2

Modes of Sampling and Statistical Inference

As explained in Sect. 1.4 the methods for estimating global or local quantities in space, time or space–time, are grouped in design-based and model-based methods. With 'method' we mean a combination of a method for selecting sampling units and a method for statistical inference, for instance estimation of a spatial mean or prediction of the values at points. A design-based method is defined as a method in which sampling units are selected by probability sampling and in which statistical inference is based on the sampling design, i.e., design-based inference, see Table 2.1. A model-based method is defined as a method in which the statistical inference is based on the model. There are no requirements on the selection of the sampling units, but in general probability sampling is sub-optimal for model-based inference, and purposive sampling of units is more efficient. Typical examples of purposive sampling for model-based inference are Centred Grid Sampling, Spatial Coverage Sampling and Geostatistical Sampling. In the following sections we elaborate on these modes of sampling unit selection and on modes of statistical inference.

2.1 Modes of Sampling Unit Selection

Three possible modes of sampling unit selection can be distinguished: *convenience* sampling, *purposive* sampling and *probability* or *random* sampling. The concept of convenience sampling is self-explanatory. An obvious example is when sampling is limited to roadsides or other easily accessible spots. The advantage of this mode is that it saves time and costs. The disadvantage is that the statistical properties are inferior to those of the other modes. For instance, estimates from a convenience sample have to be regarded as biased unless one is willing to accept specific assumptions about the sampling process and the spatial and temporal variation. These assumptions are often debatable, and this may or may not be acceptable, depending on the context of the project.

Table 2.1. Definition of design-based and model-based method as a combination of a method for selection of sampling units and a method for statistical inference

Type of method	Selection method	Inference method
Design-based method	Probability sampling	Design-based
Model-based method	Purposive sampling	Model-based

Purposive sampling tries to select the sampling units such that a given purpose is served best. An example is the 'free survey' method of mapping soil classes, whereby the surveyor selects the sampling locations that are expected to be most informative with respect to soil class delineation. In this example, the locations are selected in a subjective manner, using experience, visible landscape features and pedogenetic hypotheses, such that the surveyor expects the most useful information from his observations. Another example is where a centred regular grid or a zigzag transect is projected on a field in order to obtain a 'representative' sample.

Purposive sampling can also be done by formally optimizing an objective function related to the purpose. For instance, if the purpose is to map a spatial distribution by kriging and if geographical boundary effects are disregarded, it can be shown that the prediction-error variance is minimized by a centred triangular grid of sampling locations, under assumptions of stationarity and isotropy (McBratney et al., 1981).

If prior to the sampling a statistical model for the variation in the universe can be postulated and the prediction-error variance is a function of the sampling events only, i.e., independent of the sample data, then one could use this model to optimize the spacing of a regular grid, given a quality requirement on the mean or maximum kriging variance. The model can also be used to optimize the sampling events in the universe given the sample size. Such samples are referred to as model-based samples, and more specific as geostatistical samples when the postulated model is a geostatistical model. These methods are discussed in Sects. 8.3.2 and 8.3.4, respectively. A technique of intermediate complexity is that of Spatial Coverage Sampling, which optimizes an objective function of distances only (Sect. 8.3.3).

Probability sampling, unlike the other modes, selects sampling units at random. If this is done properly (according to a well-defined sampling design) the probabilities of selecting the units are known, and these probabilities provide the basis for statistical inference from the data. As discussed in Sects. 7.2 and 15.2, there are many techniques for random selection of sampling units. Collectively, this approach to sampling is referred to as the design-based approach, as opposed to the model-based approach, where the sampling units are fixed instead of random, and statistical inference is based on a model of the variation in the universe. The difference between design-based and model-

based inference is further explained in Sect. 2.2.1. The choice between these two approaches is an important statistical issue, which is dealt with separately in Sect. 4.1.

2.2 Modes of Statistical Inference from Sample Data

2.2.1 Design-Based versus Model-Based Inference

There are two fundamentally different approaches to sampling: the *design-based* approach, followed in classical survey sampling, and the *model-based* approach, followed in geostatistics and in time series analysis. Differences and relationships between these two approaches are extensively addressed in Särndal et al. (1992). De Gruijter and ter Braak (1990) discuss the issue in the spatial context, but the distinction also holds, and is equally relevant, for sampling in time and in space–time. The difference between the two approaches is illustrated in Fig. 2.1 by a simple example (Brus and de Gruijter, 1997): a square area is sampled at 25 locations and a 0/1 indicator variable z is measured to estimate the fraction of the area with value 1. Figure 2.1**a** shows a spatial distribution of z and 25 sampling locations. Averaging the observed values at these locations yields an estimate of the fraction.

Both approaches quantify the uncertainty of such an estimate by considering what would happen if sampling were repeated many times in a hypothetical experiment. Obviously, if neither the pattern of values nor the sampling locations were changed in this experiment, there would be no variation, so one or the other has to be varied. The two approaches differ as to which one is varied.

The design-based approach evaluates the uncertainty by repeated sampling with different sets of sampling locations, while regarding the pattern of values in the area as unknown but fixed. The sets of sampling locations are generated according to a chosen random sampling design. The *row* of figures (**a**, **b** and **c**) represents three possible outcomes.

By contrast, the model-based approach evaluates the uncertainty by repeated sampling with a fixed set of sampling locations, while varying the pattern of values in the area according to a chosen random model of the spatial variation. In this approach, the *column* of figures (**a**, **d** and **e**) represents three possible outcomes.

The experiment can remain truly hypothetical in most instances, because probability calculus enables one to determine what happens on average over all possible realizations. In more intricate situations, however, this is infeasible and repeated sampling has to be simulated numerically, varying either the sampling locations or the pattern of values, or both.

The fact that the two approaches use a different source of randomness has several important practical as well as theoretical consequences. Here we briefly discuss three of them. The main consequence is that the statistical inference

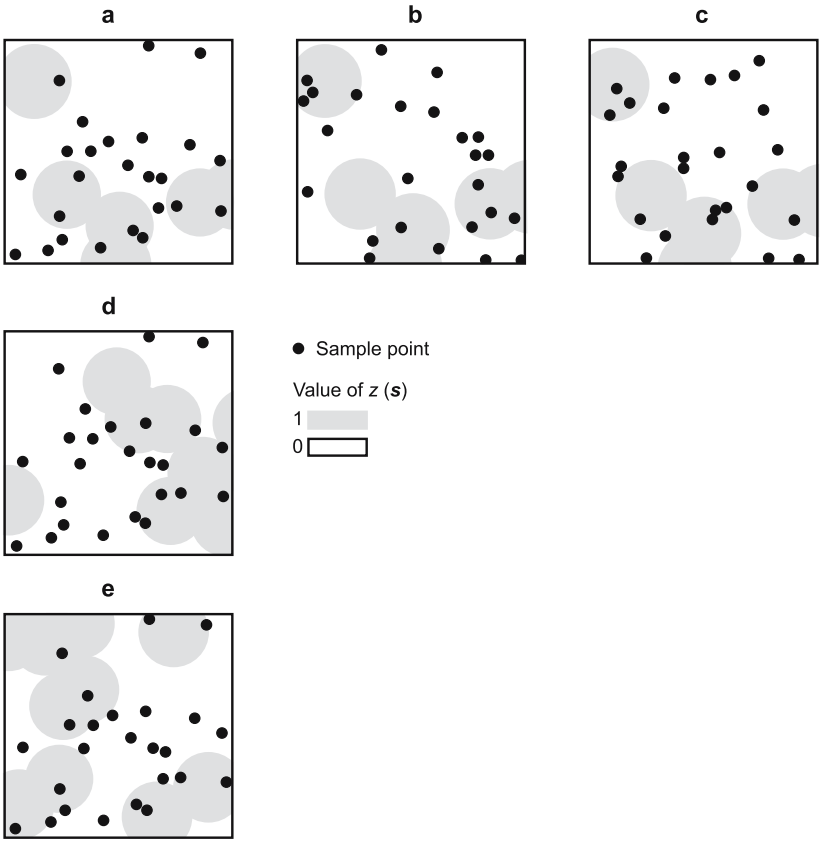

Fig. 2.1. Repeated sampling in the design-based approach (**a**, **b**, **c**) and in the model-based approach (**a**, **d**, **e**). In the design-based approach, the pattern of values is regarded as fixed and the sampling locations are random. In the model-based approach, the sampling locations are fixed and the pattern of values is regarded as random. (from Brus and de Gruijter, 1997)

from sample data is entirely different. In the design-based approach, inference is based on the selection probabilities of sampling locations as determined by the random sampling design. This means that in calculating weighted averages, for instance, the data are assigned weights that are determined by the selection probabilities of the sampling locations, not by their geographical coordinates. In the model-based approach, inference is based on a stochastic model of the variation in the universe. Here the weights of the data are determined by the covariances between the observations, which are given by the model as a function of the coordinates of the sampling locations.

Another important difference between design-based and model-based inference, apart from weighting, is the way they consider the quantity about which inference is made. This 'target quantity', a concept discussed in more detail in Sect. 3.1, is regarded as unknown but fixed in design-based inference. See, for instance, the constant areal fraction in Figs. 2.1**a**, **b** and **c**. In the model-based approach, however, there are two kinds of target quantities that can be chosen for inference: functions of parameters of the adopted stochastic model, and functions of realizations from that model. Model parameters and functions thereof are fixed by definition, while quantities of realizations are random.

A common and practically relevant example is that the model-based approach allows inference to be made about the model mean, denoted as μ, or about the mean defined by summation or integration over space and/or time, denoted by $\overline{Z}$. The former is fixed while the latter is random. For instance, the realizations in Figs. 2.1**a**, **d** and **e** have different areal fractions but the same underlying model mean. The difference between a fixed and a random target quantity is not merely a theoretical subtlety, but has direct consequences for the quantification and even the definition of the precision and reliability of results, as will be explained later in this section.

Finally, the choice between design-based and model-based inference (discussed in Sect. 4.1) has major consequences for sampling. Design-based inference requires some form of probability (random) sampling, while model-based inference typically builds on purposive sampling, and needs a model of the spatial variation. Acquisition of such a model will usually require an extra sampling effort. Furthermore, random sampling optimization is methodologically different from purposive sampling optimization, as discussed in Chap. 5.

It should also be realized that optimal sampling for model-based inference on model parameters differs from optimal sampling for model-based inference on realization quantities. For instance, inference about a model mean often requires a considerably larger sample than inference about a spatial, temporal or spatio–temporal mean with the same level of precision. Also, these two target quantities generally require different sampling patterns. Hence, different equations should be used to calculate sample sizes and different methods used for sample optimization.

It is useful to distinguish between inference modes not only in terms of whether the target quantity is considered as fixed or random, but also whether the inference result is quantitative (number or interval) or qualitative, for instance, that the target quantity exceeds a given level. Quantitative and qualitative results need their own methods of inference, and have their own quality measures for precision and reliability, so the sample size needed to meet a quality requirement has to be determined in different ways.

Combining the distinctions 'random versus fixed target quantity' with 'qualitative versus quantitative inference result' yields four different modes

Table 2.2. Summary of inference modes for different combinations of target quantities and types of results

Basis of inference	Target quantities	Type of result	Inference mode
Sampling design	Statistics defined on a *fixed* field of z-values, (e.g., the deterministic spatial mean $\bar{z}$)	Quantitative Qualitative	Estimation Testing
Stochastic model	Statistics defined on a *random* field of Z-values, (e.g., the stochastic spatial mean $\overline{Z}$), or:	Quantitative Qualitative	Prediction Classification
	Model parameters or functions thereof (*fixed* by definition)	Quantitative Qualitative	Estimation Testing

of inference: *estimation*, *prediction*[1], *testing* and *classification*. These modes are summarized in Table 2.2. A fifth mode, *detection*, should be applied when the question is whether at any point in the universe some critical condition is present, without asking where or when. The answer to this type of question can be coded as a 0/1 indicator variable. As this relates to the universe as a whole, it is a global quantity.

It is important to choose carefully between estimation and prediction on the one hand, and hypothesis testing, classification and detection on the other. This choice should be made prior to the actual design of the sampling scheme, because it is determined by the kind of information that is required. This issue is therefore discussed in Sect. 3.1 as part of the design information. Here we repeat as a guideline that whenever a choice must be made between two alternatives, e.g., whether or not to take a particular action or to draw a particular conclusion, and when this choice is to be made on the basis of sample data, hypothesis testing, classification or detection is in order. In all other cases, the original problem can be formulated as an estimation or prediction problem.

The statistical literature provides a huge variety of inference methods. In the following sections we only recapitulate the concepts that are most relevant to sampling. We illustrate the four modes of inference with the mean as target quantity, and the case depicted in Fig. 2.1 as an example. The

[1] Prediction should not be confused with forecasting. Prediction is quantitative inference on a random target quantity that may be spatial, temporal or spatio-temporal. Forecasting is a special case of prediction, i.e., when a (spatio-)temporal quantity is related to the future.

target variable in Fig. 2.1 is a 0/1 variable, indicating the presence or absence of some condition, e.g., the occurrence of a given pathogen in the soil. So in this example the mean can be interpreted as the areal fraction of infected soil.

2.2.2 Estimation

Estimation is quantitative inference on a fixed target quantity, however, as explained in the previous section, there are two kinds of fixed quantities: statistics defined on a fixed field of z-values, and (functions of) model parameters (Table 2.2). So, in the example of the areal fraction of infected soil, one first has to decide what is to be estimated: the *spatial* mean or a *model* mean. If the interest is focused on the infection actually present in the particular study region, then one would estimate the spatial mean. If, on the other hand, the interest is broader and relates to infection that may be present on average in a hypothetical ensemble of similar regions (of which the study region is only one example), then one would estimate a model mean.

Estimation of the Spatial Mean

When estimating the spatial mean, we consider this as fixed, as in the row of Figs. 2.1**a**, **b**, and **c**. The true value of the mean, denoted by $\bar{z}$, equals 0.30, but is unknown in practice. The design-based *estimator* of $\bar{z}$, denoted by $\hat{\bar{z}}$, is generally a function of the sample data without their coordinates in space and/or time. If the sample is drawn according to Simple Random Sampling (see Sect. 7.2.3), the usual estimator is the unweighted sample mean (unweighted because the sampling locations were selected with equal probability, not because of a model assumption):

$$\hat{\bar{z}} = \frac{1}{n} \sum_{i=1}^{n} z_i \ , \tag{2.1}$$

where n denotes the sample size (here 25), and z_i denotes the value measured at the i-th sampling location.

The data of the random sample depicted in Fig. 2.1**a** are: 8 'infected' out of 25, giving an *estimate* $\hat{\bar{z}} = 0.32$, with *estimation error* $\hat{\bar{z}} - \bar{z} = 0.02$. The other, equally probable samples depicted in Figs. 2.1**b** and **c** would yield 0.40 and 0.32, respectively, with estimation errors 0.10 and 0.02. In practice, the estimation error is unknown for any specific sample (otherwise there would be no need for sampling), but probability theory enables us to make statements about the estimation error in terms of averages.

First, it can be shown that the estimator, averaged over samples, equals the true value. In other words, $\hat{\bar{z}}$ is *p-unbiased*:

$$E_p\big(\hat{\bar{z}}\big) = \bar{z} \ , \tag{2.2}$$

where $E_p(\cdot)$ denotes the *p-expectation*, i.e., the average over a large (strictly: infinite) number of samples drawn according to the random sampling design p, here Simple Random Sampling with $n = 25$.

Second, the p-expectation of the squared estimation error equals the *sampling variance* of the estimator:

$$V_p\big(\hat{\bar{z}}\big) = E_p\big\{\hat{\bar{z}} - E_p\big(\hat{\bar{z}}\big)\big\}^2 = E_p\big(\hat{\bar{z}} - \bar{z}\big)^2 \ . \tag{2.3}$$

Because the sampling locations were selected independently from each other, the number of 'infected' in the data follows a binomial distribution, and the sampling variance equals

$$V_p\big(\hat{\bar{z}}\big) = \frac{\bar{z}\,(1-\bar{z})}{n} \ . \tag{2.4}$$

The *standard error* of the estimator is a usual statistical quality measure. It equals the square root of the sampling variance:

$$S_p\big(\hat{\bar{z}}\big) = \sqrt{\frac{\bar{z}\,(1-\bar{z})}{n}} \ . \tag{2.5}$$

In our example V_p equals $0.30 \times 0.70/25 = 0.0084$, and S_p equals 0.0917.

In practice these true values would be unknown because $\bar{z}$ is unknown but, from data obtained by Simple Random Sampling, the sampling variance can be estimated (again p-unbiased) by

$$\widehat{V}_p\big(\hat{\bar{z}}\big) = \frac{1}{n\,(n-1)} \sum_{i=1}^{n} \big(z_i - \hat{\bar{z}}\big)^2 \ . \tag{2.6}$$

In our example, with 0/1 data, this results in:

$$\widehat{V}_p\big(\hat{\bar{z}}\big) = \frac{\hat{\bar{z}}\,\big(1-\hat{\bar{z}}\big)}{n-1} \ . \tag{2.7}$$

The estimated standard error follows as its square root:

$$\widehat{S}_p\big(\hat{\bar{z}}\big) = \sqrt{\frac{\hat{\bar{z}}\,\big(1-\hat{\bar{z}}\big)}{n-1}} \ . \tag{2.8}$$

From the sample data in Fig. 2.1**a** we calculate $\widehat{V}_p = 0.00907$, and $\widehat{S}_p = 0.0952$.

Apart from *point* estimation as discussed above, giving a single number or point on the scale of possible values, one can also perform *interval* estimation. The result is a *confidence interval*. Confidence intervals are constructed in such a way that they contain the true value of the target quantity with probability $(1-\alpha)$, referred to as the *confidence level*. For instance, the data from the sample in Fig. 2.1**a**, 8 'infected' out of 25, result in (0.14; 0.54) as a 95% confidence interval for the areal fraction of infected soil. (This interval is based on the fact that the number of 'infected' follows a binomial distribution, but a Normal or Student distribution is usually applicable, especially with quantitative target variables and medium size or large samples.)

Estimation of the Model Mean

In practical applications one usually has to choose a model for the inference on the basis of sample data and possibly ancillary data. In the present example, however, we happen to have exact knowledge a priori of the model underlying the patterns of Figs. 2.1**a**, **b** and **c**, simply because we used this model to generate those patterns.

The model was discussed by Matérn (1986) in the context of the so-called 'bombing problem': bombs are dropped at random over a region and each bomb devastates the area within a given distance r from the point of impact. It can be shown (Matérn, 1986, Eq. 3.4.2) that under this model the expectation of the 0/1 variable Z at location $\mathbf{s}$ equals:

$$E_\xi\{Z(\mathbf{s})\} = 1 - \mathrm{e}^{-\lambda\pi r^2} \;, \tag{2.9}$$

where $E_\xi(\cdot)$ denotes the ξ-*expectation*, i.e., the average over a large (strictly: infinite) number of random realizations from the chosen model, and where λ is the intensity of the assumed stationary Poisson process by which the centres of the infection circles are spread over the region.

Because of the spatial stationarity, meaning that the expectation at locations is constant over the region, the model mean μ equals $E_\xi[Z(\mathbf{s})]$. For our example we have chosen $\lambda = 6$ and $r = 0.15$. From (2.9) we calculate for the true value of the model mean: $\mu = 0.346$, which is the areal fraction 'infected' averaged over a large (strictly infinite) number of realizations.

To estimate the model mean, we need the covariance between any two observations on Z. The covariance between two observations at locations h apart equals (Matérn, 1986, Eq. 3.4.3):

$$C(h) = \mathrm{e}^{-2\lambda\pi r^2}\left(\mathrm{e}^{\lambda\, A(h,r)} - 1\right) \;, \tag{2.10}$$

where $A(h,r)$ is the area of the intersection common to two circles with radius r and centres h apart. This area equals:

$$A(h,r) = r^2\left[2\arccos\left(\frac{h}{2r}\right) - \sin\left\{2\arccos\left(\frac{h}{2r}\right)\right\}\right] \;. \tag{2.11}$$

The model mean can be estimated with the so-called Best Linear Unbiased Estimator (Searle, 1997):

$$\hat{\mu} = \left(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\mathbf{C}^{-1}\mathbf{z} \;, \tag{2.12}$$

where $\mathbf{z}$ is a vector of n observations, $Z(\mathbf{s}_i)$, $\mathbf{C}$ is the matrix of their variances and covariances, and $\mathbf{1}$ is the n-vector of ones. This estimator is ξ-unbiased, a property defined by averaging over model realizations (given the sampling locations), not over the sampling process, i.e., $E_\xi(\hat{\mu}) = \mu$. The variance of $\hat{\mu}$ equals:

$$V_\xi(\hat{\mu}) = \left(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}\right)^{-1} . \tag{2.13}$$

From the sample depicted in Fig. 2.1**a**, and Eqs. 2.10 – 2.13, we calculate $\hat{\mu} = 0.353$ with standard error $S_\xi(\hat{\mu}) = \sqrt{V_\xi(\hat{\mu})} = 0.132$. Note that, although the estimation error in this case is only 0.007, the standard error of estimation of the model mean is 44% larger than the standard error of estimation of the spatial mean.

2.2.3 Prediction

With the example of Section 2.2.2 we illustrated design-based estimation of the spatial mean $\bar{z}$ and model-based estimation of the model mean μ, both means considered as fixed. Using the same data and the same model, we now illustrate model-based prediction of the spatial mean $\overline{Z}$, considered as random. Analogous to the Best Linear Unbiased Estimator of μ, one can predict $\overline{Z}$ with the Best Linear Unbiased Predictor:

$$\widetilde{\overline{Z}} = \boldsymbol{\lambda}'\mathbf{z} , \tag{2.14}$$

where $\boldsymbol{\lambda}$ is a vector of optimal weights, and $\mathbf{z}$ is again the vector of observations. These weights are chosen such that the prediction is ξ-unbiased, i.e., $E_\xi\left(\widetilde{\overline{Z}} - \overline{Z}\right) = 0$, and the variance of the prediction error, $E_\xi\left(\widetilde{\overline{Z}} - \overline{Z}\right)^2$, is minimized. These optimal weights can be calculated by:

$$\boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{r} - \mathbf{C}^{-1}\mathbf{1}\left(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}\right)^{-1}\mathbf{1}'\mathbf{C}^{-1}\mathbf{r} + \left(\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}\right)^{-1}\mathbf{C}^{-1}\mathbf{1} , \tag{2.15}$$

where $\mathbf{r}$ is the vector of mean covariances between each of the actual sampling locations and all potential locations in the region. This can be approximated numerically by the mean covariances between the sampling locations and the nodes of a fine grid.

The variance of the prediction error equals

$$V_\xi\left(\widetilde{\overline{Z}} - \overline{Z}\right) = E_\xi\left(\widetilde{\overline{Z}} - \overline{Z}\right)^2 = \overline{C}_{G,G} + \boldsymbol{\lambda}'\mathbf{C}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\mathbf{r} , \tag{2.16}$$

where $\overline{C}_{G,G}$ is the mean covariance between all potential sampling locations of the region, which can be approximated numerically by the mean covariance between the nodes of a fine grid.

From the sample depicted in Fig. 2.1**a** and Eqs. 2.14 – 2.16 we calculate $\widetilde{\overline{Z}} = 0.386$, with standard error (square root of the prediction-error variance) 0.078. Note that this standard error of the spatial mean is only 59% of the standard error of the model mean, using the same data and the same model. (A considerably larger sample would be needed to estimate the model mean with the same precision as for predicting the spatial mean.)

Apart from *point* prediction as discussed above, one can also perform *interval* prediction. Just as an estimation interval (Sect. 2.2.2), a prediction

interval is constructed in such a way that it contains the true value of the target quantity with probability $(1 - \alpha)$. If on the basis of the Central Limit Theorem it can be assumed that the prediction error is approximately normally distributed, then for the boundaries of a 95% prediction interval, for instance, one can simply take the prediction plus and minus 1.96 times the standard error. The data from the sample in Fig. 2.1**a**, 8 'infected' out of 25, then give (0.233; 0.538) as 95% prediction interval for the areal fraction of infected soil. Note that this model-based prediction interval is 25% narrower than the design-based estimation interval calculated in Sect. 2.2.2 for the same target quantity, using the same sample data and the same confidence level. (One should realize, however, that in practical applications the variogram is not known and must be estimated, which is an additional source of error.)

In cases where the assumption of approximate normality does not hold, the distribution of $\overline{Z}$ may be evaluated numerically by Monte Carlo simulation of realizations from the chosen model. Deutsch and Journel (1998) and Pebesma (2004) provide simulation software.

2.2.4 Testing

When decisions are based on an estimate of the global mean, the areal fraction or a percentile (e.g., median or P95), one must take account of the uncertainty in the estimate. For instance, in environmental studies a threshold value for the concentration of a pollutant is often defined on the basis of regulatory requirements, risk assessments or a reference standard for deciding whether or not to take action. The threshold value is referred to as the Action Level. Due to sampling and measurement error, there is always some degree of uncertainty whether the true value is above or below the Action Level, which means that there is a chance of taking a wrong decision. A statistical tool for decision making in the presence of uncertainty is statistical testing of hypotheses. A distinction can be made between one-sample tests and two-sample tests. In one-sample tests, sample data from a target area are compared with an absolute criterion such as a regulatory threshold. In two-sample tests sample data from a target area are compared with other sample data, for instance from another region or from another period for the same target area. In the two-sample case, the target quantity is usually the difference between the two (spatial or temporal) means, and the Action Level is often zero.

For an example of a one-sample test we go back to Fig. 2.1 and suppose that some remediation action is to be taken if the areal fraction of infected soil, $\bar{z}$, is larger than Action Level 0.2. The null-hypothesis, H_0, is then: $\bar{z} \leq 0.2$, which is tested against the alternative hypothesis H_1: $\bar{z} > 0.2$. Acceptance or rejection of H_0 is determined by the value of a *test statistic*, in this case the sample mean $\hat{\bar{z}}$, in the following way. If the value of the test statistic falls in a specific interval, called the *critical region*, then H_0 is rejected, otherwise it is accepted. This critical region is constructed such that the probability of falsely rejecting H_0 is limited to a value α, chosen by the user.

Table 2.3. The four possible combinations of the state of nature and test results.

	Test result	
State of nature	H_0	H_1
H_0	No error Prob= $1-\alpha$ Confidence	Type I error Prob= α
H_1	Type II error Prob= β	No error Prob= $1-\beta$ Power

Table 2.3 shows the four possible combinations of the state of nature and test results. Wrongly accepting H_1 is called a type I error, the probability of which is α. The probability of accepting H_0 when H_0 is true is called the confidence level of the test, $(1-\alpha)$. Wrongly accepting H_0 is called a type II error, the probability of which is denoted by β. The probability of accepting H_1 when H_1 is true is called the power of the test, $1-\beta$. Given the sampling design, the power is a function of the sample size, the confidence level $(1-\alpha)$, and the H_1 hypothesis itself.

When we use the data from the sample in Fig. 2.1**a** for a one-sided binomial test of H_0: $\bar{z} \leq 0.2$ at confidence level 0.95, we find as critical region [0.36; 1], which does not include the value of the test statistic (0.32), so that H_0 is not rejected.

There is a close connection between estimation and testing: a confidence interval contains all possible values of the target quantity that would not be rejected on testing at the same confidence level. However, the statistical quality measures differ (standard error or half-width of confidence interval versus confidence level and power), and so do methods to determine the sample size needed to reach a given quality level.

2.2.5 Classification

The term classification is used here for assigning an object to a class under uncertainty about the properties of the object. In the present context the object is the universe of interest or a part of it, and the uncertainty about its properties is represented by a stochastic model. For a simple illustration we take once more the case of Fig. 2.1**a**. Suppose that the region is to be classified in one of two classes: 'low infection' defined as $\overline{Z} \leq 0.5$, and 'high infection' defined as $\overline{Z} > 0.5$. After sampling and measuring one could decide to classify the region simply by comparing the predicted spatial mean with the class boundary 0.5. However, due to the uncertainty about the true value

of the spatial mean, there will be a risk of false 'low' classification and a risk of false 'high' classification. So, just as in testing, there are two types of possible errors, and in view of the consequences of either one it may be appropriate to classify in such a way that the probability of the most important error type, for instance false 'low', is not too high. This can be done by classifying according to a rule in terms of error probability instead of the prediction itself.

For example, the rule could be that the region will only be classified as 'low' if the probability of false 'low' is smaller than 0.05, and otherwise as 'high'. Assuming that the prediction error is normally distributed, the probability of false 'low' can be calculated from the prediction (0.386) and its standard error (0.078; see Sect. 2.2.3). This turns out to be 0.07, which is larger than the chosen threshold 0.05, hence the region is classified as 'high', although the predicted value falls in class 'low'.

As with prediction, if the assumption of normality does not hold, the distribution of $\overline{Z}$ must be evaluated numerically by Monte Carlo (geostatistical) simulation of realizations from the chosen model (Deutsch and Journel, 1998; Pebesma, 2004).

2.2.6 Detection

An example of a detection problem in the case of Fig. 2.1**a** is the question whether there is soil infection *at all* in the area, regardless where. The assessment method is often designed such that if the critical condition is observed in any of the sampling units, then it surely exists. In that case the probability of a false positive is zero, and inference only needs to quantify the probability of a false negative or its complement, the detection probability. Sampling for detection should therefore aim at maximizing the detection probability for a given budget, or minimizing the sampling costs under the condition of a given minimal detection probability. This will naturally lead to grid sampling or spatial coverage sampling.

The detection probability can in principle be calculated from the geometry of the sampling pattern used, and certain assumptions about the occurrence and extent of the condition in space and/or time. These assumptions may be captured in a stochastic model of the variation, e.g., the second-order stationary model with covariance function (2.10) underlying Fig. 2.1. The detection probability for any sampling pattern can then be calculated by geostatistical simulation (Deutsch and Journel, 1998; Pebesma, 2004). Alternatively, knowledge about the occurrence and extent of the critical condition in space and/or time may be embodied in a physical process model.

Detection of critical conditions sometimes asks for highly specialized and application-dependent sampling and inference methods. Such methods are not treated in this book.