$$\mu = \mathbb{E}[X] = \sum_{i \geq 1} x_i \mathbb{P}[X = x_i] \qquad \sigma^2 = \mathbb{VAR}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[(X - \mu)^2\right]$$

$$\gamma = \frac{\mathbb{E}\left[(X-\mathbb{E}[X])^3\right]}{(\mathbb{VAR}[X])^{\frac{3}{2}}} = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \qquad \kappa = \frac{\mathbb{E}\left[(X-\mathbb{E}[X])^4\right]}{(\mathbb{VAR}[X])^2} = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

$$\begin{aligned} \mathbb{E}[a] &= a, \\ \mathbb{VAR}[a] &= 0, \\ \mathbb{VAR}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2, \\ \mathbb{E}[aX + bY] &= a\mathbb{E}[X] + b\mathbb{E}[Y], \\ \mathbb{VAR}[aX + bY] &= a^2\mathbb{VAR}[X] + 2ab\mathbb{COV}[X, Y] + b^2\mathbb{VAR}[Y], \end{aligned}$$

$$\begin{aligned} \mathbb{COV}[a, Z] &= 0, \\ \mathbb{COV}[X, Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y], \\ \mathbb{COV}[X, X] &= \mathbb{VAR}[X], \\ \mathbb{COV}[Y, X] &= \mathbb{COV}[X, Y], \\ \mathbb{COV}[aX + bY, Z] &= a\mathbb{COV}[X, Z] + b\mathbb{COV}[Y, Z]. \end{aligned}$$

	2.5%	97.5%
(Intercept)	-73.48864	711.26388
area	10.46882	19.05758

Hence we are 95 percent confident that the rent of an apartment in Vaduz increases between 10.46882[CHF/Mth] and 19.05758[CHF/Mth] if its area increases by 1[sqm]. The null hypothesis $H_0 : \beta_1 = 0$ is significantly rejected since it is not included in this 95% confidence interval for β_1.

Scaling of variables can be used to choose an appropriate model. Note the different interpretations of the parameter β_1 as

- constant marginal effect in the LIN-LIN-model $y = \beta_0 + \beta_1 x + u$ (if x increases c.p. by one unit, y changes by β_1 units),
- constant semi-elasticity in the LOG-LIN-model $\log(y) = \beta_0 + \beta_1 x + u$ (if x increases c.p. by one unit, y changes approximately by $\beta_1 \cdot 100\%$ or exactly by $(\exp(\beta_1) - 1) \cdot 100\%$),
- constant elasticity in the LOG-LOG-model $\log(y) = \beta_0 + \beta_1 \log(x) + u$ (if x increases c.p. by 1%, y changes by $\beta_1\%$)

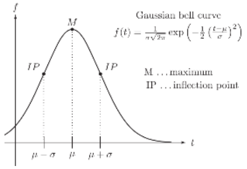
Calculation of the variance of $X \sim BER(p)$ using the definition:

$$\begin{aligned} \mathbb{E}[X] &= 1 \cdot p + 0 \cdot (1-p) = p, \\ \mathbb{VAR}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) \\ &= \left((1-p)p + (0-p)^2\right)(1-p) = p(1-p). \end{aligned}$$

$$\mathbb{P}_p[S_n = k] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is **normally distributed** with expectation μ and variance σ^2 if its probability density function is of the form

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} = \frac{1}{\sigma}\phi\left(\frac{t-\mu}{\sigma}\right)$$



The standardization of a normal distribution leads to a standard normal distribution:

$$X \sim \mathcal{N}(\mu, \sigma^2) \Longrightarrow \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

Using standardizations and statistical tables we find the probabilities

$$\begin{aligned} \mathbb{P}[X \leq -2] &= \mathbb{P}\left[\frac{X-\mu}{\sigma} \leq \frac{-2-3}{4}\right] = \Phi(-1.25) = 1 - \Phi(1.25) \\ &= 1 - 0.8944 = 0.1056 = 10.56\%, \end{aligned}$$

Similarly we find the 95%-quantile $q_{0.95}$:

$$\mathbb{P}[X \leq q_{0.95}] = 95\% \Longrightarrow \mathbb{P}\left[\frac{X-\mu}{\sigma} \leq \frac{q_{0.95} - 3}{4}\right] = 0.95$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \Longrightarrow \mathbb{E}[X] = \mu \text{ and } \mathbb{VAR}[X] = \sigma^2$$

Moreover the skewness and the kurtosis are given by

$$X \sim \mathcal{N}(\mu, \sigma^2) \Longrightarrow \gamma[X] = 0 \text{ and } \kappa[X] = 3.$$

The sample variance of a normally distributed random sample is after some adjustment χ^2 distributed:

$$\chi^2_{\nu, 1-\alpha}.$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \text{ with } \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \text{ and } X_i \sim \mathcal{IN}(\mu, \sigma^2)$$

$$\Longrightarrow \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{SS_X}{\sigma^2} \sim \chi^2_{n-1}.$$

$$\widehat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{k \cdot S_{xx}}\right)$$

$$k = 4: \quad \widehat{\beta}_1 \sim \mathcal{N}\left(3, \frac{16}{4 \cdot 10}\right) = \mathcal{N}\left(3, 0.632456^2\right)$$

$$k = 4: \quad f(t) = \frac{1}{0.632456\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-3}{0.632456}\right)^2}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i \text{ and } S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2$$

$$T = \frac{Z}{\sqrt{X/\nu}} \quad \text{with} \quad Z \sim \mathcal{N}(0, 1),\ X \sim \chi^2_\nu \text{ independent}$$

is **t distributed with ν degrees of freedom**, i.e. $T \sim t_\nu$, where its $(1-\alpha)$ quantile is denoted by $t_{\nu,1-\alpha}$ and has the symmetry property $t_{\nu,\alpha} = -t_{\nu,1-\alpha}$.

If the unknown population variance is substituted by the sample variance in order to standardize the mean of a normal sample we get a t distribution:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \text{ and } S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \text{ with } X_i \sim \mathcal{IN}(\mu, \sigma^2)$$

$$\Longrightarrow \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t_{n-1}.$$

$$T \sim t_\nu \Longrightarrow \begin{cases} \mathbb{E}[T] = 0, & \\ \mathbb{VAR}[T] = \frac{\nu}{\nu - 2} & \text{for } \nu > 2, \\ \gamma[T] = 0, & \\ \kappa[T] = \frac{3\nu - 6}{\nu - 4} & \text{for } \nu > 4. \end{cases}$$

The random variable

$$F = \frac{X_1/m}{X_2/n} \quad \text{with} \quad X_1 \sim \chi^2_m, X_2 \sim \chi^2_n \text{ independent}$$

is **F distributed with m degrees of freedom in the numerator and n degrees of freedom in the denominator**, i.e. $F \sim F_{m,n}$.

The $(1-\alpha)$ quantile of $F \sim F_{m,n}$ is denoted by $F_{m,n,1-\alpha}$ and has the symmetry property $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$ or $F_{m,n,\alpha} = \frac{1}{F_{n,m,1-\alpha}}$.

$$\left.\begin{array}{lll} X_1, X_2, \ldots, X_m & \text{with} & X_i \sim \mathcal{N}(\mu_X, \sigma_X^2), \\ Y_1, Y_2, \ldots, Y_n & \text{with} & Y_j \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \end{array}\right\} \text{independent,}$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \overline{X})^2 \text{ and } S_Y^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$$

$$\Longrightarrow F = \frac{\frac{m-1}{\sigma_X^2} S_X^2}{\frac{n-1}{\sigma_Y^2} S_Y^2} = \frac{SS_X/\sigma_X^2}{SS_Y/\sigma_Y^2} \sim F_{m-1,n-1}.$$

The **binomially distributed random variable** $X \sim BIN(n, p)$ counts the number of successes in a Bernoulli chain with length n and probability of success p so that its probability mass function is given by

$$\mathbb{P}[X = k] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \text{for } k \in \{0, 1, \ldots, n\}$$

and $\mathbb{P}[X = k] = 0$ for $k \notin \{0, 1, \ldots, n\}$. Due to this definition we can express $X \sim BIN(n, p)$ as

$$X = X_1 + X_2 + \ldots + X_n = \sum_{k=1}^n X_i \quad \text{with} \quad X_i \sim IBER(p)$$

where X_i is an indicator for a success in the i-th trail. **Hence binomial distributions typically show up in statistics if we want to infer from a random sample on the proportion of individuals in the population having a certain property.**

$$X \sim BIN(n, p) \Longrightarrow \begin{cases} \mathbb{E}[X] = n \cdot p, \\ \mathbb{VAR}[X] = n \cdot p \cdot (1-p), \\ \gamma[X] = \frac{1-2 \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}, \\ \kappa[X] = 3 - \frac{6}{n} + \frac{1}{np(1-p)}. \end{cases}$$

If n is large we can calculate $\mathbb{P}[x_1 < X \leq x_2]$ for $X \sim BIN(n, p)$ - at least approximately - as follows:

1. **Check the rule of thumb** $\sigma^2 = np(1-p) > 9$ for "good" approximations that are useful in practice.
2. **Standardize the random variable** in the expression for the probability:
$$\mathbb{P}[x_1 < X \leq x_2] = \mathbb{P}\left[\frac{x_1 - np}{\sqrt{np(1-p)}} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - np}{\sqrt{np(1-p)}}\right].$$
3. **Express the probability in terms of the Gaussian integral function**
$$\mathbb{P}[x_1 < X \leq x_2] \approx \Phi\left(\frac{x_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{x_1 - np}{\sqrt{np(1-p)}}\right).$$

$X \sim BIN(500, p)$ with $p \leq 5\%$

$$\begin{aligned} \mathbb{P}[X > 30 | p = 5\%] &= 1 - \mathbb{P}[X \leq 30 | p = 5\%] \\ &= 1 - \mathbb{P}\left[\frac{X-\mu}{\sigma} \leq \frac{30 - 500 \cdot 0.05 + 0.5}{\sqrt{500 \cdot 0.05 \cdot 0.95}} \middle| p = 5\%\right] \\ &\approx 1 - \Phi(1.129) = 1 - 0.8706 = 0.1294 = 12.94\%. \end{aligned}$$

We can evaluate the probabilities $\mathbb{P}[X = k] = \binom{10}{k} \cdot 0.2^k \cdot 0.8^{10-k}$ in

---

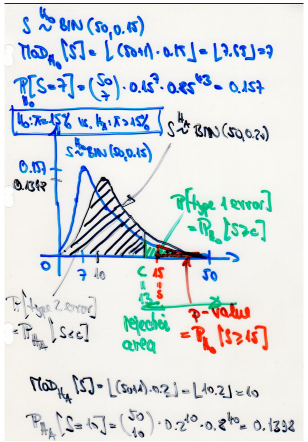
		decision based on empirical data	
		H_0 not rejected	H_0 rejected
unknown	H_0 true	OK	**type 1 error**
reality	H_0 false	**type 2 error**	OK

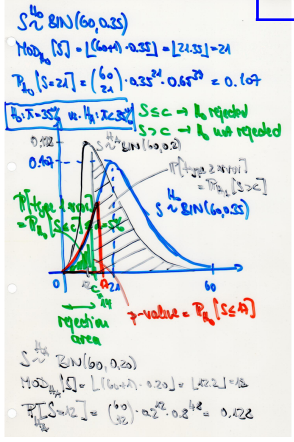
$\mathbb{P}[\text{type 1 error}] \leq \alpha$

In this sense we only learn something from the data about the population of interest if the null hypothesis is rejected.

The "null hypothesis" typically stands for "no effect", "no difference" or "no change". Hence the "alternative" stands for an "effect", a "difference" or a "change".

The **power** $1 - \beta$ denotes the probability that the null hypothesis will be rejected when it is actually false and means the probability to significantly prove an "effect", a "difference" or a "change" when it is really present.





We assume the data structure

$$X_1, X_2, \ldots, X_n \text{ i.i.d. with } X_i \sim BER(\pi)$$

The hypotheses can be formulated in terms of the population parameter:

	(1)	(2)	(3)
H_0 :	$\pi \overset{(\leq)}{=} \pi_0$	$\pi \overset{(\geq)}{=} \pi_0$	$\pi = \pi_0$
H_A :	$\pi > \pi_0$	$\pi < \pi_0$	$\pi \neq \pi_0$

The given context indicates whether the alternative is one-sided as in (1) and (2) or two-sided as in (3).

The test statistic

$$S = X_1 + X_2 + \ldots + X_n \sim BIN(n, \pi)$$

counts the number of units in the random sample that have the property of interest so that its distribution under the null hypothesis is therefore given by $S \overset{H_0}{\sim} BIN(n, \pi_0)$.

The critical values[a] c_l and c_u in the decision rules

(1)	(2)	(3)	
$S \geq c_u$	$S \leq c_l$	$S \leq c_l$ or $S \geq c_u$	H_0 rejected
$S < c_u$	$S > c_l$	$c_l < S < c_u$	H_0 not rejected

are defined by the condition that the probability of a type 1 error is limited to the significance level:

(1)	(2)	(3)
$\mathbb{P}_{H_0}[S \geq c_u] \leq^* \alpha$	$\mathbb{P}_{H_0}[S \leq c_l] \leq^* \alpha$	$\mathbb{P}_{H_0}[S \leq c_l] \leq^* \frac{\alpha}{2}$ and $\mathbb{P}_{H_0}[S \geq c_u] \leq^* \frac{\alpha}{2}$

corresponding confidence interval:

(1)	$\left[\widehat{\pi} - z_{1-\alpha} \cdot \frac{\sqrt{\widehat{\pi}(1-\widehat{\pi})}}{\sqrt{n}}, 1\right]$
(2)	$\left[0, \widehat{\pi} + z_{1-\alpha} \cdot \frac{\sqrt{\widehat{\pi}(1-\widehat{\pi})}}{\sqrt{n}}\right]$
(3)	$\left[\widehat{\pi} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\widehat{\pi}(1-\widehat{\pi})}}{\sqrt{n}}, \widehat{\pi} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{\widehat{\pi}(1-\widehat{\pi})}}{\sqrt{n}}\right]$

$n \cdot p$

$n \cdot p \cdot (1-p)$

determines the minimal sample size

$$n = \left(\frac{z_{1-\alpha}\sqrt{\pi_0(1-\pi_0)} - z_\beta\sqrt{\pi_A(1-\pi_A)}}{\pi_A - \pi_0}\right)^2$$

Election Consultant

Test Statistic and its Distribution:

For the test statistic

$$S = X_1 + X_2 + \ldots + X_n \sim BIN(n, \pi),$$

e.g. **the number of persons in the random sample that would vote for the party one year after the last election**, we can specify its distribution

$$S \overset{H_0}{\sim} BIN(50, 0.15)$$

under the null hypothesis.

Decision Rule:

$S \geq c \Longrightarrow H_0$ rejected
$S < c \Longrightarrow H_0$ not rejected

Condition for the Critical Value:

$$\mathbb{P}[\text{type 1 error}] \leq 5\% \Longleftrightarrow \mathbb{P}_{H_0}[S \geq c] \leq^* 5\%$$

$$\Longleftrightarrow \sum_{i=c}^{50} \binom{50}{i} \cdot 0.15^i \cdot 0.85^{50-i} \leq^* 5\%$$

c	$\mathbb{P}_{H_0}[S \geq c]$	remark
20	0.000016	$< 5\% \Rightarrow$ decr
10	0.2089	$> 5\% \Rightarrow$ incre
15	0.005287	
13	0.030061	$\leq 5\%$ for the
12	0.06281	$> 5\%$ for the

$$\text{p-value} = \mathbb{P}_{H_0}[S \geq 15] = \sum_{i=15}^{50} \binom{50}{i} \cdot 0.15^i \cdot 0.85^{50-i} = 0.005287$$

$X_i \sim \mathcal{IN}(\mu, \sigma^2)$ with number of viewers X_i on day i in sample of size $n = 40$

In a t-test setup a data structure of the type $X_i \sim \mathcal{IN}(\mu, \sigma^2)$ is assumed so that the sample mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

is normally distributed.

Assuming $\sigma^2 = 12^2$ and simplifying to a z test scenario, we can express the distribution of the sample mean $\overline{X}$ under the null hypothesis $H_0 : \mu = 5$ and under the alternative $H_A : \mu = 8$ as

$$\overline{X}|H_0 \sim \mathcal{N}\left(5, \frac{12^2}{n}\right) \quad \text{and} \quad \overline{X}|H_A \sim \mathcal{N}\left(8, \frac{12^2}{n}\right)$$

We can express the critical value using the distribution of the sample mean under H_0:

$$\mathbb{P}[\overline{X} > c|H_0] = 0.05 \Longrightarrow 1 - \mathbb{P}\left[\frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \le \frac{c-5}{\frac{12}{\sqrt{n}}}\middle| H_0\right] = 0.05$$

$$\Longrightarrow 1 - \Phi\left(\frac{c-5}{\frac{12}{\sqrt{n}}}\right) = 0.05 \Longrightarrow \Phi\left(\frac{c-5}{\frac{12}{\sqrt{n}}}\right) = 0.95$$

$$\Longrightarrow \frac{c-5}{\frac{12}{\sqrt{n}}} = 1.645 \Longrightarrow c = 5 + 1.645 \cdot \frac{12}{\sqrt{n}}.$$

With a special version of a F test we can **compare the variances of two normal populations**.

Data Structure:

$$\left.\begin{array}{ll} X_i \sim \mathcal{N}(\mu_X, \sigma_X^2) & (i = 1, \dots, m) \\ Y_j \sim \mathcal{N}(\mu_Y, \sigma_Y^2) & (j = 1, \dots, n) \end{array}\right\} \text{independent}$$

Hypotheses:

	(1)	(2)	(3)
H_0 :	$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 = \sigma_Y^2$	$\sigma_X^2 = \sigma_Y^2$
H_A :	$\sigma_X^2 > \sigma_Y^2$	$\sigma_X^2 < \sigma_Y^2$	$\sigma_X^2 \neq \sigma_Y^2$

Test Statistic: **Empirical value**

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{m-1,n-1} \Longrightarrow F \overset{H_0}{=} \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

Decision Rules:

	H_0 rejected	H_0 not rejected
(1)	$F > F_{m-1,n-1,1-\alpha}$	$F \le F_{m-1,n-1,1-\alpha}$
(2)	$F < F_{m-1,n-1,\alpha}$	$F \ge F_{m-1,n-1,\alpha}$
(3)	$F > F_{m-1,n-1,1-\frac{\alpha}{2}}$ or $F < F_{m-1,n-1,\frac{\alpha}{2}}$	$F_{m-1,n-1,\frac{\alpha}{2}} \le F \le F_{m-1,n-1,1-\frac{\alpha}{2}}$

With a t test for two independent samples we can **test hypotheses about the difference $\delta = \mu_X - \mu_Y$ between the mean μ_X of a first normal population and the mean μ_Y of a second normal population**, where the variances σ_X^2 and σ_Y^2 are unknown.

Data Structure:

$$\left.\begin{array}{ll} X_i \sim \mathcal{N}(\mu_X, \sigma_X^2) & (i = 1, \dots, m) \\ Y_j \sim \mathcal{N}(\mu_Y, \sigma_Y^2) & (j = 1, \dots, n) \end{array}\right\} \text{independent}$$

Hypotheses:

If the variances of the two populations are the same, e.g. $\sigma_X^2 = \sigma_Y^2$, we use the test statistic

$\sigma_{big}/\sigma_{small} \le 2$

$$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}} \overset{H_0}{\sim} t_{m+n-2}. \qquad (1)$$

If the variances of the two populations are not the same, e.g. $\sigma_X^2 \neq \sigma_Y^2$, we use the test statistic

$\sigma_{big}/\sigma_{small} > 2$

$$T = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}} \overset{H_0}{\sim} t_k \text{ with } k = \left\lfloor \frac{\left(\frac{S_X^2}{m} + \frac{S_Y^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2} \right\rfloor. \qquad (2)$$

The floor $\lfloor x \rfloor$ of x denotes the largest integer that is not greater than x.

Decision Rules:

(1)	(2)	(3)	
$T > t_{\nu,1-\alpha}$	$T < t_{\nu,\alpha}$	$\lvert T\rvert > t_{\nu,1-\frac{\alpha}{2}}$	H_0 rejected
$T \le t_{\nu,1-\alpha}$	$T \ge t_{\nu,\alpha}$	$\lvert T\rvert \le t_{\nu,1-\frac{\alpha}{2}}$	H_0 not rejected

with $\nu = m + n - 2$ under the assumption $\sigma_1^2 = \sigma_2^2$
and $\nu = k$ under the assumption $\sigma_1^2 \neq \sigma_2^2$

P-Values:

(1)	(2)	(3)
$\mathbb{P}_{H_0}[T \ge t] \le \alpha$	$\mathbb{P}_{H_0}[T \le t] \le \alpha$	$\mathbb{P}_{H_0}[\lvert T\rvert \ge \lvert t\rvert] \le \alpha$

P-Value:

(1)	(2)	(3)
$\mathbb{P}_{H_0}[T \ge t] \le \alpha$	$\mathbb{P}_{H_0}[T \le t] \le \alpha$	$\mathbb{P}_{H_0}[\lvert T\rvert \ge \lvert t\rvert] \le \alpha$

Confidence Interval for the Population Mean μ:

(1)	$\left[\overline{x} - t_{n-1,1-\alpha} \cdot \frac{s}{\sqrt{n}}, \infty\right[$
(2)	$\left]-\infty, \overline{x} + t_{n-1,1-\alpha} \cdot \frac{s}{\sqrt{n}}\right]$
(3)	$\left[\overline{x} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$

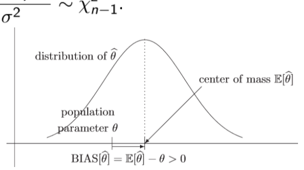
Input: DATA of a RANDOM SAMPLE drawn from the Population
ESTIMATOR in the sense of a General Formula
Output: ESTIMATE for the unknown parameter

Denote θ a single population parameter and $\boldsymbol{\theta}$ a set of population parameters.

$$\widehat{\mu} = m = \overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n). \quad \widehat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

$$\frac{\widehat{\mu} - \mu}{\widehat{\sigma}/\sqrt{n}} \sim t_{n-1} \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

$E[\hat{\pi}] = \pi$, $E[\hat{\mu}] = \mu$



$$\text{BIAS}[\widehat{\sigma}^2] = \mathbb{E}[\widehat{\sigma}^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.$$

$$\mathbb{V}AR[\widehat{\pi}] = \mathbb{V}AR\left[\frac{S}{n}\right] = \left(\frac{1}{n}\right)^2 \cdot \underbrace{\mathbb{V}AR[S]}_{=n\cdot\pi\cdot(1-\pi)} = \frac{\pi\cdot(1-\pi)}{n}$$

$$\Longrightarrow \sigma[\widehat{\pi}] = \frac{\sqrt{\pi\cdot(1-\pi)}}{\sqrt{n}} \Longrightarrow \text{SE}(\widehat{\pi}) = \frac{\sqrt{\widehat{\pi}\cdot(1-\widehat{\pi})}}{\sqrt{n}}.$$

$MSE[\hat{\theta}] = (BIAS[\hat{\theta}])^2 + \mathbb{V}AR[\hat{\theta}]$

Using the independence assumption, the calculations rules for variances and the variances $\mathbb{V}AR[\overline{X}] = \frac{\sigma_X^2}{n_1}$ and $\mathbb{V}AR[\overline{Y}] = \frac{\sigma_Y^2}{n_2}$ of the estimators $\widehat{\mu}_X = \overline{X}$ and $\widehat{\mu}_Y = \overline{Y}$, we find

$$\mathbb{V}AR[\widehat{\delta}] = \mathbb{V}AR[\overline{X} - \overline{Y}] = \mathbb{V}AR[\overline{X}] + \mathbb{V}AR[\overline{Y}] = \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}.$$

Since the variance $\mathbb{V}AR[\widehat{\delta}]$ vanishes asymptotically for increasing sample sizes n_1 and n_2, e.g. $\mathbb{V}AR[\widehat{\delta}] \longrightarrow 0$ for $n_1 \longrightarrow \infty$ and $n_2 \longrightarrow \infty$, the unbiased estimator $\widehat{\delta}$ is consistent.

$$\left[\$540 - 2.132 \cdot \frac{\$299.1}{\sqrt{5}}, \infty\right[= [\$254.8, \infty[$$

for the population mean that is bounded from below. We are 95% confident that the mean repair costs after a 10 miles per hour head on collision are at least \$254.8. Since the sample size is extremely small the accuracy of the confidence level depends heavily on the normality assumption.

sample size. **When the population in not normally distributed, the procedures are approximations. Nonetheless, we find that sample sizes of 30 or greater will provide good results in most cases. If the population is approximately normal, small sample sizes (e.g. $n < 15$) can provide acceptable results. If the population is highly skewed or contains outliers, sample sizes approaching 50 are recommended.** *(p. 356)*

x: Independent/explanatory variable
y: dependent/explained variable

(SLR.1)	$y = \beta_0 + \beta_1 x + u$	population model
(SLR.2)	(y_i, x_i) $(i = 1, \dots, n)$ with $y_i = \beta_0 + \beta_1 x_i + u_i$	random sample of size n following the population mode
(SLR.3)	$\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0$	the sample outcomes on x are not all the same
(SLR.4)	$\mathbb{E}[u\vert x] = 0$	zero conditional mean
(SLR.5)	$\mathbb{V}AR[u\vert x] = \sigma^2$	homoskedasticity
(SLR.6)	$u\vert x \sim Normal(0, \sigma^2)$	normal distributed error

$$\mathbb{E}[y|x] = \beta_0 + \beta_1 x$$

and its derivative

$$\frac{d\mathbb{E}[y|x]}{dx} = \beta_1,$$

Month	Time $x[h]$	Variable Costs $K_v[10^6 CHF]$
January	5	3
February	6	2
March	5	2
April	10	5
May	8	4
June	7	4

Technically, the value for the slope b in $K_v(x) = bx$ that minimizes the **global error**

$$GE(b) = (3 - b\cdot 5)^2 + (2 - b\cdot 6)^2 + (2 - b\cdot 5)^2 + (5 - b\cdot 10)^2 + (4 - b\cdot 8)^2 + (4 - b\cdot 7)^2$$

as the sum of the squared vertical deviations $e_i = K_{vi} - bx_i$ of the data points from the straight line through the origin delivers the ordinary least squares estimate $\widehat{\beta}$ for β:

$$\frac{dGE}{db}(b) = 2(3 - b\cdot 5)(-5) + 2(2 - b\cdot 6)(-6) + 2(2 - b\cdot 5)(-5) + 2(5 - b\cdot 10)(-10) + 2(4 - b\cdot 8)(-8) + 2(4 - b\cdot 7)(-7)$$
$$= 598b - 294$$
$$\Longrightarrow \frac{dGE}{db}(\widehat{\beta}) = 0 \Longrightarrow 598\widehat{\beta} - 294 = 0 \Longrightarrow \widehat{\beta} = \frac{147}{299} \approx 0.492.$$

SLR Model:

$$y = \beta_0 + \beta_1 x + u$$

Optimization Problem:

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{b_0, b_1} \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2$$

First Order Conditions or **Normal Equations**:

$$0 = \sum_{i=1}^{n} 2\cdot(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i))\cdot(-1) \qquad (1)$$
$$0 = \sum_{i=1}^{n} 2\cdot(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i))\cdot(-x_i) \qquad (2)$$

$$\Longrightarrow \widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i \cdot y_i - n\cdot\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\cdot\overline{x}^2}$$

Population Model:

$$y = \beta_0 + \beta_1 x + u$$

Ordinary Least Squares Estimators:

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} \qquad (3)$$
$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (4)$$

The **sample regression model**

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

x	y	$x - \overline{x}$	$y - \overline{y}$	$(x - \overline{x})^2$	$(x - \overline{x})(y - \overline{y})$
3	2	$3 - 4 = -1$	$2 - 4 = -2$	$(-1)^2 = 1$	$(-1)\cdot(-2) = 2$
3	4	$3 - 4 = -1$	$4 - 4 = 0$	$(-1)^2 = 1$	$(-1)\cdot 0 = 0$
6	6	$6 - 4 = 2$	$6 - 4 = 2$	$2^2 = 4$	$2\cdot 2 = 4$
12	12	0	0	6	6

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{6}{6} = 1,$$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x} = 4 - 1\cdot 4 = 0$$

The sample regression model can be used to calculate **the fitted values**

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

and the **residuals**

$$\widehat{u}_i = y_i - \widehat{y}_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i).$$

Using the OLS-estimators, the **sum of squared residuals**

$$SSR = \sum_{i=1}^{n}\widehat{u}_i^2 = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = GE(\widehat{\beta}_0, \widehat{\beta}_1)$$

gets as small as possible by construction.

$$\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{SSR}{n-2} \qquad (5)$$

x	y	$\widehat{y} = 0 + 1\cdot x$	$\widehat{u} = y - \widehat{y}$	$\widehat{u}^2$
3	2	$0 + 1\cdot 3 = 3$	$2 - 3 = -1$	$(-1)^2 = 1$
3	4	$0 + 1\cdot 3 = 3$	$4 - 3 = 1$	$1^2 = 1$
6	6	$0 + 1\cdot 6 = 6$	$6 - 6 = 0$	$0^2 = 0$
12	12	12	0	2

$$\overset{(5)}{\Longrightarrow} \widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{1}{3-2}\cdot 2 = 2$$
$$\Longrightarrow \widehat{\sigma} = \sqrt{2} = 1.414$$

$$\Longrightarrow \mathbb{E}[\widehat{\beta}_1|x] = \beta_1 \Longrightarrow \mathbb{E}[\widehat{\beta}_1] = \beta_1 \;/\; \mathbb{E}[\widehat{\sigma}^2] = \sigma^2.$$

$$\widehat{\mathbb{V}AR}[\widehat{\beta}_0|x] = \widehat{\sigma}^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right),$$
$$\widehat{\mathbb{V}AR}[\widehat{\beta}_1|x] = \frac{\widehat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Orthogonality Relations:

$$\sum_{i=1}^{n}\widehat{u}_i = 0, \qquad (6)$$
$$\sum_{i=1}^{n}x_i\widehat{u}_i = 0, \qquad (7)$$
$$\sum_{i=1}^{n}\widehat{y}_i\widehat{u}_i = 0. \qquad (8)$$

Under the **Gauss-Markov assumptions** (SLR.1) to (SLR.5) the least squares estimators are the **best linear unbiased estimators (BLUE)** for β_0 and β_1: There are no other estimators in the class of linear in y_i and unbiased estimators for these parameters that have smaller variances than the OLS estimators. This efficiency theorem justifies the use of the OLS method rather than using a variety of competing estimating techniques.

Under the classical linear model (CLM) assumptions (SLR.1) to (SLR.6) the OLS estimators are the **minimum variance unbiased estimators (MVUE)** for β_0 and β_1: There are no other estimators in the class of unbiased estimators for these parameters that have smaller variances than the OLS estimators.

Analysis of Variance:
The relationship

$$SST = SSE + SSR$$

between

- $SST = \sum_{i=1}^{n}(y_i - \overline{y})^2$ - the **total sum of squares**,
- $SSE = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2$ - the **explained sum of squares**, and
- $SSR = \sum_{i=1}^{n}\widehat{u}_i^2$ - the **residual sum of squares**

shows the decomposition of the variation in the y-data.

The R^2-**coefficient** or **coefficient of determination**

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \Longrightarrow 0 \le R^2 \le 1$$

is a measure for the **goodness-of-fit** since it denotes the proportion of variation in the data for the dependent variable that can be explained by the model and that are in this way related to the variations in the data for the independent variable.

Values $R^2 \approx 1$ indicate that the adapted line fits the data very well and that the model explains a large share of the variations in the explained variable.

Values $R^2 \approx 0$ indicate that the adapted line fits the data very badly and that the model explains a small share of the variations in the explained variable.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Quotient
Regression	SSE	1	$MSE = \frac{SSE}{1}$	$F = \frac{MSE}{MSR}$
Residual	SSR	$n-2$	$MSR = \frac{SSR}{(n-2)}$	
Total	SST	$n-1$		

$$\left[\widehat{\beta}_0 - t_{n-2,1-\frac{\alpha}{2}} se[\widehat{\beta}_0], \widehat{\beta}_0 + t_{n-2,1-\frac{\alpha}{2}} se[\widehat{\beta}_0]\right],$$
$$\left[\widehat{\beta}_1 - t_{n-2,1-\frac{\alpha}{2}} se[\widehat{\beta}_1], \widehat{\beta}_1 + t_{n-2,1-\frac{\alpha}{2}} se[\widehat{\beta}_1]\right],$$
$$\left[\frac{(n-2)\widehat{\sigma}^2}{\chi^2_{n-2,1-\frac{\alpha}{2}}}, \frac{(n-2)\widehat{\sigma}^2}{\chi^2_{n-2,\frac{\alpha}{2}}}\right]$$

for the model parameters β_0, β_1 and σ^2 at the confidence level $(1-\alpha)$.

Confidence Interval for the Mean Response $\mathbb{E}[y|x_0]$ of y given x_0:

$$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 \pm t_{n-2,1-\frac{\alpha}{2}}\widehat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

Prediction Interval for a Future Observation y_0 given x_0:

$$\widehat{\beta}_0 + \widehat{\beta}_1 x_0 \pm t_{n-2,1-\frac{\alpha}{2}}\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

```
Residuals:
    Min      1Q  Median      3Q     Max
-333.81 -146.18  -88.97  196.71  550.24

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  318.888    187.469   1.701    0.105
area          14.763      2.052   7.195  7.8e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 262.5 on 19 degrees of freedom
Multiple R-Squared: 0.7315, Adjusted R-squared: 0.7174
F-statistic: 51.77 on 1 and 19 DF, p-value: 7.795e-07
```

$$\widehat{rent} = \underset{(187.469)}{318.888} + \underset{(2.052)}{14.763} area$$
with $n = 21$, $R^2 = 0.7315$ and $\widehat{\sigma} = 262.5$.

According to the F-quotient $F = 51.77$ on 1 and 19 degrees of freedom with the p-value$= 7.795\cdot 10^{-7} < \alpha = 5\%$, the null hypothesis in the overall significance test is rejected.[a] Hence the SLR model $rent = \beta_0 + \beta_1 area + u$ explaines the rent of apartments significantly better than the constant only model $rent = \beta_0 + u$.

The value of the F-quotient can as well be found in the ANOVA-table:

Response: rent

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
area	1	3567438	3567438	51.774	7.795e-07	***
Residuals	19	1309186	68905			

According to the R-squared coefficient $R^2 = 0.7315$, 73.15 percent of the variations in the rents of the observed apartments can be explained by the variations of their areas via the optimally adapted simple linear regression

The variable *area* is significant because of its p-value $7.8\cdot 10^{-7}$. (The null hypothesis $H_0 : \beta_1 = 0$ is rejected in favour of its alternative $H_A : \beta_1 \neq 0$.) Therefore we could prove statistically (beyond a reasonable doubt) that there is an effect of the area of an apartment on its rent.

According to $\widehat{\beta}_1 = 14.763$, we estimate that the rent of an apartment in Vaduz increases by $14.763[CHF/Mth]$ if its area increases by $1[sqm]$. Hence the effect of the area on the rent is practically important.