Chapter 2

Designing a Sampling Program

2.1 Introduction

The first and most important step of any environmental study is to design the sampling program. This chapter discusses the basics of designing a sampling program, and shows you how to use EnvStats to help you determine required sample sizes. For a more in-depth discussion of sampling design and sample size calculation, see Millard et al. (2014).

2.2 The Necessity of a Good Sampling Design

A study is only as good as the data upon which it is based. No amount of advanced, cutting-edge statistical theory and techniques can rescue a study that has produced poor quality data, not enough data, or data irrelevant to the question it was meant to answer. From the very start of an environmental study, there must be a constant dialog between the data producers (field and lab personnel, data coders, etc.), the data users (scientists and statisticians), and the ultimate decision maker (the person for whom the study was instigated in the first place). All persons involved in the study must have a clear understanding of the study objectives and the limitations associated with the chosen physical sampling and analytical (measurement) techniques before anyone can make any sense of the resulting data.

2.3 What Is a Population and What Is a Sample?

In everyday language, the word "population" refers to all the people or organisms contained within a specific country, area, region, etc. When we talk about the population of the United States, we usually mean something like "the total number of people who currently reside in the U.S."

In the field of statistics, however, the term *population* is defined operationally by the question we ask: it is the entire collection of measurements about which we want to make a statement (Zar 2010; Berthouex and Brown 2002; Gilbert 1987).

For example, if the question is "What is the concentration of dissolved oxygen in this stream?", the question must be further refined until a suitable population can be defined: "What is the average concentration of dissolved oxygen in a particular section of a stream at a depth of 0.5 m over a particular 3-day period?" In this case, the population is the set of all possible measurements of dissolved oxygen in

that section of the stream at 0.5 m within that time period. The section of the stream, the time period, the method of taking water samples, and the method of measuring dissolved oxygen all define the population.

A *sample* is defined as some subset of a population (Zar 2010; Berthouex and Brown 2002; Gilbert 1987). If the sample contains all the elements of the population, it is called a *census*. Usually, a population is too large to take a census, so a portion of the population is sampled. The statistical definition of the word sample (a selection of individual population members) should not be confused with the more common meaning of a *physical sample* of soil (e.g., 10 g of soil), water (e.g., 5 ml of water), air (e.g., 20 cc of air), etc.

2.4   Random Versus Judgment Sampling

Judgment sampling involves subjective selection of the population units by an individual or group of individuals (Gilbert 1987). For example, the number of samples and sampling locations might be determined based on expert opinion or historical information. Sometimes, public opinion might play a role and samples need to be collected from areas known to be highly polluted. The uncertainty inherent in the results of a judgment sample cannot be quantified and statistical methods cannot be applied to judgment samples. Judgment sampling does *not* refer to using prior information and the knowledge of experts to define the area of concern, define the population, or plan the study. Gilbert (1987) also describes "haphazard" sampling, which is a kind of judgment sampling with the attitude that "any sample will do" and can lead to "convenience" sampling, in which samples are taken in convenient places at convenient times.

Probability sampling or *random sampling* involves using a random mechanism to select samples from the population (Gilbert 1987). All statistical methods used to quantify uncertainty assume some form of random sampling has been used to obtain a sample. At the simplest level, a simple random sample is used in which each member of the population has an equal chance of being chosen, and the selection of any member of the population does not influence the selection of any other member. Other probability sampling methods include stratified random sampling, composite sampling, and ranked set sampling.

2.5   Common Mistakes in Environmental Studies

The most common mistakes that occur in environmental studies include the following:

- **Using Judgment Sampling to Obtain Samples.** When judgment sampling is used to obtain samples, there is no way to quantify the precision and bias of any type of estimate computed from these samples.
- **Lack of Samples from Proper Control Populations**. If one of the objectives of an environmental study is to determine the effects of a pollutant on some specified population, then the sampling design must

include samples from a proper control population. This is a basic tenet of the scientific method. If control populations were not sampled, there is no way to know whether the observed effect was really due to the hypothesized cause, or whether it would have occurred anyway.

- **Failing to Randomize over Potentially Influential Factors.** An enormous number of factors can influence the final measure associated with a single sampling unit, including the person doing the sampling, the device used to collect the sample, the weather and field conditions when the sample was collected, the method used to analyze the sample, the laboratory to which the sample was sent, etc. A good sampling design controls for as many potentially influencing factors as possible, and randomizes over the factors that cannot be controlled. For example, if data are collected from two sites, and two laboratories are used to analyze the results, you should not send all the samples from site 1 to laboratory A and all the samples from site 2 to laboratory B, but rather send samples collected at each site to each of the laboratories.

- **Collecting Too Few Samples to Have a High Degree of Confidence in the Results.** The ultimate goal of an environmental study is to answer one or more basic questions. These questions should be stated in terms of hypotheses that can be tested using statistical procedures, as well as what constitutes an important scientific effect since statistically significant effects are not always scientifically important. In this case, you can determine the probability of rejecting the null hypothesis when in fact it is true (a Type I error), and the probability of not rejecting the null hypothesis when in fact it is false (a Type II error). Usually, the Type I error is set in advance, and the probability of correctly rejecting the null hypothesis when in fact it is false (the power), or the width of a confidence, prediction, or tolerance interval, is calculated for various sample sizes and assumed amounts of variability. Too often, this step of determining power and/or interval width versus sample size is neglected, resulting in a study from which no conclusions can be drawn with any great degree of confidence.

2.6   The Data Quality Objectives Process

The Data Quality Objectives (DQO) process is a systematic planning tool based on the scientific method that has been developed by the U.S. Environmental Protection Agency (USEPA 2006b). The DQO process provides an easy-to-follow, step-by-step approach to decision-making in the face of uncertainty. Each step focuses on a specific aspect of the decision-making process. Data Quality Objectives are the qualitative and quantitative statements that:

- Clarify the study objective.
- Define the most appropriate type of data to collect.
- Determine the most appropriate conditions under which to collect the data.

- Specify acceptable levels of decision errors that will be used as the basis for establishing the quantity and quality of data needed to support the decision.

The seven steps in the DQO process are: (1) state the problem, (2) identify the goals of the study, (3) identify information inputs, (4) define boundaries of the study, (5) develop the analytic approach, (6) specify performance or acceptance criteria, and (7) develop the plan for obtaining the data (see Millard et al. 2014, for more details). Steps 5 and 6 involve deciding what statistical methods you will use and trading off limits on Type I and Type II errors and sample size.

2.7 Power and Sample Size Calculations

EnvStats contains several functions to assist you in determining how many samples you need for a given degree of confidence in the results of a sampling program (see the help file *Power and Sample Size*). These functions are based on the ideas of confidence intervals, prediction intervals, tolerance intervals, and hypothesis tests. If you are unfamiliar with these concepts, please see Millard et al. (2014).

A very important point to remember is that no matter what you come up with for estimates of required sample sizes, it is always a good idea to assume you will lose some percentage of your observations due to sample loss, sample contamination, database issues, etc.

2.8 Sample Size for Confidence Intervals

Table 2.1 lists the functions available in EnvStats for computing required sample sizes, half-widths, and confidence levels associated with a confidence interval. For the normal and binomial distributions, you can compute the half-width of the confidence interval given the user-specified sample size, compute the required sample size given the user-specified half-width, and plot the relationship between sample size and half-width. For a nonparametric confidence interval for a percentile, you can compute the required sample size for a specified confidence level, compute the confidence level associated with a given sample size, and plot the relationship between sample size and confidence level. Chapter 5 gives more details on computing confidence intervals once you have your data.

Bacchetti (2010) presents strong arguments against the current convention in scientific research for computing sample size that is based on formulas that use a fixed Type I error (usually 5 %) and a fixed minimal power (often 80 %) without regard to costs. He notes that a key input to these formulas is a measure of variability (usually a standard deviation) that is difficult to measure accurately "unless there is so much preliminary data that the study isn't really needed." Also, study designers often avoid defining what a scientifically meaningful difference is by presenting sample size results in terms of the effect size (i.e., the difference of interest divided by the elusive standard deviation). Bacchetti (2010) encourages

study designers to use simple tables in a sensitivity analysis to see what results of a study may look like for low, moderate, and high rates of variability and large, intermediate, and no underlying differences in the populations or processes being studied.

Distribution	Function	Output
Normal	ciTableMean	Confidence intervals for mean of normal distribution, or difference between two means, following Bacchetti (2010)
	ciNormHalfWidth	Half-width of confidence interval for mean of normal distribution or difference between two means
	ciNormN	Required sample size for specified half-width of confidence interval for mean of normal distribution or difference between two means
	plotCiNormDesign	Plots for sampling design based on confidence interval for mean of normal distribution or difference between two means
Binomial	ciTableProp	Confidence intervals for binomial proportion, or difference between two proportions, following Bacchetti (2010)
	ciBinomHalfWidth	Half-width of confidence interval for binomial proportion or difference between two proportions
	ciBinomN	Required sample size for specified half-width of confidence interval for binomial proportion or difference between two proportions
	plotCiBinomDesign	Plots for sampling design based on confidence interval for binomial proportion or difference between two proportions
Nonparametric	ciNparConfLevel	Confidence level of confidence interval for a percentile, given the sample size
	ciNparN	Required sample size for specified confidence level of confidence interval for a percentile
	plotCiNparDesign	Plots for sampling design based on confidence interval for a percentile

Table 2.1 Sample size functions for confidence intervals

2.8.1   Confidence Interval for the Mean of a Normal Distribution

The EnvStats function ciTableMean produces a table similar to Table 1 of Bacchetti (2010) for looking at how the confidence interval for the mean of a normal distribution or the difference between two means varies with various levels of variability and the value of the estimated mean or difference between two means, given the sample size and confidence level.   The EnvStats function ciNormHalfWidth computes the half-width associated with the confidence interval, given the sample size, estimated standard deviation, and confidence level. The function ciNormN computes the sample size required to achieve a specified half-width, given the estimated standard deviation and confidence level.   The function plotCiNormDesign plots the relationships between sample size, half-width, estimated standard deviation, and confidence level.

    The data frame EPA.09.Ex.16.1.sulfate.df contains sulfate concentrations (ppm) at one background and one downgradient well. The estimated mean and standard deviation for the background well are 536 and 27 ppm, respectively, based on a sample size of $n = 8$ quarterly samples take over 2 years. A two-sided 95 % confidence interval for this mean is [514, 559], which has a half-width of 23 ppm.

```
> EPA.09.Ex.16.1.sulfate.df

   Month Year     Well.type Sulfate.ppm
1    Jan 1995    Background          560
2    Apr 1995    Background          530
…
15   Jul 1996 Downgradient          610
16   Oct 1996 Downgradient          630

> Sulfate.back <- with(EPA.09.Ex.16.1.sulfate.df,
    Sulfate.ppm[Well.type == "Background"])

> enorm(Sulfate.back, ci = TRUE)

Results of Distribution Parameter Estimation
-------------------------------------------


Assumed Distribution:          Normal

Estimated Parameter(s):        mean = 536.25000
                               sd   =  26.69270

Estimation Method:             mvue

Data:                          Sulfate.back

Sample Size:                   8
```

```
Confidence Interval for:          mean

Confidence Interval Method:       Exact

Confidence Interval Type:         two-sided

Confidence Level:                 95%

Confidence Interval:              LCL = 513.9343
                                  UCL = 558.5657
```

Suppose we are planning a future study and are interested in the size of the confidence interval. Initially we plan to take eight quarterly samples taken over 2 years, as in the previous study. We could assume an estimated standard deviation of about 25 or 30 ppm, but based on the 95 % confidence interval for the variance, which is [311, 2,951] ppm, the true standard deviation may be as small as about 18 ppm or as large as about 54 ppm.

```
> enorm(Sulfate.back, ci = TRUE,
    ci.param = "variance")$interval
```

```
Confidence Interval for:          variance

Confidence Interval Method:       Exact

Confidence Interval Type:         two-sided

Confidence Level:                 95%

Confidence Interval:              LCL =  311.4703
                                  UCL = 2951.4119
```

Letting the estimated standard deviation vary from 15 to 60 ppm shows that the width of the confidence interval varies between about 13 and 50 ppm:

```
> ciNormHalfWidth(n.or.n1 = 8, sigma.hat = c(15, 30, 60))
```

```
[1] 12.54031 25.08063 50.16126
```

Assuming a standard deviation of about 30 ppm, if in a future study we take only four observations, the half-width of the confidence interval should be about 48 ppm:

```
> ciNormHalfWidth(n.or.n1 = 4, sigma.hat = 30)
```

```
[1] 47.73669
```

Also, if we want the confidence interval to have a half-width of 10 ppm, we would need to take $n = 38$ observations (i.e., quarterly samples taken over more than 9 years).

```
> ciNormN(half.width = 10, sigma.hat = 30)
```

```
[1] 38
```

Figure 2.1 displays the half-width of the confidence interval as a function of the sample size for various confidence levels, again assuming a standard deviation of about 30 ppm.
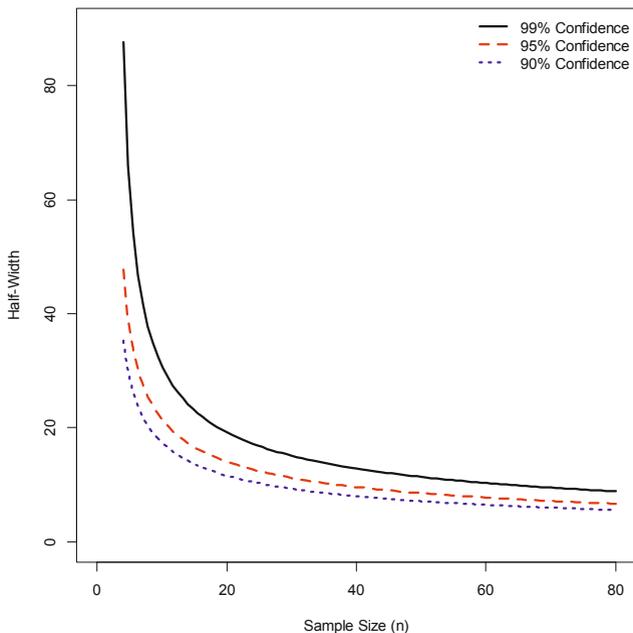

Fig. 2.1 The half-width of the confidence interval for the mean of background sulfate concentration (ppm) as a function of sample size and confidence level, assuming a standard deviation of 30 ppm

To create this plot, type these commands:

```
> plotCiNormDesign(sigma.hat = 30, range.x.var = c(4, 80),
    conf = 0.99, xlim = c(0, 80), ylim = c(0, 90), main = "")

> plotCiNormDesign(sigma.hat = 30, range.x.var = c(4, 80),
    conf = 0.95, plot.col = 2, plot.lty = 2, add = TRUE)

> plotCiNormDesign(sigma.hat = 30, range.x.var = c(4, 80),
    conf = 0.90, plot.col = 4, plot.lty = 3, add = TRUE)
```

```
> legend("topright",
    paste(c("99%", "95%", "90%"), "Confidence")
    col = c(1, 2, 4), lty = 1:3, lwd = 2, bty = "n")
```

Considering the data frame EPA.09.Ex.16.1.sulfate.df again, the estimated mean and standard deviation for the *downgradient* well are 608 and 18 ppm, respectively, based on a sample size of $n = 6$ quarterly samples. A two-sided 95 % confidence interval for the difference between this mean and the background mean is [44, 100] ppm.

```
> Sulfate.down <- with(EPA.09.Ex.16.1.sulfate.df,
    Sulfate.ppm[Well.type == "Downgradient"])

> enorm(Sulfate.down)

Results of Distribution Parameter Estimation
--------------------------------------------

Assumed Distribution:            Normal

Estimated Parameter(s):          mean = 608.33333
                                 sd   =  18.34848

Estimation Method:               mvue

Data:                            Sulfate.down

Sample Size:                     6

Number NA/NaN/Inf's:             2

> t.test(Sulfate.down, Sulfate.back,
    var.equal = TRUE)$conf.int

[1] 44.33974 99.82693
attr(,"conf.level")
[1] 0.95
```

We can use ciTableMean to look how the confidence interval for the difference between the background and downgradient means in a future study using eight quarterly samples at each well varies with assumed value of the pooled standard deviation and the observed difference between the sample means. Our current estimate of the pooled standard deviation is 24 ppm:

```
> summary(lm(Sulfate.ppm ~ Well.type,
    data = EPA.09.Ex.16.1.sulfate.df))$sigma

[1] 23.57759
```

We see that if this is overly optimistic and in our next study the pooled standard deviation is around 50 ppm, then if the observed difference between the means is 50 ppm, the lower end of the confidence interval for the difference between the two means will include 0, so we may want to increase our sample size.

```
> ciTableMean(n1 = 8, n2 = 8, diff = c(100, 50, 0),
    SD = c(15, 25, 50), digits = 0)

        Diff=100      Diff=50      Diff=0
SD=15 [ 84, 116] [ 34,  66] [-16,  16]
SD=25 [ 73, 127] [ 23,  77] [-27,  27]
SD=50 [ 46, 154] [ -4, 104] [-54,  54]
```

2.8.2   Confidence Interval for a Binomial Proportion

The EnvStats functions `ciTableProp` produces a table similar to Table 1 of Bacchetti (2010) for looking at how the confidence interval for a binomial proportion or the difference between two proportions varies with the value of the estimated proportion(s), given the sample size, confidence level, and method of computing the confidence interval. The function `ciBinomHalfWidth` computes the half-width associated with the confidence interval for the proportion or difference between two proportions, given the sample size, estimated proportion(s), confidence level, and method of computing the confidence interval. The function `ciBinomN` computes the sample size required to achieve a specified half-width, given the estimated proportion(s) and confidence level. The EnvStats function `plotCiBinomDesign` plots the relationships between sample size, half-width, estimated proportion(s), and confidence level.

The data frame `EPA.92c.benzene1.df` contains observations on benzene concentrations (ppb) in groundwater from six background wells sampled monthly for 6 months. Nondetect values are reported as "<2."

```
> EPA.92c.benzene1.df

   Benzene.orig Benzene Censored Month Well
1            <2       2     TRUE     1    1
2            <2       2     TRUE     2    1
...
35           10      10    FALSE     5    6
36           <2       2     TRUE     6    6
```

Of the 36 values, 33 are nondetects. Based on these data, the estimated probability of observing a nondetect is 92 %, and the two-sided 95 % confidence interval for the binomial proportion based on using the normal score approximation with continuity correction is [0.76, 0.98]. The half-width of this interval is 0.11, or 11 % points.

```
> with(EPA.92c.benzene1.df , ebinom(Censored, ci = TRUE))
```

```
Results of Distribution Parameter Estimation
--------------------------------------------

Assumed Distribution:          Binomial

Estimated Parameter(s):        size = 36.0000000
                               prob =  0.9166667

Estimation Method:             mle/mme/mvue for 'prob'

Data:                          Censored

Sample Size:                   36

Confidence Interval for:       prob

Confidence Interval Method:    Score normal approximation
                               (With continuity correction)

Confidence Interval Type:      two-sided

Confidence Level:              95%

Confidence Interval:           LCL = 0.7640884
                               UCL = 0.9782279
```

Suppose we are planning a future study and are interested in the size of the confidence interval. Initially we plan to take 36 samples as in the previous study. Letting the estimated percentage of nondetects vary from 75 % to 95 % shows that the width of the confidence interval varies between about 15 % and 10 % points.

```
> ciBinomHalfWidth(n.or.n1 = 36, p.hat = c(0.75, 0.85, 0.95))

$half.width
[1] 0.14907011 0.12529727 0.09523133

$n
[1] 36 36 36
$p.hat
[1] 0.7500000 0.8611111 0.9444444

$method
[1] "Score normal approximation, with continuity correction"
```

Assuming an estimated proportion of 90 %, if we take only $n = 10$ observations, the half-width of the confidence interval would be about 23 % points:

```
> ciBinomHalfWidth(n.or.n1 = 10, p.hat = 0.9)

$half.width
[1] 0.2268019

$n
[1] 10

$p.hat
[1] 0.9

$method
[1] "Score normal approximation, with continuity correction"
```

Also, if we want the confidence interval to have a half-width of 0.03 (3 % points), we would need to take $n = 319$ observations (a sample size probably not feasible for many environmental studies!).

```
> ciBinomN(half.width = 0.03, p.hat = 0.9)

$n
[1] 319

$p.hat
[1] 0.8996865

$half.width
[1] 0.03466104

$method
[1] "Score normal approximation, with continuity correction""
```

Figure 2.2 displays the half-width of the confidence interval as a function of the sample size for various confidence levels, based on using the score normal approximation with continuity correction to construct the confidence interval.

```
> plotCiBinomDesign(p.hat = 0.9, range.x.var = c(10, 200),
    conf = 0.99, xlim = c(0, 200), ylim = c(0, 0.3),
    main = "")

> plotCiBinomDesign(p.hat = 0.9, range.x.var = c(10, 200),
    conf = 0.95, plot.col = 2, plot.lty = 2, add = TRUE)

> plotCiBinomDesign(p.hat = 0.9, range.x.var = c(10, 200),
    conf = 0.90, plot.col = 4, plot.lty = 3, add = TRUE)
```

```
> legend("topright",
   paste(c("99%", "95%", "90%"), "Confidence"),
   col = c(1, 2, 4), lty = 1:3, lwd = 3, bty = "n")
```
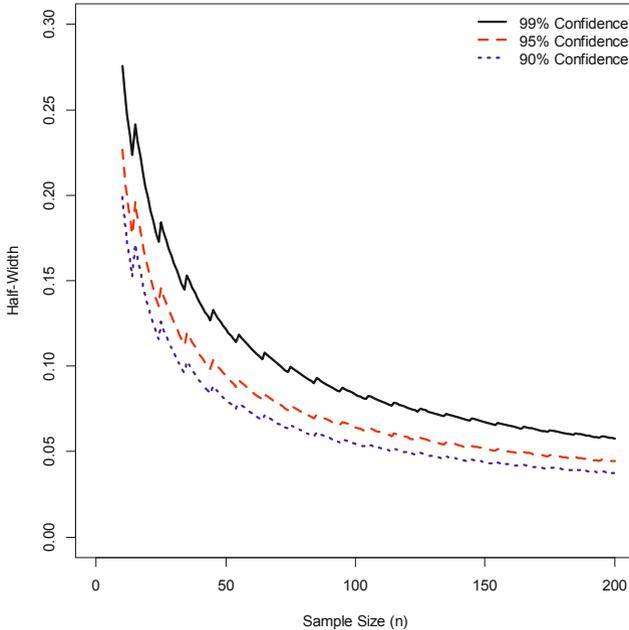

Fig. 2.2 The half-width of the confidence interval for the probability of a nondetect as a function of sample size and confidence level, assuming an estimated nondetect proportion of 90 %

If we are planning a study to compare the proportion of nondetects at a background and downgradient well, we can use `ciTableProp` to look how the confidence interval for the difference between the two proportions using say 36 quarterly samples at each well varies with the observed estimated proportions. Here we'll let the argument `p1.hat` denote the proportion of nondetects observed at the downgradient well and set this equal to 20 %, 40 % and 60 %. The argument `p2.hat.minus.p1.hat` represents the proportion of nondetects at the background well minus the proportion of nondetects at the downgradient well.

```
> ciTableProp(n1 = 36, p1.hat = c(0.2, 0.4, 0.6),
   n2 = 36, p2.hat.minus.p1.hat = c(0.3, 0.15, 0))

                Diff=0.31      Diff=0.14        Diff=0
P1.hat=0.19 [ 0.07, 0.54] [-0.09, 0.37] [-0.18, 0.18]
P1.hat=0.39 [ 0.06, 0.55] [-0.12, 0.39] [-0.23, 0.23]
P1.hat=0.61 [ 0.09, 0.52] [-0.10, 0.38] [-0.23, 0.23]
```

We see that even if the observed difference in the proportion of nondetects is about 15 % points, all of the confidence intervals for the difference between the proportions of nondetects at the two wells contain 0, so if a difference of 15 % points is important to substantiate, we may need to increase our sample sizes.

2.8.3   Nonparametric Confidence Interval for a Percentile

The function `ciNparConfLevel` computes the confidence level associated with a nonparametric confidence interval for the pth quantile (the pth quantile is same as the $100p$th percentile, where $0 \leq p \leq 1$), given the sample size and value of p. The function `ciNparN` computes the sample size required to achieve a specified confidence level, given the value of p. The function `plotCiNparDesign` plots the relationships between sample size, confidence level, and p.

The data frame `EPA.92c.copper2.df` contains copper concentrations (ppb) at three background wells and two compliance wells.

```
> EPA.92c.copper2.df

   Copper.orig Copper Censored Month Well  Well.type
1           <5    5.0     TRUE     1    1 Background
2           <5    5.0     TRUE     2    1 Background
3          7.5    7.5    FALSE     3    1 Background
...
38          <5    5.0     TRUE     6    5 Compliance
39         5.6    5.6    FALSE     7    5 Compliance
40          <5    5.0     TRUE     8    5 Compliance
```

There are eight observations associated with each of the three background wells. Of the 24 observations at the three background wells, 15 are nondetects recorded as "< 5". The other nine observations at the background wells are: 5.4, 5.9, 6.0, 6.1, 6.4, 6.7, 7.5, 8.0, and 9.2. The estimated 95th percentile of copper concentration at the background wells is 7.925 ppb.

```
> Cu.Bkgrd <- with(EPA.92c.copper2.df,
    Copper[Well.type == "Background"]

> eqnpar(Cu.Bkgrd, p = 0.95)

Results of Distribution Parameter Estimation
-------------------------------------------

Assumed Distribution:            None

Estimated Quantile(s):           95'th %ile = 7.925

Quantile Estimation Method:      Nonparametric
```

```
Data:                              Cu.Bkgrd
```

```
Sample Size:                       24
```

If we use the largest observed value of 9.2 as the upper confidence limit of the 95th percentile of the copper concentration, the associated confidence level is 71 %.

```
> ciNparConfLevel(n = 24, p = 0.95, ci.type = "upper")
```

```
[1] 0.708011
```

If only four observations had been taken at each well for a total sample size of $n = 12$, the associated confidence level would have been 46 %.

```
> ciNparConfLevel(n = 12, p = 0.95, ci.type = "upper")
```

```
[1] 0.4596399
```

If we want to construct a nonparametric confidence interval for the 95th percentile of copper concentration with an associated confidence level of at least 95 %, we would need $n = 59$ observations (about 20 observations at each background well).

```
> ciNparN(p = 0.95, ci.type = "upper")
```
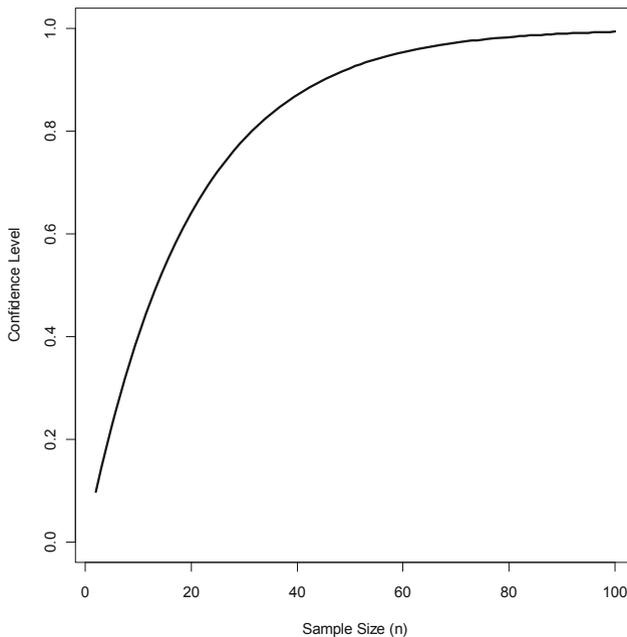
```
[1] 59
```


Fig. 2.3 Confidence level for the one-sided upper nonparametric confidence interval for the 95th percentile versus sample size, using the maximum value as the upper confidence limit

Figure 2.3 displays the confidence level of the one-sided upper confidence interval for the 95th percentile as a function of the sample size. To create this plot, type this command:

```
> plotCiNparDesign(p = 0.95, ci.type = "upper",
   range.x.var = c(2, 100), ylim = c(0, 1))
```

2.9   Sample Size for Prediction Intervals

Table 2.2 lists the functions available in EnvStats for computing required sample sizes, half-widths, and confidence levels associated with a prediction interval. For the normal distribution, you can compute the half-width of the prediction interval given the user-specified sample size, compute the required sample size given the user-specified half-width, and plot the relationship between sample size and half-width. For a nonparametric prediction interval, you can compute the required sample size for a specified confidence level, compute the confidence level associated with a given sample size, and plot the relationship between sample size and confidence level.

Distribution	Function	Output
Normal	predIntNormHalfWidth	Half-width of prediction interval for normal distribution
	predIntNormN	Required sample size for specified half-width of prediction interval for normal distribution
	plotPredIntNormDesign	Plots for sampling design based on prediction interval for normal distribution
Nonparametric	predIntNparConfLevel predIntNparSimultaneousConfLevel	Confidence level of prediction interval, given sample size
	predIntNparN predIntNparSimultaneousN	Required sample size for specified confidence level of prediction interval
	plotPredIntNparDesign plotPredIntNparSimultaneousDesign	Plots for sampling design based on prediction interval

Table 2.2 Sample size functions for prediction intervals

2.9.1   Prediction Interval for a Normal Distribution

The function predIntNormHalfWidth computes the half-width associated with the prediction interval for a normal distribution, given the sample size, number of future observations the prediction interval should contain, estimated

standard deviation, and confidence level. The function `predIntNormN` computes the sample size required to achieve a specified half-width, given the number of future observations, estimated standard deviation, and confidence level. The function `plotPredIntNormDesign` plots the relationships between sample size, number of future observations, half-width, estimated standard deviation, and confidence level.

The data frame `EPA.92c.arsenic3.df` contains arsenic concentrations (ppb) collected quarterly for 3 years at a background well and quarterly for 2 years at a compliance well.

```
> EPA.92c.arsenic3.df

   Arsenic Year  Well.type
1     12.6    1 Background
2     30.8    1 Background
...
19     2.6    5 Compliance
20    51.9    5 Compliance
```

The estimated mean and standard deviation for the background well are 28 and 17 ppb, respectively. The exact two-sided 95 % prediction limit for the next $k = 4$ future observations is $[-25, 80]$, which has a half-width of 52.5 ppb and includes values less than 0, which are not possible to observe.

```
> As.Bkgrd <- with(EPA.92c.arsenic3.df,
    Arsenic[Well.type == "Background"])

> predIntNorm(As.Bkgrd, k = 4, method = "exact")

Results of Distribution Parameter Estimation
--------------------------------------------

Assumed Distribution:            Normal

Estimated Parameter(s):          mean = 27.51667
                                 sd   = 17.10119

Estimation Method:               mvue

Data:                            As.Bkgrd

Sample Size:                     12

Prediction Interval Method:      exact

Prediction Interval Type:        two-sided
```

```
Confidence Level:                95%

Number of Future Observations:   4

Prediction Interval:             LPL = -24.65682
                                 UPL =  79.69015
```

In fact, given an assumed standard deviation of $s = 17$, the smallest half width you can achieve for a prediction interval for the next $k = 4$ future observations is 42 ppb, based on an infinite sample size. ***Unlike a confidence interval, the half-width of a prediction interval does not approach 0 as the sample size increases.*** Figure 2.4 shows a plot of sample size versus half-width for a 95 % prediction interval for a normal distribution for various values of k (the number of future observations), assuming a standard deviation of $s = 17$.
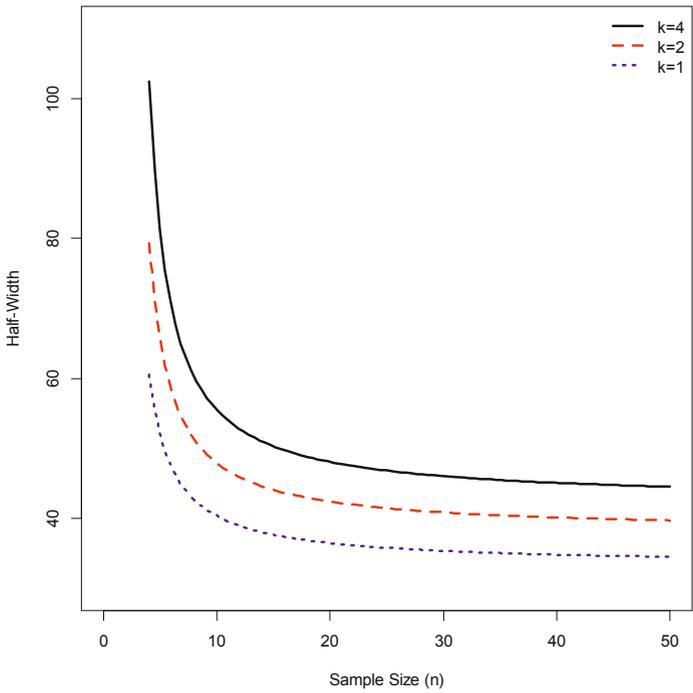

Fig. 2.4 The half-width of a 95 % prediction interval for arsenic concentrations (ppb) as a function of sample size and number of future observations (k), assuming a standard deviation of 17 ppb

Type these commands to create the plot:

```
> plotPredIntNormDesign(range.x.var = c(4, 50), k = 4,
   sigma.hat = 17, xlim = c(0, 50), ylim = c(30, 110),
   main = "")

> plotPredIntNormDesign(range.x.var = c(4, 50), k = 2,
   sigma.hat = 17, plot.col = 2, plot.lty = 2, add = TRUE)

> plotPredIntNormDesign(range.x.var = c(4, 50), k = 1,
   sigma.hat = 17, plot.col = 4, plot.lty = 3, add = TRUE)

> legend("topright", c("k=4", "k=2", "k=1"),
   col = c(1, 2, 4), lty = 1:3, lwd = 3, bty = "n")
```

2.9.2   Nonparametric Prediction Interval

The function `predIntNparConfLevel` computes the confidence level associated with a nonparametric prediction interval, given the minimum number of future observations the interval should contain (k), the number of future observations (m), and sample size. The function `predIntNparN` computes the sample size required to achieve a specified confidence level, given the number of future observations. The function `plotPredIntNparDesign` plots the relationships between sample size, confidence level, and number of future observations.

Table 2.3 shows the required sample size for a two-sided nonparametric prediction interval for the next m future observations (assuming $k = m$) for various values of m and required confidence levels, assuming we are using the minimum and maximum values as the prediction limits. The values for the table are generated using this command:

```
> predIntNparN(m = rep(c(1, 5, 10), 2),
   conf.level = rep(c(0.9, 0.95), each = 3))
```

Confidence level (%)	# future observations (*m*)	Required sample size (*n*)
90	1	19
	5	93
	10	186
95	1	39
	5	193
	10	386

Table 2.3 Required sample sizes for a two-sided nonparametric prediction interval, using the minimum and maximum values as the prediction limits

Figure 2.5 displays the confidence level of a two-sided nonparametric prediction interval as a function of sample size for various values of m, using the minimum and maximum values as the prediction limits. To create this figure, type these commands:

```
> plotPredIntNparDesign(range.x.var = c(2, 100), k = 1,
    m = 1, xlim = c(0, 100), ylim = c(0, 1), main = "")

> plotPredIntNparDesign(range.x.var = c(2, 100), k = 5,
    m = 5, plot.col = 2, plot.lty = 2, add = TRUE)

> plotPredIntNparDesign(range.x.var = c(2, 100),
    k = 10, m = 10, plot.col = 4, plot.lty = 3, add = TRUE)

> legend("bottomright", c("m=  1", "m=  5", "m=10"),
    col = c(1, 2, 4), lty = 1:3, lwd = 3)
```
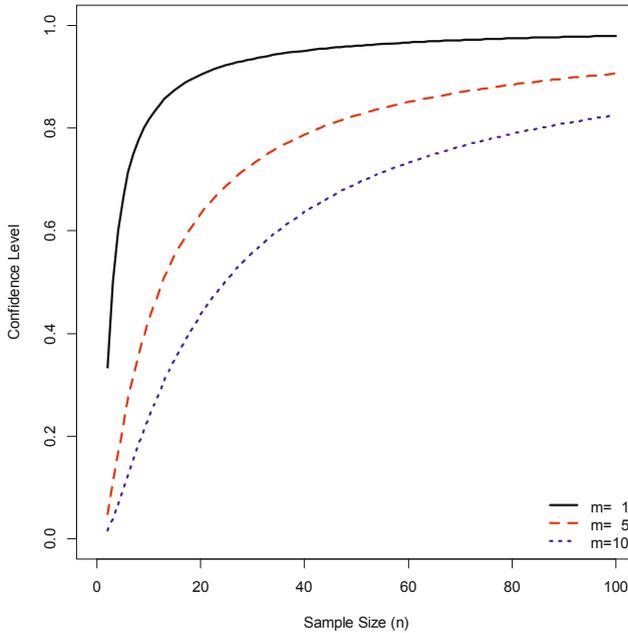

Fig. 2.5 The confidence level of a two-sided nonparametric prediction interval as a function of sample size, for various values of the number of future observations (*m*)

2.10 Sample Size for Tolerance Intervals

Table 2.4 lists the functions available in EnvStats for computing required sample sizes, half-widths, coverage, and confidence levels associated with a tolerance interval. For the normal distribution, you can compute the half-width of the tolerance interval given the user-specified sample size and coverage, compute the required sample size given the user-specified half-width and coverage, and plot the relationship between sample size, half-width, and coverage. For a non-parametric prediction interval, you can compute the required sample size for a specified confidence level and coverage, compute the confidence level associated

with a given sample size and coverage, compute the coverage associated with a given sample size and confidence level, and plot the relationship between sample size, confidence level, and coverage.

Distribution	Function	Output
Normal	`tolIntNormHalfWidth`	Half-width of tolerance interval for normal distribution
	`tolIntNormN`	Required sample size for specified half-width of tolerance interval for normal distribution
	`plotTolIntNormDesign`	Plots for sampling design based on tolerance interval for normal distribution
Nonparametric	`tolIntNparConfLevel`	Confidence level of tolerance interval, given the coverage and sample size
	`tolIntNparCoverage`	Coverage of tolerance interval, given confidence level and sample size
	`tolIntNparN`	Required sample size for specified confidence level and coverage of a tolerance interval
	`plotTolIntNparDesign`	Plots for sampling design based on a tolerance interval

Table 2.4 Sample size functions for tolerance intervals

2.10.1 Tolerance Interval for a Normal Distribution

The function `tolIntNormHalfWidth` computes the half-width associated with a tolerance interval for a normal distribution, given the sample size, coverage, estimated standard deviation, and confidence level. The function `tolIntNormN` computes the sample size required to achieve a specified half-width, given the coverage, estimated standard deviation, and confidence level. The function `plotTolIntNormDesign` plots the relationships between sample size, half-width, coverage, estimated standard deviation, and confidence level.

Again using the data frame `EPA.92c.arsenic3.df` containing arsenic concentrations, we saw in Sect. 2.9.1 that the estimated mean and standard deviation for the background well are 28 and 17 ppb, respectively, based on a sample size of $n = 12$ quarterly samples. The two-sided β-content tolerance limit with 95 % coverage and associated confidence level of 99 % is [−39, 94], which has a half-width of 66.5 ppb and includes values less than 0, which are not possible to observe.

```
> tolIntNorm(As.Bkgrd, coverage = 0.95,
    cov.type = "content", conf.level = 0.99)
```

```
Results of Distribution Parameter Estimation
--------------------------------------------


Assumed Distribution:           Normal

Estimated Parameter(s):         mean = 27.51667
                                sd   = 17.10119

Estimation Method:              mvue

Data:                           As.Bkgrd

Sample Size:                    12

Tolerance Interval Coverage:    95%

Coverage Type:                  content

Tolerance Interval Method:      Wald-Wolfowitz Approx

Tolerance Interval Type:        two-sided

Confidence Level:               99%

Tolerance Interval:             LTL = -38.66445
                                UTL =  93.69778
```

In fact, given an assumed standard deviation of $s = 17$, the smallest half width you can achieve for a tolerance interval with 95 % coverage and 99 % confidence is 33 ppb, based on an infinite sample size. ***Unlike a confidence interval, the half-width of a tolerance interval does not approach 0 as the sample size increases.*** Figure 2.6 shows a plot of sample size versus half-width for a β-content tolerance interval for a normal distribution with confidence level 99 % for various values of coverage, assuming a standard deviation of $s = 17$. It was created with these commands:

```
> plotTolIntNormDesign(range.x.var = c(5, 50),
   sigma.hat = 17, coverage = 0.99, conf = 0.99,
   xlim = c(0, 50), ylim = c(0, 200), main = "")

> plotTolIntNormDesign(range.x.var = c(5, 50),
   sigma.hat = 17, coverage = 0.95, conf = 0.99,
   plot.col = 2, plot.lty = 2, add = TRUE)
```

```
> plotTolIntNormDesign(range.x.var = c(5, 50),
   sigma.hat = 17, coverage = 0.90, conf = 0.99,
   plot.col = 4, plot.lty = 3, add = TRUE)

> legend("topright",
   paste(c("99%", "95%", "90%"), "Coverage"),
   col = c(1, 2, 4), lty=1:3, lwd=3, bty = "n")
```
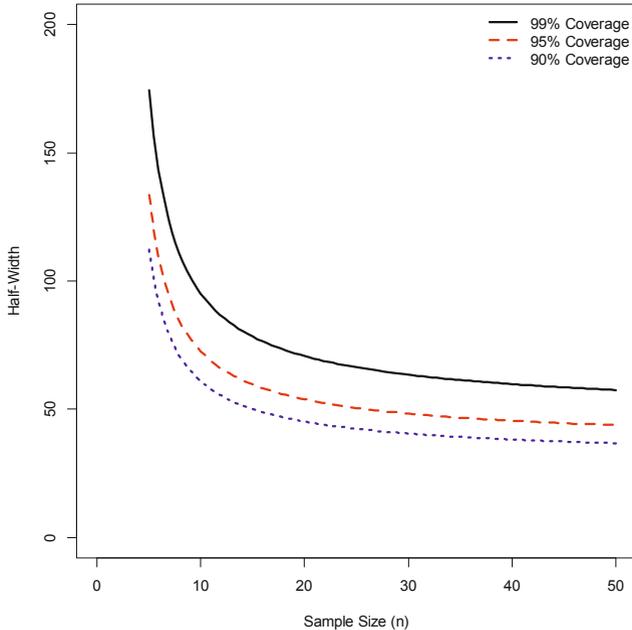

Fig. 2.6 The half-width of a tolerance interval for arsenic concentrations (ppb) as a function of sample size and coverage, assuming a standard deviation of 17 ppb

2.10.2 Nonparametric Tolerance Interval

The function `tolIntNparConfLevel` computes the confidence level associated with a nonparametric tolerance interval, given the coverage and sample size. The function `tolIntNparCoverage` computes the coverage associated with the tolerance interval, given the confidence level and sample size. The function `tolIntNparN` computes the sample size required to achieve a specified confidence level, for a given coverage. The function `plotTolIntNparDesign` plots the relationships between sample size, confidence level, and coverage.

Table 2.5 shows the required sample size for a two-sided nonparametric tolerance interval for various values of coverage and required confidence levels, assuming we are using the minimum and maximum values as the tolerance limits. The values for the table are generated using this command:

```
> tolIntNparN(coverage = rep(c(0.8, 0.9, 0.95), 2),
    conf.level = rep(c(0.9, 0.95), each = 3))
```

Confidence level (%)	Coverage (%)	Required sample size (*n*)
90	80	18
	90	38
	95	77
95	80	22
	90	46
	95	93

Table 2.5 Required sample sizes for a two-sided nonparametric tolerance interval, using the minimum and maximum values as the tolerance limits

Figure 2.7 displays the confidence level of a two-sided nonparametric tolerance interval as a function of sample size for various values of coverage, using the minimum and maximum values as the tolerance limits.
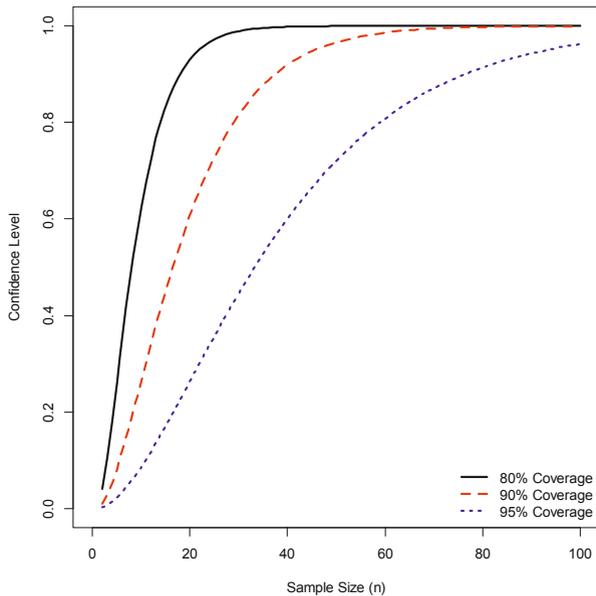

Fig. 2.7 The confidence level of a two-sided nonparametric tolerance interval as a function of sample size, for various values of coverage

To create this plot, type these commands:

```
> plotTolIntNparDesign(range.x.var = c(2, 100),
    coverage = 0.8, xlim = c(0, 100), ylim = c(0, 1),
    main = "")
```

```
> plotTolIntNparDesign(range.x.var = c(2, 100),
    coverage = 0.90, plot.col = 2, plot.lty = 2, add = TRUE)

> plotTolIntNparDesign(range.x.var = c(2, 100),
    coverage = 0.95, plot.col = 4, plot.lty = 3, add = TRUE)

> legend("bottomright",
    paste(c("80%", "90%", "95%"), "Coverage"),
    col = c(1, 2, 4), lty = 1:3, lwd = 3, bty = "n")
```

2.11 Sample Size and Power for Hypothesis Tests

Table 2.6 lists the functions available in EnvStats for computing required sample sizes, powers, and minimal detectable differences associated with several different hypothesis tests. In this section, we will illustrate how to use EnvStats functions to explore the relationship between sample size and power for testing the mean of a normal distribution, testing a binomial proportion, and using simultaneous prediction limits with retesting. See Millard et al. (2014) and the help files for the functions listed in Table 2.6 for more examples of exploring the relationship between sample size and power for other kinds of hypothesis tests.

Test	Function	Output
Student's t-test	tTestPower	Power of t-test
	tTestAlpha	Type I error of t-test
	tTestN	Required sample size for specified power of t-test
	tTestScaledMdd	Required scaled minimal detectable difference (δ/σ) for specified power of t-test
	plotTTestDesign	Plots for sampling design based on t-test
Student's t-test, lognormal distribution	tTestLnormAltPower	Power of one- or two-sample t-test assuming lognormal distribution
	tTestLnormAltN	Required sample size for specified power for one- or two-sample t-test assuming lognormal distribution
	tTestLnormAltRatioOfMeans	Required ratio of means for specified power for one- or two-sample t-test assuming lognormal distribution
	plotTTestLnormAltDesign	Plots for sampling design based on one- or two-sample t-test assuming lognormal distribution

Table 2.6 Sample size and power functions for hypothesis tests

Test	Function	Output
ANOVA F-test	`aovPower`	Power of F-test for one-way ANOVA
	`aovN`	Required sample size for specified power of F-test for one-way ANOVA
	`plotAovDesign`	Plots for sampling design based on F-test for one-way ANOVA
Proportion test, binomial distribution	`propTestPower`	Power of one- or two-sample proportion test
	`propTestN`	Required sample size for specified power for one- or two-sample proportion test
	`propTestMdd`	Required minimal detectable difference for specified power for one- or two-sample proportion test
	`plotPropTestDesign`	Plots for sampling design based on one- or two-sample proportion test
Linear trend t-test	`linearTrendTestPower`	Power of test for non-zero slope
	`linearTrendTestN`	Required sample size for specified power for test of non-zero slope
	`linearTrendTestScaledMds`	Required minimal detectable slope for specified power for test of non-zero slope
	`plotLinearTrendTestDesign`	Plots for sampling design based on test for non-zero slope
Prediction interval, normal distribution	`predIntNormTestPower`	Power of test based on prediction interval for normal distribution
	`plotPredIntNormTestPowerCurve`	Power curve for test based on prediction interval for normal distribution
	`predIntNormSimultaneousTestPower`	Power of test based on simultaneous prediction interval for normal distribution
	`plotPredIntNormSimultaneousTestPower Curve`	Power curve for test based on simultaneous prediction interval for normal distribution

Table 2.6 (continued). Sample size and power functions for hypothesis tests

Test	Function	Output
Prediction interval, lognormal distribution	`predIntLnormAltTestPower`	Power of test based on prediction interval for lognormal distribution
	`plotPredIntLnormAltTestPowerCurve`	Power curve for test based on prediction interval for lognormal distribution
	`predIntLnormAltSimultaneousTestPower`	Power of test based on simultaneous prediction interval for lognormal distribution
	`plotPredIntLnormAltSimultaneousTest PowerCurve`	Power curve for test based on simultaneous prediction interval for lognormal distribution
Prediction interval, nonparametric	`predIntNparSimultaneousTestPower`	Power of test based on nonparametric simultaneous prediction interval
	`plotPredIntNparSimultaneousTestPower Curve`	Power curve for test based on nonparametric simultaneous prediction interval

Table 2.6 (continued). Sample size and power functions for hypothesis tests

2.11.1 Testing the Mean of a Normal Distribution

Power and sample size calculations based on Student's t-test involve four quantities:

1. The fixed type I error (also called the α-level).
2. The desired power of the test.
3. The sample size(s).
4. The scaled minimal detectable difference (scaled MDD), also often called the effect size. For the one-sample case, the scaled MDD is the difference between the true population mean and the population mean hypothesized under the null hypothesis, divided by the population standard deviation. For the two-sample case, the scaled MDD is the difference between the true population means for the two groups minus the difference between the population means hypothesized for the two groups under the null hypothesis, divided by the population standard deviation (the standard deviation is assumed to be the same for both groups). Because the term "effect size" is sometimes used to denote simply the difference between the means, we always use the term scaled MDD to denote this quantity.

The EnvStats function tTestPower computes the power associated with the Student's t-test to perform a hypothesis test for the mean of a normal distribution or the difference between two means, given the sample size, scaled MDD, and α-level. The function tTestAlpha computes the α-level given the power, sample size, and scaled MDD. The function tTestN computes the sample size required to achieve a specified power, given the scaled MDD and α-level. The function tTestScaledMdd computes the scaled MDD associated with user-specified values of power, sample size, and α-level. The function plotTTestDesign plots the relationships between sample size, power, scaled MDD, and α-level.

The guidance document *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Unified Guidance* (USEPA 2009) contains an example on pages 22–6 to 22–8 that uses vinyl chloride (ppb) concentrations at two different compliance wells. There are 4 years of quarterly observations at each of the two wells. The first year of data corresponds to the background period and the subsequent 3 years correspond to the compliance period. The data in this example are stored in the data frame EPA.09.Ex.22.1.VC.df.

```
> EPA.09.Ex.22.1.VC.df

   Year Quarter       Period Well VC.ppb
1     1           1 Background GW-1    6.3
2     1           2 Background GW-1    9.5
...
31    4           3 Compliance GW-2    7.5
32    4           4 Compliance GW-2    9.7
```

The groundwater protection standard (GWPS) has been set to 5 ppb. During compliance monitoring, we want to test the null hypothesis that the mean vinyl chloride concentration is less than or equal to 5 ppb versus the alternative that it is greater than 5 ppb based on using 1 year of data (i.e., four quarterly observations). We want to have 80 % power of detecting an increase of twice the GWPS (i.e., detecting a true mean vinyl chloride concentration of 10 ppb, a difference of 5 ppb between the assumed mean under the null hypothesis and the mean under the alternative hypothesis).

In this example, first we'll use the first year (background period) of monitoring to estimate the standard deviation of vinyl chloride measurements to determine the required α-level. Then we'll see how changing the α-level and sample size affects the power.

For the first year (background period) of monitoring, the observed means and standard deviations are 8.9 and 2.4 ppb for Well 1, and 7.4 and 3.9 ppb for Well 2, and the pooled estimate of standard deviation (assuming the standard deviation is the same at the two wells) is 3.2 ppb.

```
> summaryStats(VC.ppb ~ Well, data = EPA.09.Ex.22.1.VC.df,
    subset = Period == "Background", digits = 1)
```

```
     N Mean  SD Median Min  Max
GW-1 4  8.9 2.4    8.8 6.3 11.9
GW-2 4  7.4 3.9    7.4 3.0 12.0
```

```
> VC.lm.fit <- lm(VC.ppb ~ Well,
   data = EPA.09.Ex.22.1.VC.df,
   subset = Period == "Background")
```

```
> summary(VC.lm.fit)$sigma
```

```
[1] 3.200976
```

However, if we compute a two-sided 95 % confidence interval for the true standard deviation based on the background period data, we see that it may be as high as about 6 ppb:

```
> sqrt(enorm(VC.lm.fit$residuals, ci = TRUE,
   ci.param = "variance")$interval$limits)
```

```
     LCL      UCL
1.959408 6.031586
```

Assuming population standard deviations of 3.2 and 6 ppb, basing the one-sample t-test on $n = 4$ observations, we need to set the type I error level to 0.057 or 0.23 respectively in order to achieve 80 % power of detecting a true concentration of vinyl chloride of 10 ppb:

```
> tTestAlpha(n.or.n1 = 4, delta.over.sigma = 5 / c(3, 6),
   power = 0.8, sample.type = "one.sample",
   alternative = "greater")
```

```
[1] 0.05763283 0.22936065
```

If we set the significance level to 1 % and assume a standard deviation of 3.2 ppb, we can see how the power varies with sample size:

```
> tTestPower(n.or.n1 = c(4, 8, 12),
   delta.over.sigma = 5 / 3.2, alpha = 0.01,
   alternative = "greater")
```

```
[1] 0.3173891 0.8839337 0.9911121
```

If we set the significance level to 1 %, the desired power to 90 %, and assume a standard deviation of 6 ppb, we would need a sample size of at least $n = 22$ to detect an average vinyl chloride concentration that is 5 ppb above the GWPS:

```
> tTestN(delta.over.sigma = 5 / 6, alpha = 0.01,
   power = 0.9, alternative = "greater")
```
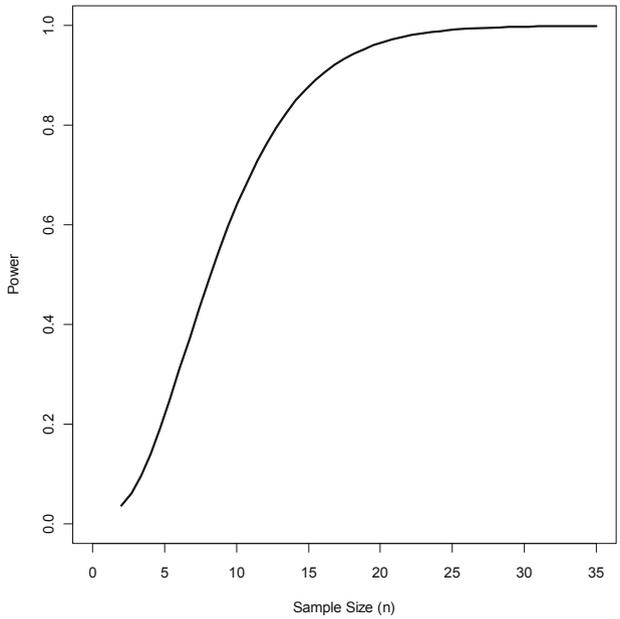
```
[1] 22
```

Fig. 2.8 Power versus sample size for a one-sample t-test with a significance level of 1 %, assuming a scaled minimal detectable difference of $\delta/\sigma = 1$
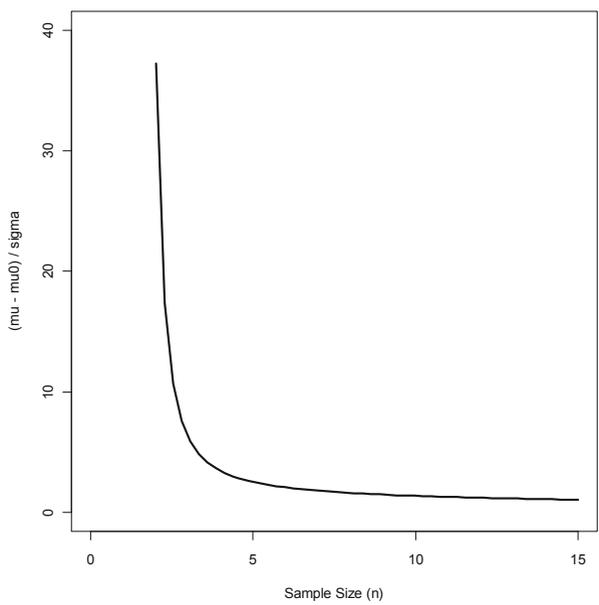

Fig. 2.9 Scaled minimal detectable difference versus sample size for a one-sample t-test with a significance level of 1 %, assuming a power of 90 %

Figure 2.8 plots power as a function of sample size for a significance level of 1 %, assuming a scaled minimal detectable difference of 1. Use this command to produce it:

```
> plotTTestDesign(alpha = 0.01, delta.over.sigma = 1,
    range.x.var = c(2, 35), xlim = c(0, 35), ylim = c(0, 1),
    alternative="greater", approx = FALSE)
```

Figure 2.9 plots the scaled minimal detectable difference as a function of sample size for a significance level of 1 %, assuming a power of 90 %.

```
> plotTTestDesign(y.var = "delta.over.sigma",
    alpha = 0.01, power = 0.9, range.x.var = c(2, 15),
    xlim = c(0, 15), ylim = c(0, 40), alternative="greater",
    approx. = FALSE)
```

2.11.2 Testing a Binomial Proportion

The guidance document *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Unified Guidance* (USEPA 2009) contains an example on page 22–20 that involves determining whether more than 10 % of chlorine gas containers are stored at pressures above a compliance limit. We want to test the one-sided null hypothesis that 10 % or fewer of the containers are stored at pressures greater than the compliance limit versus the alternative that more than 10 % are stored at pressures greater than the compliance limit. We want to have at least 90 % power of detecting a true proportion of 30 % or greater, using a 5 % Type I error level. The example in the guidance document uses the normal approximation to the binomial distribution (without a continuity correction) to determine we need to check 30 containers:

```
> propTestN(p.or.p1 = 0.3, p0.or.p2 = 0.1, alpha = 0.05,
    power = 0.9, sample.type = "one.sample",
    alternative = "greater", approx = TRUE, round.up = TRUE)

[1] 30
```

However, a quick simulation shows that the true Type I error of the hypothesis test based on the normal approximation without using the continuity correction is inflated above 5 % and is really about 7 %:

```
> set.seed(274)

> N <- 10000

> Reject.vec <- logical(N)
```

```
> for(i in 1:N) {
    Reject.vec[i] <- prop.test(
    x = rbinom(n = 1, size = 30, prob = 0.1), n = 30, p = 0.1,
    alternative = "greater", correct = FALSE)$p.value < 0.05
    }

> mean(Reject.vec)

[1] 0.071
```

The 95 % confidence interval for the true Type I error level based on our simulation of 10,000 trials is [6.6 %, 7.6 %]:

```
> binom.test(x = sum(Reject.vec), n = length(Reject.vec),
    p = 0.05)$conf.int

[1] 0.06604181 0.07620974
```

We could try basing our sample size calculation on the test based on the normal approximation *with* the continuity correction, but simulation shows that the continuity correction makes the true Type I error rate about 2.5 % with a 95 % confidence interval of [2.2 %, 2.8 %] for the true Type I error rate:

```
> set.seed(538)

> N <- 10000

> Reject.vec <- logical(N)

> for(i in 1:N) {
    Reject.vec[i] <- prop.test(
    x = rbinom(n = 1, size = 30, prob = 0.1), n = 30, p = 0.1,
    alternative = "greater", correct = TRUE)$p.value < 0.05
    }

> mean(Reject.vec)

[1] 0.0248

> binom.test(x = sum(Reject.vec), n = length(Reject.vec),
    p = 0.05)$conf.int

[1] 0.02184098 0.02803999
```

If we base our sample size calculation on the exact binomial test instead of the test based on the normal approximation, we can set how much the actual Type I error rate can deviate from what we specify by using the tol.alpha argument to propTestN. By default, tol.alpha is equal to 10 % of the value of alpha, so in this case tol.alpha=0.005 which means the smallest the true Type I error rate can be is 0.045, and the required sample size is 34:

```
> propTestN(p.or.p1 = 0.3, p0.or.p2 = 0.1, alpha = 0.05,
   power = 0.9, sample.type = "one.sample",
   alternative = "greater", approx = FALSE, round.up = TRUE)
$n
[1] 34

$power
[1] 0.9214717

$alpha
[1] 0.04814433

$q.critical.upper
[1] 6
```

If we allow the true Type I error to deviate by 0.01, the required sample size is 33:

```
> propTestN(p.or.p1 = 0.3, p0.or.p2 = 0.1, alpha = 0.05,
   power = 0.9, sample.type = "one.sample", tol.alpha = 0.01,
   alternative = "greater", approx = FALSE, round.up = TRUE)
$n
[1] 33

$power
[1] 0.9055545

$alpha
[1] 0.04170385

$q.critical.upper
[1] 6
```

2.11.3 Testing Multiple Wells for Compliance with Simultaneous Prediction Intervals

The guidance document *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Unified Guidance* (USEPA 2009) contains an example on page 19–23 that involves monitoring $n_w = 100$ compliance wells at a large facility with minimal natural spatial variation every 6 months for $n_c = 20$ separate chemicals. There are $n = 25$ background measurements for each chemical to use to create simultaneous prediction intervals. We would like to determine which kind of resampling plan based on normal distribution simultaneous prediction intervals to use (1-of-m, 1-of-m based on means, or Modified California) in order to have adequate power of detecting an increase in chemical concentration at any of the 100 wells while at the same time maintaining a site-wide false positive rate

(SWFPR) of 10 % per year over all 4,000 (100 wells × 20 chemicals × semi-annual sampling) comparisons.

The EnvStats functions for computing power based on simultaneous prediction limits include the argument r that is the number of future sampling occasions ($r=2$ in this case because we are performing semi-annual sampling), so to compute the individual test Type I error level α_{test} (and thus the individual test confidence level), we only need to worry about the number of wells (100) and the number of constituents (20): $\alpha_{test} = 1-(1-\alpha)^{1/(nw \times nc)}$. The individual confidence level is simply $1-\alpha_{test}$. Plugging in 0.1 for α, 100 for nw, and 20 for nc yields an individual test confidence level of $1-\alpha_{test} = 0.9999473$.

```
> nc <- 20

> nw <- 100

> conf.level <- (1 - 0.1)^(1 / (nc * nw))

> conf.level

[1] 0.9999473
```

Now we can compute the power of any particular sampling strategy using the EnvStats function predIntNormSimultaneousTestPower. For example, to compute the power of detecting an increase of three standard deviations in concentration using the prediction interval based on the "1-of-2" resampling rule, type this command:

```
> predIntNormSimultaneousTestPower(n = 25, k = 1,
    m = 2, r = 2, rule = "k.of.m", delta.over.sigma = 3,
    pi.type = "upper", conf.level = conf.level)

[1] 0.3900202
```

The following commands will reproduce the table shown in Step 2 on page 19–23 of the EPA guidance document:

```
> rule.vec <- c(rep("k.of.m", 3), "Modified.CA",
    rep("k.of.m", 3))

> m.vec <- c(2, 3, 4, 4, 1, 2, 1)

> n.mean.vec <- c(rep(1, 4), 2, 2, 3)

> n.scenarios <- length(rule.vec)

> K.vec <- numeric(n.scenarios)

> Power.vec <- numeric(n.scenarios)

> K.vec <- predIntNormSimultaneousK(n = n, k = 1, m = m.vec,
    n.mean = n.mean.vec, r = r, rule = rule.vec,
    pi.type = "upper", conf.level = conf.level)
```

```
> Power.vec <- predIntNormSimultaneousTestPower(n = n, k = 1,
   m = m.vec, n.mean = n.mean.vec, r = r, rule = rule.vec,
   delta.over.sigma = 3, pi.type = "upper",
   conf.level = conf.level)

> data.frame(Rule = rule.vec, k = rep(1, n.scenarios),
   m = m.vec, N.Mean = n.mean.vec, K = round(K.vec, 2),
   Power = round(Power.vec, 2),
   Total.Samples = m.vec * n.mean.vec)

        Rule k m N.Mean    K Power Total.Samples
1      k.of.m 1 2      1 3.16  0.39             2
2      k.of.m 1 3      1 2.33  0.65             3
3      k.of.m 1 4      1 1.83  0.81             4
4 Modified.CA 1 4      1 2.57  0.71             4
5      k.of.m 1 1      2 3.62  0.41             2
6      k.of.m 1 2      2 2.33  0.85             4
7      k.of.m 1 1      3 2.99  0.71             3
```

The above table shows the κ-multipliers for each prediction interval, along with the power of detecting a change in concentration of three standard deviations at any of the 100 wells during the course of a year, for each of the sampling strategies considered. The last three rows of the table correspond to sampling strategies that involve using the mean of two or three observations.

Figure 2.10 shows the power curves for the first four sampling strategies. It was created with these commands:

```
> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
   k = 1, m = 4, r = 2, rule="k.of.m", pi.type = "upper",
   conf.level = conf.level,
   xlab = "SD Units Above Background", main = "")

> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
   k = 1, m = 3, r = 2, rule="k.of.m", pi.type = "upper",
   conf.level = conf.level, add = TRUE, plot.col = 2,
   plot.lty = 2)

> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
   k = 1, m = 2, r = 2, rule="k.of.m", pi.type = "upper",
   conf.level = conf.level, add = TRUE, plot.col = 3,
   plot.lty = 3)

> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
   r = 2, rule="Modified.CA", pi.type = "upper",
   conf.level = conf.level, add = TRUE, plot.col = 4,
   plot.lty = 4)
```

```
> legend(0, 1, c("1-of-4", "Modified CA", "1-of-3",
    "1-of-2"), col = c(1, 4, 2, 3), lty = c(1, 4, 2, 3),
    lwd = 2, bty = "n")
```
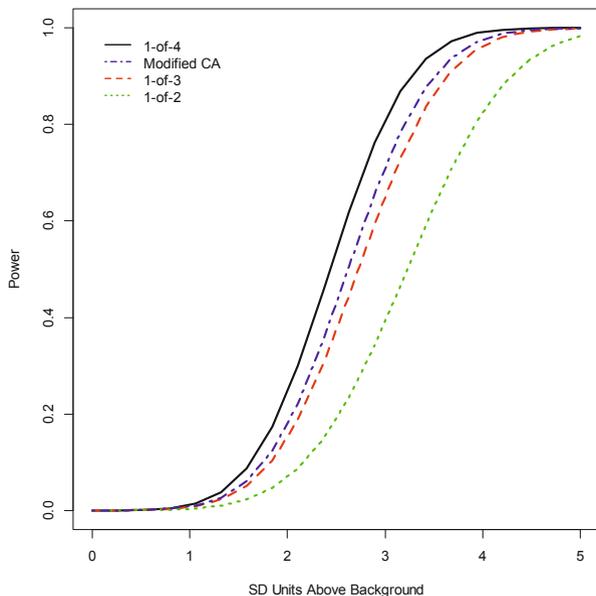

Fig. 2.10 Power versus increase in concentration for various testing strategies based on simultaneous prediction limits, with 100 wells, 20 chemicals, semi-annual sampling, and an annual SWFPR of 10 %

Figure 2.11 shows the power curves for the last three sampling strategies. It was created with these commands:

```
> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
    k = 1, m = 2, n.mean = 2, r = 2, rule="k.of.m",
    pi.type = "upper", conf.level = conf.level,
    xlab = "SD Units Above Background", main = "")

> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
    k = 1, m = 1, n.mean = 2, r = 2, rule="k.of.m",
    pi.type = "upper", conf.level = conf.level, add = TRUE,
    plot.col = 2, plot.lty = 2)

> plotPredIntNormSimultaneousTestPowerCurve(n = 25,
    k = 1, m = 1, n.mean = 3, r = 2, rule="k.of.m",
    pi.type = "upper", conf.level = conf.level, add = TRUE,
    plot.col = 3, plot.lty = 3)
```

```
> legend(0, 1, c("1-of-2, Order 2", "1-of-1, Order 3",
    "1-of-1, Order 2"), col = c(1, 3, 2), lty = c(1, 3, 2),
    lwd = 2, bty="n")
```
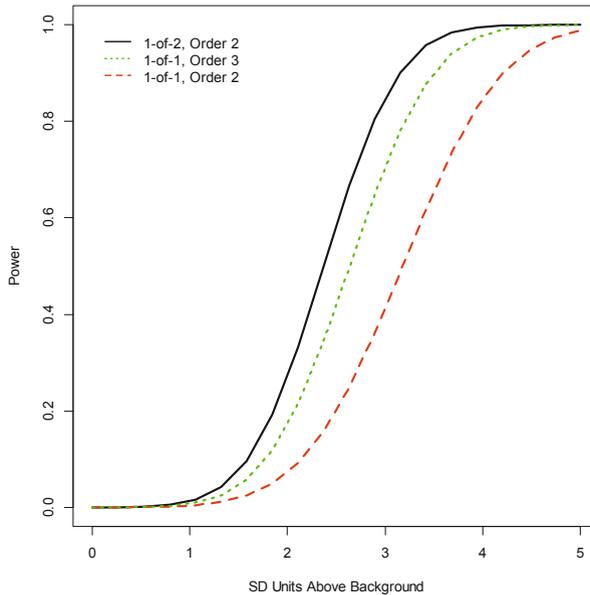

Fig. 2.11 Power versus increase in concentration for various testing strategies based on simultaneous prediction limits using the mean of two measures, with 100 wells, 20 chemicals, semi-annual sampling, and an annual SWFPR of 10 %

2.12 Summary

- The first and most important step of any environmental study is to design the sampling program.
- *Probability sampling* or *random sampling* involves using a random mechanism to select samples from the population. All statistical methods used to quantify uncertainty assume some form of random sampling was used to obtain the sample.
- The Data Quality Objectives (DQO) process is a systematic planning tool based on the scientific method. The last two steps involve trading off limits on Type I and Type II errors and sample size.
- You can use the EnvStats functions listed in Tables 2.1, 2.2, 2.4, and 2.6 (and in the help file *Power and Sample Size*) to estimate required samples sizes for an environmental study.