Stichprobenverteilug einer Normalverteilung und Intervalschätzung für den Mittelwert

Grundgesamtheit generieren und darstellen

Eine Grundgesamtheit $N(\mu, \sigma^2)$ (Universum) als size_universe Zufallszahlen generieren. Der Mittelwert ist mu, die Standardabeichung ist sigma.

```
size_universe=10000
mu <- 10
sigma <- 2
universe = rnorm(size_universe, mean=mu, sd=sigma)
cat("First 5 random numbers", head(universe, 5), "\n" )
```

```
## First 5 random numbers 7.317525 8.829013 14.19566 6.435334 8.232714
```

Durchschnitt und Standardabweichung aus den Daten "schätzen". Verteilung graphisch darstellen:

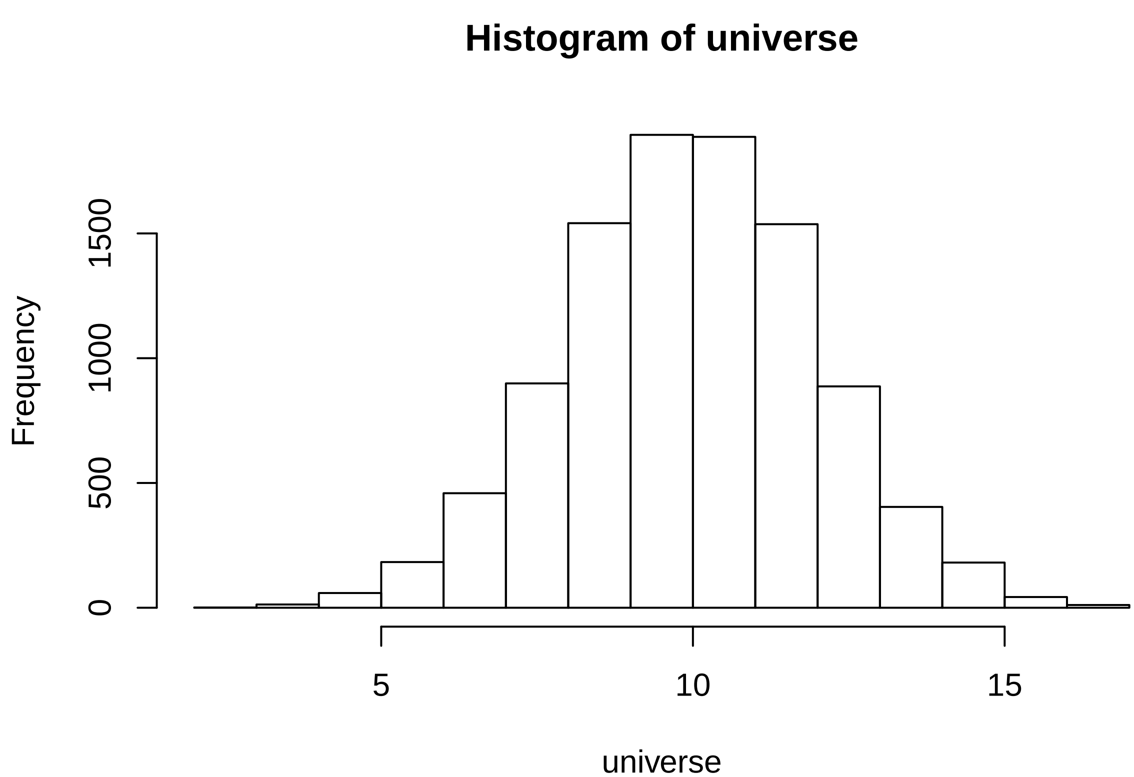
```
cat("Mean of the universe:", mean(universe), "\n")
```

```
## Mean of the universe: 9.968137
```

```
cat("Standard Deviation of the universe:", sd(universe), "\n")
```

```
## Standard Deviation of the universe: 2.007045
```

```
hist(universe)
```



Eine Stichprobe ziehen

Jetzt ziehen wir eine kleine Stichprobe:

```
sample_size <- 16 # Sample size
my_sample <- sample(universe, sample_size)
head(my_sample, 5) # Print the first 5 elements of the sample on the screen
```

```
## [1]  8.976426 10.883864  6.729218 11.756061 10.695479
```

Der Mittelwert und Standardabweichung der Stichprobe sind "nah" zu mu und sigma aber nicht genau gleich:

```
cat("Mean of the sample:", mean(my_sample), "\n")
```

```
## Mean of the sample: 10.06437
```

```
cat("Standard Deviation of the sample:", sd(my_sample), "\n")
```

```
## Standard Deviation of the sample: 1.618137
```

Stichprobeverteilung

Wir können jetzt die Stichprobeziehung mehrmals durchführen. Der Mittelwert der sample_size Zufallsvariabeln ist auch eine Zufallsvariabel. Schauen wir mal, wie diese verteilt ist.:

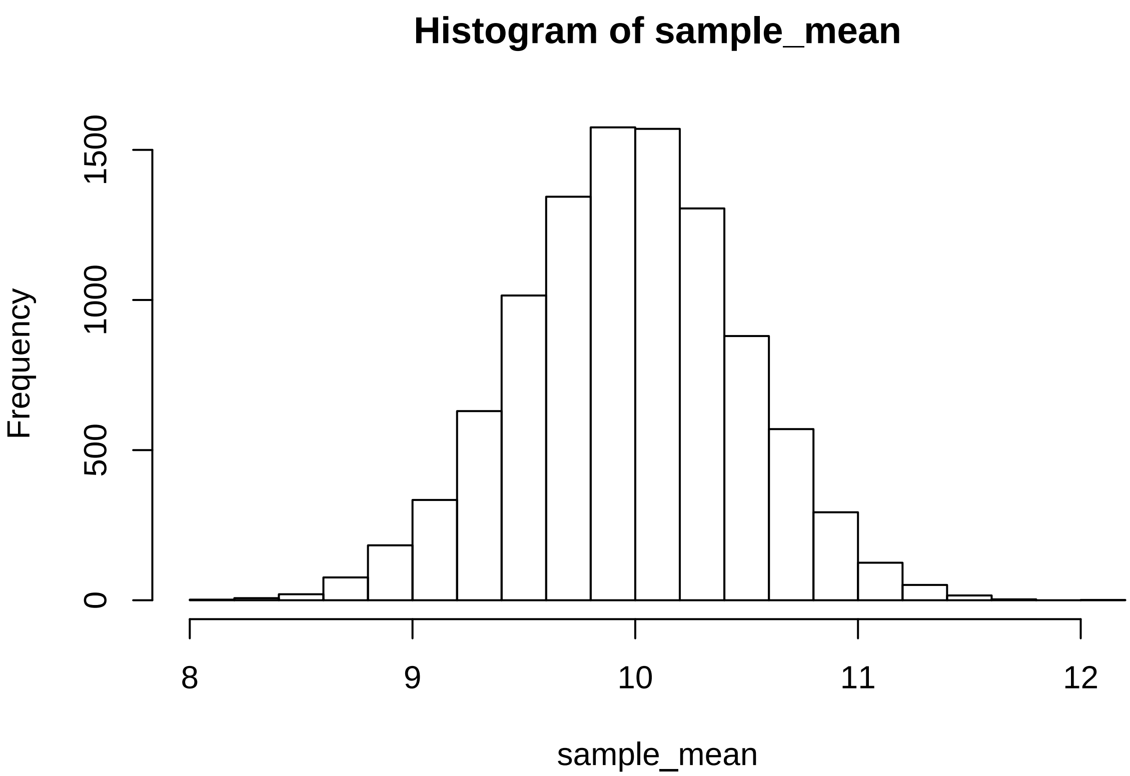
```
n_iterations <-  10000 # Number of times that we repeat
sample_mean <-  NULL
for (i in 1:n_iterations) {
    my_sample <- sample(universe, sample_size)
    x_bar <- mean(my_sample)
    sample_mean <- c(sample_mean, x_bar)
}
cat("Mean of the sampling distribution:", mean(sample_mean), "\n")
```

```
## Mean of the sampling distribution: 9.972593
```

```
cat("Standard Deviation of the sampling distribution:", sd(sample_mean), "\n")
```

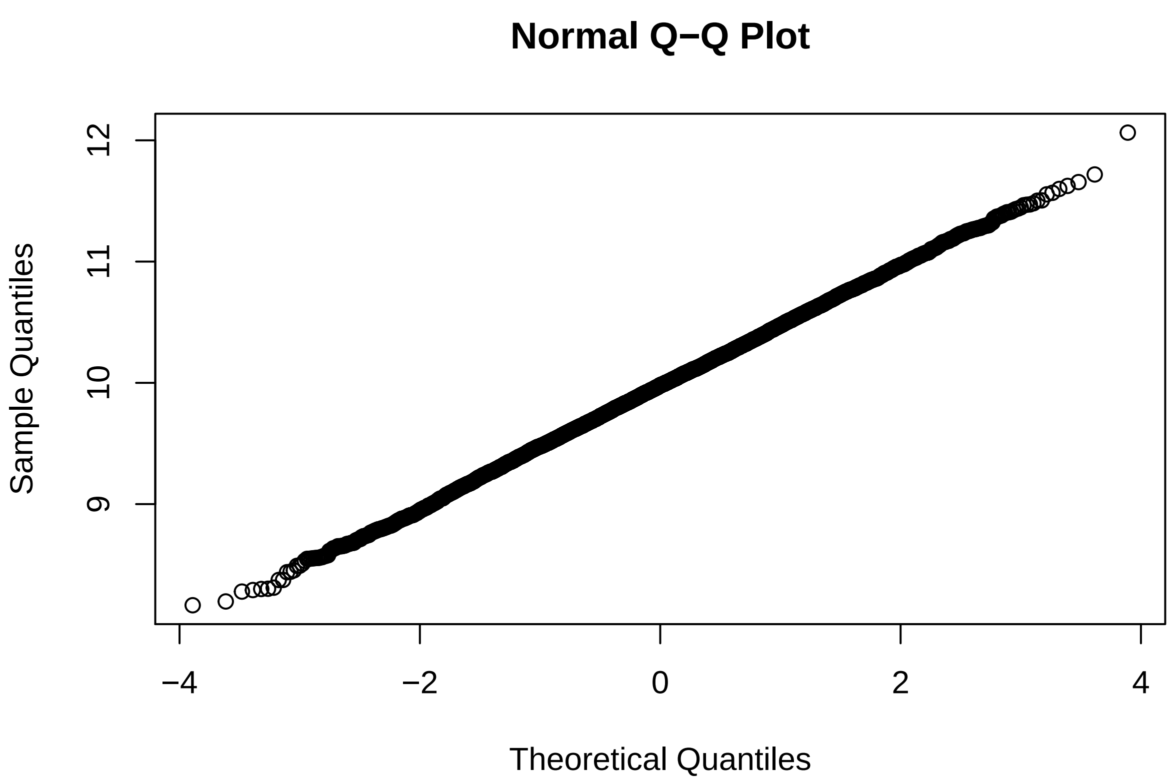
```
## Standard Deviation of the sampling distribution: 0.5004873
```

```
hist(sample_mean)
```



Um zu schauen, ob dieser normalverteilt ist, können wir die qqnorm Funktion verwenden. Fallen die dargestellent Punkte entlang einer Gerade, so ist die Verteilung normal.

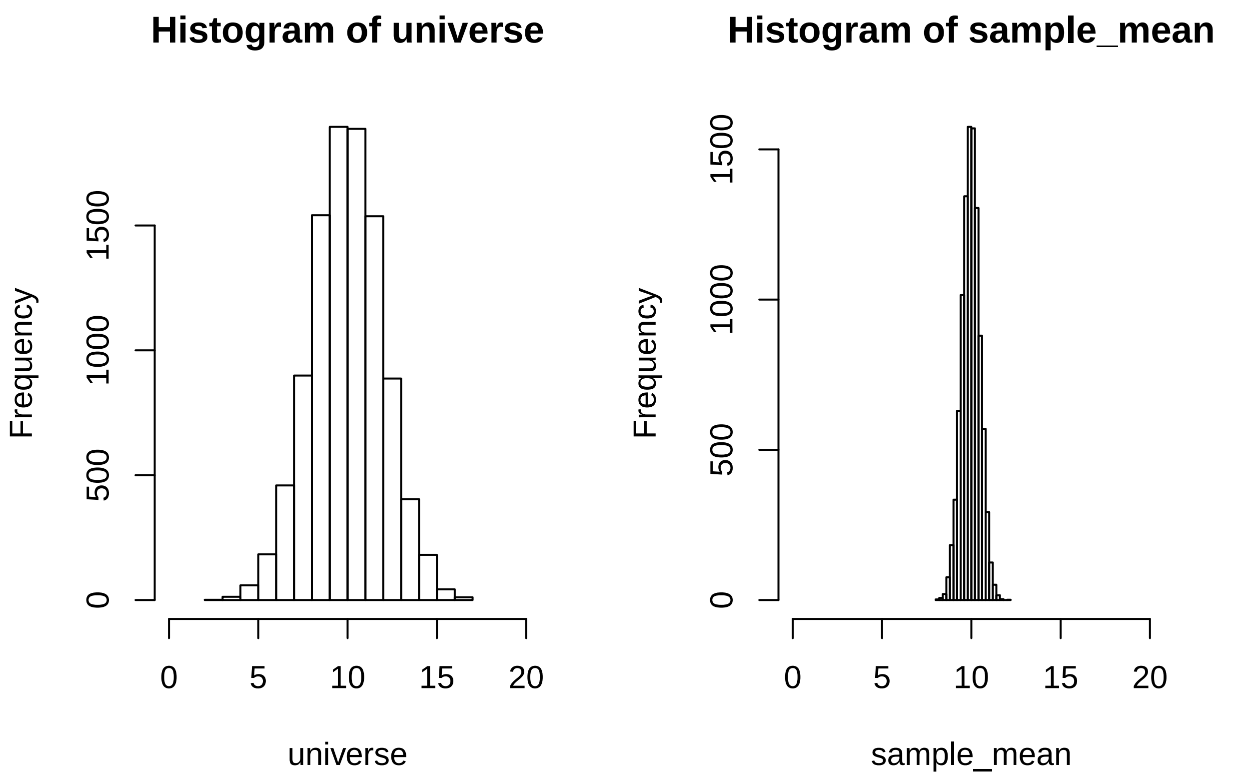
```
qqnorm(sample_mean)
```



Schlussfolgerung (was nutzt uns das Ganze)

Stellen wir das Universum und die Stichprobenverteilung nebeneinander dar:

```
par(mfrow = c(1, 2))
hist(universe, freq = TRUE, xlim = c(mu - 5 * sigma, mu + 5 * sigma))
hist(sample_mean, freq = TRUE, xlim = c(mu - 5 * sigma, mu + 5 * sigma))
```




```
par(mfrow = c(1, 1))
```

Gemäss CLT (Zentralergrenzwertsatz) erwarten wir, dass der (Mittelwert der Stichprobe) $\bar{X}$ wie N(μ, $\frac{\sigma^2}{\sqrt{n}}$) verteilt ist. Ist es so?

```
cat("Mean of the universe:", mean(universe), "\n")
```

```
## Mean of the universe: 9.968137
```

```
cat("Mean of the sampling distribution:", mean(sample_mean), "\n")
```

```
## Mean of the sampling distribution: 9.972593
```

Diese Werte sind praktisch gleich. Das ist gut.

```
cat("Standard Deviation of the sampling distribution:", sd(sample_mean), "\n")
```

```
## Standard Deviation of the sampling distribution: 0.5004873
```

```r
cat("Theoretical value of its standard deviation:", sigma / sqrt(sample_size), "\n")
```

```
## Theoretical value of its standard deviation: 0.5
```

Auch diese Werte sind praktisch gleich. Das ist gut. Die Simulation steht im Einklang zur Theorie.

Anders gezeigt. Die Verteilung $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ist N(mu, 1) verteilt.

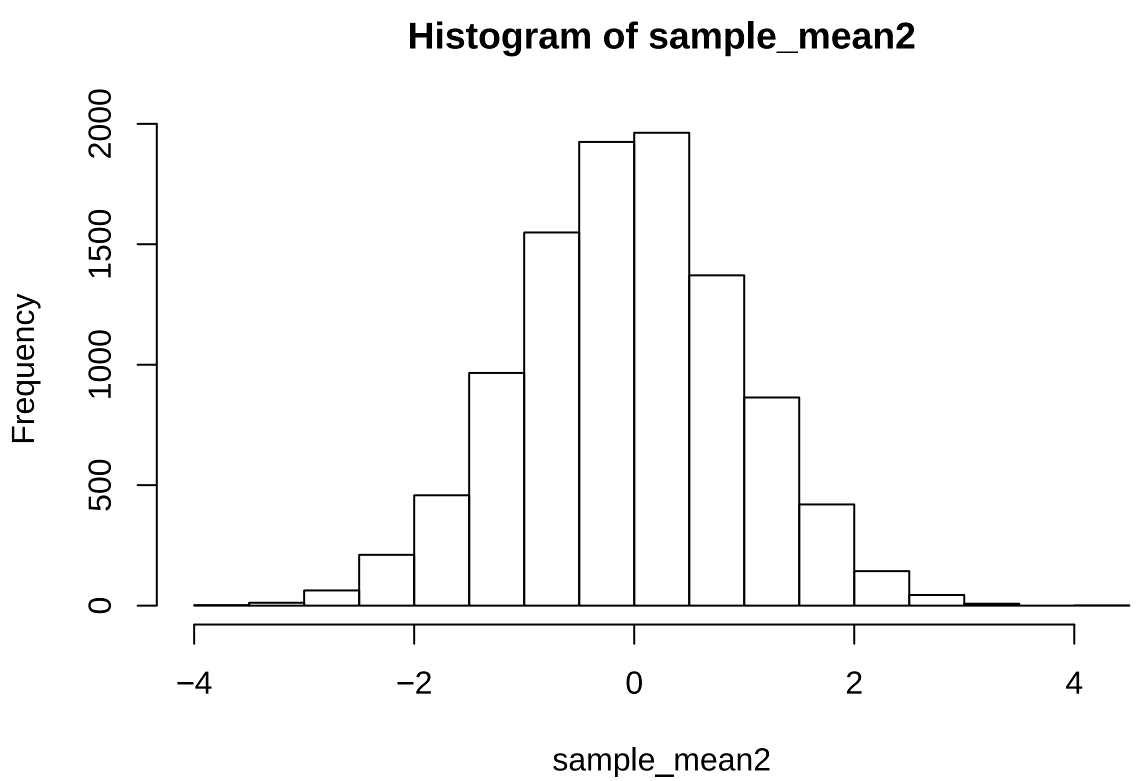
```r
sample_mean2 <- (sample_mean - mu) / (sigma / sqrt(sample_size) )
cat("Mean of sample_mean2:", mean(sample_mean2), "\n")
```

```
## Mean of sample_mean2: -0.054813
```

```r
cat("SD of sample_mean2:", sd(sample_mean2), "\n")
```

```
## SD of sample_mean2: 1.000975
```

```r
hist(sample_mean2)
```



wie aus der Formel erwartet.

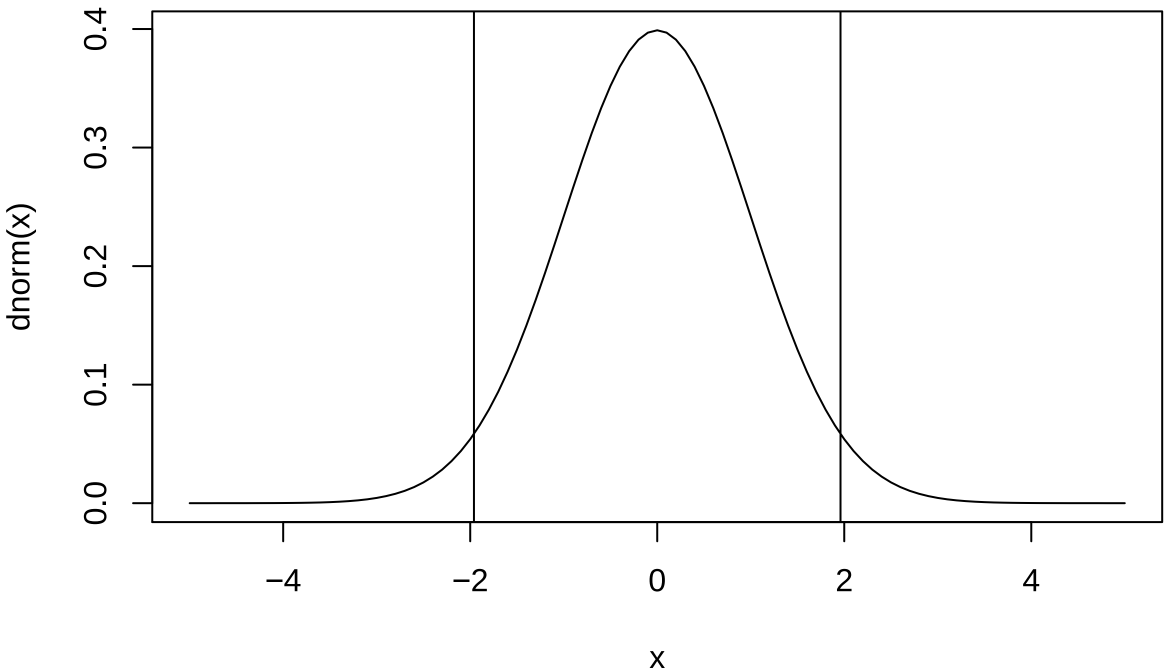
Konfidenzinterval wenn σ bekannt ist

Wir sind jetzt in der Lage, die Schätzung vom μ zu quantifizieren. Setzen wir ein Konfidenzniveau 1 - α für die Schätzung. α ist die Irrtumswahrscheinlichkeit. Angenommen die Standardabweichung der Grundgesamtheit σ ist bekannt, verwenden wir folgende Formel: Das Konfidenzinterval berechnen wir dann als:

$$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Der Wert Z_α berechnet sich als qnorm(1 - alpha / 2) mir R.

```
alpha = 0.05
curve(dnorm(x), from = -5, to = 5)
abline(v = qnorm(alpha / 2)); abline(v = qnorm(1 -alpha / 2))
```



Statt immer wieder Zahlen in die Formel einzustecken (fehleranfällig), können wir folgende Funktion verwenden:

```
# Define a function for the confidence interval for estimating the mean of a normal distribution
# when we have a sample of n data points and we know the mean sigma of the universe
conf.int.z <- function(xbar, sigma, n, alpha) {

    d <- qnorm(1 - alpha / 2) * sigma / sqrt(n)

    cat("With", 1-alpha, "certainty the mean mu lies between", xbar, "+/-", d,  "\n")
    cat("With", 1-alpha, "certainty the mean mu lies between", xbar -d, "and", xbar + d, "\n")

    return(d)

}


# Test the function. Example 10.4, Dürr, Meyer.
conf.int.z(xbar = 80.50, sigma = 2.2, n = 40, alpha = 0.05)
```

```
## With 0.95 certainty the mean mu lies between 80.5 +/- 0.6817745
## With 0.95 certainty the mean mu lies between 79.81823 and 81.18177

## [1] 0.6817745
```