Confidence Intervals

Krzysztof Podgorski

Department of Mathematics and Statistics

University of Limerick

September 21, 2009

Large sample distribution of sample mean

Confidence limits of the mean for large samples

- Now that we know the form of the sampling distribution of the mean we can

return to the problem of using a sample to define a range which we may reason-

ably assume includes the true value. (Remember that in doing this we are assum»

ing systematic errors to be absent.) Such a range is known as a confidence interval

and the extreme values of the range are called the confidence limits. The term

'confidence' implies that we can assert with a given degree of confidence, i.e. a

' certain probability, that the confidence interval does include the true value. The

` size of the confidence interval will obviously depend on how certain we want to

be that it includes the true value: the greater the certainty, the greater the interval

required.

— Figure 2.6 shows the sampling distribution of the mean for samples of size n. lf

i we assume that this distribution is normal then 95% of the sample means will lie in

the range given by:

iw1.96(o‘/~/h)<Y<ii+l.96(cr/~G) (2.6)

(The exact value 1.96 has been used in this equation rather than the approximate

Page 3

EIVQG Sélmp 9 COI'] I GIWCG IHIGYVH HSS OD Sélmp G

m€8I'l

(The exact value 1.96 has been used in this equation rather than the approximate

value, 2, quoted in Section 2.2. The reader can use Table A.1 to check that the pro—

portion of values between z= -1.96 and z = 1.96 is indeed 0.95,)

In practice, however, we usually have one sample, of known mean, and

we require a range for rr, the true value. Equation (2.6) can be rearranged to give

this:

Y - 1.96(o/W) < it < Y + 1.96(cr/NG) (2.7)

Equation (2.7) gives the 95% confidence interval of the mean. The 95% confidence

limits are Y x 1.96a/¤/H.

In practice we are unlikely to know o· exactly. However, provided that the sample is

large, o· can be replaced by its estimate, s.

Page 4

Example of computations using R

Finding confidence intervals for the mean for the nitrate ion

concentrations in Table 2.1.

#reading data

x=scan("Table2\_l.txt")

#setting the confidence level

CL=O.95

#computing confidence interval

n=length(x)

pm=sd(x)\*c(qnozm(0.025),qnorm(O.975))/sqrt(n)

CI:mean(x)+pm

Page 5

SmaII—Samp|e Case (n < 30)

If the data have a normal probability distribution and the sample

standard deviation s is used to estimate the population

standard deviation s, the interval estimate is given by:

X it to/2S/mv

where ta/2 is the value providing an area of a/2 in the upper tail

of a Student's tdistribution with n — 1 degrees of freedom.

Small sample confidence intervals

2.7 Confidence limits of the mean for small samples

As the sample size gets smaller, s becomes less reliable as an estimate of rx This can

be seen by again treating each column of the results in Table 2.2 as a sample of size

tive. The standard deviations of the 10 columns are 0.009, 0.015, 0.026, 0.021

0.013, 0.019, 0.013, 0.017, 0.010 and 0.018. We see that the largest value of s is

nearly three times the size ofthe smallest. To allow for this, equation (2.8) must be

modified.

For small samples, the confidence limits of the mean are given by

2 x r,,,,s/~/Z (2.9)

Student t-distribution

[ Tate as Values ¤t Hur eunmlenne intervals

2, i Degrees offreedom values et [fur ccnhdence interval of

E 95% 99%

' 7

I { 2 4.30 9,92

s 2.57 4.03

I 1 I0 2.23 3.17

{ 20 2.09 2.zs \_

E I so 2.01 2.6s

; i 100 1.98 2.63 1

{ The subscript (n — 1) indicates that r depends on this quantity, which is known as the

\_ number of degrees of freedom, d.f. (usually given the symbol v). The term \*degrees

I uf freedom' refers to the number of independent deviations (xp?) which are used in

` mlculating s, In this case the number is (n— 1), because when (n - 1) deviations are

F hauwn the last can be deduced since Z (x, — 1) = 0. The value of t also depends on the

T degree of confidence required. Some values of t are given in Table 2.3. A more com-

plete version of this table is given in Table A.Z in Appendix 2.

i For large rt, the values of z,, \_ , for confidence intervals of 95% and 99% respectively

l. arevery close to the values 1996 and 2.58 used in Example 2.6.1. The following exam-

P ple illustrates the use of equation (2.9).

Why "Student"?

Lovely day for a

Guinness n

William S. Gosset was a statistician , rr yi"

employed by the Guinness brewing

company which had stipulated that

he not publish under his own name. \_\_\_\_ A M V (

He therefore wrote under the pen ` ” { 1\*

name "Student." His main contribu- $ ? " °

tion was published in 1908. A 4 \_` v` .

Example of computations using R

Finding confidence intervals for the mean for the nitrate ion

concentrations in Example 2.7.1.

#Typing data in

x=c(lO2,97,99,98,lOl,lO6)

mean(x)

sd(x)

n;length(x)

#setting the confidence level

CL=O.95

#computing confidence interval

pm=sd(x)\*c(qt(0.025,n—l),qt(O.975,n—l))/Sqrt(n)

CI=mean(x)+pm