Chapter 8. Confidence Intervals.

In Chapter 5 we discussed what are called "point estimates"; the estimation of a parameter by a single number derived from the data, by using the principle of maximum likelihood estimation or some other method. And we discussed how each such estimate could be accompanied by a measure of its anticipated accuracy, such as its mean squared error. In practice this is often accomplished by presenting each estimate paired with an estimate of its standard deviation, termed its "standard error" (SE). If the estimate is unbiased, its standard error is the square root of its mean squared error, and if the estimate is approximately normally distributed, the probability it is within one SE of the parameter is about 0.65; within 2 SEs about 0.95.

An alternative approach to estimation is the use of Confidence Intervals. Briefly, the idea is that instead of estimating a parameter by a single number (and stating its SE), you estimate the parameter by stating an interval of possible values and giving the "level of confidence" you would attach to the proposition that the interval contains the true value of the parameter. The precise meaning of the term "level of confidence" is subtle and often misunderstood. We shall explain it terms of a simple example, which also illustrates a method (the "pivotal method") often used to find confidence intervals.

Suppose that $X_1, X_2, \ldots, X_n$ are data which may be assumed to be normally distributed with expectation θ. We will also suppose that the variance of the normal distribution is known, say to be 25; that is, the X_i are independent, $N(\theta,25)$. The MLE of θ is then $\overline{X}$, and since it is itself normally distributed, $N(\theta, 25/n)$, we know that regardless of the value of θ,

$$P(-1.96 \times (5/\sqrt{n}) \leq \overline{X} - \theta \leq 1.96 \times (5/\sqrt{n})) = 0.95\,.$$

If we do a simple algebraic manipulation of this set of inequalities (subtract $\overline{X}$ throughout and multiply by –1, being careful to reverse the inequalities), we arrive at:

$$P(\overline{X} - 1.96 \times (5/\sqrt{n}) \leq \theta \leq \overline{X} + 1.96 \times (5/\sqrt{n})) = 0.95\,.$$

This states that the probability the interval $[\overline{X} - 1.96 \times (5/\sqrt{n}), \overline{X} + 1.96 \times (5/\sqrt{n})]$ contains the parameter θ is 0.95, and on that basis the interval could be presented as a 95% confidence interval for θ. It is important to keep in mind what is being treated as random in this calculation: the data, not the parameter. It would be an error, admittedly a tempting one, to interpret this statement as meaning "the probability θ falls in the interval given the data values as observed is 0.95," which is a Bayesian statement. But no prior distribution for θ has been used and without a prior for θ such a statement is not justified. In this calculation, θ is fixed and the data are treated as random. The interval is random, and the probability 0.95 is the probability the random interval captures the fixed parameter. If the experiment were to be repeated many times with the same θ, a considerable number of different intervals would be generated and about 95% of them would in fact contain the parameter. The "level of confidence" 95% refers to the confidence we have in our procedure: 95% of the time the interval will contain the parameter, but in a given instance we cannot be sure; we can only state that we are using a generally reliable procedure. The possibility of misunderstanding is greatest after the data are observed and (for example) $n = 25$ and $\overline{X} = 347.3$ is computed, and the interval is found to be [345.34, 349.26]. We cannot state that the probability is 0.95 that $345.34 \leq \theta \leq 349.26$, only that we have 95% confidence in the procedure that led to this interval. To make an a posteriori (after the data) probability statement about θ it would be necessary to make an a priori statement about the probability distribution of θ, that is, to adopt a Bayesian approach as in Chapter 4. The virtue of confidence intervals is that they combine an assessment of accuracy with the estimate; their drawback is the propensity of the statement to be misinterpreted as a Bayesian statement when it is in fact somewhat weaker than that.

Confidence Interval Example. Eighty samples of size n = 25 were taken from an N(350, 25) distribution and the eighty 95% confidence intervals for the mean θ = 350 were computed. In this example, 5 of the 80 missed the target, about as expected.

X-bar	Lower	Upper	Covers?
348.17	346.21	350.13	
351.21	349.25	353.17	
350.15	348.19	352.11	
350.69	348.73	352.65	
348.51	346.55	350.47	
350.69	348.73	352.65	
352.94	350.98	354.90	No
350.35	348.39	352.31	
349.11	347.15	351.07	
348.77	346.81	350.73	
349.88	347.92	351.84	
349.40	347.44	351.36	
349.60	347.64	351.56	
349.39	347.43	351.35	
350.82	348.86	352.78	
350.38	348.42	352.34	
349.62	347.66	351.58	
349.77	347.81	351.73	
350.02	348.06	351.98	
349.81	347.85	351.77	
349.14	347.18	351.10	
349.10	347.14	351.06	
348.47	346.51	350.43	
349.73	347.77	351.69	
348.79	346.83	350.75	
350.43	348.47	352.39	
350.65	348.69	352.61	
349.29	347.33	351.25	
349.17	347.21	351.13	
350.00	348.04	351.96	
349.97	348.01	351.93	
349.60	347.64	351.56	
351.41	349.45	353.37	
350.86	348.90	352.82	
351.28	349.32	353.24	
351.14	349.18	353.10	
349.54	347.58	351.50	
350.59	348.63	352.55	
351.58	349.62	353.54	
350.93	348.97	352.89	

X-bar	Lower	Upper	Covers?
350.44	348.48	352.40	
349.52	347.56	351.48	
347.75	345.79	349.71	No
349.10	347.14	351.06	
349.44	347.48	351.40	
348.47	346.51	350.43	
348.60	346.64	350.56	
349.37	347.41	351.33	
351.37	349.41	353.33	
350.10	348.14	352.06	
349.15	347.19	351.11	
350.97	349.01	352.93	
350.46	348.50	352.42	
350.16	348.20	352.12	
351.29	349.33	353.25	
350.37	348.41	352.33	
348.92	346.96	350.88	
349.25	347.29	351.21	
349.31	347.35	351.27	
351.23	349.27	353.19	
349.99	348.03	351.95	
350.29	348.33	352.25	
350.88	348.92	352.84	
347.41	345.45	349.37	No
349.91	347.95	351.87	
348.53	346.57	350.49	
350.03	348.07	351.99	
352.13	350.17	354.09	No
349.99	348.03	351.95	
350.81	348.85	352.77	
350.14	348.18	352.10	
350.39	348.43	352.35	
349.50	347.54	351.46	
351.29	349.33	353.25	
349.74	347.78	351.70	
351.14	349.18	353.10	
349.89	347.93	351.85	
350.80	348.84	352.76	
347.98	346.02	349.94	No
349.07	347.11	351.03	