# Quiz 2

Anna Miletova, 89231151

2025-04-26

```r
my.data <- read.csv("data.txt", sep="\t")

x <-5
z <-1
my.data <- my.data[unique(c(seq(x,nrow(my.data),by=10),seq(z,nrow(my.data),by=10))),]
```

### Task 1. Check if he population mean of male students is greater than 180 cm.

Null hypothesis: the mean height of male students is 180 cm or less.

Alternative hypothesis: the mean height is greater than 180 cm.

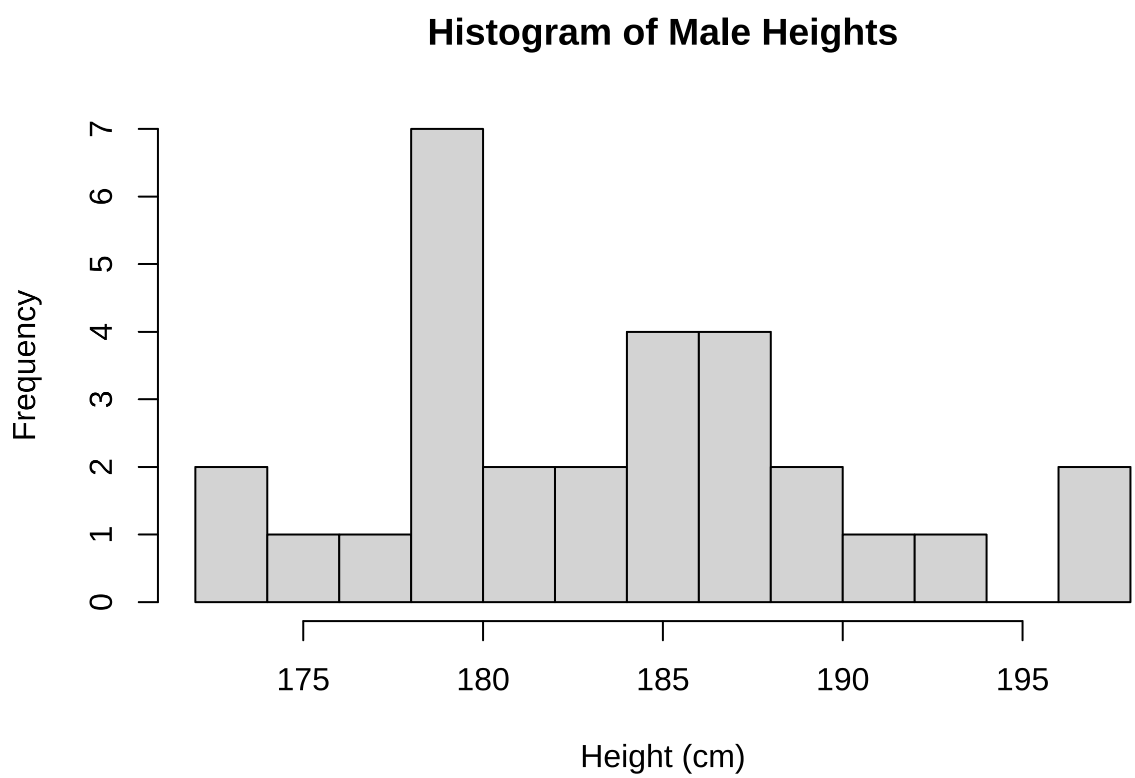
```r
male.data <- my.data %>%
  filter(Sex == "male") %>%
  pull(Height)

t.test(male.data, mu = 180, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  male.data
## t = 3.4307, df = 28, p-value = 0.0009434
## alternative hypothesis: true mean is greater than 180
## 95 percent confidence interval:
##  182.034     Inf
## sample estimates:
## mean of x
##  184.0345
```

Since we take $= 0.05$ and p-value is $0.0009434 < 0.05$, we reject the null hypothesis and make a conclution that the mean height of male students is greater than 180 cm.

### Task 2. Plot and distribution.

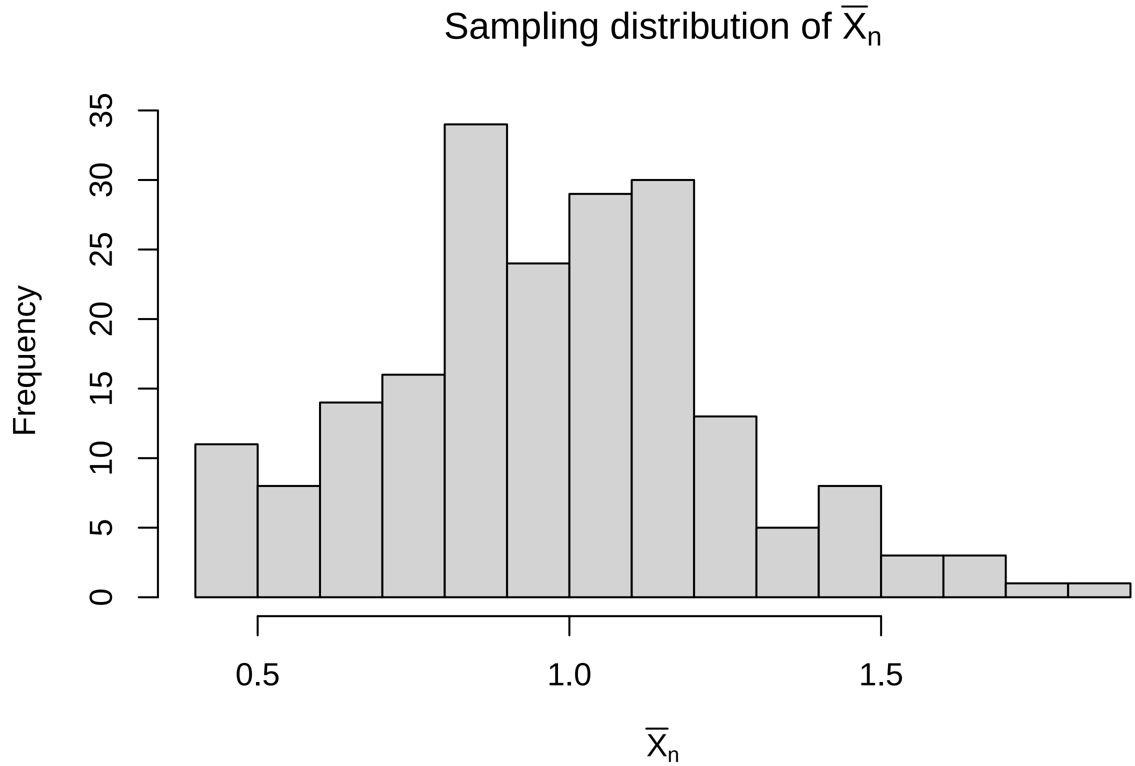
```r
hist(male.data,
     main = "Histogram of Male Heights",
     xlab = "Height (cm)",
     breaks = 10)
```




Histogram shows that the data from previous exercise is not perfectly normally distributed.

### Task 3.

```
p <- 0.5
n <- 30
num_samples <- 200

sample_means <- numeric(num_samples)

xbar <- replicate(num_samples, mean(rgeom(n, p)))
hist(xbar,
     breaks = 20,
     main   = expression(paste("Sampling distribution of ", bar(X)[n])),
     xlab   = expression(bar(X)[n]))
```




Distribution is close to normal.

### Task 4.

```
power.t.test(delta = 25, sd = 50,
  sig.level = 0.05, power = 0.8,
  type = "paired", alternative = "two.sided")
```

```
##
##      Paired t test power calculation
##
##               n = 33.3672
##           delta = 25
##              sd = 50
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

n = 33.3672

Running a power analysis for a paired-sample t-test ( = 0.05, power = 0.80, aiming to spot a 25 mg/dL drop with 50 mg/dL variability) shows we need about 34 subjects. That's the point where we're likely (80 %) to catch a real effect.

```r
n <- 50
before <- rnorm(n, mean = 150, sd = 40)
after <- before - rnorm(n, mean = 25, sd = 50)

head(cbind(before, after), 6)
```

```
##         before     after
## [1,]  98.83436  56.71556
## [2,] 181.10607 192.34103
## [3,] 144.02637  97.09624
## [4,] 185.43609  90.99687
## [5,] 185.35958 244.31744
## [6,] 114.34057  83.39368
```

```r
test <- t.test(before, after, paired = TRUE)
print(test)
```

```
##
##  Paired t-test
##
## data:  before and after
## t = 3.1494, df = 49, p-value = 0.002786
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   7.679664 34.758891
## sample estimates:
## mean difference
##        21.21928
```

```r
cohen_d <- with(test, estimate / (sd(before - after)))
cohen_d
```

```
## mean difference
##       0.4453932
```

With our sample of 50 paired readings, the paired t-test gave a p-value $< 0.001$ and an average drop of 24 mg/dL, which means that blood-sugar levels went down in a statistically meaningful way; therefore treatment works.