MATH 450: Mathematical statistics

Oct 13th, 2020

Lecture 13: Confidence intervals

Countdown to midterm: 16 days

Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	Chapter 7: Point Estimation
Week 7	**Chapter 8: Confidence Intervals**
Week 10	Chapter 9: Test of Hypothesis
Week 11	Chapter 10: Two-sample inference
Week 13	Regression

Chapter 8: Confidence intervals

Overview

8.1 Basic properties of confidence intervals (CIs)

- Interpreting CIs
- General principles to derive CI

8.2 Large-sample confidence intervals for a population mean

- Using the Central Limit Theorem to derive CIs

8.3 Intervals based on normal distribution

- Using Student's t-distribution

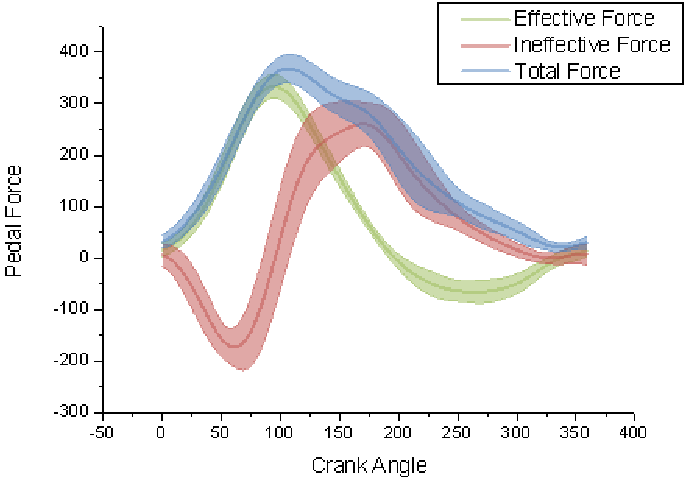
8.4 CIs for standard deviation

Framework

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $f(x, \theta)$
- In Chapter 7, we learnt methods to construct an estimate $\hat{\theta}$ of θ
- Goal: we want to indicate the degree of uncertainty associated with this random prediction
- One way to do so is to construct a *confidence interval* $[\hat{\theta} - a, \hat{\theta} + b]$ such that

$$P[\theta \in [\hat{\theta} - a, \hat{\theta} + b]] = 95\%$$

Confidence interval



Principles for deriving CIs

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution $f(x, \theta)$, then

- Find a random variable $Y = h(X_1, X_2, \ldots, X_n; \theta)$ such that the probability distribution of Y does not depend on θ or on any other unknown parameters.
- Find constants a, b such that

$$P\left[a < h(X_1, X_2, \ldots, X_n; \theta) < b\right] = 0.95$$

- Manipulate these inequalities to isolate θ

$$P\left[\ell(X_1, X_2, \ldots, X_n) < \theta < u(X_1, X_2, \ldots, X_n)\right] = 0.95$$

Confidence interval: Example 1

Problem

Suppose the sediment density (g/cm) of a randomly selected specimen from a certain region is normally distributed with mean μ *(unknown) and standard deviation* $\sigma = 0.85$*. A random sample of* $n = 25$ *specimens is selected with sample average* $\bar{X}$*.*

- *(Step 1) What is the distribution of*

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- *(Step 2) Find a number c such that*

$$P\left[-c < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < c\right] = 0.95$$

$\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997

- Assumptions:
  - Normal distribution
  - σ is known
- 95% confidence interval

  If after observing $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$, we compute the observed sample mean $\bar{x}$. Then

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

  is a 95% confidence interval of μ

z-critical value

NOTATION

z_α will denote the value on the measurement axis for which α of the area under the z curve lies to the right of z_α. (See Figure 4.19.)

For example, $z_{.10}$ captures upper-tail area .10 and $z_{.01}$ captures upper-tail area .01.

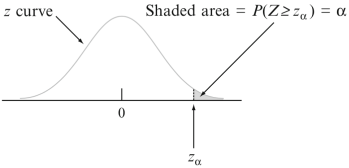


Figure 4.19 z_α notation illustrated

Since α of the area under the standard normal curve lies to the right of z_α, $1 - \alpha$ of the area lies to the left of z_α. Thus z_α *is the* $100(1-\alpha)$*th percentile of the standard normal distribution*. By symmetry the area under the standard normal curve to the left of $-z_\alpha$ is also α. The z_α's are usually referred to as **z critical values**. Table 4.1 lists the most useful standard normal percentiles and z_α values.

$100(1-\alpha)\%$ confidence interval

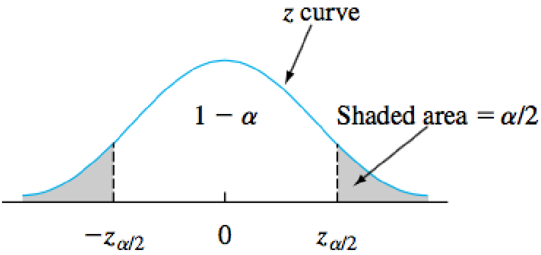


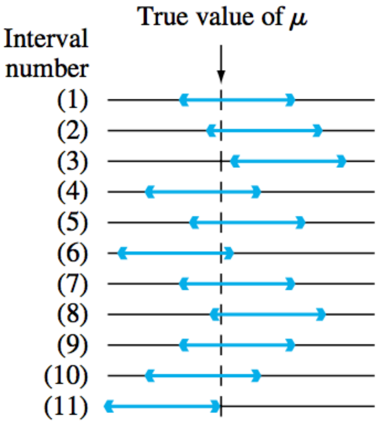
Figure 8.4 $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$

A **100(1 − α)% confidence interval** for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{8.5}$$

or, equivalently, by $\bar{x} \pm z_{\alpha/2} \cdot \sigma / \sqrt{n}$.

Interpreting confidence intervals



95% confidence interval: If we repeat the experiment many times, the interval contains μ about 95% of the time

Interpreting confidence intervals

- Writing
$$P[\mu \in (\bar{X} - 1.7, \bar{X} + 1.7)] = 95\%$$
is okay.
- If $\bar{x} = 2.7$, writing
$$P[\mu \in (1, 4.4)] = 95\%$$
is NOT okay.
- Saying $\mu \in (1, 4.4)$ with confidence level 95% is okay.
- Saying "if we repeat the experiment many times, the interval contains μ about 95% of the time" is perfect.

Example 2

Example

Assume that the helium porosity (in percentage) of coal samples taken from any particular seam is normally distributed with true standard deviation $\sigma = .75$.

- Compute a 95% CI for the true average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85.
- How large a sample size is necessary if the width of the 95% interval is to be .40?

One-sided CIs (Confidence bounds)

Example 1b: One-sided confidence interval

Problem

Suppose the sediment density (g/cm) of a randomly selected specimen from a certain region is normally distributed with mean μ (unknown) and standard deviation $\sigma = 0.85$. A random sample of $n = 25$ specimens is selected with sample average $\bar{X}$.

- *(Step 1) What is the distribution of*

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- *(Step 2) Find a number b such that*

$$P\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < b\right] = 0.95$$

CIs vs. one-sided CIs

CIs:

- $100(1-\alpha)\%$ confidence

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

- 95% confidence

$$\left(\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}\right)$$

One-sided CIs:

- $100(1-\alpha)\%$ confidence

$$\left(-\infty, \bar{x} + z_{\alpha}\frac{s}{\sqrt{n}}\right)$$

- 95% confidence

$$\left(-\infty, \bar{x} + 1.64\frac{s}{\sqrt{n}}\right)$$