```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(ggplot2)
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.1
```

```
## Carregando pacotes exigidos: carData
```

```
## Warning: package 'carData' was built under R version 4.3.1
```

```
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(EnvStats)
```

```
## Warning: package 'EnvStats' was built under R version 4.3.1
```

```
##
## Attaching package: 'EnvStats'
##
## The following object is masked from 'package:car':
##
##     qqPlot
##
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```r
getwd()
```

```
## [1] "C:/Users/maris/Downloads"
```

```r
eua.escolas <- read_xlsx("C:/Users/maris/Downloads/dados.xlsx")
```

```r
fit_ingles <- lm(testscr ~ el_pct, data = eua.escolas)
res_ingles <- residuals(fit_ingles)

summary(fit_ingles) ##teste t beta_1: beta_1 diferente de zero, existe associação significante
```

```
##
## Call:
## lm(formula = testscr ~ el_pct, data = eua.escolas)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.861 -10.183  -0.807   9.004  45.183
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 664.73944    0.94064  706.69   <2e-16 ***
## el_pct       -0.67116    0.03898  -17.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.59 on 418 degrees of freedom
## Multiple R-squared:  0.4149, Adjusted R-squared:  0.4135
## F-statistic: 296.4 on 1 and 418 DF,  p-value: < 2.2e-16
```

```r
anova_fit_ingles <- anova(fit_ingles)  ##teste f beta_1: mesmo resultado de antes

shapiro.test(res_ingles)  ##teste de normalidade: nao rejeito a hipotese nula de normalidade
```

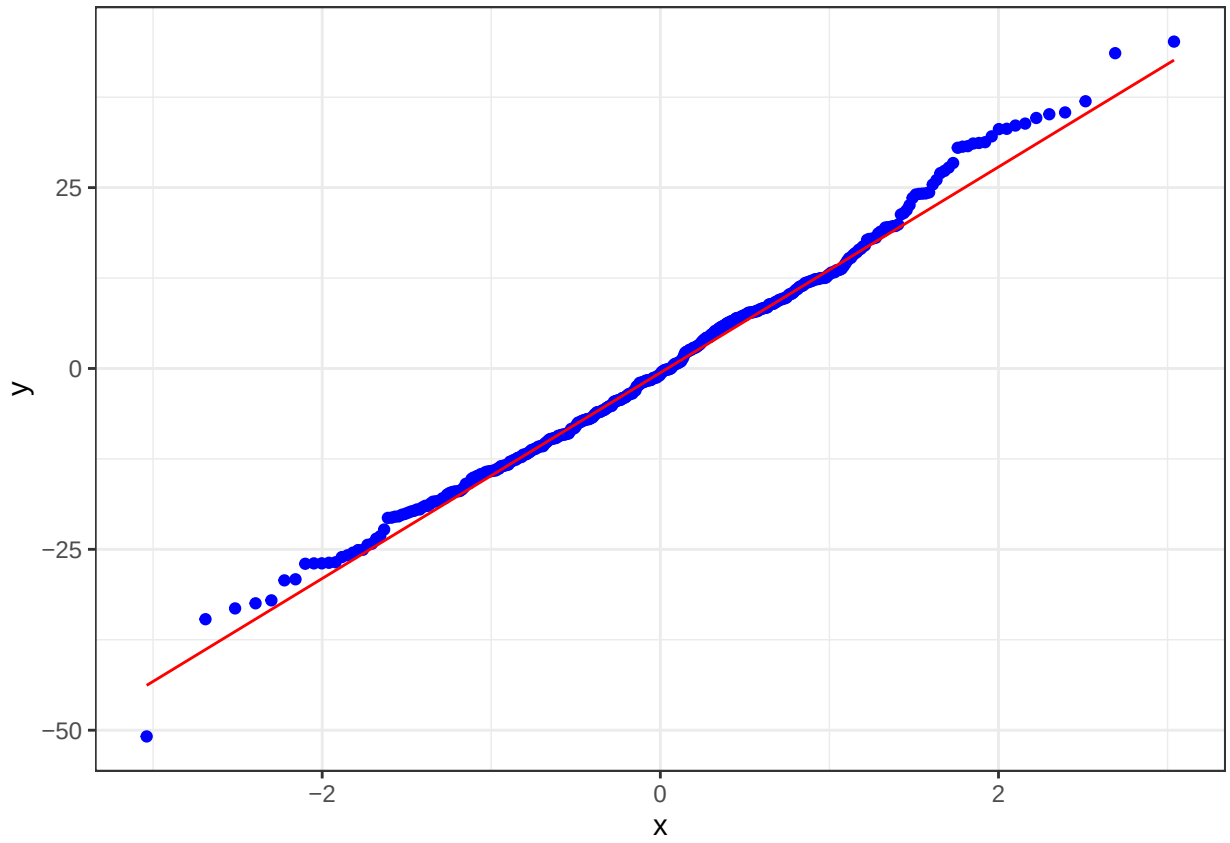
```
##
##  Shapiro-Wilk normality test
##
## data:  res_ingles
## W = 0.99336, p-value = 0.06159
```

```r
ks.test(res_ingles, y = pnorm)  ##teste de normalidade: nao rejeito a hipotese nula de normalidade
```

```
## Warning in ks.test.default(res_ingles, y = pnorm): ties should not be present
## for the Kolmogorov-Smirnov test
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  res_ingles
## D = 0.44166, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

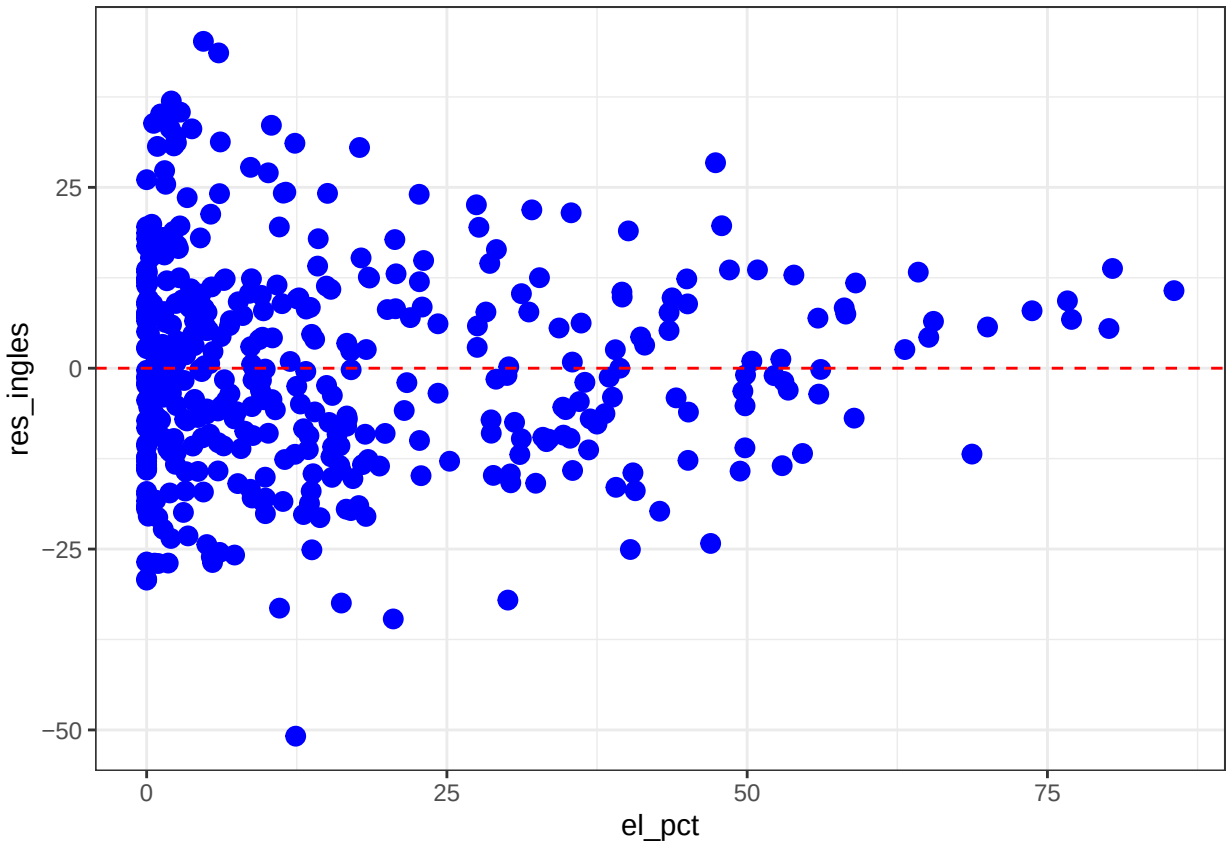
```
eua.escolas %>% ggplot(aes(sample = res_ingles)) +
  stat_qq(color = "blue") +
  stat_qq_line(color = "red") +
  theme_bw()  ## parece ter um pouco de cauda pesada, nem tao relevante
```




```
ncvTest(fit_ingles)  ##teste de homocedasticidade: rejeito a hipotese nula, nao é homocedastico
```

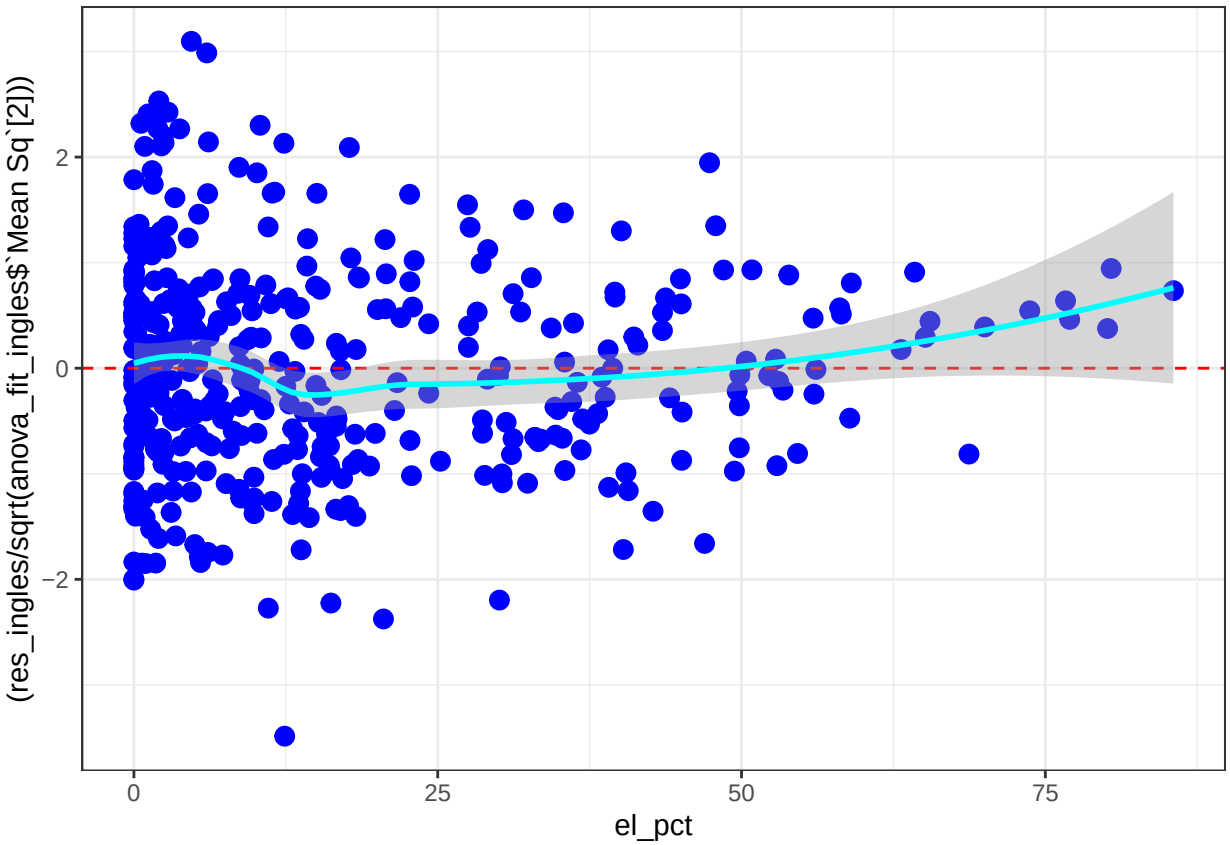
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 9.407229, Df = 1, p = 0.0021613
```

```
eua.escolas %>% ggplot(aes(x = el_pct, y = res_ingles)) +
  geom_point(color = "blue", size = 3) +
  theme_bw() +
  geom_hline(yintercept = 0, color = "red", linetype = 2) ##grafico dos residuos x preditora
```




```
eua.escolas %>% ggplot(aes(x = el_pct, y = (res_ingles/sqrt(anova_fit_ingles$`Mean Sq`[2])))) +
  geom_point(color = "blue", size = 3) +
  theme_bw() +
  geom_hline(yintercept = 0, color = "red", linetype = 2) +
  geom_smooth(color = "cyan") ##grafico residuos padronizados x preditora
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
anovaPE(fit_ingles)  ##teste de falta de ajuste: não rejeito a hipotese nula, o modelo linear é razoave
```

```
##                    Df Sum Sq Mean Sq  F value Pr(>F)
## el_pct              1  63110   63110 319.5906 <2e-16 ***
## Lack of Fit       370  79521     215   1.0884 0.3709
## Pure Error         48   9479     197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```