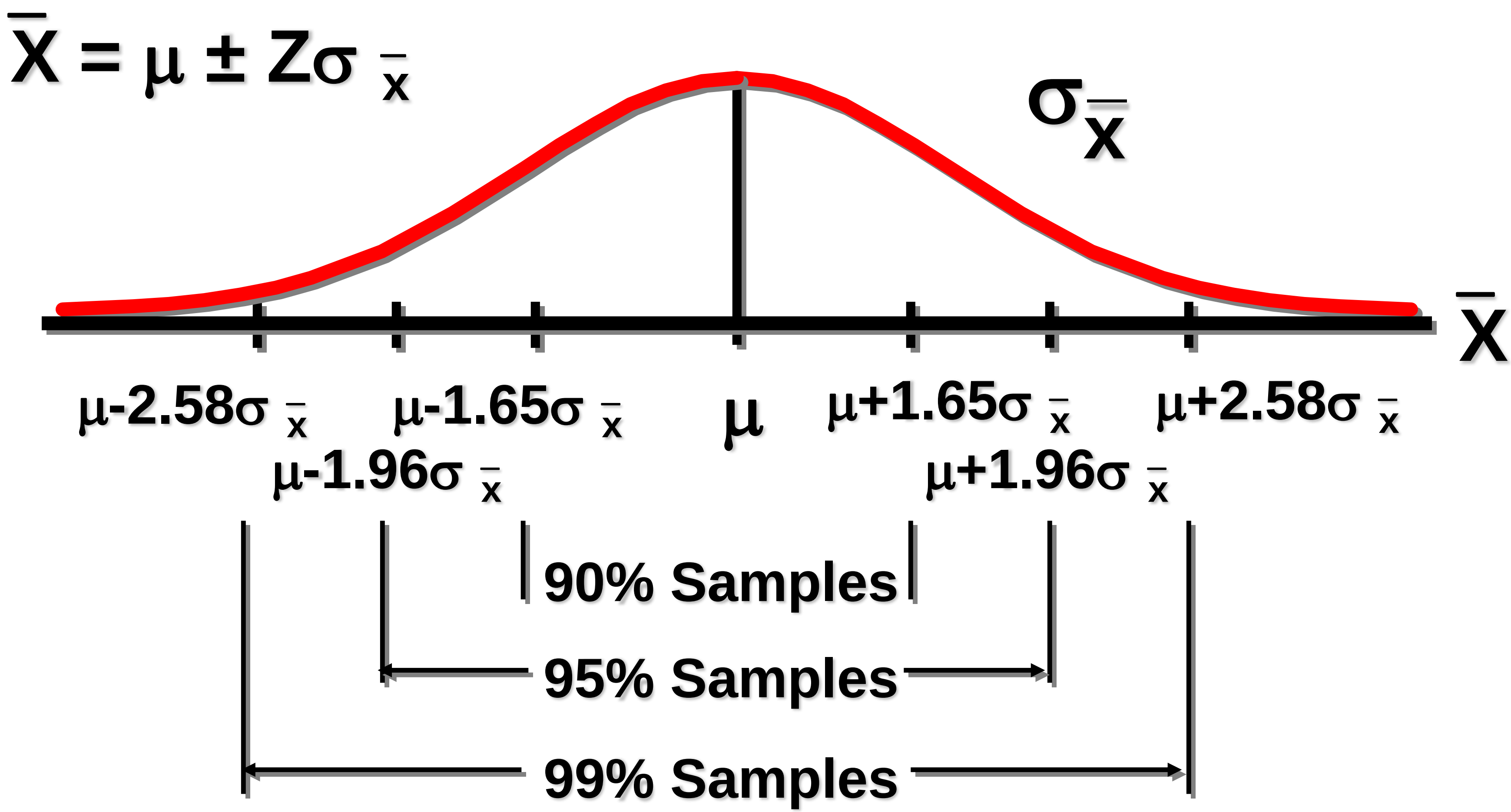
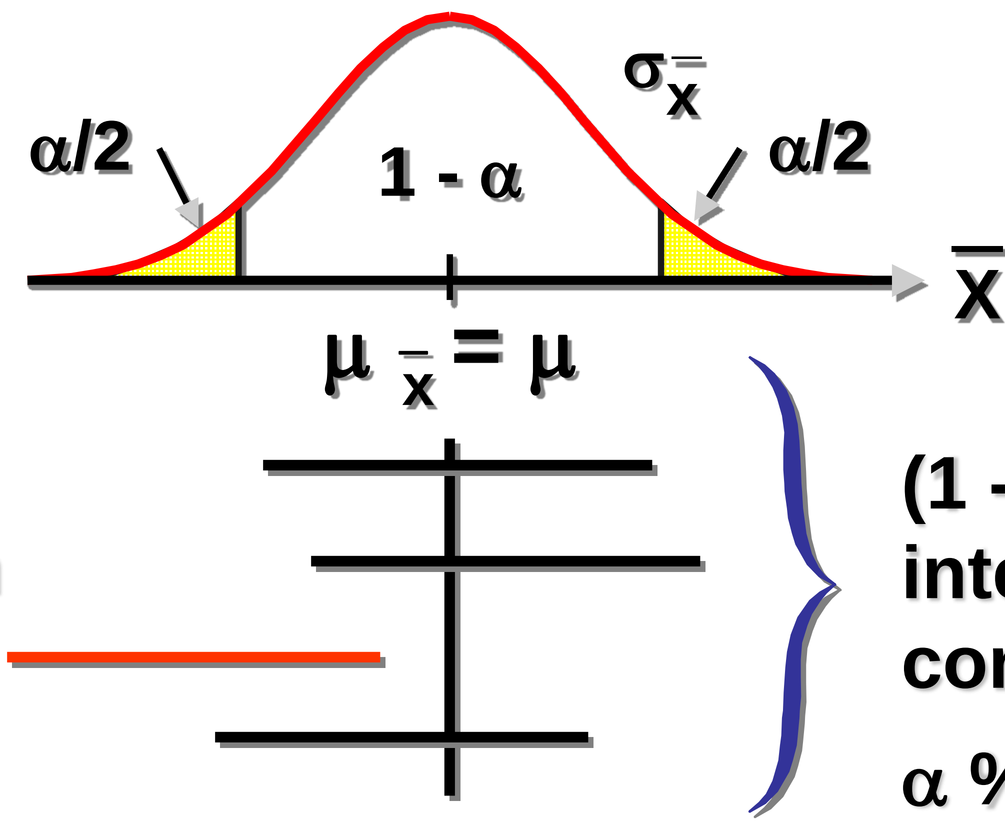
Confidence depends on interval (z)



Intervals & Confidence Level

Sampling Distribution of Mean



Intervals extend from
$\bar{X} - Z\sigma_{\bar{x}}$ **to**
$\bar{X} + Z\sigma_{\bar{x}}$

$(1 - \alpha)$ % of intervals contain μ.

α % do not.

Intervals derived from many samples

True or False?

$$P\left(\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95 \quad ?$$

false

μ is a fix number not a random variable!

Exercise

A large airline wants to estimate the average number of unoccupied seats per flight over the past year (we assume the number of occupied seats is normally distributed and has a known standard deviation equal to 4.1). The records of 225 flights are randomly selected from the files, and the number of unoccupied seats is noted for each flight. The sample mean results 11.6.

Compute the 95% confidence interval for average number of unoccupied seats per flight.

$$1-\alpha = 0.95 \quad 1-\alpha/2 = 0.975 \quad z_{1-\alpha/2} = 1.96$$

$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{225}}$$

$$11.6 \pm 1.96 \times \frac{4.1}{\sqrt{225}} = 11.6 \pm 1.96 \times 0.2733 = [11.0643;\ 12.1357]$$

We are <u>95% confident</u> that the interval

$[11.0643;\ 12.1357]$

contains the population mean.

Esercizio

A city should consider whether to suspend the movement of the cars next Sunday. For the purposes of the decision, it is important to observe the level of particulate matter in the air which is described by a random variable X with expected value E (X) = μ ("average level of powders in the air") and variance equal to $^2 = 540$.

Measurements made on 120 central units report an average value equal to 56.

- Assume X is Normally distributed

Esercizio

- Propose an estimator for μ ("average level of powders in the air") and evaluate the estimate in the observed sample.
- Compute the variance of the estimator in 1).
- Build a confidence interval for μ at the confidence level 90%;

$$90\%CI = \left[\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \right]$$

$$90\%CI = \left[56 - 1.645\sqrt{\frac{540}{120}}, 56 + 1.645\sqrt{\frac{540}{120}} \right] = [52.5104, 59.4896]$$

Second case:
$X \sim N(\mu,\sigma^2)$ μ, σ^2 unknown

REMEMBER:

a possible estimator for σ^2: $$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

S^2 is a biased estimator!!!

An unbiased estimator is S^2_c:

$$S_C^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$S_C^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

The previous construction of confidence interval was based on Normal standardized:

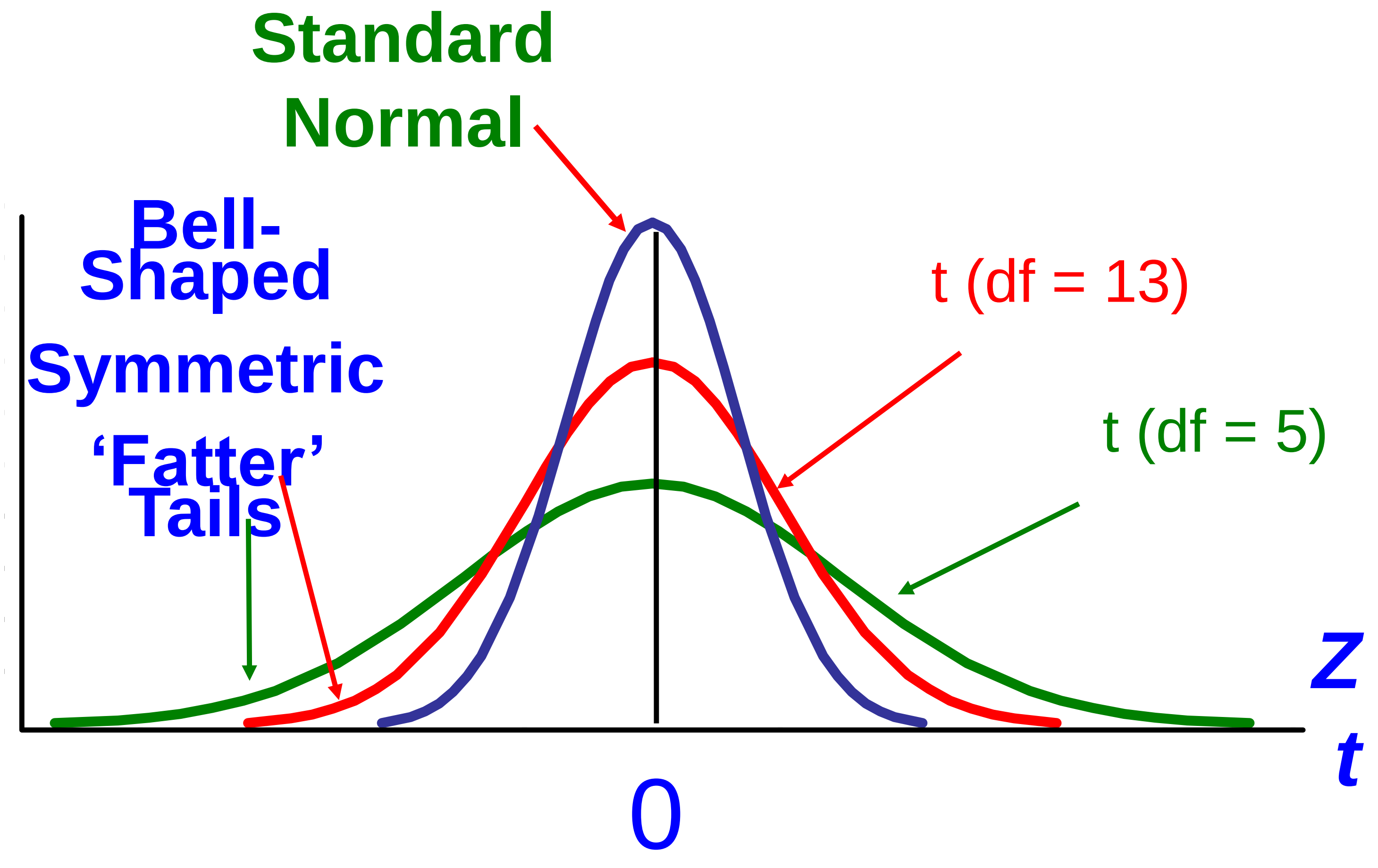
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{standard} \quad \text{Normal}$$

$$T_n = \frac{\bar{X}_n - \mu}{\frac{S_C}{\sqrt{n}}} \quad \text{is} \quad \text{not} \quad \text{standard} \quad \text{Normal!!!}$$

T_n is distributed as t-student’s distribution with (n-1)-degrees of freedom: n sample size

How does t-student’s distribuition look like?

Student's *t* Distribution



Student's *t* distribution

Student's *t* Distribution is a bell-shaped distribution, it is simmetric and it depends on one parameter indicated as n.

Bigger is n, closer t-students distribution to a Gaussian distribution

Distributions' values are available in statistical tables!

	Valori della funzione di ripartizione della distribuzione $T^{(k)}$						
k	**0.75**	**0.9**	**0.95**	**0.975**	**0.99**	**0.995**	**0.9995**
1	1.000	3.078	6.314	12.706	31.821	63.656	636.578
2	0.816	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.689
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.660
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	1.299	1.676	2.009	2.403	2.678	3.496
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
70	0.678	1.294	1.667	1.994	2.381	2.648	3.435
80	0.678	1.292	1.664	1.990	2.374	2.639	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	3.390
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.675	1.282	1.645	1.960	2.327	2.576	3.291

Confidence Intervals (σ Unknown)

- Assumptions
  - Population standard deviation (or variance) is unknown
  - Population is normally distributed
  - If not normal, use large samples (central limit theorem)
- Use student's *t* distribution

Confidence Intervals

(σ Unknown)

$$\bar{X} - t_{\alpha/2,\,n-1} \times \frac{S_C}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,\,n-1} \times \frac{S_C}{\sqrt{n}}$$

Degrees of Freedom (df)

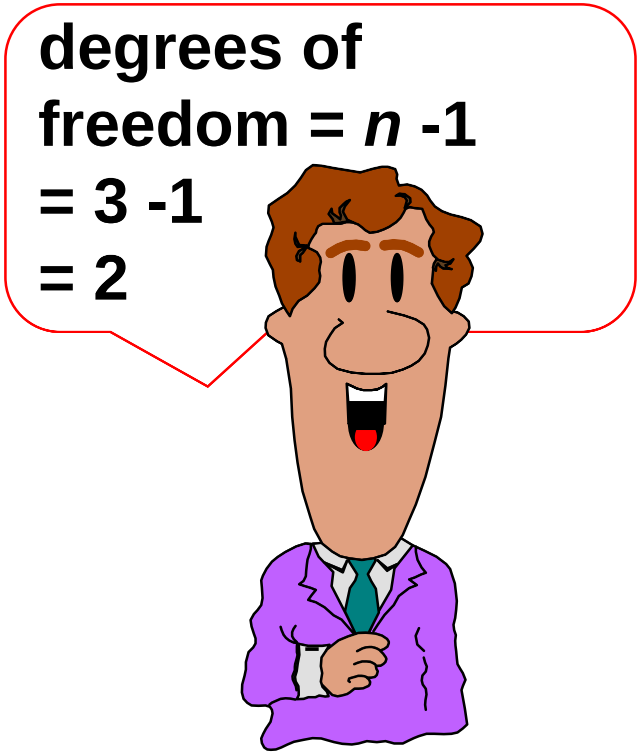
- Number of observations that are free to vary after sample mean has been calculated
- Example
  - Mean of 3 numbers is 2

    X_1 = 1 (or any number)

    X_2 = 2 (or any number)

    X_3 = 3 (cannot vary)

    mean = 2



Example: Interval Estimation
σ Unknown

A random sample of n = 25 has $\bar{x} = 50$ and $S_C = 8$.

Set up a 95% confidence interval estimate for μ (assume $X \sim N(\mu,\sigma^2)$)

$$\bar{X} - t_{\alpha/2,n-1} \times \frac{S_C}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1} \times \frac{S_C}{\sqrt{n}}$$

$$50 - 2.064 \times \frac{8}{\sqrt{25}} < \mu < 50 + 2.064 \times \frac{8}{\sqrt{25}}$$

$$95\%CI = [46.69, \quad 53.30]$$

Confidence Intervals for population mean μ

- $X \sim N(\mu,\sigma^2)$ σ^2 known

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- $X \sim N(\mu,\sigma^2)$ σ^2 UNknown

$$\left(\bar{x}_n - t_{1-\alpha/2,n-1} \frac{s_C}{\sqrt{n}} ; \bar{x}_n + t_{1-\alpha/2,n-1} \frac{s_C}{\sqrt{n}} \right) \qquad S_c^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)}{n-1}$$

Confidence Intervals for population mean μ

- X any distribution $E(X)=\mu$

  $Var(X)=\sigma^2$ known – n large

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- σ^2 UNknown- n large

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{s_C}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{s_C}{\sqrt{n}} \right) \qquad S_c^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)}{n-1}$$

Confidence Intervals for population proportion p

- X ~Bernoulli(p)
- E(X)=p Var(X)=p(1-p)   n large

$$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \; \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Exercise

The following table is a random sample of number of shares of 15 stocks traded on the NASDAQ market as reported in the Wall Street Journal.

Suppose the number of shared traded by each stock is normally distributed.

Set un a 95% confidence interval estimate of the population average number of shares traded for a company on the NASDAQ market.

Company	Shares	Company	Shares
Air Method	4,7	Fd GP	116,7
Brio Tech	34,3	Photronics	433,1
Concrd Camm	40,9	Prodigy	69,4
Exe Tech	43,3	Safenet	81,0
Good Guys	18,2	Tractor Supply	9,7
IKOS	27,3	Trikon	0,600
JMAR	27,6	World Com	23,1
Norstan	23,0		

$$\sum_{i=1}^{15} x_i = 952.9 \quad \sum_{i=1}^{15} x_i^2 = 220313.5$$

$$\bar{x} = 63.53$$

$$s^2 = \frac{\sum_{i=1}^{15} x_i^2}{14} - \left(\frac{15}{14}\right)\bar{x}^2 = 11412.78$$

$$t_{1-\alpha/2}^{14} = 2.145$$

$$95\%CI = \left(\bar{x} - t_{\alpha/2}^{n-1}\frac{s}{\sqrt{15}}; \bar{x} - t_{\alpha/2}^{n-1}\frac{s}{\sqrt{15}}\right) = \left(4.36\,,122.70\right)$$

Exercise

Let X be the waste recycled per person per day, and let consider a random sample of 13 American adults, and consider the following result

$$\sum_{i=1}^{13} x_i = 55.9 \quad \sum_{i=1}^{13} x_i^2 = 241.45$$

Assume the variable X is normally distributed, with unknown mean μ and unknow standard deviation σ.

1. Provide an unbiased estimator of μ and the corresponding estimate
2. Provide an unbiased estimator of σ^2 and the corresponding estimate
3. Provide a 90% confidence interval for μ.

1. Unbiased estimator of μ

Sample mean or $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

estimate for $\bar{x} = \dfrac{55.9}{13} = 4.3$

2. Unbiased estimator of σ²

$$S_C^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

estimate for $s_C^2 = \dfrac{1}{12}\left(241.45 - 13 \times 4.3^2\right) = \dfrac{1.08}{12} = 0.09$

Confidence Interval

1. The point estimate is $\bar{X}$ = 4.3 pounds
2. The estimate standard error of mean is $\frac{s_C}{\sqrt{13}} = \frac{0.3}{\sqrt{13}} = 0.0832$

$$1-\alpha = 0.90 \quad \alpha/2 = 0.05 \quad t_{0.5\,,\,12} = 1.78$$

$$4.3 - 1.78 \times 0.0832 \; < \; \mu \; < 4.3 + 1.78 \times 0.0832$$
$$4.1519 < \mu < 4.4481$$

With 90% confidence, you can say the mean waste recycled per person per day is between 4.15 and 4.45 pounds.

Exercise

There is a suspect that young children living in Harlem, NY, have a low risk of hypertension due to a low blood pressure.

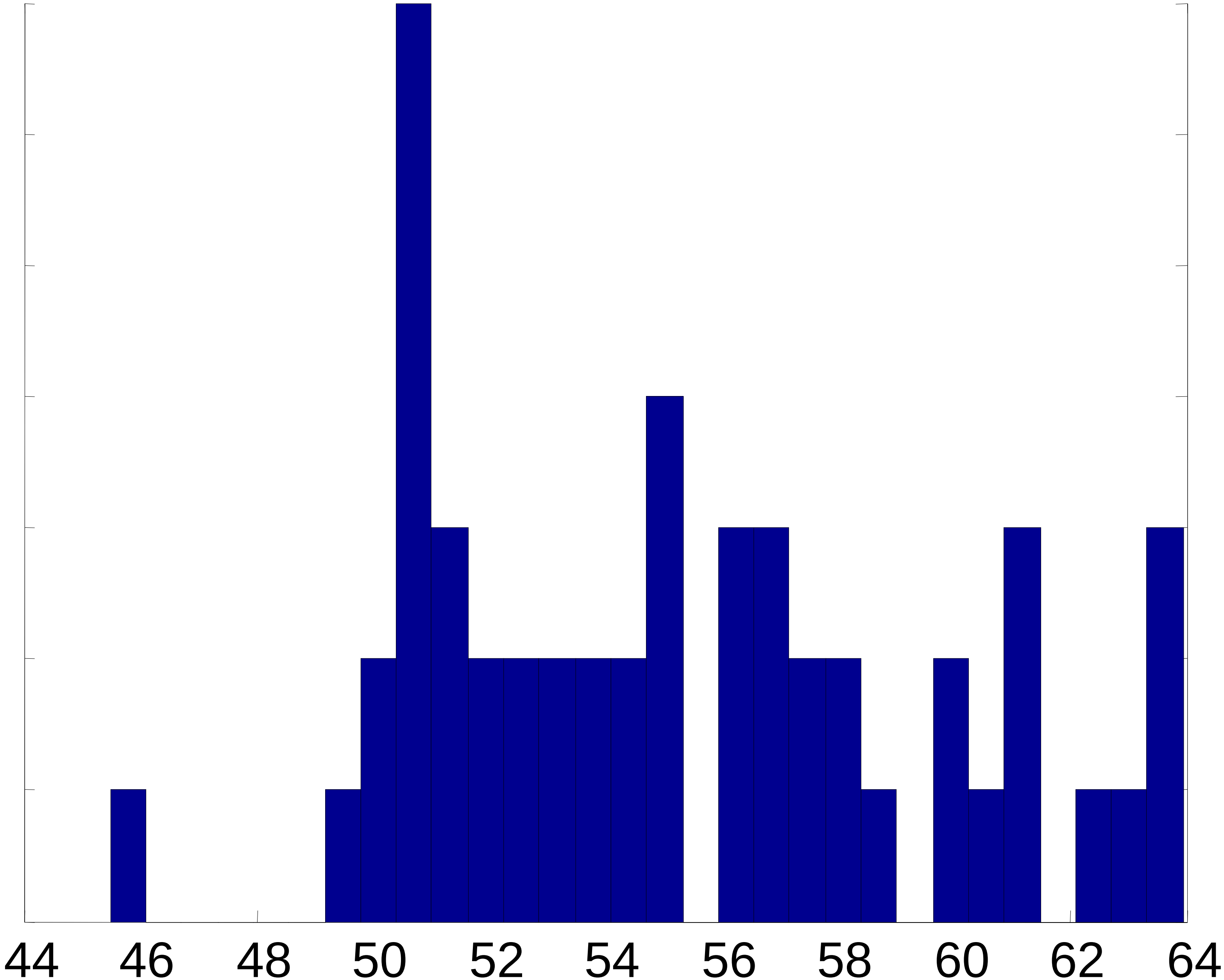
Let blood pressure be a random variable with unknown distribution, unknown expected value m and unknown variance s^2.

50 children are randomly selected and their blood pressure measured, the following values are observed:

$$\sum_{i=1}^{50} x_i = 2766.3 \quad \sum_{i=1}^{50} (x_i - \bar{x})^2 = 998.7181$$

1. Provide an unbiased estimator and the corresponding estimate of population mean μ
2. Evaluate variance of estimator proposed at point 1)
3. Provide a 90% Confidence interval for μ.

Sample distribution



1. Unbiased estimator of μ

Sample mean or $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

estimate for $\bar{x} = 55.326$

Unbiased estimator of σ^2

$$S_C^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}$$

estimate for σ^2

$$s_C^2 = \frac{998.7181}{49} = 20.382$$

2. Variance of sample mean

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Estimator of sample mean variance

$$V\hat{a}r(\overline{X}) = \frac{\hat{\sigma}^2}{n} = \frac{s_C^2}{n}$$

estimate for sample mean variance

20.382/50=0.4076

How can we build confidence intervals
If we DO not have distribution of X?

For large n, thanks to central limit theorem, we can apply the formula for confidence interval without normal assumption of X

For large n, distribution of $\bar{X}$ is approximately normal

$$T_n = \frac{\bar{X}_n - \mu}{\frac{S_C}{\sqrt{n}}}$$

T_n is distributed as t-student's distribution with 49-degrees of freedom for large n,

t-student distribuition is asymptotically normal, T_n is a standard normal distribution

Confidence Intervals for μ

unknown distribution, σ Unknown, n large

$$\bar{X} - z_{1-\alpha/2} \times \frac{S_C}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \times \frac{S_C}{\sqrt{n}}$$

$$\bar{X} - z_{1-\alpha/2} \times \frac{S_C}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \times \frac{S_C}{\sqrt{n}}$$

$$55.3261 - 1.645 \times \frac{\sqrt{20.382}}{\sqrt{50}} < \mu < 55.3261 + 1.645 \times \frac{\sqrt{20.382}}{\sqrt{50}}$$

$$90\% CI = [54.2758, \quad 56.3764]$$

Exercise

The fill on a random sample of 150 bottles of Pepsi from bottling machine yielded a sample mean of 19.5 oz, and a sample standard deviation s=1.4 oz.

1. Give a point estimate for the average fill on all bottles
2. Give a 95% confidence interval estimate for the average fill on all bottles
3. Give a 99% confidence interval estimate for the average fill on all bottles

1. Point estimate for μ: 19.5oz

$$S_C^2 = \frac{n}{n-1} S^2$$

$$s_C = \sqrt{\frac{n}{n-1}} s = \sqrt{\frac{150}{149}} \times 1.4 = 1.409 \approx 1.4$$

2. 95%CI:

   $19.5 \pm 1.96 \times 1.409/\sqrt{150}$

   95%CI=[19.2745, 19.7255]

3. 99%CI:

   $19.5 \pm 2.576 \times 1.409/\sqrt{150}$

   99%CI=[19.2036, 19.7964]

Example

A random sample of 35 airfare prices (in dollars) for a one-way ticket from Atlanta to Chicago. Find a point estimate for the population mean, μ. (suppose X ?'airfare price' are Normally distributed

99	102	105	105	104	95	100	114	108	103	94	105
101	109	103	98	96	98	104	87	101	106	103	90
107	98	101	107	105	94	111	104	87	117	101	

The sample mean is

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3562}{35} = 101.77$$

The point estimate for the price of all one way tickets from Atlanta to Chicago is $101.77.

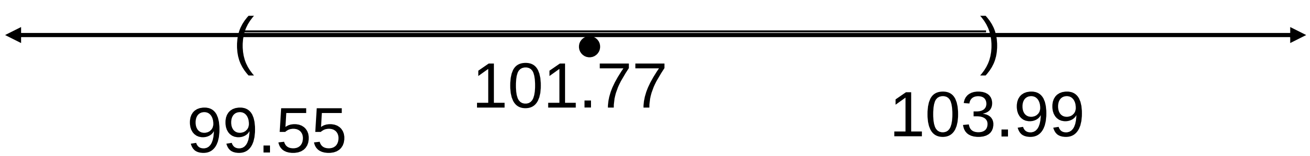
$$\sum_{i=1}^{35} x_i^2 = 364032$$

$$S_C^2 = \frac{1}{34}\left(\sum_{i=1}^{35} X_i^2 - 35 \times \overline{X}^2\right)$$

$$s_C^2 = \frac{1}{34}\left(364032 - 35 \times 101.77^2\right) = 45.0691$$

$$95\%CI = \left[101.77 - 1.96 \times \sqrt{\frac{45.0691}{35}};\quad 101.77 + 1.96 \times \sqrt{\frac{45.0691}{35}}\right]$$

$$95\%CI = \left[99.5459;\quad 103.9941\right]$$



With 95% confidence, you can say the mean one-way fare from Atlanta to Chicago is between \$99.5459 and \$103.9941.

Confidence interval for a population proportion

An interval of values, computed from sample data, that is almost sure to cover the true population proportion.

A commission on crime is interested in estimation the proportion of crimes to firearms in an area with one of the highest crime rates in a country. The commission selects a random sample of 300 files of recently committed crimes in the area and determines that a firearm was reportedly used in 180 of them.
The best estimator of the population proportion *p* is sample proportion given by the sample proportion

$$\hat{p} = \frac{180}{300} = 0.6$$

The Rule for Sample Proportions

If numerous samples or repetitions of size *n* are taken, the frequency curve of the sample proportions $(\hat{p})$ from various samples will be *approximately bell-shaped*. The *mean* of those sample proportions will be *p* (the population proportion). The *standard deviation* will be:

$$\sqrt{\frac{p(1-p)}{n}}$$

How can we estimate a confidence intervals for the parameter *p*?

Central limit theroem tells us that the sample proportion is normally distributed with mean *p* and variance p(1-p)/n

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n} \right)$$

Formula for a 100×(1-α)% Confidence Interval for the Population Proportion

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$95\% CI = \left(\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

$$\left(0.6 - 1.96 \times \sqrt{\frac{0.6 \times 0.4}{300}}, 0.6 + 1.96 \times \sqrt{\frac{0.6 \times 0.4}{300}} \right)$$

(0.5446, 0.6554)

We are 95% confident that the interval from 0.54 to 0.66 contains the true proportion of crimes committed in the area that are related to firearms.
That is, in repeated construction of 95% confidence intervals, 95% of all samples would produce confidence interval that enclose *p*.

Example

A random sample of 400 voters showed 32 preferred candidate A. Set up a 99% confidence interval estimate for *p*

$$99\%CI=\left(\hat{p}-2.576\times\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\hat{p}+2.576\times\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$$99\%CI=\left(0.08-2.576\times\sqrt{\frac{0.08(1-0.08)}{400}},\hat{p}+2.576\times\sqrt{\frac{0.08(1-0.08)}{400}}\right)$$

99%CI=[0.045, 0.115]

Example: Political Science

Political Science, Inc. (PSI) specializes in voter polls and surveys designed to keep political office seekers informed of their position in a race. Using telephone surveys, interviewers ask registered voters who they would vote for if the election were held that day.

In a recent election campaign, PSI found that 220 registered voters, out of 500 contacted, favored a particular candidate. PSI wants to develop a 95% confidence interval estimate for the proportion of the population of registered voters that favors the candidate.

$$\hat{p} \pm 1.96 \times \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}}$$

$$.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{500}}$$

where: n = 500, $\hat{p}$ = 220/500 = 0.44, $z_{\alpha/2}$ = 1.96

0.44 ± .0435=[0.3965, 0.4835]

PSI is 95% confident that the proportion of all voters that favors the candidate is between 0.3965 and 0.4835.

Thought Questions

In a Yankelovich Partners poll of 1000 adults (*USA Today*, 20 April 1998), 45% reported that they believed in "faith healing." Based on this survey, a "95% confidence interval" for the proportion in the population who believe is about 42% to 48%. If this poll had been based on 5000 adults instead, do you think the "95% confidence interval" would be wider or narrower than the interval given?

You can see that if you increase the sample size or reduce the population standard deviation, you will get a smaller standard error and a narrower confidence interval.

Conversely, if you reduce the sample size or increase the population standard deviation, you will get a larger standard error and a bigger interval.

Length of a confidence interval

- For population mean

$$l = 2z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- For populaton proportion

$$l = 2z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- With α fixed, if n increase, confidence interval length decreases
- With α fixed, if n decreases, confidence interval length increases
- With n fixed, if $(1-\alpha)$ increases, $z_{1-\alpha}$ increases and confidence interval length increases
- With n fixed, if $(1-\alpha)$ decreases, $z_{1-\alpha}$ decreases and confidence interval length decreases

Exercise

In a random sample of 500 families owning television sets in the city of Hamilton, Canada, it was found that 340 owned color sets. Find a 92% confidence interval for the actual proportion of families in this city with color sets

$$92\%CI = \left[\frac{340}{500} - 1.75\sqrt{\frac{0.68 \times 0.32}{500}}, \quad \frac{340}{500} + 1.75\sqrt{\frac{0.68 \times 0.32}{500}}\right] = [0.6435, \quad 0.7165]$$

Exercise

President Bush wants to estimate the percentage of Americans who do not approve his war in Iraq. 200 American citizens randomly chosen are interviewed and 120 answer that they oppose the war. Compute a 95% Confidence Interval (**CI**) for the proportion of American against the war.

1. Confidence Interval (**CI)** equation

$$95\%CI = \left[\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \, , \, \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

2. 95% numerical CI

$$z_{1-\alpha/2} = z_{0.975} = 1.96$$

$$95\%CI = [0.5321 \, , \, 0.6679]$$

Sample Size

Too Big:
Requires too much resources

Too small:

Won't do the job

Exercise

Children’s weekly water consumption (X, in litres) is a variable that is normally distributed with dispersion $\sigma = 3.5$. In a simple random sample of 45 children, the weekly average consumption of water is 17.72 litres.

a) Obtain a confidence interval (90%) for the expected consumption.

b) If water consumption is not normally distributed, what happens with the confidence

interval?

c) Determine the minimum sample size to estimate the expected consumption with a margin of error of 1 litre and a confidence level of 95%.

Solution

$$90\% CI = \left[17.72 - 1.64 \frac{3.5}{\sqrt{45}}, 17.72 + 1.64 \frac{3.5}{\sqrt{45}} \right] = [16.8643, 18.5757]$$

$$l_\alpha = 2 z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad \sqrt{n} \geq z_{1-\alpha/2} \frac{\sigma}{l_\alpha}$$

$$\sqrt{n} \geq z_{1-\alpha/2} \frac{\sigma}{l_\alpha} = 1.96 \times 3.5 = 6.86$$

N almeno 46

Example: Minimum Sample Size

You wish to estimate the proportion of fatal accidents that are alcohol related at a 99% level of confidence. Find the minimum sample size needed to be be accurate to within 2% of the population proportion. Use a preliminary estimate of *p* = 0.235.

$$z_{1-\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 0.05 \qquad \sqrt{n} = z_{1-\alpha/2}\,\frac{\sqrt{\hat{p}\left(1-\hat{p}\right)}}{0.02}$$

$$n = 2.576^2\,\frac{0.235\times(1-0.235)}{0.02^2} = 2982.37 \cong 2983$$

With a preliminary sample you need at least ***n*** = 2983 for your sample.

Sporting Survey

A National Sporting Goods Association survey of 1,000 adult Americans indicated that 34% of Americans surveyed used home exercise equipment during the past year. You are asked to prepare a 95% confidence interval for the population proportion of Americans who used home exercise equipment last year.

SOLUTION:

$$0.34 \pm 1.96\sqrt{\frac{0.34 \cdot (1-0.34)}{1000}}$$

or $$CI=[\ 0.311\ \ ;\ \ 0.369]$$

Work Sampling

In a production environment, a supervisor wanted to determine the percentage of time a particular machine operator was idle. A random sample of 500 observations indicated the employee was idle 15% of the time. Construct a 95% confidence interval for the percent of idle time for this operator.

SOLUTION:

$$0.15 \pm 1.96\sqrt{\frac{0.15 \cdot (1-0.15)}{500}}$$

or, CI=[0.1187; 1813]

A Political Question

A state representative would like to determine the average amount earned last summer by foreign student interns working during summer vacation in Indiana. She requires **95%** confidence that the sample mean is within **$50** of the population mean. Based on studies in previous years, the population standard deviation is estimated to be **$400**. How large a sample is necessary?

SOLUTION:

$$n = \frac{1.96^2 \cdot 400^2}{50^2} = 245.9 = 246$$

Confidence Interval (reflects sampling error) should always be reported along with the Point Estimate

An interpretation of the Confidence Interval estimate should also be provided

The sample size should be reported

The confidence level should always be reported

Frequentist interpretation of CI's

- In an infinitely long series of trials in which repeated samples of size *n* are drawn from the same population and 95% CI's for μ are calculated using the same method, the proportion of intervals that actually include μ will be 95% (coverage probability).
- However, for any particular CI, it is not known whether or not the CI includes μ, but the probability that it includes μ is either 0 or 1, that is, either it does or it doesn't.
- It is incorrect to say that the probability is 0.95 that the true $\mu$is in a particular CI.

95% CI, 50 samples from unit normal distribution