Chapter 8: Confidence Intervals

MATH 450

October 12^{th}, 2017

Countdown to mid-term exam: 14 days

Week 1	Chapter 1: Descriptive statistics
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 13	Two-sample inference, ANOVA, regression

Overview

95% confidence interval

- 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

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

A **$100(1-\alpha)\%$ 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 $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

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

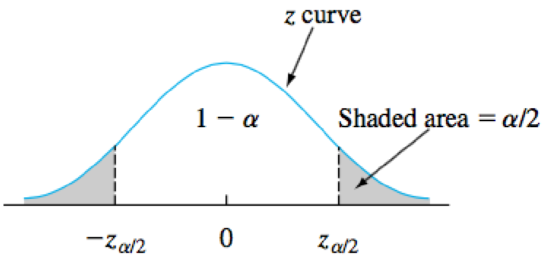


Figure 8.4 $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$

$\Phi(z)$

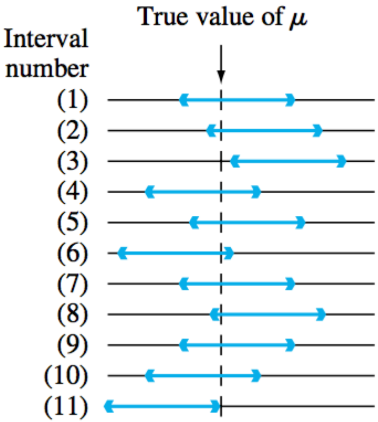
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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

- Section 8.1
  - Normal distribution
  - σ is known
- Section 8.2
  - ~~Normal distribution~~
    $\rightarrow$ Using Central Limit Theorem $\rightarrow$ needs $n > 30$
  - ~~σ is known~~
    $\rightarrow$ needs $n > 40$
- Section 8.3
  - Normal distribution
  - ~~σ is known~~
  - n is small

$\rightarrow$ Introducing t-distribution

Interpreting confidence interval



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

How NOT to interpret confidence interval

- Writing
$$P[\mu \in (\bar{X} - 1.7, \bar{X} + 1.7)] = 95\%$$
is ok.
- If $\bar{x} = 2.7$, writing
$$P[\mu \in (1, 4.4)] = 95\%$$
is NOT ok.
- Saying "there is a 95% chance that $\mu \in (1, 4.4)$" is generally NOT ok.
- Saying $\mu \in (1, 5.4)$ with confidence level 95% is ok.
- Saying "if we repeat the experiment many times, the interval contains μ about 95% of the time" is perfect.

Large-sample CIs of the population mean and proportion

Principles

- Central Limit Theorem
$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
is approximately normal when $n > 40$
- Moreover, when n is sufficiently large $S \approx \sigma$
- Conclusion:
$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$
is approximately normal when n is sufficiently large

If $n > 40$**, we can ignore the normal assumption and replace** σ **by** S

95% confidence interval

If after observing $X_1 = x_1$, $X_2 = x_2, \ldots$, $X_n = x_n$ ($n > 40$), we compute the observed sample mean $\bar{x}$ and sample standard deviation S. Then

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

is a 95% confidence interval of μ

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

If after observing $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$ ($n > 40$), we compute the observed sample mean $\bar{x}$ and sample standard deviation S. Then

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

is a 95% confidence interval of μ

One-sided CIs (Confidence bounds)

The central example

Problem

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

- *Find a number a such that*

$$P[-a < \bar{X} - \mu < a] = 0.95$$

- *Find a number b such that*

$$P[\bar{X} < b] = 0.95$$

One-sided CIs

A **large-sample upper confidence bound for** μ is

$$\mu < \bar{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$$

and a **large-sample lower confidence bound for** μ is

$$\mu > \bar{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$$

CIs vs. one-sided CIs

CIs:

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

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

- 95% confidence

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

One-sided CIs:

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

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

- 95% confidence

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

CI for a population proportion

Population and sample proportion

- Let p denote the proportion of successes in a population
- A random sample of n individuals is to be selected, and X is the number of successes in the sample
- Sample proportion

$$\hat{p} = \frac{X}{n}$$

Central limit theorem for sample proportion

- $E(\hat{p}) = p$, $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$
- When $n > 30$,
$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$
is approximately standard normal
- To derive 95% CI
$$P[-a < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < a] = 0.95$$

A confidence interval for a population proportion p with confidence level approximately $100(1-\alpha)\%$ has

$$\text{lower confidence limit} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + (z_{\alpha/2}^2)/n}$$

and (8.10)

$$\text{upper confidence limit} = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + (z_{\alpha/2}^2)/n}$$

where $\hat{q} = 1 - \hat{p}$.