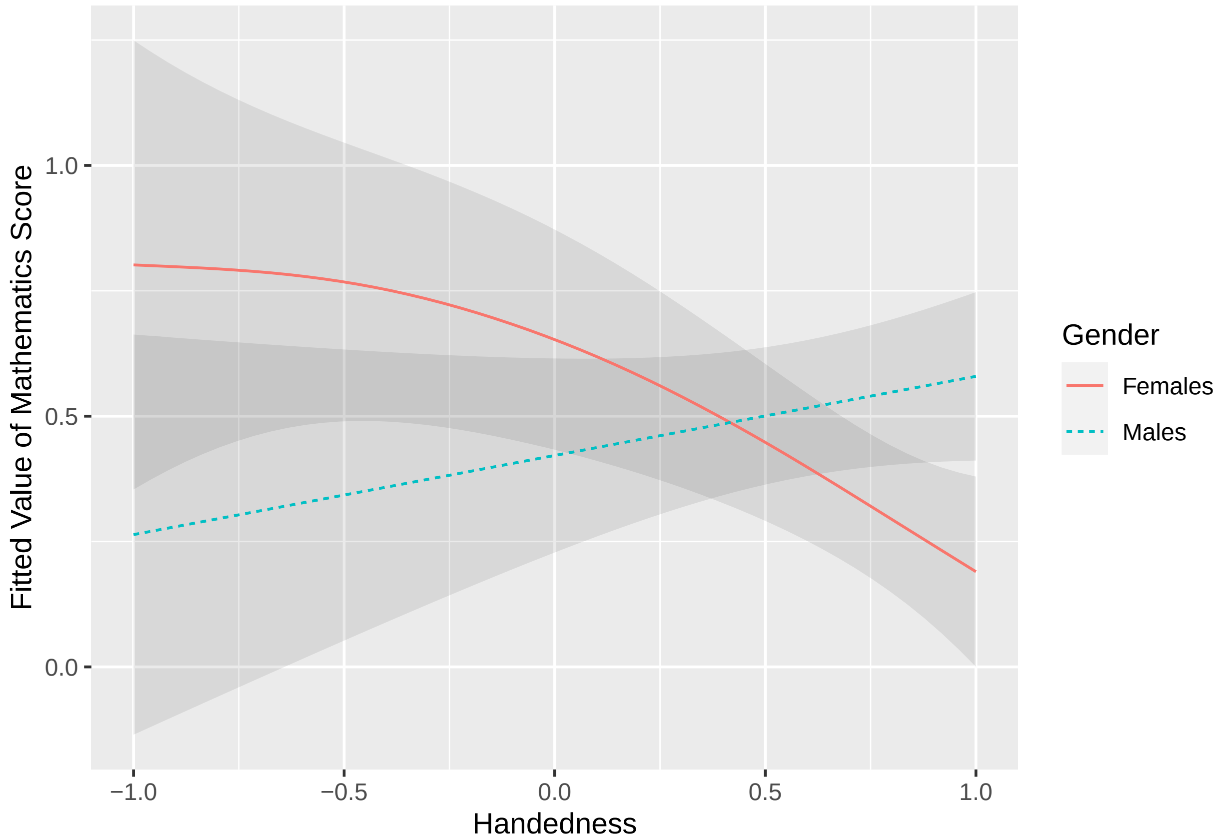
```r
dropterm(gamlss_1_int, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math ~ pvc(h, by = gender) + random(school, df = df1)
##                              Df    AIC     LRT  Pr(Chi)
## <none>                          1248.2
## pvc(h, by = gender)      3.8663 1256.9 16.4402 0.002197 **
## random(school, df = df1) 2.7922 1249.3  6.7152 0.069907 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
GAIC(gamlss_1, gamlss_1_int, k = 2.5)
```

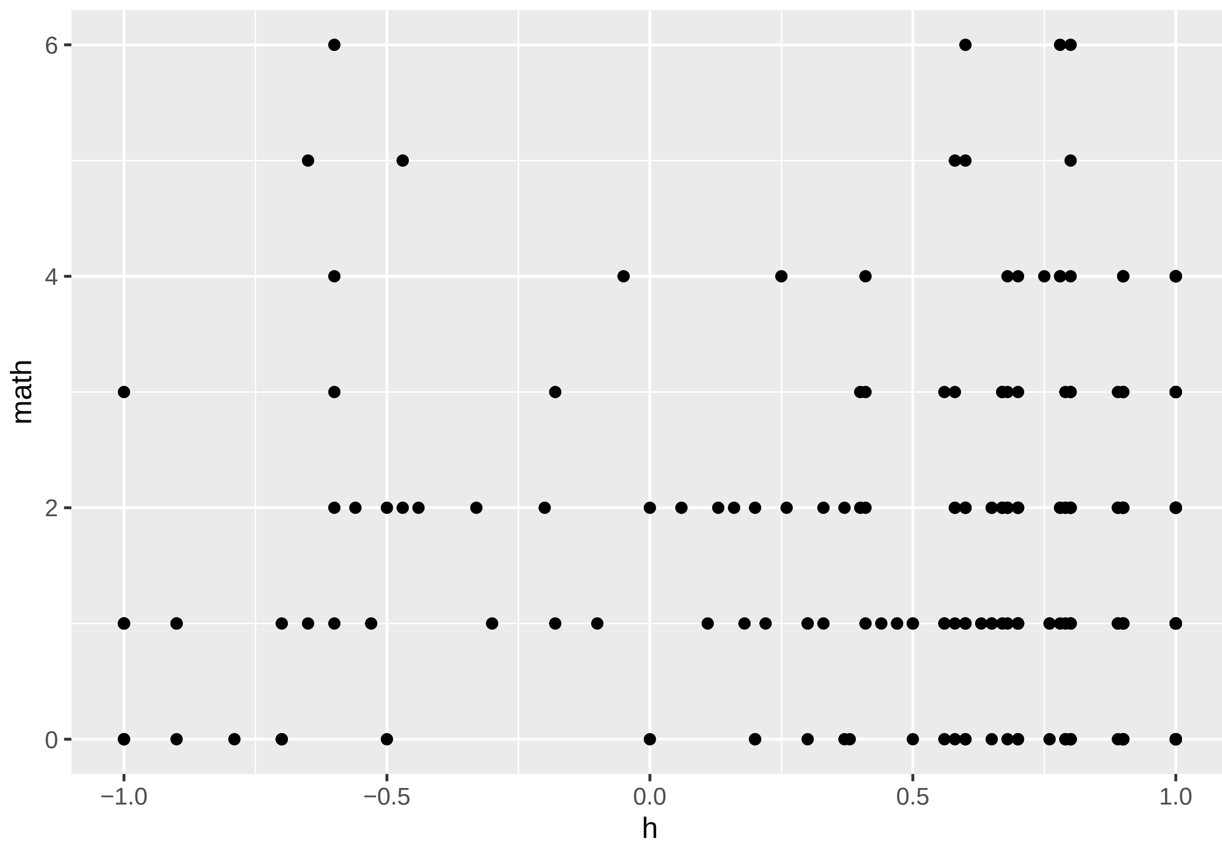
```
##                    df      AIC
## gamlss_1_int 8.634012 1252.491
## gamlss_1     7.592816 1259.684
```

```r
model_1 <- gam(math ~ s(h, by = gender) +
               gender + s(school , bs = "re"),
               data = dat_1, family = nb())

model_1p <- predict_gam(model_1, exclude_terms = "s(school)")
Fig1 <- model_1p %>%
  filter(school == "1") %>%
  mutate(Gender = case_when(gender == 0 ~ "Females",
                            gender == 1 ~ "Males")) %>%
  ggplot(aes(h, fit)) +
  geom_smooth_ci(Gender) +
  xlab("Handedness") +
  ylab("Fitted Value of Mathematics Score")

print(Fig1)
```



#Study 2

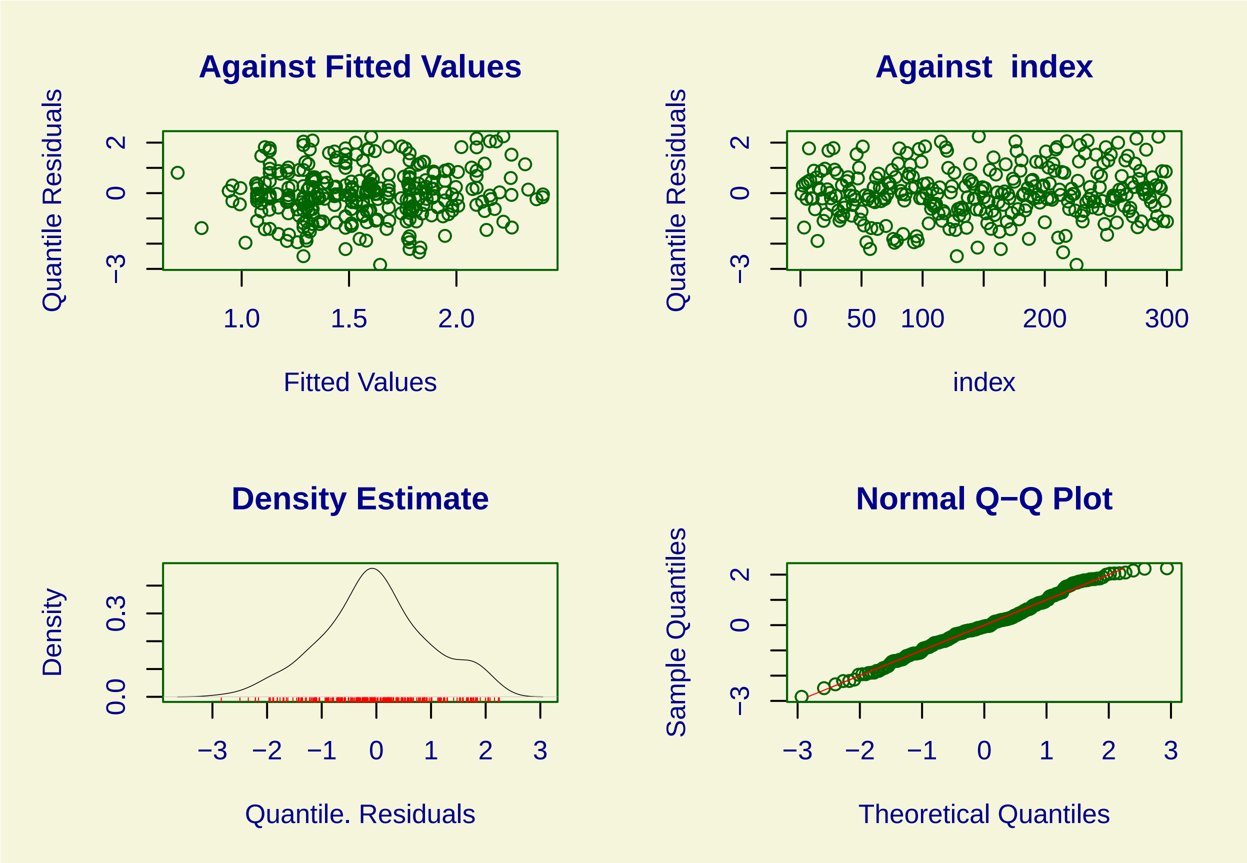
```
ggplot(dat_2, aes(x = h, y = math)) + geom_point()
```




```
gamlss_2 <- gamlss(math ~ pb(h) + gender + random(school),
                   sigma.formula =~ 1,
                   nu.formula =~ 1,
                   c.crit = 0.001,
                   data = dat_2, n.cyc = 100, family = NBI())
```

```
## GAMLSS-RS iteration 1: Global Deviance = 924.4631
## GAMLSS-RS iteration 2: Global Deviance = 924.2683
## GAMLSS-RS iteration 3: Global Deviance = 924.2651
## GAMLSS-RS iteration 4: Global Deviance = 924.2659
```

```
plot(gamlss_2)
```




```
## ******************************************************************
##   Summary of the Randomised Quantile Residuals
##                            mean   =  -0.009325386
##                        variance   =  0.9925761
##                coef. of skewness   =  0.01913409
##                coef. of kurtosis   =  2.826237
## Filliben correlation coefficient   =  0.9957679
## ******************************************************************
```

Figure S2. From the upper-left to the bottom-right: the residuals against the fitted values (*mu* parameter); the residuals against participants" index; the residuals' Kernel density estimate; and the QQ-normal plot comparing estimated and theoretical residuals. The upper plots show no relationship between residuals and other variables (e.g., heteroscedasticity). The bottom plots indicate that the distribution of the residuals is approximately normal.

```
dropterm(gamlss_2, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math ~ pb(h) + gender + random(school)
##                    Df    AIC     LRT   Pr(Chi)
## <none>                944.35
## pb(h)          3.2210 949.25 11.3348 0.0122729 *
## gender         1.4920 954.64 13.2659 0.0006397 ***
## random(school) 3.2876 946.53  8.7555 0.0410932 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

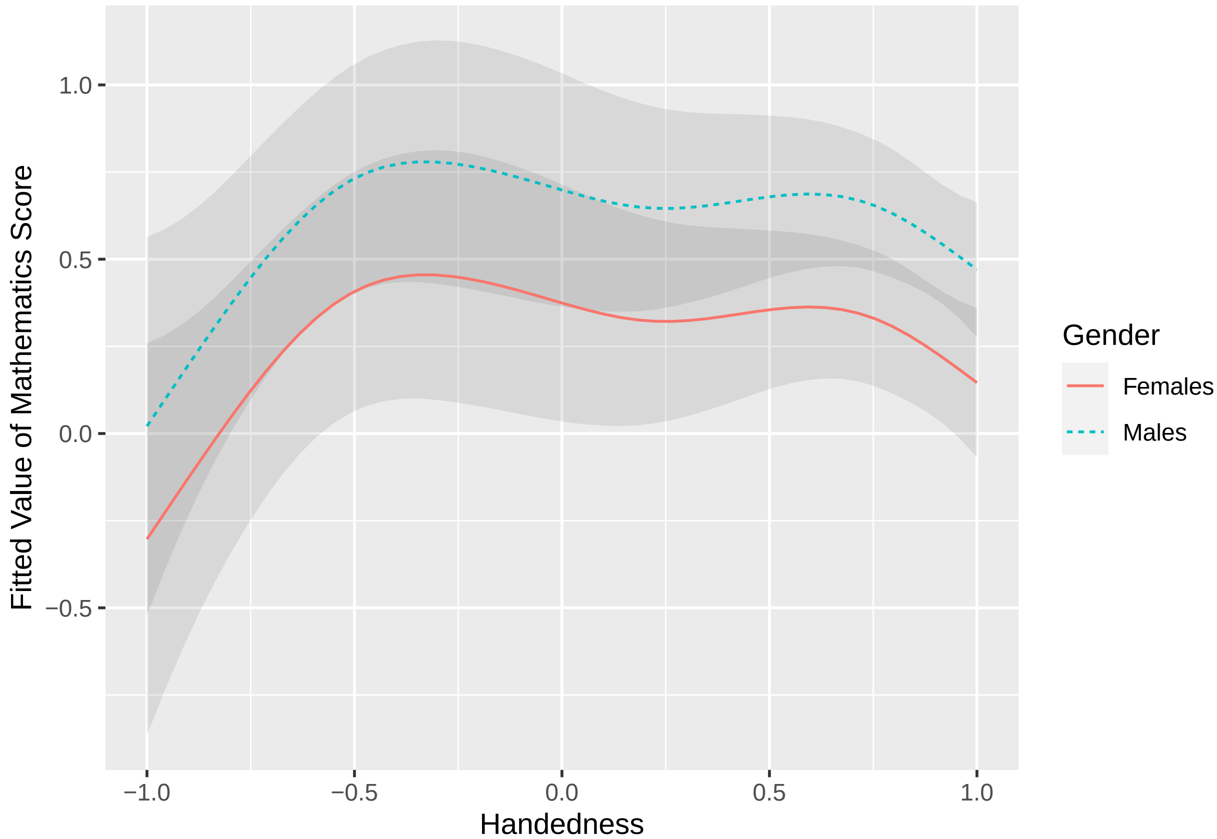
```
dropterm(gamlss_2_int, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math ~ pvc(h, by = gender) + random(school, df = df2)
##                                Df    AIC     LRT  Pr(Chi)
## <none>                             947.72
## pvc(h, by = gender)        5.1774 956.24 18.8724 0.002346 **
## random(school, df = df2) 3.0449 950.35  8.7186 0.034517 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
GAIC(gamlss_2, gamlss_2_int, k = 2.5)
```

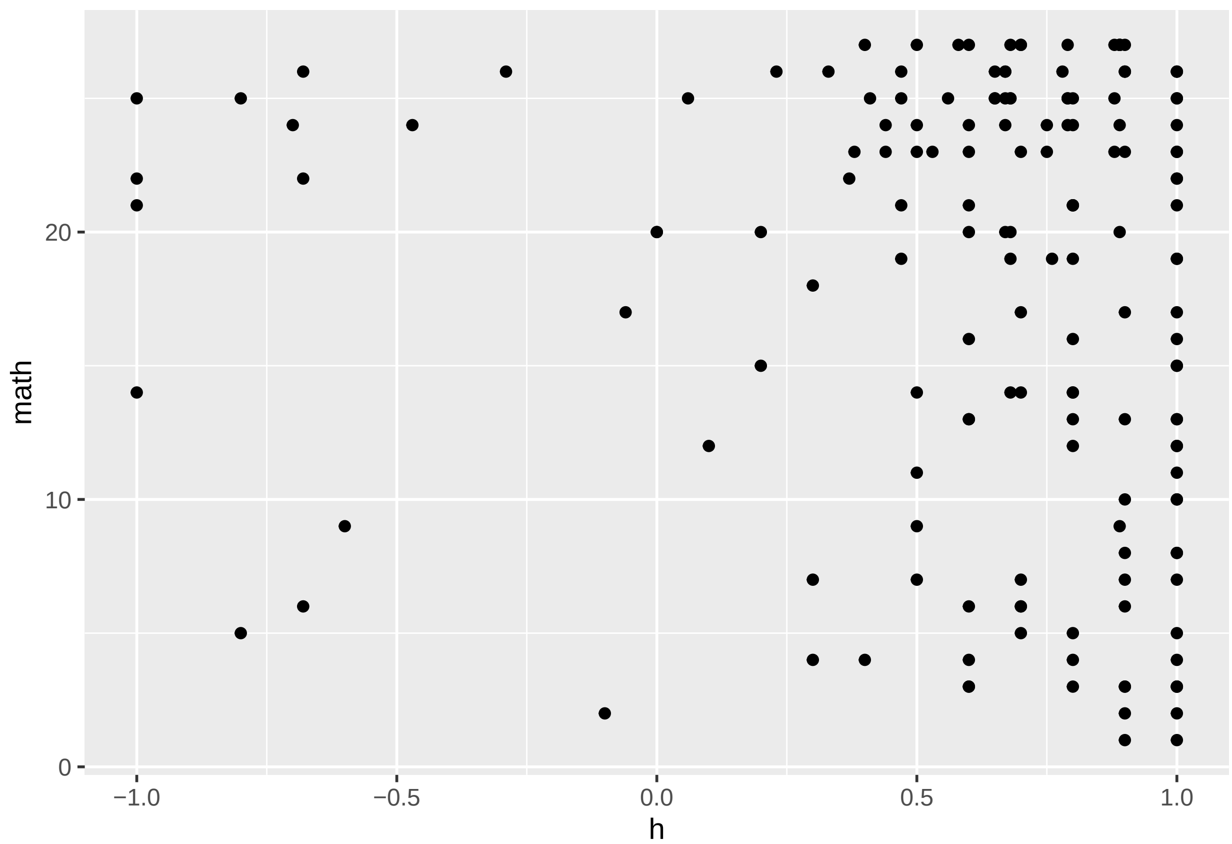
```
##                    df      AIC
## gamlss_2     10.04442 949.3769
## gamlss_2_int 10.64673 953.0415
```

```
model_2 <- gam(math ~ s(h) +
               gender + s(school , bs = "re"),
               data = dat_2, family = nb())


model_2p <- predict_gam(model_2, exclude_terms = "s(school)")
Fig2 <- model_2p %>%
  filter(school == "1") %>%
  mutate(Gender = case_when(gender == 0 ~ "Females",
                            gender == 1 ~ "Males")) %>%
  ggplot(aes(h, fit)) +
  geom_smooth_ci(Gender) +
  xlab("Handedness") +
  ylab("Fitted Value of Mathematics Score")

print(Fig2)
```



#Study 3

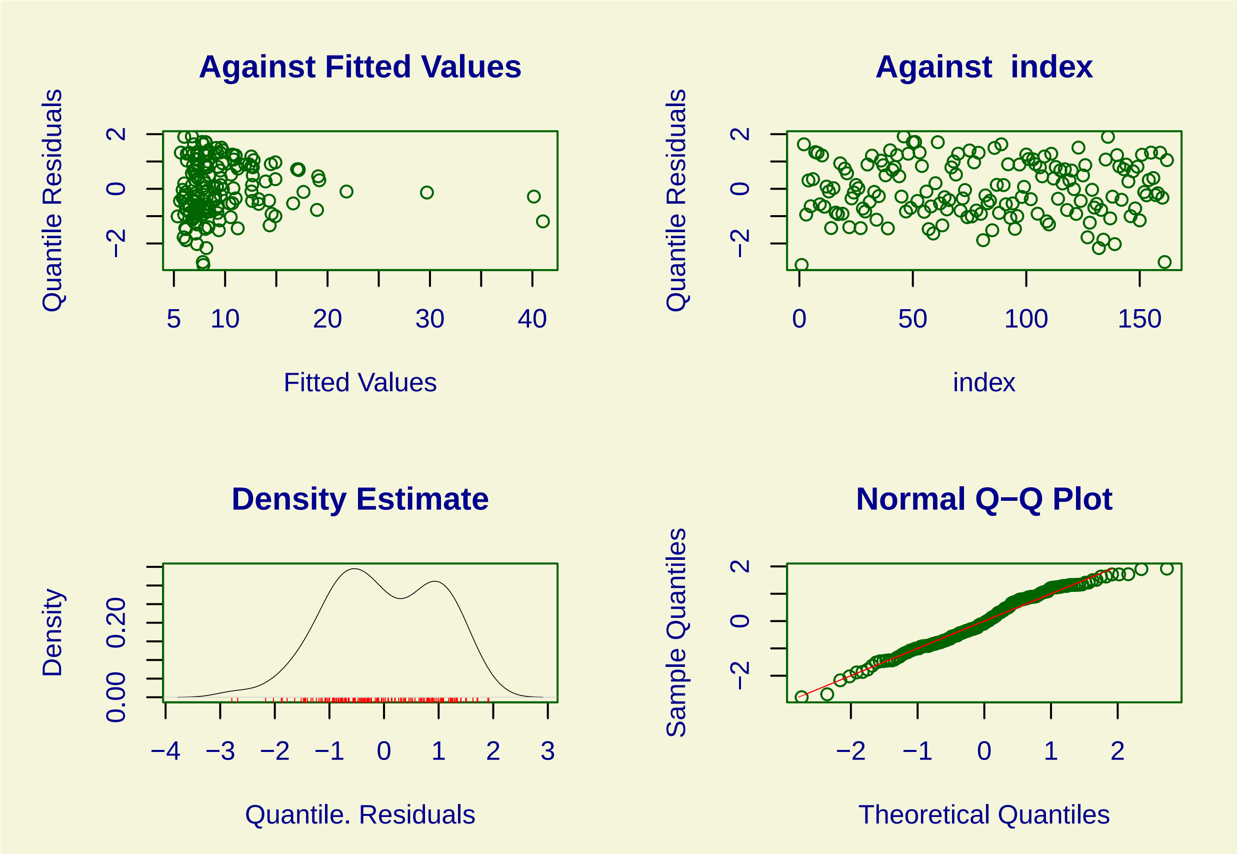
```
ggplot(dat_3, aes(x = h, y = math)) + geom_point()
```




```
gamlss_3 <- gamlss(math_inv ~ pb(h) + spat + gender,
                   sigma.formula =~ 1,
                   nu.formula =~ 1,
                   c.crit = 0.001,
                   data = dat_3, n.cyc = 100, family = NBI())
```

```
## GAMLSS-RS iteration 1: Global Deviance = 1051.906
## GAMLSS-RS iteration 2: Global Deviance = 1051.905
```

```
plot(gamlss_3)
```




```
## ******************************************************************
##   Summary of the Randomised Quantile Residuals
##                            mean   =  -7.413124e-05
##                        variance   =  1.032807
##                coef. of skewness   =  -0.1901872
##                coef. of kurtosis   =  2.322315
## Filliben correlation coefficient   =  0.9891712
## ******************************************************************
```

Figure S3. From the upper-left to the bottom-right: the residuals against the fitted values (mu parameter); the residuals against participants" index; the residuals' Kernel density estimate; and the QQ-normal plot comparing estimated and theoretical residuals. The upper plots show no relationship between residuals and other variables (e.g., heteroscedasticity). The bottom plots indicate that the distribution of the residuals is approximately normal (yet slightly platykurtic).

```
dropterm(gamlss_3, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math_inv ~ pb(h) + spat + gender
##             Df     AIC     LRT   Pr(Chi)
## <none>            1061.9
## pb(h)  1.00026 1060.5  0.5766    0.4478
## spat   0.16517 1078.0 16.3898 3.134e-06 ***
## gender 1.00000 1061.0  1.0936    0.2957
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
dropterm(gamlss_3_int, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## math_inv ~ pvc(h, by = gender) + spat
##                          Df     AIC    LRT   Pr(Chi)
## <none>                      1062.4
## pvc(h, by = gender) 3.07548 1059.5  3.248    0.3674
## spat                0.37725 1079.5 17.812 4.346e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
GAIC(gamlss_3, gamlss_3_int, k = 2.5)
```

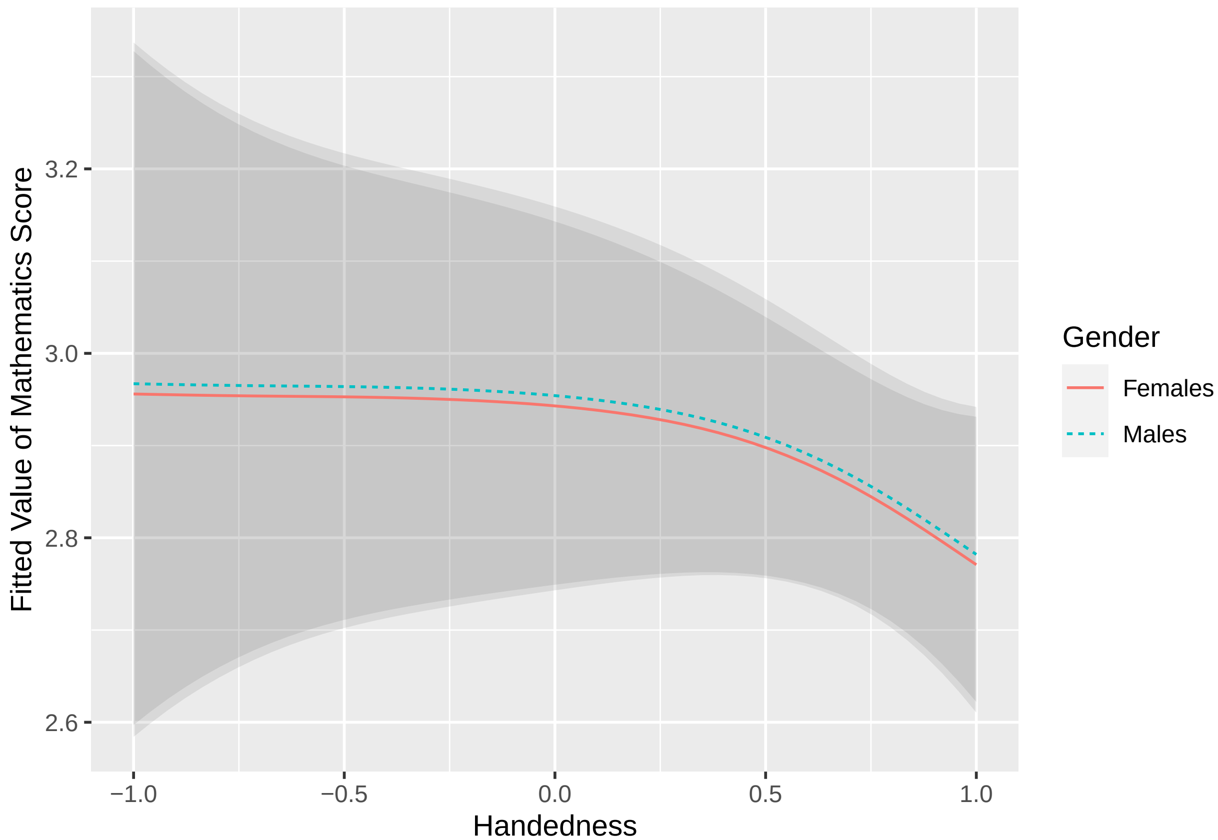
```
##                    df      AIC
## gamlss_3     5.000256 1064.406
## gamlss_3_int 6.075484 1065.434
```

```
model_3 <- gam(math ~ s(h) +
               gender, data = dat_3,
               family = nb())

summary(model_3)
```

```
##
## Family: Negative Binomial(3.744)
## Link function: log
##
## Formula:
## math ~ s(h) + gender
##
## Parametric coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.84987    0.06225  45.782   <2e-16 ***
## gender1      0.01113    0.08975   0.124    0.901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##        edf Ref.df Chi.sq p-value
## s(h) 1.614  1.984  2.015    0.33
##
## R-sq.(adj) =  0.0088   Deviance explained = 1.68%
## -REML = 592.99  Scale est. = 1         n = 162
```

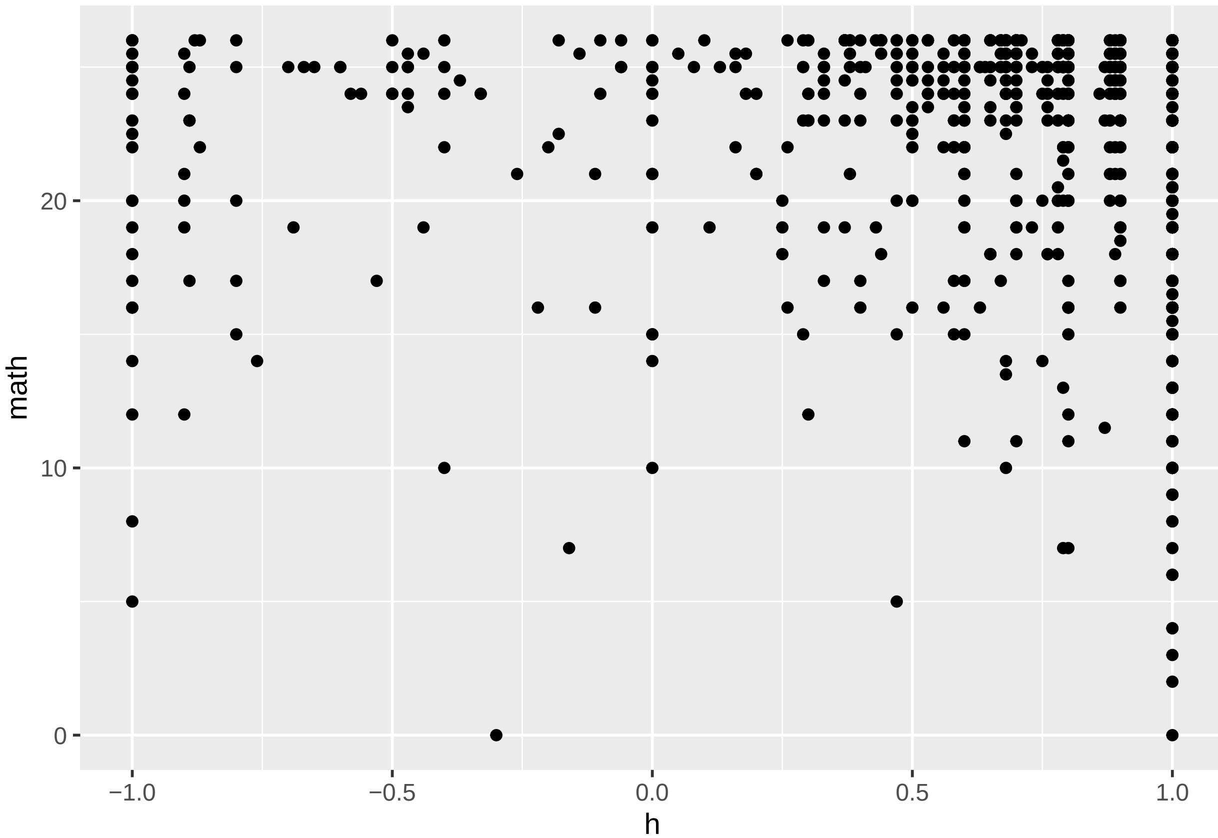
```
model_3p <- predict_gam(model_3)
Fig3 <- model_3p %>%
  mutate(Gender = case_when(gender == 0 ~ "Females",
                            gender == 1 ~ "Males")) %>%
  ggplot(aes(h, fit)) +
  geom_smooth_ci(Gender) +
```

```
  xlab("Handedness") +
  ylab("Fitted Value of Mathematics Score")

print(Fig3)
```



#Study 4

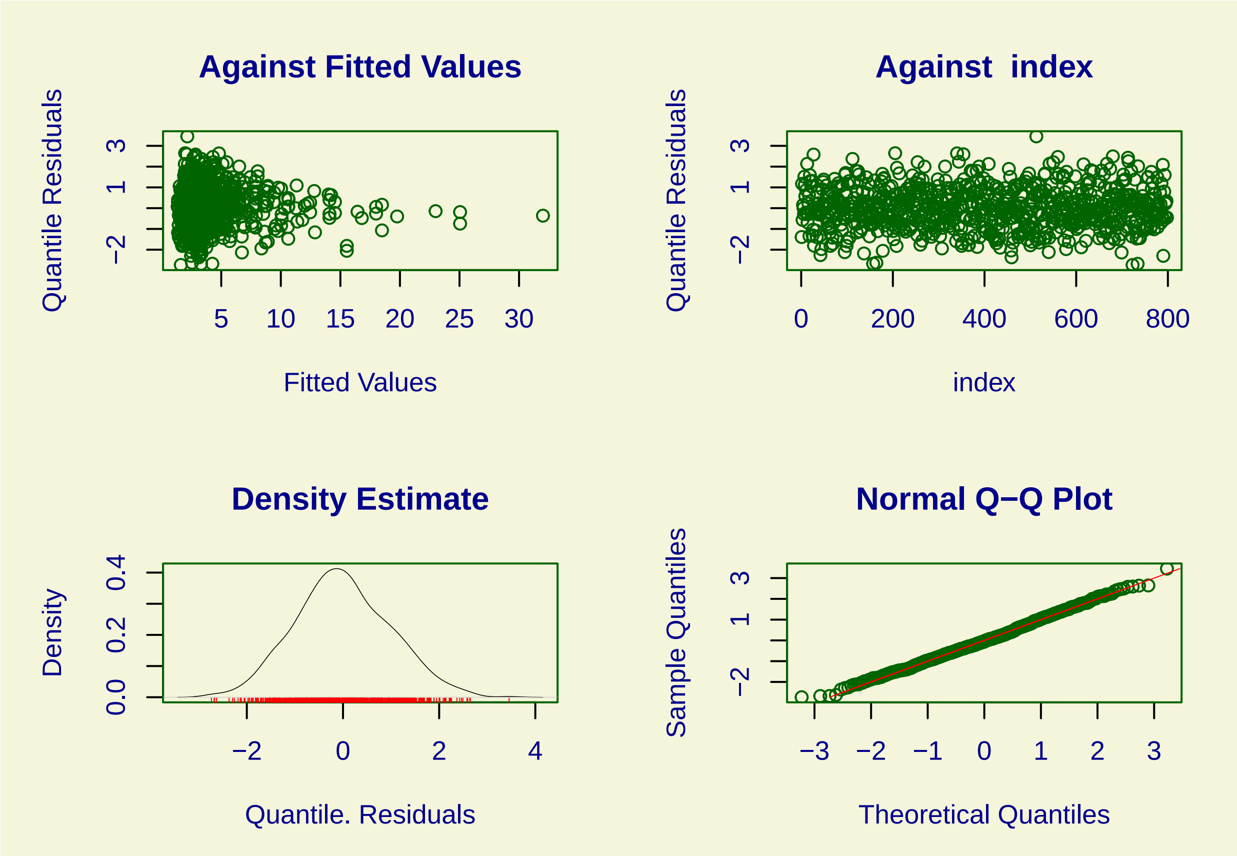
```
ggplot(dat_4, aes(x = h, y = math)) + geom_point()
```




```
gamlss_4 <- gamlss(round(math_inv) ~ pb(h) + gender + spat + random(school),
                   sigma.formula =~ 1,
                   nu.formula =~ 1,
                   c.crit = 0.001,
                   data = dat_4, n.cyc = 100, family = NBI())
```

```
## GAMLSS-RS iteration 1: Global Deviance = 3861.414
## GAMLSS-RS iteration 2: Global Deviance = 3861.365
## GAMLSS-RS iteration 3: Global Deviance = 3861.364
```

```
plot(gamlss_4)
```




```
## ******************************************************************
##   Summary of the Randomised Quantile Residuals
##                            mean   =  -0.001323925
##                        variance   =  0.9822149
##                coef. of skewness   =  0.1468449
##                coef. of kurtosis   =  2.892707
## Filliben correlation coefficient   =  0.998793
## ******************************************************************
```

Figure S4. From the upper-left to the bottom-right: the residuals against the fitted values (mu parameter); the residuals against participants" index; the residuals' Kernel density estimate; and the QQ-normal plot comparing estimated and theoretical residuals. The upper plots show no relationship between residuals and other variables (e.g., heteroscedasticity). The bottom plots indicate that the distribution of the residuals is approximately normal.

```
dropterm(gamlss_4, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## round(math_inv) ~ pb(h) + gender + spat + random(school)
##                     Df    AIC    LRT   Pr(Chi)
## <none>                 3887.9
## pb(h)          4.35880 3914.2 35.082 6.928e-07 ***
## gender         0.89492 3889.8  3.688   0.04668 *
## spat           0.58044 3976.2 89.496 < 2.2e-16 ***
```

```
## random(school) 4.49447 3920.1 41.194 4.633e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

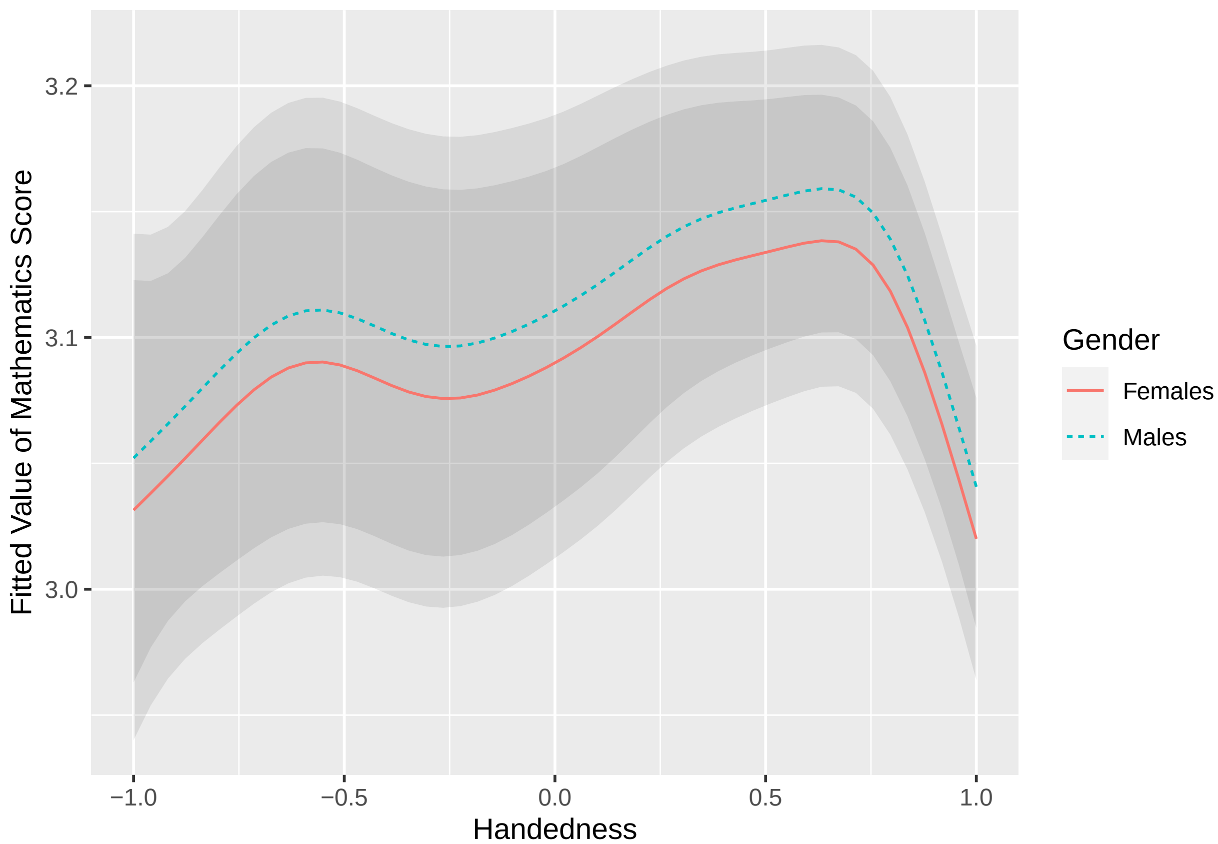
```
dropterm(gamlss_4_int, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## round(math_inv) ~ pvc(h, by = gender) + spat + random(school,
##     df = df4)
##                               Df    AIC    LRT   Pr(Chi)
## <none>                            3889.8
## pvc(h, by = gender)      7.81908 3917.9 43.734 5.382e-07 ***
## spat                     0.15294 3974.6 85.105 < 2.2e-16 ***
## random(school, df = df4) 4.61068 3924.9 44.279 1.252e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
GAIC(gamlss_4, gamlss_4_int, k=2.5)
```

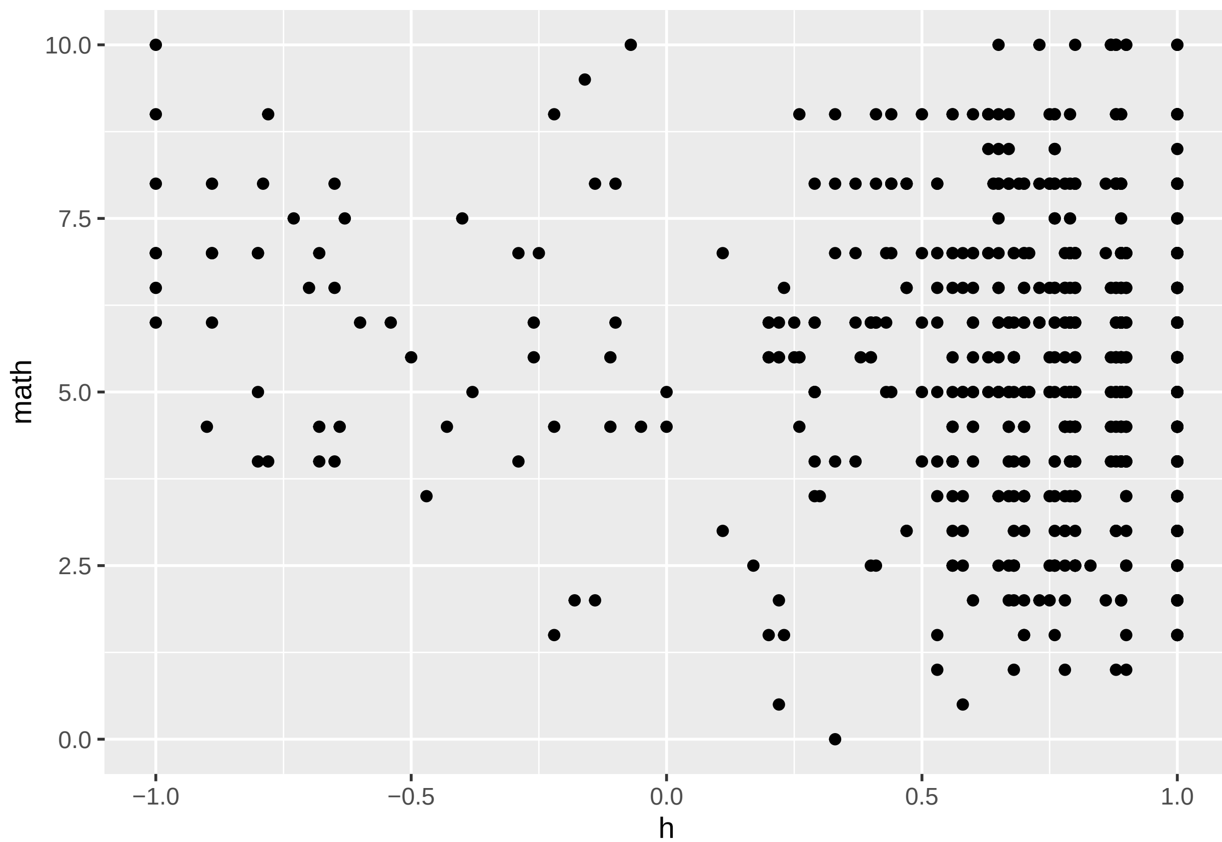
```
##                    df      AIC
## gamlss_4     13.24503 3894.477
## gamlss_4_int 15.60931 3897.638
```

```
model_4 <- gam(math ~ s(h) +
               gender + s(school, bs = "re"),
               data = dat_4, family = nb())


model_4p <- predict_gam(model_4, exclude_terms = "s(school)")
Fig4 <- model_4p %>%
  filter(school == "1") %>%
  mutate(Gender = case_when(gender == 0 ~ "Females",
                            gender == 1 ~ "Males")) %>%
  ggplot(aes(h, fit)) +
  geom_smooth_ci(Gender) +
  xlab("Handedness") +
  ylab("Fitted Value of Mathematics Score")

print(Fig4)
```



#Study 5

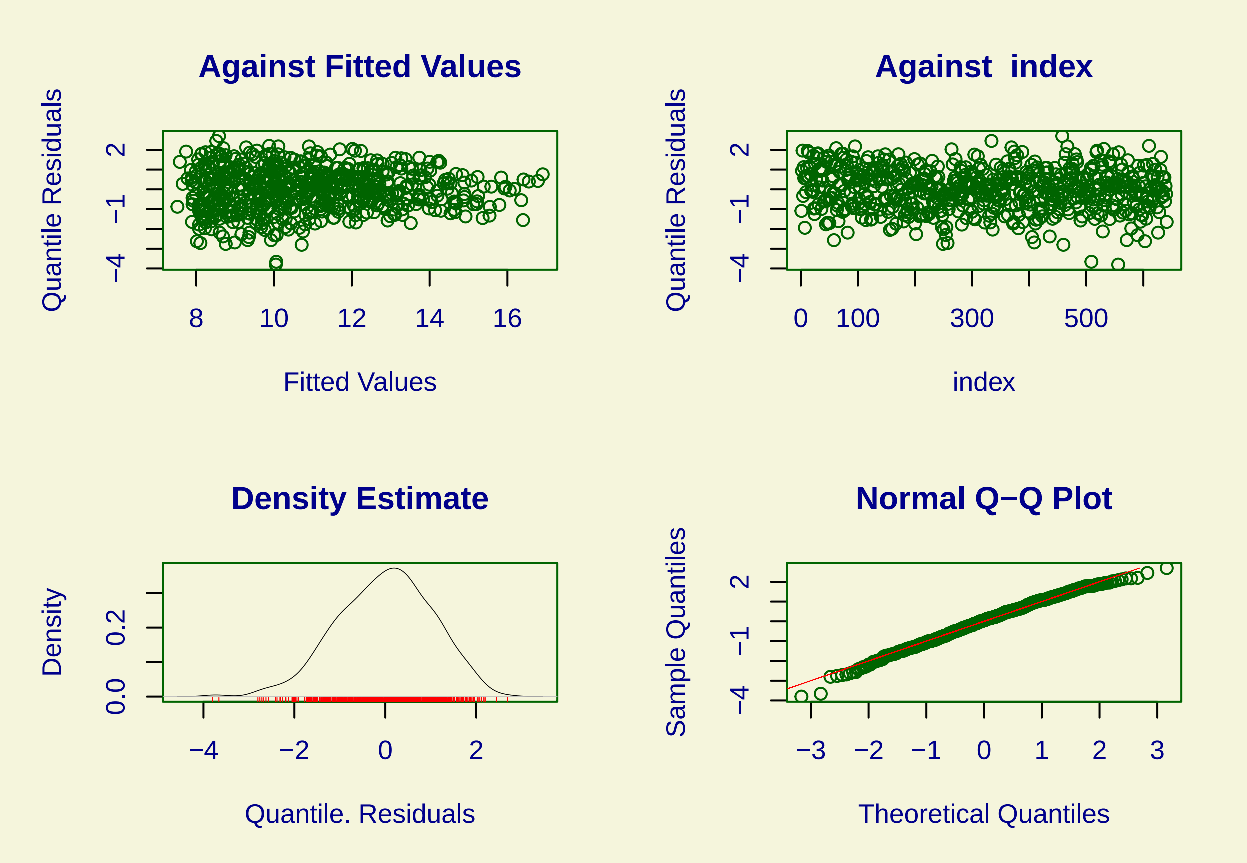
```
ggplot(dat_5, aes(x = h, y = math)) + geom_point()
```




```
gamlss_5 <- gamlss(I(math * 2) ~ pb(h) + gender + spat +
                   random(school),
                   sigma.formula =~ 1,
                   nu.formula =~ 1,
                   c.crit = 0.001,
                   data = dat_5, n.cyc = 100, family = NBI())
```

```
## GAMLSS-RS iteration 1: Global Deviance = 3455.536
## GAMLSS-RS iteration 2: Global Deviance = 3455.298
## GAMLSS-RS iteration 3: Global Deviance = 3455.298
```

```
plot(gamlss_5)
```




```
## ******************************************************************
##   Summary of the Randomised Quantile Residuals
##                            mean   =  -0.004694524
##                        variance   =  1.07981
##                coef. of skewness   =  -0.2783742
##                coef. of kurtosis   =  2.963451
## Filliben correlation coefficient   =  0.9968823
## ******************************************************************
```

Figure S5. From the upper-left to the bottom-right: the residuals against the fitted values (*mu* parameter); the residuals against participants" index; the residuals' Kernel density estimate; and the QQ-normal plot comparing estimated and theoretical residuals. The upper plots show no relationship between residuals and other variables (e.g., heteroscedasticity). The bottom plots indicate that the distribution of the residuals is approximately normal.

```
dropterm(gamlss_5, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## I(math * 2) ~ pb(h) + gender + spat + random(school)
##                    Df    AIC    LRT   Pr(Chi)
## <none>                3473.1
## pb(h)         1.97397 3473.5  4.346    0.1114
## gender        0.99837 3473.5  2.407    0.1205
## spat          1.47692 3529.5 59.348 4.267e-14 ***
```

```
## random(school) 2.82973 3554.0 86.572 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

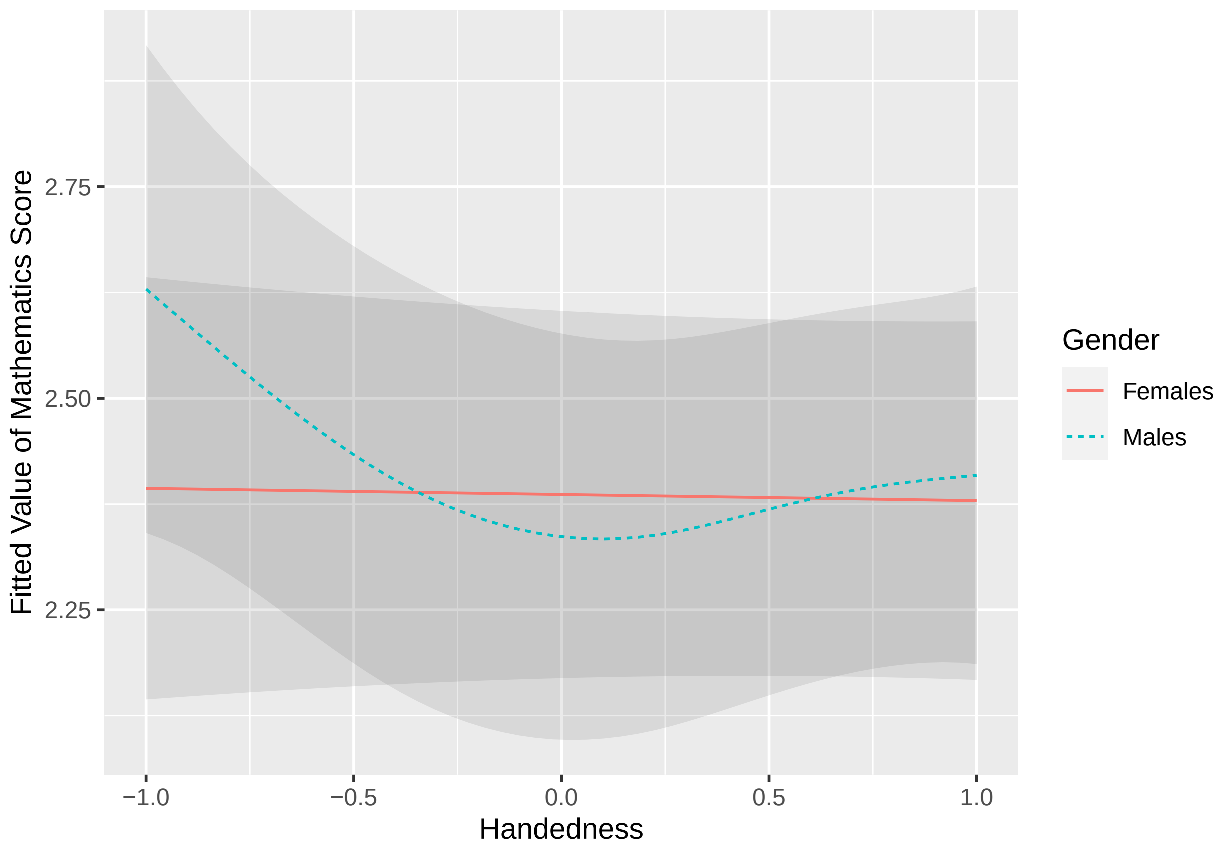
```
dropterm(gamlss_5_int, test = "Chisq")
```

```
## Single term deletions for
## mu
##
## Model:
## I(math * 2) ~ pvc(h, by = gender) + spat + random(school, df = df5)
##                             Df    AIC    LRT   Pr(Chi)
## <none>                          3471.3
## pvc(h, by = gender)      4.6779 3473.8 11.860   0.02985 *
## spat                     1.2134 3527.6 58.723 3.142e-14 ***
## random(school, df = df5) 2.6303 3551.7 85.612 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
GAIC(gamlss_5, gamlss_5_int, k = 2.5)
```

```
##                     df      AIC
## gamlss_5_int 10.594038 3476.628
## gamlss_5      8.892091 3477.528
```

```
model_5 <- gam(I(math * 2) ~ s(h, by = gender) +
               gender + s(school, bs = "re"),
               data = dat_5, family = nb())


model_5p <- predict_gam(model_5, exclude_terms = "s(school)")
Fig5 <- model_5p %>%
  filter(school == "1") %>%
  mutate(Gender = case_when(gender == 0 ~ "Females",
                            gender == 1 ~ "Males")) %>%
  ggplot(aes(h, fit)) +
  geom_smooth_ci(Gender) +
  xlab("Handedness") +
  ylab("Fitted Value of Mathematics Score")

print(Fig5)
```


Fitted Value of Mathematics Score
2.75
2.50
2.25
−1.0
−0.5
0.0
0.5
1.0
Handedness
Gender
Females
Males