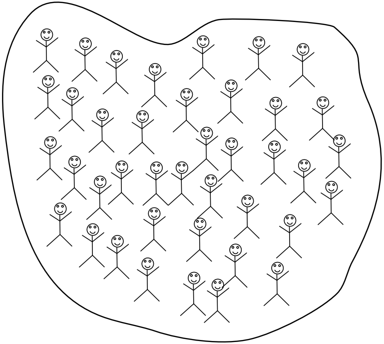
UNIVERSITY OF TWENTE.

Statistical Techniques

Confidence Intervals (Ch 3.1 and 3.2)

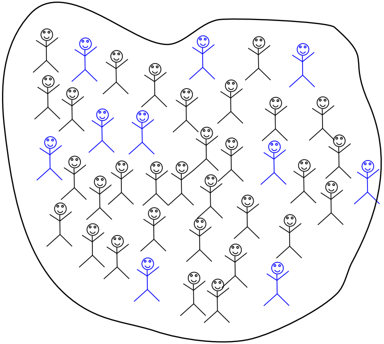
Motivation

Population



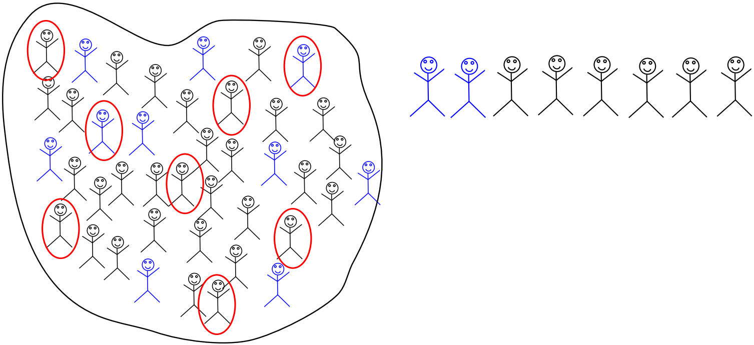
Motivation

Population



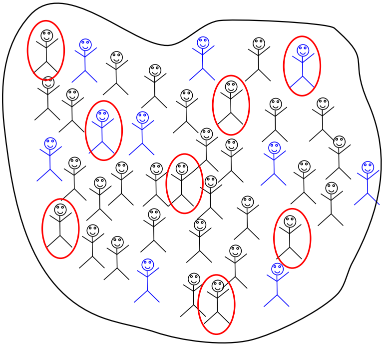
Motivation

Population **Sample 1**

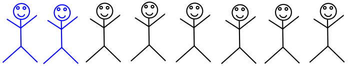


Motivation

Population



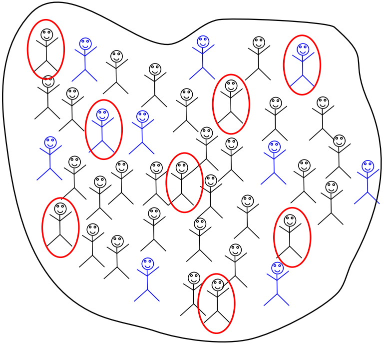
Sample 1



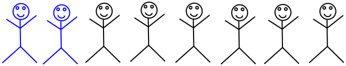
- $\hat{p} = 0.25$

Motivation

Population



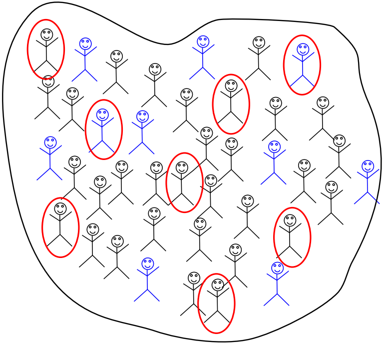
Sample 1



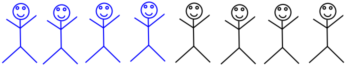
- $\hat{p} = 0.25$
- Average IQ: $\bar{x} = 107.8$

Motivation

Population

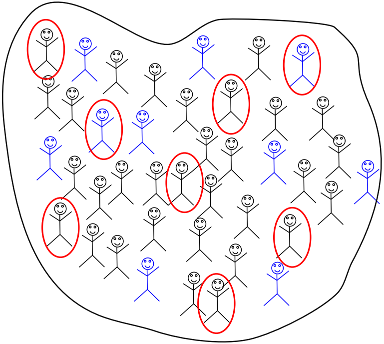


Sample 2

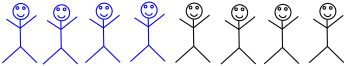


Motivation

Population



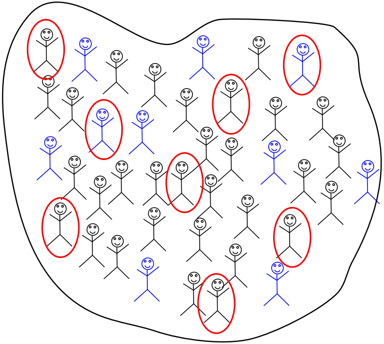
Sample 2



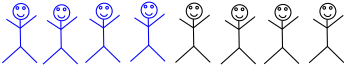
- $\hat{p} = 0.5$

Motivation

Population



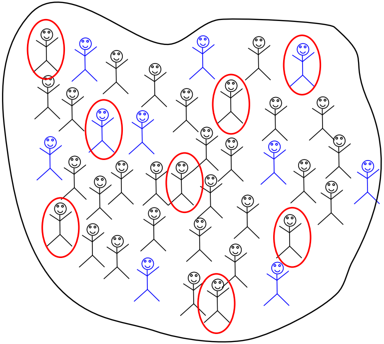
Sample 2

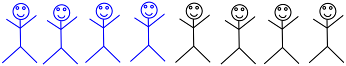


- $\hat{p} = 0.5$
- Average IQ: $\bar{x} = 105.3$

Motivation

Population | **Sample 2**





- $\hat{p} = 0.5$
- Average IQ: $\bar{x} = 105.3$

- $\hat{p}$ and $\bar{X}$ are **random variables**.
- Only **after sampling**, the values are fixed.
- Different samples may result in different values of $\hat{p}$ and $\bar{x}$.

Accounting for uncertainty

- $\hat{p}$ and $\bar{X}$ are **point estimators**.
- **Idea:** Attach error bars to $\hat{p}$ and $\bar{X}$ capturing uncertainty.
- Report

$$\hat{p} \pm \Delta_{\hat{p}} \quad \text{and} \quad \hat{X} \pm \Delta_{\bar{X}},$$

such that the **probabilities**

$$P\left(p \in (\hat{p} - \Delta_{\hat{p}}, \hat{p} + \Delta_{\hat{p}})\right) \quad \text{and} \quad P\left(\mu \in (\bar{X} - \Delta_{\bar{X}}, \bar{X} + \Delta_{\bar{X}})\right)$$

are **close to one**.
- **Question:** How to choose $\Delta_{\bar{X}}$ and $\Delta_{\hat{p}}$?

Accounting for uncertainty

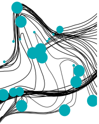
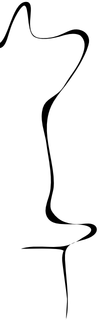
- $\hat{p}$ and $\bar{X}$ are **point estimators**.
- **Idea:** Attach error bars to $\hat{p}$ and $\bar{X}$ capturing uncertainty.
- Report

$$\hat{p} \pm \Delta_{\hat{p}} \quad \text{and} \quad \bar{X} \pm \Delta_{\bar{X}},$$

such that the **probabilities**

$$P\left(p \in (\hat{p} - \Delta_{\hat{p}}, \hat{p} + \Delta_{\hat{p}})\right) \quad \text{and} \quad P\left(\mu \in (\bar{X} - \Delta_{\bar{X}}, \bar{X} + \Delta_{\bar{X}})\right)$$

are **close to one**.
- **Question:** How to choose $\underbrace{\Delta_{\bar{X}}}_{\text{today}}$ and $\underbrace{\Delta_{\hat{p}}}_{\text{on Thursday}}$?


Questions?

Confidence interval for the population mean μ

- Assume that $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$-distributed.

Confidence interval for the population mean μ

- Assume that $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$-distributed.
- The choice of $\Delta_{\bar{X}}$ should depend on:

Confidence interval for the population mean μ

- Assume that $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$-distributed.
- The choice of $\Delta_{\bar{X}}$ should depend on:
  - the sample size.
  - the variability.
  - the degree of desired certainty.

Confidence interval for the population mean μ

- Assume that $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$-distributed.
- The choice of $\Delta_{\bar{X}}$ should depend on:
  - the sample size.
  - the variability.
  - the degree of desired certainty.
- We know

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Confidence interval for the population mean μ

- μ should be within $\pm z \times \frac{\sigma}{\sqrt{n}}$ of $\bar{X}$ with high probability if z-is large enough.
- For given $\alpha \in (0, 1)$, we want

$$P\left(\mu \in \left(\bar{X} - z \times \frac{\sigma}{\sqrt{n}}, \bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right)\right) \geq 1 - \alpha$$

Confidence interval for the population mean μ

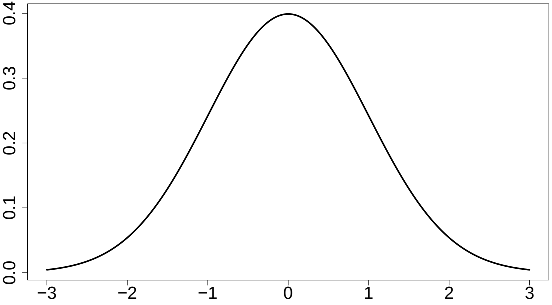
- μ should be within $\pm z \times \frac{\sigma}{\sqrt{n}}$ of $\bar{X}$ with high probability if z-is large enough.
- For given $\alpha \in (0, 1)$, we want

$$P\left(\mu \in \left(\bar{X} - z \times \frac{\sigma}{\sqrt{n}}, \bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right)\right) \geq 1 - \alpha$$

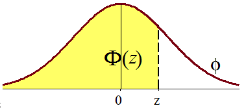
- How to choose z?

How to choose z?

$$P\left(\mu \in \left(\bar{X} - z \times \frac{\sigma}{\sqrt{n}}, \bar{X} + z \times \frac{\sigma}{\sqrt{n}}\right)\right) = P\Big(-z \le \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{\sim\mathcal{N}(0,1)} \le z\Big)$$



Standard normal probabilities



The table gives the distribution function Φ for a N(0,1)-variable Z

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx$$

Last column: N(0,1)-density function (z in 1 dec.): $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

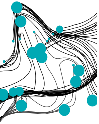
	Second decimal of **z**										
z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**	**φ(z)**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0.3989
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.3970
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.3910
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.3814
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.3683
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	0.3521
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.3332
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.3123
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.2897
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	0.2661
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	0.2420
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	0.2179
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	0.1942
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	0.1714
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	0.1497
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	0.1295
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	0.1109
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	0.0940
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	0.0790
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	0.0656
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0.0540

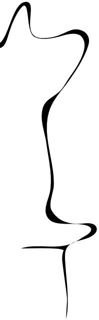
Confidence interval for the population mean μ

$(1-\alpha)$-CI for μ when σ is known

$$\mathrm{CI}(\mu, \sigma, \alpha) = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right),$$

- $z_{1-\frac{\alpha}{2}}$ is the $(1-\alpha)$-quantile of the $\mathcal{N}(0,1)$- distribution.





Questions?

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$\sigma = 7.11$

- Determine a 95%-confidence interval for the average height of Dutch men!

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$\sigma = 7.11$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$\sigma = 7.11$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$
- $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$\sigma = 7.11$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$
- $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$
- $n = 10$, $\bar{x} = 183.91$ and $\sigma = 7.11$.
- The 95%-confidence interval is given by

$$(179.503\,,\ 188.317).$$

True or false:

The interval (179.503 , 188.317) contains the true mean μ with probability 0.95.

Confidence intervals for μ when σ **is unknown**

Confidence intervals for μ when σ **is unknown**

- **Idea:** Replace σ by S and do everything as before.

Confidence intervals for μ when σ is unknown

- **Idea:** Replace σ by S and do everything as before.

Confidence intervals for μ when σ **is unknown**

- **Idea:** Replace σ by S and do everything as before.
- **We need to be careful here!**

Confidence intervals for μ when σ **is unknown**

- **Idea:** Replace σ by S and do everything as before.
- **We need to be careful here!**

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \not\sim \mathcal{N}(0, 1).$$

Confidence intervals for μ when σ is unknown

- **Idea:** Replace σ by S and do everything as before.
- **We need to be careful here!**

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \not\sim \mathcal{N}(0, 1).$$

- We have that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} :$$

$\frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a **t-distribution with $n-1$ degrees of freedom**.

The t-distribution

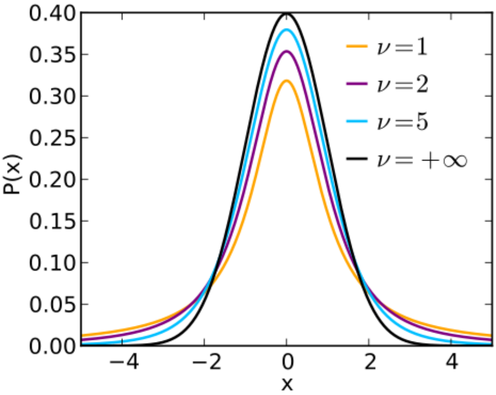
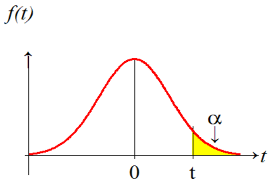


Table *t*-distribution

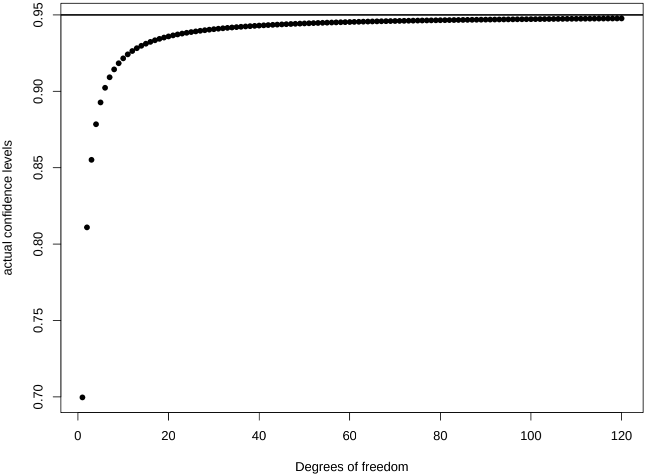
In the table you find the critical values *t* for the upper-tailed probabilities such that

$$P(T \geq t) = \alpha$$



Number of degrees of freedom	α							
	0.25	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792

Using the normal distribution instead of the *t*-distribution

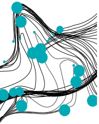


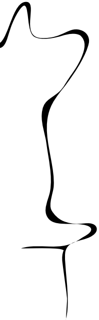
Confidence interval for the population mean μ

$(1-\alpha)$-CI for μ when σ is unknown

$$\text{CI}(\mu, S, \alpha) = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right),$$

- $t_{1-\frac{\alpha}{2},n-1}$ is the $(1-\alpha)$-quantile of the t_{n-1}- distribution.





Questions?

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$s = 8.292$

- Determine a 95%-confidence interval for the average height of Dutch men!

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$s = 8.292$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right).$

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$s = 8.292$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right).$
- $t_{n-1,1-\frac{\alpha}{2}} = t_{9,0.975} = 2.262$

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2

$n = 10$
$\bar{x} = 183.91$
$s = 8.292$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$.
- $t_{n-1,1-\frac{\alpha}{2}} = t_{9,0.975} = 2.262$
- $n = 10$, $\bar{x} = 183.91$ and $s = 7.9980$.
- The 95%-confidence interval is given by

$$(177.978\,,\ 189.842).$$

Example: Average height of Dutch men

- The confidence intervals were:

$$(179.503\,,\,188.317) \quad \text{and} \quad (177.978\,,\,189.842).$$

Example: Average height of Dutch men

- The confidence intervals were:

$$(179.503\,,\,188.317) \quad \text{and} \quad (177.978\,,\,189.842).$$

- The true mean That was used to generate the data was $\mu = 183.9$.