# Tool: Data Quality and Bias Mitigation

**Gather and Qualify Data**

**Instructions:**

Use the tables below to determine where bias may be influencing your sample. Consider whether your sample is representative and whether the results you obtain will be informative.

**Is the sample representative?**

Degree of randomness: Move as far from convenience sampling as you can.

|  |  |  |
| --- | --- | --- |
| **Randomness** | | |
| **Factor to Consider** | **Possible Bias or Limitation** | **Mitigation Action** |
| Did/will all cases in the population have a chance to appear in the sample? | Data may not be valid when considering certain cases or groups of cases. | Define range of data collection. Identify collectable data that lies outside this range.  Identify cases or subgroups of cases outside the range of collected data and use quota sampling to add data points to the sample. |
| Did/will all cases or subgroups in the population have the same chance of appearing in the sample? | Data may artificially skew so that some cases or subgroups dominate or are overlooked in the results. | Use stratified random sampling so that all subgroups are well represented. |

Non-response bias: Make sure your actual sample is as close as possible to random.

|  |  |  |
| --- | --- | --- |
| **Non-response bias** | | |
| **Factor to Consider** | **Possible Bias or Limitation** | **Mitigation Action** |
| As you are polling or sampling, what is the response rate?  What proportion of the data is coming back blank or unusable? | Data may skew toward self-selection, causing a bias toward people who have an interest, willingness, or ability to respond. | If polling self-selecting respondents, work to maximize response rate.   * Consider what may be preventing respondents from getting back to you. Increase ease of responding. * If using a mail survey, send reminder mailers or follow up by telephone. * If using a telephone survey, call back, leave messages, and consider in-person follow-up.   If you are rejecting cases or responses with incomplete data, maximize usable data.   * Consider ways to get valid results from these cases even if they are incomplete.   Consider how you could revisit the case or response to fill in missing data. For example, you can fill in missing data with zero values or with values that are the average of the data you have. |

**Is the sample informative?**

|  |  |  |
| --- | --- | --- |
| **Information value** | | |
| **Factor to Consider** | **Possible Bias or Limitation** | **Mitigation Action** |
| How large is the actual sample?  What proportion of the population is actually being sampled? | Data may fall below a threshold of statistical significance. | Define both an absolute minimum threshold (sample >30) and a relative minimum threshold for sample size, and resample or continue sampling until the larger of these values is achieved.  For automated processes, consider whether the sampling rate takes into account factors of production that may introduce bias into the sample for certain rates of sampling. |