Q22. What Are the Types of Biases That Can Occur During Sampling?

- Selection bias
- Under coverage bias
- Survivorship bias

Q23. What is Survivorship Bias?

It is the logical error of focusing aspects that support surviving some process and casually overlooking those that did not work because of their lack of prominence. This can lead to wrong conclusions in numerous different means.

Q24. What is selection Bias?

Selection bias occurs when the sample obtained is not representative of the population intended to be analysed.

Q: Explain selection bias (with regard to a dataset, not variable selection). Why is it important? How can data management procedures such as missing data handling make it worse?

Selection bias is the phenomenon of selecting individuals, groups, or data for analysis in such a way that proper randomization is not achieved, ultimately resulting in a sample that is not representative of the population.

Understanding and identifying selection bias is important because it can significantly skew results and provide false insights about a particular population group.

Types of selection bias include:

- **Sampling bias**: a biased sample caused by non-random sampling
- **Time interval**: selecting a specific time frame that supports the desired conclusion. e.g. conducting a sales analysis near Christmas.
- **Exposure**: includes clinical susceptibility bias, protopathic bias, indication bias. *Read more here.*
- **Data**: includes cherry-picking, suppressing evidence, and the fallacy of incomplete evidence.
- **Attrition**: attrition bias is similar to survivorship bias, where only those that 'survived' a long process are included in an analysis, or failure bias, where those that 'failed' are only included
- **Observer selection**: related to the Anthropic principle, which is a

Q22. What Are the Types of Biases That Can Occur During Sampling?

- Selection bias
- Under coverage bias
- Survivorship bias

Q23. What is Survivorship Bias?

It is the logical error of focusing aspects that support surviving some process and casually overlooking those that did not work because of their lack of prominence. This can lead to wrong conclusions in numerous different means.

Q24. What is selection Bias?

Selection bias occurs when the sample obtained is not representative of the population intended to be analysed.

75. Do we have different Selection Biases, if yes, what are they?

Answer: Sampling Bias: This bias arises when you select only particular people or when non-random selection of samples happened. In general terms, it is nothing but a selection of the majority of the people belong to one group.

Time Interval: sometimes a trial may be terminated earlier than actual time (probably due to some ethical reasons) but the extreme value finally taken into consideration is the most significant value even though all other variables have similar Mean.

Data: We can name it as a Data bias when a separate set of data is taken to support a conclusion or eliminates terrible data based on the arbitrary grounds, instead of generally relying on generally stated criteria.

Attrition bias: Attrition bias is defined as an error that occurs due to Unequal loss of participants from a randomized controlled trial (RCT).

84. What is Selection Bias?

Answer: Selection bias is a kind of error that occurs when the researcher decides who is going to be studied. It is usually associated with research where the selection of participants isn't random. It is sometimes referred to as the selection effect. It is the distortion of statistical analysis, resulting from the method of collecting samples. If the selection bias is not taken into account, then some conclusions of the study may not be accurate.

The types of selection bias include:

Sampling bias: It is a systematic error due to a non-random sample of a population causing some members of the population to be less likely to be included than others resulting in a biased sample.

Time interval: A trial may be terminated early at an extreme value (often for ethical reasons), but the extreme value is likely to be reached by the variable with the largest variance, even if all variables have a similar mean.

Data: When specific subsets of data are chosen to support a conclusion or rejection of bad data on arbitrary grounds, instead of according to previously stated or generally agreed criteria.

Attrition: Attrition bias is a kind of selection bias caused by attrition (loss of participants) discounting trial subjects/tests that did not run to completion.

100. What is selection bias and why does it matter?

Answer: Selection bias is a product of inadequately or improperly randomized data leading to data sets that are not representative of the whole. In an interview, you should express the importance of this in terms of its effect on your solution. If your data is not representative, your solutions likely are not either.

104. What do you understand by the Selection Bias? What are its various types?

Answer: Selection bias is typically associated with research that doesn't have a random selection of participants. It is a type of error that occurs when a researcher decides who is going to be studied. On some occasions, selection bias is also referred to as the selection effect.

In other words, selection bias is a distortion of statistical analysis that results from the sample collecting method. When selection bias is not taken into account, some conclusions made by a research study might not be accurate. Following are the various types of selection bias:

Sampling Bias: A systematic error resulting due to a non-random sample of a populace causing certain members of the same to be less likely included than others that results in a biased sample.
Time Interval – A trial might be ended at an extreme value, usually due to ethical reasons, but the extreme value is most likely to be reached by the variable with the most variance, even though all variables have a similar mean.
Data – Results when specific data subsets are selected for supporting a conclusion or rejection of bad data arbitrarily.
Attrition – Caused due to attrition, i.e. loss of participants, discounting trial subjects or tests that didn't run to completion.

Q2. What is Selection Bias?

Selection bias is a kind of error that occurs when the researcher decides who is going to be studied. It is usually associated with research where the selection of participants isn't random. It is sometimes referred to

as the selection effect. It is the distortion of statistical analysis, resulting from the method of collecting samples. If the selection bias is not taken into account, then some conclusions of the study may not be accurate.

The types of selection bias include:

1. **Sampling bias**: It is a systematic error due to a non-random sample of a population causing some members of the population to be less likely to be included than others resulting in a biased sample.
2. **Time interval**: A trial may be terminated early at an extreme value (often for ethical reasons), but the extreme value is likely to be reached by the variable with the largest variance, even if all variables have a similar mean.
3. **Data**: When specific subsets of data are chosen to support a conclusion or rejection of bad data on arbitrary grounds, instead of according to previously stated or generally agreed criteria.
4. **Attrition**: Attrition bias is a kind of selection bias caused by attrition (loss of participants) discounting trial subjects/tests that did not run to completion.

11. What is selection bias, why is it important and how can you avoid it?

Answer by **Matthew Mayo**.

Selection bias, in general, is a problematic situation in which error is introduced due to a non-random population sample. For example, if a given sample of 100 test cases was made up of a 60/20/15/5 split of 4 classes which actually occurred in relatively equal numbers in the population, then a given model may make the false assumption that probability could be the determining predictive factor. Avoiding non-random samples is the best way to deal with bias; however, when this is impractical, techniques such as resampling, boosting, and weighting are strategies which can be introduced to help deal with the situation.

5. Explain selection bias (with regard to a dataset, not variable selection). Why is it important? How can data management procedures such as missing data handling make it worse?

- Selection of individuals, groups or data for analysis in such a way that proper randomization is not achieved

Types:
- Sampling bias: systematic error due to a non-random sample of a population causing some members to be less likely to be included than others
- Time interval: a trial may terminated early at an extreme value (ethical reasons), but the extreme value is likely to be reached by the variable with the largest variance, even if all the variables have similar means
- Data: “cherry picking”, when specific subsets of the data are chosen to support a conclusion (citing examples of plane crashes as evidence of airline flight being unsafe, while the far more common example of flights that complete safely)
- Studies: performing experiments and reporting only the most favorable results
- Can lead to unaccurate or even erroneous conclusions
- Statistical methods can generally not overcome it

Why data handling make it worse?
- Example: individuals who know or suspect that they are HIV positive are less likely to participate in HIV surveys
- Missing data handling will increase this effect as it’s based on most HIV negative
-Prevalence estimates will be unaccurate