---

title: “BasicStatistic” author: “Maksym Rud”’ date: “13 09 2020” output: pdf_document

---

Father’s Hight Statistics

The table below gives the heights of fathers and their sons, based on a famous experiment by Karl Pearson around 1903. The number of cases is 1078. Random noise was added to the original data, to produce heights to the nearest 0.1 inch.

```
my_data <- read.delim(file.choose())

knitr::kable(head(my_data[1:5,]), "simple")
```

Father	Son
65.0	59.8
63.3	63.2
65.0	63.3
65.8	62.8
61.1	64.3

Here are some basic evaluations of the data

```
print(c("Mean", mean(my_data$Father)))
```

```
## [1] "Mean"             "67.686827458256"
```

```
print(c("Median", median(my_data$Father)))
```

```
## [1] "Median" "67.8"
```

```
print(c("Variance", var(my_data$Father)))
```

```
## [1] "Variance"         "7.53956634160375"
```

```
print(c("Standart deviation", sd(my_data$Father)))
```

```
## [1] "Standart deviation" "2.74582707787722"
```

```
print(c("coef of variance", sd(my_data$Father)/mean(my_data$Father)))
```

```
## [1] "coef of variance"   "0.0405666387536132"
```

```r
print(c("Range of the set of data", max(my_data$Father) - min(my_data$Father)))
```

```
## [1] "Range of the set of data" "16.4"
```

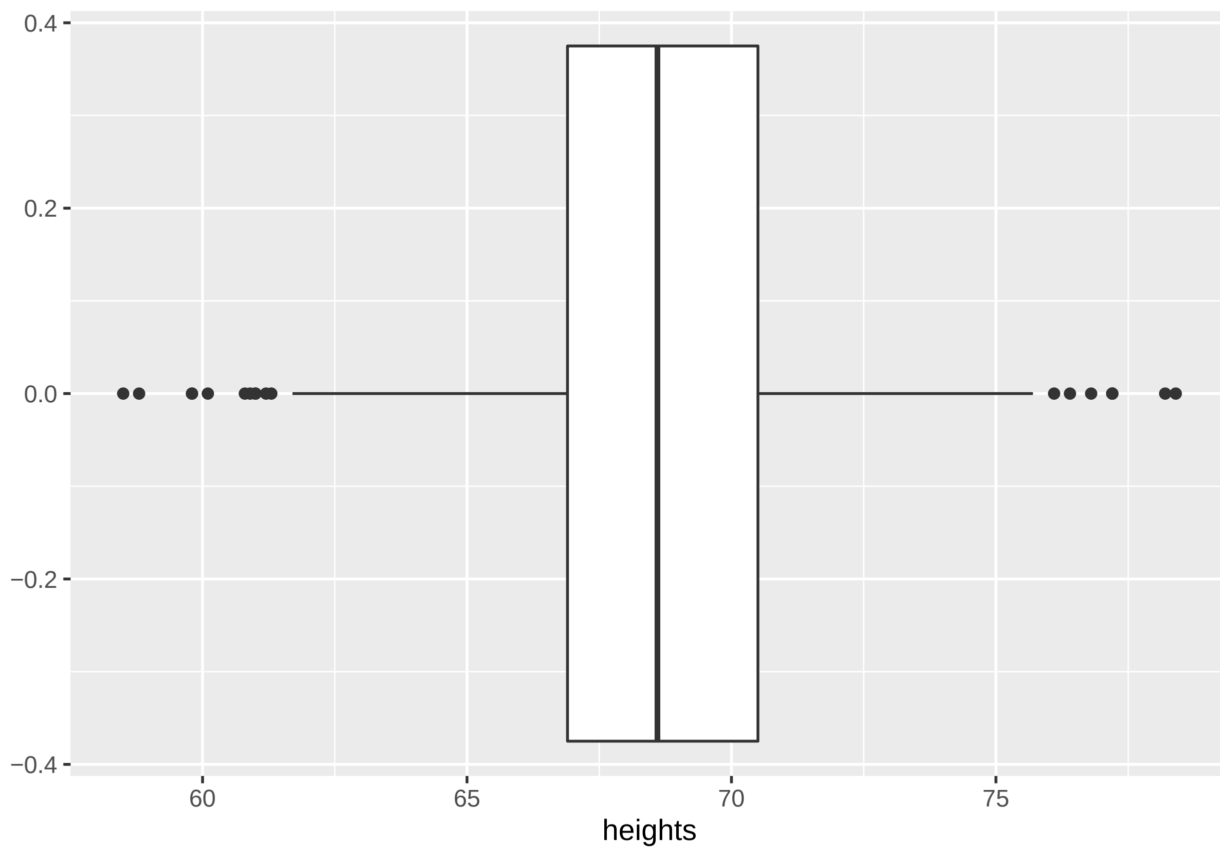
```r
print(c("interquartile range", IQR(my_data$Father, na.rm = FALSE)))
```

```
## [1] "interquartile range" "3.8"
```

Box Plot

The plot is representing interquartile range, mean and “vityk” values and informal information about density distribution.

```r
ggplot(my_data) +
  aes(y = unlist(my_data[2])) +
  geom_boxplot() +
  coord_flip() +
  labs(
    y = "heights"
  )
```



Quartiles

```r
quantile(my_data$Father)
```

```
##   0%  25%  50%  75% 100%
## 59.0 65.8 67.8 69.6 75.4
```

Summary

```r
summary(my_data$Father)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   59.00   65.80   67.80   67.69   69.60   75.40
```

First and Ninth Deciles

```r
quantile(my_data$Father, prob = seq(0, 1, length = 11), type = 5)[2]
```

```
##  10%
## 64.3
```

```r
quantile(my_data$Father, prob = seq(0, 1, length = 11), type = 5)[10]
```

```
##  90%
## 71.3
```

Skewness

```r
skewness(my_data$Father)
```

```
## [1] -0.0881151
```

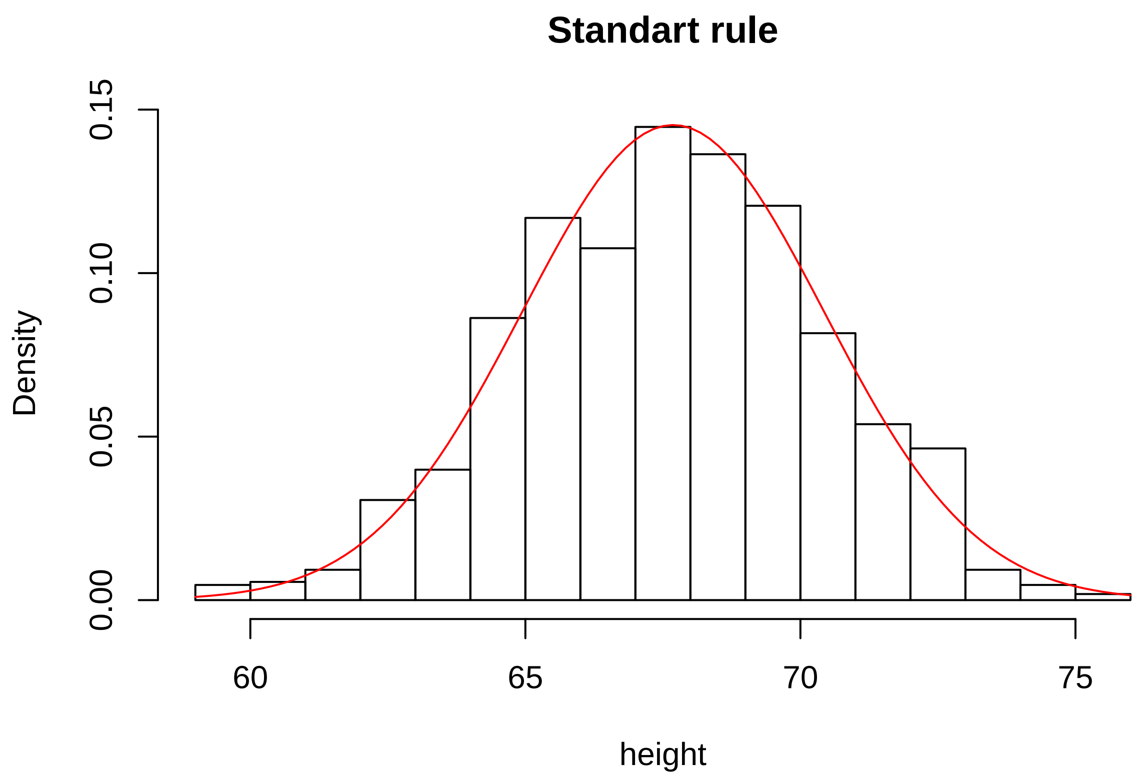
Kurtosis

```r
kurtosis(my_data$Father)
```

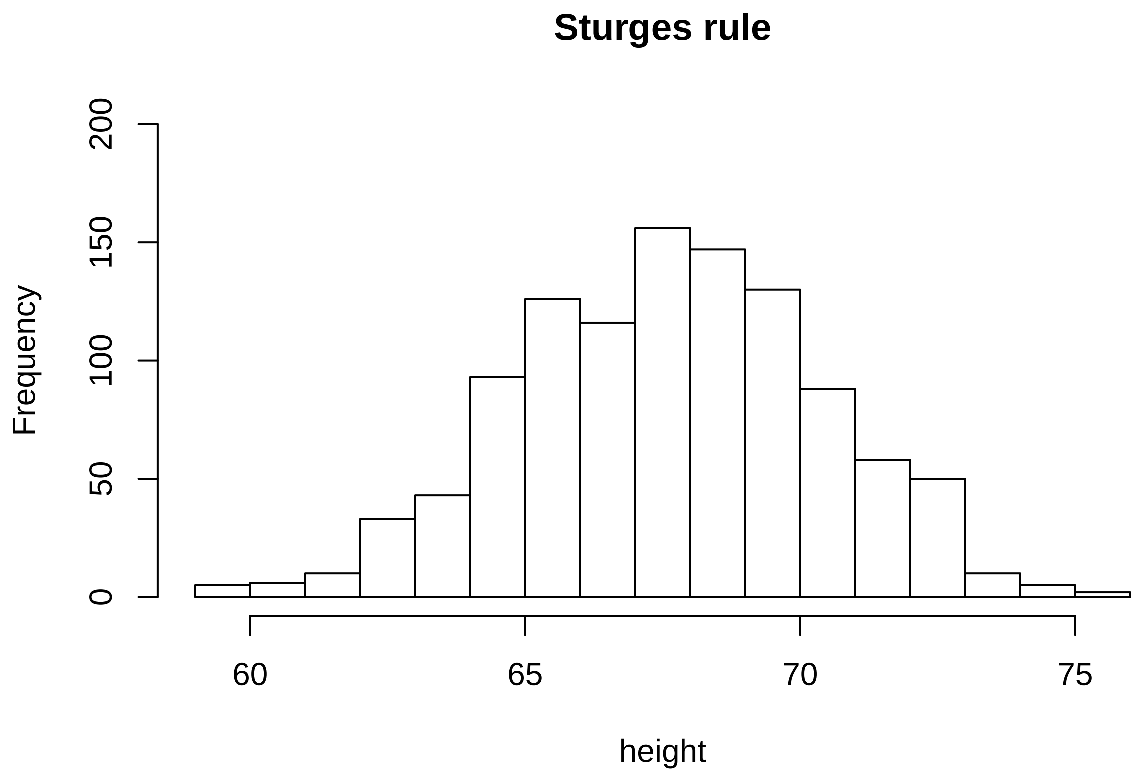
```
## [1] -0.1645894
```

Histograms based on

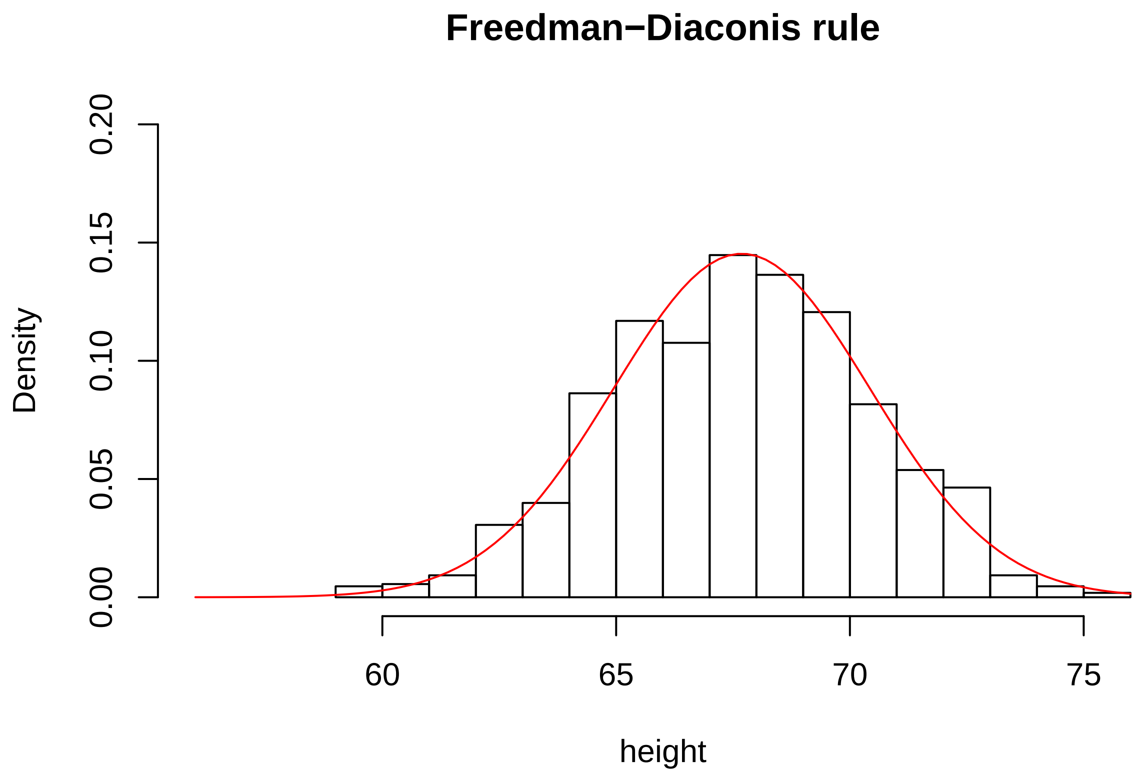
```r
hist(my_data$Father, main = "Standart rule", xlab = "height", freq = FALSE)
std = sqrt(var(my_data$Father))
curve(dnorm(x,mean = mean(my_data$Father), sd = std), add = TRUE, col="red")
```



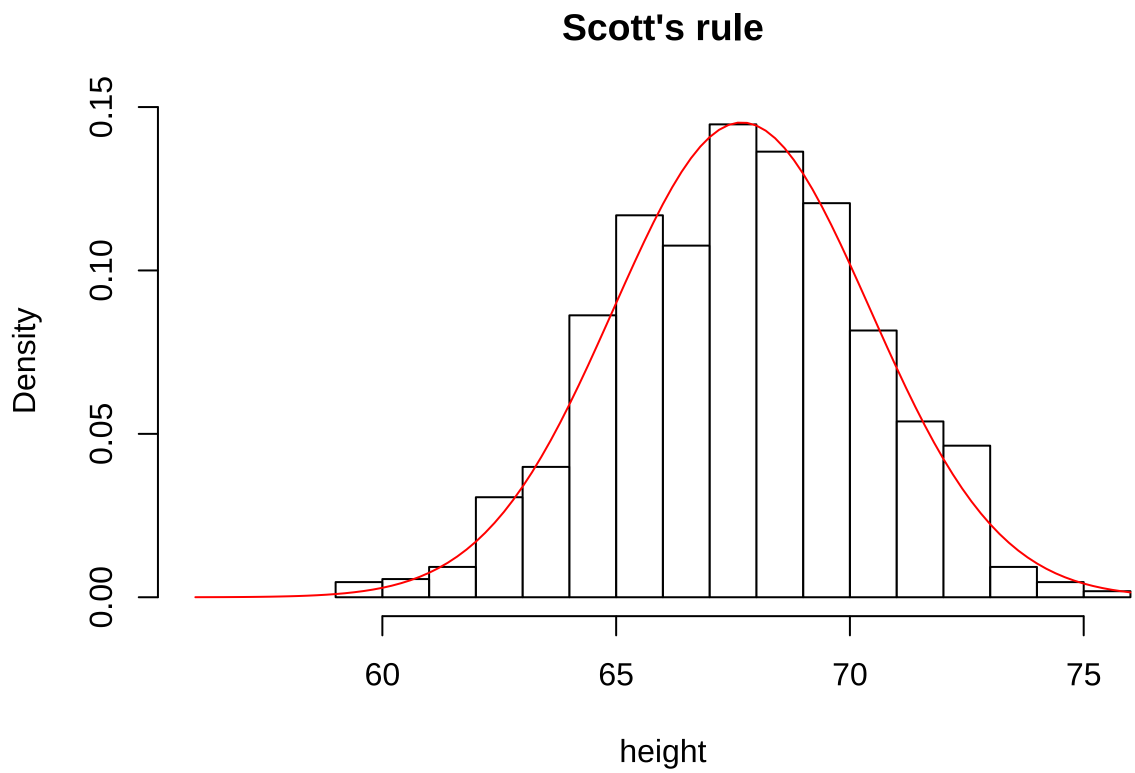

```r
hist(my_data$Father, main = "Sturges rule", xlab = "height", ylim = c(0, 200))
```




```
hist(my_data$Father,main = "Freedman-Diaconis rule", xlab = "height", breaks = "FD", freq = FALSE,
     xlim = c(56, 76), ylim = c(0, 0.2))
curve(dnorm(x,mean = mean(my_data$Father), sd = std), col="red", add = TRUE)
```



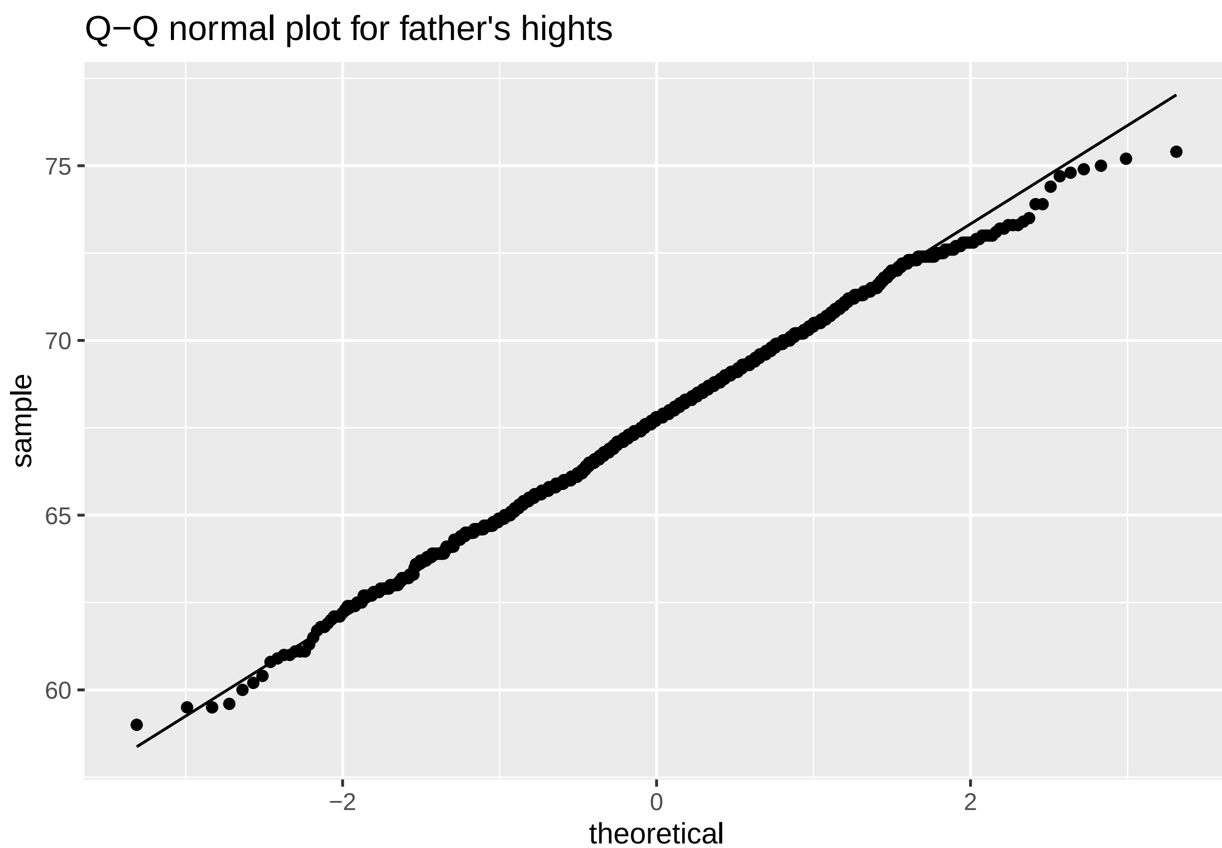

```
hist(my_data$Father,main = "Scott’s rule", xlab = "height",  breaks = "Scott", freq = FALSE, xlim = c(5
curve(dnorm(x,mean = mean(my_data$Father), sd = std), col="red", add = TRUE)
```



Creating a Normal Quantile–Quantile Plot

Sometimes it's important to know if your data is normally distributed. A quantile–quantile (Q–Q) plot is a good first check.

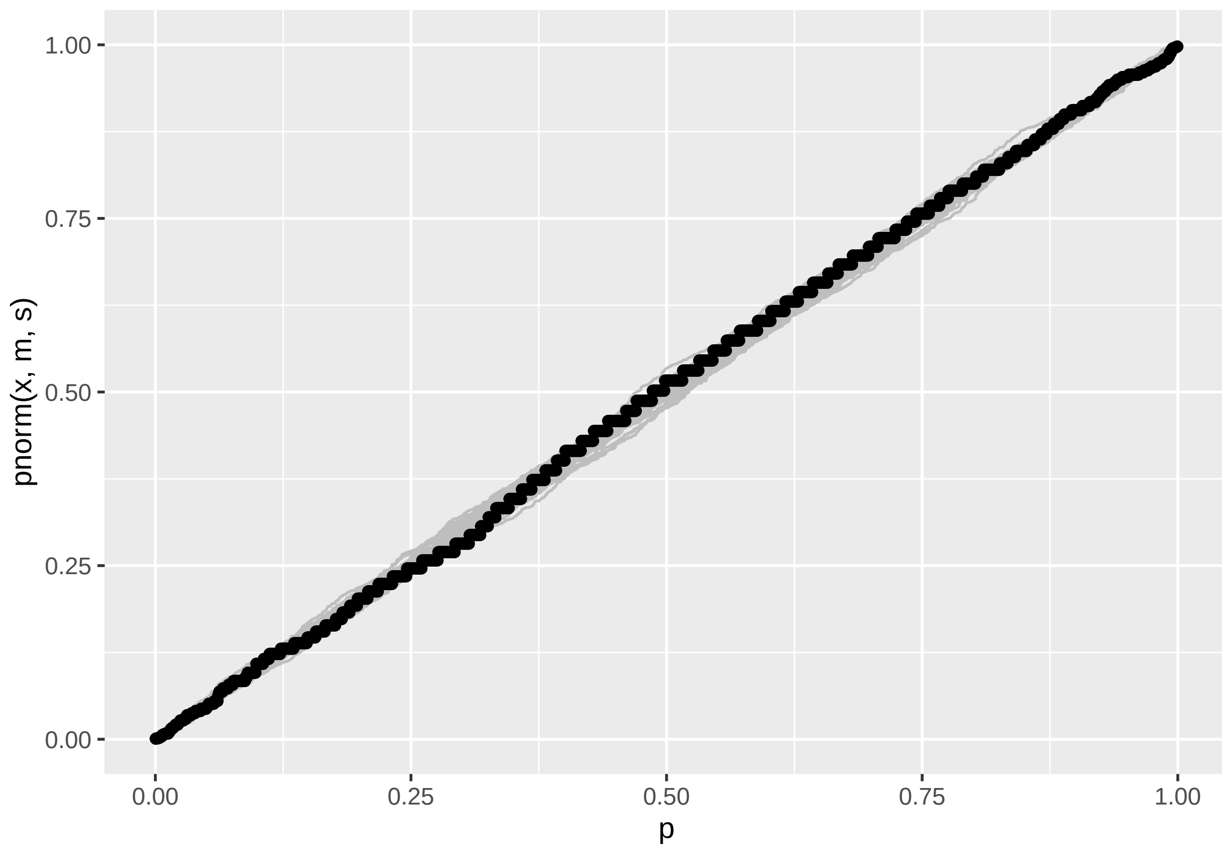
```
ggplot(my_data, aes(sample = Father)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "Q-Q normal plot for father's hights")
```



If the data had a perfect normal distribution, then the points would fall exactly on the diagonal line. Many points are close, especially in the middle section in particular.

Creating a Normal Probability-Probability plot

```
m <- mean(my_data$Father)
s <- sd(my_data$Father)
n <- nrow(my_data)
p <- (1 : n) / n - 0.5 / n
gb <- nboot(my_data$Father, 50)
pp <- ggplot() +
geom_line(aes(x = p, y = pnorm(x, m, s), group = sim),
          color = "gray", data = gb)
pp +
geom_point(aes(x = p, y = sort(pnorm(Father, m, s))), data = (my_data))
```



Shapiro, Anderson-Darling, Cramer-von Mises, Pearson chi-square, Kolmogorov normality test

```
shapiro.test(my_data$Father)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  my_data$Father
## W = 0.99779, p-value = 0.1594
```

```
ad.test(my_data$Father)
```

```
##
##  Anderson-Darling normality test
##
## data:  my_data$Father
## A = 0.45093, p-value = 0.2741
```

```
cvm.test(my_data$Father)
```

```
##
```

```
##  Cramer-von Mises normality test
##
## data:  my_data$Father
## W = 0.06596, p-value = 0.3156
```

```r
pearson.test(my_data$Father)
```

```
##
##  Pearson chi-square normality test
##
## data:  my_data$Father
## P = 76.633, p-value = 5.954e-06
```

```r
ks.test(x = my_data$Father, y = pnorm(nrow(my_data), m, s))
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  my_data$Father and pnorm(nrow(my_data), m, s)
## D = 1, p-value = 0.2705
## alternative hypothesis: two-sided
```