Introduction to statistics

(Day 2)

Agenda

- **Where:**
  - Mar 4: ✅
  - Mar 5: AULA DARWIN - MBC
  - Mar 6 : Online
- **When:**
  - 14-17
  - 1 coffee break
- **Who:**
  - Paola Dalmasso paola.dalmasso@unito.it
  - Alessia Visconti alessia.visconti@unito.it
- **How (to pass):**
  - Attend at least 2 lessons

How to ask questions/give feedback

- Interrupt me
- Take advantage of end/start/breaks
- Send emails alessia.visconti@unito.it
- Use the shared pad: https://etherpad.wikimedia.org/p/intro_stats_2024_specialita (or https://t.ly/vRbvy)

Recap

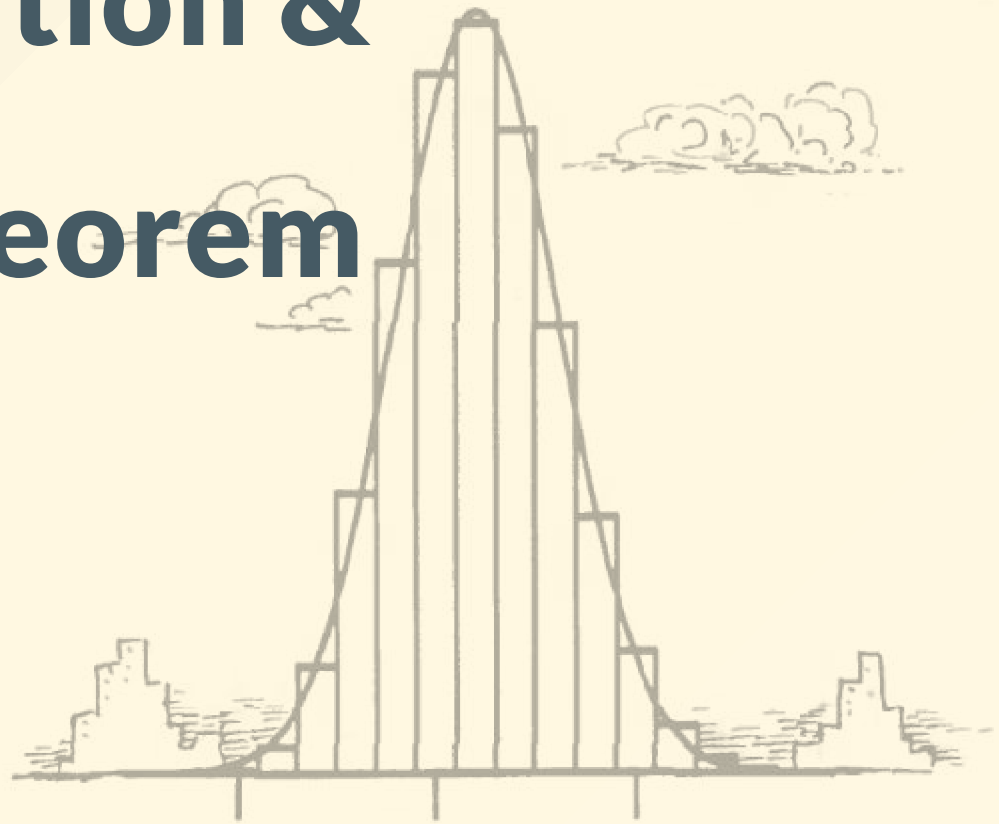
Recap

- The collection, organisation, summarisation, and analysis of data
  → *Descriptive* statistics
- The drawing of inferences about a body of data when only a part of the data is observed
  → *Inferential* statistics

Recap

- When can't study a population, we select a representative sample
- There are different sampling strategies
- Data are described with measures of centrality (mode, median, mean) and dispersion (range, IQR, standard deviation)
- Parameters (calculated on the population) *vs* statistics (calculated on the sample)

The Normal Distribution & the Central Limit Theorem

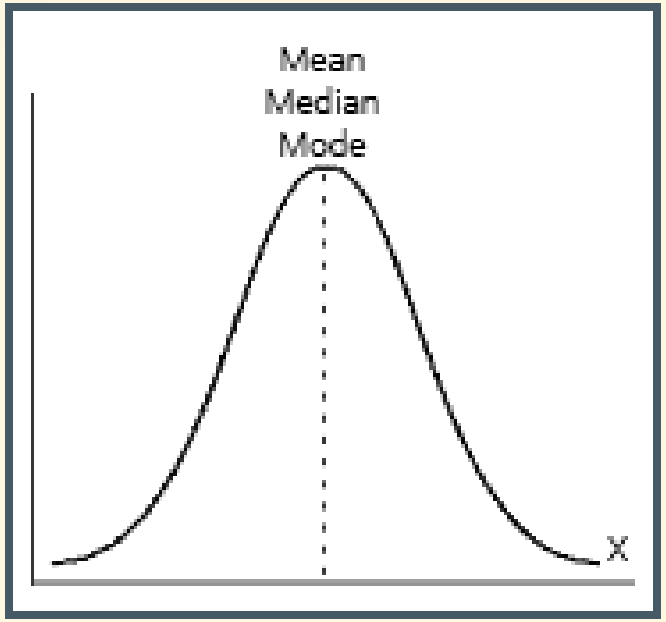


Learning objectives

- Know the characteristics of the Normal distribution
- Understand the Central Limit Theorem

The Normal distribution

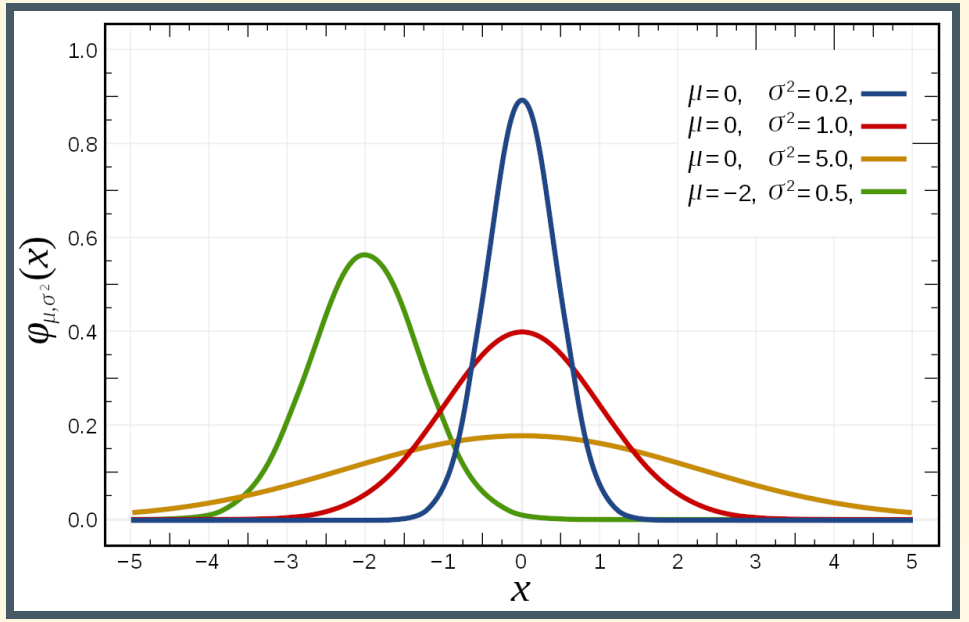
- Symmetrical
- $\mathcal{N} = (\mu, \sigma^2)$



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

The Normal distribution

- Symmetrical
- $\mathcal{N} = (\mu, \sigma^2)$

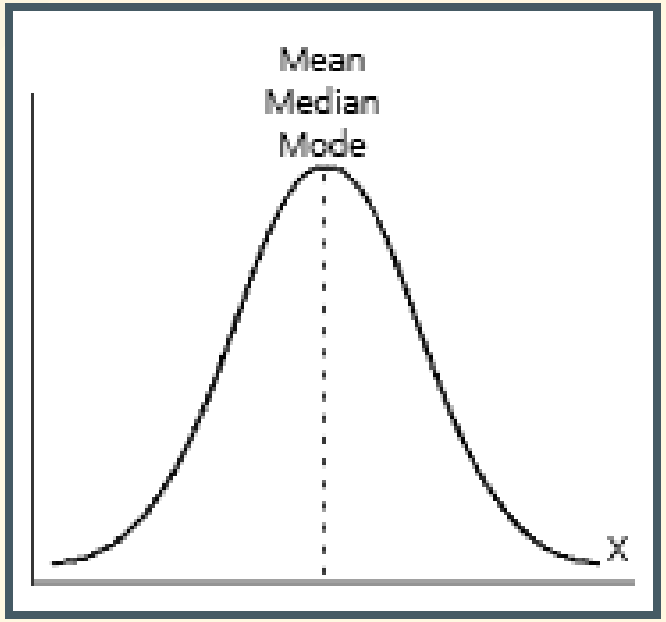


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The Normal distribution

- Symmetrical
- $\mathcal{N} = (\mu, \sigma^2)$
- $\text{Area under the curve} = 1$

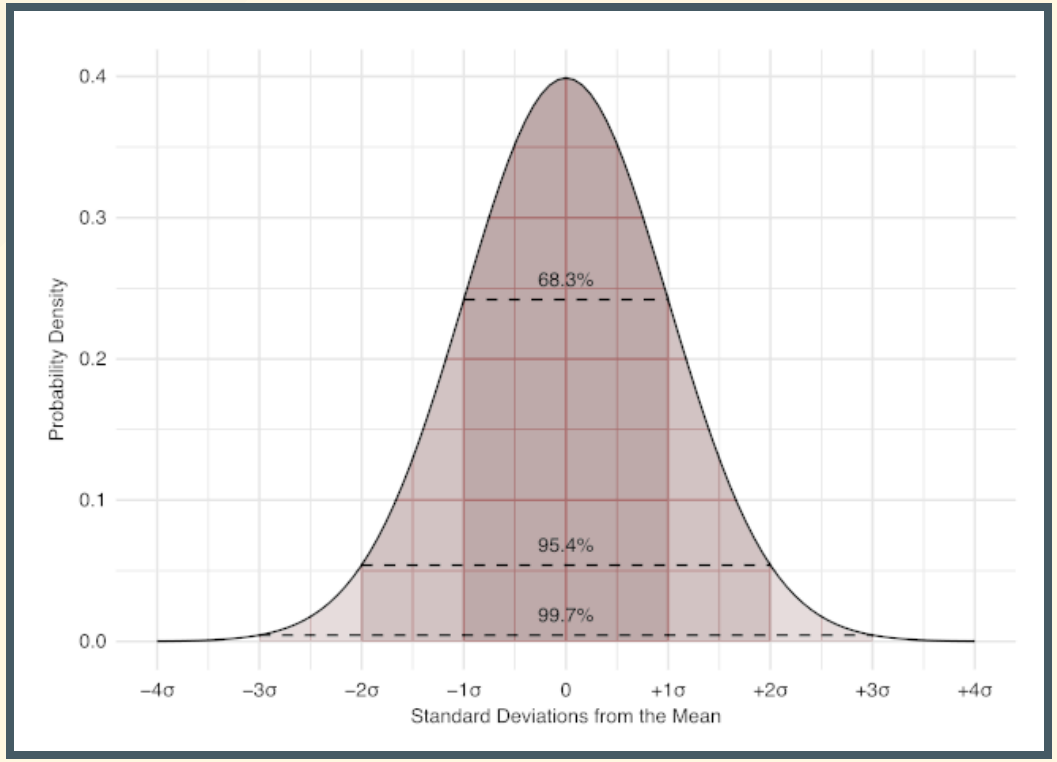
? What is the area left of the median? And right?



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

The Normal distribution

- Symmetrical
- $\mathcal{N} = (\mu, \sigma^2)$
- $\text{Area under the curve} = 1$
- *3-sigma* (or *68–95–99.7*) rule

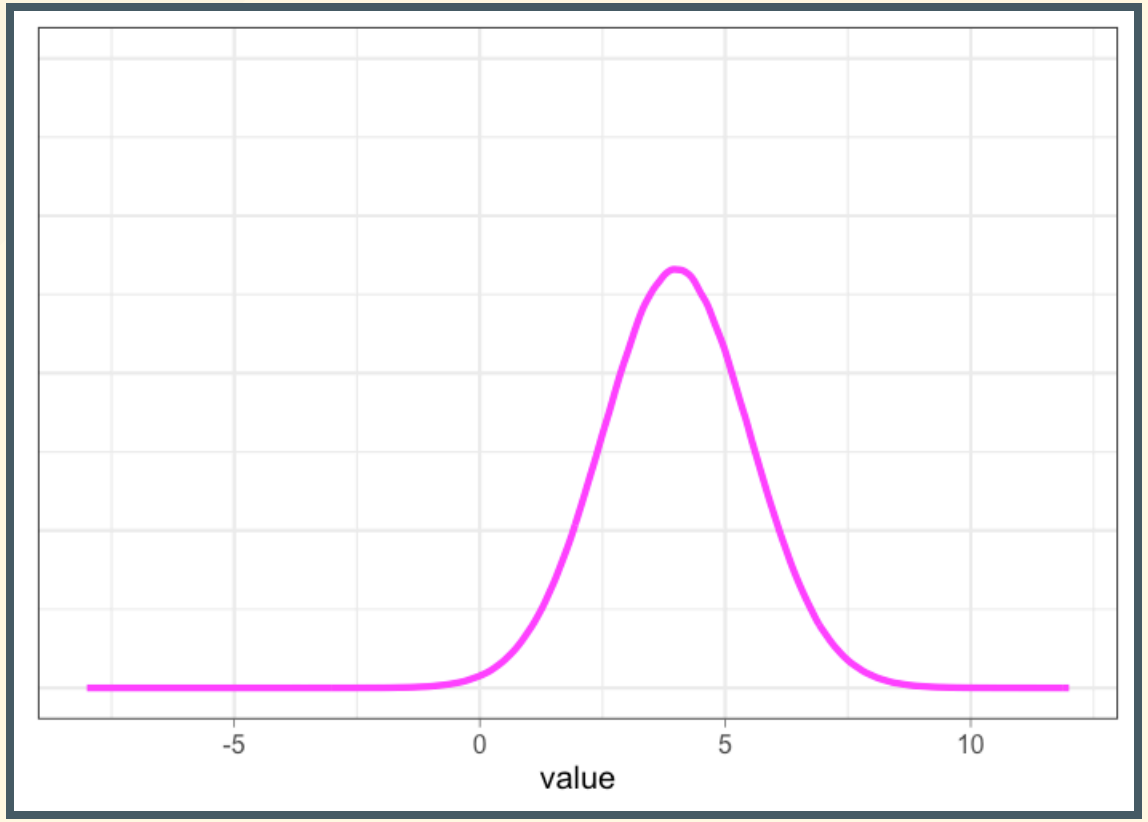


The Standard Normal distribution

- $\mathcal{N} = Z = (0, 1)$

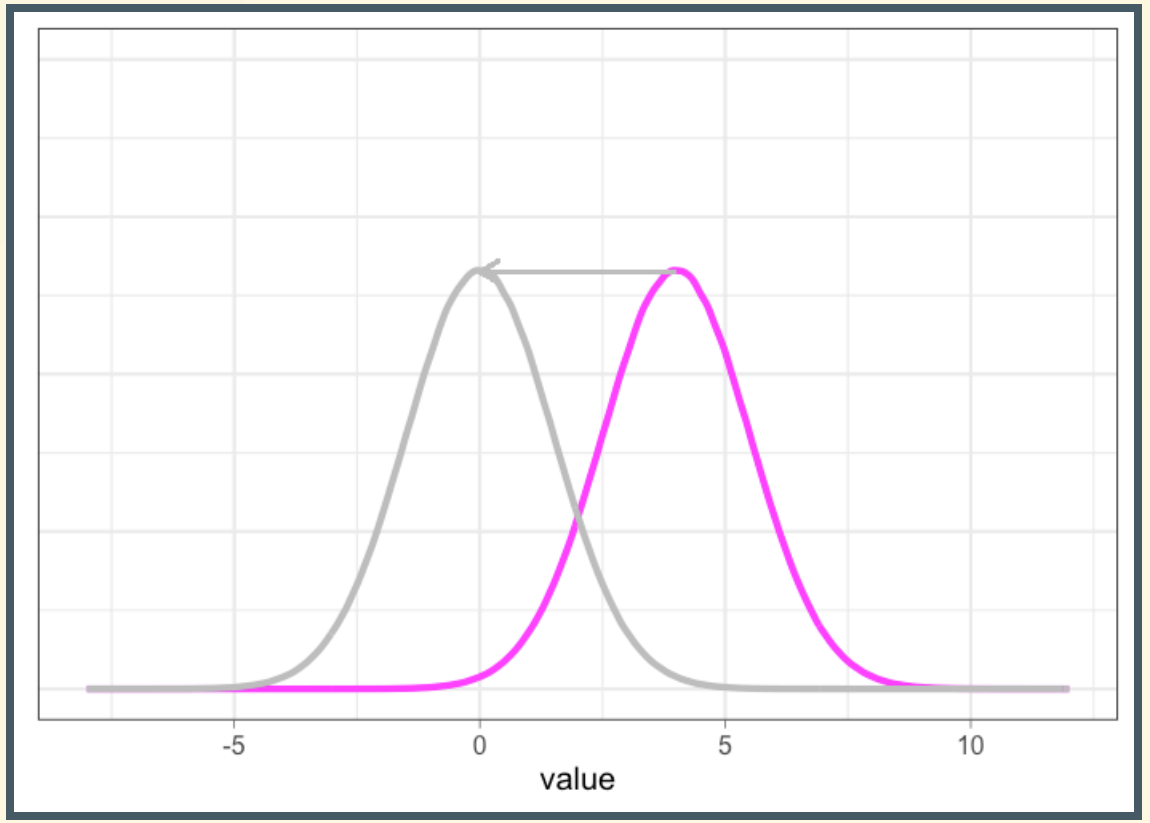
The Standard Normal distribution

- $\mathcal{N} = (\mu, \sigma^2) \rightarrow Z = (0, 1)$



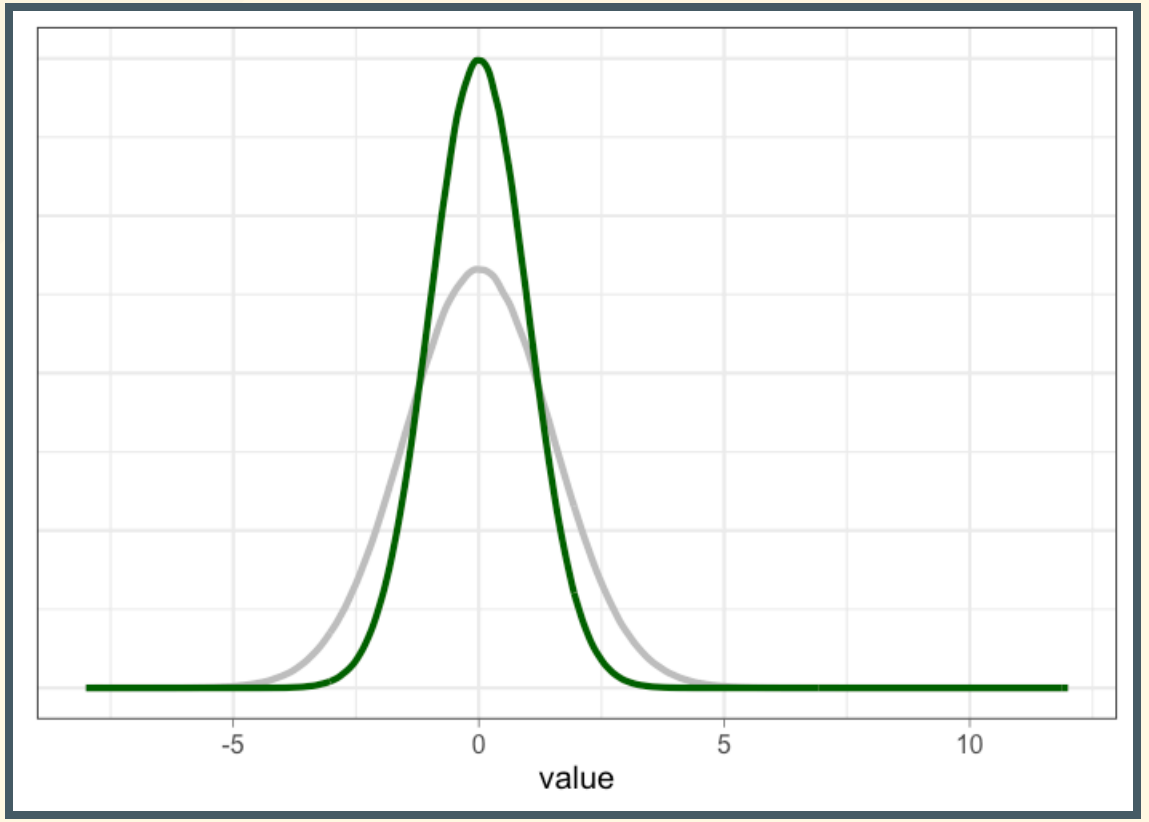
The Standard Normal distribution

- $\mathcal{N} = (\mu, \sigma^2) \rightarrow Z = (0, 1)$
- $z = \frac{x - \mu}{}$



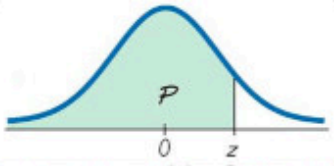
The Standard Normal distribution

- $\mathcal{N} = (\mu, \sigma^2) \rightarrow Z = (0, 1)$
- $z = \frac{x - \mu}{\sigma}$



The Standard Normal distribution

- $\mathcal{N} = (\mu, \sigma^2) \rightarrow Z = (0, 1)$
- $z = \frac{x - \mu}{\sigma}$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

The Standard Normal distribution in practice

📌 $\mu = 170\ \text{cm} \quad \sigma = 9.5\ \text{cm}$

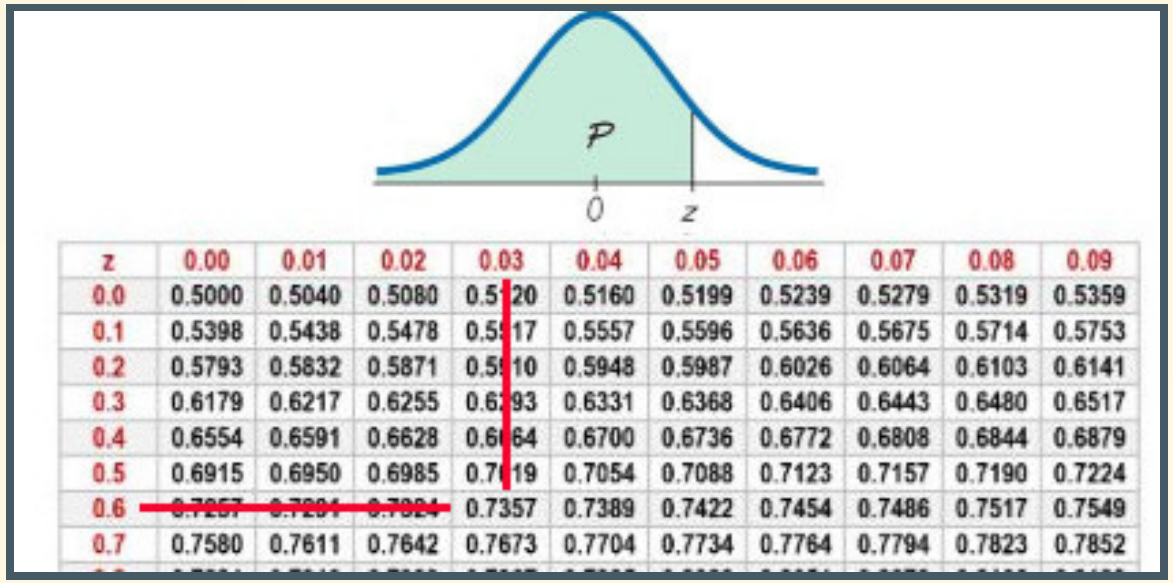
$\mathcal{P}(x \geq 176) = \ ?$

The Standard Normal distribution in practice

📌 $\mu = 170\ \text{cm} \quad \sigma = 9.5\ \text{cm}$

$$z = \frac{x-\mu}{\sigma} = \frac{176-170}{9.5} = 0.63$$

$$\mathcal{P}(x \geq 176) = 1 - 0.7357 = \\ = 0.26$$

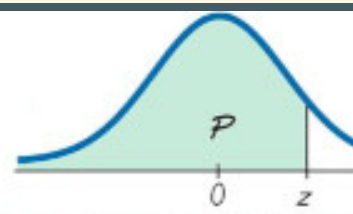


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852

Exercise #7

📌 $\mu = 170\ \text{cm} \quad \sigma = 9.5\ \text{cm}$

$\mathcal{P}(x \geq 179.5) = \ ?$



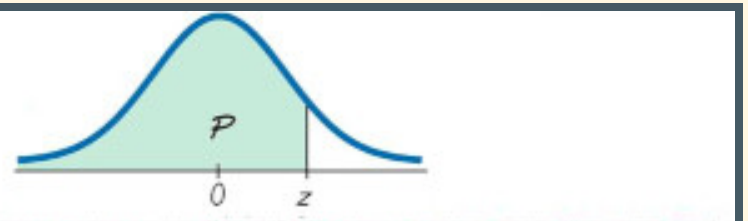
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

Exercise #7 -- Solution

📌 $\mu = 170\ \text{cm} \quad \sigma = 9.5\ \text{cm}$

$$z = \frac{x-\mu}{\sigma} = \frac{179.5-170}{9.5} = 1$$

$$\mathcal{P}(x \geq 179.5) = 1 - 0.8413 = \\ = 0.16$$



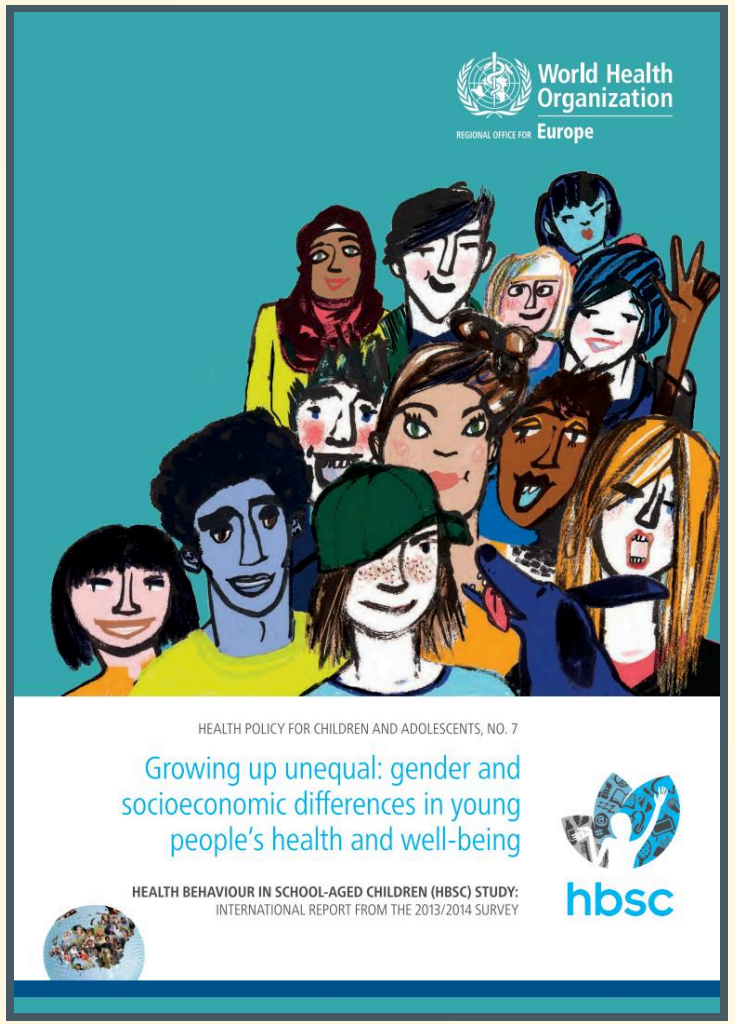
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

The Standard Normal distribution in practice

📌 $n = 6,705$ 15 y.o. males
$\bar{x}_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$s_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 25$ in the population?

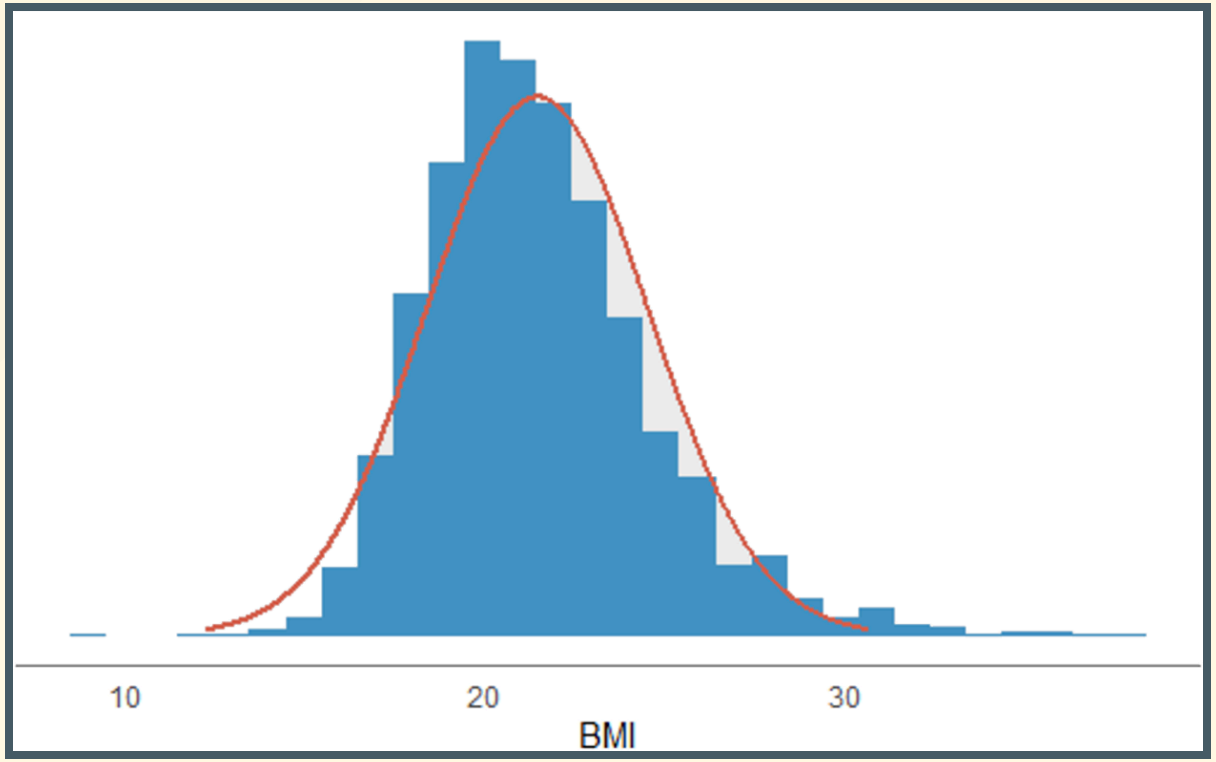
https://hbsc.org



The Standard Normal distribution in practice

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 25$ in the population?



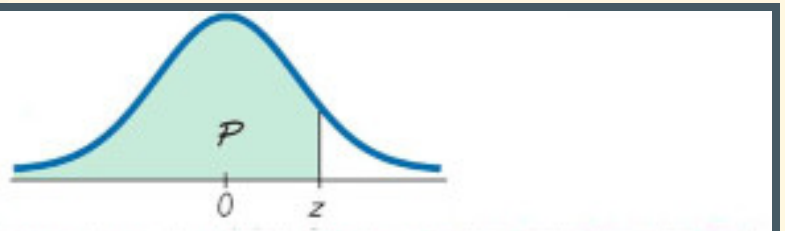
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The Standard Normal distribution in practice

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 25$ in the population?

$$z = \frac{x - \mu}{\sigma} = \frac{25 - 21.5}{3.1} = 1.12$$

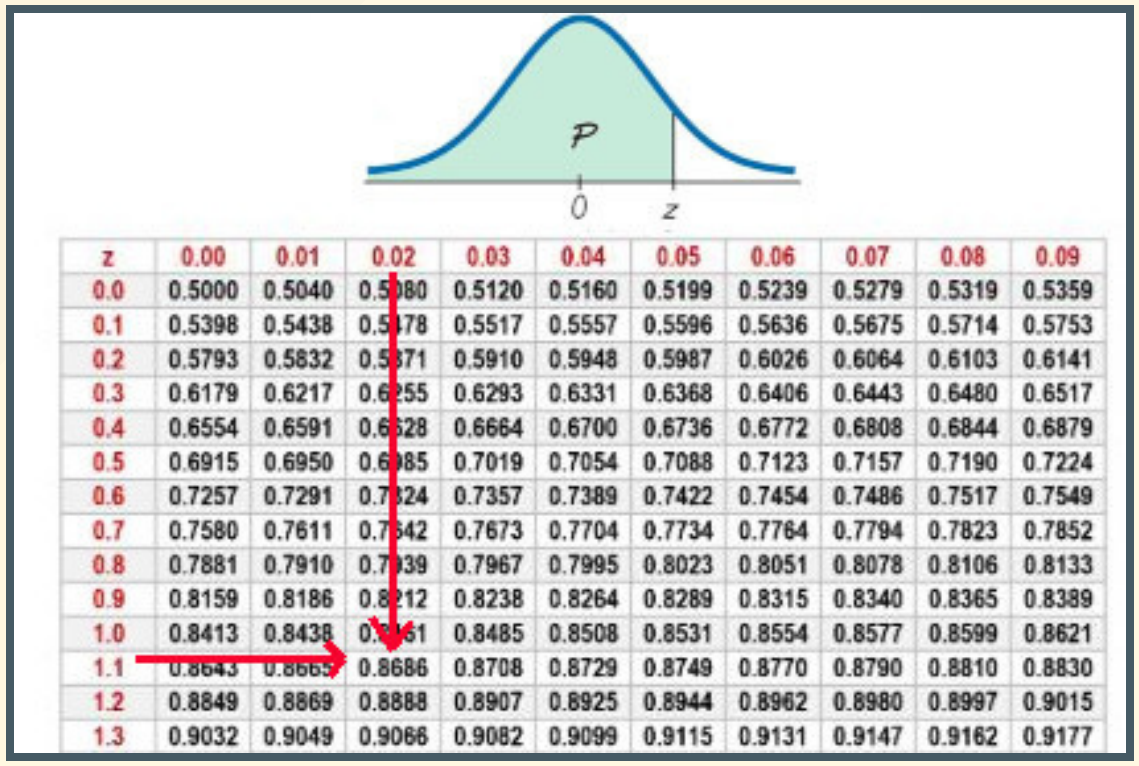


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

The Standard Normal distribution in practice

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 25$ in the population?



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

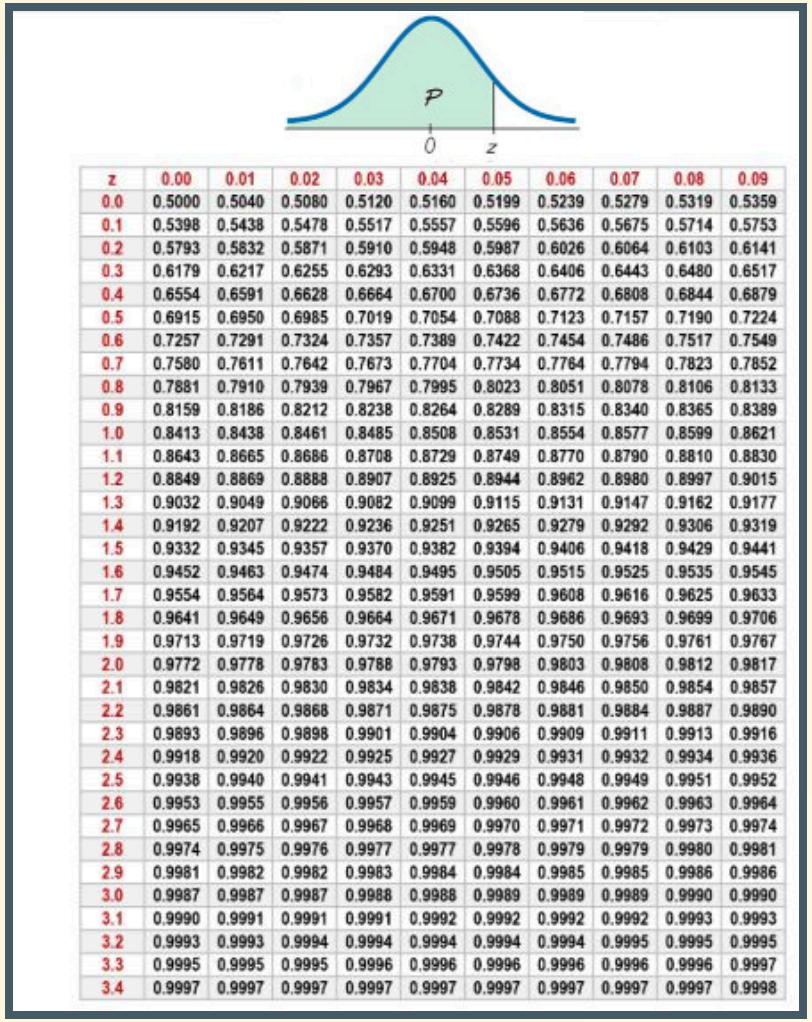
$$z = \frac{x - \mu}{\sigma} = \frac{25 - 21.5}{3.1} = 1.12$$

$$\mathcal{P}(\text{BMI} > 25) = 1 - 0.8686 = 0.1314 \rightarrow 13.1\%$$

Exercise #8

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 30$ in the population?



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Exercise #8 -- Solution

📌 $n = 6,705$ 15 y.o. males

$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$

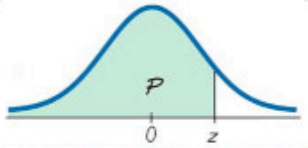
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 30$ in the population?

$$z = \frac{x-\mu}{\sigma} = \frac{30-21.5}{3.1} = 2.74$$

$$\mathcal{P}(\text{BMI} > 30) = 1 - 0.9969$$

$$= 0.0031 \rightarrow 0.31\%$$

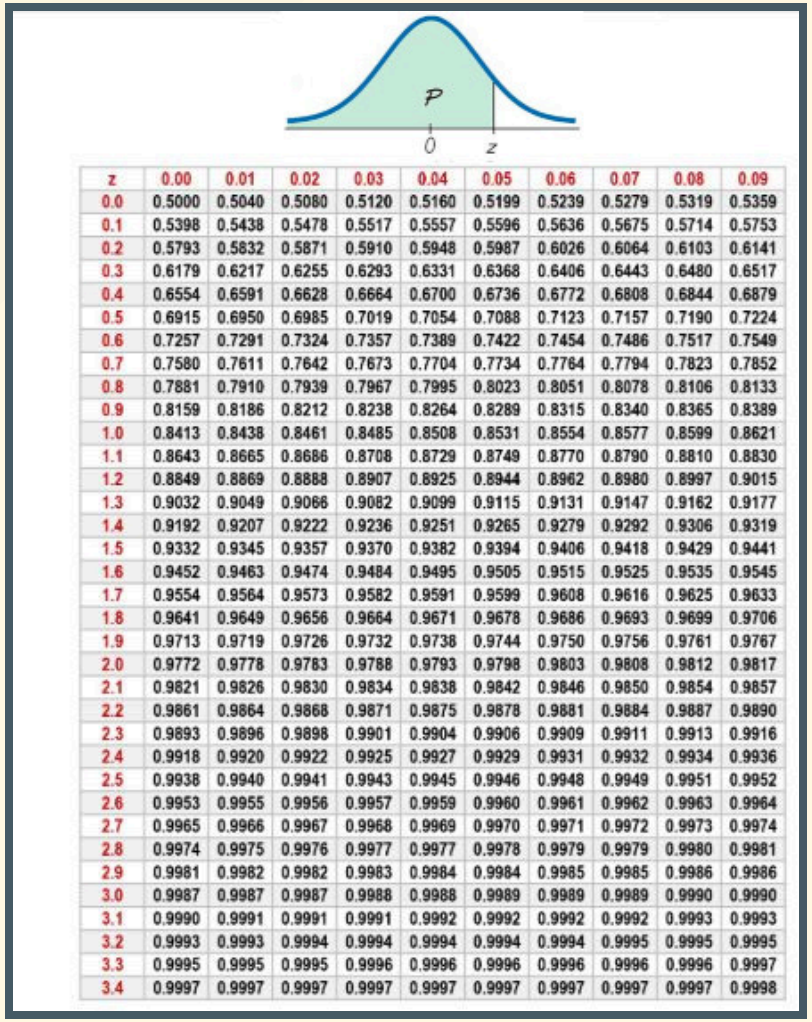


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Exercise #9

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

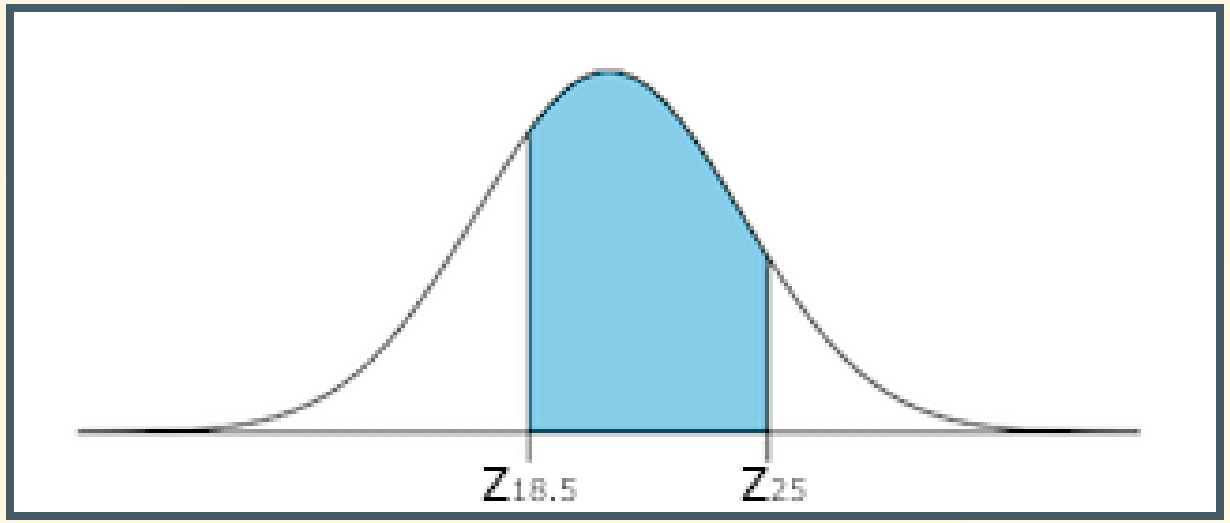


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

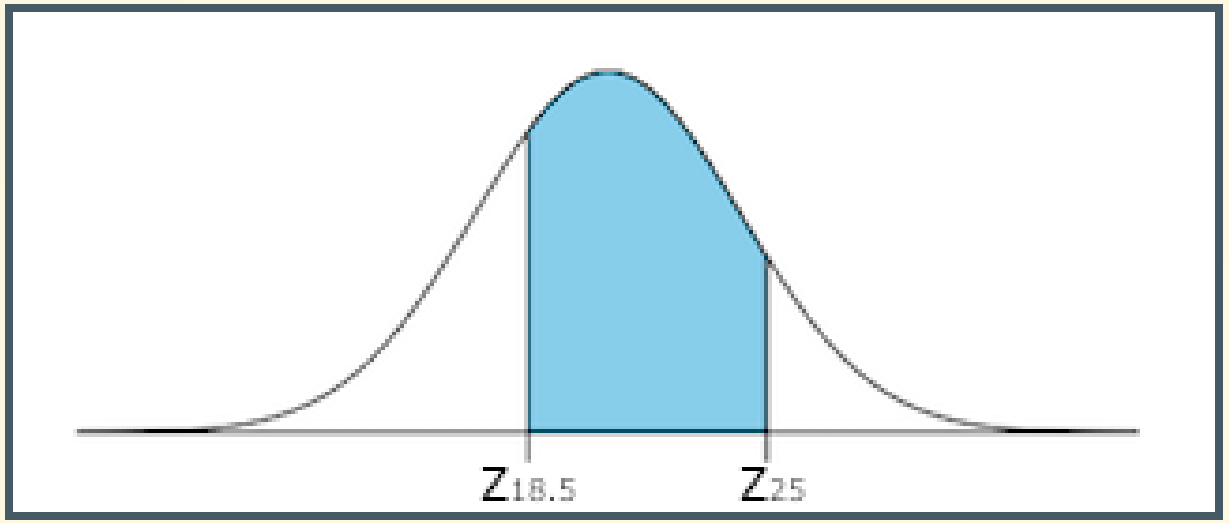


Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

$$z_{25} = \frac{x-\mu}{\sigma} = \frac{25-21.5}{3.1} = 1.12$$
$$\mathcal{P}(\text{BMI} > 25) = 1 - 0.8686 =$$
$$= 0.131$$



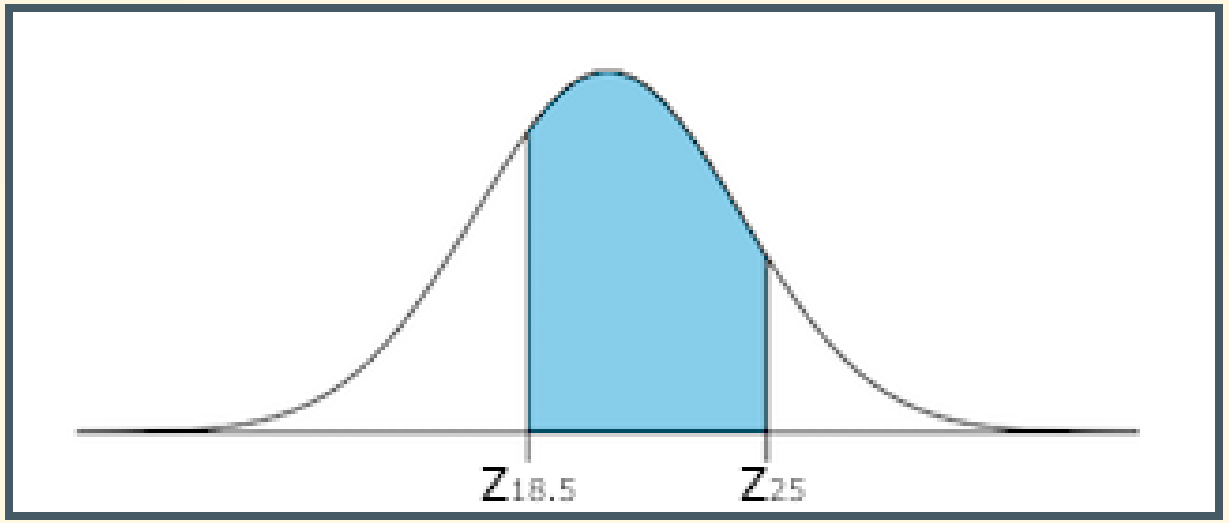
Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

$$\mathcal{P}(\text{BMI} > 25) = 0.131$$

$$z_{18.5} = \frac{x-\mu}{\sigma} = \frac{18.5-21.5}{3.1} = -0.97$$



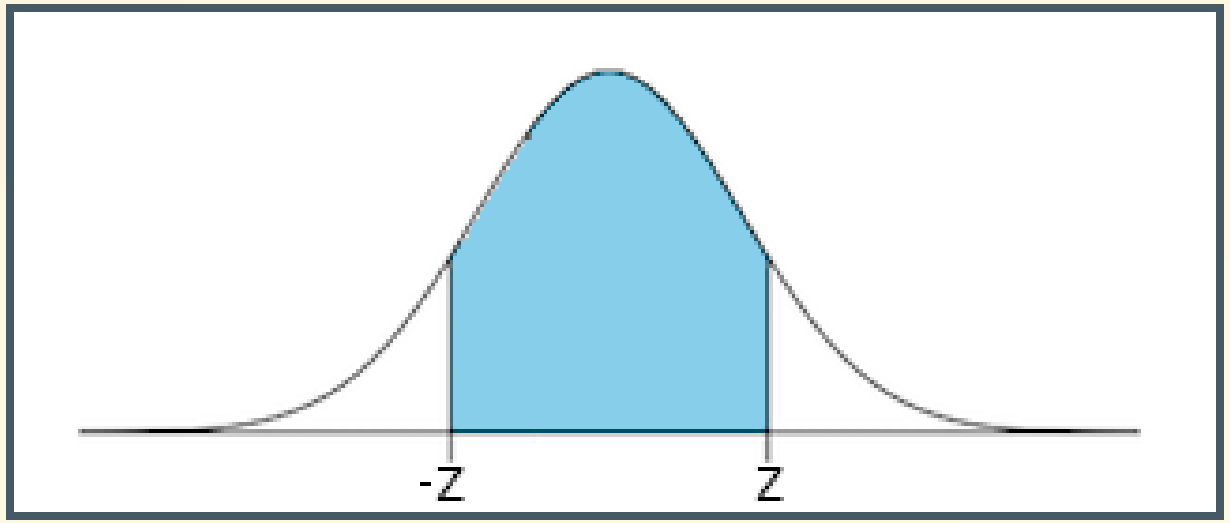
Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

$$\mathcal{P}(\text{BMI} > 25) = 0.131$$

$$z_{18.5} = \frac{x-\mu}{\sigma} = \frac{18.5-21.5}{3.1} = -0.97$$



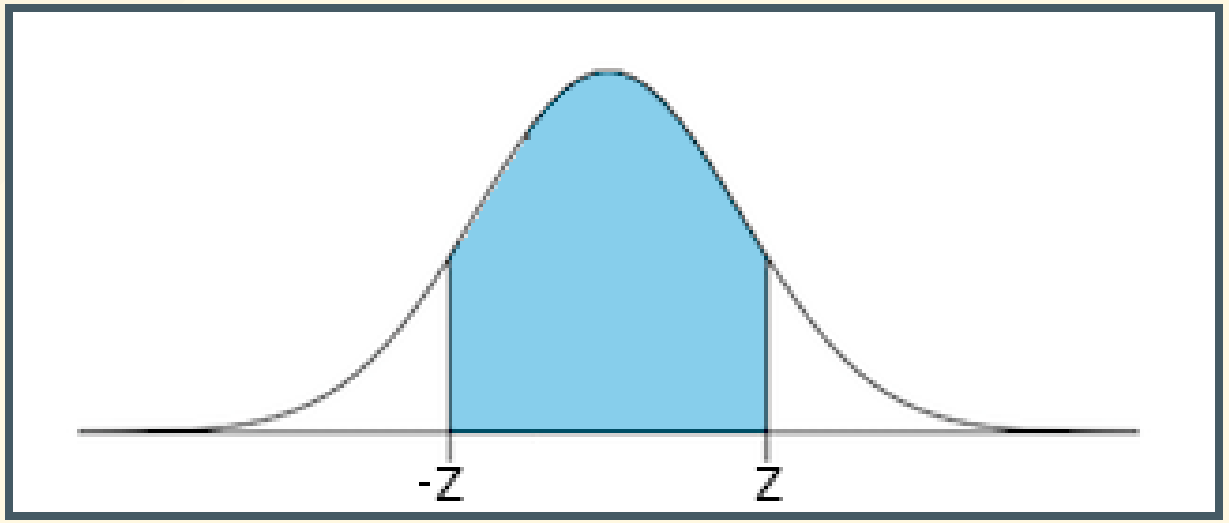
Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

$$\mathcal{P}(\text{BMI} > 30) = 0.131$$

$$z_{18.5} = \frac{x-\mu}{\sigma} = \frac{18.5-21.5}{3.1} = -0.97$$
$$\mathcal{P}(\text{BMI} < 18.5) = 1 - 0.8340 = 0.166$$

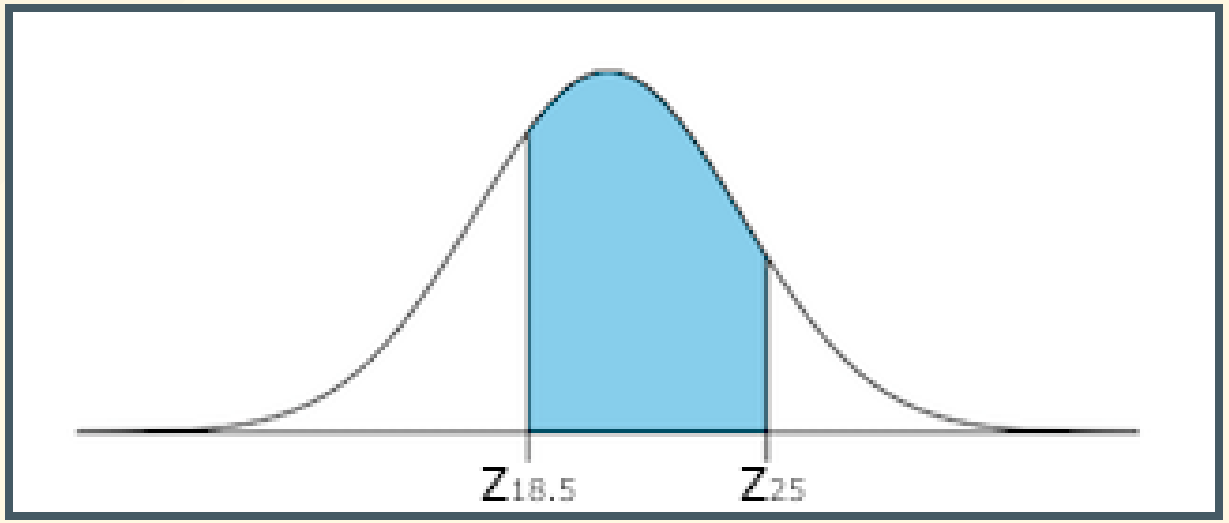


Exercise #9 -- Solution

📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $18.5 < \text{BMI} < 25$ in the population?

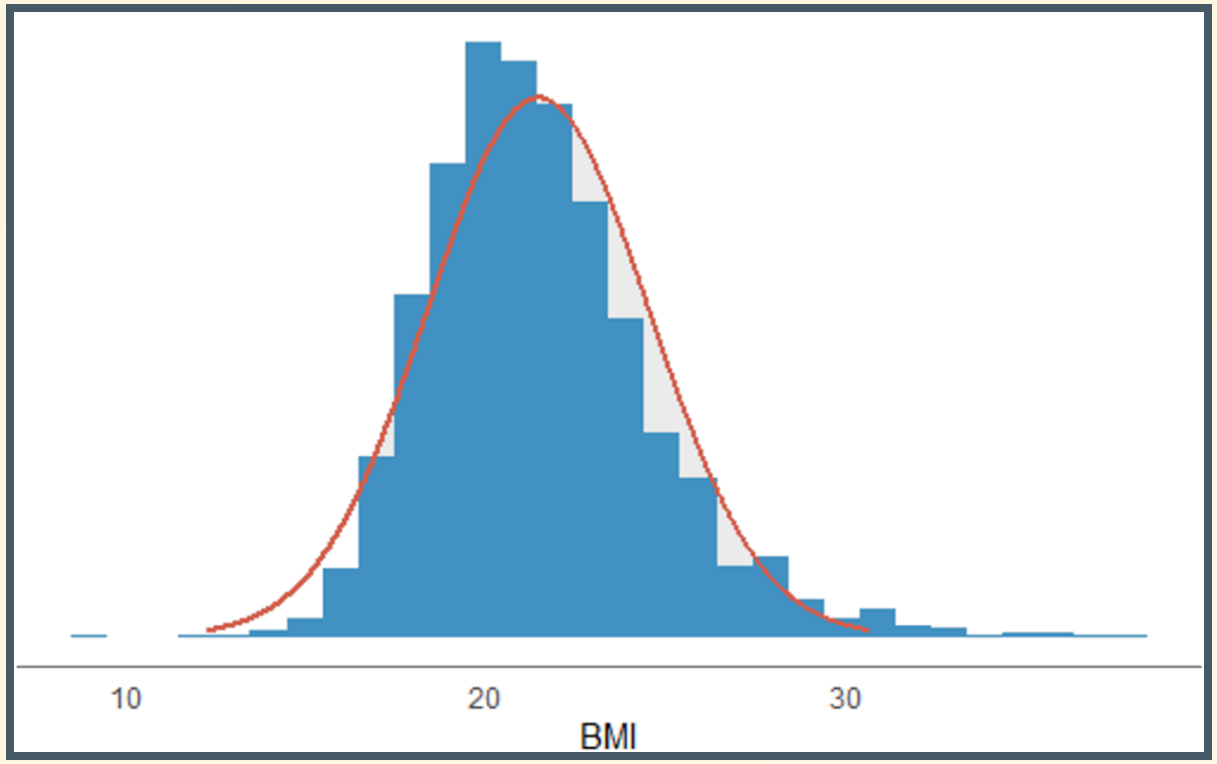
$\mathcal{P}(\text{BMI} > 30) = 0.131$
$\mathcal{P}(\text{BMI} < 18.5) = 0.166$



$$\mathcal{P}(18.5 < \text{BMI} < 25) = 1 - 0.131 - 0.166 = 0.703 \rightarrow 70.3\%$$

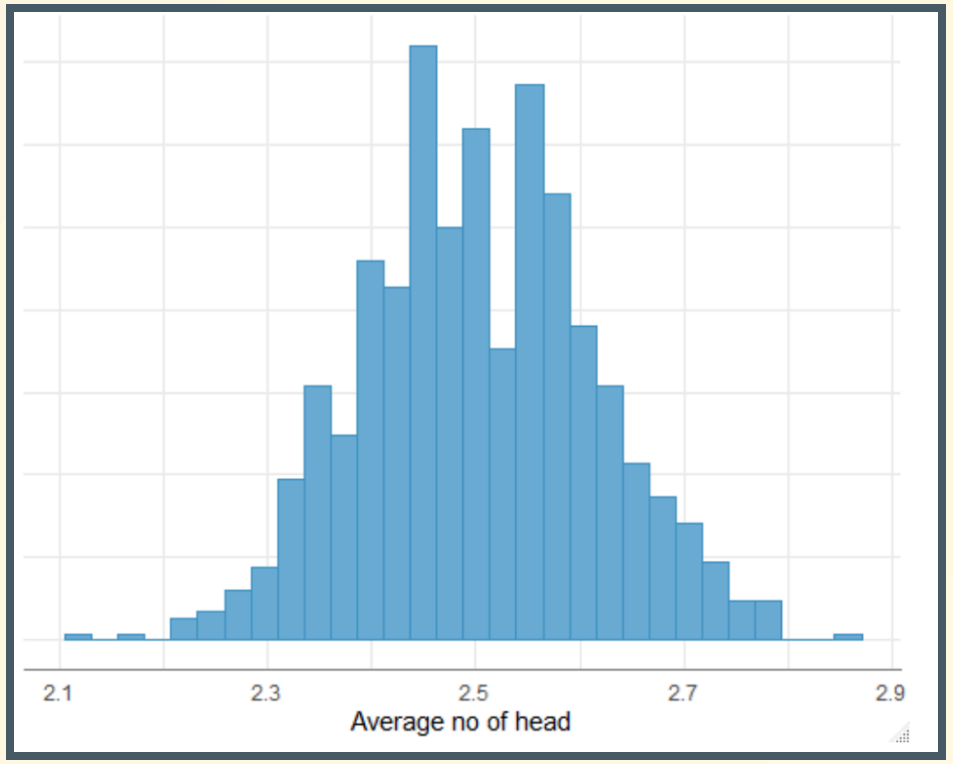
The Normal distribution in the wild

- Several natural phenomena approximate a Normal distribution

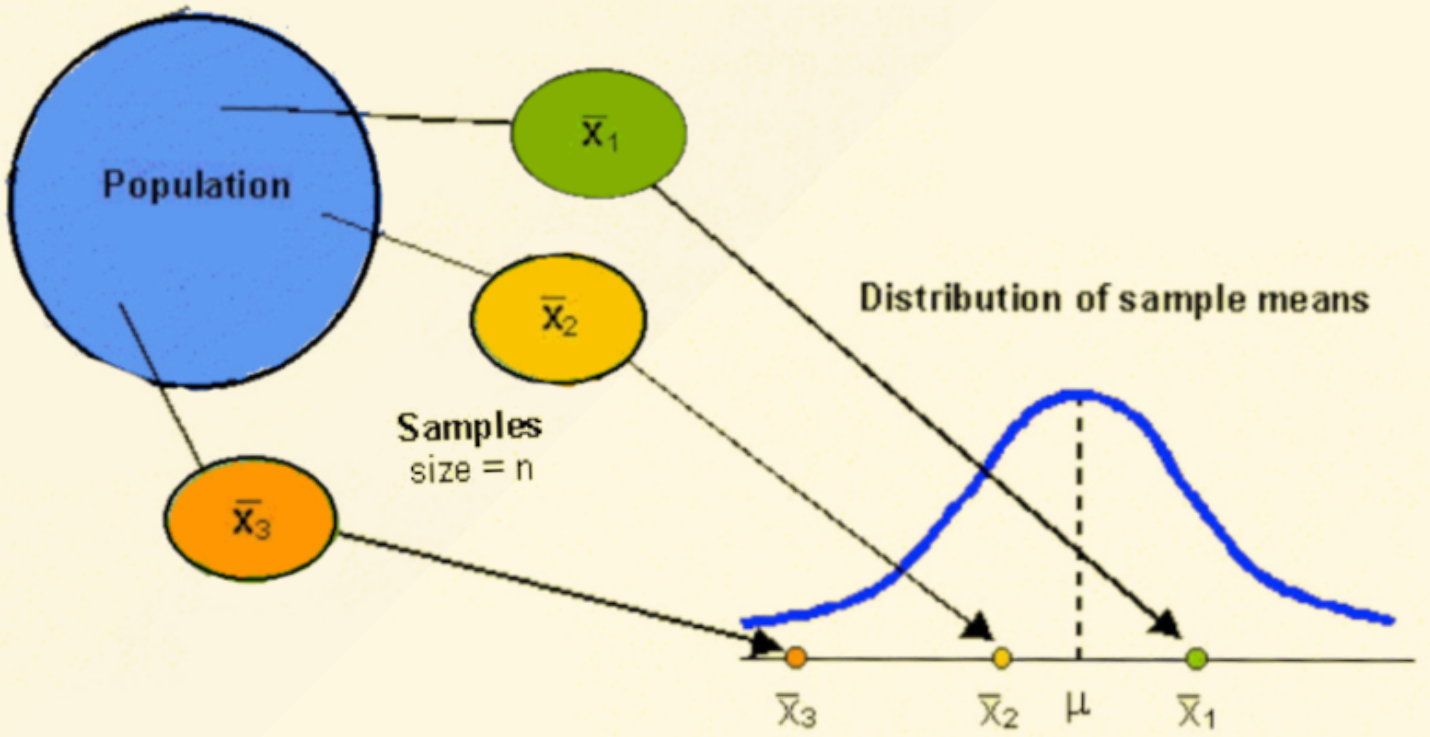


The Normal distribution in the wild

- Several natural phenomena approximate a Normal distribution
- Several probability distributions approximate a Normal distribution



Sampling distribution



The Central Limit Theorem (CLT)

🎯 As the sample size gets larger, the sampling distribution of the sample means approaches a normal distribution $\mathcal{N} = (\mu, \frac{\sigma^2}{n})$

$$\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \rightarrow \text{ standard error (SE)}$$

The Central Limit Theorem (CLT)

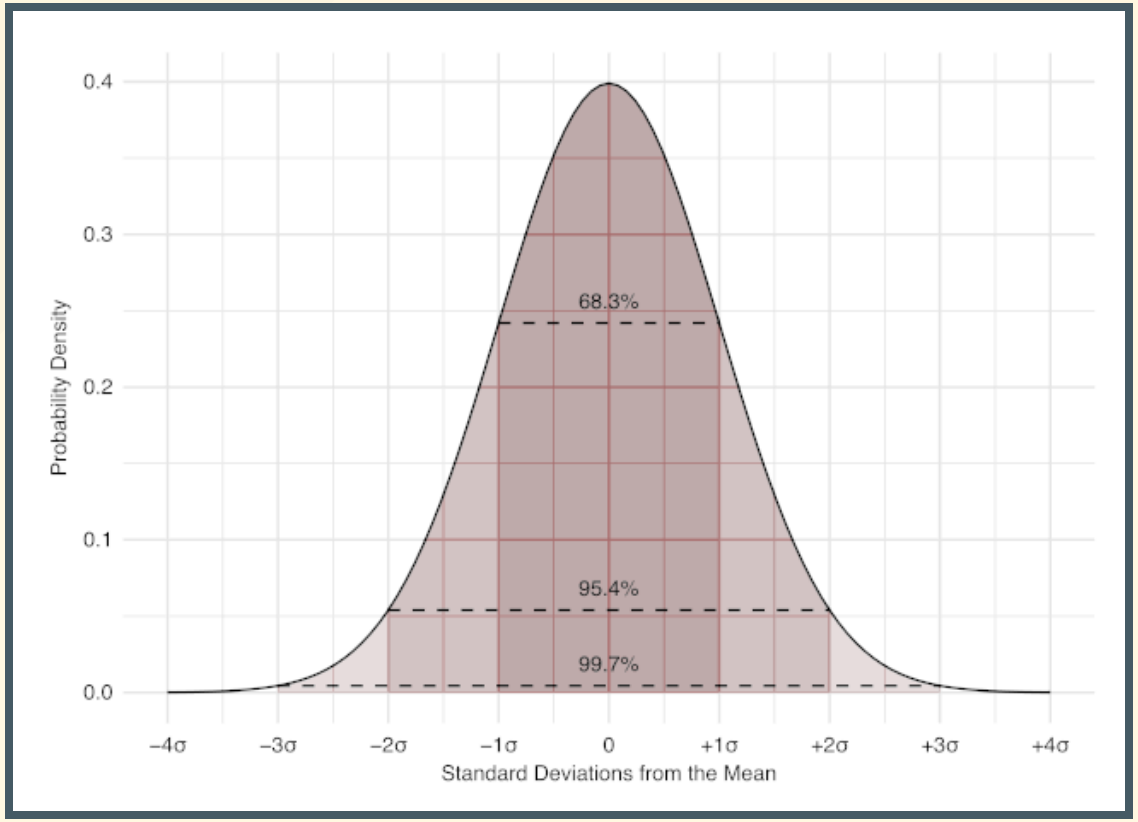
- As the sample size gets larger, the sampling distribution of the sample means approaches a normal distribution $\mathcal{N} = (\mu, \frac{\sigma^2}{n})$

$$\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \rightarrow \text{ standard error (SE)}$$

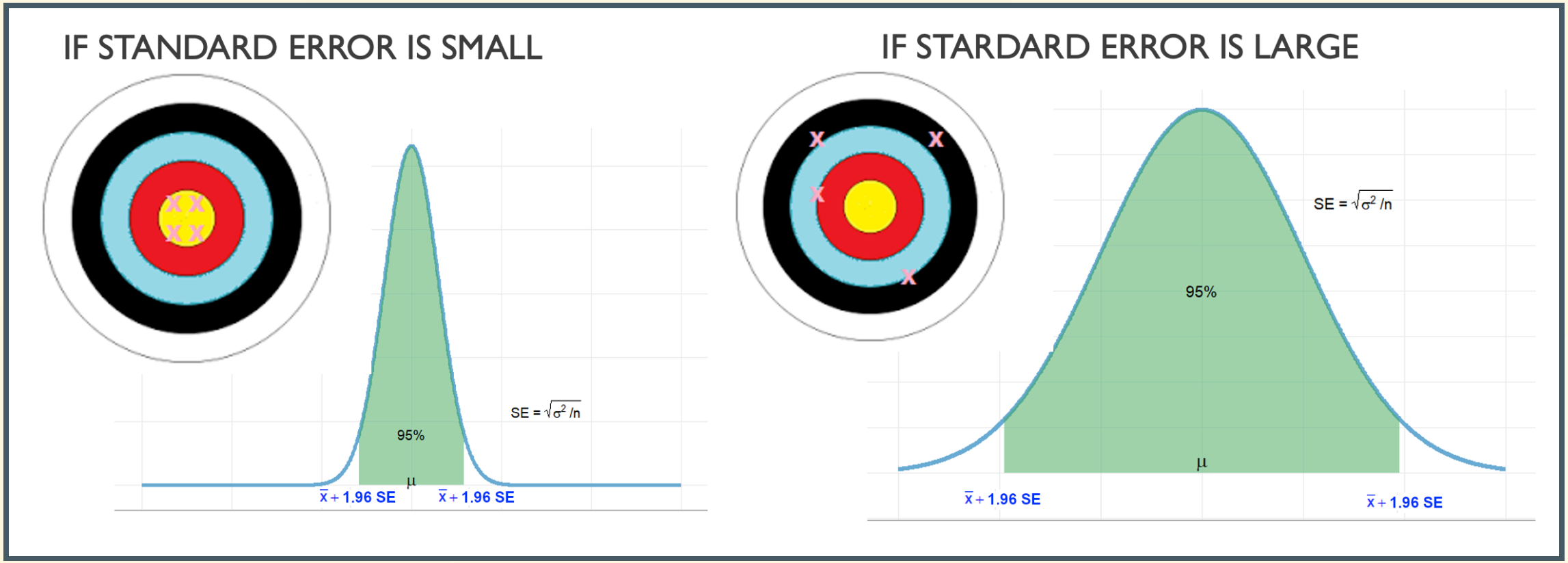
- standard error $\neq$ standard deviation
  standard deviation (SD) $\rightarrow$ spread in the collected data
  standard error (SE) $\rightarrow$ error in the estimation

The Standard Error

🎯 95% of the sample means are within 1.96 SE from the population mean

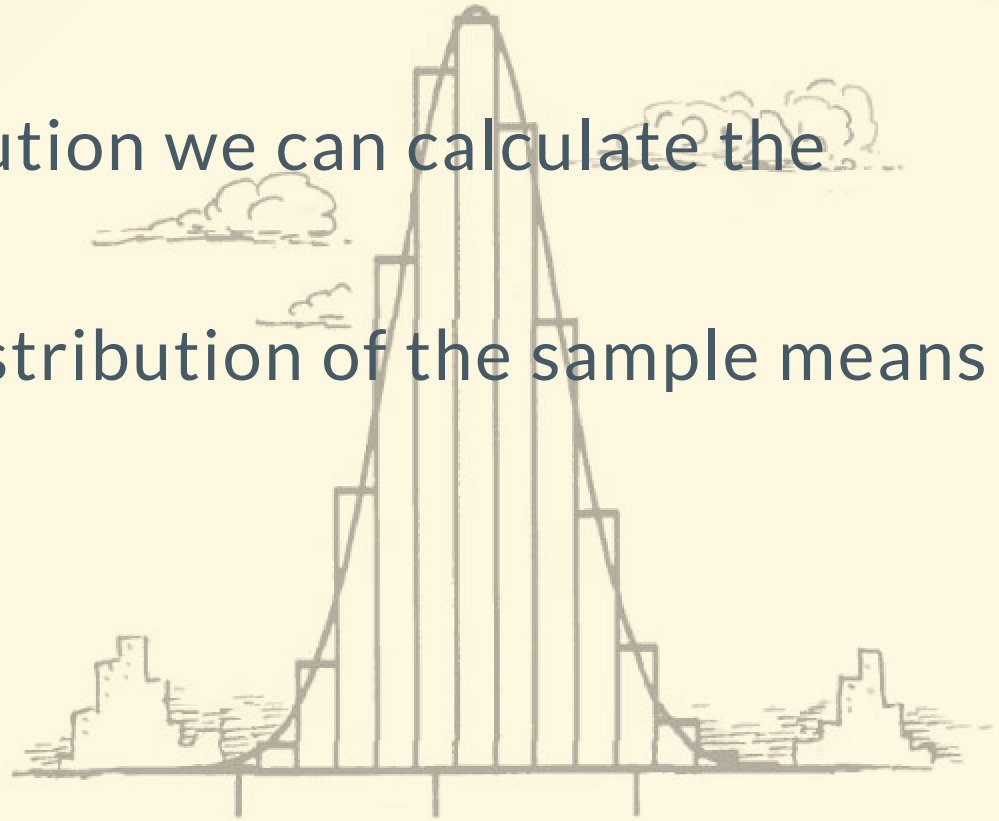


The Standard Error



Summary

- Multiple phenomena and statistical distributions are normally distributed
- Using the standard normal distribution we can calculate the probability of an observation
- For large samples, the sampling distribution of the sample means will be normally distributed (CLT)

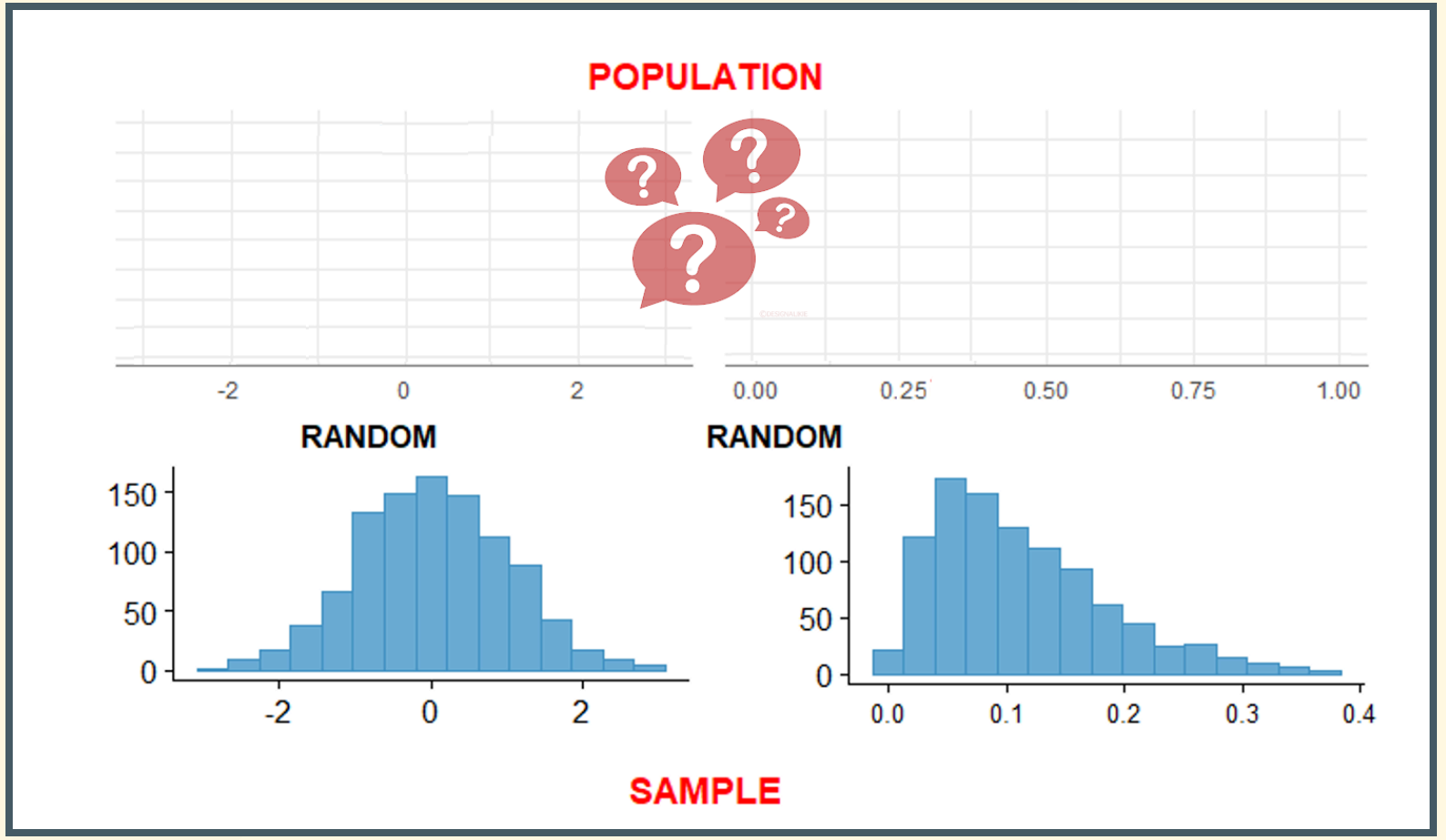


Inferential statistics

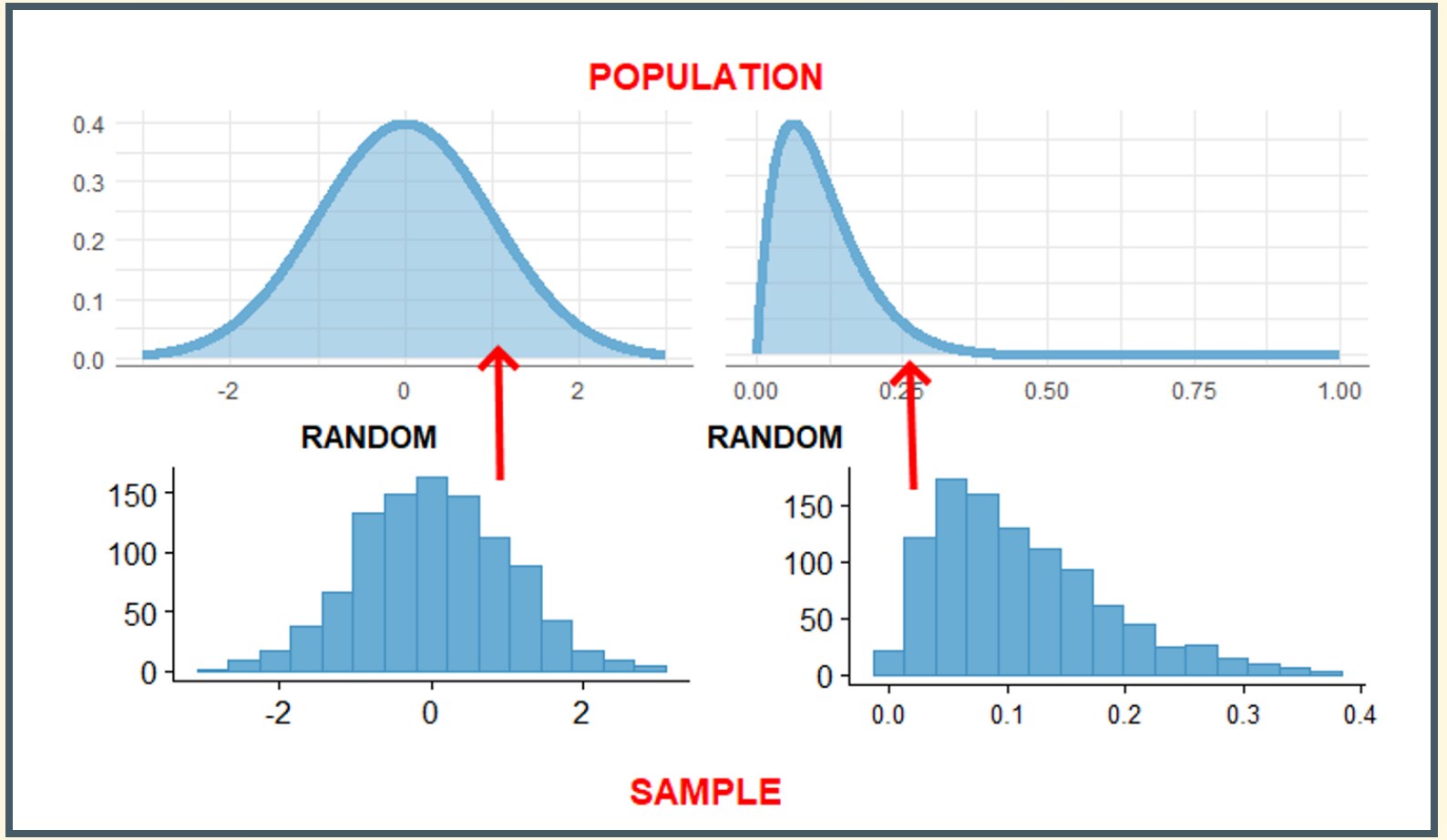
Learning objectives

- Understanding how to move from empirical to theoretical distributions
- Be able to calculate and interpret point and interval estimates (confidence intervals)

From sample to population



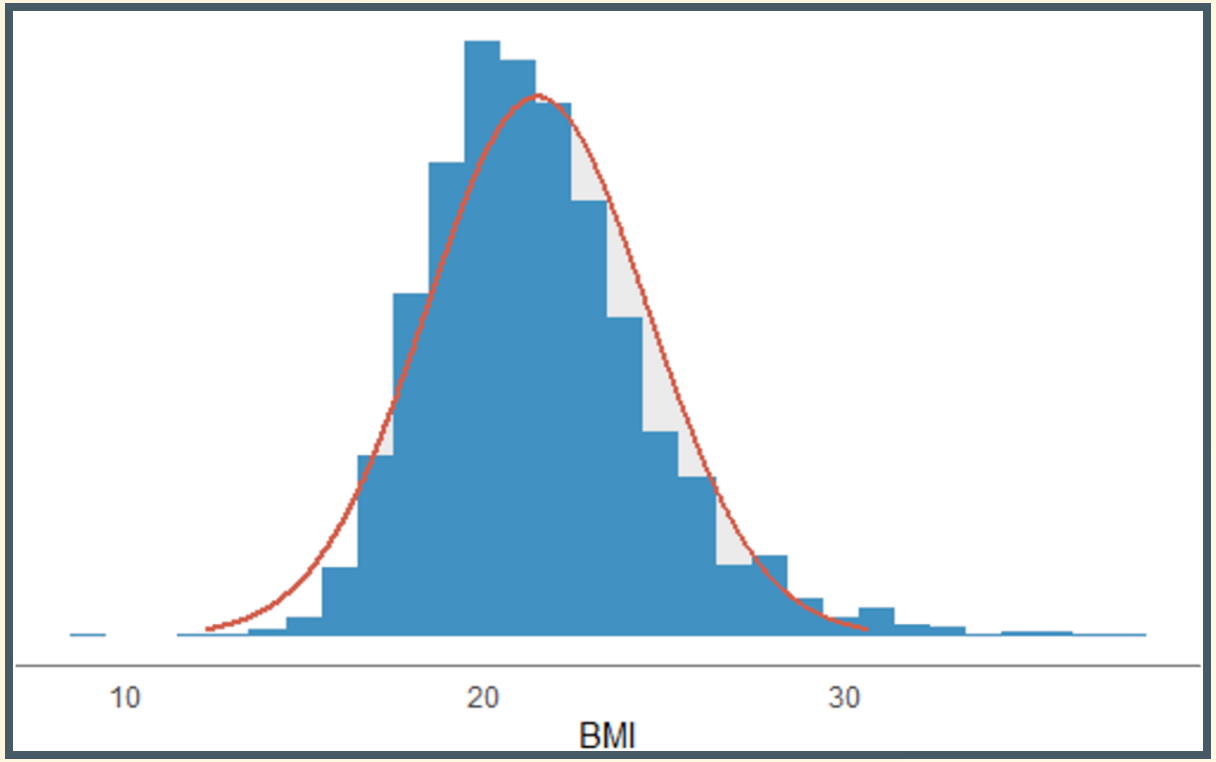
From sample to population



Point estimates

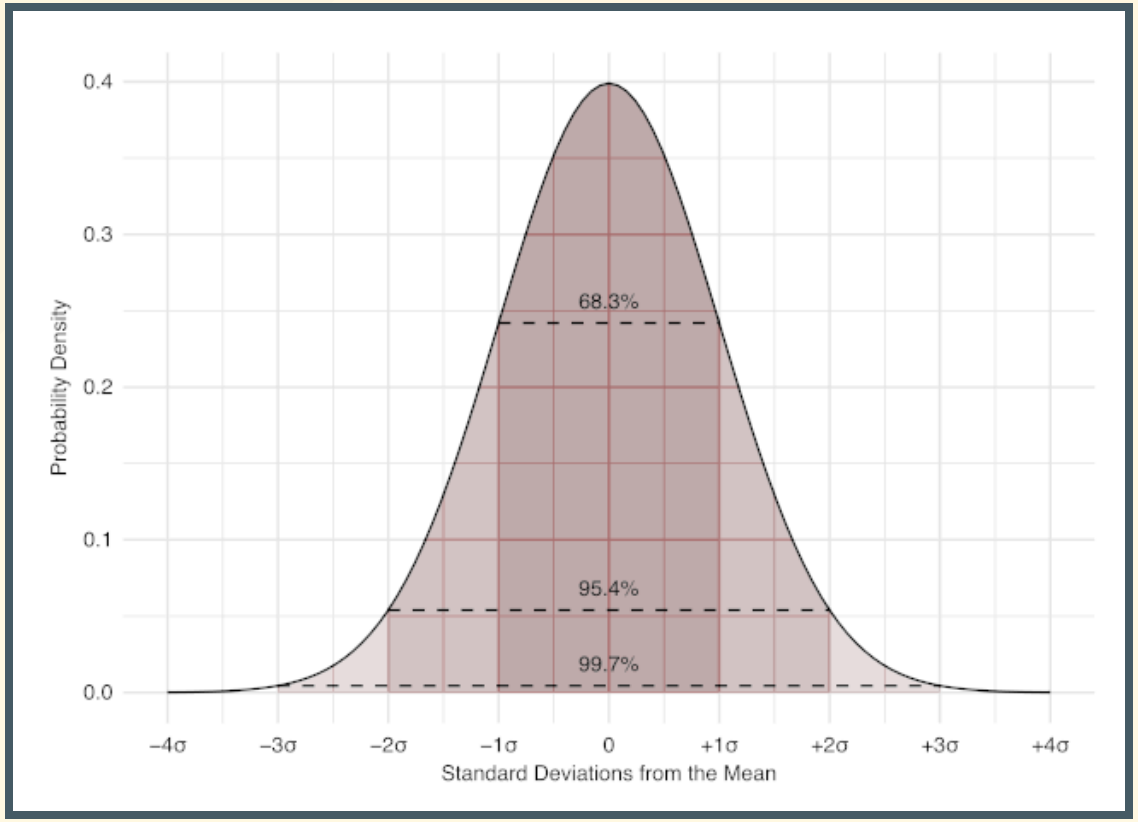
📌 $n = 6,705$ 15 y.o. males
$\mu_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$\sigma_{\text{BMI}} = 3.1\ \text{kg/m}^2$

What percentage of 15 y.o. males have $\text{BMI} > 25$ in the population?



Confidence intervals

🎯 A confidence interval (CI) is a range of values which includes the estimated parameter with a given degree of confidence

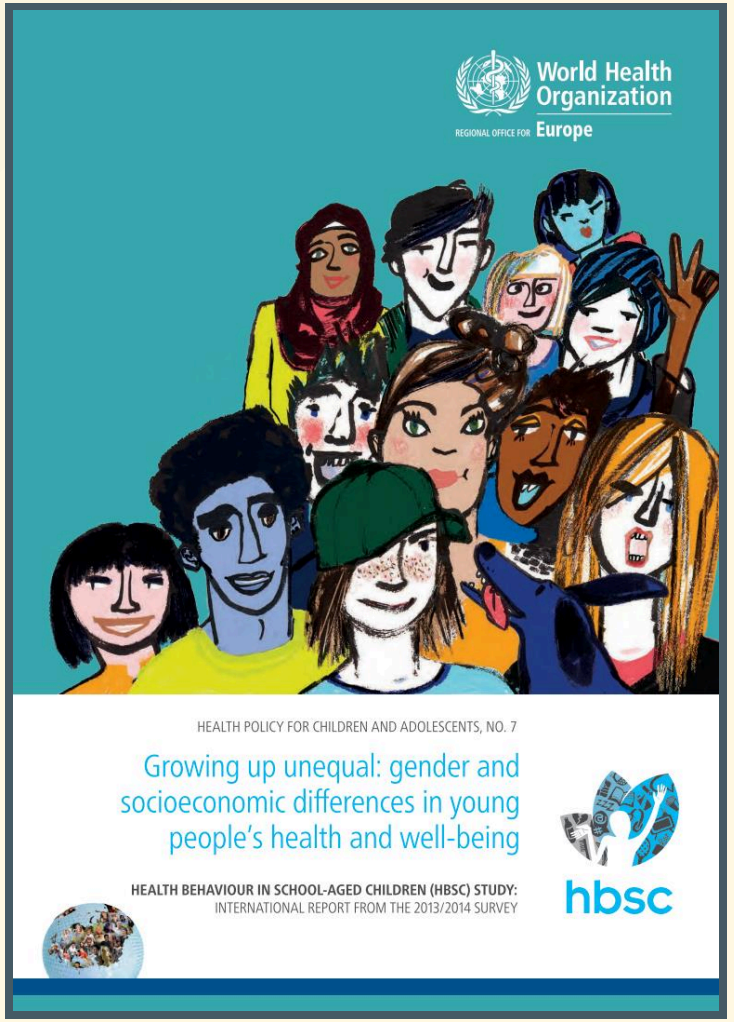


Confidence intervals for means

📌 $n = 6,705$ 15 y.o. males

$\bar{x}_{\text{BMI}} = 21.5\ \text{kg/m}^2$

$s_{\text{BMI}} = 3.1\ \text{kg/m}^2$



https://hbsc.org

Confidence intervals for means

📌 $n = 6,705$ 15 y.o. males

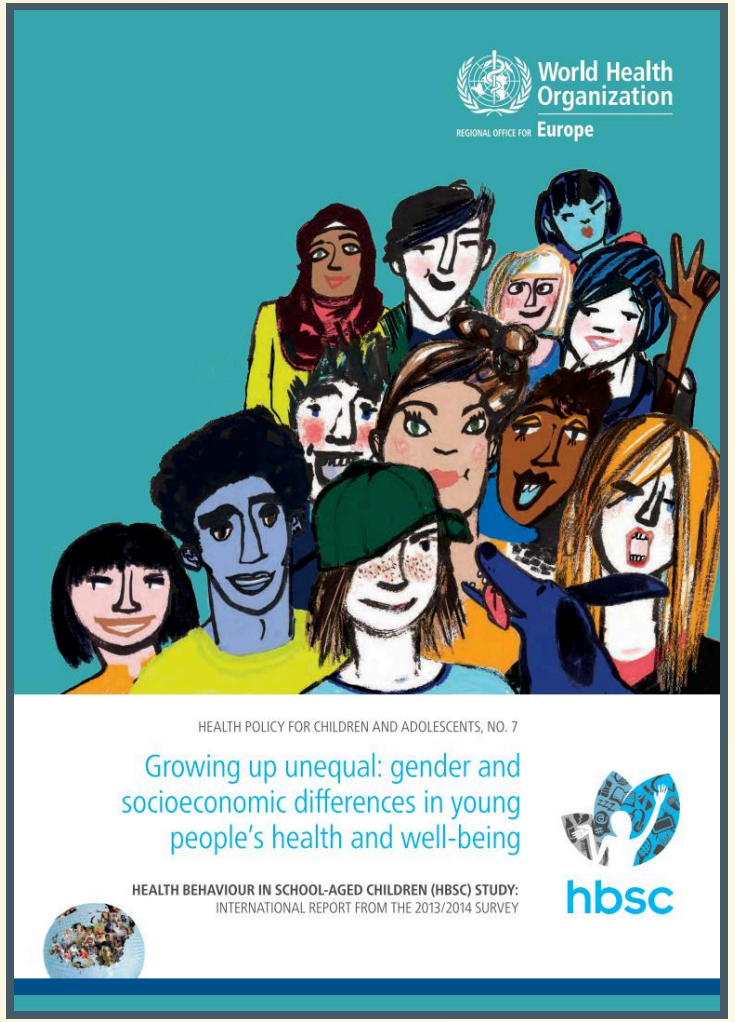
$\bar{x}_{\text{BMI}} = 21.5\ \text{kg/m}^2$

$s_{\text{BMI}} = 3.1\ \text{kg/m}^2$

$\text{SE} = \sigma/\sqrt{n}$, where $\sigma =$?

$\hat{\text{SE}} = s/\sqrt{n} = \frac{3.1}{\sqrt{(6{,}705)}} = 0.038$

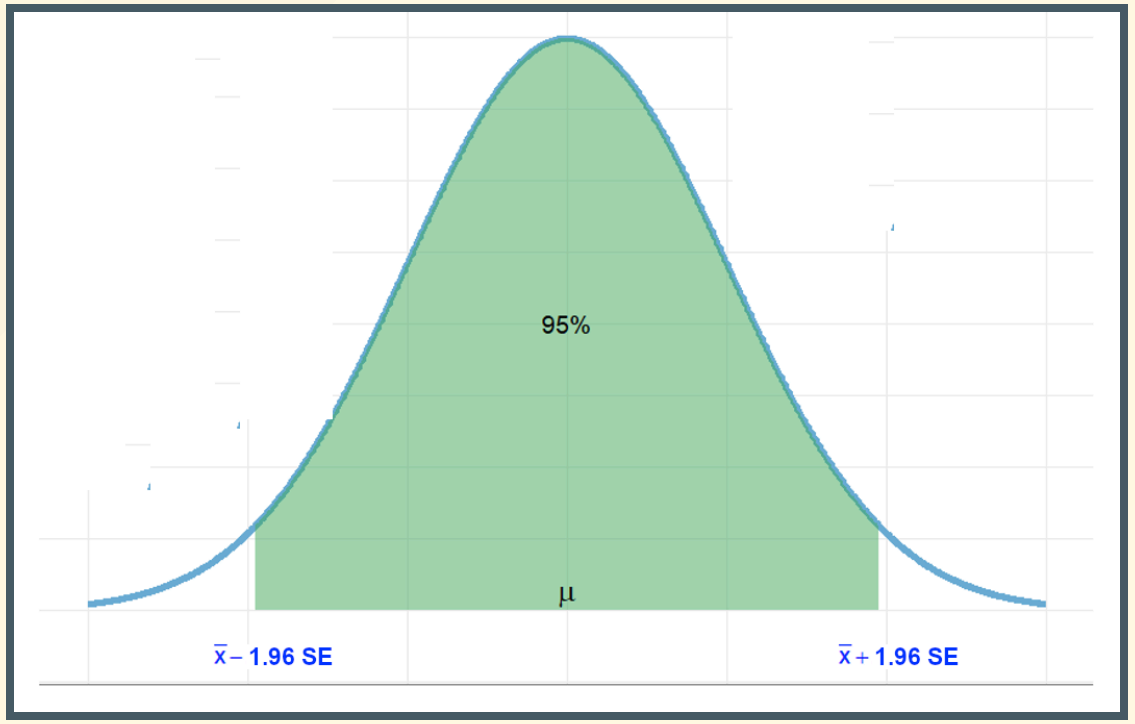
https://hbsc.org



Confidence intervals for means

📌 $n = 6,705$ 15 y.o. males
$\bar{x}_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$s_{\text{BMI}} = 3.1\ \text{kg/m}^2$

$$\hat{\text{SE}} = s/\sqrt{n} = \frac{3.1}{\sqrt{(6{,}705)}} = 0.038$$



$$\mathcal{P}(\bar{x} - 1.96 \times SE \leq \mu \leq \bar{x} + 1.96 \times \hat{SE}) = 95\%$$

Confidence intervals for means

📌 $n = 6,705$ 15 y.o. males
$\bar{x}_{\text{BMI}} = 21.5\ \text{kg/m}^2$
$s_{\text{BMI}} = 3.1\ \text{kg/m}^2$

$$\hat{\text{SE}} = s/\sqrt{n} = \frac{3.1}{\sqrt{(6{,}705)}} = 0.038$$

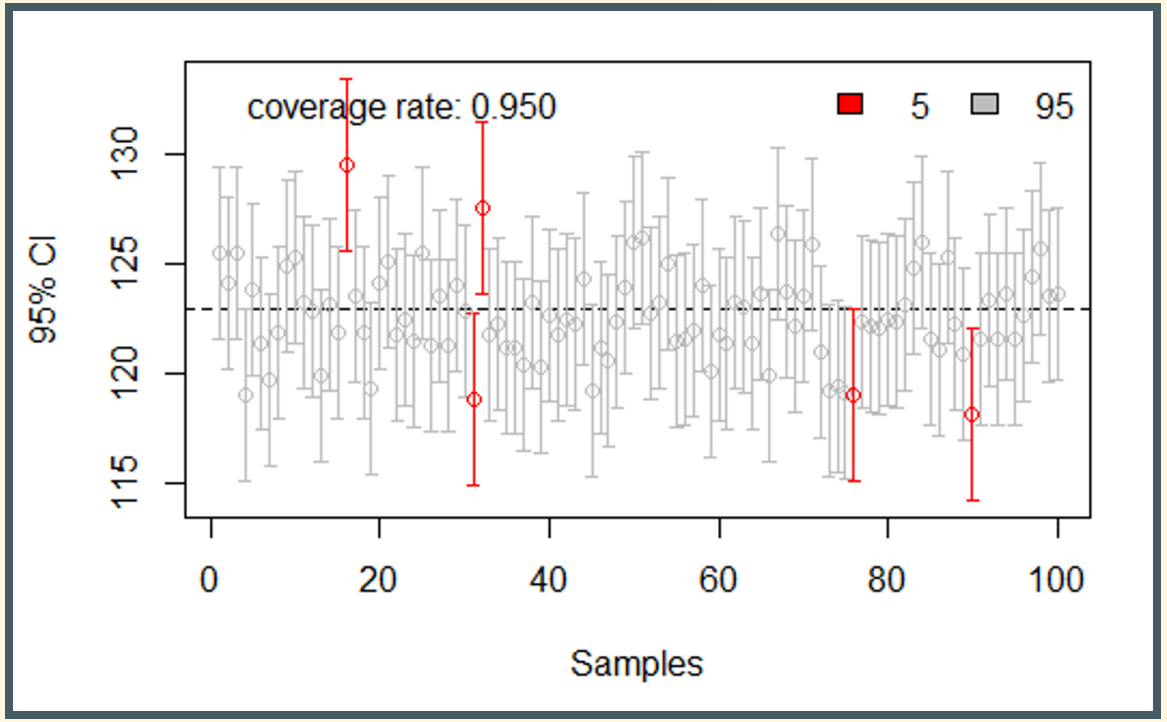
$$\mathcal{P}(\bar{x} - 1.96 \times \hat{SE} \leq \mu \leq \bar{x} + 1.96 \times \hat{SE}) =$$
$$= \mathcal{P}(21.5 - 1.96 \times 0.038 \leq \mu \leq 21.5 + 1.96 \times 0.038) =$$
$$= \mathcal{P}(21.42 \leq \mu \leq 21.58) = 95\%$$

$$\rightarrow 95\%\ \text{Confidence Interval (CI)} = (21.42; 21.58)$$

Interpreting confidence intervals

📌 Population: Italian women 25-74 years old

$\mu = 123$ mmHg



Exercise #10

? We take a random sample of 500 Italian women aged 25-74 with the following summary statistics

$$\bar{x} = 122.1 \text{ mmHg}$$
$$s = 19.8 \text{ mmHg}$$

What is the 95% CI for the true mean μ?

Exercise #10 -- Solution

? $n = 500$
$\bar{x} = 122.1 \text{ mmHg}$
$s = 19.8 \text{ mmHg}$

What is the 95% CI for the true mean μ?

$$\hat{\text{SE}} = s/\sqrt{n} = \frac{19.8}{\sqrt{(500)}} = 0.89$$

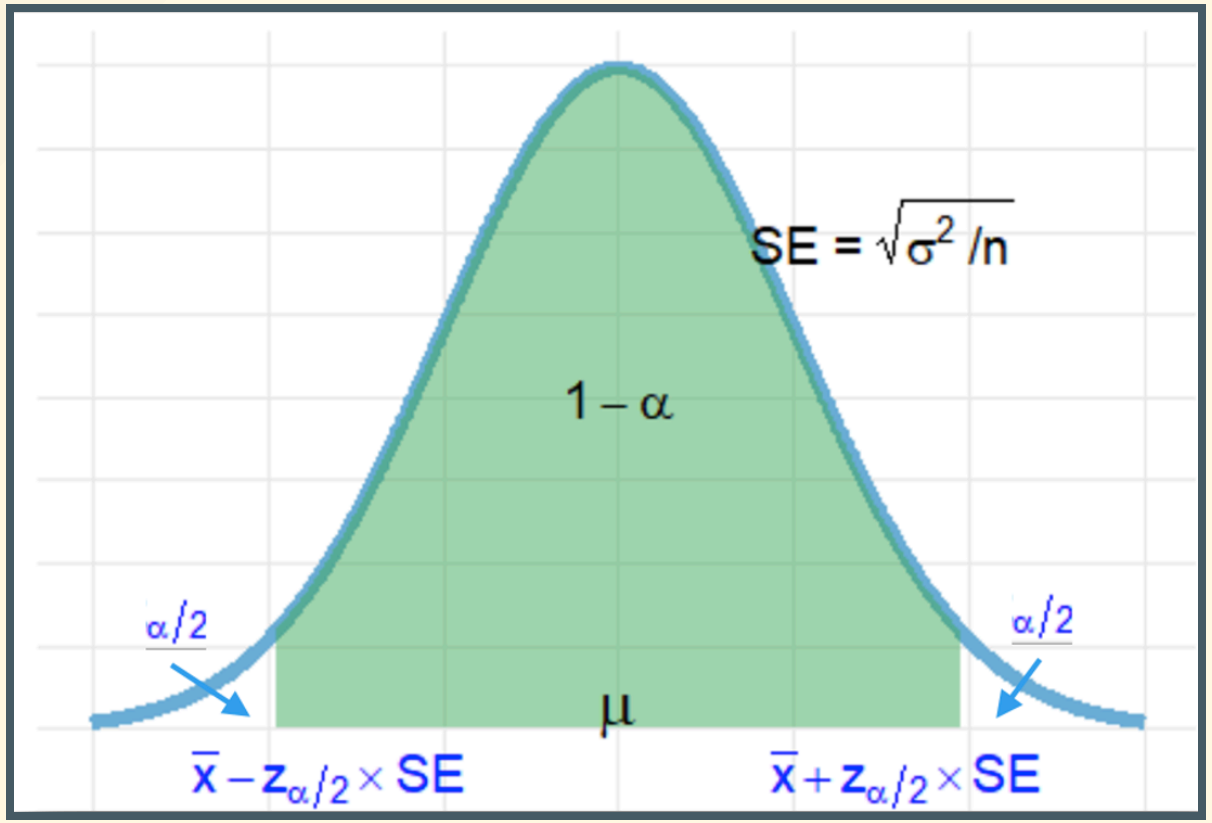
$$95\% \text{ CI} = (\bar{x} - 1.96 \times \hat{SE} \, ; \bar{x} + 1.96 \times \hat{SE}) =$$
$$= (122.1 - 1.96 \times 0.89 \, ; 122.1 + 1.96 \times 0.89)$$
$$= (120.34; 123.86)$$

The α level

🎯 $95\%\ \text{CI} = (\bar{x} - 1.96 \times \hat{SE}\ ;\ \bar{x} + 1.96 \times \hat{SE})$

1.96 ?

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	



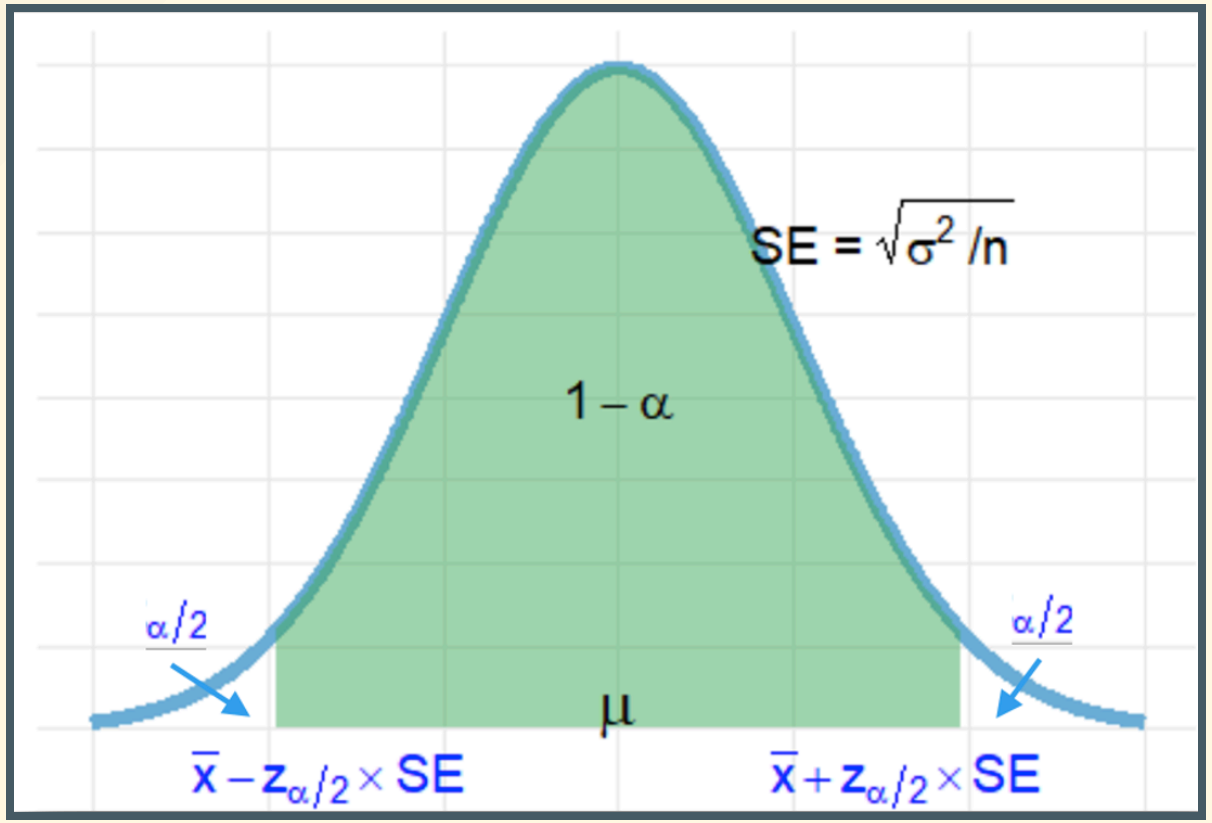
The α level

🎯 $95\%\ \text{CI} = (\bar{x} - 1.96 \times \hat{SE}\ ;\ \bar{x} + 1.96 \times \hat{SE})$

1.96 ?

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	

$100\% - 2.5\% = 97.5\%$



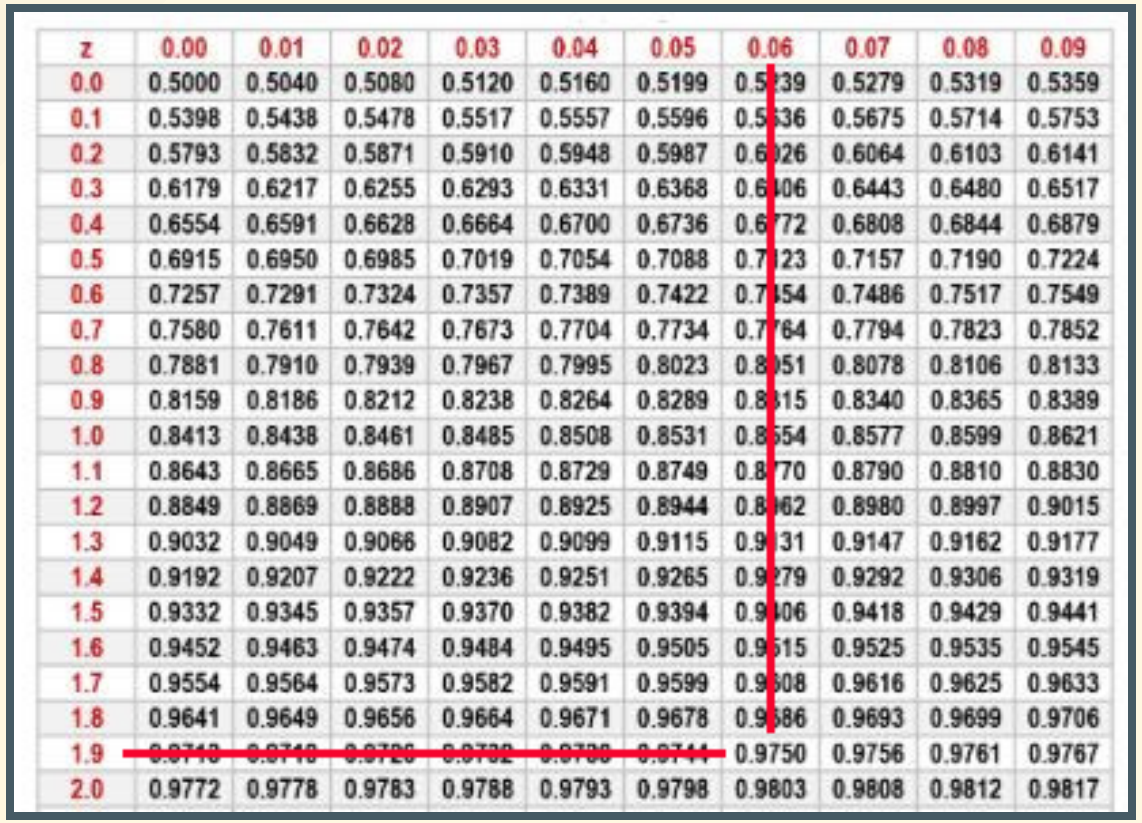
The α level

🎯 $95\%\ \text{CI} = (\bar{x} - 1.96 \times \hat{SE}\ ;\ \bar{x} + 1.96 \times \hat{SE})$

1.96 ?

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	1.96

$100\% - 2.5\% = 97.5\% \rightarrow z = 1.96$

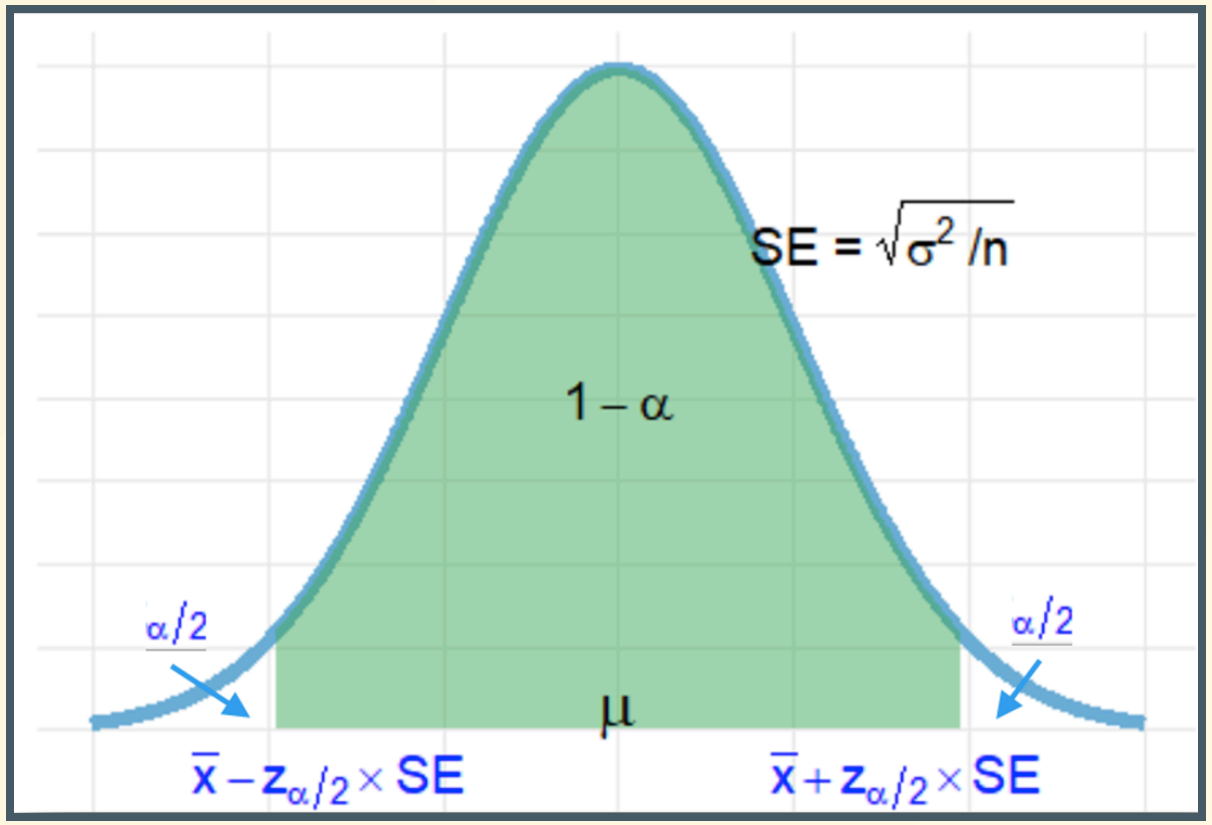


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293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

The α level

Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	1.96
90%	10%	5.0%	1.65
99%	1%	0.5%	2.58

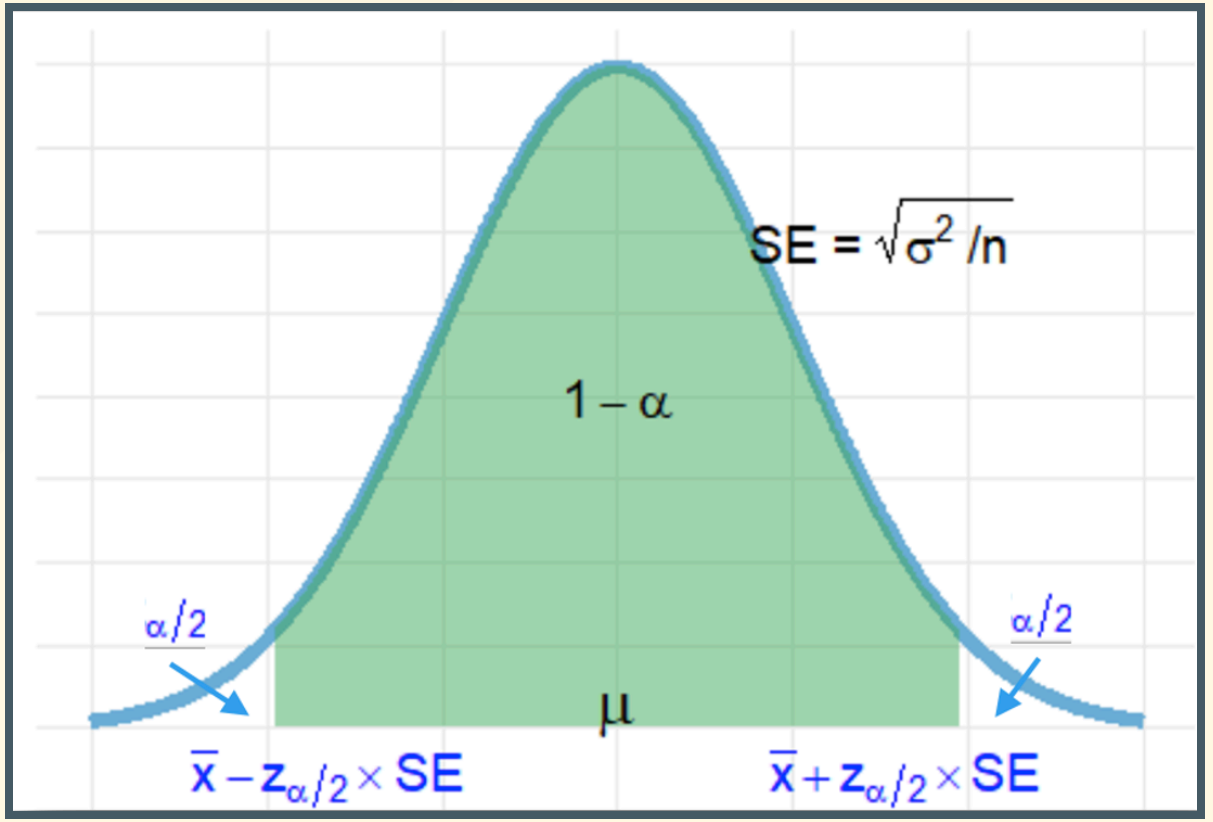
$$100\% - 2.5\% = 97.5\% \rightarrow z = 1.96$$
$$100\% - 5.0\% = 95.0\% \rightarrow z = 1.65$$
$$100\% - 0.5\% = 99.5\% \rightarrow z = 2.58$$



Exercise #11

? If the CI is large we are...

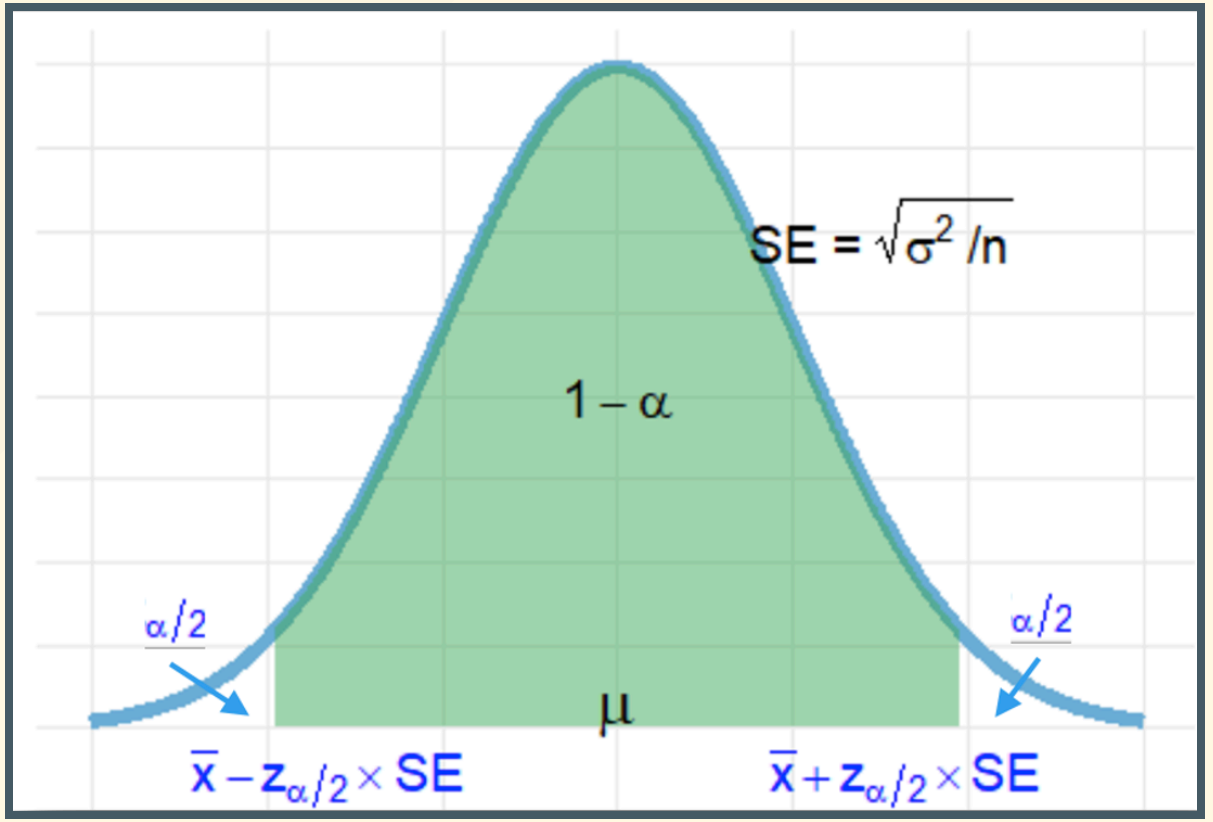
a) more likely of including μ
b) less likely of including μ
c) there is no difference



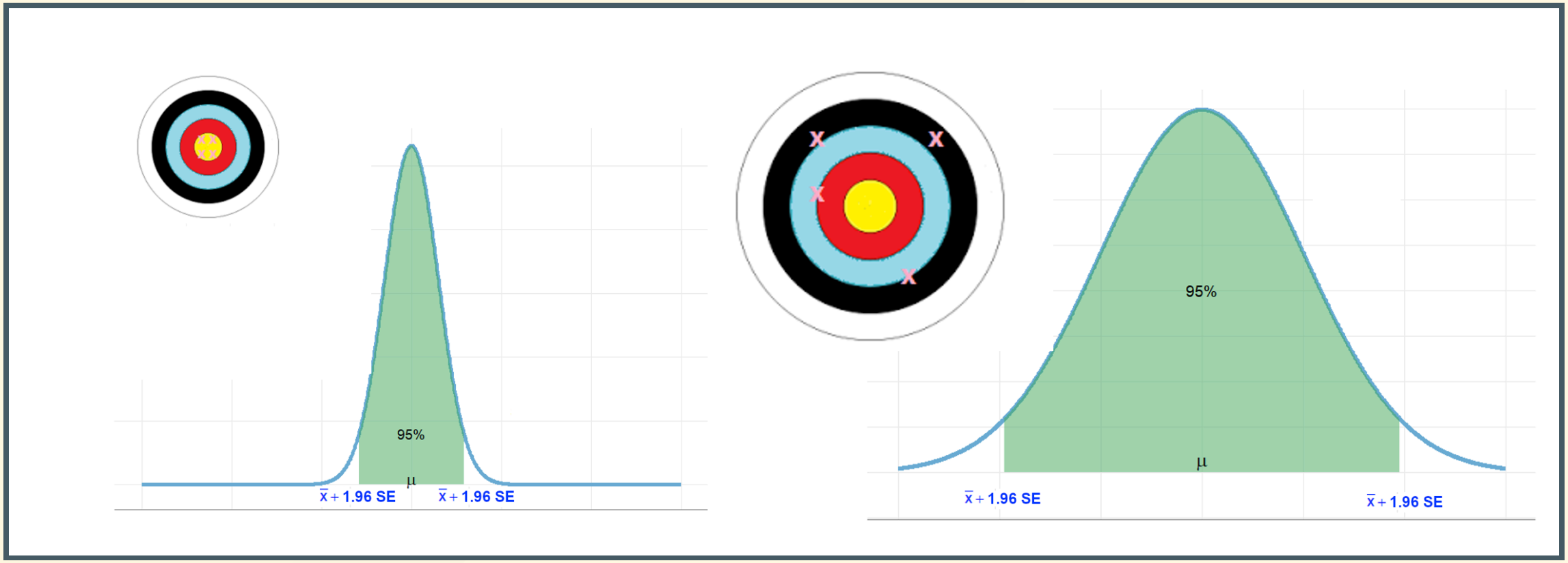
Exercise #11 -- Solution

? If the CI is large we are...

a) more likely of including μ ✅
b) less likely of including μ
c) there is no difference



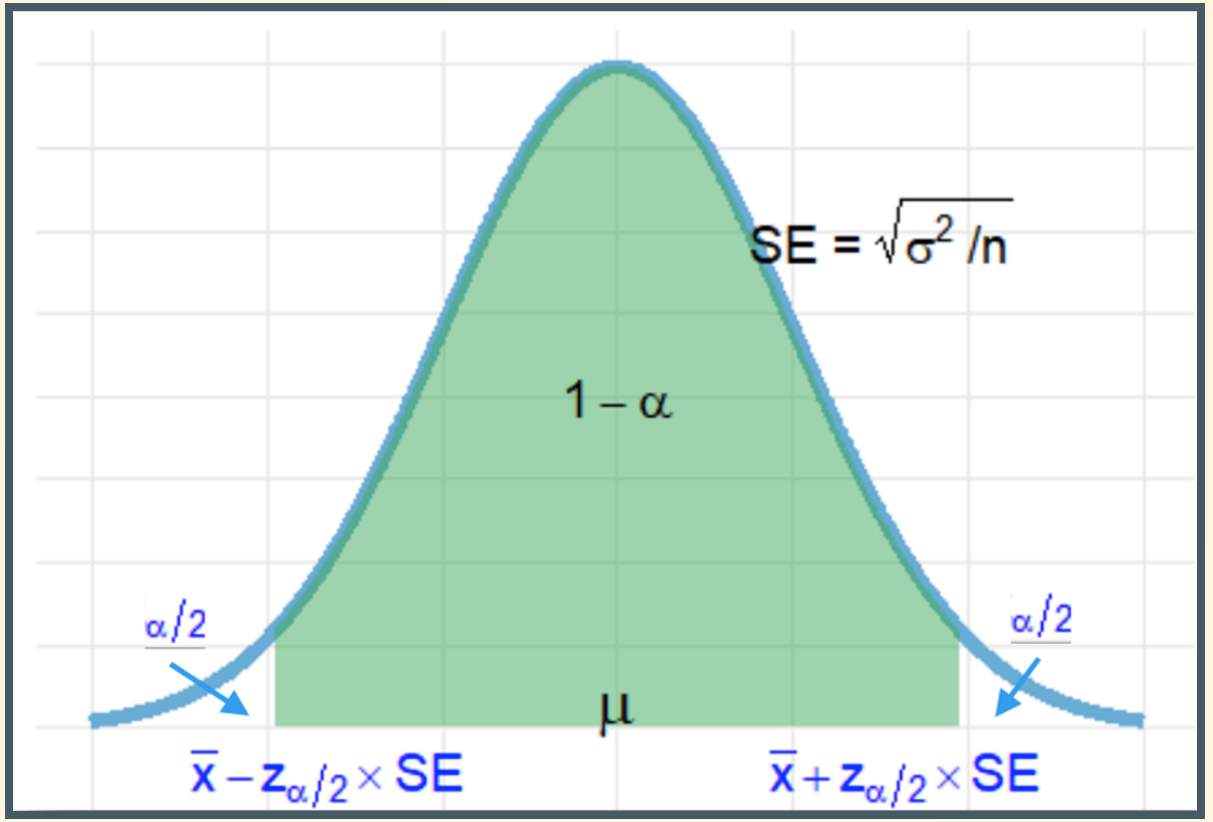
Exercise #11 -- Solution



Exercise #12

? If the CI is large we are...

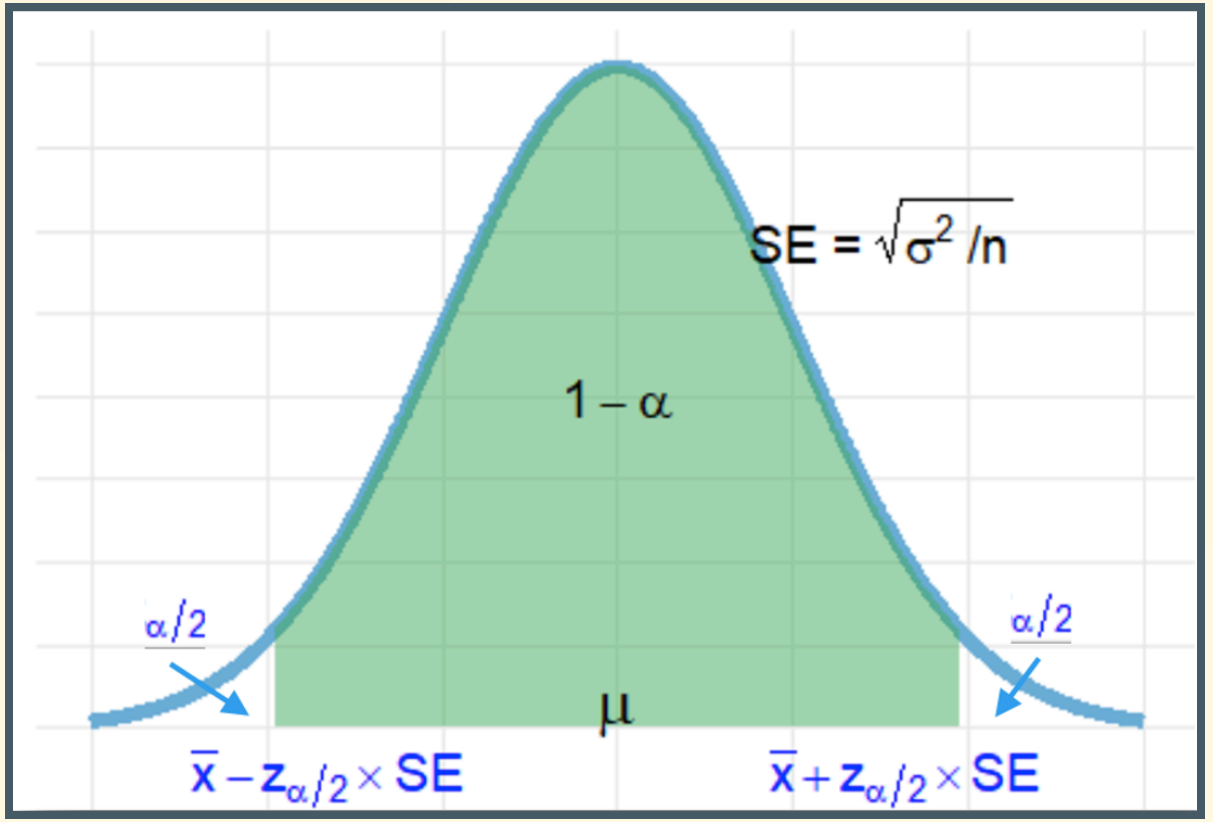
a) more precise
b) less precise
c) there is no difference



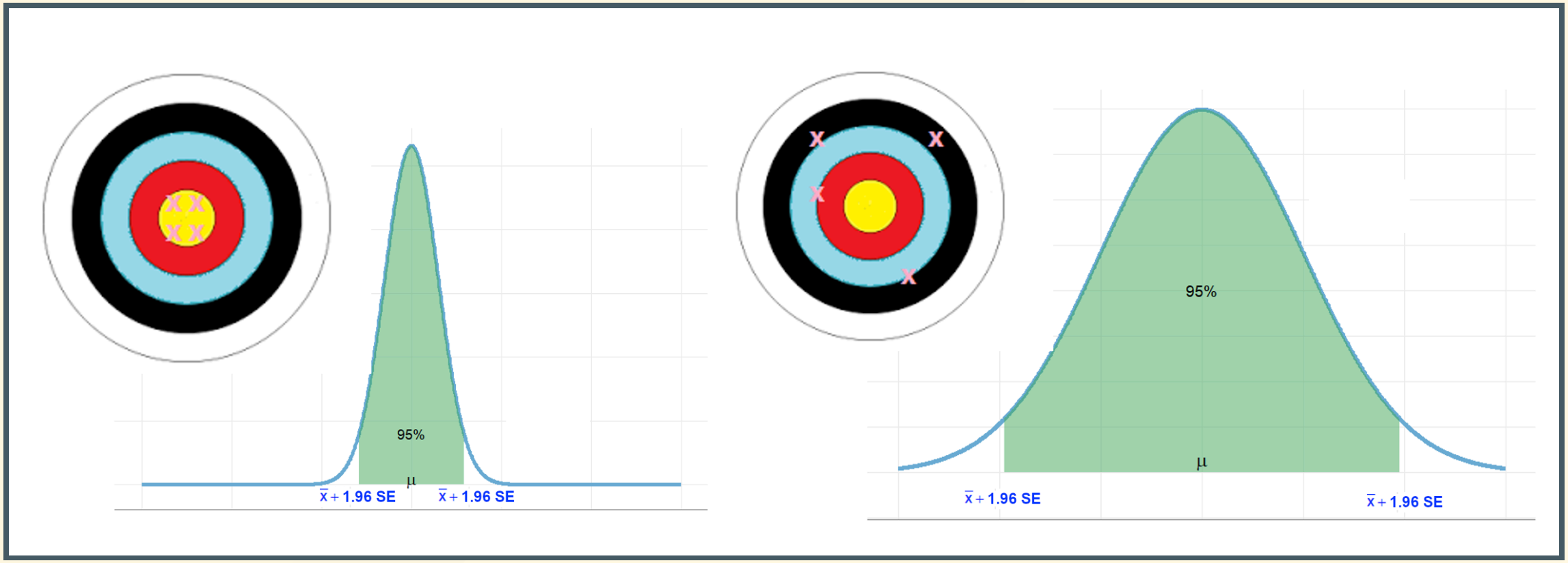
Exercise #12 -- Solution

? If the CI is large we are...

a) more precise
b) less precise ✅
c) there is no difference



Exercise #12 -- Solution



Confidence intervals for differences of means

📌 Are the intervention (i) and the control (c) group different?

Confidence intervals for differences of means

📌 Are the intervention (i) and the control (c) group different?

$$\mathcal{N} = (\mu_{\mathrm{i}} - \mu_{\mathrm{c}}, \sqrt{\frac{\sigma_{\mathrm{i}}^2}{n_{\mathrm{i}}} + \frac{\sigma_{\mathrm{c}}^2}{n_{\mathrm{c}}}})$$
$$\rightarrow \mathrm{CLT}$$

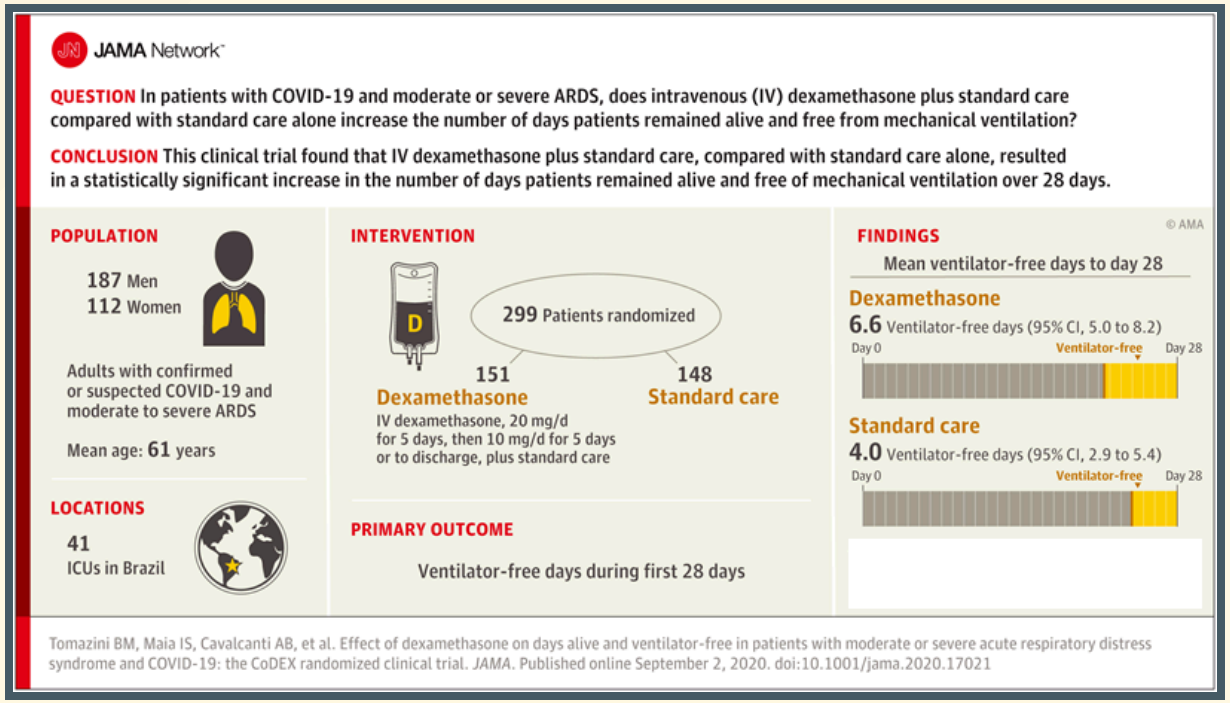
$$\hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}} + \frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}}}$$

Confidence intervals for differences of means

📌 Are the intervention (i) and the control (c) group different?

$$n_\mathrm{i} = 151, \bar{x}_\mathrm{i} = 6.6, s_\mathrm{i} = 10.0$$
$$n_\mathrm{c} = 148, \bar{x}_\mathrm{c} = 4.0, s_\mathrm{c} = 8.7$$

$$95\%\ \mathrm{CI} = \ ?$$



Confidence intervals for differences of means

📌 $n_\text{i} = 151, \bar{x}_\text{i} = 6.6, s_\text{i} = 10.0$
$n_\text{c} = 148, \bar{x}_\text{c} = 4.0, s_\text{c} = 8.7$

$\bar{x}_\text{i} - \bar{x}_\text{c} = 6.6 - 4.0 = 2.6$

$$\hat{\text{SE}} = \sqrt{\frac{s_\text{i}^2}{n_\text{i}} + \frac{s_\text{c}^2}{n_\text{c}}} = \sqrt{\frac{10.0^2}{151} + \frac{8.7^2}{148}} = 1.08$$

Confidence intervals for differences of means

📌 $n_\mathrm{i} = 151, \bar{x}_\mathrm{i} = 6.6, s_\mathrm{i} = 10.0$
$n_\mathrm{c} = 148, \bar{x}_\mathrm{c} = 4.0, s_\mathrm{c} = 8.7$

$$\bar{x}_\mathrm{i} - \bar{x}_\mathrm{c} = 6.6 - 4.0 = 2.6$$
$$\hat{\mathrm{SE}} = \sqrt{\frac{s_\mathrm{i}^2}{n_\mathrm{i}} + \frac{s_\mathrm{i}^2}{n_\mathrm{i}}} = \sqrt{\frac{10.0^2}{151} + \frac{8.7^2}{148}} = 1.08$$

$$\begin{aligned} 95\%\ \mathrm{CI} &= (\bar{x}_\mathrm{i} - \bar{x}_\mathrm{c}) - 1.96 \times \hat{SE};\ (\bar{x}_\mathrm{i} - \bar{x}_\mathrm{c}) + 1.96 \times \hat{SE}) = \\ &= (2.6 - 1.96 \times 1.08;\ 2.6 + 1.96 \times 1.08) = \\ &= (0.48;\ 4.72) \end{aligned}$$

Exercise #13

? $n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$
$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$

$$\bar{x}_{\mathrm{i}} - \bar{x}_{\mathrm{c}} = 6.6 - 4.0 = 2.6$$

$$\hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}} = \sqrt{\frac{10.0^2}{151} + \frac{8.7^2}{148}} = 1.08$$

$$90\%\ \mathrm{CI} = \ ?$$ $(z_{\alpha/2}=z_{5/2}=1.65)$

Exercise #13 -- Solution

? $n_\text{i} = 151, \bar{x}_\text{i} = 6.6, s_\text{i} = 10.0$
$n_\text{c} = 148, \bar{x}_\text{c} = 4.0, s_\text{c} = 8.7$

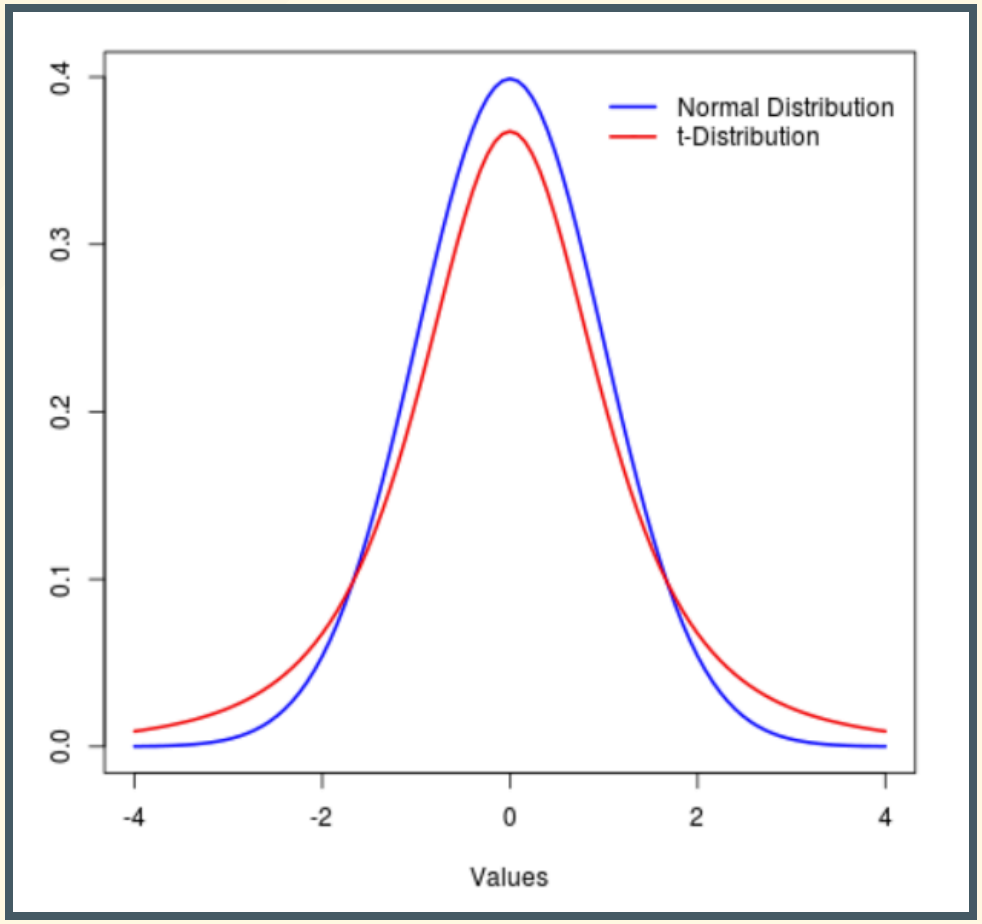
$$\bar{x}_\text{i} - \bar{x}_\text{c} = 6.6 - 4.0 = 2.6$$

$$\hat{\text{SE}} = \sqrt{\frac{s_\text{i}^2}{n_\text{i}} + \frac{s_\text{i}^2}{n_\text{i}}} = \sqrt{\frac{10.0^2}{151} + \frac{8.7^2}{148}} = 1.08$$

$$90\%\ \text{CI} = (2.6 - 1.65 \times 1.08;\ 2.6 + 1.65 \times 1.08) = (0.82;\ 4.38)$$

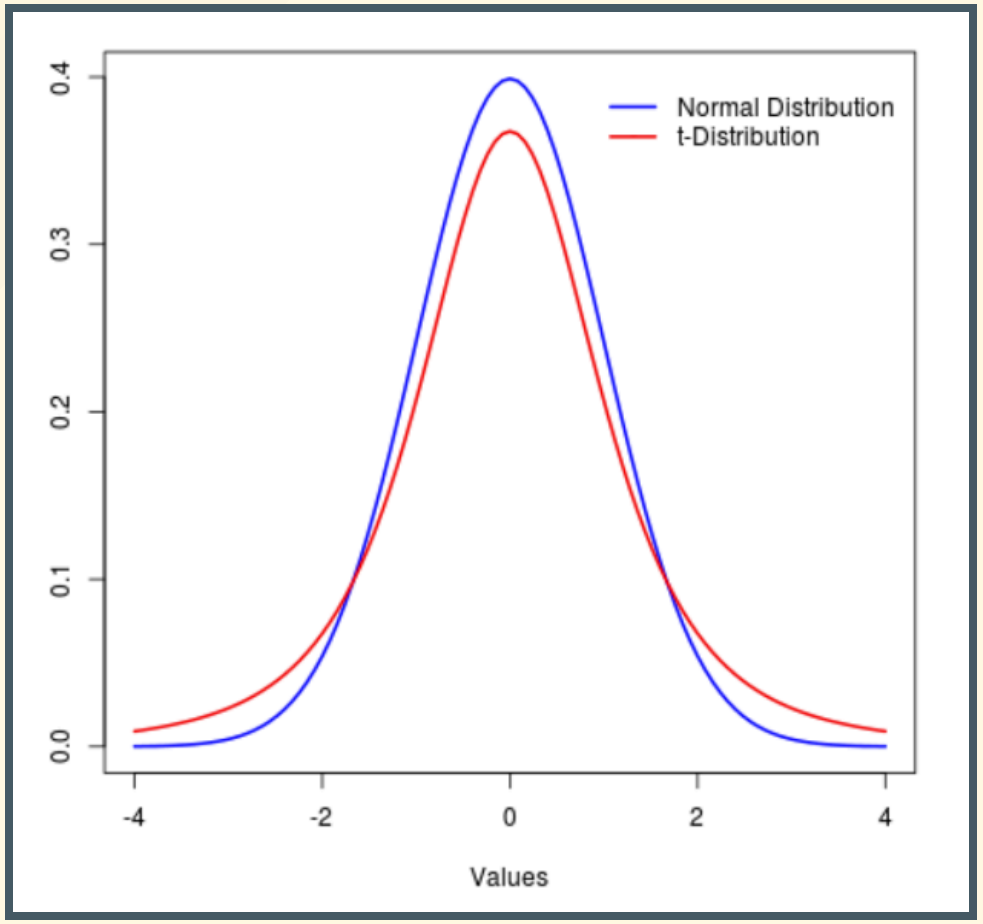
CTL and small sample size

- CTL works for large sample size
- t distribution



CTL and small sample size

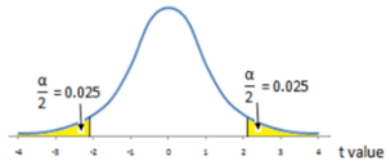
- CTL works for large sample size
- t distribution
  - keeps into account the degree of freedom (df)
  - one sample of size $n \rightarrow df = n - 1$
  - two samples of size $n_1 \wedge n_2 \rightarrow df = n_1 - 1 + n_2 - 1 = \\ = n_1 + n_2 - 2$



CTL and small sample size

- CTL works for large sample size
- t distribution
  - keeps into account the degree of freedom (df)
  - one sample of size $n \rightarrow df = n - 1$
  - two samples of size $n_1 \wedge n_2 \rightarrow df = n_1 - 1 + n_2 - 1 = n_1 + n_2 - 2$

$$95\%\ \text{CI} = (\bar{x} - t \times \hat{SE};\ \bar{x} + t \times \hat{SE})$$



Confidence	90	95	95.45	99	99.73
alpha	0.1000	0.0500	0.0455	0.0100	0.0027
df					
1	6.314	12.706	13.968	63.657	235.784
2	2.920	4.303	4.527	9.925	19.206
3	2.353	3.182	3.307	5.841	9.219
4	2.132	2.776	2.869	4.604	6.620
5	2.015	2.571	2.649	4.032	5.507
6	1.943	2.447	2.517	3.707	4.904
7	1.895	2.365	2.429	3.499	4.530
8	1.860	2.306	2.366	3.355	4.277
9	1.833	2.262	2.320	3.250	4.094
10	1.812	2.228	2.284	3.169	3.957
11	1.796	2.201	2.255	3.106	3.850
12	1.782	2.179	2.231	3.055	3.764
13	1.771	2.160	2.212	3.012	3.694
14	1.761	2.145	2.195	2.977	3.636
15	1.753	2.131	2.181	2.947	3.586
16	1.746	2.120	2.169	2.921	3.544
17	1.740	2.110	2.158	2.898	3.507
18	1.734	2.101	2.149	2.878	3.475
19	1.729	2.093	2.140	2.861	3.447
20	1.725	2.086	2.133	2.845	3.422
21	1.721	2.080	2.126	2.831	3.400
22	1.717	2.074	2.120	2.819	3.380
23	1.714	2.069	2.115	2.807	3.361
24	1.711	2.064	2.110	2.797	3.345
25	1.708	2.060	2.105	2.787	3.330
26	1.706	2.056	2.101	2.779	3.316
27	1.703	2.052	2.097	2.771	3.303
28	1.701	2.048	2.093	2.763	3.291
29	1.699	2.045	2.090	2.756	3.280
30	1.697	2.042	2.087	2.750	3.270
40	1.684	2.021	2.064	2.704	3.199
50	1.676	2.009	2.051	2.678	3.157
60	1.671	2.000	2.043	2.660	3.130
70	1.667	1.994	2.036	2.648	3.111
80	1.664	1.990	2.032	2.639	3.096
90	1.662	1.987	2.028	2.632	3.085
100	1.660	1.984	2.025	2.626	3.077
1000	1.646	1.962	2.003	2.581	3.007
∞	1.645	1.960	2.000	2.576	3.000

CTL and small sample size

📌 $n = 58$ patients with T2D

$\bar{x}_{\text{BMI}} = 25.0\ \text{kg/m}^2$

$s_{\text{BMI}} = 2.7\ \text{kg/m}^2$

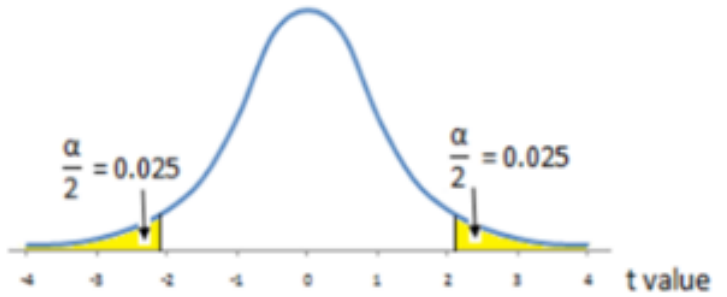
$95\%\ \text{CI} = \ ?$

CTL and small sample size

- $n = 58$ patients with T2D
  $\bar{x}_{\text{BMI}} = 25.0\ \text{kg/m}^2$
  $s_{\text{BMI}} = 2.7\ \text{kg/m}^2$

$$\hat{\text{SE}} = s/\sqrt{n} = \frac{2.7}{\sqrt{(58)}} = 0.36$$

$$\text{df} = n - 1 = 58 - 1 = 57$$



Confidence	90	95	95.45	99	99.73
alpha	0.1000	0.0500	0.0455	0.0100	0.0027
df					
1	6.314	12.706	13.968	63.657	235.784
2	2.920	4.303	4.527	9.925	19.206
3	2.353	3.182	3.307	5.841	9.219
4	2.132	2.776	2.869	4.604	6.620
29	1.699	2.045	2.090	2.756	3.280
30	1.697	2.042	2.087	2.750	3.270
40	1.684	2.021	2.064	2.704	3.199
50	1.676	2.009	2.051	2.678	3.157
60	1.671	2.000	2.043	2.660	3.130
70	1.667	1.994	2.036	2.648	3.111
80	1.664	1.990	2.032	2.639	3.096
90	1.662	1.987	2.028	2.632	3.085
100	1.660	1.984	2.025	2.626	3.077
1000	1.646	1.962	2.003	2.581	3.007
∞	1.645	1.960	2.000	2.576	3.000

CTL and small sample size

📌 $n = 58$ patients with T2D

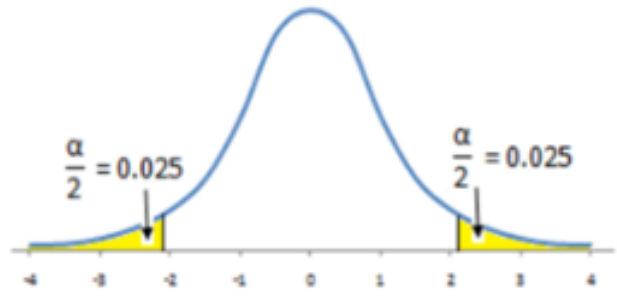
$\bar{x}_{\text{BMI}} = 25.0\ \text{kg/m}^2$

$s_{\text{BMI}} = 2.7\ \text{kg/m}^2$

$$\hat{\text{SE}} = s/\sqrt{n} = \frac{2.7}{\sqrt{(58)}} = 0.36$$

$$\text{df} = n - 1 = 58 - 1 = 57$$

$$95\%\ \text{CI} = (\bar{x} - t \times \hat{SE};\ \bar{x} + t \times \hat{SE}) =$$
$$= (25.0 - t \times 0.36;\ 25.0 + t \times 0.36) =$$
$$= (25.0 - 2 \times 0.36;\ 25.0 + 2 \times 0.36) =$$
$$= (25.0 - 0.72;\ 25.0 + 0.72) =$$
$$= (24.28; 25.72)$$



Confidence	90	95	95.45	99	99.73
alpha	0.1000	0.0500	0.0455	0.0100	0.0027
df					
1	6.314	12.706	13.968	63.657	235.784
2	2.920	4.303	4.527	9.925	19.206
3	2.353	3.182	3.307	5.841	9.219
4	2.132	2.776	2.869	4.604	6.620
29	1.699	2.045	2.090	2.756	3.280
30	1.697	2.042	2.087	2.750	3.270
40	1.684	2.021	2.064	2.704	3.199
50	1.676	2.009	2.051	2.678	3.157
60	1.671	2.000	2.043	2.660	3.130
70	1.667	1.994	2.036	2.648	3.111
80	1.664	1.990	2.032	2.639	3.096
90	1.662	1.987	2.028	2.632	3.085
100	1.660	1.984	2.025	2.626	3.077
1000	1.646	1.962	2.003	2.581	3.007
∞	1.645	1.960	2.000	2.576	3.000

Confidence intervals for proportions

📌 How many patients show gastrointestinal side effects from a new medication?

$$\mathcal{N} = (\pi, \frac{\pi \times (1-\pi)}{n}) \rightarrow \text{CLT}$$

$$\hat{SE} = \sqrt{\frac{\bar{p} \times (1-\bar{p})}{n}}, \text{ where } \bar{p} = \frac{m}{n}$$

Confidence intervals for proportions

📌 How many patients show gastrointestinal side effects from a new medication?

$n = 100, \;\; m = 69$

$95\%\ \mathrm{CI} = \;\; ?$

Confidence intervals for proportions

📌 $n = 100, \;\; m = 69$

$$\bar{p} = \frac{m}{n} = \frac{69}{100} = 0.69$$

$$\hat{SE} = \sqrt{\frac{\bar{p} \times (1 - \bar{p})}{n}} = \sqrt{\frac{0.69 \times (1 - 0.69)}{100}} = 0.046$$

Confidence intervals for proportions

📌 $n = 100, \;\; m = 69$

$$\bar{p} = \frac{m}{n} = \frac{69}{100} = 0.69$$

$$\hat{SE} = \sqrt{\frac{\bar{p} \times (1 - \bar{p})}{n}} = \sqrt{\frac{0.69 \times (1 - 0.69)}{100}} = 0.046$$

$$\begin{aligned} 95\%\ \text{CI} &= (\bar{p} - 1.96 \times \hat{SE};\ \bar{p} + 1.96 \times \hat{SE}) = \\ &= (0.69 - 1.96 \times 0.046;\ 0.69 + 1.96 \times 0.046) = \\ &= (0.60;\ 0.78) \end{aligned}$$

Exercise #14

? $n = 100, \;\; m = 69$

$$\bar{p} = \frac{m}{n} = \frac{69}{100} = 0.69$$

$$\hat{SE} = \sqrt{\frac{\bar{p} \times (1 - \bar{p})}{n}} = \sqrt{\frac{0.69 \times (1 - 0.69)}{100}} = 0.046$$

$99\%\ \text{CI} = \ ?$ $(z_{\alpha/2} = z_{0.5/2} = 2.58)$

Exercise #14 -- Solution

? $n = 100, \;\; m = 69$

$$\bar{p} = \frac{m}{n} = \frac{69}{100} = 0.69$$

$$\hat{SE} = \sqrt{\frac{\bar{p} \times (1 - \bar{p})}{n}} = \sqrt{\frac{0.69 \times (1 - 0.69)}{100}} = 0.046$$

$$99\%\ \text{CI} = (\bar{p} - 2.58 \times \hat{SE};\ \bar{p} + 2.58 \times \hat{SE}) =$$
$$= (0.69 - 2.58 \times 0.046;\ 0.69 + 2.58 \times 0.046) = (0.57;\ 0.81)$$

Confidence intervals for differences of proportion

📌 Are the intervention (i) and the control (c) group different?

Confidence intervals for differences of proportion

📌 Are the intervention (i) and the control (c) group different?

$$\mathcal{N} = (\pi_{\mathrm{i}} - \pi_{\mathrm{c}}, \frac{\pi_i \times (1-\pi_i)}{n_i} + \frac{\pi_c \times (1-\pi_c)}{n_c})$$
$$\rightarrow \mathrm{CLT}$$

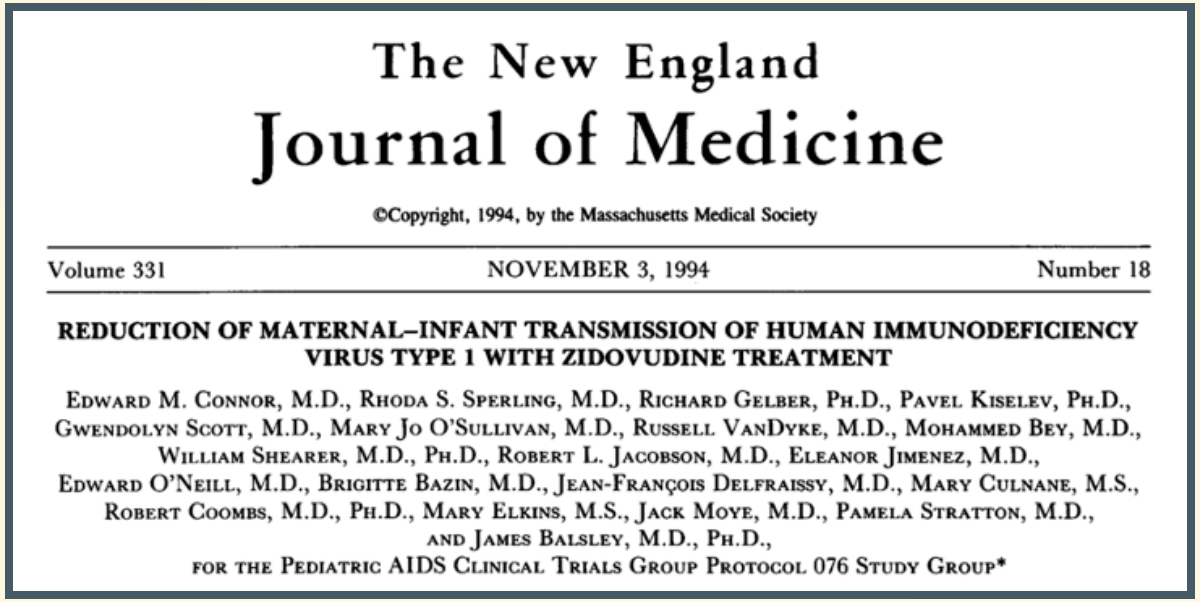
$$\hat{\mathrm{SE}} = \sqrt{\frac{\bar{p}_i \times (1-\bar{p}_i)}{n_i} + \frac{\bar{p}_c \times (1-\bar{p}_c)}{n_c}}$$

Confidence intervals for differences of proportion

📌 Are the intervention (i) and the control (c) group different?

$$n_\mathrm{i} = 180, m_\mathrm{i} = 13$$
$$n_\mathrm{c} = 183, m_\mathrm{c} = 40$$

$$95\%\ \mathrm{CI} = \ ?$$



The New England
Journal of Medicine

Volume 331 NOVEMBER 3, 1994 Number 18

REDUCTION OF MATERNAL–INFANT TRANSMISSION OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 WITH ZIDOVUDINE TREATMENT

EDWARD M. CONNOR, M.D., RHODA S. SPERLING, M.D., RICHARD GELBER, PH.D., PAVEL KISELEV, PH.D., GWENDOLYN SCOTT, M.D., MARY JO O'SULLIVAN, M.D., RUSSELL VANDYKE, M.D., MOHAMMED BEY, M.D., WILLIAM SHEARER, M.D., PH.D., ROBERT L. JACOBSON, M.D., ELEANOR JIMENEZ, M.D., EDWARD O'NEILL, M.D., BRIGITTE BAZIN, M.D., JEAN-FRANÇOIS DELFRAISSY, M.D., MARY CULNANE, M.S., ROBERT COOMBS, M.D., PH.D., MARY ELKINS, M.S., JACK MOYE, M.D., PAMELA STRATTON, M.D., AND JAMES BALSLEY, M.D., PH.D., FOR THE PEDIATRIC AIDS CLINICAL TRIALS GROUP PROTOCOL 076 STUDY GROUP*

Confidence intervals for differences of proportion

📌 Are the intervention (i) and the control (c) group different?

$$n_{\mathrm{i}} = 180, m_{\mathrm{i}} = 13$$
$$n_{\mathrm{c}} = 183, m_{\mathrm{c}} = 40$$

$$\bar{p}_{\mathrm{i}} - \bar{p}_{\mathrm{c}} = \frac{m_i}{n_i} - \frac{m_c}{n_c} = \frac{13}{180} - \frac{40}{183} = 0.07 - 0.22 = -0.15$$
$$\hat{SE} = \sqrt{\frac{0.07 \times (1-0.07)}{180} + \frac{0.22 \times (1-0.22)}{183}} = 0.036$$

Confidence intervals for differences of proportion

📌 Are the intervention (i) and the control (c) group different?

$$n_{\mathrm{i}} = 180, m_{\mathrm{i}} = 13$$
$$n_{\mathrm{c}} = 183, m_{\mathrm{c}} = 40$$

$$\bar{p}_{\mathrm{i}} - \bar{p}_{\mathrm{c}} = \frac{m_i}{n_i} - \frac{m_c}{n_c} = \frac{13}{180} - \frac{40}{183} = 0.07 - 0.22 = -0.15$$
$$\hat{SE} = \sqrt{\frac{0.07 \times (1-0.07)}{180} + \frac{0.22 \times (1-0.22)}{183}} = 0.036$$

$$95\%\ \mathrm{CI} = (\bar{p}_{\mathrm{i}} - \bar{p}_{\mathrm{c}} - 1.96 \times \hat{SE};\ \bar{p}_{\mathrm{i}} - \bar{p}_{\mathrm{c}} + 1.96 \times \hat{SE}) =$$
$$= (-0.15 - 0.007;\ \ -0.015 + 0.07) = (-0.22; -0.08)$$

Summary

- We use statistics to estimate parameters (point estimates)
- We can take advantage of the CLT to calculate interval estimates (CI)
- 95% confidence intervals tell us the the true value has 95% probability of being inside the given range

See you tomorrow