Parameter Estimation

Ananda Mysore

SJSU

Parameter Estimation

- In statistical terms, a **parameter** of a population is just a quantitative expression that characterizes it.
- The two most common parameters are the mean μ and the standard deviation σ.
- The corresponding **estimates** are the sample mean $\bar{x}$ and the sample standard deviation S.
- An effective way to express the uncertainty associated with these estimates is to use **confidence intervals**.

Confidence Intervals

- A **confidence interval** designates the bounds within which a parameter is expected to lie.
- For example, an expression for confidence that the mean μ would state that it is equal to an estimate $\bar{x}$, with some uncertainty $\pm\delta$ is:

$$\bar{x} - \delta \leq \mu \leq \bar{x} + \delta$$

- The 100(1-α)% **confidence level** expresses the probability that the parameter does lie within a specified interval, where α is the **significance level**.

$$P\{x - \delta \leq \mu \leq \bar{x} + \delta\} = 1 - \alpha$$

- For example, a 95% confidence level corresponds to $\alpha = \underline{???}$

Normal Distribution (Review)

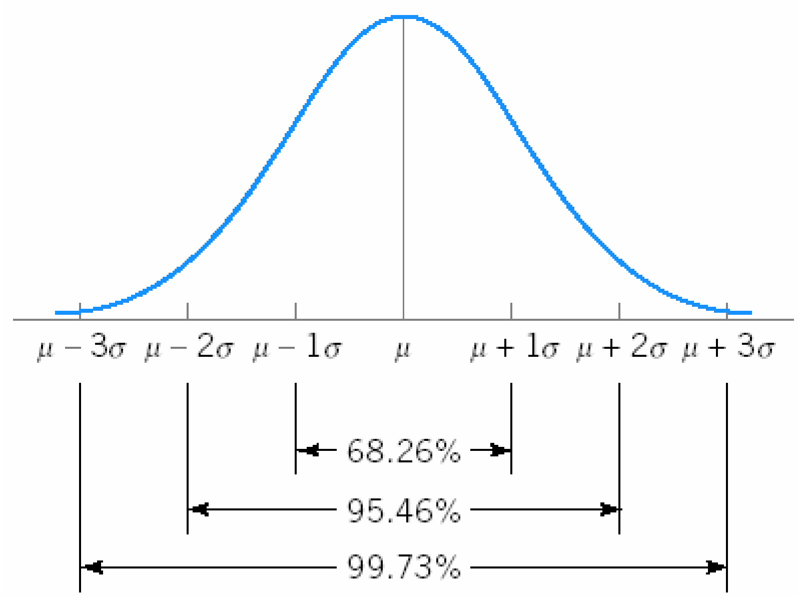
Definition

The **normal distribution** is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty \qquad \text{(2-21)}$$

The mean of the normal distribution is μ ($-\infty < \mu < \infty$) and the variance is $\sigma^2 > 0$.

- Relevant for "normal" random variables x.
- Most common and arguably most important distribution in applied statistics.
- Abbreviated notation $N(\mu,\sigma^2)$.



Central Limit Theorem (Review)

Definition: The Central Limit Theorem

If $x_1, x_2, \ldots, x_n$ are independent random variables with mean μ_i and variance σ_i^2, and if $y = x_1 + x_2 + \ldots + x_n$, then the distribution of

$$\frac{y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

approaches the $N(0, 1)$ distribution as n approaches infinity.

- The sum y of n independent random variables x has a distribution that is approximately normal, regardless of the distribution of each individual random variable x_i in the sum.
- The approximation improves as n increases.
- In many circumstances this theorem is often used to justify the assumption of a normal distribution regardless of underlying distribution

Standard Normal Distribution

- A standard normal distribution converts an $N(\mu,\sigma^2)$ random variable to an $N(0,1)$ random variable. Why is this useful?

$$z = \frac{x-\mu}{\sigma}$$

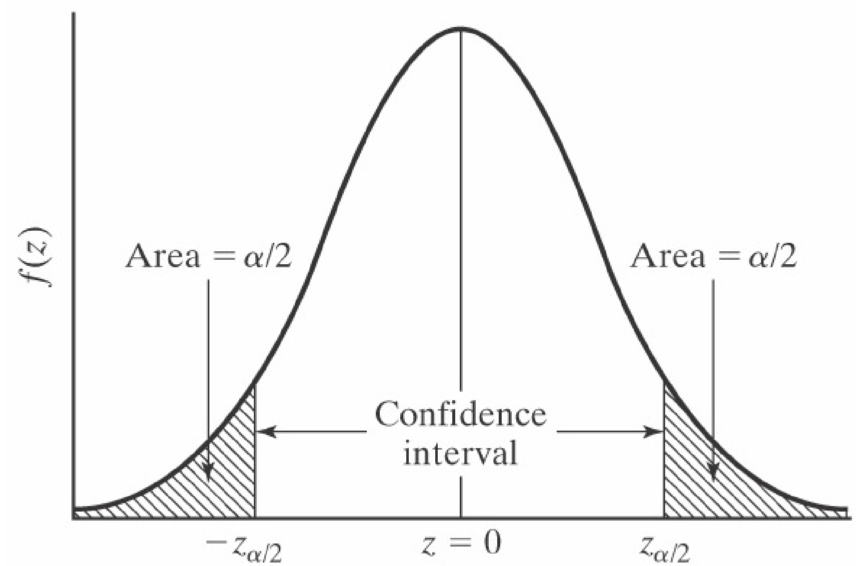
- The probability that the normal random variable x is less that or equal to a threshold a can be determined from the solution to the following integral expression.

$$P\{x \le a\} = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

- Results are tabulated (thankfully) based on a single input, z, in what is called a cumulative standard normal distribution table.
- Also: $P\{x \ge a\} = 1 - P\{x \le a\}$

Confidence and the Standard Normal Distribution

- For a standard normal distribution, the probability of lying within any specified portion of the distribution is explicitly known in terms of fractional area under the probability distribution curve.
- Using z and α provides a non-dimensional way of specifying probabilities, confidence intervals, and confidence levels.



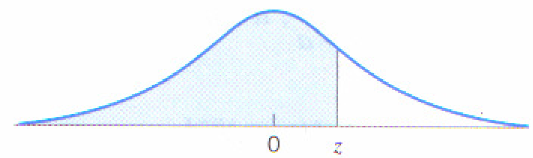
$$P\left\{-z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} = 1-\alpha$$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Standard Normal Distribution Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	
0.1	0.53983	0.54379	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	
0.3	0.61791	0.62172	0.62551	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	
0.4	0.65542	0.65910	0.62276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68438	
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	
0.7	0.75803	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	
0.8	0.78814	0.79103	0.79389	0.79673	0.79954	0.80234	0.80510	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83397	0.83646	0.83891
1.0	0.84134	0.84375	0.84613	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87285	0.87493	0.87697	0.87900	0.88100	0.88297
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89616	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91465	0.91621	0.91773
1.4	0.91924	0.92073	0.92219	0.92364	0.92506	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95448
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96637	0.96711	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$



Interpret as
"Area under curve"

Images from Introduction to Statistical Quality Control, 5th Ed.
by Douglas C. Montgomery, ©2005 John Wiley & Sons, Inc.

Confidence Interval with Known Variance

- In engineering applications, "population" variance σ is often never known exactly, but if the number of measurements is sufficiently high (n > 30), *s* serves as a good estimate.
- The *z*-statistic provides a way to express a confidence interval on the population mean, using the mean from a random sample of *n* observations on *x*:

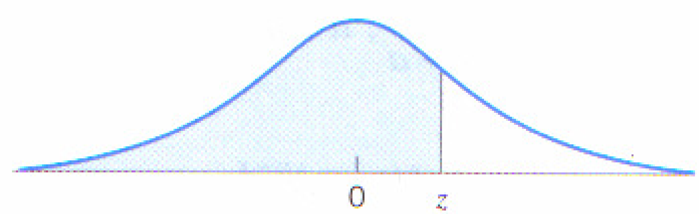
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \qquad \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Although the population mean is unknown, it is estimated to be within an uncertainty interval based on available measurements.
- What happens to the interval as σ, α and n change?

Standard Normal Distribution Example

Description	Symbol	Units	Value	Microsoft Excel Formula
Number of observations	n		36	
Sample average	x_bar	Ω	25	
Population standard deviation	σ	Ω	0.5	
Significance level	α		0.100	
Lower critical percent	$\alpha/2$		0.050	
Lower critical z-value	$-z_{\alpha/2}$		-1.645	NORMSINV(α/2)
Upper critical percent	$1-\alpha/2$		0.950	
Upper critical z-value	$z_{\alpha/2}$		1.645	NORMSINV(1-α/2)
Lower limit of confidence interval		Ω	24.86	
Upper limit of confidence interval		Ω	25.14	
Plus/minus difference from mean		Ω	0.14	
100(1-α)% confidence interval			90	100*(NORMSDIST($z_{\alpha/2}$)-NORMSDIST($-z_{\alpha/2}$))

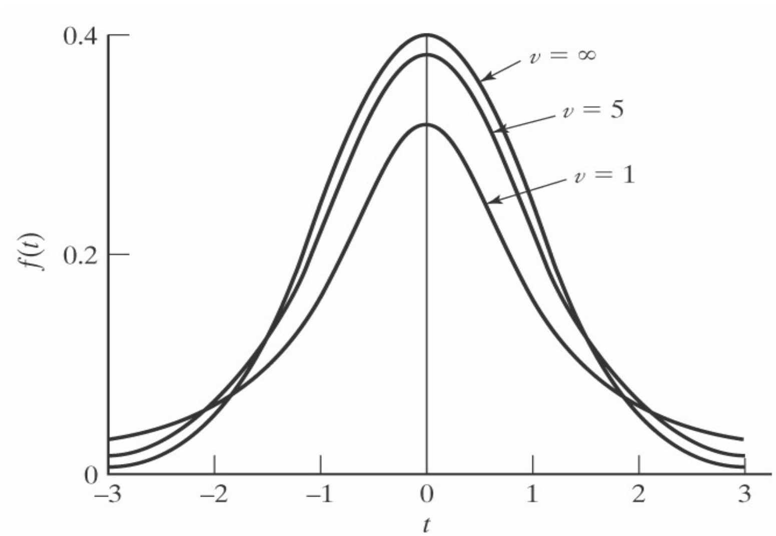
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

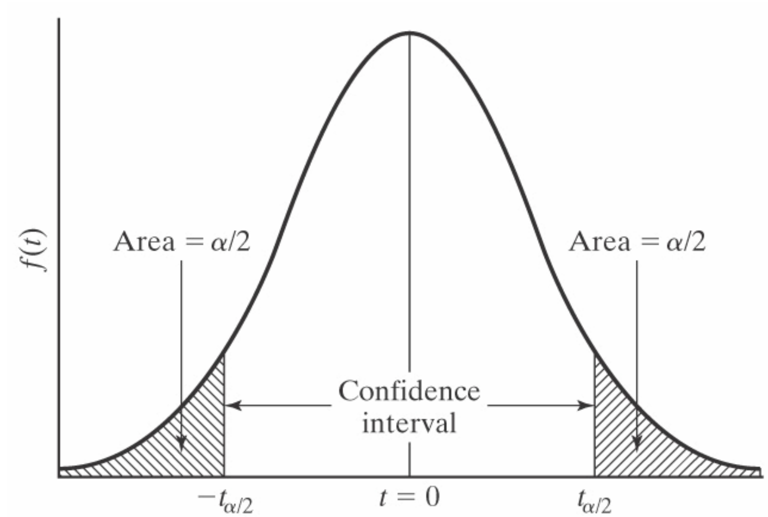


Confidence Interval with Unknown Variance

- In many practical applications, it is often difficult or impossible to obtain a large number of measurements to estimate variance.
- Like the *z*-statistic, the *t*-statistic also provides a way to express a confidence interval on the population mean, but the critical value $t_{\alpha/2}$ depends on degrees of freedom $\nu = n - 1$.

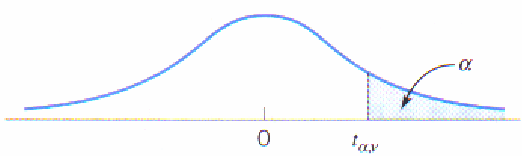
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$





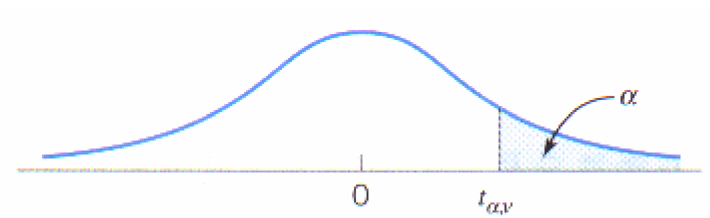
Using t-Distribution Tables

				α			
v	0.40	0.25	0.10	0.05	0.025	0.01	0.005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.49
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.20	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576

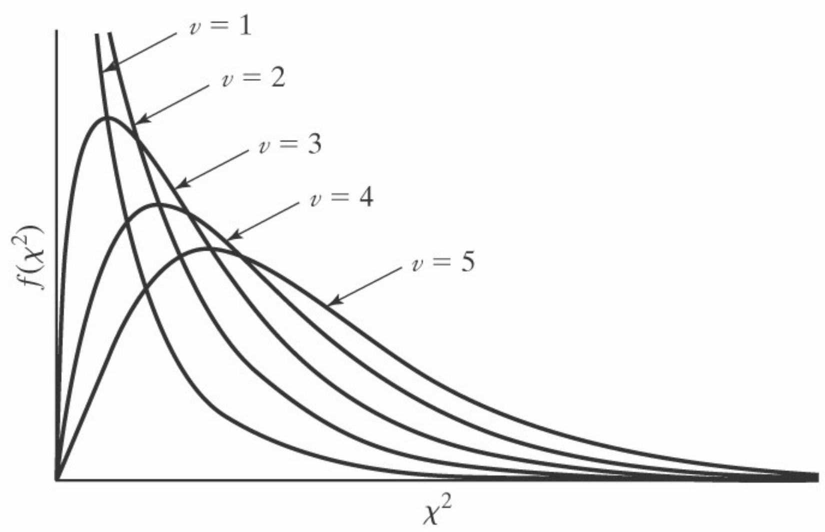


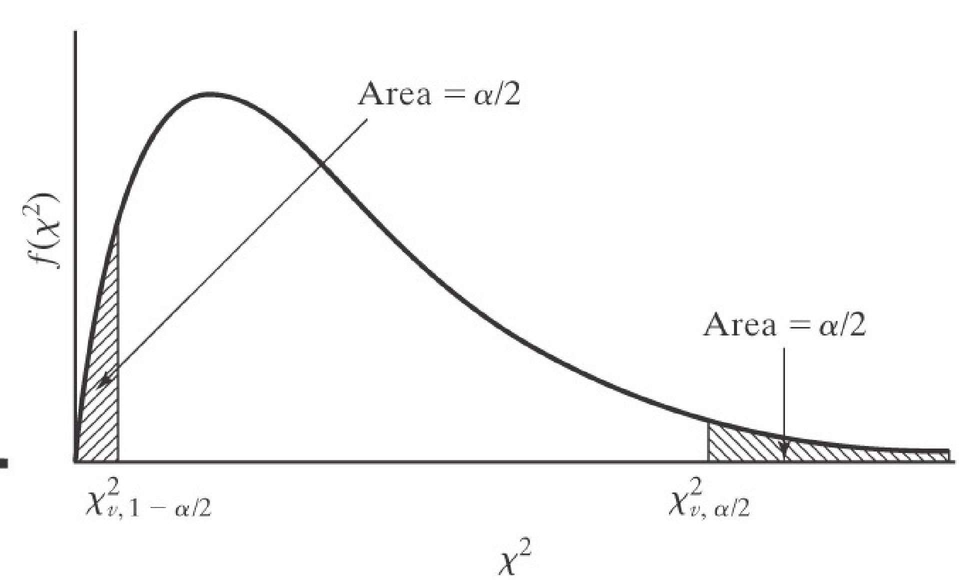
t-Distribution Example

Data	Description	Symbol	Value	Excel Formula
1250	Sample average	x_bar	1372	
1320	Number of observations	n	6	
1542	Sample standard deviation	s	114	STDEV(data)
1464	Confidence interval	$(1-\alpha)$	0.95	
1275	Significance level	α	0.050	
1383	Degrees of freedom	ν	5	n-1
	Lower critical percent	$\alpha/2$	0.025	
	Lower critical t-value	$-t_{\alpha/2,\nu}$	-2.571	$-TINV(\alpha,\nu)$
	Upper critical percent	$1-\alpha/2$	0.975	
	Upper critical t-value	$t_{\alpha/2,\nu}$	2.571	$TINV(\alpha,\nu)$
	Lower limit of confidence		1253	
	Upper limit of confidence		1492	
	Plus/minus difference from mean		119	



χ^2 (Chi-Squared) Distribution





- The chi-squared statistic is used analogously to the z-statistic and t-statistic for estimating population mean with respect to sample mean, but is used for the estimate of population variance σ^2 with respect to sample variance s^2.