Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

Statistical Confidence Intervals

So far, sampled data

2) Assumed it was generated from a probability distribution

3) Estimated the unknown parameters of the distribution

But the estimation problem is not yet solved

- How good is our estimate?
- How can we measure its quality?

We measure quality using **CONFIDENCE INTERVALS**

Confidence Statements

- Fortune Teller

"I believe the answer is 75.6 meters"

- Scientist

"I believe the answer is 75.6 meters plus or minus 2.0"

Confidence interval

Let θ = unknown population parameter

Let (a < b) be two numbers, where

the interval [a, b] contains the parameter θ

with confidence level of 1- α , then

[a, b] is a 1- a <u>confidence interval</u> for θ

Typically a, and b depend on random samples,
so they are random variables. Hence confidence interval means:

$$P(\theta \in [a,b]) = 1 - \alpha$$

Interpretation of confidence level: For θ = 0.05, we expect that 95% of all observed samples would give an interval that includes the true parameter θ .

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

How to construct confidence interval

Example: Say $X_1,\ldots X_n \sim N\left(\mu,\ \sigma^2\right)$ and μ is unknown. Suppose we need a confidence interval for μ :

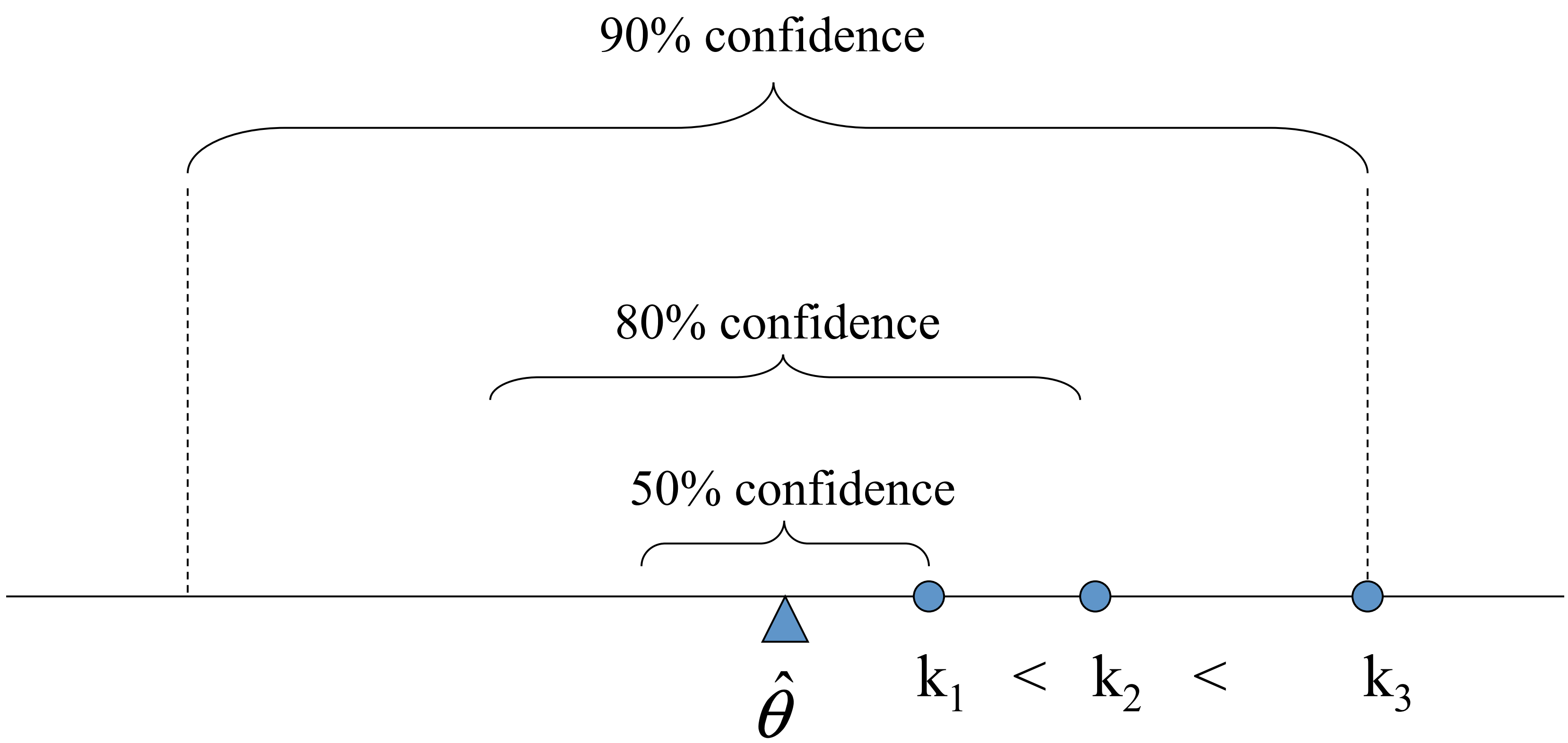
Maybe $\overline{\mathrm{X}}$ should be in the interval, perhaps in the middle?

Using this intuition, lets make the interval of the form

$$\left(\bar{X}-k,\ \ \bar{X}+k\right)$$

Now, how do we figure out what k should be?

Constructing a Confidence Interval



Confidence Interval – Example

Suppose we specify some $0 < \alpha < 1$ for a 1-α confidence level, then we want to find ***k*** such that

$$P(\bar{X} - k < \mu < \bar{X} + k) = 1 - \alpha$$

$$or$$

$$P(-k < \bar{X} - \mu < k) = 1 - \alpha$$

Example - continued

$$Note: \bar{X} \sim N(\mu, \sigma^2 / n) \quad \Rightarrow \quad \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim Z$$

$$P\left(\frac{-k}{\sigma / \sqrt{n}} < Z < \frac{k}{\sigma / \sqrt{n}} \right) = 1 - \alpha$$

$$P\left(Z < \frac{k}{\sigma / \sqrt{n}} \right) - P\left(Z < \frac{-k}{\sigma / \sqrt{n}} \right) = 1 - \alpha$$

$$\Phi\left(\frac{k}{\sigma / \sqrt{n}} \right) - \left(1 - \Phi\left(\frac{k}{\sigma / \sqrt{n}} \right) \right) = (1 - \alpha / 2) - (\alpha / 2)$$

$$\Phi\left(\frac{\mathrm{k}}{\sigma / \sqrt{\mathrm{n}}} \right) = 1 - \alpha / 2$$

$$\Rightarrow \quad \frac{\mathrm{k}}{\sigma / \sqrt{\mathrm{n}}} = z_{\alpha/2} = \Phi^{-1}(1 - \alpha / 2)$$

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

Z-interval

$$\frac{k}{\sigma/\sqrt{n}} = z_{\alpha/2} \quad or \quad k = \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

The two-sided confidence interval for m assuming σ known is called **z-interval** and is given by

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

Example 1: z-interval

You want to rent an unfurnished 1Br apartment for the next semester. The mean monthly rent for a sample of 20 apartments in the local newspaper is \$540 for a community where you want to move. Historically, the standard deviation of the rent is \$80. Find a 90% confidence interval for the mean monthly rent in this community.

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

Unknown variance

$$Note: \bar{X} \sim N(\mu, \sigma^2 / n)$$

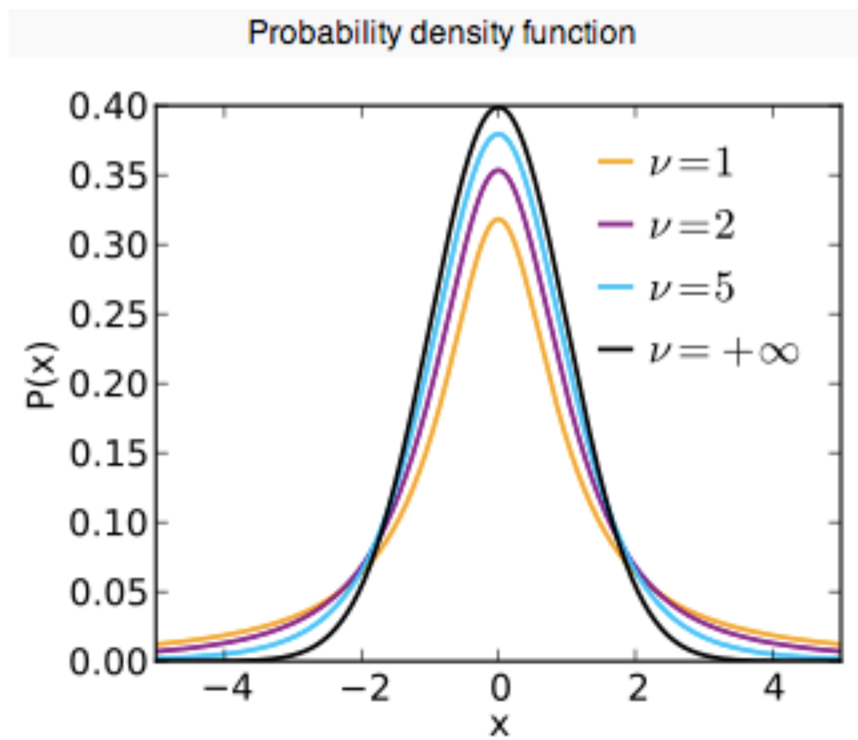
$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim Z$$

(but we don’ t know s^2)

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}$$

Student's t distribution

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$



Example - continued

$$P(-k < \mu - \overline{x} < k) =$$

$$P(-k < \overline{x} - \mu < k) =$$

$$P\left(\frac{-k}{s/\sqrt{n}} < T_{n-1} < \frac{k}{s/\sqrt{n}}\right)$$

t random variable
with *n-1* degree of freedom

T-interval

$$\frac{k}{s/\sqrt{n}} = t_{\frac{\alpha}{2},n-1} \qquad or \qquad k = \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2},n-1}$$

The two-sided confidence interval for m is assuming σ^2 unknown is called **t-interval** and is

$$\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2},n-1}, \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2},n-1} \right)$$

Example 2: t-interval

A study of the career path of hotel managers sent questionnaires to random sample of 61. According to these responses, the average time the general managers spent with their current company was 11.7 years and the sample standard deviation of time with the company is 3.2 years. Give a 99% confidence interval for the mean number of years general managers have spent with their current company.

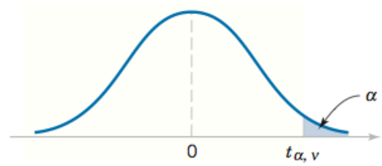


Table V Percentage Points $t_{\alpha,\nu}$ of the t Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

One Sided Confidence Interval

If you hear the words "confidence interval", the default interval is two sided (plus or minus).

In many cases a one-sided interval is more appropriate
Let θ = unknown population parameter

Let a be one number, where

the interval (-∞, a) contains the parameter θ

with confidence level of 1-α, then

(a, b) is a 1- α upper <u>confidence interval</u> for θ

Typically a depends on random samples and it is a random variables:

$$P(\theta \in (-\infty, a]) = 1 - \alpha$$

One sided confidence interval for t-random variable

A 1-a **upper confidence bound** for m is

$$\bar{x} + \frac{s}{\sqrt{n}} t_{\alpha, n-1}$$

A 1-a **lower confidence bound** for m is

$$\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha, n-1}$$

Example: 1-sided confidence interval

GT is running a survey on starting salaries for graduate students from GT. Based on survey of 20 students, the sample average salary is $60,000 with a sample standard deviation of $5500. What is the minimum starting salary that current students may use in negotiating their salaries when they take their first job with confidence level 0.9?

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

Confidence interval sample proportion

Our data is X ~ Binomial(n,p)

Point Estimation of p → $\hat{p} = \frac{x}{n}$

Sampling Distribution of $\hat{p}$

E($\hat{p}$)=p

V($\hat{p}$)=p(1-p)/n

By Central Limit Theorem (CLT),

$\hat{p}$ ~ N (p, p(1-p)/n)

Confidence Intervals for p

Using a normal approximation (sampling distribution), tests and confidence intervals are:

1-a C.I. for p: $\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n}$

1-a lower bound for p: $\hat{p} - z_{\alpha} \sqrt{\hat{p}(1-\hat{p})/n}$

1-a upper bound for p: $\hat{p} + z_{\alpha} \sqrt{\hat{p}(1-\hat{p})/n}$

Example

During election campaigns, it is important to assess the chance of winning before the election in order to identify weak points or areas with higher or lower electoral votes. For this many surveys are performed. Problem: estimating the probability of winning the election for two parties.

Randomly observe 50,000 individuals and their vote options. Suppose 25,264 of them vote for candidate A.

What is a 99% interval for the probability that individuals will vote for candidate A?

Outline

- Concept
- Construct confidence interval
- Two-sided z-interval
- Two-sided t-interval
- One-sided confidence interval: t-interval
- Confidence interval for sample proportion
- Sample size determination

Sample Size Determination

A confidence interval depends upon:

1. The confidence level (1 -*a)* through the *a/2* normal or t-distribution critical points – *a decreases so L increases, and therefore more **reliable***

2. The sample size has also an effect on the length of the confidence interval – *n increases so L increases, and therefore more **accurate***

$$L = \frac{2z_{\alpha/2}\sigma}{\sqrt{n}} \text{ (z-interval)}$$

$$L = \frac{2t_{\alpha/2}S}{\sqrt{n}} \text{ (t-interval)}$$

Sample Size Determination

Example: Say $X_1, \ldots X_n$ ~ $N(\mu, \ \sigma^2)$ and μ is unknown. Suppose we need a confidence interval for μ :

Goal: construct a 100(1- α)% C.I. for α with a specified length (L)

If we know σ^2, we use $(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}, \quad \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}})$

$$L = \frac{2 z_{\alpha/2} \sigma}{\sqrt{n}}$$

If (L, a) are specified, then there is a minimal n needed to assure the interval is no wider than L.

Example: determine sample size

Digital thermometer measurements. The standard error of digital thermometer measurement is $\sigma = 0.05$

We want confidence level $1 - \alpha = 0.90$

Suppose we want our estimation accuracy to be 0.1 degrees, how many measurements should be taken?

Sample Size Determination

$$L = \frac{2Z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} \Rightarrow n \geq \frac{4\sigma^2 Z_{\frac{\alpha}{2}}^2}{L^2}$$

$$\alpha = 0.10, \sigma = 0.05 \; and \; L = 0.1$$

$$n = \frac{4(1.645)^2(0.05)^2}{(0.1)^2}$$

$$= 2.7$$

We need to sample at least 3 observations