Confidence Interval for μ (Known Population σ)

Wir suchen den μ, und wissen die Standardabweichung der Population σ

$$\left[\bar{X}_{(n)} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Falls wir ein Sample haben, können wir den z-Test nutzen. Wenn nicht, müssen wir die Formel per Hand benutzen.

Nur Two sided

```r
library(TeachingDemos)
sample <- c(8, 9, 10, 13, 14, 16, 17, 20, 21)
sample_mean <- mean(sample)
pop_sd <- 2.8
alpha <- 0.05
q <- qnorm(1 - alpha / 2)
n <- length(sample)
L <- sample_mean - q * (pop_sd / sqrt(n))
U <- sample_mean + q * (pop_sd / sqrt(n))
z.test(x = sample, stdev = pop_sd, alternative =
    "two.sided", conf.level = 1-alpha)$conf.int
z.test(x = sample_mean, n = n, stdev = pop_sd,
    alternative = "two.sided", conf.level = 1-
    alpha)$conf.int # Wenn sample.mean und n
```

Nur Upper ODER Lower

Wir müssen alpha NICHT mehr teilen, da sich die Prozente auf eine Seite konzentrieren.

```r
L_alleine <- sample_mean - qnorm(1 - alpha) * (
    sample_sd/sqrt(n))
U_alleine <- sample_mean + qnorm(1 - alpha) * (
    sample_sd/sqrt(n))
```

Umformungen

$\bar{X}_{(n)}$: Für den Sample Mean $\bar{X}_{(n)}$ Umstellen

$$\bar{X}_{(n)} = \frac{\text{obere Grenze} + \text{untere Grenze}}{2}$$

```r
sample_mean_umgestellt <- (L + U) / 2
```

$u_{1-\frac{\alpha}{2}}$: Für Quantile der Normalverteilung Umformen

Aus der Intervalllänge:

$$u_{1-\frac{\alpha}{2}} = \frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot \sigma}$$

Aus der oberen Grenze:

$$u_{1-\frac{\alpha}{2}} = \frac{\text{obere Grenze} - \bar{X}_{(n)}}{\frac{\sigma}{\sqrt{n}}}$$

Aus der unteren Grenze:

$$u_{1-\frac{\alpha}{2}} = \frac{\bar{X}_{(n)} - \text{untere Grenze}}{\frac{\sigma}{\sqrt{n}}}$$

```r
z <- (leange * sqrt(n)) / 2 * sampel_sd
q_umgestellt <-(U - sample_mean) / (pop_sd /
    sqrt(n))
q_umgestellt <-(sample_mean - L) / (pop_sd /
    sqrt(n))
```

σ: Für die Standardabweichung σ

Aus der oberen Grenze:

$$\sigma = \frac{\left(\text{obere Grenze} - \bar{X}_{(n)}\right) \cdot \sqrt{n}}{u_{1-\frac{\alpha}{2}}}$$

Aus der unteren Grenze:

$$\sigma = \frac{\left(\bar{X}_{(n)} - \text{untere Grenze}\right) \cdot \sqrt{n}}{u_{1-\frac{\alpha}{2}}}$$

```r
pop_sd_umgestellt <- ((U - sample_mean) * sqrt(n
    )) / qnorm(1 - alpha / 2)
pop_sd_umgestellt <- ((sample_mean - L) * sqrt(n
    )) / qnorm(1 - alpha / 2)
```

n: Für die Stichprobengröße n

Aus der oberen oder unteren Grenze:

$$n = \left(\frac{\sigma \cdot u_{1-\frac{\alpha}{2}}}{\text{obere Grenze} - \bar{X}_{(n)}}\right)^2 = \left(\frac{\sigma \cdot u_{1-\frac{\alpha}{2}}}{\bar{X}_{(n)} - \text{untere Grenze}}\right)^2$$

$$n \geq \left(\frac{2 \cdot u_{1-\frac{\alpha}{2}} \cdot \sigma}{\text{Länge}}\right)^2$$

```r
n_umgestellt <- round(((pop_sd * q) / (U -
    sample_mean))^2)
n_umgestellt <- round(((pop_sd * q) / (
    sample_mean - L))^2)
n <- ceiling(((2 * q * sd) / length)^2)
n_aus_moe <- ceiling(((q * pop_sd) / moe)^2)
```

(MOE): Mit Margin of Error

$$MOE = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$n = \left(\frac{u_{1-\frac{\alpha}{2}} \cdot \sigma}{MOE}\right)^2$$

```r
moe <- qnorm(1 - alpha / 2) * (pop_sd / sqrt(n))
```

α: Für das Signifikanzniveau α UND Confidence Level

$$\alpha = 2 \cdot \left(1 - \Phi(u_{1-\frac{\alpha}{2}})\right)$$

$$\alpha = 2 \cdot \left(1 - \Phi\left(\frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot \sigma}\right)\right)$$

⚠ Um den Confidence Level zu bekommen: 1 - alpha ⚠

```r
z_value <- (length * sqrt(n)) / (2 * sigma)
alpha <- 2 * (1 - pnorm(z_value))
```

Intervalllänge: Formel für die Intervalllänge

$$\text{Intervalllänge} = U - L$$

$$\text{Intervalllänge} = 2 \cdot E = 2 \cdot u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

```r
intervallleange <- 2 * qnorm(1 - alpha / 2) * (
    sample_sd / sqrt(n))
```

MOE mit Intervalllänge

$$E = \frac{\text{Intervalllänge}}{2} = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Interval for μ (Unknown Population σ)

Wir suchen den μ, und wissen die Standardabweichung des Samples $S_{(n)}$

$$\left[\bar{X}_{(n)} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_{(n)}}{\sqrt{n}}, \bar{X}_{(n)} + t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_{(n)}}{\sqrt{n}}\right]$$

Falls wir ein Sample haben, können wir den t-Test nutzen.

Nur Two sided

```r
library(TeachingDemos)
sample <- c(8, 9, 10, 13, 14, 16, 17, 20, 21)
n <- length(sample)
sample_mean <- mean(sample)
sample_sd <- sd(sample)
alpha <- 0.05
t <- qt(1 - (alpha / 2), n - 1)
L <- sample_mean - t * (sample_sd / sqrt(n))
U <- sample_mean + t * (sample_sd / sqrt(n))
t.test(x = sample, conf.level = 1 - alpha,
    alternative = 'two.sided')
```

Nur Upper oder Lower

Wir müssen alpha nicht mehr teilen, da sich die Prozente auf eine Seite konzentrieren

```r
L_alleine <- sample_mean - qt(1 - alpha, n - 1)
    * (sample_sd / sqrt(n))
U_alleine <- sample_mean + qt(1 - alpha, n - 1)
    * (sample_sd / sqrt(n))
```

Umformungen

$\bar{X}_{(n)}$: Für den Sample Mean $\bar{X}_{(n)}$ Umstellen

$$\bar{X}_{(n)} = \frac{\text{obere Grenze} + \text{untere Grenze}}{2}$$

```r
sample_mean_umgestellt <- (L + U) / 2
```

$t_{1-\frac{\alpha}{2},n-1}$: Für Quantile der t-Verteilung $t_{1-\frac{\alpha}{2},n-1}$

Aus der Intervalllänge:

$$t_{1-\frac{\alpha}{2},n-1} = \frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot S_{(n)}}$$

Aus der oberen Grenze:

$$t_{1-\frac{\alpha}{2},n-1} = \frac{\text{obere Grenze} - \bar{X}_{(n)}}{\frac{S_{(n)}}{\sqrt{n}}}$$

Aus der unteren Grenze:

$$t_{1-\frac{\alpha}{2},n-1} = \frac{\bar{X}_{(n)} - \text{untere Grenze}}{\frac{S_{(n)}}{\sqrt{n}}}$$

```r
t_umgestellt_1 <- (U - L) * sqrt(n) / (2 *
    sample_sd)
t_umgestellt_2 <- (U - sample_mean) / (sample_sd
    / sqrt(n))
t_umgestellt_3 <- (sample_mean - L) / (sample_sd
    / sqrt(n))
```

$S_{(n)}$: Für die Sample Standardabweichung $S_{(n)}$

Aus der oberen Grenze:

$$S_{(n)} = \frac{\left(\text{obere Grenze} - \bar{X}_{(n)}\right) \cdot \sqrt{n}}{t_{1-\frac{\alpha}{2},n-1}}$$

Aus der unteren Grenze:

$$S_{(n)} = \frac{\left(\bar{X}_{(n)} - \text{untere Grenze}\right) \cdot \sqrt{n}}{t_{1-\frac{\alpha}{2},n-1}}$$

```r
sample_sd1 <- (U - sample_mean) * (sqrt(n)) / t
sample_sd2 <- (sample_mean - L) * (sqrt(n)) / t
```

n: Für die Sample size n

Aus der oberen Grenze:

$$n = \left(\frac{S_{(n)} \cdot t_{1-\frac{\alpha}{2},n-1}}{\text{obere Grenze} - \bar{X}_{(n)}}\right)^2 = \left(\frac{S_{(n)} \cdot t_{1-\frac{\alpha}{2},n-1}}{\bar{X}_{(n)} - \text{untere Grenze}}\right)^2$$

$$n \geq \left(\frac{2 \cdot t_{1-\frac{\alpha}{2}} \cdot \sigma}{\text{Länge}}\right)^2$$

$$n = \left(\frac{t_{1-\frac{\alpha}{2},n-1} \cdot S_{(n)}}{MOE}\right)^2$$

```r
n_umgestellt_1 <- round(((sample_sd * t_value) /
    (U - sample_mean))^2)
n_umgestellt_2 <- round(((sample_sd * t_value) /
    (sample_mean - L))^2)
n <- ceiling(((2 * t * sample_sd) / length)^2)
n_aus_moe <- ceiling(((t_value * sample_sd) /
    moe)^2)
```

(MOE): Mit Margin of Error

$$MOE = t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_{(n)}}{\sqrt{n}}$$

```r
moe <- t_value * (sample_sd / sqrt(n))
```

α: Für das Signifikanzniveau α

$$\alpha = 2 \cdot \left(1 - \Phi\left(t_{1-\frac{\alpha}{2},n-1}\right)\right)$$

$$\alpha = 2 \cdot \left(1 - \Phi\left(\frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot s}\right)\right)$$

Um den Confidence Level zu bekommen: 1 - alpha

```r
new_t <- (length*sqrt(n)) /(2 * sample_sd)
alpha <- 2 * (1-pt(new_t, n - 1))
```

Intervalllänge: Formel für die Intervalllänge

$$\text{Intervalllänge} = U - L$$

$$\text{Intervalllänge} = 2 \cdot MOE = 2 \cdot t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_{(n)}}{\sqrt{n}}$$

```r
length <- U - L
length <- 2 * qt(1 - (alpha / 2), df = n - 1) *
    (sample_sd / sqrt(n))
```

MOE mit Intervalllänge

$$MOE = \frac{\text{Intervalllänge}}{2} = t_{1-\frac{\alpha}{2},n-1} \cdot \frac{S_{(n)}}{\sqrt{n}}$$

```r
moe_aus_intervalllaenge <- intervalllaenge / 2
```

Confidence Interval for Variance σ^2, mean μ_0 known:

Wir suchen die Variance σ^2, und kennen den Mean der Population μ

$$\left[\frac{Q_{(n)}}{\chi^2_{n;1-\frac{\alpha}{2}}}, \frac{Q_{(n)}}{\chi^2_{n;\frac{\alpha}{2}}}\right] \quad \text{with} \quad Q_{(n)} = \sum_{i=1}^{n}(X_i - \mu_0)^2$$

Wir brauchen ein Sample oder einen Wert für $Q_{(n)}$.

```
sample <- c(247.4, 249.0, 248.5, ..., 249.4)
mean <- 250
alpha <- 0.05
n <- length(sample) #20
qn <- sum((sample - mean)^2)
L_var <- qn / (qchisq(1 - (alpha / 2),n))
U_var <- qn / qchisq(alpha / 2, n)
```

Confidence Interval for σ^2, mean μ_0 UNKNOWN:

Wir suchen die Variance σ^2, und kennen den Mean der Population μ NICHT

$$\left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Wir brauchen N und Sample sd.
⚠ sigma.test funktioniert nur mit Sample ⚠

```
sample <- c(247.4, 249.0, 248.5, ..., 249.4)
alpha <- 0.05
sample_sd <- sd(sample)
n <- length(sample) #20
b <- (n - 1) * sample_sd^2
L_var <- b / qchisq(1 - (alpha / 2), n-1)
U_var <- b / qchisq(alpha / 2, n - 1)
sigma.test(x = sample, conf.level = 1 - alpha,
    alternative = 'two.sided')
```

1 Biased Estimator

Question:
Show that the relative frequency is an unbiased point estimators for the proportion of voters preferring A in the whole population.

Answer:
Let p be the probability that a randomly chosen voter supports A. Then X = number of voters preferring A in a sample of size n follows a B(n, p)-distribution.
The expected value of the relative frequency
p.hat = x / n is n * p.ha / n = p.ha, ie p.hat is an unbiased estimator of p

Confidence Interval for $\hat{p}$ (Proportion)(Stichprobenanteil)

We are estimating $\hat{p}$, the sample proportion

$$\left[\hat{p} - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$\hat{p} = \frac{x}{n}$$

Info nur für den Mert Kaan::::::

$u_{1-\frac{\alpha}{2}}$ = qnorm(1 - (alpha/2))

```r
prob <- 0.7
alpha <- 0.05
n <- 250
p_hut <- (prob*n)/n
q <- qnorm(1-(alpha/2))
L <- p_hut - q * sqrt((p_hut * (1-p_hut))/n)
U <- p_hut + q * sqrt((p_hut * (1-p_hut))/n)
binom.test(x=0.7 * n, n=n, conf.level = 1-alpha,
    alternative = "two.sided")$conf.int
#[0.6390569, 0.7561285]
```

Einschätzung: Wir sind uns zu 95% sicher, dass zwischen 64% und 75% der Wähler ja gestimmt haben.

Nur obere oder untere Grenze berechnen:

Hier teilen wir α nicht mehr, da sich die Prozente auf eine Seite konzentrieren:

```r
q <- qnorm(1 - alpha)
L_alleine <- p_hut - q * sqrt((p_hut * (1 -
    p_hut))/n)
L_exact <- binom.test(x=0.7*n, n=n, conf.level =
    1-alpha, alternative = 'greater')$conf.int
U_alleine <- p_hut + q * sqrt((p_hut * (1 -
    p_hut))/n)
U_exact <- binom.test(x=0.7*n, n=n, conf.level =
    1-alpha, alternative = 'less')$conf.int
```

Umformungen

$\hat{p}$: Solving for $\hat{p}$

$$\hat{p} = \frac{\text{upper limit} + \text{lower limit}}{2}$$

```r
p_hut <- (L + U) / 2
```

α: Für alpha umstellen

⚠ Um das Confidence Level zu bekommen: 1 - alpha ⚠

$$\alpha = 2 \cdot \left(1 - \Phi\left(u_{1-\frac{\alpha}{2}}\right)\right)$$

$$\alpha = 2 \cdot \left(1 - \Phi\left(\frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot \sqrt{\hat{p}(1-\hat{p})}}\right)\right)$$

```r
z_value <- (leange * sqrt(n)) /
  (2 * sqrt(p_hut * (1 - p_hut)))
alpha <- 1-(2 * (1 - pnorm(z_value)))
confidence_level <- 2 * pnorm(length/(2*sqrt(
    p_hut*(1-p_hut)/n)))-1
```

$u_{1-\frac{\alpha}{2}}$: Für Quantile der Normalverteilung Umformen

Aus der Intervalllänge:

$$u_{1-\frac{\alpha}{2}} = \frac{\text{Intervalllänge} \cdot \sqrt{n}}{2 \cdot \sqrt{\hat{p}(1-\hat{p})}} =$$

Aus der oberen und unteren Grenze:

$$u_{1-\frac{\alpha}{2}} = \frac{\hat{p} - \text{untere Grenze}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{\text{obere Grenze} - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

```r
#nach q umstellen
leange <- U - L
q_umgestellt_1 <- (leange * sqrt(n)) / (2 * sqrt
    (p_hut * (1 - p_hut)))
q_umgestellt_2 <- (U - p_hut) / (sqrt(p_hut * (1
    - p_hut) / n))
q_umgestellt_3 <- (p_hut - L) / (sqrt(p_hut * (1
    - p_hut) / n))
```

MOE (Margin of Error):

$$MOE = u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{Intervalllänge}{2}$$

```r
moe <- q * (sqrt(p_hut * (1 - p_hut) / n))
moe_2 <- laenge / 2
```

Intervalllänge:

$$\text{Intervalllänge} = 2 \cdot u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = U - L$$

```r
length <- 2 * q * (sqrt(p_hut * (1 - p_hut) / n)
    )
length <- U - L
```

Sample Länge n:

```r
n <- ceiling((qnorm(1-alpha/2) / length)^2)
```

$\hat{p}$ wird auf 0.5 gesetzt, wenn wir den Wert nicht kennen.

Wenn wir einen Two sided Confidence Interval haben dann müssen wir unser alpha teilen:

$$n = \frac{(u_{1-\frac{\alpha}{2}})^2 \cdot \hat{p}(1-\hat{p})}{MOE^2}$$

Wenn wir einen Upper oder Lower Bound haben:

$$n = \frac{(u_{1-\alpha})^2 \cdot \hat{p}(1-\hat{p})}{MOE^2}$$