10

Confidence intervals

10.1 General principles

10.2 A more formal approach

10.3 Some illustrative simulations

In this chapter we continue and deepen the theory of *statistical inference* and introduce the notion of *confidence interval*. We will find that a confidence interval incorporates all the numerical information of the data and presents the results as an interval (or a region if it is multidimensional). We will also see that we at the same time perform a test of hypothesis.

A confidence interval can be considered to be a very honest way to report the information in a given situation. However, there is nothing dishonest in reporting only a point estimate, such as a calculated average or a calculated fault percentage, but it does not reveal anything about the uncertainty in the point estimate. By calculating and reporting an interval we repair this drawback.

10.1 General principles

Anyone that has studied the earlier chapters about mean and standard deviation, linear combinations of variables, models, distributions, etc will now have a good idea of parameters and the need for estimating these parameters. We have calculated *point estimates* such as the calculated average but it is now time to introduce the rather natural idea of confidence intervals. As usual we start from general principles and common experiences.

In our daily lives we experience the idea of using an interval when estimating e.g. temperatures or times: 'The temperature tomorrow afternoon will be 12 – 15 degrees' is a common statement in the weather forecast. In the newspaper it is stated that 'It will take 2 – 3 months to complete the job' where 'the job' is to repair the traffic junction in city.

A person that meets another adult for the first time will usually have no problem to identify the other person's sex. However, when it comes to age he will estimate something like '50 – 55 years of age'. Giving such a statement he knows that he most likely has covered the other person's age with this interval. Suppose that he later on learns that the other person actually was 66 old last autumn. He then realises that his interval was completely wrong; his 'measurements' led him to the incorrect interval.

We can transfer this line of reasoning to situations where we have a set of numerical values and we want to estimate a parameter such as the true mean. Here we will give all necessary mathematics, but if we insist on firm ground for all statements we would be forced to dive deep into the theory. Therefore we will stick to the level of the earlier chapters of this book.

10.1.1 An intuitive reasoning around p and μ

Guessing p in a binomial distribution. If we study n situations or products and report the number of such situations or products with a certain feature, say OK or not OK, we study a binomially distributed variable (see also chapter 8). We also require that the n situations are OK or not OK independently of one another and that the fault rate p (i.e. the probability of being not OK) is constant. The idea is to design a confidence interval for the unknown value of p using the number not OK products. We can of course state the following confidence interval for p:

$$0 \le p \le 1$$

This is however a silly interval and of no help. If we do not find *any* faulty products we also can rule out that $p = 1$ (because then all products are faulty) and state that $0 \le p < 1$. If we find *all* products faulty, we can rule out that $p = 0$ (because such a fault rate would not create any faulty products) and state that $0 < p \le 1$. Still, this kind of reasoning does not lead us anywhere.

Suppose that we call the number of faulty products y and that we after an inspection of $n = 20$ products find one faulty product, i.e. $y = 1$. Our point estimate for p will be $1/20 = 0.05$. Now we will investigate how *high p* can be and still have a fair probability of giving this 'low' outcome (i.e. $y \ge 1$). This will be our upper limit for p. Then we will investigate how *low p* can be and still give this 'high' result (i.e. $y \le 1$). This will be our lower limit for p.

We summarise these last sentences:

A	How high might p be and still give an outcome of ≤ 1 fault?
B	How low might p be and still give an outcome of ≥ 1 fault?

We investigate this and show the result in two tables, A and B. We see that we have to increase our guess about p in order to decrease the risk of being wrong:

A	Possible high values of p	Probability of $y \leq 1$ in a sample of $n = 20$
	0.10	0.392
	0.15	0.176
	0.20	0.069
	0.25	0.024
	0.248735	0.025

We see from table A that if p is as high as 0.248735 we have the probability 0.025 of getting $y \leq 1$. We regard this probability fairly small and therefore we will call 0.248735 our upper limit for p.

Table B reports in the same manner possible low values of p:

B	Possible low values of p	Probability of $y \geq 1$ in a sample of $n = 20$
	0.010	0.182
	0.005	0.095
	0.003	0.058
	0.001	0.020
	0.00126509	0.025

We see from table B that if p is as low as 0.00126509 we have the probability 0.025 of getting $y \geq 1$. We regard this probability fairly small and therefore we will call 0.00126509 our lower limit for p. The result from table A and B is summarised in figure 10.1.1.
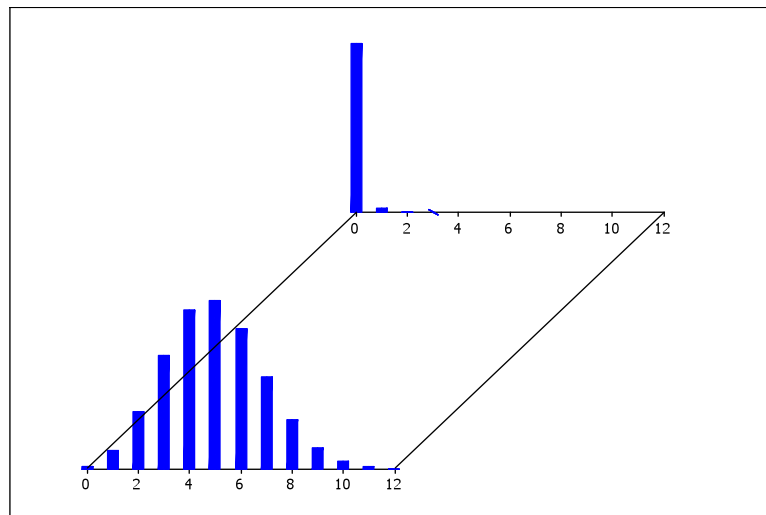
Figure 10.1.1 The p-value might be as low as 0.001265 (but hardly lower) and still produce 1 (or more) faulty product in a sample of 20 products. (The distribution in the back.)

On the other hand, the p-value might be as high as 0.2487 (but hardly higher) and still produce 1 (or less) faulty product in a sample of 20 products. (The distribution in the front.)

These two extreme points will therefore be the endpoints in a confidence interval for p calculated on the given result i.e. $y = 1$ and $n = 20$.

We now have a confidence interval [0.00126509; 0.248735] that with the probability 0.95 covers the true value of the parameter. There is a risk of 0.05 (i.e. 0.025 + 0.025) that the interval misses the true value. There are at least three things to notice:

1. The large number of decimal places is of no practical value but shows that we can get an interval with exactly 0.95 confidence. (The two endpoints of the interval are actually calculated using formulas found in section 10.2.2.)

2. The interval is astonishingly wide. 1 fault in n = 20 can come from a process with $p \approx 0$ to $p \approx 0.25$!

3. The interval is not symmetrical around the point estimate 0.05. See figure 10.1.2!
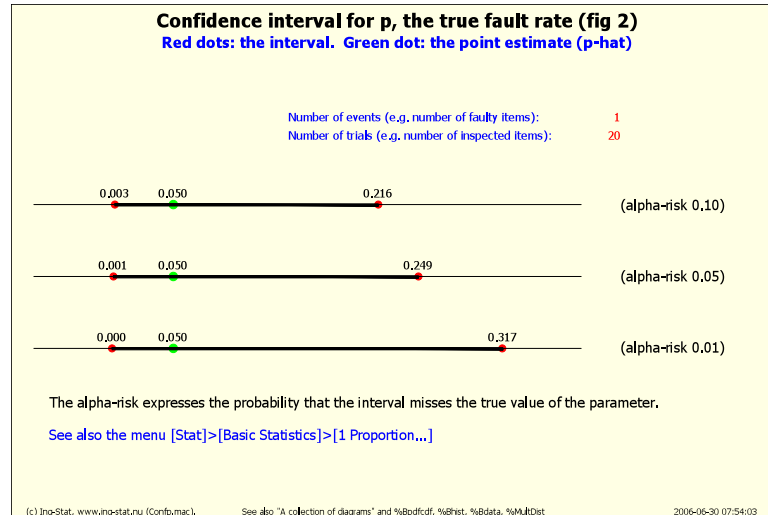

Confidence interval for p, the true fault rate (fig 2)
Red dots: the interval. Green dot: the point estimate (p-hat)

Number of events (e.g. number of faulty items):	1
Number of trials (e.g. number of inspected items):	20

0.003 0.050 0.216 (alpha-risk 0.10)

0.001 0.050 0.249 (alpha-risk 0.05)

0.000 0.050 0.317 (alpha-risk 0.01)

The alpha-risk expresses the probability that the interval misses the true value of the parameter.

See also the menu [Stat]>[Basic Statistics]>[1 Proportion...]

(c) Ing-Stat, www.ing-stat.nu (Confp.mac).   See also "A collection of diagrams" and %Bpdfcdf, %Bhist, %Bdata, %MultDist   2006-06-30 07:54:03

Figure 10.1.2 An 95-percent confidence interval for p based on 1 fault found in a sample n = 20.

(The middle interval.)

N.B. the length of the interval and that it is not symmetrical about the point estimate.

%Confp
%Hypo

ECS – Ex: 6.6

Guessing μ in a normal distribution. We can carry out a similar discussion for a confidence interval for μ in a normal distribution where we consider σ as known. We use the calculated average as an estimator of μ and from chapter 8 we know that the average is a normal random variable with theoretical mean and standard deviation according to the following table. Suppose that the calculated average is 45 using n = 30 and σ = 4:

True mean of $\overline{X}$	Theoretical standard deviation of $\overline{X}$
μ	$\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4}{\sqrt{30}} \approx 0.73$

Suppose that the calculated average is 45 using 30 measurements, and σ = 4. We will carry out reasoning close to the one above for p. We summarise our problem as before:

C   How high can μ be and still give an outcome of $\overline{x} \leq 45$ if $\sigma_{\overline{X}} \approx 0.73$?

D   How low can μ be and still give an outcome of $\overline{x} \geq 45$ if $\sigma_{\overline{X}} \approx 0.73$?

We investigate this and show the result in two tables, C and D:

C	Possible high values of μ	Probability of an average $\leq$ 45, n = 30
	46.0	0.085
	46.1	0.066
	46.2	0.050
	46.3	0.038
	46.43	0.025

We see from table C that if μ is as high as 46.43 we have the probability 0.025 of getting an average $\leq$ 45. Let us call 46.43 our upper limit for μ.

D	Possible low values of μ	Probability of an average ≥ 45, $n = 30$
	44.0	0.085
	43.9	0.066
	43.8	0.050
	43.7	0.038
	43.57	0.025

We see from table D that if μ is as low as 43.57 we have the probability 0.025 of getting an average ≥ 45. Let us call 43.57 our lower limit for μ.

We now have a confidence interval [43.57; 46.43] that with the probability 0.95 covers the true value of the parameter μ. There is a risk of 0.05 (i.e. 0.025 + 0.025) that the interval misses the true value. (The two endpoints of the interval are actually calculated using formulas found in section 10.2.1.)

The result as a figure.   Figure 10.1.3 shows a graphical presentation and reasoning around the confidence interval for μ. The dotted distributions are the normal distribution for the sample average. By 'sliding' the distribution first to the right until the calculated average, shown using a dot, just enters the 2.5 % area in the left tail of the distribution, we get the upper end of the confidence interval (i.e. 46.43).

Then by 'sliding' the distribution to the left, until the calculated area just enters the 2.5 % area in the right tail of the distribution, we get the lower end of the confidence interval. The length of this total 'sliding' is shown using a wider line in the figure, and constitutes the total length of the confidence interval (i.e. 46.43 - 43.57 = 2.86).

NB that while sigma of the process is 4, sigma for the averages is 0.73 if n = 30. See previous page.
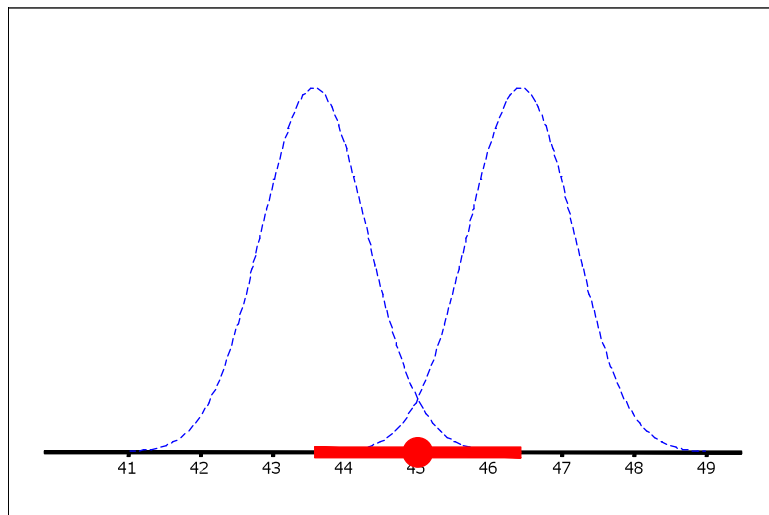

Figure 10.1.3 The left dotted distribution shows how far to the *left* one can slide the whole distribution and still embrace the calculated average (the filled circle).

Correspondingly, the right distribution shows how far to the *right* one can slide the distribution and still embrace the calculated average. The thick line shows the total movement and is then called our confidence interval.

Additional note.   We have used some reasoning to get the desired confidence interval. In following sections we use some mathematical formulas to directly calculate the confidence intervals. In general we will not give any deep explanation for the formulas.

Instead we will use simulation to support the claim for the formulas to create confidence interval with the desired properties. In the discussion above we supposed that σ was known. In reality we usually know only an estimate of σ (i.e. s). However, the formulas will take care of any difficulty caused by this fact.

10.2 A more formal approach

A more formal approach means a further use of mathematics. The main idea is that we use several parts of the data i.e. the calculated average, the estimated standard deviation, and the number of values *n*.
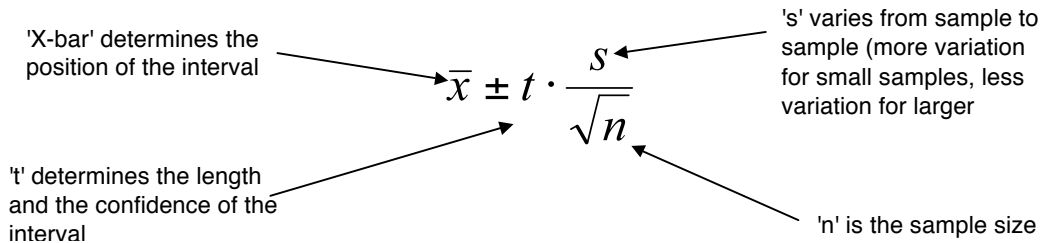
In many calculations of a confidence interval the theory leans on the use of the normal distribution and its features. The reason is that many times the mathematics for an exact calculation of a confidence interval might be too difficult to derive. However, this approximation must not be misunderstood: we do *not* approximate the distribution of interest with a normal distribution, but we do approximate the distribution of the *estimates* of the parameter with a normal distribution. This means that if we do repeated estimates of a parameter, by e.g. simulation, we very often find that these estimates strongly resemble a normal distribution.

%ConfInt

ECS – Ex: 6.1 – 6.5

10.2.1 Calculation of a confidence interval for μ

The interval in figure 10.1.3 can be calculated using a simple formula (N.B. several of the formulas used below, are exactly those used to test a hypothesis):

'X-bar' determines the position of the interval

's' varies from sample to sample (more variation for small samples, less variation for larger

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

't' determines the length and the confidence of the interval

'n' is the sample size

The *t*-value. The *t*-value can be considered as a constant that gives the confidence interval the desired confidence. The higher the *t*-value the larger the confidence is of embracing the true μ. This higher confidence has to be paid for by a longer confidence interval, unfortunately. (The only way to decrease the length of the interval is to increase the sample size *n*).

The *t*-value comes from a so-called *t-distribution*. However, we will not discuss the *t*-distribution at all. When using a computer for the calculations, the proper *t*-value is of course used. Here we introduce just a small table with some common *t*-values just to illustrate a simple calculation:

	The desired confidence level		
n – 1	0.90	0.95	0.99
30	1.6972	2.0423	2.7500
40	1.6839	2.0211	2.7045
50	1.6759	2.0085	2.6778
60	1.6707	2.0003	2.6604
70	1.6669	1.9944	2.6480
80	1.6641	1.9901	2.6387
90	1.6620	1.9867	2.6316
98	1.6606	1.9843	2.6270
100	1.6602	1.9840	2.6259
110	1.6588	1.9818	2.6213
120	1.6577	1.9799	2.6174
∞	1.6449	1.9600	2.5758

A simple example. Suppose that we have measured a number of manufacturing times in e.g. hours and that we want to calculate a 95 percent confidence interval for the true mean μ of the process. Such a confidence interval has a probability of 0.95 to cover the true μ. The calculation gave the following estimates of μ and σ:

$$\bar{x} = 127 \qquad s = 4.3 \qquad n = 81$$

The calculation of the interval gives us the following. The *t*-value is the shaded value in the table above for $n - 1 = 80$ and confidence level 0.95:

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}} = 127 \pm 1.9901 \cdot \frac{4.3}{\sqrt{81}} = 127 \pm 0.951 \implies [126.05, 127.95]$$

This interval has the probability of 0.95 to embrace the true but unknown value of μ. In section 10.3.3 we simulate some intervals for a similar process.

10.2.2 Calculation of a confidence interval for *p*

%Confp

%Hypo

ECS – Ex: 6.6

The parameter *p* in the binomial distribution seems to be an innocent parameter, although there are some difficulties. However, there are a several ways of calculating a confidence interval for the parameter *p*. We will show two of them, one approximate and one exact way (the latter needs a computer program).

An approximate calculation. An approximate confidence interval for *p* can be calculated using the following formula:

$$\hat{p} \pm t \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \quad \text{where} \qquad \hat{p} = \frac{x}{n} \qquad \text{(the number of faulty units (}x\text{) found in a sample divided by number of units } n.)$$

Suppose that we have inspected 500 units (mechanical parts, rows in a computer program, etc.) and found 10 incorrect units. Obviously we can rule out $p = 0$ (because some units are faulty) and $p = 1$ (because not all units are faulty). We are interested in calculating a confidence interval for the unknown parameter *p*:

$$\hat{p} = \frac{x}{n} = \frac{10}{500} = 0.02 \quad \text{This estimate (called } p\text{-hat) is now placed in the next formula:}$$

$$\hat{p} \pm t \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} = 0.02 \pm 1.96 \cdot \sqrt{\frac{0.02 \cdot (1 - 0.02)}{500}} = 0.02 \pm 0.012 \implies [0.008, 0.032]$$

The calculated interval [0.008, 0.032] is now a 95 percent confidence interval that with the probability 0.95 covers the true value of *p*.

Some comments. If the lower endpoint of the interval becomes negative, then it is set to 0 (a negative estimate of an interval for p is silly because $p \geq 0$, i.e. p is never negative). The formula above is an approximation only and the literature often states the following rule-of-thumb:

If $n \cdot \hat{p} \cdot (1 - \hat{p}) > 10$ then the approximate method above can be applied.

Another, maybe more serious, problem arises if $x = 0$ which means that we did not find any incorrect units at all. If we, despite the rule-of-thumb above, use $x = 0$ in the formulas we get zero everywhere. Does this mean that $p = 0$? Of course not. Therefore we need an exact method of calculation the confidence interval.

An exact calculation (I). An exact confidence interval for p can be calculated using the following macro from a computer program. We apply it first to the reasoning of section 10.1.1 and then to the example above:

Code	Description
`let k1 = 1`	Number of incorrect items (y).
`let k2 = 20`	Number of items inspected, n.
`let k3 = 2*(k2-k1+1)`	Number of degrees of freedom (1).
`let k4 = (k1=0)*2+(k1>0)*2*k1`	Number of degrees of freedom (2).
`let k5 = 2*(k1+1)`	Number of degrees of freedom (3).
`let k6 = 2*(k2-k1)`	Number of degrees of freedom (4).
`let k7 = 0.05`	Confidence level α. Here 5 %.
`let k8 = 1-k7/2`	$\alpha/2$, the upper tail area.
`invcdf k8 k9;` `f k3 k4.`	Obtains a value from an F-distribution. Used in the calculations below.
`invcdf k8 k10;` `f k5 k6.`	Obtains a value from an F-distribution. Used in the calculations below.
`let k11 = k1/(k1+(k2-k1+1)*k9)`	The lower endpoint of the interval.
`let k12 = (k1+1)/(k1+1+(k2-k1)/k10)`	The upper endpoint of the interval.
`print k11 k12`	Prints the two endpoints of the interval.

Result: K11   0.00126509   We can see that the result is exactly the
           K12   0.248735   ones in section 10.1.1.

If we have a computer program to calculate a confidence interval for p, we will most likely use that on all occasions and disregard any approximate methods. However it is always handy to have a method that can be used together with a pen and paper and a pocket calculator.

An exact calculation (II).   Here we use the example in **An approximate calculation** above. We put in $x = 10$ and $n = 500$ in the macro and run it again. Then we get the following:

Result:
```
K11        0.0096
K12        0.0367
```

The approximate result, calculated earlier ([0.008, 0.032]), is similar to the exact result ([0.0096, 0.0367]). It is of course better to use a procedure that gives an exact result, given the input and proper assumptions, although the main drawback is the need for a computer macro.

Several observations.   If we have a number of observations from a process where we want to calculate an interval for p, we will of course pool the result. Suppose that we have the following result: $x_1 = 3$, $n_1 = 210$, $x_2 = 0$, $n_2 = 240$, and $x_3 = 4$, $n_3 = 310$, we will of course add these figures to $x = 3 + 0 + 4 = 7$, $n = 210 + 240 + 310 = 760$ and calculate the confidence interval as shown above.

10.2.3 Calculation of a confidence interval for λ

%Conflam

ECS – Ex: 6.7

The parameter λ, used in the Poisson distribution, is usually called the *intensity* of the distribution. There are a number of different ways of calculating a confidence interval for the parameter and we will show two of them, one approximate and one exact way (the latter one needs the support of a computer).

An approximate calculation.   An approximate confidence interval for λ can be calculated using the following formula:

$$\hat{\lambda} \pm t \cdot \sqrt{\hat{\lambda}}$$

where $\hat{\lambda}$ is the number of observed events and t is a constant to give the interval correct confidence level. Suppose that we have observed a certain process for 1 hour and that we found 8 events. Obviously we can rule out $\lambda = 0$ (because we did observe some events). We are interested in calculating a confidence interval for the unknown parameter λ and we use the formulas above:

$$\hat{\lambda} \pm t \cdot \sqrt{\hat{\lambda}\cdot} = 8 \pm 1.96 \cdot \sqrt{8} = 8 \pm 5.54 \Rightarrow [2.46,\ 13.54]$$

The calculated interval [2.46, 13.54] is now a approximately 95 percent confidence interval that with the approximate probability 0.95 covers the true value of λ.

Some comments.   If the lower endpoint of the interval becomes negative, then it is set to 0 (a negative estimate of an interval for λ is silly because $\lambda \geq 0$, i.e. λ is never negative). The formula above is an approximation only and there are several other, more mathematical formulated approximations. The one above is to be used when only pen an paper is available. For exact intervals a computer program is the easiest way.

Another problem arises if we did not observe any events at all. If we use $\hat{\lambda} = 0$ in the formulas we get zero everywhere. Does this mean that $\lambda = 0$? Of course not. Therefore we need an exact method of calculation the confidence interval.

An exact calculation (I). An exact confidence interval for λ can be calculated using the following macro from a computer program:

```let k1 = 1```	Number of events.
```let k2 = 2*k1+1*(k1=0)```	Number of degrees of freedom, lower limit.
```let k3 = 2*(k1+1)```	Number of degrees of freedom, upper limit.
```let k4 = 0.05```	Confidence level, α. Here 5 %.
```let k5 = k4/2```	The lower tail area, $\alpha/2$.
```let k6 = 1–k4/2```	The upper tail area, $1 - \alpha/2$.
```invcdf k5 k7;``` ```chisquare k2.```	Used for calculating the lower limit.
```invcdf k6 k8;``` ```chisquare k3.```	Used for calculating the upper limit.
```let k9 = 0.5*k7*(k1>0)```	The lower endpoint of the interval.
```let k10 = 0.5*k8```	The upper endpoint of the interval.
```print k9 k10```	Prints the two endpoints of the interval.

Result:
```
K9        0.0
K10       3.68888
```

A simple check of the result. We can check this result easily by using some computer commands:

```pdf;``` ```poisson 3.68888.```	Calculates the probabilities for 0, 1, 2 etc events when λ = 3.6888.

Executing this ```pdf```-command, we will see that there is a small probability (0.0250 i.e. exactly $\alpha/2$) that we get an outcome of zero events when λ = 3.68888. This is therefore our upper limit.

An exact calculation (II). Suppose that we found 8 events (faults) during a certain time or on a certain area etc. What is a 95 percent confidence interval for λ? We use the computer program again:

Result:
```
K9        3.45383
K10       15.7632
```

A simple check of the result. We can check this result easily by using some computer commands:

```cdf;``` ```poisson 3.45383.```	Calculates the accumulated probabilities for 0, 1, 2 etc events when λ = 3.45383.
```cdf;``` ```poisson 15.7632.```	Calculates the accumulated probabilities for 0, 1, 2 etc events when λ = 15.7632.

Executing these ```cdf```-commands, we will see that if λ = 3.45383 there is a small probability (0.0250 i.e. exactly $\alpha/2$) that we get an outcome of > 8 events. This is therefore our lower limit. We will also see that if λ = 15.7632 there is a small probability (0.0250 i.e. exactly $\alpha/2$) that we get an outcome of ≤ 8 events. This is therefore our upper limit.

Several observations. If we have a number of observations from a process where we want to calculate an interval for λ, we will of course pool the result. Suppose that we have the following result: $x_1 = 6$, $x_2 = 2$, $x_3 = 4$ events, each value during e.g. 24 hours. Then we add these figures to $x = 6 + 2 + 4 = 12$ events during 72 hours:
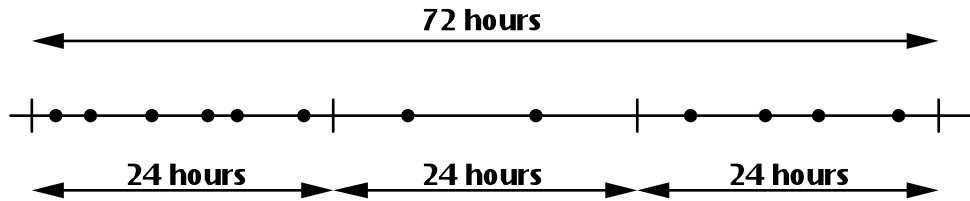
However, from the theory about the Poisson distribution, we know that a sum of e.g. three observations from a Poisson distribution with parameter λ, is an observation from a Poisson distribution with the parameter $\lambda + \lambda + \lambda = 3\lambda$. If we then use 12 events for calculating a confidence interval for the parameter, we do that for Po(3λ), using the designation from chapter 8. Therefore, we have to divide the values for the end points by 3 to get the confidence interval for the parameter λ. This is done in the simulation in section 10.3.3.

10.2.4 Calculation of a confidence interval for $\mu_1 - \mu_2$

%t-test

Sometimes we want to compare the true mean μ from two different processes in order to understand how large the difference might be. It can be two different methods of doing the same job, the result from two different machines etc. This need to compare outcomes from processes is usually the same in many areas of statistics: design of experiments, analysis of variance, regression analysis, etc. The jargon is unfortunately different which disguises the similarities.

An example. The following calculations are based on section 9.3 where we perform some tests of hypotheses. While the result from a test of hypothesis is whether or not we can reject the stated hypothesis, a calculate confidence interval shows how large the difference between might be. Thus the confidence interval is of more practical value. We repeat some of the printout from section 9.3:

```
Two-sample T for C1 vs C2
      N       Mean      StDev    SE Mean
C1   50      44.77       4.27       0.60
C2   50      46.60       3.98       0.56

95% CI for difference: (-3.46; -0.19)
```

The printout from section 9.3 although slightly shorter.

The following formula is used to calculate a confidence interval for the difference between two means. N.B. that the formulas are exactly those used in section 9.3 but there we calculated a *t*-value to test a hypothesis. Here we pick a *t*-value from a table to give the interval correct length:

$$\bar{x}_1 - \bar{x}_2 \pm t \cdot \sqrt{\left[\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{(n_1 - 1) + (n_2 - 1)} \right] \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Using the data from the printout we get the following:

$$44.77 - 46.60 \pm 1.98 \cdot \sqrt{\left[\frac{(50-1)\cdot 4.27^2 + (50-1)\cdot 3.98^2}{(50-1)+(50-1)}\right]\cdot \left(\frac{1}{50}+\frac{1}{50}\right)}$$

The resulting confidence interval becomes [–3.46, –0.19] which is exactly the same interval as given in the printout above (and in section 9.3). This interval embraces the true difference between μ_1 and μ_2 with the probability of 95 %. In section 10.3.6 we simulate this problem.

10.2.5 Calculation of a confidence interval for $p_1 - p_2$

Sometimes we want to compare two different proportions from e.g. two different processes, two different programming environments, two different machines, two different feeding devices etc. The idea is that we let the processes, machines or whatever we have, carry out a task that is judged as either 'OK' or 'not OK'. In the programming example every task is a line of code, in the machine example we might have an assembly that the machine puts together, in the feeding example we might have a component that is either positioned correctly or incorrectly.

Sometimes we might find a better variable that we can measure in a different way. The result of the feeding operation might e.g. be measured as a deviation distance from a certain target, and we would then calculate a confidence interval for μ.

However, suppose that we only have the possibility to record the result as the binary outcome. Of course, we have made sure that the only difference in the result is the different ability of the processes. It would be incorrect to allocate a simple type of assembly to one machine and a difficult assembly to the other machine. Such an arrangement would make it impossible, without using a more complicated analysis, to evaluate any difference between the machines.

We call the two different machines A and B respectively. Suppose that we did get the following result:

	Number of faulty items	Number of inspected items	Estimated fault rate
A	x_1	n_1	$\hat{p}_1 = \dfrac{x_1}{n_1}$
B	x_2	n_2	$\hat{p}_2 = \dfrac{x_2}{n_2}$

If $n_1 \cdot \hat{p}_1 \cdot (1-\hat{p}_1) > 10$ and $n_2 \cdot \hat{p}_2 \cdot (1-\hat{p}_2) > 10$ then we can calculate a confidence interval for the true difference between p_1 and p_2. We use the following formula:

$$\hat{p}_1 - \hat{p}_2 \pm t \cdot \sqrt{\frac{\hat{p}_1 \cdot (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1-\hat{p}_2)}{n_2}} \quad \text{where} \quad \hat{p}_1 = \frac{x_1}{n_1} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

An example. Suppose that we have inspected 500 items from process A and that we found 30 incorrect items. The corresponding figures from process B are 400 and 50 respectively. What is a 95 per cent confidence interval for the true but unknown difference in proportion of faults of process A and B? We get using the formulas above. The t-value comes from the table above with confidence level 0.95:
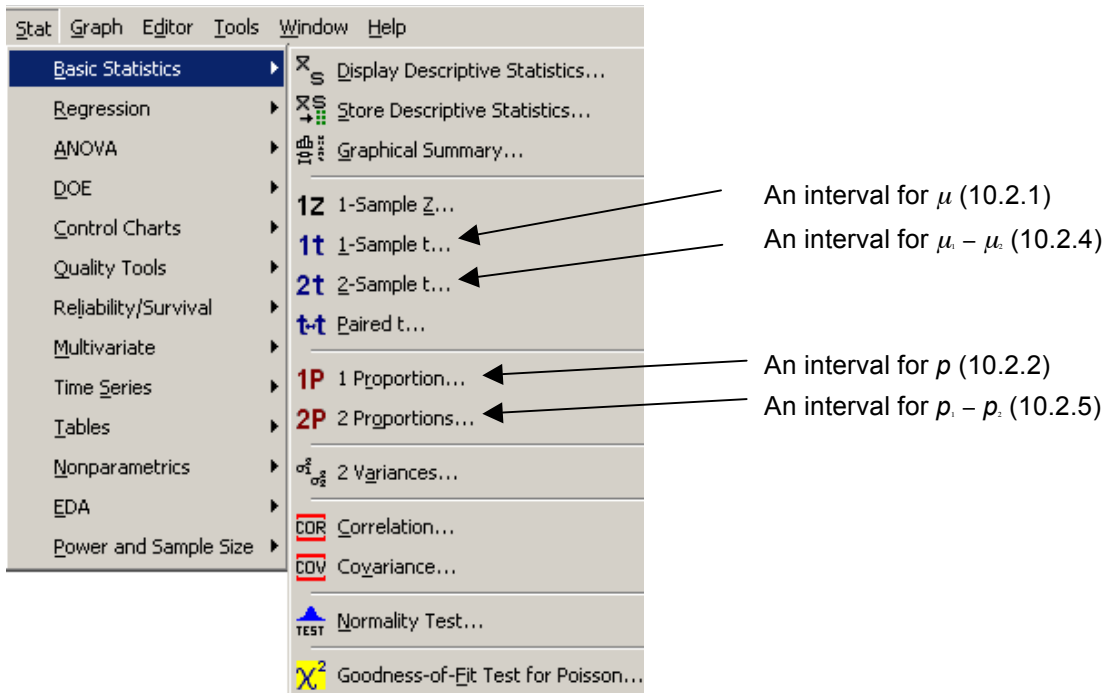
$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{30}{500} = 0.06 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{50}{400} = 0.125$$

$$\hat{p}_1 - \hat{p}_2 \pm t \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}} =$$

$$= 0.06 - 0.125 \pm 1.96 \cdot \sqrt{\frac{0.06 \cdot (1 - 0.06)}{500} + \frac{0.125 \cdot (1 - 0.125)}{400}} = -0.065 \pm 0.039$$

The calculated interval is then [–0.026, –0.104]. This means that we conclude that process B has an inferior quality. Unfortunately the confidence interval becomes rather long in spite of the 900 observations. This is a drawback using this kind of variables. In section 10.3.7 we simulate this problem.

The menus of Minitab for tests of hypothesis and calculation of confidence interval

An interval for μ (10.2.1)

An interval for $\mu_1 - \mu_2$ (10.2.4)

An interval for p (10.2.2)

An interval for $p_1 - p_2$ (10.2.5)

10.3 Some illustrative simulations

In order to illustrate the calculation and use of confidence intervals, we carry out some simulations. We will show how the interval is influenced by different number of observations but also how different confidence levels can be used. We simulate confidence intervals with as low as 50 percent confidence.

In reality nobody states conclusions with such a low confidence. The temptation is of course the short length of the interval. It means however that, on the average, every second confidence interval will miss the true value of the parameter. Nobody will trust a person using 50 percent confidence intervals!

10.3.1 Simulation of confidence intervals with different *n*

In section 10.2.1 we calculated an interval for μ. From the formulas one can see that larger *n* will produce shorter confidence interval. Below we simulate data sets with $n = 50$, $n = 250$, $n = 450$ and $n = 650$ and calculate a confidence interval for each of the 20 sets of data. The results are shown in figure 10.3.1:

%ConfInt

```
let k2 = 120
```
The expected value μ.

```
let k3 = 10
```
The standard deviation σ.

```
random 50 c1-c5;
normal k2 k3.
```
Stores 50 random values from a normal distribution in the columns c1-c5.

```
random 250 c6-c10;
normal k2 k3.
```
Stores 250 random values from a normal distribution in the columns c6–c10.

```
random 450 c11-c15;
normal k2 k3.
```
Stores 450 random values from a normal distribution in the columns c11–c15.

```
random 650 c16-c20;
normal k2 k3.
```
Stores 650 random values from a normal distribution in the columns c16–c20.

```
tint c1-c20
```
Obtains a 95 percent confidence interval for each of the columns c1–c20.
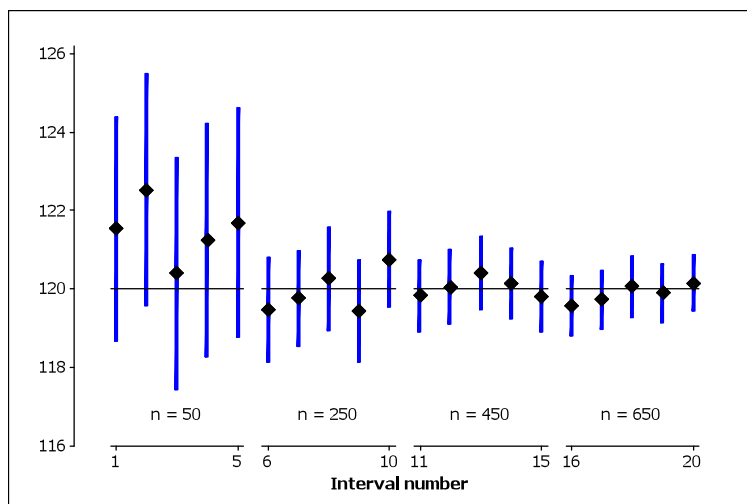

Figure 10.3.1 20 different confidence intervals. There are four groups of five intervals based on 50, 250, 450 and 650 observations. The horizontal line at 120 is the true value of μ used in the simulation.

The intervals become shorter when *n* increases.

We actually expect 1 interval in 20 (or 5 in 100) to miss the true value (95 % confidence).

We can see that the interval becomes shorter when *n* increases.

10.3.2 Simulation of confidence intervals with different confidence

In section 10.2.1 we calculated an interval for μ. From the formulas one can see that larger t will produce confidence interval with larger confidence. Unfortunately, the length thereby increases. Below we simulate data sets with 99, 95, 90 and 50 percent confidence. As expected, if we lower the confidence more intervals will miss the true value of the parameter. The results are shown in figure 10.3.2:

```
let k1 = 200
```
The number of values per column.

```
let k2 = 120
```
The expected value μ.

```
let k3 = 10
```
The standard deviation σ.

```
random k1 c1-c20;
normal k2 k3.
```
Stores k1 random values from a normal distribution in the columns c1–c20.

```
tint 99 c1-c5
```
Obtains a 99 percent confidence interval for each of the columns c1–c5.

```
tint 95 c6-c10
```
Obtains a 95 percent confidence interval for each of the columns c6–c10.

```
tint 90 c11-c15
```
Obtains a 90 percent confidence interval for each of the columns c11–c15.

```
tint 50 c16-c20
```
Obtains a 50 percent confidence interval for each of the columns c16–c20.
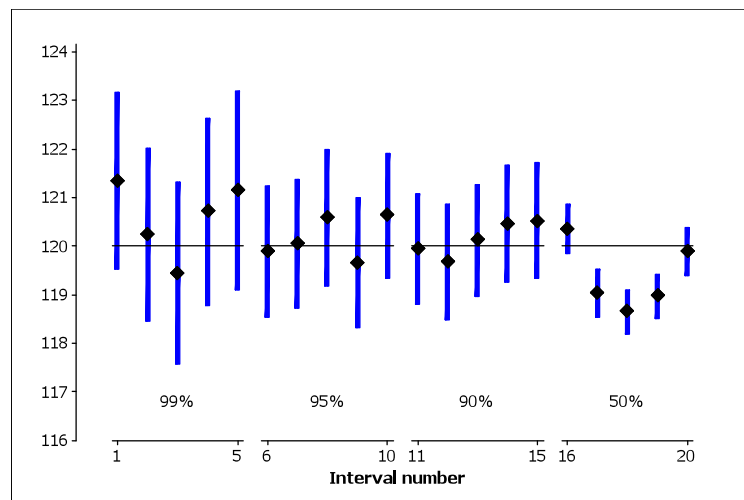
%ConfInt

Figure 10.3.2 20 different confidence intervals. There are 4 groups 5 intervals based on 99, 95, 90 and 50% confidence.

The horizontal line at 120 is the true value of μ used in the simulation.

The intervals become shorter when the confidence is lowered. But as expected, more intervals miss the true value. In the last group (50 percent confidence) 3 out of 5 intervals miss the true value!

If we lower the confidence level, more intervals will miss the true value. Intervals based on as low as 50 percent confidence are of no value. Short but useless! On the other hand, if we demand too high a confidence, the intervals might become too long.

10.3.3 Simulation of confidence intervals for μ

In section 10.2.1 we showed how to calculate a confidence interval for μ. Here we will, as in section 10.3.1 and 10.3.2, simulate 20 sets of data and use the computer to calculate the intervals. In reality, treating real values we would of course calculate just one such interval using all the values, at least if we think the data comes from the same process. Therefore we will calculate an interval for all the data.

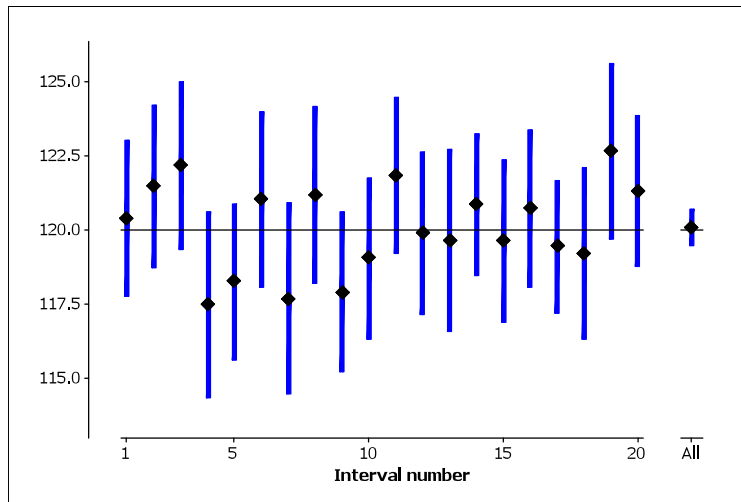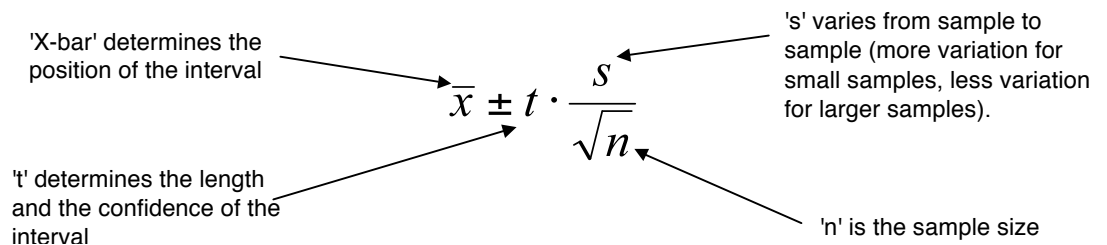|	
```let k1 = 50```	Number of simulated values in each sample.
```let k2 = 120```	Set μ = 120.
```let k3 = 10```	Sets σ = 10.
```random k1 c1-c20;``` ```normal k2 k3.```	Generates k1 values in c1-c20
```stack c1-c21```	Stacks all data into column c21.
```tint c1-c21```	Calculates intervals for the 21 columns.

%ConfInt

Figure 10.3.3 20 different confidence intervals for μ with 95 % confidence.

The interval labelled 'All data' is based on all 1000 values. This interval becomes shorter. However one must consider the price for this shorter interval.

Figure 10.3.3 shows many of the important features of calculating confidence intervals. Firstly, we expect 1 interval in 20 (i.e. 5%) will miss the true value of the parameter of interest. We happened to get 0 of the 20 intervals outside the true parameter value (120).

We can also see, at least if one uses a ruler, that both the position and length of the intervals vary. This is natural as the interval is positioned using the calculated average, which varies from sample to sample. In calculating the length of the interval we use a t-value, n and s. Here s is estimated from the sample and thus s varies from sample to sample:

'X-bar' determines the position of the interval

's' varies from sample to sample (more variation for small samples, less variation for larger samples).

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

't' determines the length and the confidence of the interval

'n' is the sample size

10.3.4 Simulation of confidence intervals for *p*

In section 10.2.2 we showed an exact method to calculate a confidence interval for *p*. Here we will use that macro program to calculate intervals based upon simulated observations from a process with *p* = 0.02. We simulate five samples with *n* = 100, 150, 200, 250 respectively and expect that 1 in 20 of the intervals will miss the true value (*p* = 0.02) if we calculated intervals with 95 percent confidence.

```
random 5 c1;
binomial 100 0.02.
```
Generates 5 values from Bin(100, 0.02).

```
random 5 c2;
binomial 150 0.02.
```
Generates 5 values from Bin(150, 0.02).

```
random 5 c3;
binomial 200 0.02.
```
Generates 5 values from Bin(200, 0.02).

```
random 5 c4;
binomial 250 0.02.
```
Generates 5 values from Bin(250, 0.02).

```
stack c1-c5
```
Stacks all data into column c5.

%Confp

Now for every value in column c5 we use the macro for calculating an interval for *p*. The two end points and the estimate of *p* are shown in figure 10.3.4. If we in a real situation have a number of values from the process, we of course add up the number of faulty items to one figure and the number of inspected items to another figure and calculate a confidence interval from this. We have done so in this simulation and the result is shown in the diagram as 'All data'.
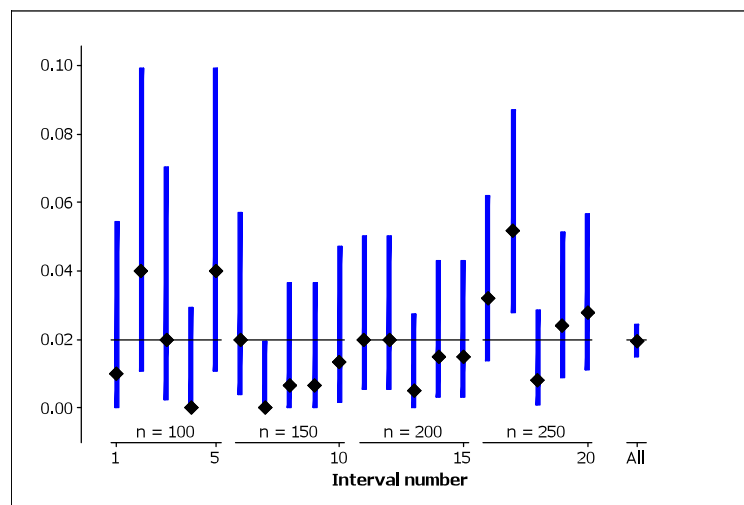

Figure 10.3.4 20 different confidence intervals for *p* with 95 % confidence.

We see that one interval (number 17) misses the true value (the line at 0.02) altogether.

The interval labelled 'All data' is based on all 3500 values. This interval becomes shorter. However one must consider the price for this shorter interval.

We have here intervals are based on *n* = 100, 150, 200 and 250 values. The intervals become shorter when *n* increases but there is a large variation. See e.g. the first and second interval where the first one is much shorter. This is typical for this type of variables. (The fourth interval happened to be based on zero faults found. This sets, quite naturally, the point estimate and the lower end to zero.)

This simulation gave 1 interval (number 17) that does not embrace the true value of the estimate. This is according to the theory for calculating confidence intervals.

10.3.5 Simulation of confidence intervals for λ

In section 10.2.3 we showed an exact method to calculate a confidence interval for λ. Here we will use that macro program to calculate 20 intervals based upon simulated observations from a process with e.g. $\lambda = 5.6$. We expect that 1 in 20 of the intervals will miss this true value if we calculated intervals with 95 percent confidence.

```random 20 c1;``` ```poisson 5.6.```	Generates 20 values from Po(5.6).
```let k1 = sum(c1)```	Sums all data into constant k1.

Now for every value in column c2 we use the macro for calculating an interval for λ. The two end points and the estimate of λ are shown in figure 10.3.5. If we in a real situation have a number of values from the process, we of course add up the number of events to one figure calculate a confidence interval from this. We have done so in this simulation and the result is shown in the diagram as 'All data'.

A remark. Suppose that our original observations are 'number of events during 10 minutes'. Then the sum of 20 such values means 'number of events during 200 minutes'. If we then calculate the confidence interval for the sum, we get an interval for the intensity 'number of events during 200 minutes'. However, we want the interval for 'number of events during 10 minutes', and we get this by divide the endpoints by 20. This result is shown in the diagram, and we can see that this interval is shorter, as we expect, because it is based on more information.
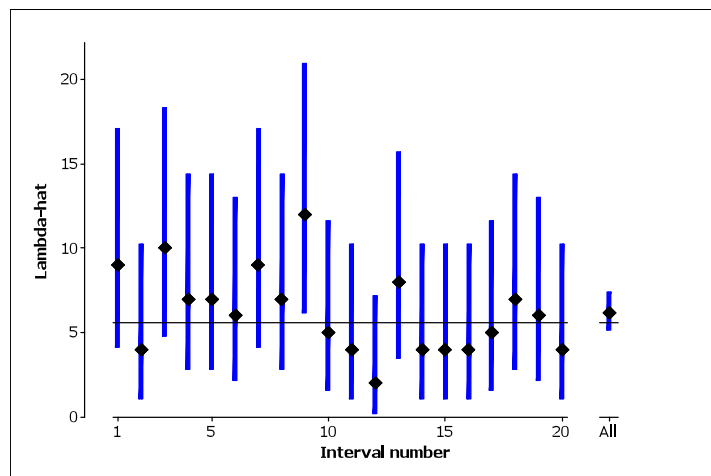

Figure 10.3.5 20 different confidence intervals for λ with 95 % confidence.

We see that the 9th interval misses the true value (the line at 5.6).

The interval labelled 'All data' is based on all 20 values. This interval becomes shorter.

The each one of intervals in figure 10.3.5 are based on one observation from the Poisson distribution with the parameter $\lambda = 5.6$. If we have several observations from this distribution that we want to use in calculating the interval, we add these values and use this sum when running the computer program for the calculation.

See also the remark above and a similar remark at the end of section 10.2.3. Chapter 8 gives a thorough treatment of the different features of the Poisson distribution.

10.3.6 Simulation of confidence intervals for $\mu_1 - \mu_2$

In section 10.2.4 we showed how to calculate a confidence interval for the difference between two different expected values μ_1 and μ_2. This a common and rather straight forward situation. In this section we perform simulation to show some of the features of a confidence interval.

Some comments to the simulation. There are enough computer commands to directly calculate the wanted confidence interval, but here we in stead made use of the following formula, found in section 10.2.4:

$$\underbrace{\overline{x}_1 - \overline{x}_2}_{\underset{\text{C103}}{\underbrace{\;\;}}} \pm \underbrace{t}_{\text{K1}} \cdot \sqrt{\underbrace{\left[\frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{(n_1 - 1) + (n_2 - 1)} \right]}_{\text{C106}} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\underset{\text{C101}}{\overline{x}_1} \quad \underset{\text{C102}}{\overline{x}_2}$$

%t-test

We simulate 20 normally distributed values in each of the columns c1-c50, where all values come from the same distribution. Then we simulate 20 normally distributed values in each of the columns c51-c100, where all values come a different distribution. Every row of column c1-c50 makes up one sample and, likewise, every row of column c51-c100 makes up 20 other samples.

Now we calculate the average for each of the 40 samples and put the results in column c101 and c102 respectively, and their differences, row by row, are placed in column c103. The standard deviation for each sample is placed into the column c104 and c105 respectively.

Finally we calculate the lower confidence limit and store the result in column c201 and the upper confidence limit is stored in column c202.

```random 20 c1-c50;``` ```normal 48 5.```	Stores 20 random values from a normal distribution in the columns c1-c50.
```random 20 c51-c100;``` ```normal 45 5.```	Stores 20 random values from a normal distribution in the columns c51–c100.
```rmean c1-c50 c101```	Stores the rowwise average of c1–c51 in c101.
```rmean c51-c100 c102```	Stores the rowwise average in c102.
```let c103 = c101-c102```	Stores difference between averages in c103.
```rstan c1-c50 c104```	Stores rowwise stand. dev. of c1–c51 in c104.
```rstan c51-c100 c105```	Stores rowwise stand. dev. in c105.
```invcdf 0.975 k1;``` ```t 98.```	Obtains the proper t-value for 100 measurements.
```let c106 = ((50-1)*c104**2 +``` ``` (50-1)*c105**2)/((50-1) + (50-1))```	Stores the pooled variance in c106, row by row.
```let c201 = c101-c102 -``` ``` k1*sqrt(c106*(1/50 + 1/50))```	Stores the lower confidence limit for each sample, row by row, in c201.
```let c202 = c101-c102 +``` ``` k1*sqrt(c106*(1/50 + 1/50))```	Stores the upper confidence limit for each sample row by row, in c202.

Figure 10.3.6 shows all the confidence limits as a diagram. The quickest way to create the interval for *all* the data is the following:

```stack c1–c51 c501```	Stacks all data from the first distribution in column c501.
```stack c51–c100 c502```	Stacks all data from the second distribution in column c502.
```twosample c501 c502```	Calculates the confidence interval for the differences.
```let c201(21) = a```	'a' is the lower confidence limit taken from the printout from the ```twosample```-command. (This value can also be written directly into the data window, row 21 of column c201.)
```let c103(21)=mean(c501) – mean(c502)```	Stores the difference in mean of the total data in column c103, row 21.
```let c202(21) = b```	'b' is the upper confidence limit taken from the printout from the ```twosample```-command. (This value can also be written directly into the data window, row 21 of column c202.)

The following commands can be used to create the diagram:

Plotting of the simulated result

|	
```stack c201 c103 c202 c301```	Stacks 'lower' on 'point' on 'upper' c301.
```stack c103 c103 c103 c302```	Stacks 'point' on 'point' on 'point' c302.

```set c303```
```3(1:20 22)```
```end```
	Creates an x-axis for the diagram. The 'All data'-interval is plotted at position 22.

```plot c301*c303;``` ```project;```	Plots the calculated points.
```base c302;```	Draws a vertical line to the centre points.

```symbol;```
```size 0.9;```
```reference 2 3.```
	Draws also a horizontal line for the true difference i.e. on the *y*-axis at *y* = 3.
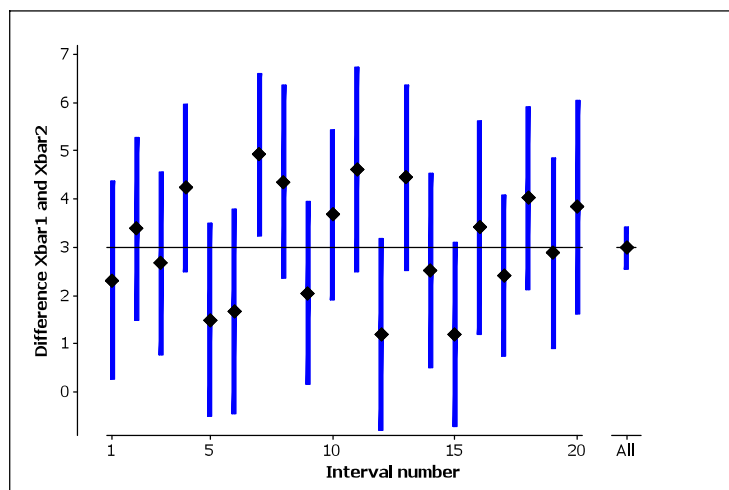
Figure 10.3.6 20 different 95 % confidence intervals for the difference between two mean values.

We see that interval 7 misses the true value (the line at 3).

The interval labelled 'All data' is based on all 2 x 1000 values. This interval becomes much shorter.

10.3.7 Simulation of confidence intervals for $p_1 - p_2$

In section 10.2.5 we calculated a confidence interval for the difference of two proportions. Here we will simulate data from two different processes. Suppose that the two processes has the true parameters $p_1 = 0.14$ and $p_2 = 0.06$ respectively and that we take the samples $n_1 = 200$ and $n_2 = 300$ respectively. The true difference between p_1 and p_2 is obviously 0.08.

If simulate a large number of samples and calculate the difference in proportions we will on the average get the difference to be 0.06. If we also use the method in 10.2.5 to calculate confidence intervals of, say, 95 percent we will observe that approximately 5 percent of the intervals do not embrace the true difference. (Using real data we will of course never know if an interval misses the true difference.)

`let k1 = 0.14`	The proportion p_1.
`let k2 = 0.06`	The proportion p_2.
`let k3 = 200`	The number in sample n_1.
`let k4 = 300`	The number in sample n_2.
`let k5 = 30`	Number of intervals to be simulated.
`let k6 = k1 – k2`	Difference between p_1 and p_2.
`random k5 c1;` `binomial k3 k1.`	Generates k5 values from a binomial distribution.
`random k5 c2;` `binomial k4 k2.`	Generates k5 values from another binomial distribution.
`let c1 = c1/k3`	Calculates observed fault rate in process 1.
`let c2 = c2/k4`	Calculates observed fault rate in process 2.
`let c3 = c1 – c2`	Calculate the observed difference.

`let c4 = c1 – c2 – 1.96 *`
`     sqrt((c1*(1–c1)/k3)`
`     + (c2*(1–c2)/k4))`
	Calculates lower end of confidence interval.

`let c5 = c1 – c2 + 1.96 *`
`     sqrt((c1*(1–c1)/k3)`
`     + (c2*(1–c2)/k4))`
	Calculates upper end of confidence interval.

`print c4 c3 c5`	Prints the lower end, middle, and upper end of the confidence interval.

Plotting of the simulated result

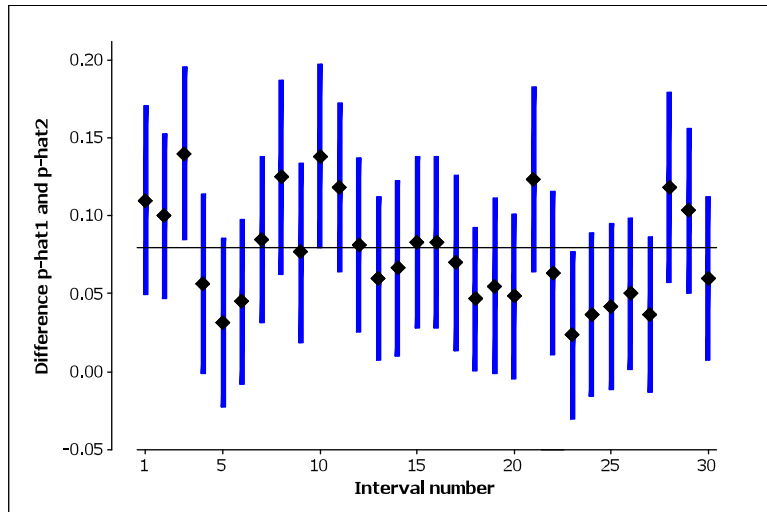|	
`stack c4 c3 c5 c11`	Stacks 'lower' on 'point' on 'upper' in c11.
`stack c3 c3 c3 c12`	Stacks 'point' on 'point' on 'point¨' in c12.

`set c13`
`3(1:k5)`
`end`
	Creates an x-axis for the diagram.

`plot c11*c13;`
`project;`
`base c12;`
`symbol;`
`size 0.9;`
`reference 2 k6.`
	Plots the calculated points. Draws a vertical line to the centre points. Draws also a horizontal line for the true difference.

Figure 10.3.7 These confidence intervals are calculated using the macro above. The theory here rests on an approximation with the normal distribution.

Here interval number 3 and 23 miss the true difference in fault rate between the two processes, 0.08.

Some of the intervals has the lower endpoint on the negative side. This is of course impossible and the lower end point should be set to zero.

We can see from figure 10.3.7 that two intervals miss the true difference of 0.08. Is this in any way an extreme outcome? This question can be answered the usual way. Which means that we want to calculate the probability of getting two or less than 2 incorrect intervals in a sample of $n = 30$ intervals when the true $p = 0.05$ i.e. we expect 5 percent of all interval to miss the true value.

This is a situation when we use the binomial distribution and we use a few computer commands to answer the question:

```
cdf;
binomial 30 0.05.
```

cdf is the cumulated distribution function and gives the probability $P(X \leq x)$.

The answer is $P(X \leq 1) = 0.5535$.

We see that this result is what is expected.

An interval for μ (10.3.3)

An interval for $\mu_1 - \mu_2$ (10.3.6)

An interval for p (10.3.4)

An interval for $p_1 - p_2$ (10.3.7)

10.3.8 Simulation of confidence intervals for the famous number π

The number π is common in many mathematical entities. Although we are used to it in e.g. trigonometry, the number also has a habit of popping up in the most unexpected places. It has been calculated with many decimals so there is actually no need to simulate a confidence interval. However, a simulation of the number π illuminates the features of this chapter in an interesting way.

In the statistical literature we find that an experiment was carried out already 1850 by a person called Volser who calculated an approximation 3.1596 as a point estimate of π (π is equal to 3.14159 with five decimals).

Suppose that we throw a pin on a plane having parallel lines. The length of the pin is half the distance between the lines. We throw the pin a number of n times and count the number of times the pin crosses any line. We call this number k. From this we can state the following (based on some theory not shown here):

$$\frac{k}{n} \approx \frac{1}{\pi} \quad \text{that can be transferred to} \quad \pi \approx \frac{n}{k}$$

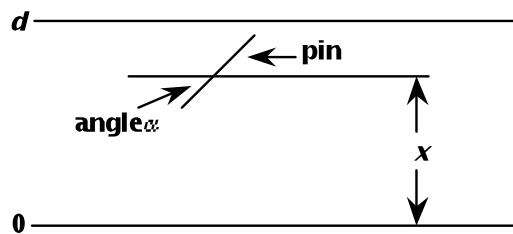This approximation becomes better for higher n.

Figure 10.3.9 The figure shows a pin thrown on a plane with parallel lines, the distance d apart.

x shows the distance from a line to the point of gravity of the pin. This distance is considered a random variable uniformly distributed on [0, d].

The angle α is likewise distributed on [0, 90°].

Call the distance between the parallel lines d. The length of the pin is then $d/2$ and x is the distance between the centre of the pin from, say, the zero-line. Then we can state the following (which can be verified by some mathematics):

If $\quad x - \dfrac{d}{4}\sin(\alpha) < 0 \quad$ the pin crosses the 0-line

$\quad x + \dfrac{d}{4}\sin(\alpha) > d \quad$ the pin crosses the d-line.

From a uniform distribution we get the x-value ([0, d]) and from another uniform distribution we get the angle i.e. the α-value ([0, 90°] which is the interval [0, $\pi/2$] when expressed in radians). By some mathematics we calculate the two statements and check if the pin crosses any of the two lines. Every such crossing is accumulated and again using some mathematics we get an estimate of π. The program that we use is the following:

```
let k1 = 4
```
The length between the lines.

```
let k2 = 500
```
Number of 'throws' to be simulated.

```
random k2 c1;
uniform 0 k1.
```
Generates k2 uniformly distributed values in the interval [0, 4]. These values are the x-distances in figure 10.3.9.

```
random k2 c2;
uniform 0 1.5708.
```
Generates k2 uniformly distributed values in the interval [0, $\pi/2$]. These values are the angle α in figure 10.3.9.

```
let k11 = sum((c1 - sin(c2) < 0)
     or (c1 + sin(c2) > k1))/k2
```
Counts the number of times a line is crossed and calculates the proportion.

```
print k11
```
Prints the proportion of 'line crossing'.

k11 is now an estimate ($\hat{p}$) of the true proportion p. By using the macro of section 10.2.2 we can calculate a lower and upper confidence limit for this proportion. By using the following relationship, we can calculate a lower and upper confidence limit for the number π together with a point estimate:

$$\pi \approx \frac{n}{k} = \frac{1}{\hat{p}}$$

If we use this several times we get different intervals. The table below shows such simulations together with a compilation of *all* data (the true p is $1/\pi \approx 0.318$):

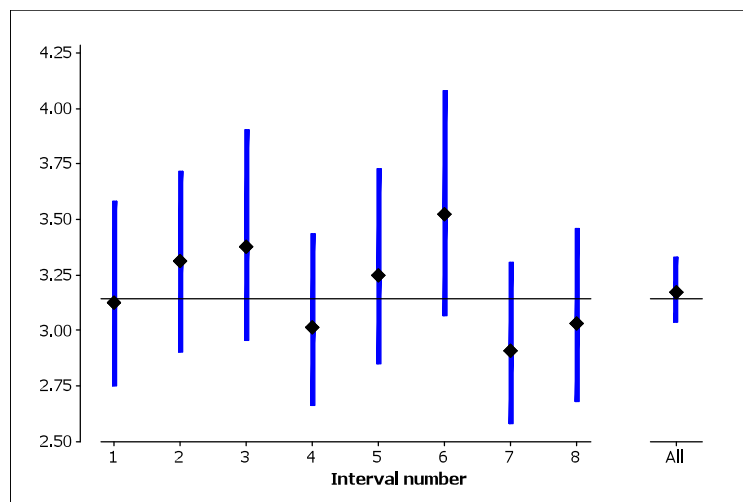|	**p = 1/π ≈ 0.318**			**Confidence limits for π**	
Simulation	p_L	$\hat{p}$	p_U	**Lower ($1/p_U$)**	**Upper ($1/p_L$)**
1	0.279	0.320	0.363	2.75	3.58
2	0.269	0.302	0.344	2.91	3.82
3	0.256	0.296	0.338	2.96	3.91
4	0.291	0.332	0.375	2.67	3.44
5	0.268	0.308	0.351	2.84	3.73
6	0.245	0.284	0.326	3.07	4.08
7	0.302	0.344	0.387	2.58	3.31
8	0.289	0.330	0.373	2.68	3.46
All	0.300	0.315	0.329	3.04	3.33

Figure 10.3.10 The figure shows eight confidence intervals for the number π.

All confidence intervals include the true value of the parameter (here 3.14159).

The rightmost interval is based on all the available data and thus become shorter.

We can see that all the intervals include the true value of π. Using much more advanced features of mathematics it is possible to show that estimating p can be made much more effective if one uses a grid of lines instead of just parallel lines. It is 12 times more efficient! The idea of effectiveness is a rather common question in statistics. However, if we have a large amount of data (of good quality!), we might not have to worry about the effectiveness too much.

See also the document "Buffon's needle, part I" at www.ing-stat.se.

10.3.9 Summary of confidence intervals

Calculation of confidence interval for _μ_

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

$\bar{x}$ is the calculated average
s is the calculated standard deviation
t comes from the _t_-distribution (a table is used)
n number of values in the sample

Calculation of confidence interval for _p_

$$\hat{p} \pm t \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \qquad \text{where} \qquad \hat{p} = \frac{x}{n}$$

and x is the number of faulty units
       n is the number of units inspected
           (if possible use the computer macro)

Calculation of confidence interval for _λ_

$$\hat{\lambda} \pm t \cdot \sqrt{\hat{\lambda}}$$

where $\hat{\lambda}$ is the number of events found
             (if possible use the computer macro)

Calculation of confidence interval for _μ₁ – μ₂_

$$\bar{x}_1 - \bar{x}_2 \pm t \cdot \sqrt{\left[\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{(n_1 - 1) + (n_2 - 1)} \right] \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$\bar{x}_1$ and $\bar{x}_2$ are the calculated averages from the two samples
s_1 and s_2 are the calculated standard deviations from the two samples
n_1 and n_2 are the number of values in the two samples
t comes from the _t_-distribution with $(n_1 - 1 + n_2 - 1)$ degrees of freedom

Calculation of confidence interval for _p₁ – p₂_

$$\hat{p}_1 - \hat{p}_2 \pm t \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}} \qquad \text{where} \qquad \hat{p}_1 = \frac{x_1}{n_1} \qquad \hat{p}_2 = \frac{x_2}{n_2}$$

x_1 and x_2 are the number of faulty items or units in the two samples
n_1 and n_2 are the number of units in the two samples