Appendix 4: Inverse probability weights

Where a weighted random sampling plan has been used to oversample priority sites or time periods (e.g., boat ramps, school holidays), or when there have been unexpected changes to the sampling plan, inverse probability weights (IPW) should be used in analyses to correct for sampling bias and increase representation across sampling time frame.

There are three steps to calculating IPW:

5.1 Identify the strata to be weighted.

Sampling site is the only strata to be weighted, if data was collected on intensive field trips, where sampling across sites was uneven (e.g., 6 samples are from Ramp A and 17 samples are from Ramp B). However, where sampling has occurred less intensively across extended periods (e.g., 2 days per week over 6 months), researchers may want to consider weighting across sampling sites, day type (weekday/weekend), school holidays (yes/no) and time (AM/PM). The number of samples within each sampling level must be known. Table 1 presents a simple example.

Table 1. A simple example of Step 5.1, where sampling has occurred over an extended period, and “Ramp” and “School Holidays” have been identified as strata to be weighted. There are 48 samples in total (48 individual sampling shifts).

Strata	Levels	Number of samples (n = 48)
Ramp	Ramp A	17
	Ramp B	31
	Yes	14

School Holidays	No	34

5.2 *Calculate the inclusion probability*

The inclusion probability is the probability of each strata being included in the sample.

a. Divide the number of samples for every level in a strata by the total number of samples (Table 2).

Table 2. Step 5.2a is presented in the grey column, where the inclusion probability for each level of the weighting strata has been calculated.

Strata	Levels	Number of samples (n = 48)	Inclusion probability
Ramp	Ramp A	17	17/48 = 0.35
	Ramp B	31	31/48 = 0.65
School Holidays	Yes	14	14/48 = 0.29
	No	34	34/48 = 0.71

b. When weighting multiple strata, their inclusion probabilities should be multiplied for every observation in the data to obtain an overall inclusion probability for each observation (Table 3).

Table 3. Step 5.2b is presented, where the weights applicable to each observation are multiplied together to obtain the total inclusion probability.

Observation	Ramp (weight)	School holidays (weight)	Total inclusion probability
1	A (0.35)	Yes (0.29)	0.35 x 0.29 = 0.10
2	B (0.65)	Yes (0.29)	0.65 x 0.29 = 0.19
3	A (0.35)	No (0.71)	0.35 x 0.71 = 0.25
4	B (0.65)	No (0.71)	0.65 x 0.71 = 0.46

5.3 *Calculate IPW*

Divide one by the total inclusion probability for each observation to calculate the IPW (Table 4). IPWs should overweight strata levels which were undersampled, and should underweight strata levels which were oversampled. The IPW should be applied to all analyses, this corrects for sampling bias and increases the representativeness of the data across the sampling time frame.

Table 4. Presents Step 5.3 in the grey column, which shows how the IPWs are calculated.

Observation	Ramp (weight)	School holidays (weight)	Total probability (ramp weight x school holiday weight)	IPW (1/total probability)
1	A (0.35)	Yes (0.29)	0.35 x 0.29 = 0.10	1/0.10 = 10
2	B (0.65)	Yes (0.29)	0.65 x 0.29 = 0.19	1/0.19 = 5.26
3	A (0.35)	No (0.71)	0.35 x 0.71 = 0.25	1/0.25 = 4
4	B (0.65)	No (0.71)	0.65 x 0.71 = 0.46	1/0.46 = 2.17