STATISTICS PART - III

LAXMINARAYEN

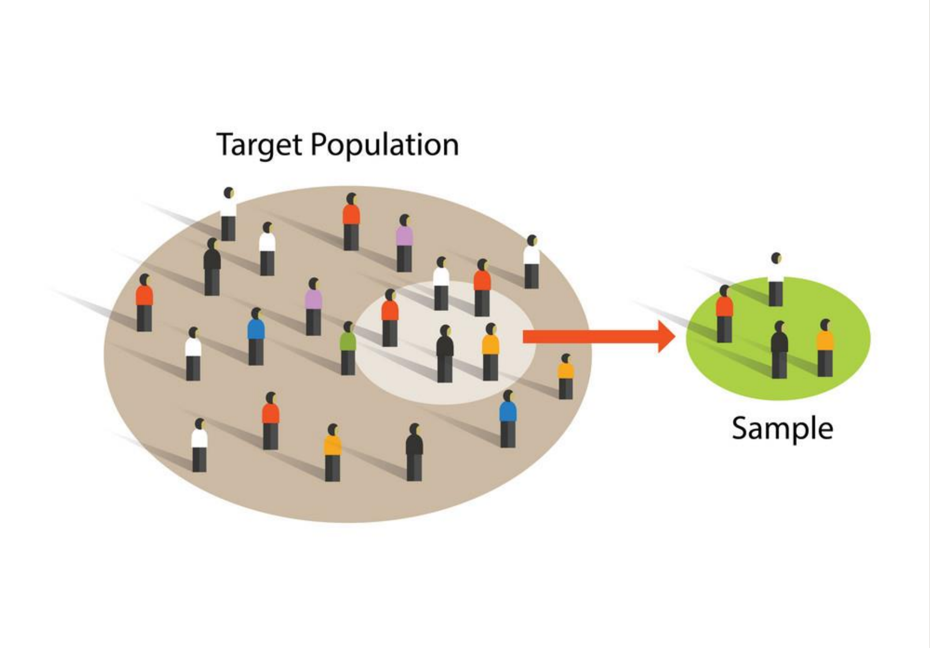
POPULATION AND SAMPLE

Population

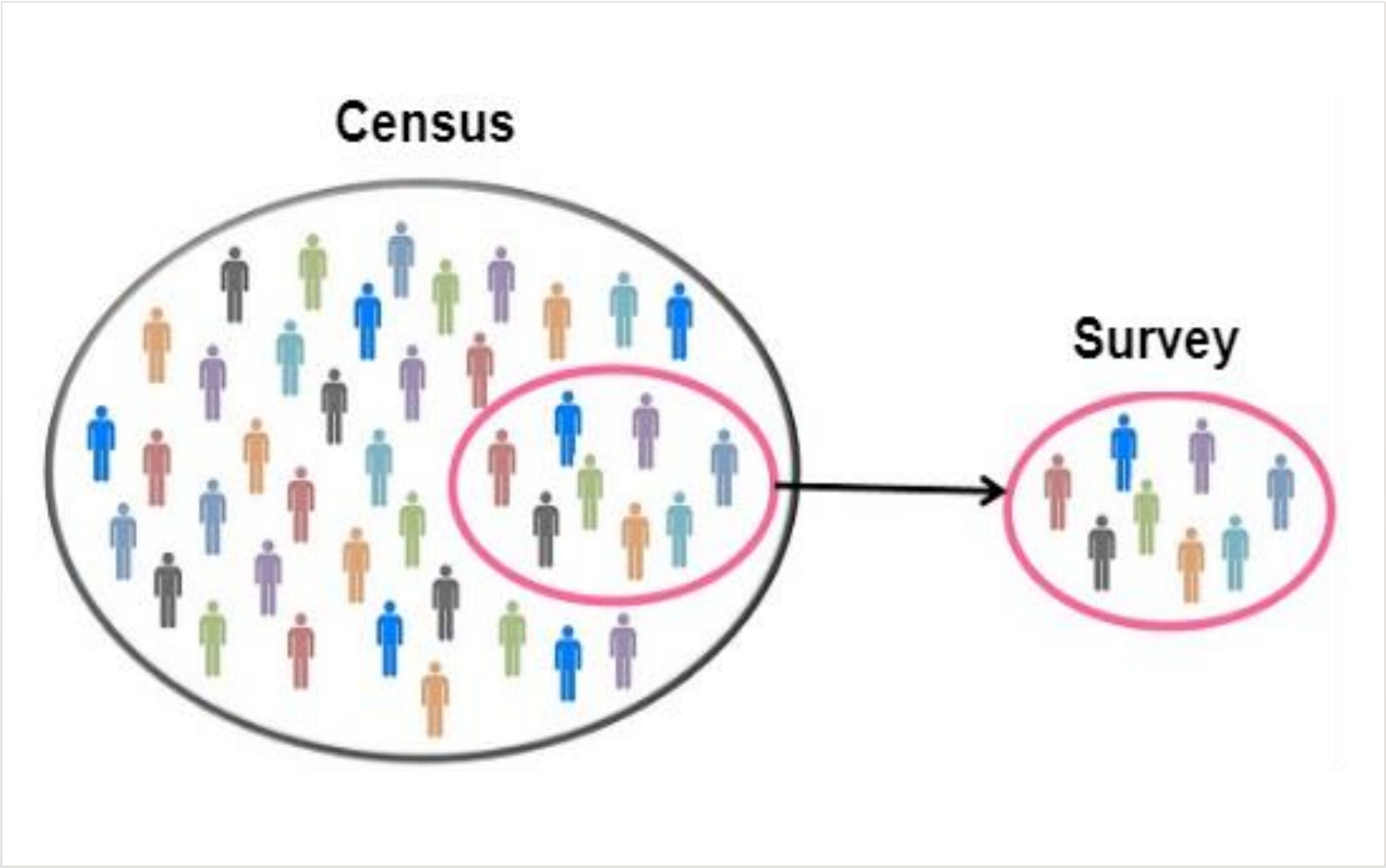
The entire set of possible cases

Sample

A subset of the population from which data are collected



EXAMPLES FOR POPULATION AND SAMPLE



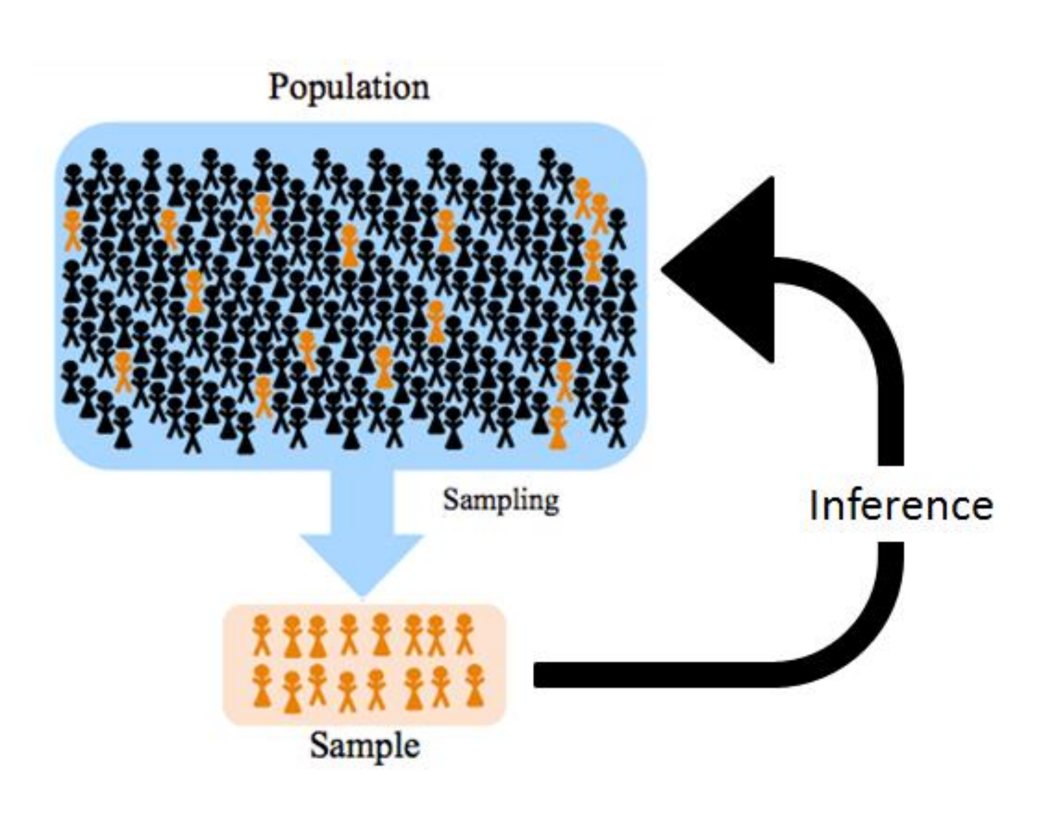
STATISTIC AND PARAMETER

Statistic

A measure concerning a sample (e.g., sample mean)

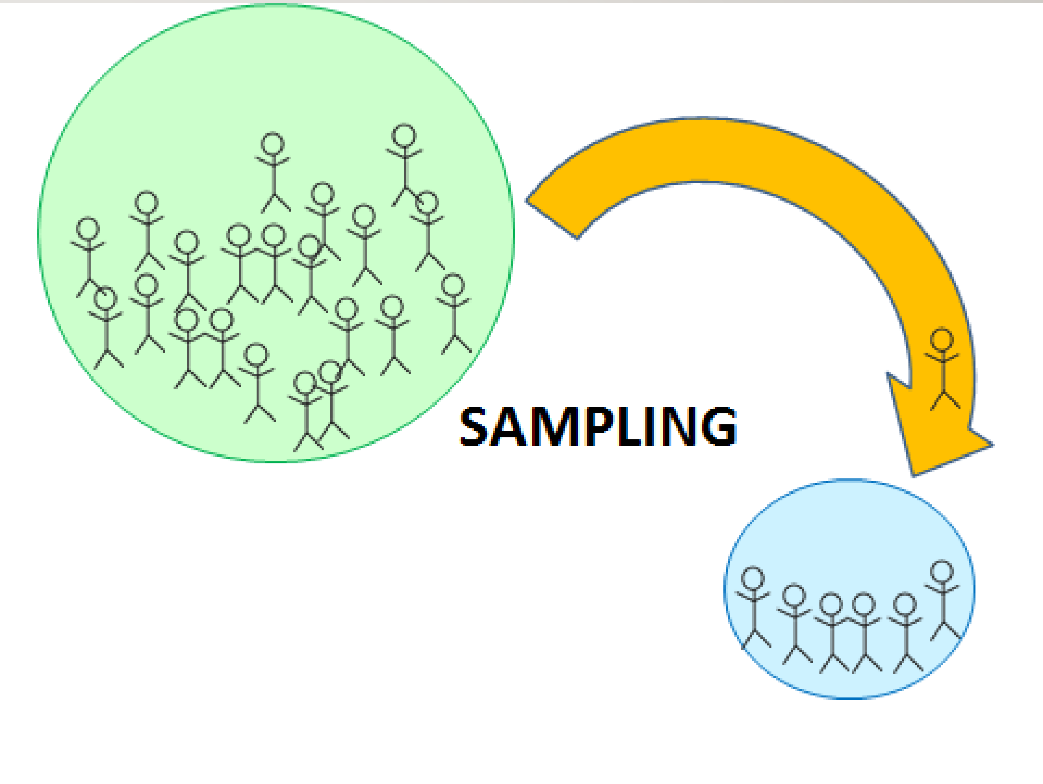
Parameter

A measure concerning a population (e.g., population mean)



SAMPLING

PROCESS OF SELECTING A SMALL NUMBER OF ELEMENTS FROM A MUCH LARGER SET



SAMPLING BIAS

- Three types of Sampling bias:

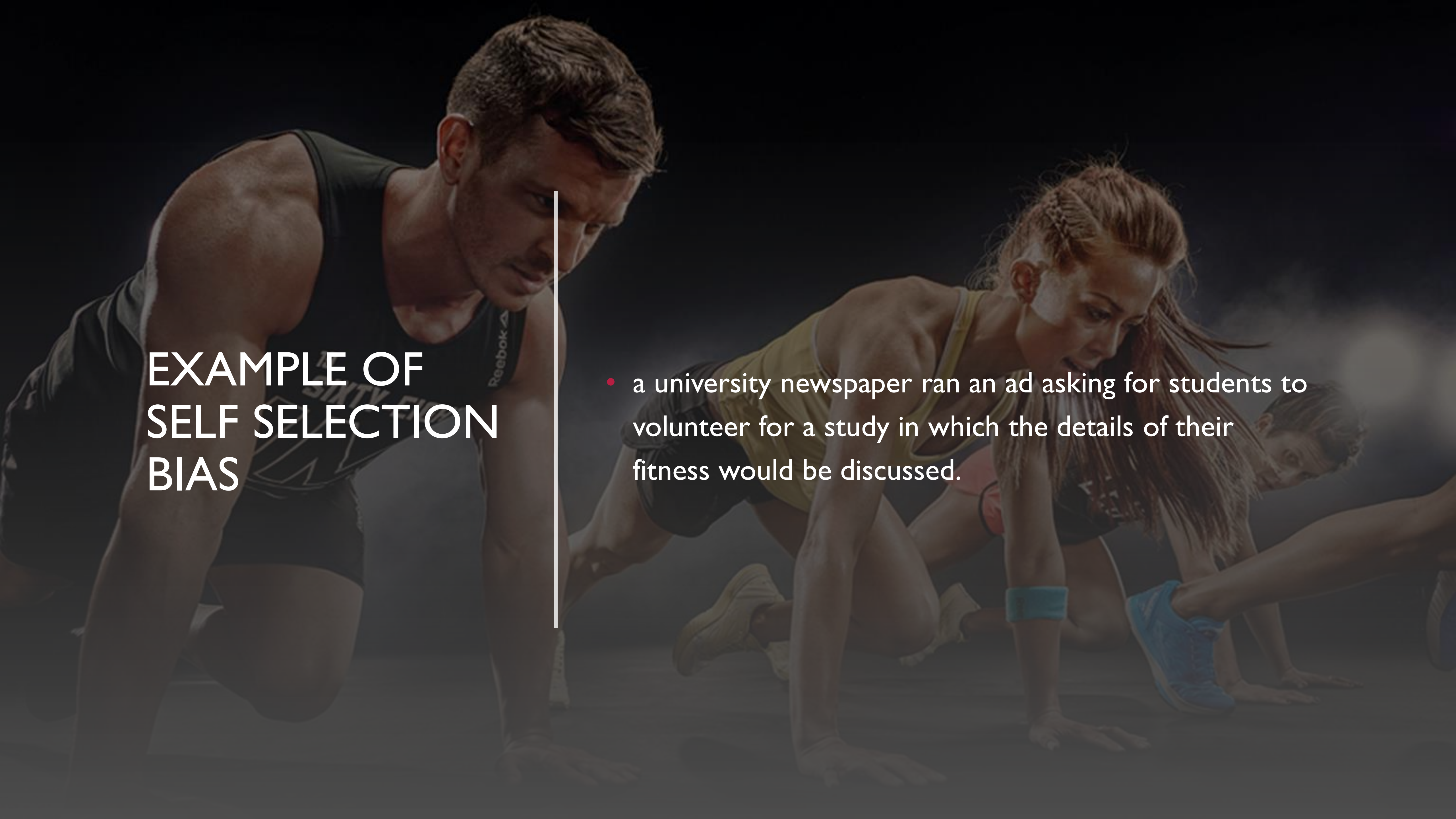
1. *SELF- SELECTION BIAS*
2. *UNDER-COVERAGE BIAS*
3. *SURVIVORSHIP BIAS*
4. *RESPONSE BIAS*
5. *NON – RESPONSE BIAS*

SELF – SELECTION BIAS

- This is the most common type of Bias
- This type of bias favors those members of a population who are more inclined and able to answer polls.

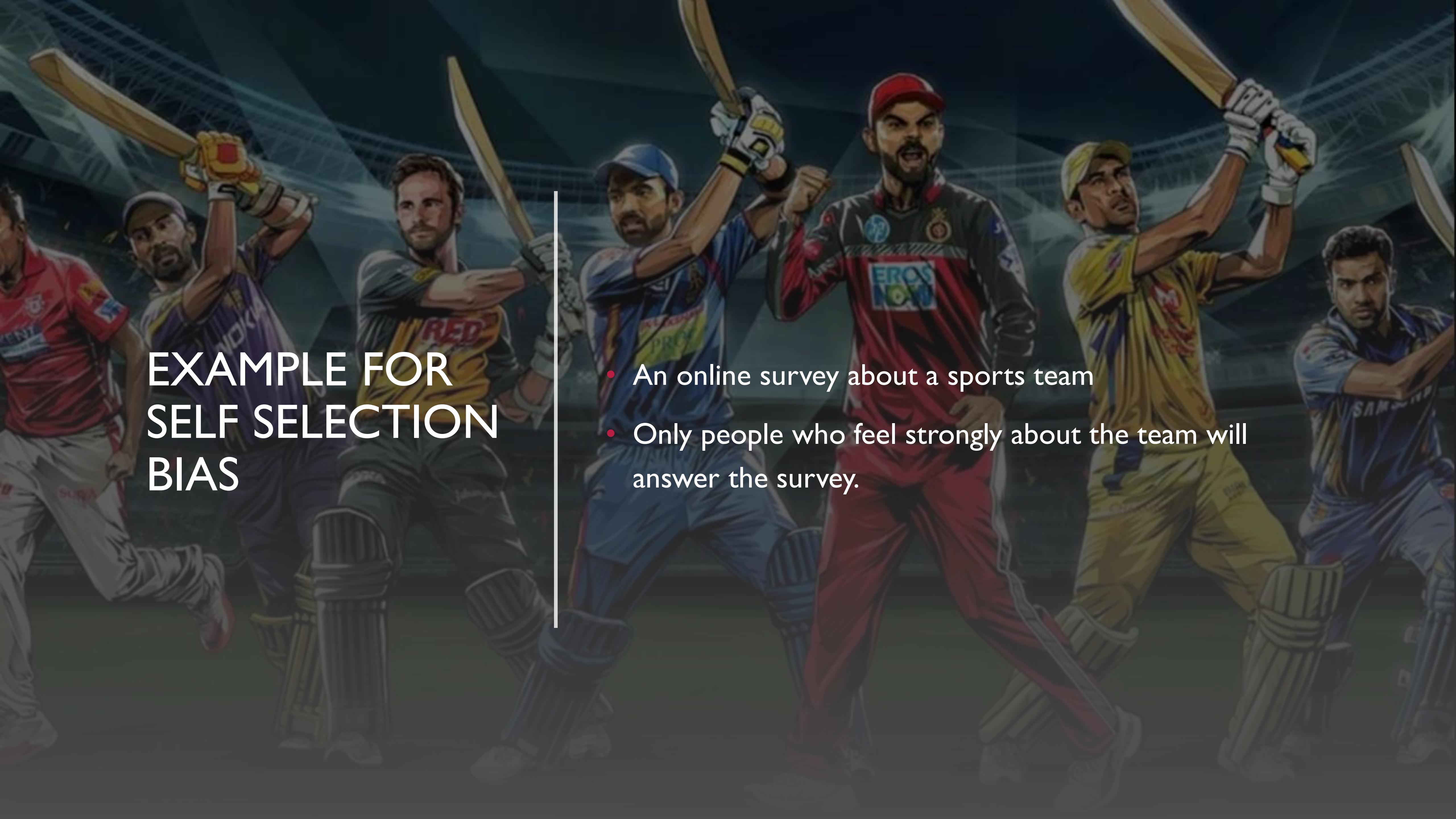
EXAMPLE OF SELF – SELECTION BIAS

- An online survey about computer use is likely to attract people more interested in technology than is typical.



EXAMPLE OF SELF SELECTION BIAS

- a university newspaper ran an ad asking for students to volunteer for a study in which the details of their fitness would be discussed.

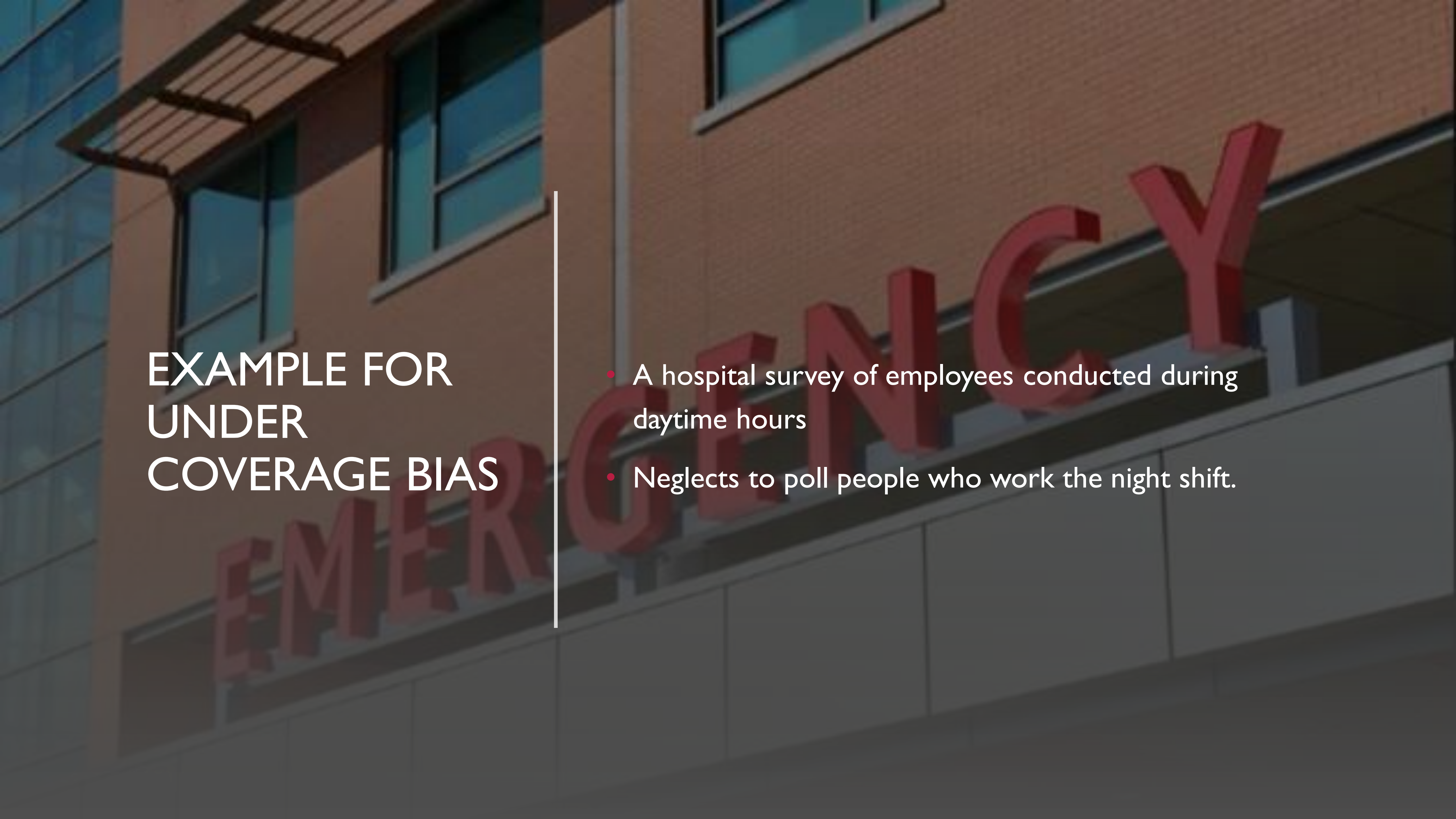


EXAMPLE FOR SELF SELECTION BIAS

- An online survey about a sports team
- Only people who feel strongly about the team will answer the survey.

UNDER COVERAGE BIAS

- A common type of sampling bias is to sample too few observations from a segment of the population.



EXAMPLE FOR UNDER COVERAGE BIAS

- A hospital survey of employees conducted during daytime hours
- Neglects to poll people who work the night shift.

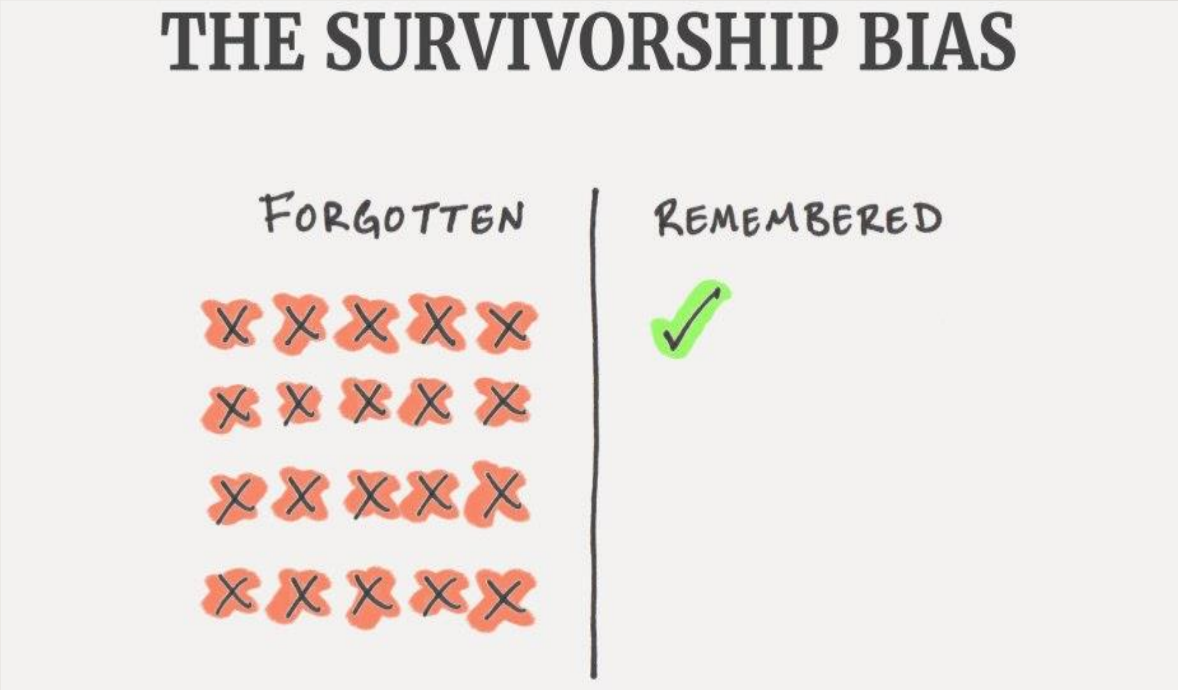
THE LITERARY DIGEST CASE STUDY

- A commonly-cited example of under coverage is the poll taken by the Literary Digest in 1936 that indicated that Landon would win an election against Roosevelt by a large margin when, in fact, it was Roosevelt who won by a large margin.

The Literary Digest
NOVEMBER 14, 1936
Thirty CENTS
IS OUR FACE RED!

SURVIVOR BIAS

- Survivorship bias is a tendency to only focus on the survivors rather than everybody.
- This can due to their lack of visibility which can, in turn, lead to false conclusions.



EXAMPLE FOR SURVIVOR BIAS

- In World War I, statistician Abraham Wald worked for America’s Statistical Research Group (SRG)

EXAMPLE FOR SURVIVOR BIAS

- One problem the SRG worked on was to examine the distribution of damage to aircraft by enemyfire and to advise the best placement ofadditional armor.

EXAMPLE FOR SURVIVOR BIAS

- Common logic was to provide greater protection to parts that received more damage.

EXAMPLE FOR SURVIVOR BIAS

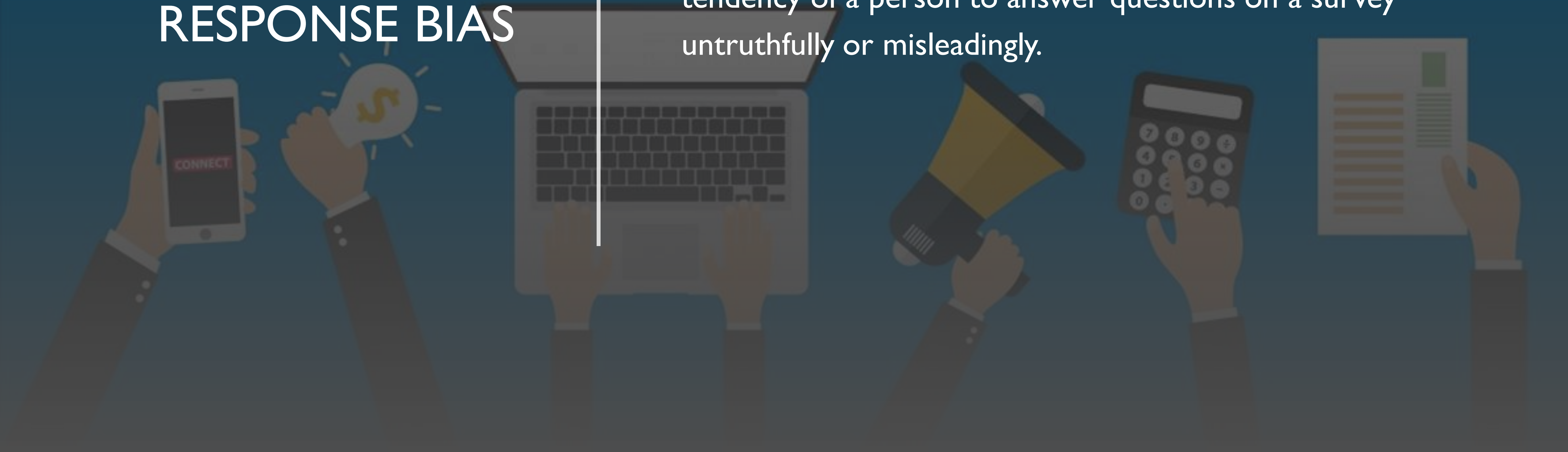
- Wald saw it differently – he felt that damage must be more uniformly distributed and that aircraft that could return had been hit in less vulnerable parts.

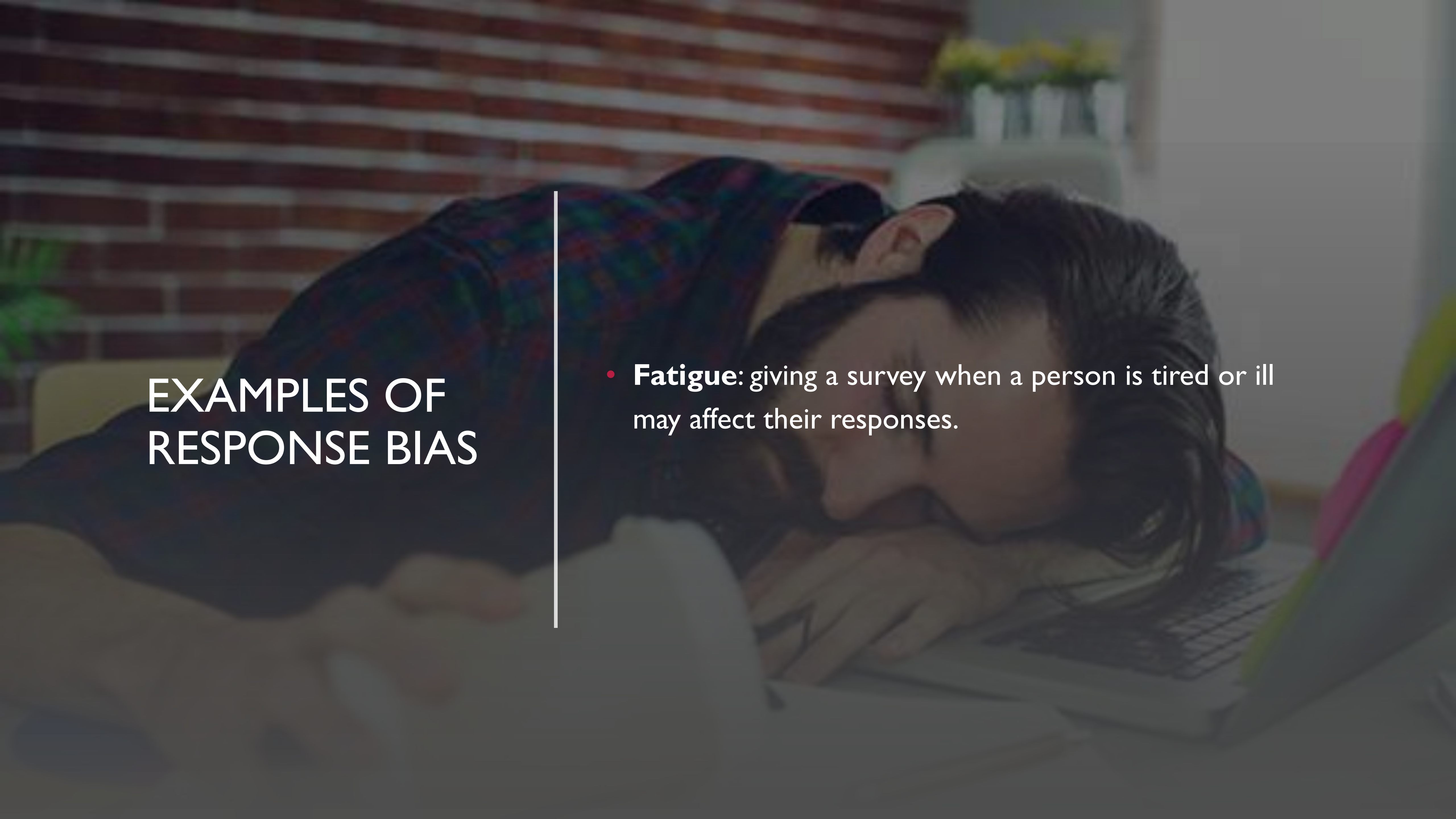
EXAMPLE FOR SURVIVOR BIAS

- Wald proposed that the Navy reinforce the areas where returning aircraft were undamaged, since those were areas that,if hit, would cause the plane to be lost!

RESPONSE BIAS

- **Response bias** (also called survey **bias**) is the tendency of a person to answer questions on a survey untruthfully or misleadingly.



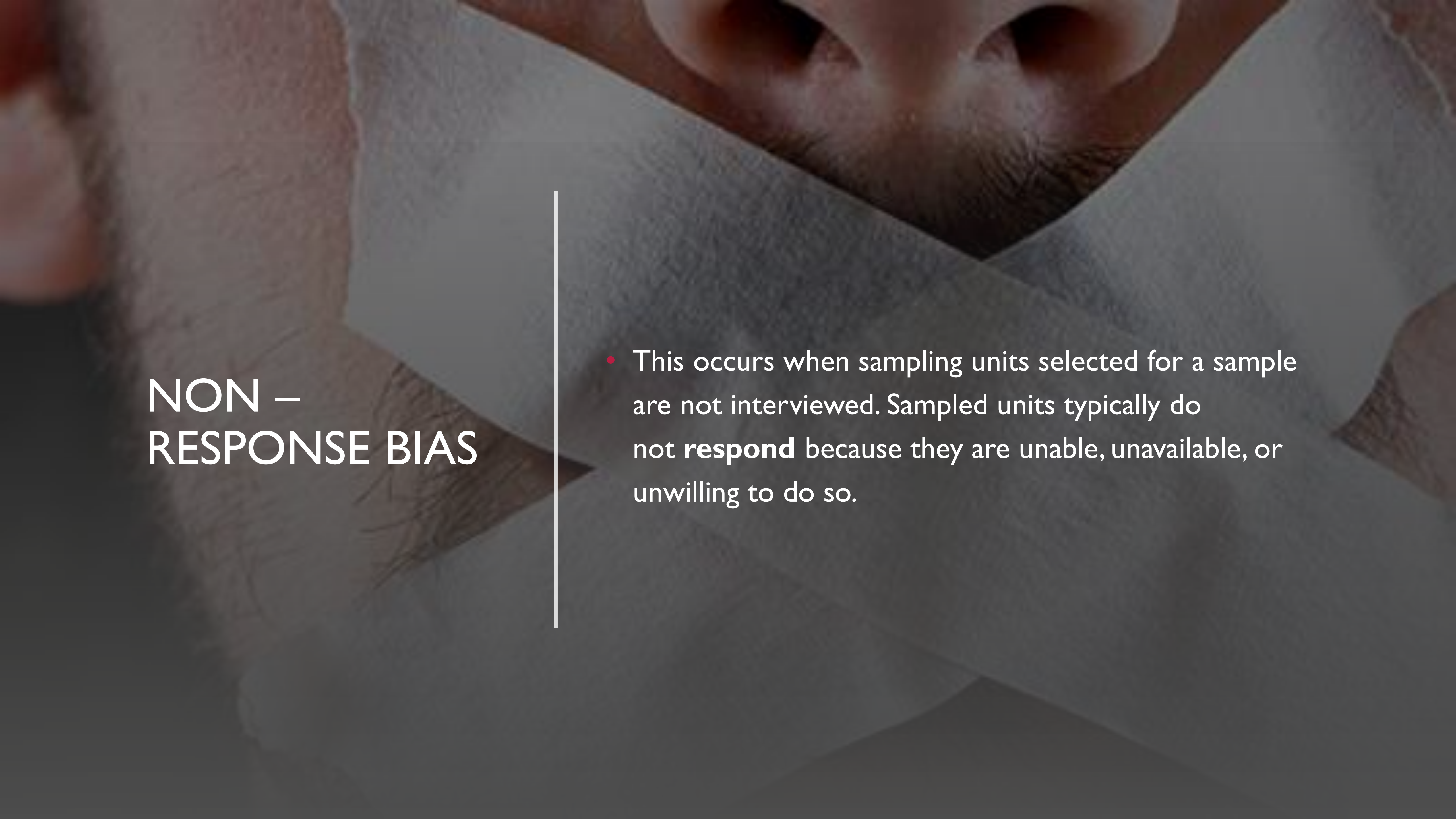


EXAMPLES OF RESPONSE BIAS

- **Fatigue**: giving a survey when a person is tired or ill may affect their responses.

EXAMPLES OF RESPONSE BIAS

- **Faulty recall:** asking a person about an event that happened in the distant past may result in erroneous responses.



NON – RESPONSE BIAS

- This occurs when sampling units selected for a sample are not interviewed. Sampled units typically do not **respond** because they are unable, unavailable, or unwilling to do so.

EXAMPLES FOR NON RESPONSE BIAS

- You have a smartphone survey for older adults.

BENEFITS OF SAMPLING

- A reasonably sized (>30) random sample will almost always reflect the population.
- But how do we select these members, and avoid bias?

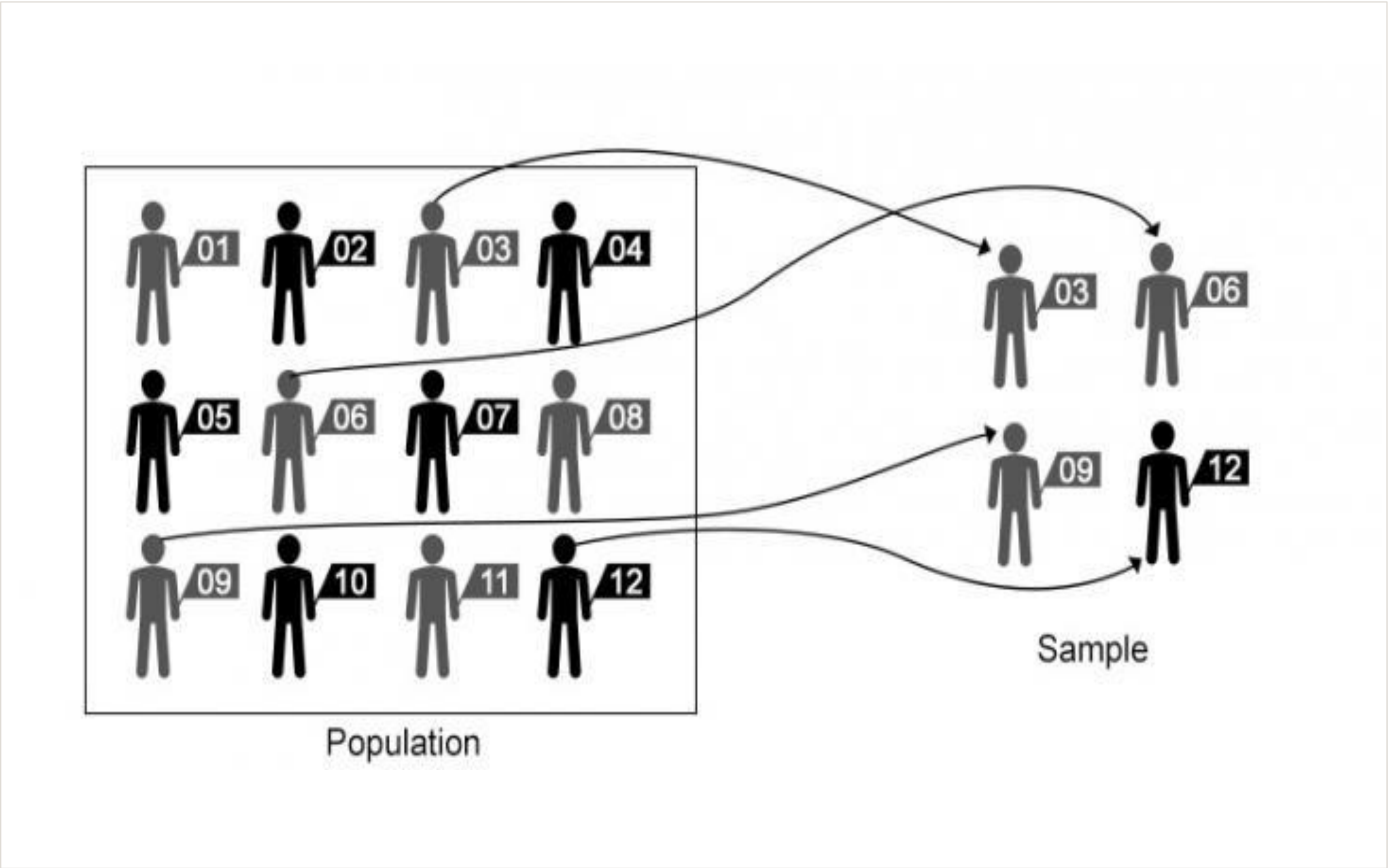
TYPES OF SAMPLING

- Random
- Stratified Random
- Cluster

RANDOM SAMPLING

- As its name suggests, random sampling means every member of a population has an equal chance of being selected.

EXAMPLES FOR RANDOM SAMPLING



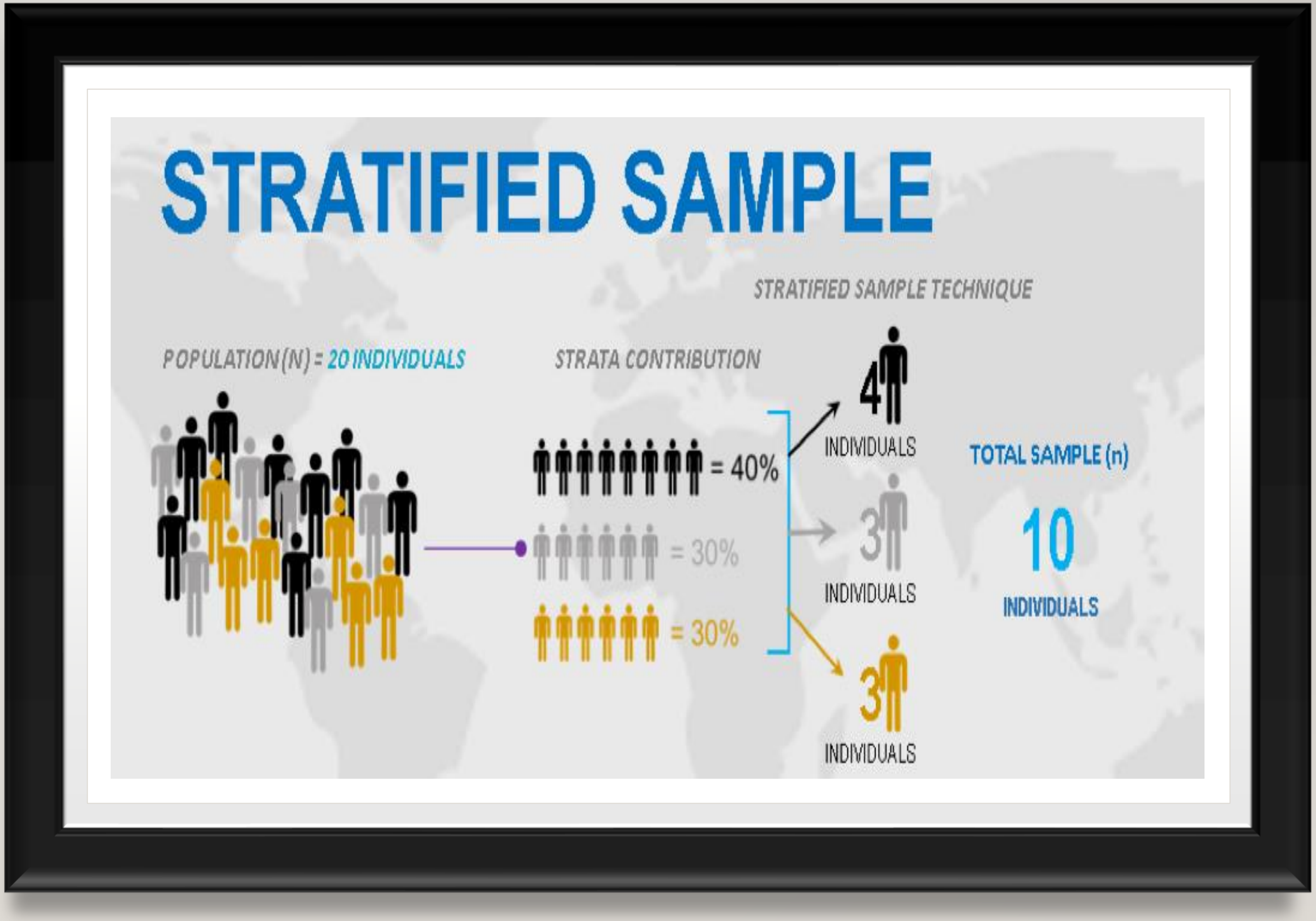
STRATIFIED RANDOM SAMPLING

- Stratified random sampling ensures that data in population are adequately represented in terms of groups.

STEPS FOR STRATIFIED RANDOM SAMPLING

- First, divide the population into segments based on some characteristic.
- Next, take random samples from each group

EXAMPLES FOR STRATIFIED RANDOM SAMPLING



EXAMPLE OF STRATIFIED RANDOM SAMPLING

A Company wanted to conduct a survey of customer satisfaction. They can only survey 10% of customers. So to ensure every age group is fairly represented

EXAMPLE OF STRATIFIED RANDOM SAMPLING

- The customer breakdown by age group is as follows:

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000

stratum

strata

EXAMPLE OF STRATIFIED RANDOM SAMPLING

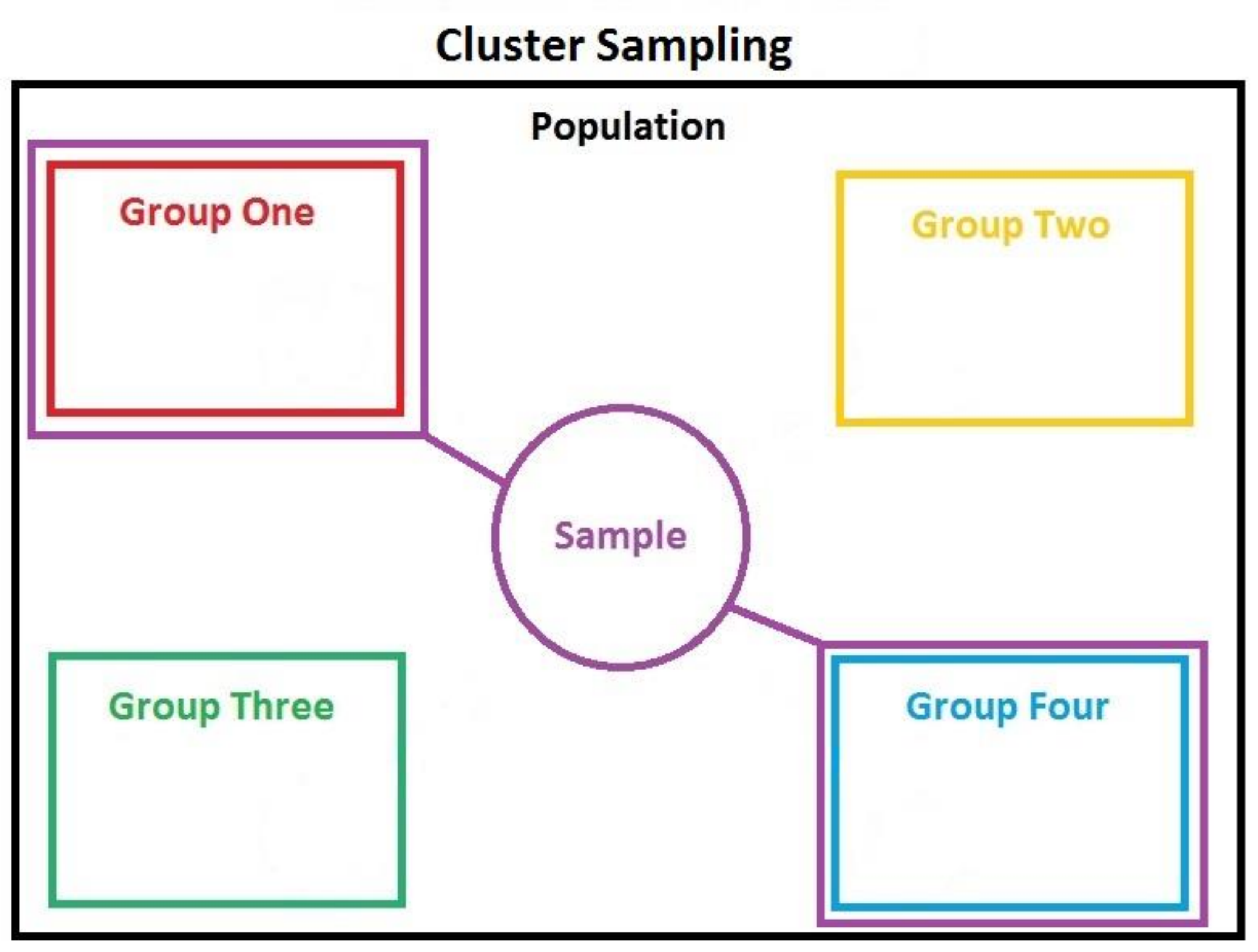
- To get the 10% we take 10% from each group

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000
140	**445**	**320**	**95**	**1,000**

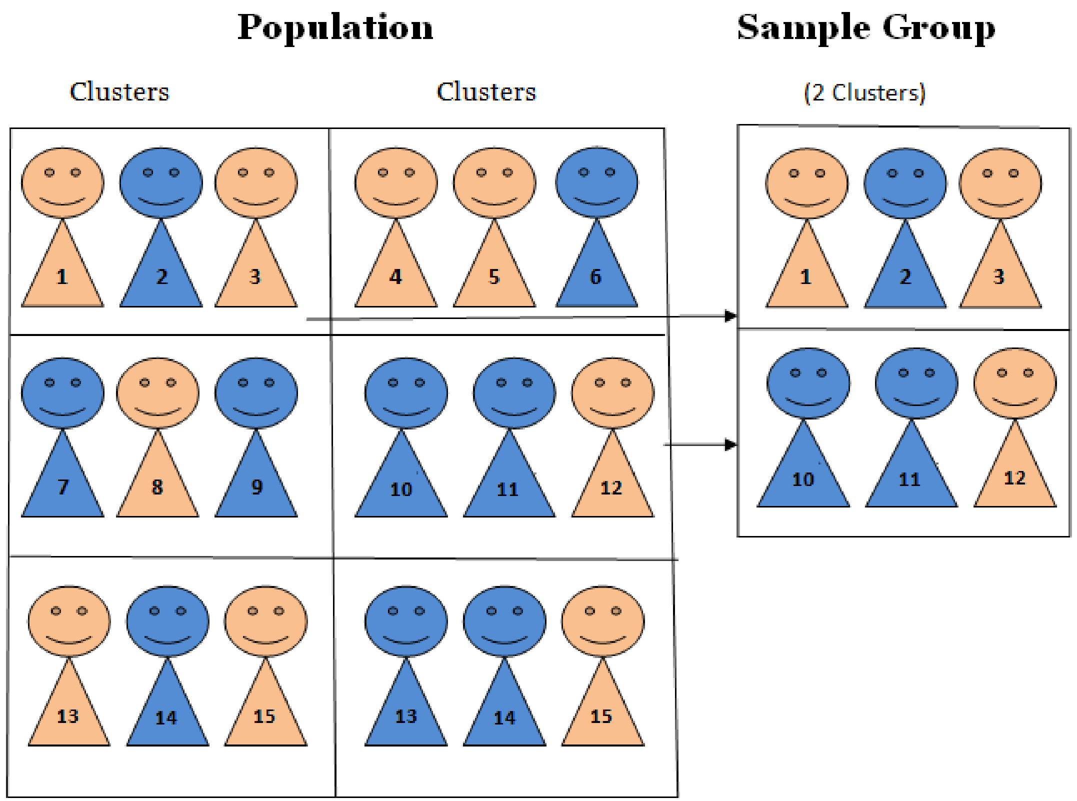
CLUSTERING

- **Cluster sampling** refers to a type of **sampling** method .
- With **cluster sampling**, the researcher divides the population into separate groups, called **clusters**. Then, a simple random **sample** of **clusters** is selected from the population.

EXAMPLE FOR CLUSTER SAMPLING

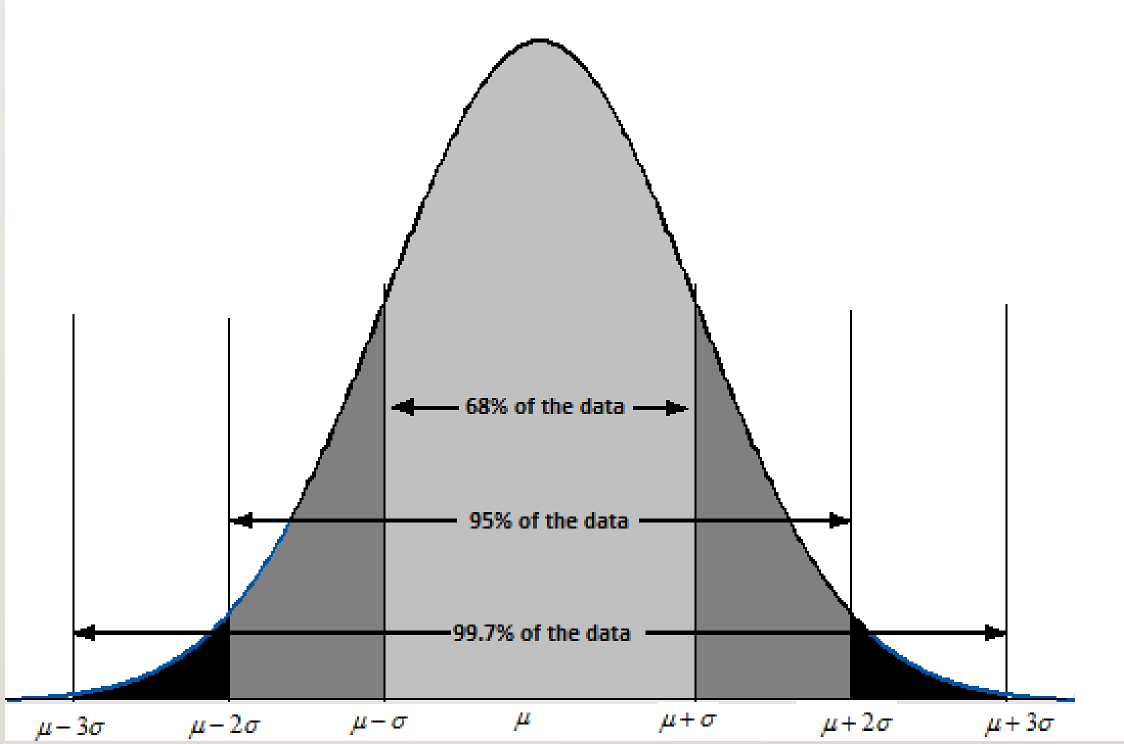


EXAMPLE FOR CLUSTER SAMPLING



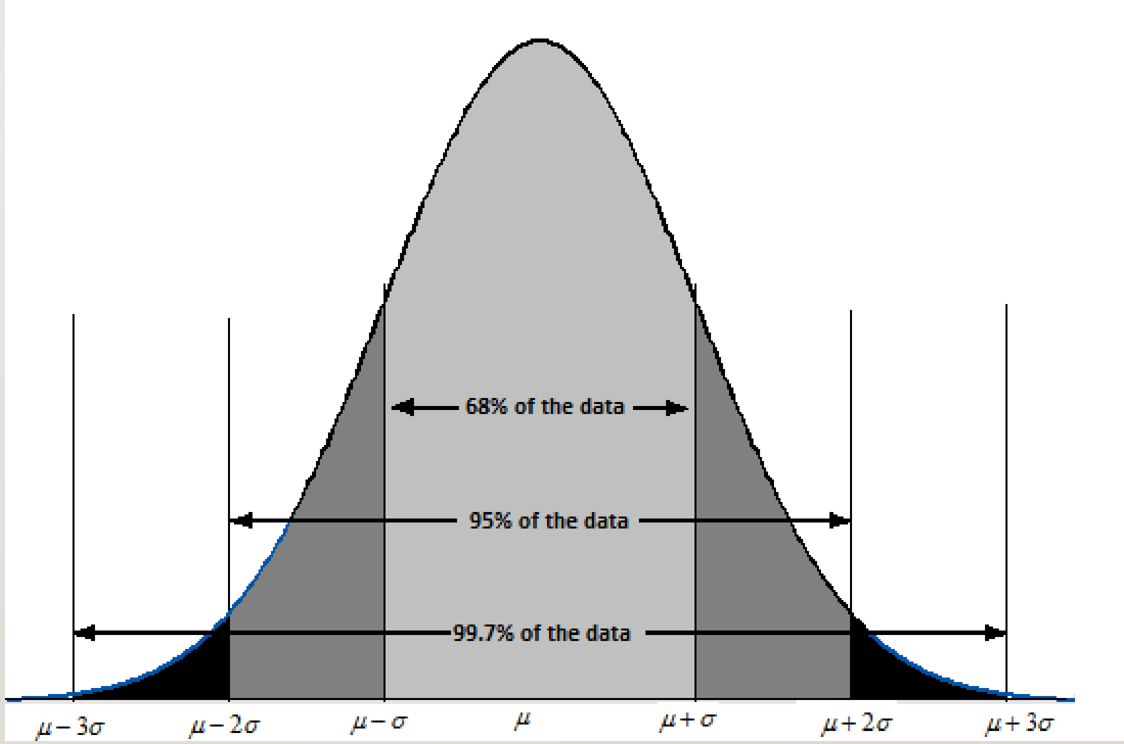
NORMAL DISTRIBUTION EXTENDED

- Characteristics of a normal curve:
- The values of mean, median and mode are same



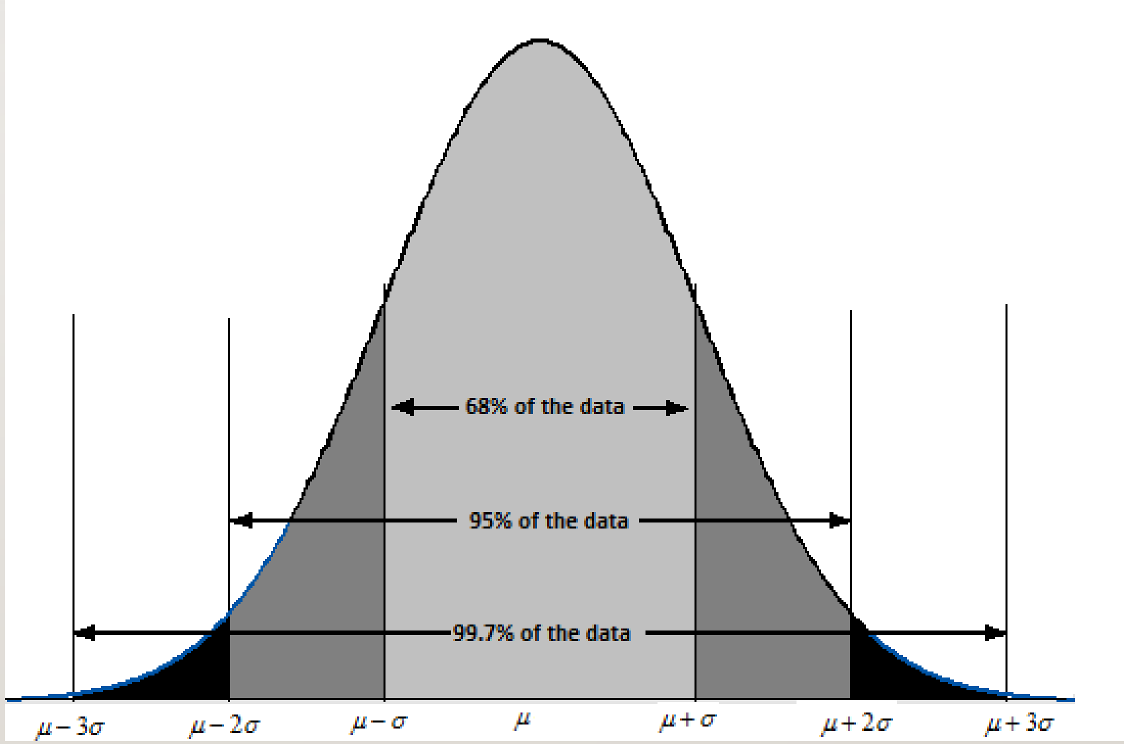
NORMAL DISTRIBUTION EXTENDED

- Characteristics of a normal curve:
- It shows a symmetric distribution as 50% of the data set lies on the left side of the mean and 50% of the data set lies on the right side of the mean.



NORMAL DISTRIBUTION EXTENDED

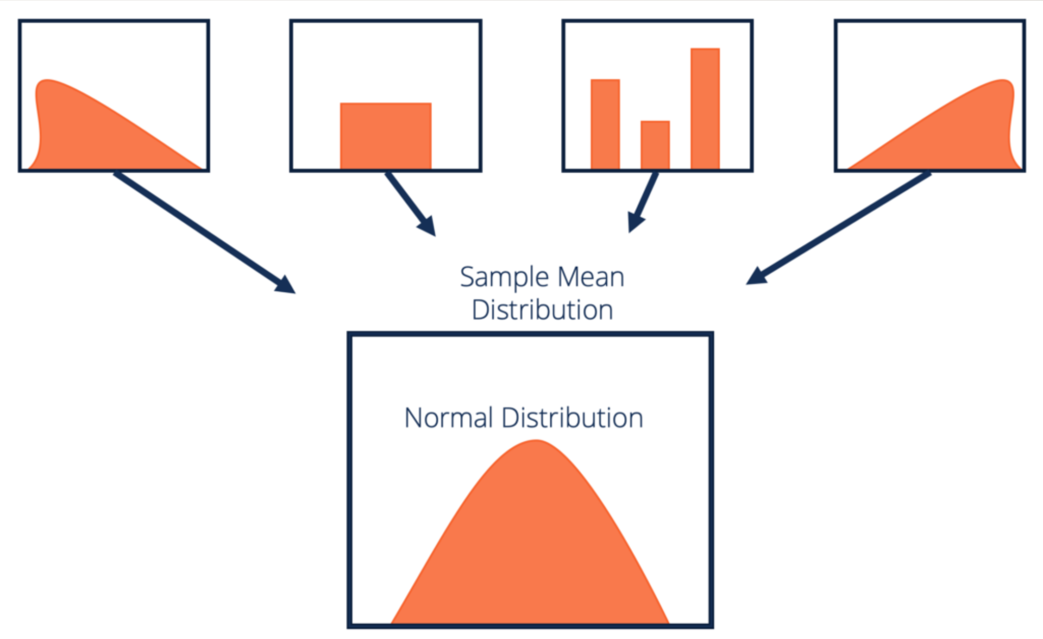
- Characteristics of a normal curve:
- Empirical rule: 68% of the data fall within μ ±σ, 95% of the data fall within μ ± 2 σ and 99.7% of the data fall within μ ± 3 σ



CENTRAL LIMIT THEOREM

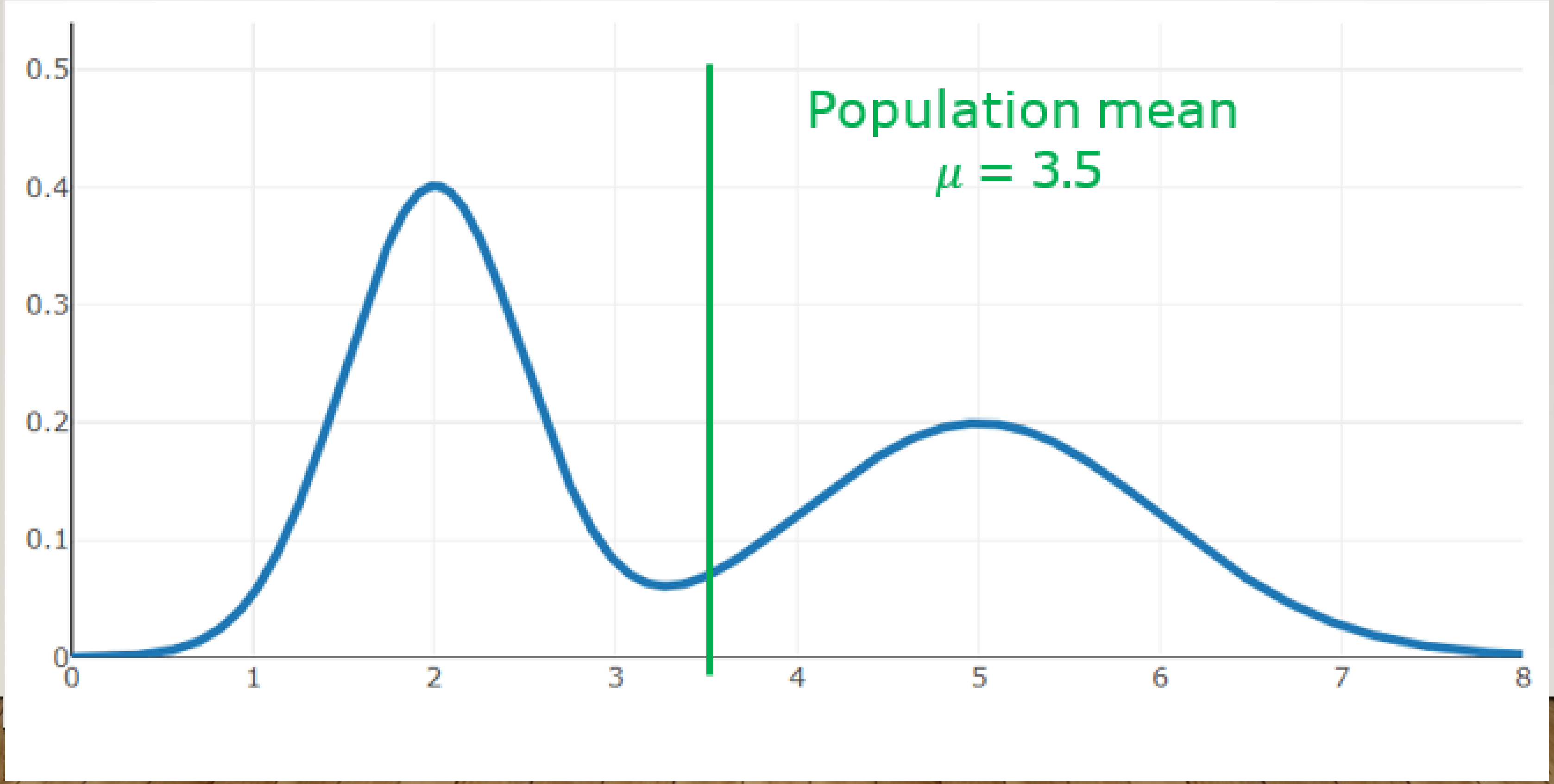
- What makes sampling such a good statistical tool is the Central Limit Theorem

- ***The CLT states that the mean values from a group of samples will be normally distributed about the population mean, even if the population itself is not normally distributed.***



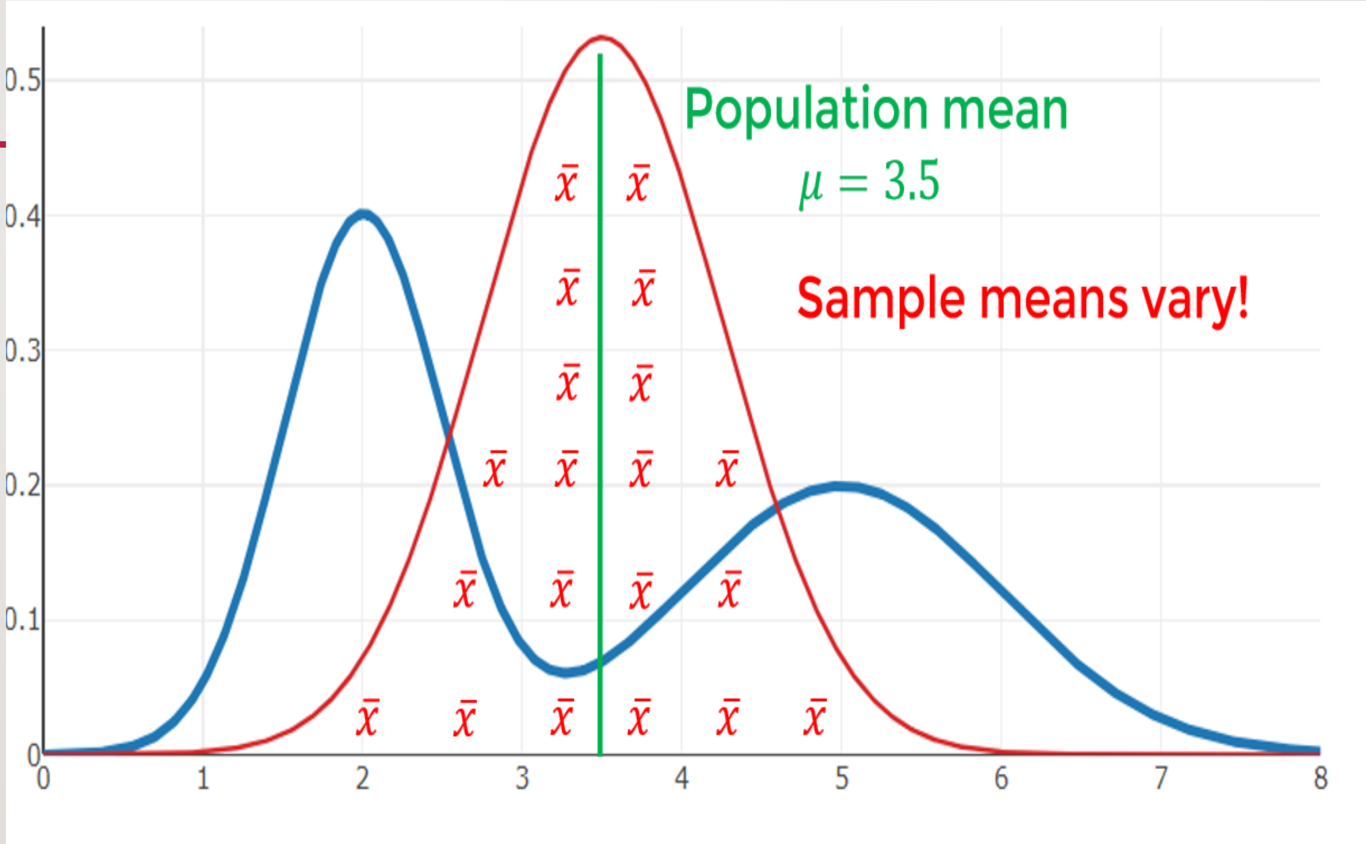
EXAMPLE FOR CLT

Consider this example distribution of the population with mean 3.5



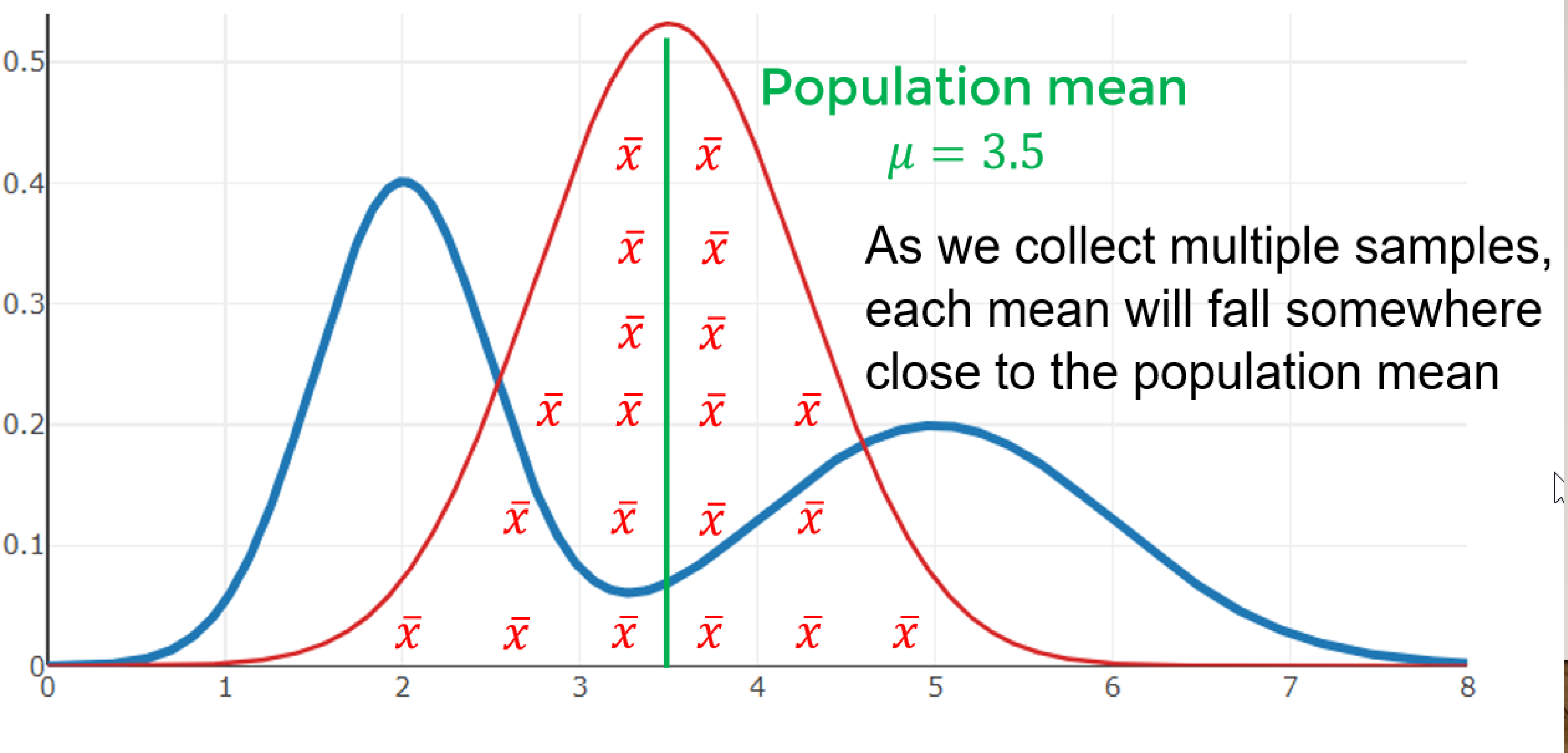
EXAMPLE FOR CLT

- Here when we take different samples from the population and find their means we would get the following distribution



EXAMPLE FOR CLT

That is 95% of all the sample means should fall within 2σ of the population mean

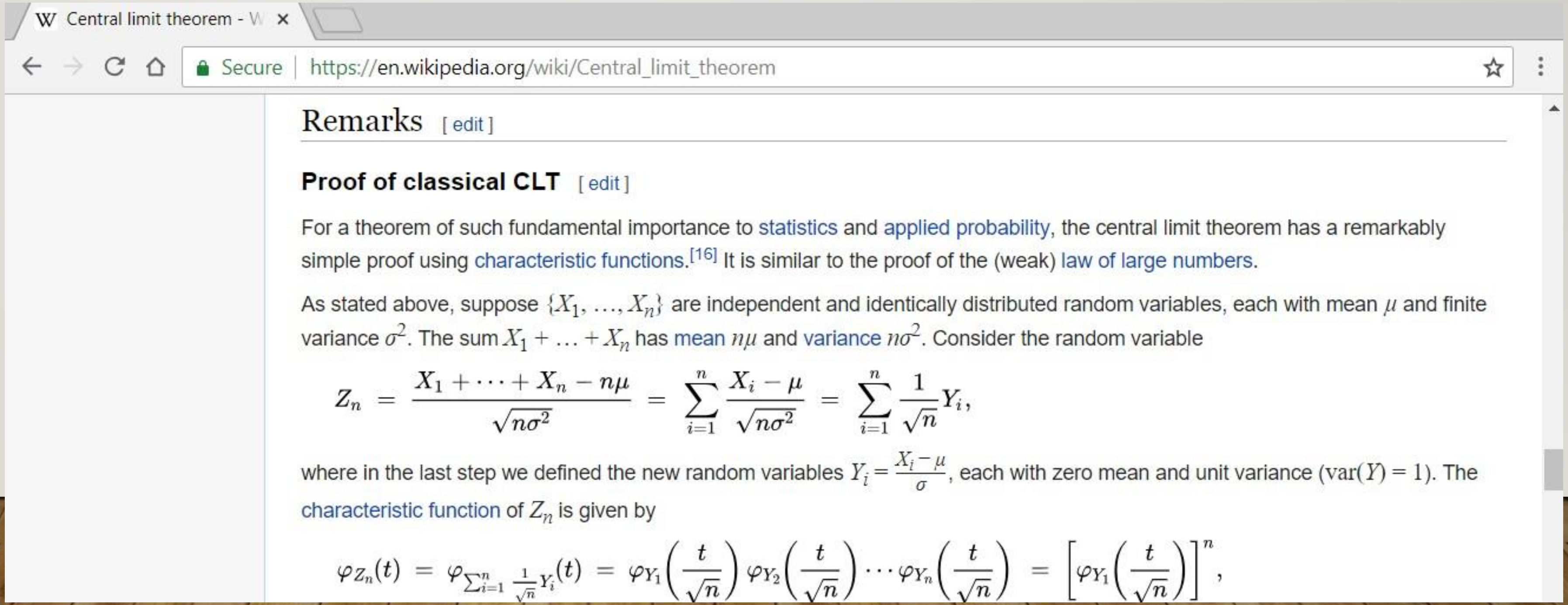


As we collect multiple samples, each mean will fall somewhere close to the population mean

Proof of CLT Available on Wikipedia

For those who are curious, the full proof of the Central Limit Theorem is available at

https://en.wikipedia.org/wiki/Central_limit_theorem


Central limit theorem - W ×
Secure
https://en.wikipedia.org/wiki/Central_limit_theorem

Remarks [edit]

Proof of classical CLT [edit]

For a theorem of such fundamental importance to statistics and applied probability, the central limit theorem has a remarkably simple proof using characteristic functions.[16] It is similar to the proof of the (weak) law of large numbers.

As stated above, suppose $\{X_1, \ldots, X_n\}$ are independent and identically distributed random variables, each with mean μ and finite variance σ^2. The sum $X_1 + \ldots + X_n$ has mean $n\mu$ and variance $n\sigma^2$. Consider the random variable

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i,$$

where in the last step we defined the new random variables $Y_i = \frac{X_i - \mu}{\sigma}$, each with zero mean and unit variance ($\operatorname{var}(Y) = 1$). The characteristic function of Z_n is given by

$$\varphi_{Z_n}(t) = \varphi_{\sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i}(t) = \varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) \varphi_{Y_2}\left(\frac{t}{\sqrt{n}}\right) \cdots \varphi_{Y_n}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

STANDARD ERROR MEAN

- What is Standard deviation?
  - Describes how wide individuals values are deviated from the population mean
- Standard error of the mean describes how far the sample mean may be deviated from the population mean

STANDARD ERROR OF MEAN

- If the population standard deviation σ is known, then the sample standard error of the mean can be calculated as:

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

EXAMPLE OF STANDARD ERROR

- An IQTest is designed to have a mean score of 100 with a standard deviation of 15points.
- If a sample of 10scores has a mean of 104,can we assume they come from the generalpopulation?

EXAMPLE OF STANDARD ERROR

- Sample of 10 IQ Test scores:
  - n = 10 x = 104 σ = 15

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4.743$$

68% of 10-item sample means are expected to fall between 95.257 and 104.743

Z-SCORE OR STANDARD SCORE

- **Simply put, a z-score is the number of standard deviations from the mean a data point is.**
- But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

$$\mathbf{z = (x - \mu) / \sigma}$$

EXAMPLE FOR Z-SCORE AND STANDARD SCORE

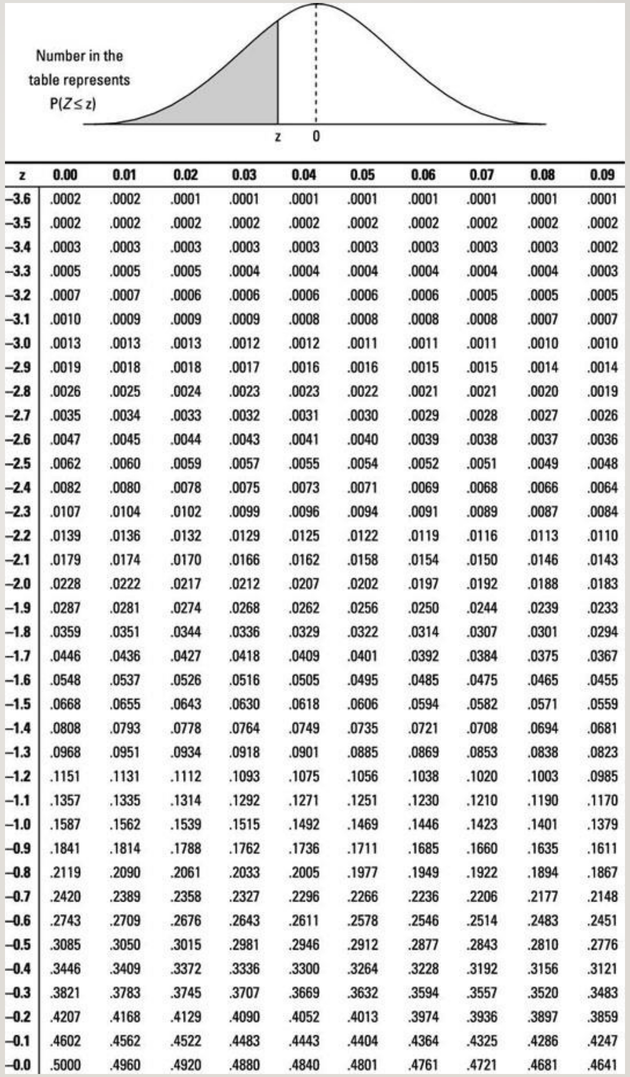
- Let's say you have a test score of 175. The test has a mean (μ) of 150 and a standard deviation (σ) of 25. Assuming a normal distribution, your z score would be and how many percent of the people scored below you:
- $z = (x – \mu) / \sigma$
  $= 175 – 150 / 25 = 1$
- 50 + 34 = 84%

EXAMPLES FOR Z-SCORE OR STANDARD SCORE

- Let’s say you have a test score of 190. The test has a mean (μ) of 150 and a standard deviation (σ) of 25. Assuming a normal distribution, your z score would be and how many percent of the people scored below you:
- $z = (x - \mu) / \sigma$
  $= 190 - 150 / 25 = 1.6.$
- We will look at the Z – Table

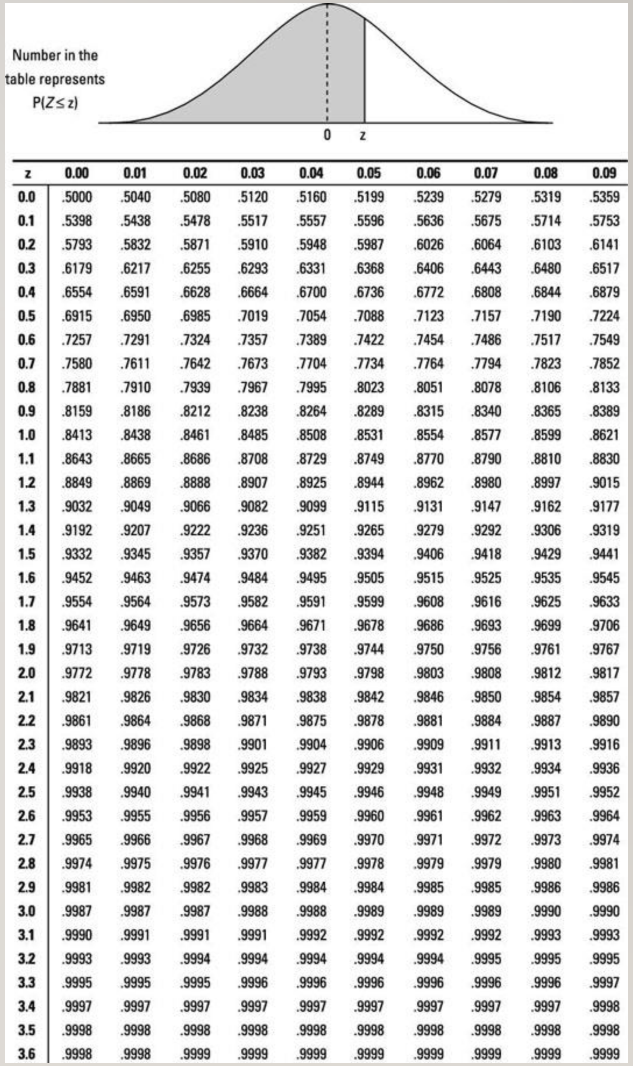
Z-TABLE

- Now from the results to understand how many people have scored below us we can look up the z-table and tell the value in percentage.



Number in the table represents $P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
–3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

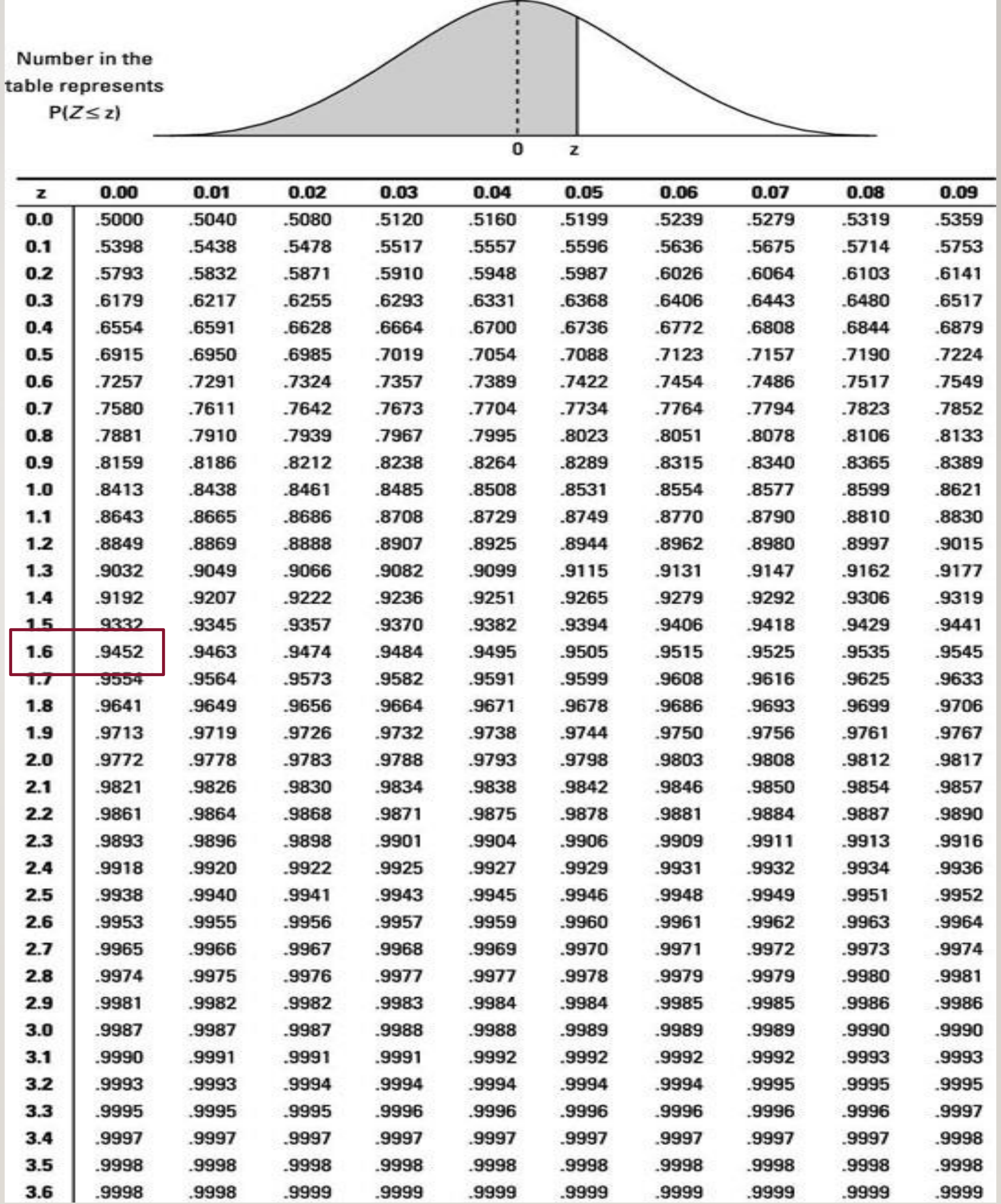


Number in the table represents $P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

EXAMPLES FOR Z-SCORE OR STANDARD SCORE

- Let's say you have a test score of 190. The test has a mean (μ) of 150 and a standard deviation (σ) of 25. Assuming a normal distribution, your z score would be and how many percent of the people scored below you:
- We will look at the Z – Table



Number in the table represents P(Z ≤ z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

EXAMPLE FOR Z – TABLE

- You take up the CAT examination and score 1100. The mean score for the SAT is 1026 and standard deviation is 209. How well did you perform than the average test taker?

MARGIN OF ERROR

- A small amount that is allowed for in case of miscalculation or change of circumstances.

ME = z * SE

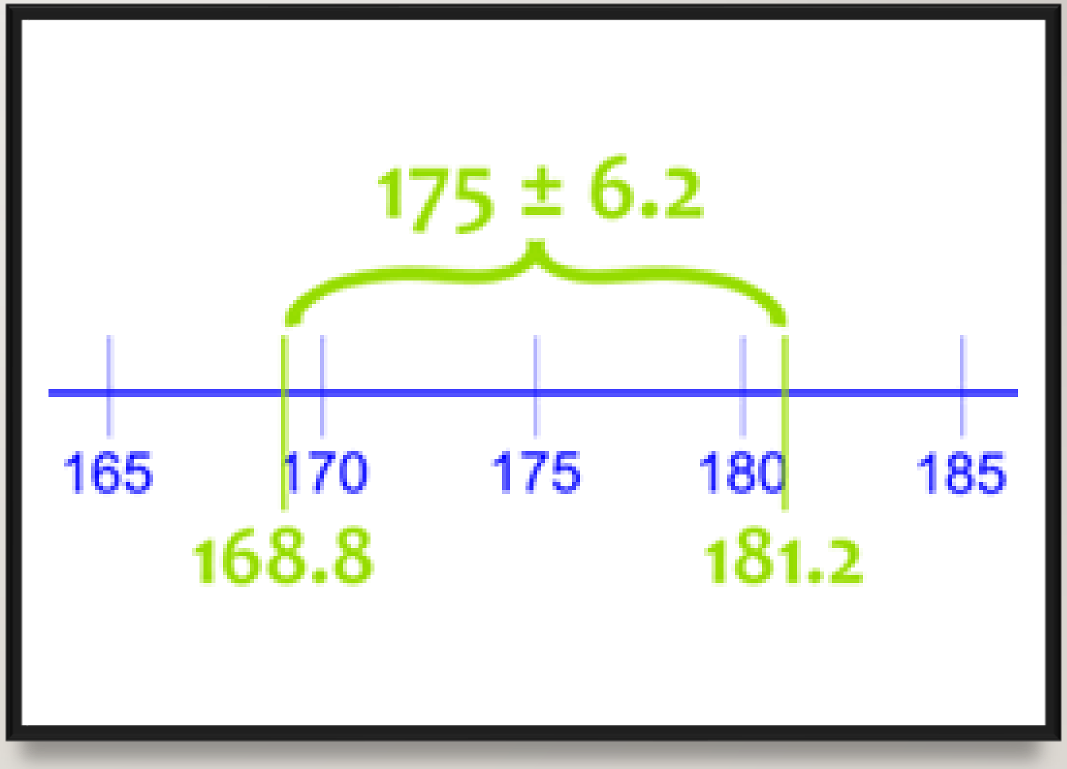
Where z is the z score and SE is the Standard Error

CONFIDENCE INTERVAL

- A Confidence Interval is a **range of values** we are fairly sure our **true value** lies in.

EXAMPLE FOR CONFIDENCE INTERVAL

- We measure the heights of **40** randomly chosen men, and get a mean height of **175cm**,
- We also know the standard deviation of men's heights is **20cm**.
- The **95% Confidence Interval** (we show how to calculate it later) is:

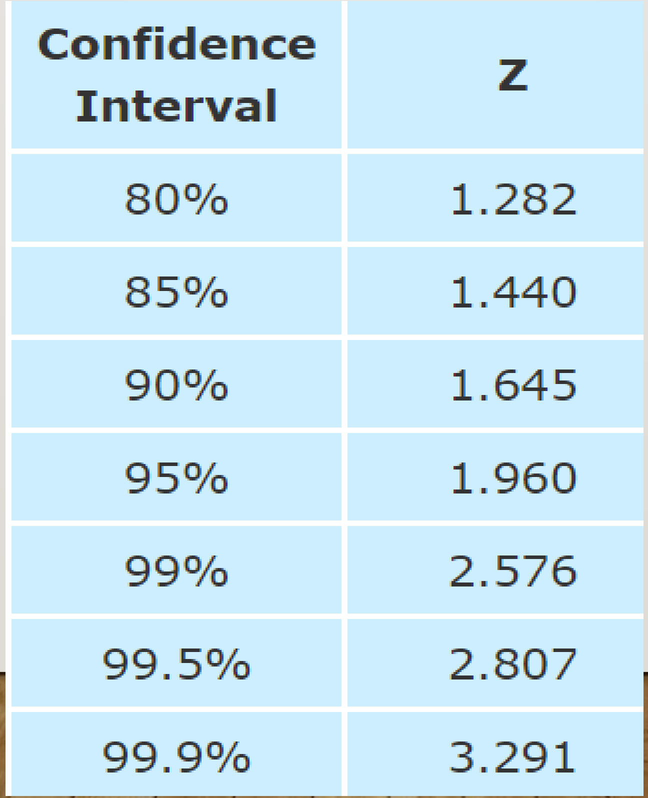


STEPS FOR FINDING CONFIDENCE INTERVAL

- **Step 1**: find the number of observations **n**, calculate their mean **X**, and standard deviation**s**
- Using our example:
- Number of observations: **n = 40**
- Mean: **X = 175**
- Standard Deviation: **s = 20**

STEPS FOR FINDING CONFIDENCE INTERVAL

- **Step 2**: decide what Confidence Interval we want: 95% or 99% are common choices. Then find the "Z" value for that Confidence Interval here:



Confidence Interval	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291

For 95% the Z value is **1.960**

STEPS FOR FINDING CONFIDENCE INTERVAL

- **Step 3**: use that Z in this formula for the Confidence Interval

$$\overline{X} \pm Z\frac{s}{\sqrt{n}}$$

And we have:

$$175 \pm 1.960 \times \mathit{20}\boldsymbol{\sqrt{40}}$$

Which is:

$$\mathbf{175cm \pm 6.20cm}$$

In other words: from 168.8cm to 181.2cm