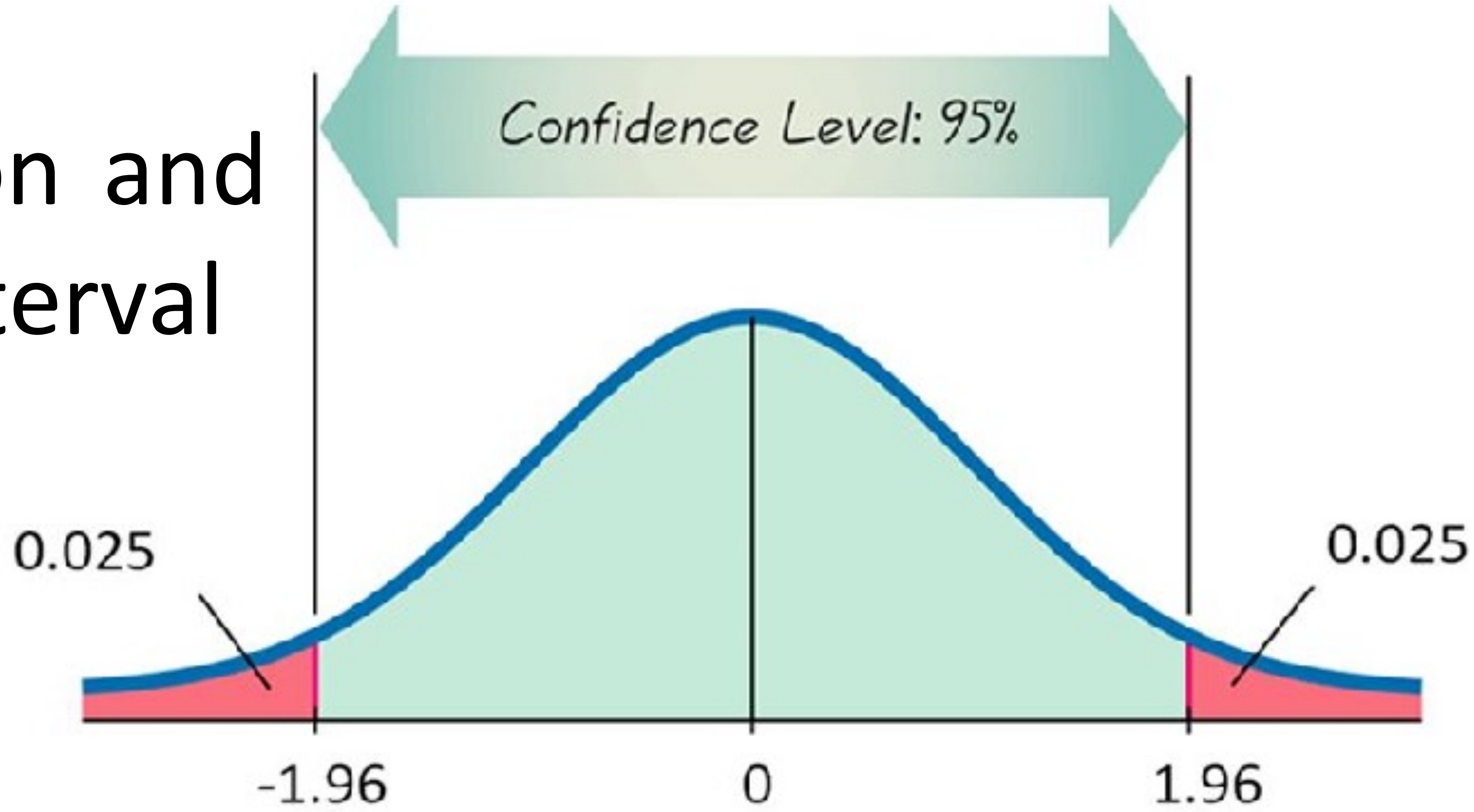
L06: Estimation and Confidence Interval

Yiming QIN
Assistant Professor
School of Energy and Environment
City University of Hong Kong



Introduction

In previous Chapters, we defined a **random variable *X* and had a new interpretation of data ----- the actual values of the random variable.**

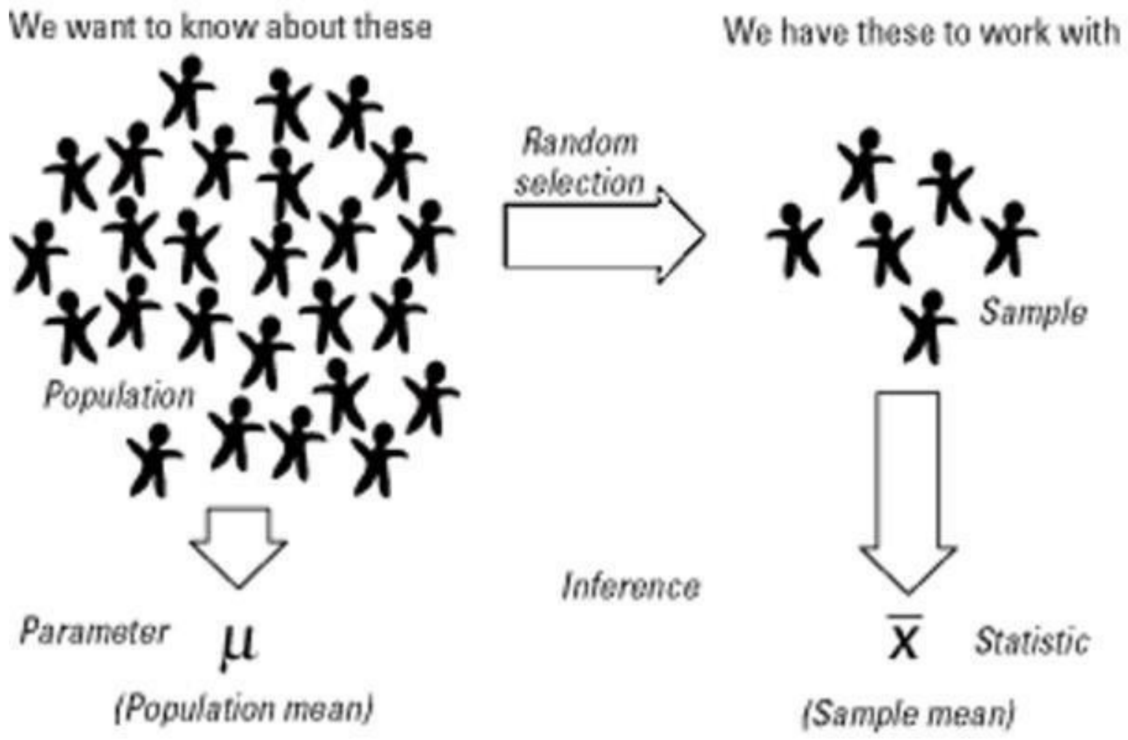
- If the distribution of X is known, then we can answer any probability problem of X.

- If the distribution is unknown, then we need to use STATISTICS to make an inference about the underlying distribution or the parameter of our interest.

Introduction

Key terms in statistics:

- **An UNKNOWN population: An UNKNOWN distribution of the r.v. *X*.**
- **A sample: A collection of data of *X*.**
- **A parameter: For example μ**
- **A statistic: For example, $\bar{x}$**



Introduction

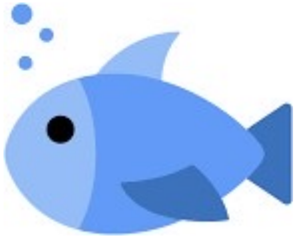
Since the distribution of X is unknown, the population mean μ_X and other parameters are also unknown

If we are interested in their true values, then we can use the collected data to guess their values statistically.

Sampling Variability

The observed value of the statistic depends on the particular sample selected from the population and it will vary from sample to sample.

Sampling Variability



Suppose there are 20 fish in the pond. The lengths of the fish (in inches) are given below

4.5	5.4	10.3	7.9	8.5	6.6	11.7	8.9	2.2	9.8
6.3	4.3	9.6	8.7	13.3	4.6	10.7	13.4	7.7	5.6

We caught fish with lengths 6.3 inches, 2.2 inches, and 13.3 inches.

$\bar{x}$ = 7.27 inches

2nd sample - 8.5, 4.6, and 5.6 inches.

$\bar{x}$ = 6.23 inches

3rd sample – 10.3, 8.9, and 13.4 inches.

$\bar{x}$ = 10.87 inches

The true mean is 8.

Notice that some sample means are closer and some farther away; some above and some below the mean

Sampling Variability

Suppose we wanted to estimate the proportion of blue candies in a very large bag of M&M

How might we go about estimating this proportion?

Sampling Variability

Suppose we wanted to estimate the proportion of blue candies in a very large bag of M&M

How might we go about estimating this proportion?

We could take a sample of candies and compute the proportion of blue candies in our sample.

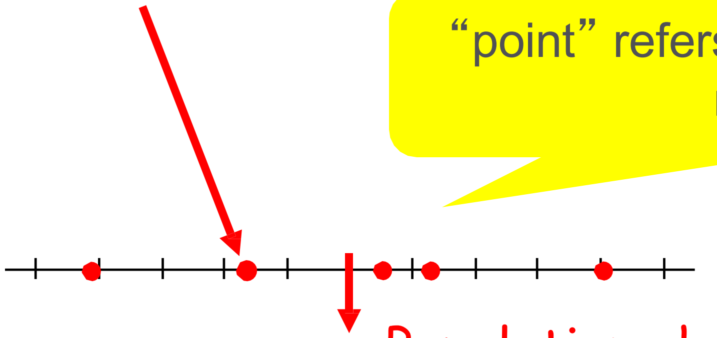
We would have a sample proportion or a statistic – a single value for the estimate.

Point Estimate

- A single number (a statistic) based on sample data that is used to estimate a population characteristic
- But not always to the population characteristic due to sampling variation

Different samples may produce different statistics.

"point" refers to the single value on a number line.

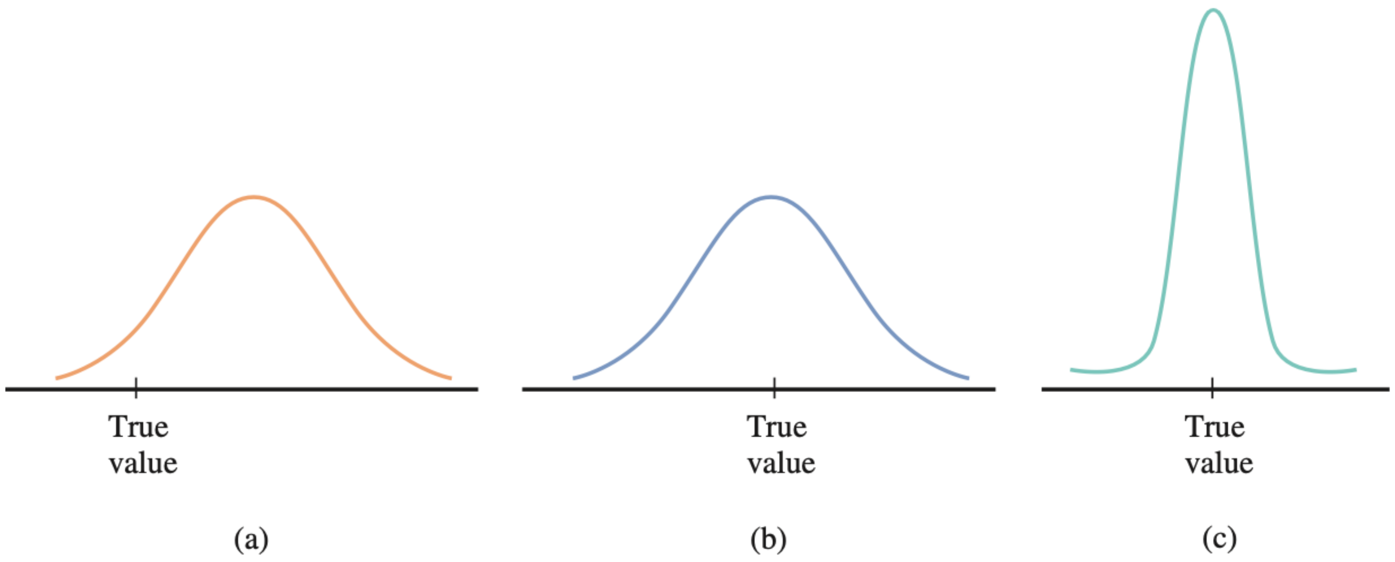


Population characteristic

Computing an Estimate

We would like to use a statistic that tends to produce an accurate estimate—that is, an estimate close to the value of the population characteristic.

Information about the accuracy of estimation for a particular statistic is provided by the statistic's sampling distribution.

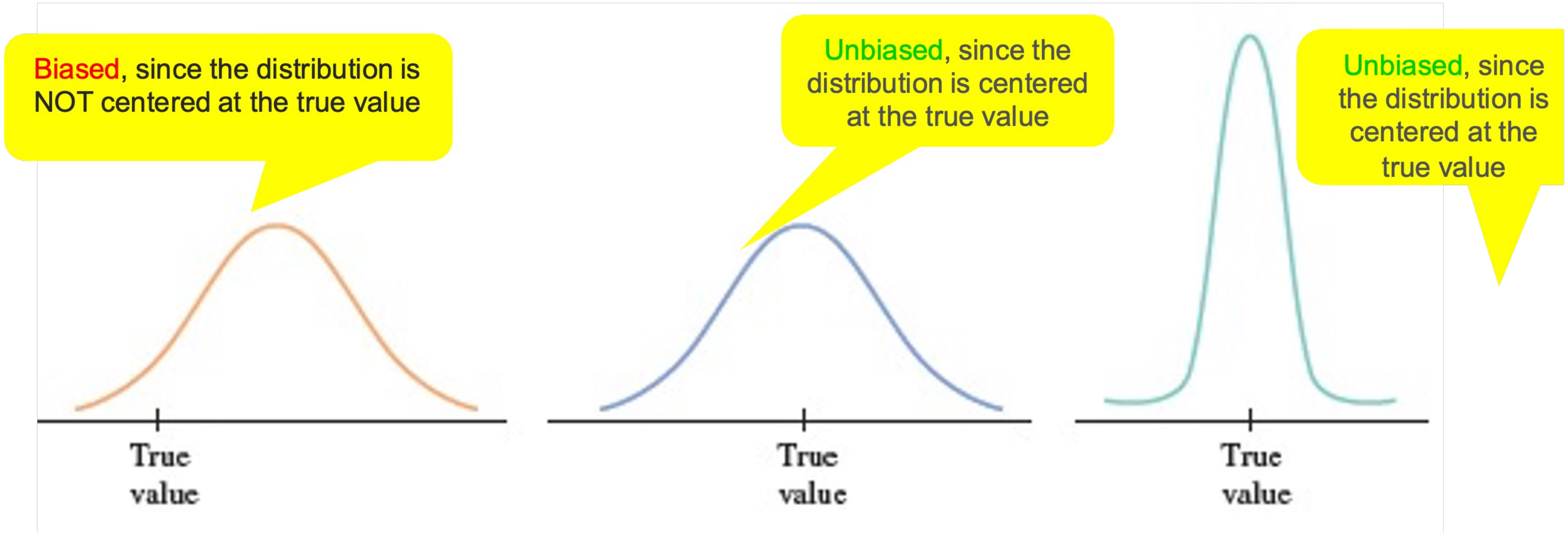


Sampling distributions of three different statistics for estimating a population characteristic.

Computing an Estimate

Definition:

- A statistic whose mean value is equal to the value of the population characteristic being estimated is said to be an **unbiased statistic**.
- A statistic that is not unbiased is said to be **biased**.



The best choice to use is the unbiased statistic with the smallest standard deviation.

M&M example revisit

Suppose we wanted to estimate the proportion of blue candies in a very large bag of M&M

We could take a sample of candies and compute the proportion of blue candies in our sample.

How much confidence do you have in the point estimate?

Would you have more confidence if your answer were an interval?

Large- sample **confidence interval** for a population proportion

We hope that the chosen statistic produces an estimate that is close, on average, to the true value.

Although a point estimate may represent our best **single-number guess** for the value of the population characteristic, it is not the only plausible value.

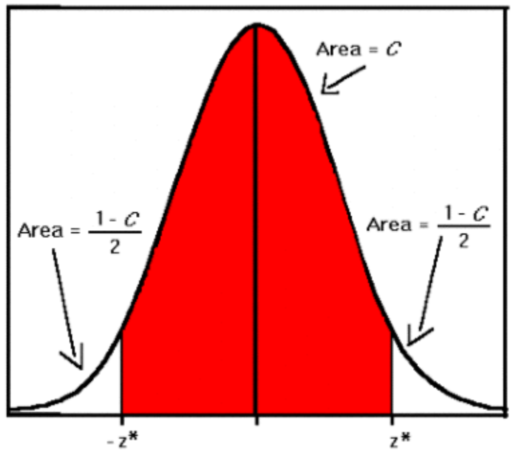
As an alternative to a point estimate, we can use the **sample data to report an interval of plausible values** for the population characteristic.

Confidence intervals

A confidence interval (CI) for a population characteristic is an interval of plausible values for the characteristic.

It is constructed so that, with a chosen degree of confidence, the actual value of the characteristic will be between the **lower and upper endpoints of the interval**.

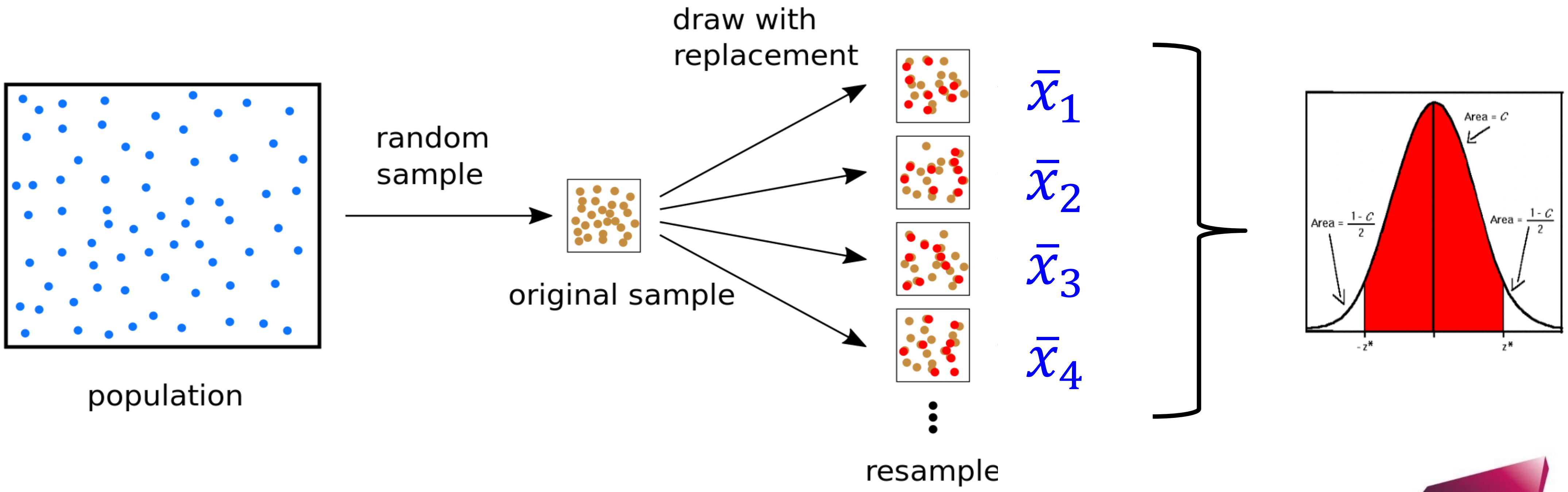
The primary goal of a confidence interval is to estimate an unknown population characteristic.



How to Construct a Confidence interval

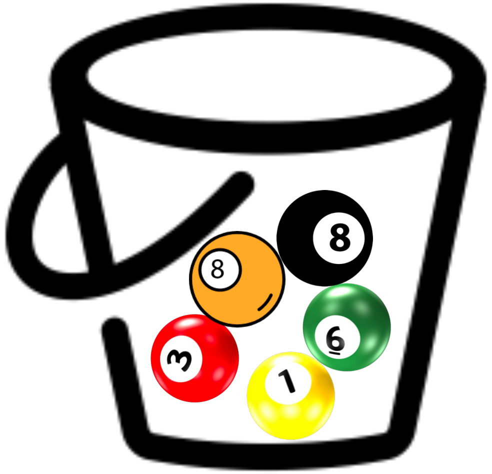
To construct a confidence interval, we need to know the **sampling distribution of the point-valued estimator**.

Bootstrap is a resampling procedure repeatedly taking random samples from the sample.



Bootstrap

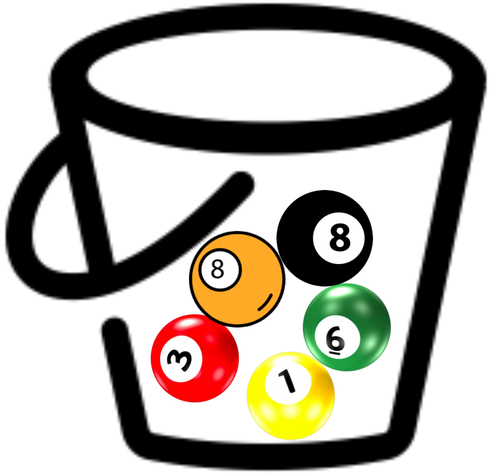
Imagine we have 5 billiard balls in a bucket.



Bootstrap

We first pick randomly a ball and replicate it. This is called **sampling with replacement.** We move the replicated ball to another bucket.

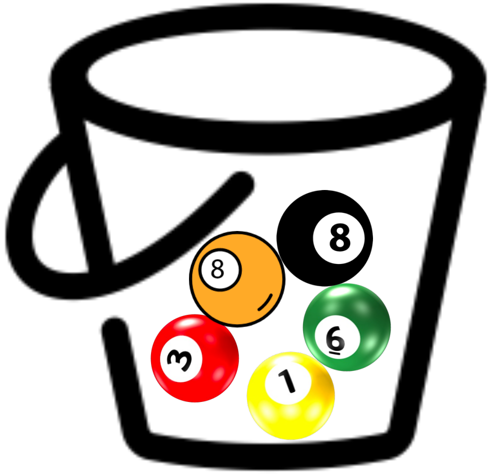
Bootstrap

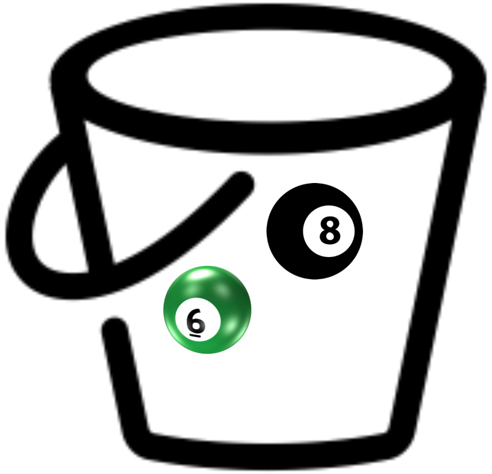


Bootstrap

We then pick randomly another ball and again we replicate it.
Again, we then move the replicated ball to the other bucket.

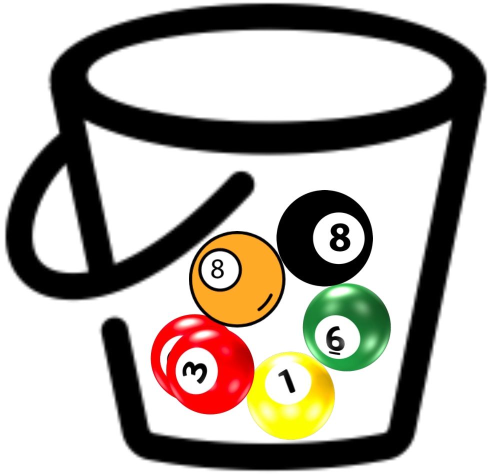
Bootstrap

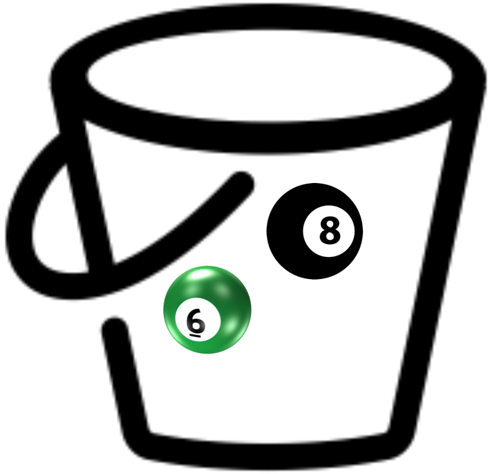




Bootstrap

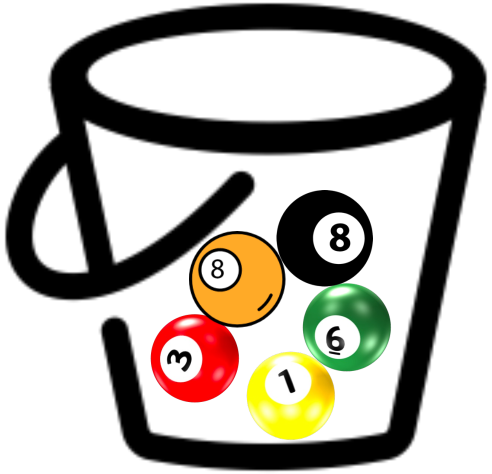
We repeat this process.

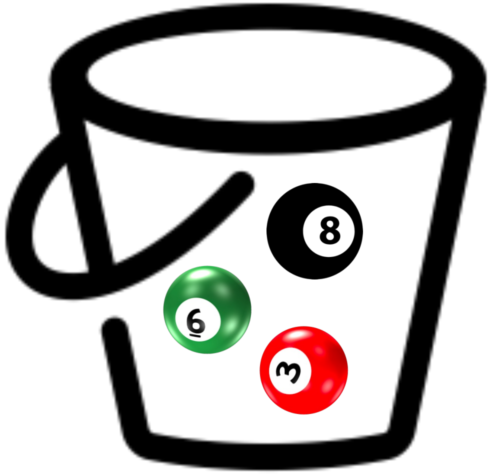




Bootstrap

Again



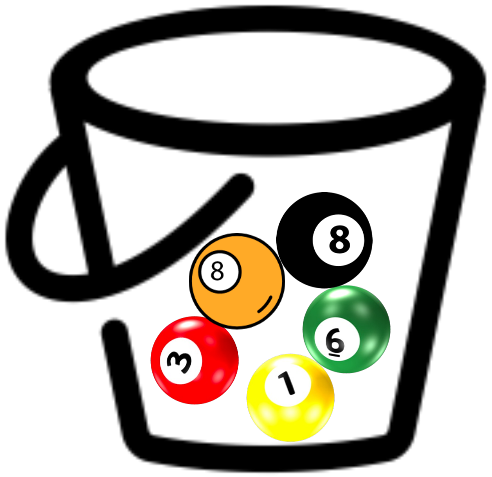


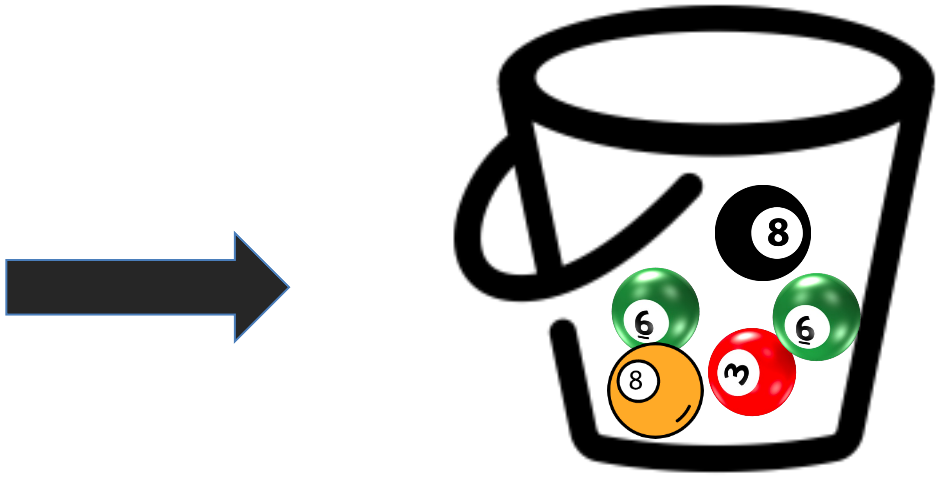
Bootstrap

And again

Bootstrap

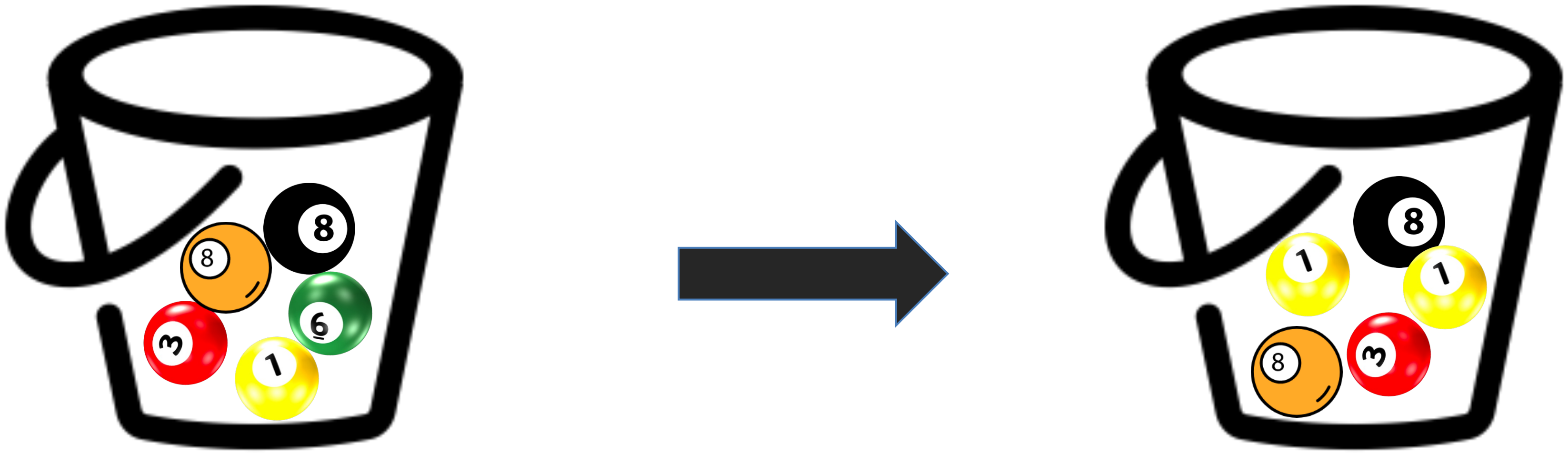
Until the other bucket has the same number of balls as the original one.





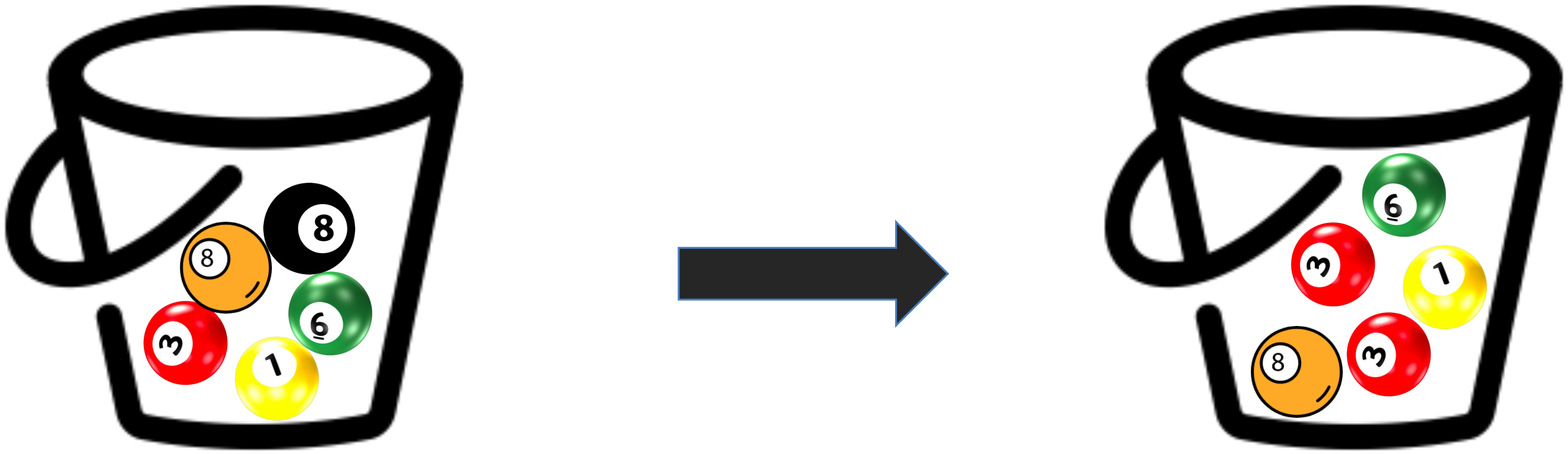
Bootstrap

We repeat the same process and acquire another sample



Bootstrap

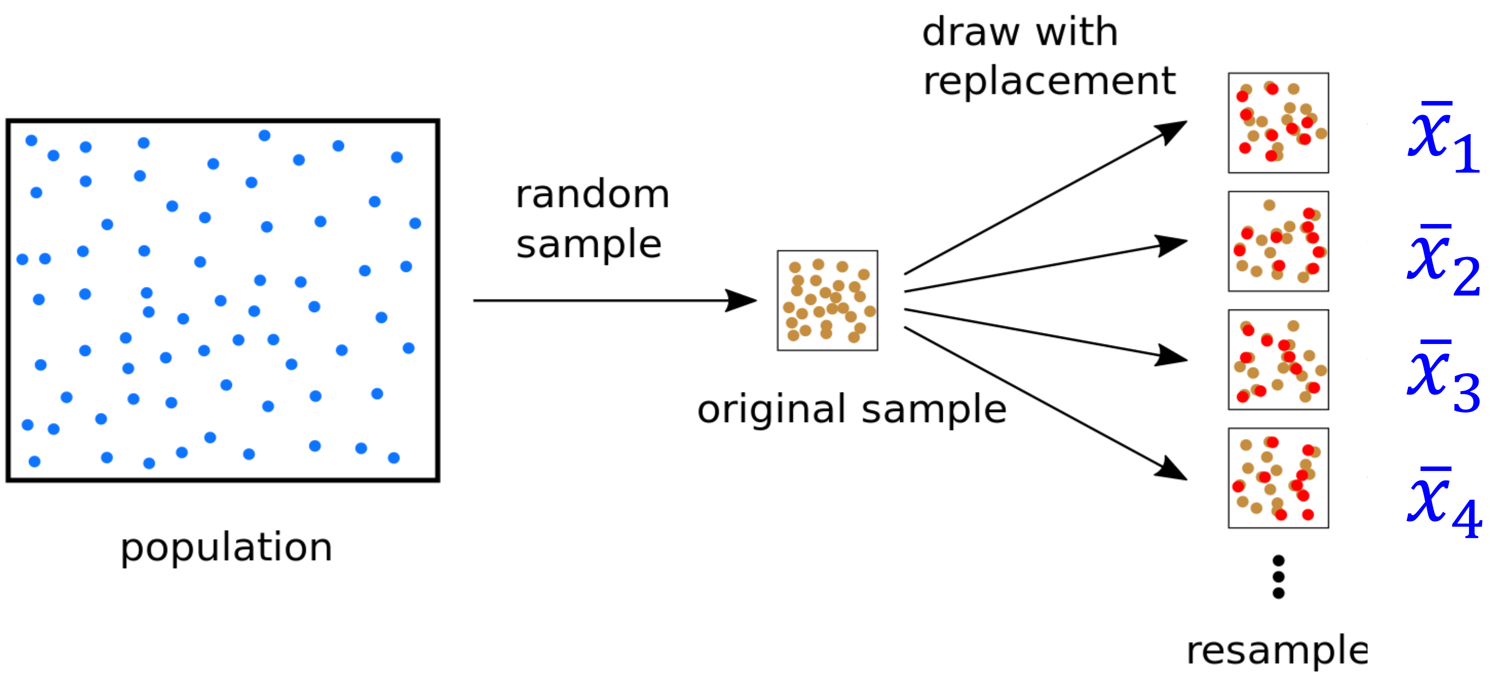
We repeat the same process and acquire another sample



How to Construct a Confidence interval

To construct a confidence interval, we need to know the **sampling distribution of the point-valued estimator**.

Bootstrap is a resampling procedure repeatedly taking random samples from the sample.



Sample Mean

The **sample mean** is defined as

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)}$$

where x_i is the i^{th} realization (**actual value**) of the rv X of our interest.

Thus, if we have B different sets of sample, then we would have B sample means, $\bar{x}_1$ $\bar{x}_2$ $\bar{x}_{3,\ldots.}$ $\bar{x}_B$.

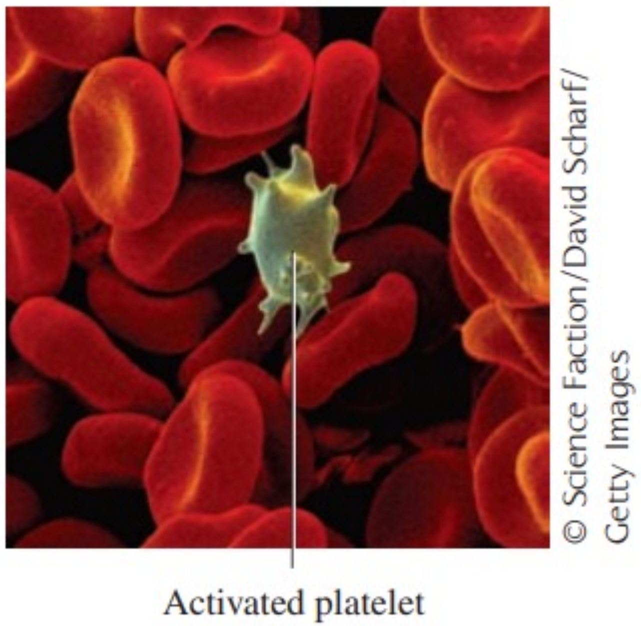
Note that all of these B sample means behave very similar to the n data $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, …, $\boldsymbol{x_n}$ of the rv X. So, we can imagine a rv whose actual values are the B sample means. **The rv is reasonably denoted by $\bar{X}$.**

How to Construct a Confidence interval

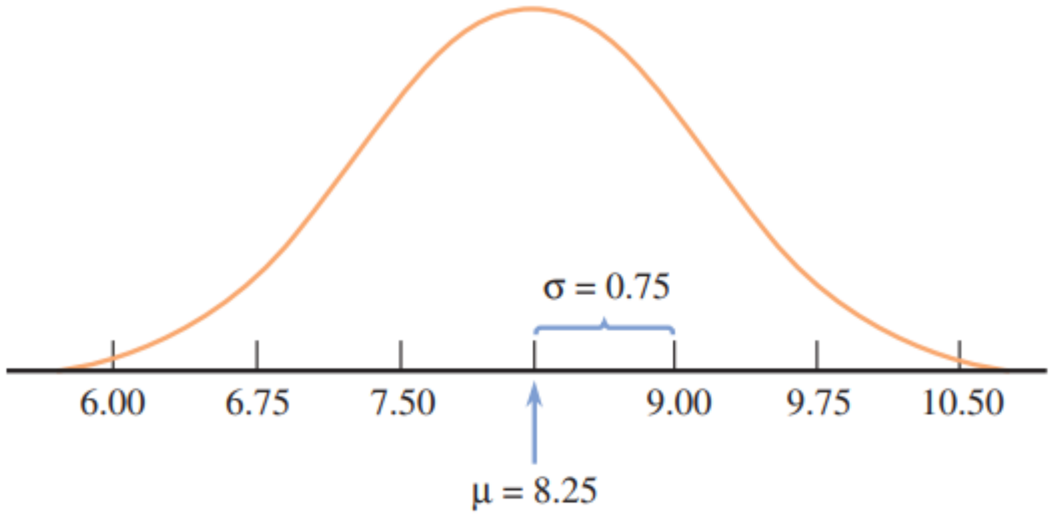
To construct a confidence interval, we need to know the **sampling distribution of the point-valued estimator**.

We need information about the sampling distribution.

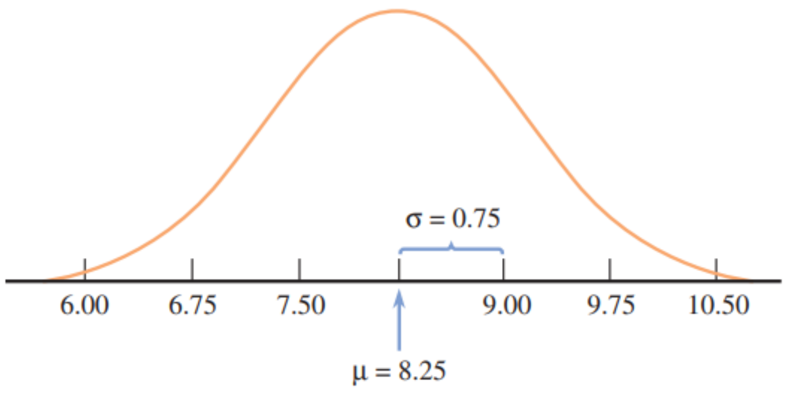
Example : Sampling Distributions

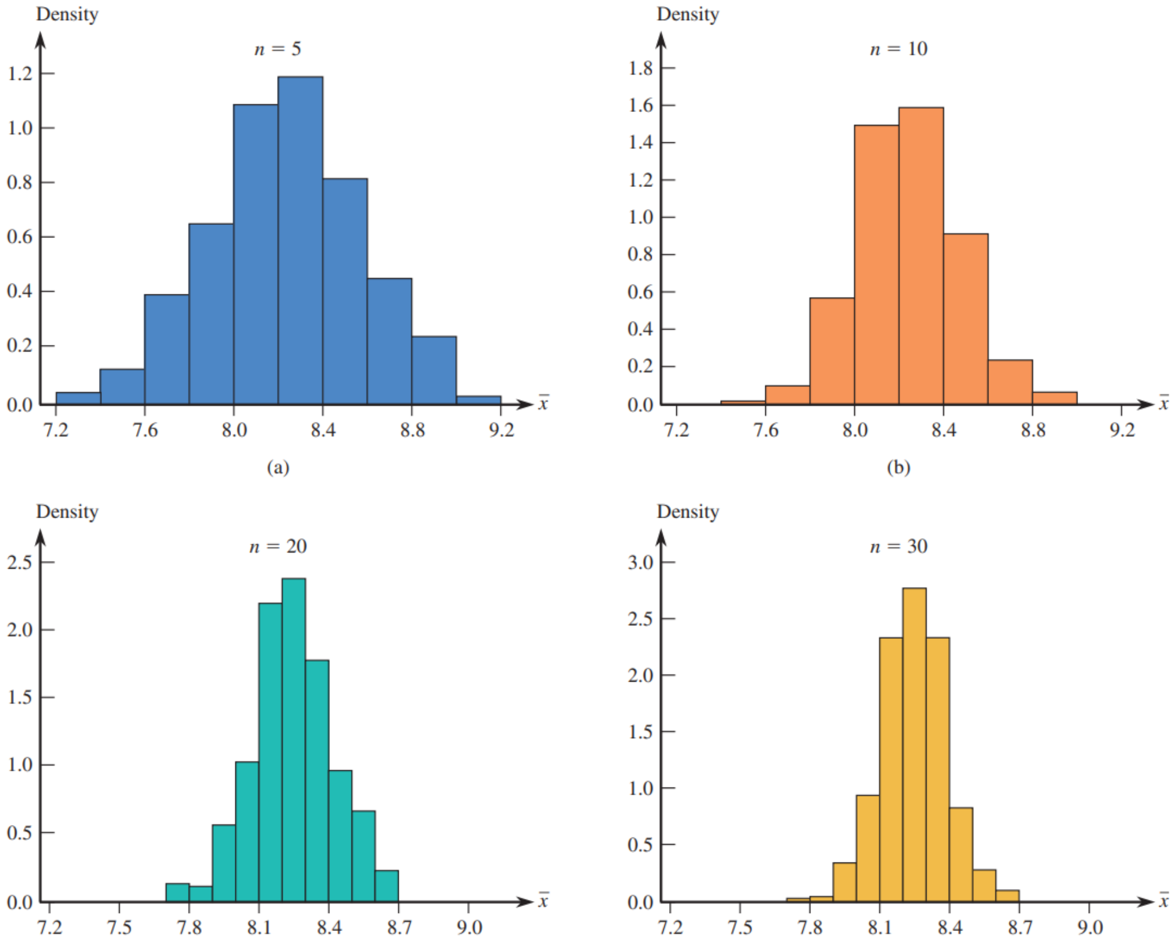


A study shows that the distribution of platelet volume for patients who do not have metabolic syndrome is approximately normal with a mean of 8.25 and a standard deviation of 0.75.



Example : Sampling Distributions

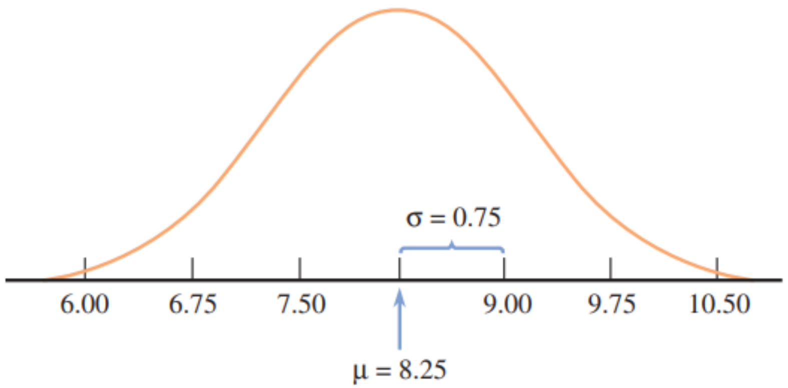




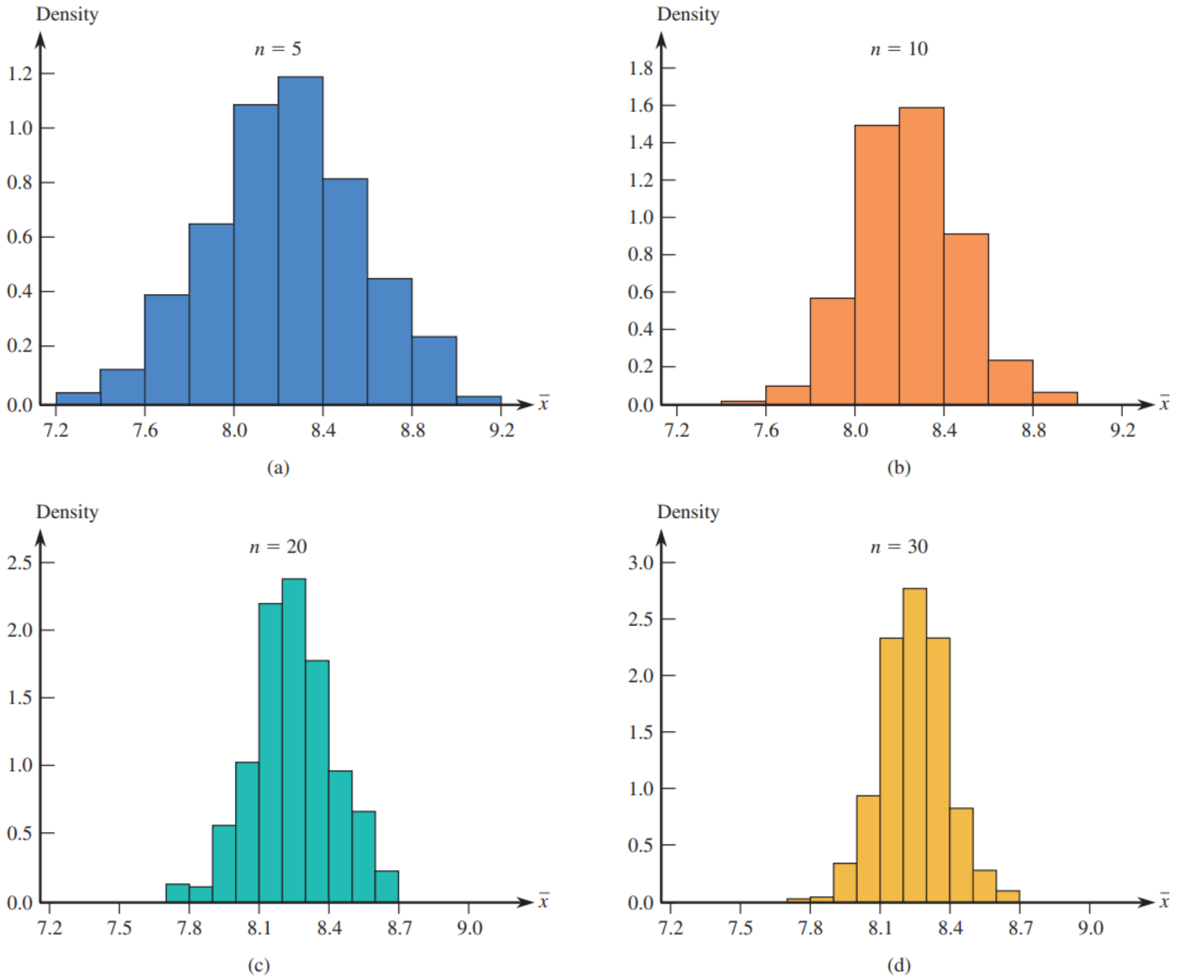
(a) (b) (c) (d)

We select 500 random samples from this normal distribution, with each sample consisting of n=5, 10, 20, and 30 observations.

Example : Sampling Distributions



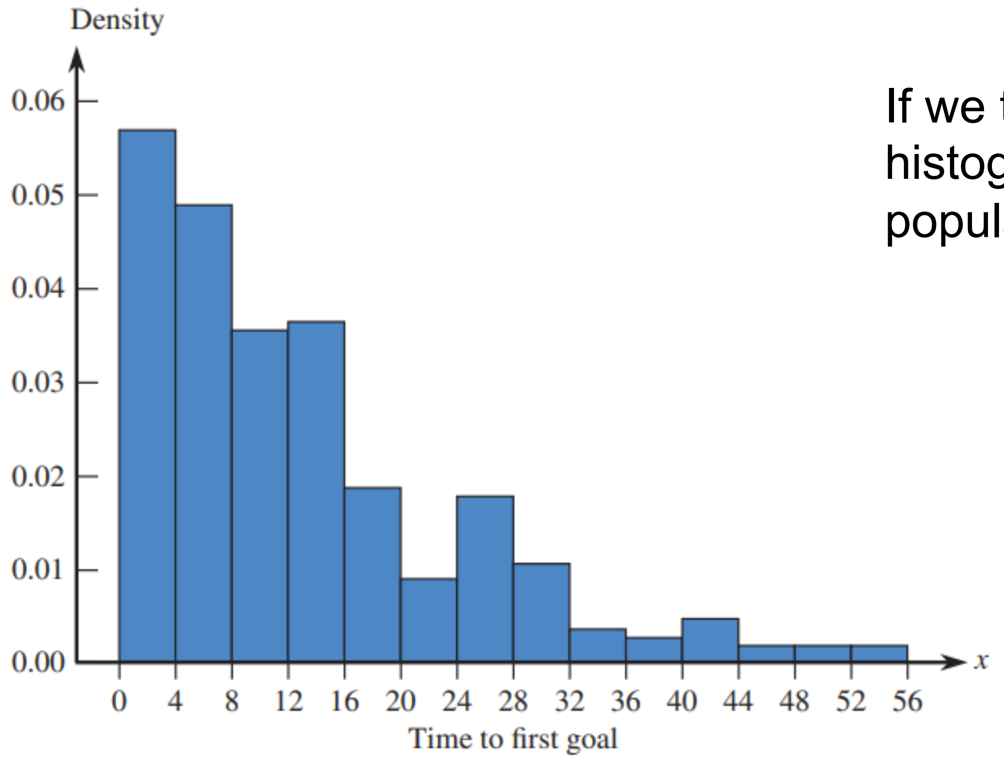
- Each of the four histograms is approximately normal in shape.
- Each histogram is centered approximately at 8.25, the mean of the population being sampled.
- The smaller the value of n, the greater the extent to which the sampling distribution spreads out about the population mean value.



Another Example : Sampling Distributions

Now consider when the population is quite skewed (and thus very unlike a normal distribution).

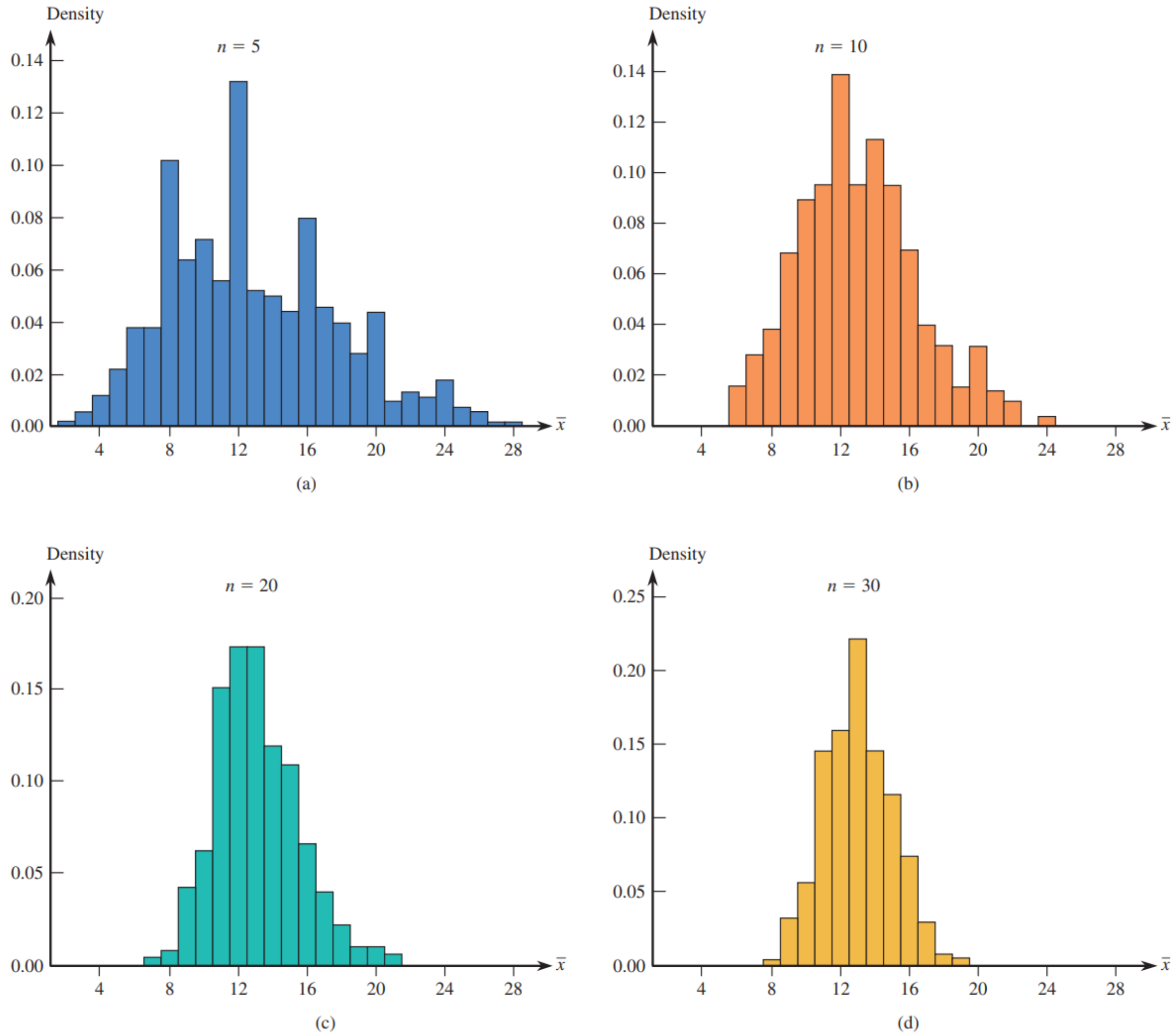
Here gave data on the time (in minutes) from the start of the game to the first goal scored for the 281 regular season games from the 2005–2006 season that went into overtime.



If we think of the 281 values as a population, the histogram in Figure shows the distribution of values in that population with a population mean of 13.

For each of the sample sizes n =5, 10, 20, and 30, we selected 500 random samples of size n.

Another Example : Sampling Distributions



- As with samples from a normal population, the averages of the 500 $\bar{x}$ values for the four different sample sizes are all close to the population mean of 13.
- As n increases, the histogram's spread about its center decreases.
- The distributions are skewed and differ in shape, but they become progressively more symmetric as the sample size increases. n =30, the histogram has a shape much like a normal curve.

General properties of the sampling distribution

Let $\bar{x}$ denote the mean of the observations in a random sample of size n from a population having mean μ and standard deviation σ. Denote the mean value of the $\bar{x}$ distribution by $\mu_{\bar{x}}$ and the standard deviation of the $\bar{x}$ distribution by $\sigma_{\bar{x}}$. Then the following rules hold:

Rule 1. $\mu_{\bar{x}} = \mu$.

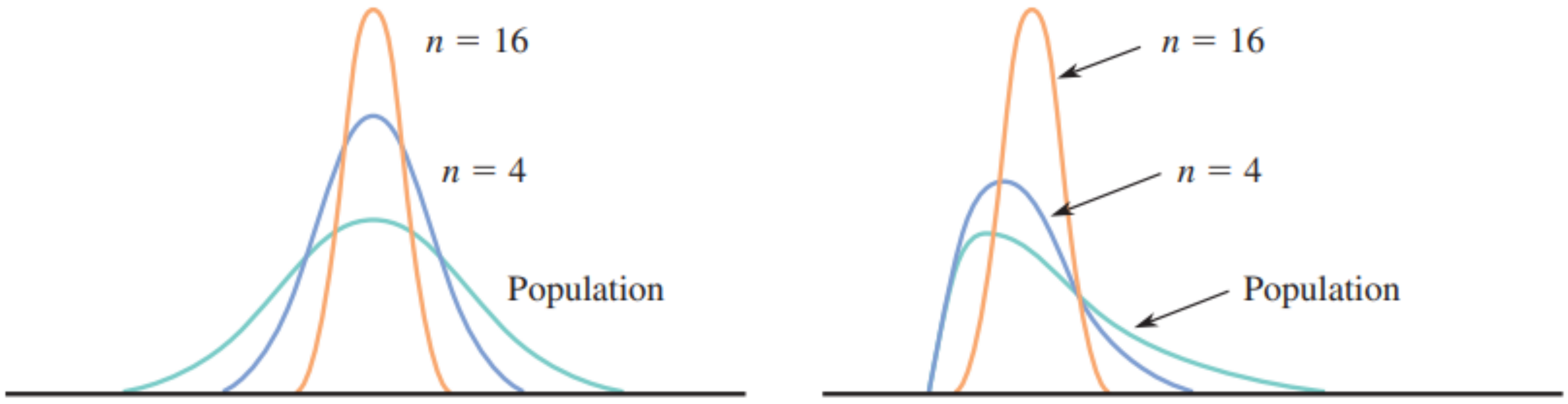
Rule 2. $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. This rule is exact if the population is infinite, and is approximately correct if the population is finite and no more than 10% of the population is included in the sample.

Rule 3. When the population distribution is normal, the sampling distribution of $\bar{x}$ is also normal for any sample size n.

Rule 4. (Central Limit Theorem) When n is sufficiently large, the sampling distribution of $\bar{x}$ is well approximated by a normal curve, even when the population distribution is not itself normal.

General properties of the sampling distribution

Population distribution and sampling distributions of x: (a) symmetric population; (b) skewed population.



(a) (b)

When n is sufficiently large, the x distribution is approximately normal, regardless of the population distribution.

This result has enabled statisticians to develop procedures for making inferences about a population mean using a large sample, even when the shape of the population distribution is unknown.

General properties of the sampling distribution

How large an n is needed for the x distribution to approximate a normal curve depends on how much the population distribution differs from a normal distribution.

Many statisticians recommend the following conservative rule:

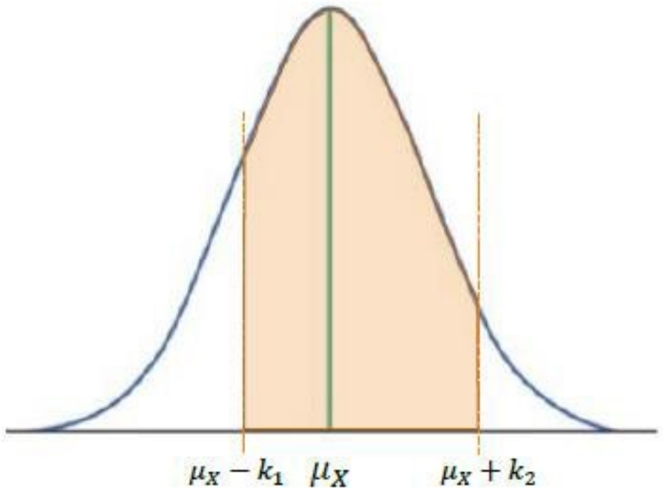
The Central Limit Theorem can safely be applied if *n* is greater than or equal to 30.

The Construction of a Confidence Interval

Recall that we have the result of the exact distribution of $\bar{X}$

$$\boxed{\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)}$$

Thus, we have the following normal density curve of this estimator:



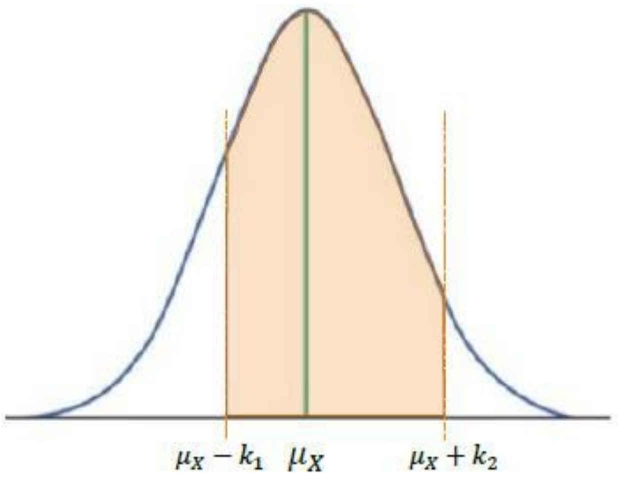
According to this curve, which interval should we use such that μ_X can be included? Obviously, **any interval in form of $[\boldsymbol{\mu_X} - \boldsymbol{k_1}, \boldsymbol{\mu_X} + \boldsymbol{k_2}]$,** where $0 < \boldsymbol{k_1} \leq +\infty$ *and* $0 < \boldsymbol{k_2} \leq +\infty$, i.e. including $(-\infty, +\infty)$, **can include the parameter $\boldsymbol{\mu_X}$.**

The Construction of a Confidence Interval

- Note that the area of the shaded region in the picture above is say C. Thus, we have

$$P(\mu_X - k_1 \le \bar{X} \le \mu_X + k_2)$$

$\boxed{\text{The random interval } [\bar{X} - k1\ ,\ \bar{X} + k2] \text{ contains } \mu_X \text{ with probability } C.}$



The Construction of a Confidence Interval

The random interval $[\bar{X}-k1\ ,\ \bar{X}+k2]$ contains μ_X with probability C.

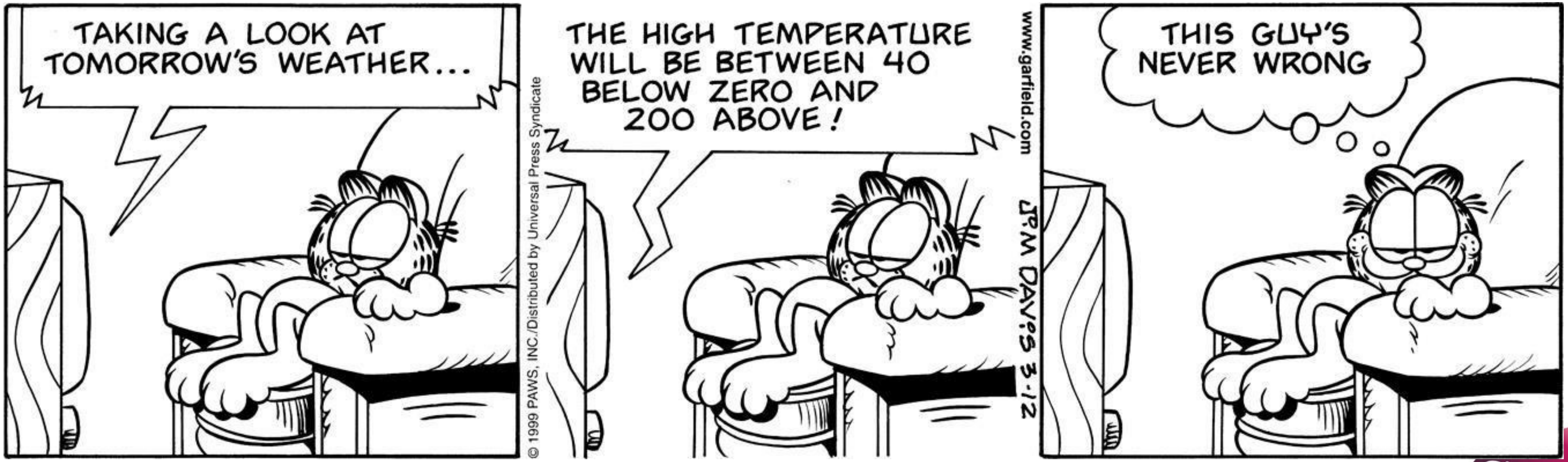
Undoubtedly, we would like to **find a random interval with probability as large as possible**.

The largest value of C is 1. Then, that's easy! We just find the interval with probability 1. That is, the interval always contains the unknown parameter!!

The Construction of a Confidence Interval

However, this “perfect” interval indeed is a very poor interval. Why?

This is because the interval with probability 1 indeed is $(-\infty, +\infty)$. It is certainly correct but a super poor interval!!

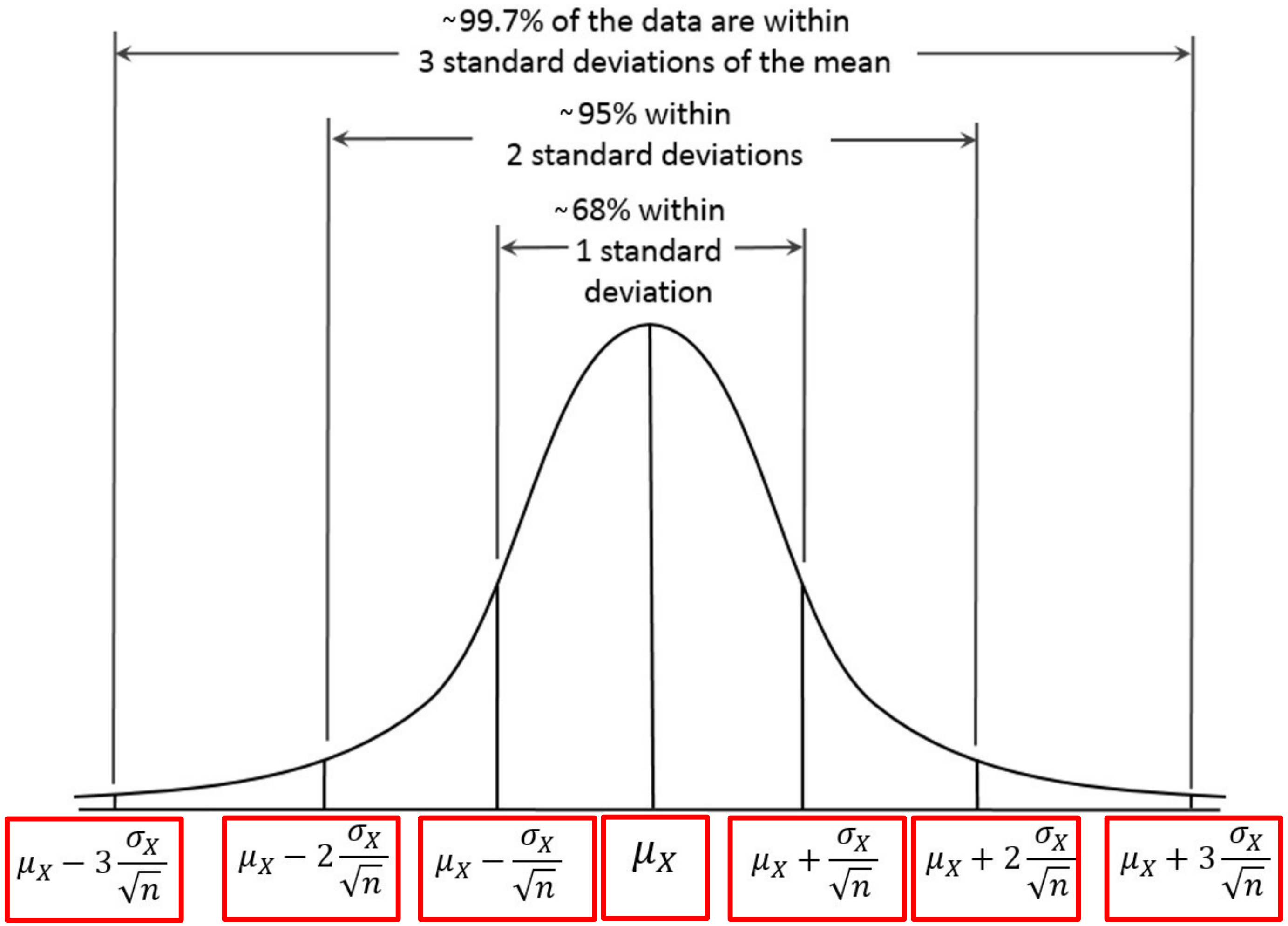


The Construction of a Confidence Interval

- Therefore, we can only consider a large value (but not 1) of C for the random interval. The commonly used values of C for random intervals are 0.9, 0.95 and 0.99.

- In the following, we would study how to get the random interval exactly with C = 0.95. First, we don't want the interval to bias towards any either side, thus we would consider the random interval with $k_1 = k_2$.

The Construction of a Confidence Interval



The Construction of a Confidence Interval

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to th

Z	.00	.01	.02	.03	.04	.05	.06
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500

Therefore,

1-0.02275*2=0.9545

$$P\left(\mu_X - 2\frac{\sigma_X}{\sqrt{n}} \le \bar{X} \le \mu_X + 2\frac{\sigma_X}{\sqrt{n}}\right) = 0.9544997\ .$$

To be more precise, we should use 1.96 instead of 2 because

$$P\left(\mu_X - 1.96\frac{\sigma_X}{\sqrt{n}} \le \bar{X} \le \mu_X + 1.96\frac{\sigma_X}{\sqrt{n}}\right) = 0.9500042\ .$$

Let’s make the value 0.9500042 correct to 4 d.p. for our following steps. Thus, we have

$$P\left(\mu_X - 1.96\frac{\sigma_X}{\sqrt{n}} \le \bar{X} \le \mu_X + 1.96\frac{\sigma_X}{\sqrt{n}}\right) = 0.9500\ (or\ 0.95)$$

and

$$P\left(\bar{X} - 1.96\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \bar{X} + 1.96\frac{\sigma_X}{\sqrt{n}}\right) = 0.95\ .$$

The Construction of a Confidence Interval

Thus, according to a particular set of collected data, we have the following general formal for the so-called **95% confidence interval of μ_X** :

$$[\bar{x} - \boxed{1.96}\frac{\sigma_x}{\sqrt{n}}, \bar{x} + \boxed{1.96}\frac{\sigma_x}{\sqrt{n}}]$$

when σ_X is known.

The Construction of a Confidence Interval

Thus, according to a particular set of collected data, we have the following general formal for the so-called **95% confidence interval of μ_X** :

Bound on error of estimation

$$\left[\bar{x} \boxed{-\boxed{1.96}\frac{\sigma_x}{\sqrt{n}}}, \bar{x} \boxed{+\boxed{1.96}\frac{\sigma_x}{\sqrt{n}}}\right]$$

Point estimator

when σ_X is known.

More About the Confidence Interval

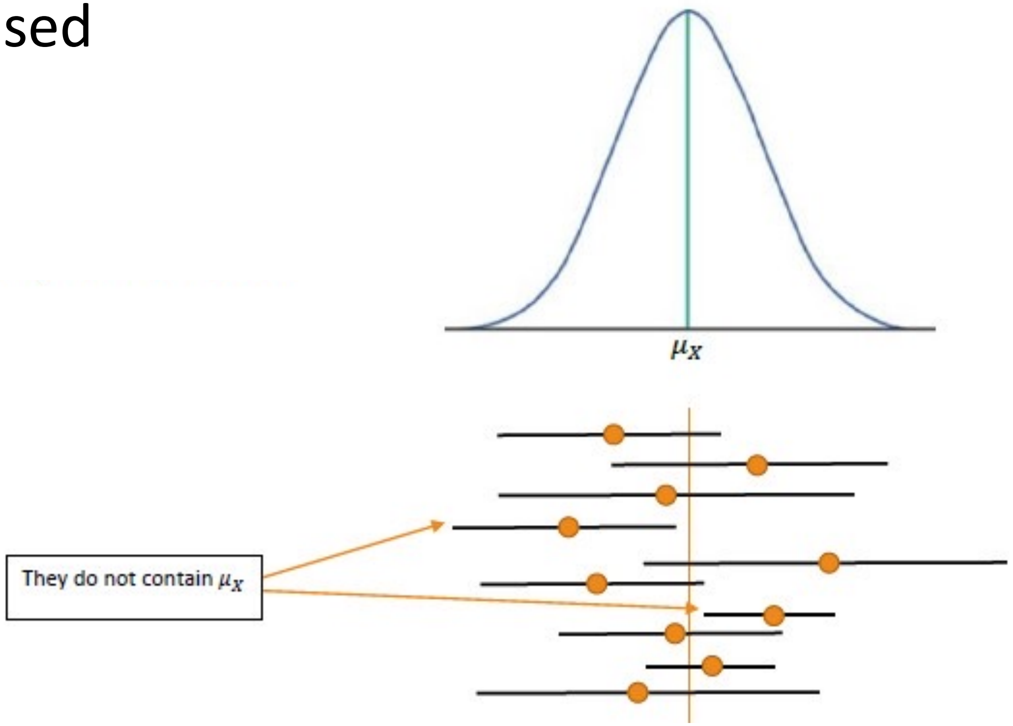
Here I would like to emphasize that the value **95% does not mean that the confidence interval contains μ_X with probability 0.95.** Indeed, the **probability** that the confidence interval **contains μ_X is either 0 (never contain) or 1 (contain certainly). That is,**

$$P(\bar{x} - 1.96\frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{x} + 1.96\frac{\sigma_x}{\sqrt{n}}) \quad \text{is either 0 or 1.}$$

More About the Confidence Interval

- We can illustrate this situation in the picture on the right:

● represents the numerical sample mean based on a particular set of data.



The One-Sample *z* Confidence Interval for μ

The general formula for a confidence interval for a population mean μ is

$$\bar{x} \pm (z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$$

When

1. $\bar{x}$ is the sample mean from a **simple random sample,**
2. the **sample size *n* is large** (generally $n > 30$)
3. $\boldsymbol{\sigma}$**, the population standard deviation, is known**

The three most used confidence levels, 90%, 95%, and 99%, use z critical values 1.645, 1.96, and 2.58, respectively.

Table of Standard Normal Probabilities for Negative Z-scores

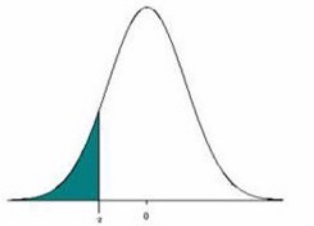
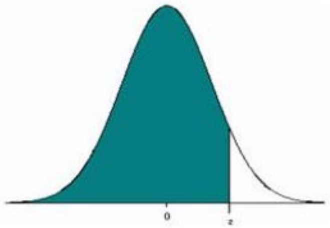


Table of Standard Normal Probabilities for Positive Z-scores



1.645(90%)

1.96(95%)

2.58(99%)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
[illegible]	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Note that the probabilities given in this table represent the area to the LEFT of the z-score.
The area to the RIGHT of a z-score = 1 – the area to the LEFT of the z-score

Example

In a stream-monitoring study, 30 chloride samples are collected in a particular location.

Past history indicates that the standard deviation of chloride concentrations in this area is 25 mg/L. If a sample of 30 readings has a mean value of 560 mg/L, how confident can we be that the average represents the true average of the stream?

Example

In a stream-monitoring study, 30 chloride samples are collected in a particular location.

Past history indicates that the standard deviation of chloride concentrations in this area is 25 mg/L. If a sample of 30 readings has a mean value of 560 mg/L, how confident can we be that the average represents the true average of the stream?

For 95% confidence interval, $[\bar{x} - \boxed{1.96}\frac{\sigma_x}{\sqrt{n}}, \bar{x} + \boxed{1.96}\frac{\sigma_x}{\sqrt{n}}]$

$$\bar{X} \pm 1.96\frac{\sigma_X}{\sqrt{n}}$$
$$= 560 \pm 1.96\frac{25}{\sqrt{30}}$$
$$= 560 \pm 8.9$$

Thus, we can be 95% certain that the true mean lies within 551.1 mg/L and 568.9mg/L. Because the 95% confidence interval is $\pm 8.9\text{mg/L}$, the true mean lies within arrange that is narrow than$\pm 10\text{mg/L}$ with at 95% assurance.

Example

In a stream-monitoring study, 30 chloride samples are collected in a particular location. Past history indicates that the standard deviation of chloride concentrations in this area is 25 mg/L. If a sample of 30 readings has a mean value of 560 mg/L, how confident can we be that the average represents the true average of the stream?

Can we say that there is 99% chance that the true mean lies within 10mg/L of the observed mean?

Example

In a stream-monitoring study, 30 chloride samples are collected in a particular location. Past history indicates that the standard deviation of chloride concentrations in this area is 25 mg/L. If a sample of 30 readings has a mean value of 560 mg/L, how confident can we be that the average represents the true average of the stream?

Can we say that there is 99% chance that the true mean lies within 10mg/L of the observed mean?

The three most commonly used confidence levels, 90%, 95%, and **99%,** use z critical values 1.645, 1.96, and **2.58**, respectively.

For 99% confidence interval,

$$\bar{X} \pm 2.58 \frac{\sigma_X}{\sqrt{n}}$$

$$= 560 \pm 2.58 \frac{25}{\sqrt{30}}$$

$$= 560 \pm 11.8$$

Therefore, the answer is "No".

For this problem, the confidence interval lies within the desired $\pm$ 10mg/L range of the observed mean at the 95% confidence interval but not at the 99% confidence interval.

Sample size

Problem: Find the **sample size** necessary in order to obtain a specified maximum error and level of confidence (assume the standard deviation is known).

The bound-on error of estimation associated with a 95% confidence interval is

$$B = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Solve this expression for *n*:

$$n = \left(\frac{1.96\sigma}{B}\right)^2$$

If the desired confidence level is something other than 95%, 1.96 is replaced by the appropriate *z* critical value (for example, 2.58 for 99% confidence).

Example

Find the sample size necessary to estimate a population mean to within 0.5 with 95% confidence if the standard deviation is 6.2.

Solution: Since we want to be 95% certain, $z = 1.96$. Applying the formula, we need at least

$$n = \left[\frac{(1.96)(6.2)}{0.5}\right]^2 = [24.304]^2 = 590.684$$

Therefore, $n = 591$.

Note: When solving for sample size n, always round up to the next largest integer.

Example

Suppose we wish to estimate the proportion of young people exposed to ‘social noise’ in certain venues (e.g. in clubs, pop-concerts, cinemas) who exhibit various forms of hearing impairment. We decide to conduct a survey in the form of a simple random sample of size *n* drawn from a relevant study population of size *N*. The variable to be measured under appropriate medical protocols is the hearing threshold, in decibels.

Suppose that in a corresponding earlier survey the population variance was 0.38. What sample size would be needed to be 99% sure of our estimate of hearing threshold being within 0.1 decibels of the true value?

Example

Suppose that in a corresponding earlier survey the population variance was 0.38. What sample size would be needed to be 99% sure of our estimate of hearing threshold being within 0.1 decibels of the true value?

Solution: Since we want to be 99 % certain, z = 2.576.

Applying the formula, we need at least

$$n = \left[\frac{z\,\sigma}{\mathrm{B}}\right]^2$$

$$n = \frac{(2.576)^2(0.38)}{(0.1)^2}$$

$$n = 252.16$$

So a sample of size 253 or larger.

CityU

Confidence Interval When σ Is Unknown

The confidence interval just developed has an obvious drawback: To compute the interval endpoints, σ must be known.

$$\bar{x} \pm (z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$$

Unfortunately, this is rarely the case in practice. We now turn our attention to the situation when σ is unknown

Confidence Interval When σ Is Unknown

The confidence interval just developed has an obvious drawback: To compute the interval endpoints, σ must be known.

$$\bar{x} \pm (z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$$

When σ is unknown, we use the sample standard deviation s to estimate σ.

In place of z-scores, we must use the following to t value:

$$t = \frac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$$

Confidence Interval When σ Is known

To understand the derivation of this confidence interval, it is instructive to begin by taking another look at the previous 95% confidence interval.

We know that $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Also, when the population distribution is normal, the $\bar{x}$ distribution is normal. These facts imply that the standardized variable Z has approximately a standard normal distribution.

$$z = \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

Because the interval from -1.96 to 1.96 captures an area of .95 under the *z* curve, **approximately 95% of all samples result in an $\bar{x}$ value that satisfies**

$$-1.96 < \frac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} < 1.96$$

Confidence Interval When σ Is known

$$-1.96 < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1.96$$

Rearrange

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

Lower endpoint of the 95% large-sample confidence

Upper endpoint of the 95% large-sample confidence

Confidence Interval When σ Is Unknown

When σ is unknown, we use the sample standard deviation ***s*** to estimate σ.

If σ is unknown, we must use the sample data to estimate σ. If we use the sample standard deviation as our estimate, the result is a different standardized variable de noted by t:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

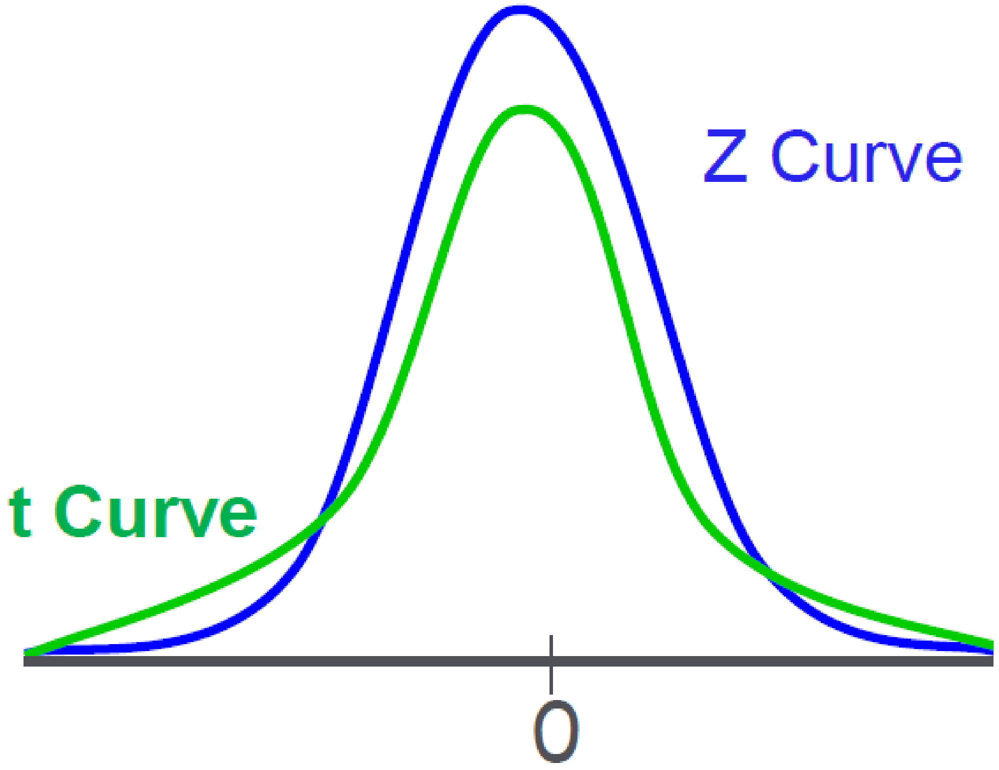
The value of **s** may not be all that close to σ, especially when n is small.

The use of the value of *s* introduces extra variability.

Therefore, the distribution of t values has more variability than a standard normal curve.

Confidence Interval When σ Is Unknown

The distribution of t is more spread out than the standard normal (z) distribution.



Important Properties of t Distributions

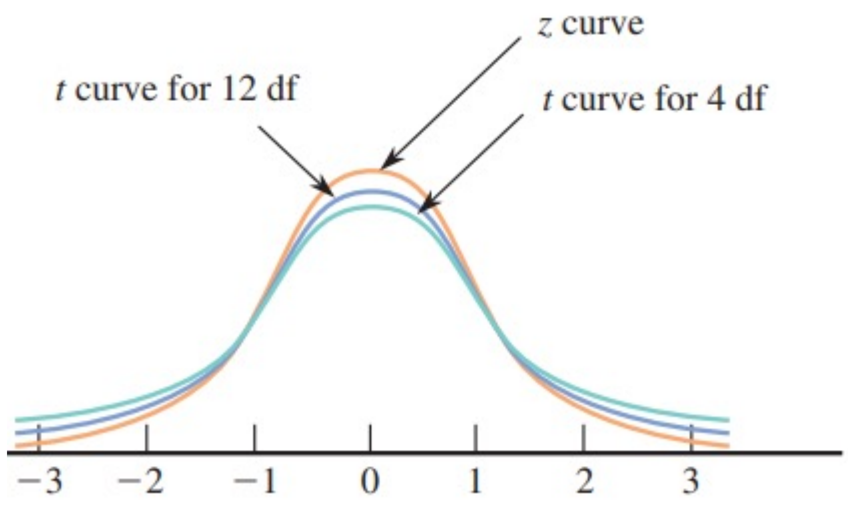
While normal distributions are distinguished from one another by their mean and standard deviation, t-distributions are distinguished by degrees of freedom (df).

1. The t distribution corresponding to any particular df is bell-shaped and centered at zero.

2. Each t distribution is more spread out than the standard normal (z) distribution.

3. As the **number of degrees of freedom increases**, the **spread of the corresponding t distribution decreases**.

4. As the number of degrees of freedom increases, the corresponding sequence of t distributions approaches the standard normal (z) distribution.



t distribution critical values

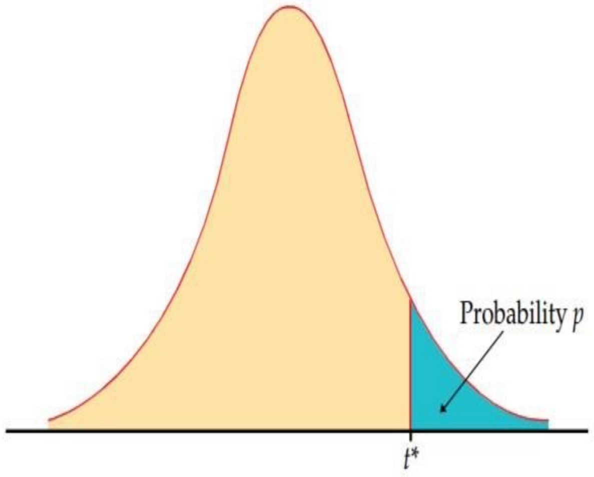


Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

To find a particular critical value, go down the left margin of the table to the row labeled with the desired number of degrees of freedom.

Then move over in that row to the column headed by the desired area.

TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Example

Find the critical t value for df =9,

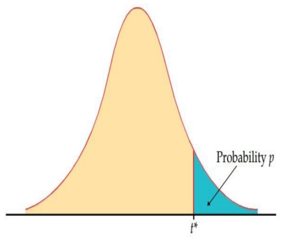
$P\{T_9 \geq t_{0.025,9}\}$=0.025

Critical value =2.262

TABLE D

t distribution critical values

Table entry for *p* and *C* is the critical value t^* with probability *p* lying to its right and probability *C* lying between $-t^*$ and t^*.



	Upper-tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Example

1. Determine $\mathrm{P}(T_{10} \geq 2.764)$

Solution:

$$\mathrm{P}(T_{10} \geq 2.764) = \mathrm{P}\big(T_{10} \geq t_{0.01,10}\big) = 0.01$$

2. Find x, for $\mathrm{P}(T_{10} \geq x)$=0.05.

Solution:

$$x = t_{0.05,10} = 1.812$$

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

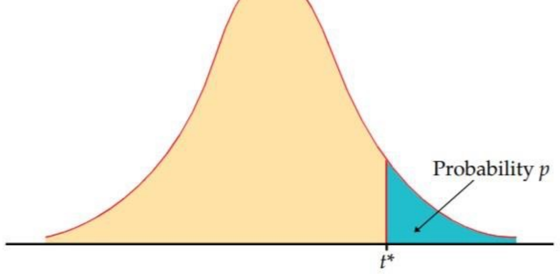
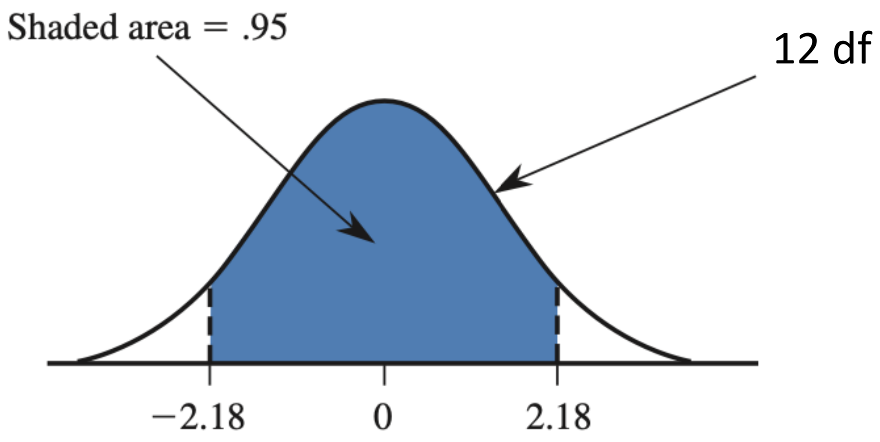


TABLE D

t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Example



The value in the 12-df row under the column corresponding to central area .95 is 2.179(~2.18), so 95% of the area under the *t* curve with 12 df lies between -.18 and 2.18

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	t
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.00
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.0
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.
	Confidence Level										

Example

The larger the number of degrees of freedom, the more closely the *t* curve resembles the *z* curve.

Furthermore, once the number of degrees of freedom exceeds 30, the critical values change little as the number of degrees of freedom increases.

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
					Confidence Level						

One-Sample t Confidence Interval

The fact that the t distribution is approximately the z (standard normal) distribution **<u>when n is large led</u>** to the z confidence interval when σ is known.

In the same way, the following proposition provides the key to obtaining a confidence interval when the population distribution is normal but σ is unknown.

The One-Sample *z* Confidence Interval for μ

The general formula for a confidence interval for a population mean μ based on a sample of size n when is

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

When

1. $\bar{x}$ is the sample mean from a **simple random sample,**
2. the population distribution is normal**, or** the sample size n is large (generally n > 30)
3. $\boldsymbol{\sigma}$**, the population standard deviation, is unknown**

where the *t* critical value is based on df = *n* - 1. t Table gives critical values appropriate for each of the confidence levels 90%, 95%, and 99%, as well as several other less frequently used confidence levels.

z Confidence Interval

To see how this result leads to the desired 95% confidence interval, consider the case *n* =25.

df= *n* -1=24.

t critical value of 2.06

The interval between -2.06 and 2.06 captures a central area of 95% under the *t* curve with 24 df.

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

z Confidence Interval

To see how this result leads to the desired 95% confidence interval, consider the case n =25.

df= n -1=24.

t critical value of 2.06

The interval between -2.06 and 2.06 captures a central area of 95% under the t curve with 24 df.

z critical value 1.96
The extra uncertainty that results from estimating s causes the *t* interval to be wider than the *z* interval.

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Example

A study is conducted to learn how long it takes the typical tax payer to complete their federal income tax return. **A random sample of 17 income** tax filers showed a **mean time (in hours) of 7.8** and a standard deviation **of 2.3. Find a 95% confidence interval** for the true mean time required to complete a federal income tax return. **Assume the time to complete the return is normally distributed.**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%

Confidence Level

Example

A study is conducted to learn how long it takes the typical tax payer to complete their federal income tax return. **A random sample of 17 income** tax filers showed a **mean time (in hours) of 7.8** and a standard deviation **of 2.3. Find a 95% confidence interval** for the true mean time required to complete a federal income tax return. **Assume the time to complete the return is normally distributed.**

Parameter of Interest: the mean time required to complete a federal income tax return.

Confidence Interval Criteria:

a) Assumptions: Sampled population assumed normal, σ unknown.
b) Test statistic: t will be used.
c) Confidence level: 0.95

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Example

A study is conducted to learn how long it takes the typical tax payer to complete their federal income tax return. **A random sample of 17 income** tax filers showed a **mean time (in hours) of 7.8** and a standard deviation **of 2.3. Find a 95% confidence interval** for the true mean time required to complete a federal income tax return. **Assume the time to complete the return is normally distributed.**

Parameter of Interest: the mean time required to complete a federal income tax return.

Confidence Interval Criteria:

a) Assumptions: Sampled population assumed normal, σ unknown.

b) Test statistic: t will be used.

c) Confidence level: 0.95

The Sample Evidence: n = 17, df =n-1=16, $\bar{x}$ =7.8, and s=2.3

The Confidence Interval:

Confidence coefficients: t= 2.12

$$\bar{x} \pm \boxed{(t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)} = (2.12)\frac{2.3}{\sqrt{17}} = (2.12)(0.5578) = 1.18$$

So, the Confidence limits:

7.8−1.18 to 7.8+1.18

6.62 to 8.98

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Example

In a study, chimpanzees learned to use an apparatus that dispensed food when either of two ropes was pulled. When one of the ropes was pulled, only the chimp controlling the apparatus received food. When the other rope was pulled, food was dispensed both to the chimp controlling the apparatus and also to a chimp in the adjoining cage. The accompanying data represent the number of times out of 36 trials that each of seven chimps chose the option that would provide food to both chimps (the “charitable” response).

23 22 21 24 19 20 20

Compute a 99% confidence interval for the mean number of charitable responses for the population of all chimps.

Example

In a study, chimpanzees learned to use an apparatus that dispensed food when either of two ropes was pulled. When one of the ropes was pulled, only the chimp controlling the apparatus received food. When the other rope was pulled, food was dispensed both to the chimp controlling the apparatus and also to a chimp in the adjoining cage. The accompanying data represent the number of times out of 36 trials that each of seven chimps chose the option that would provide food to both chimps (the “charitable” response).

23 22 21 24 19 20 20

Compute a 99% confidence interval for the mean number of charitable responses for the population of all chimps.

Check the normal probability plot of these data

Example

One way to see whether an assumption of population normality is plausible is to construct a **normal probability plot** of the data.

DEFINITION

A **normal probability plot** is a scatterplot of the ($\boxed{\text{normal score,}}$ observed value) pairs.

A strong linear pattern in a normal probability plot suggests that population normality is plausible. On the other hand, systematic departure from a straight-line pattern (such as curvature in the plot) indicates that it is not reasonable to assume that the population distribution is normal.

$$z \text{ score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

23	22	21	24	19	20	20

Sample mean $\bar{x}$= 21.29

Sample standard deviation s = 1.80

Normal probability plot

One way to see whether an assumption of population normality is plausible is to construct a **normal probability plot** of the data.

DEFINITION

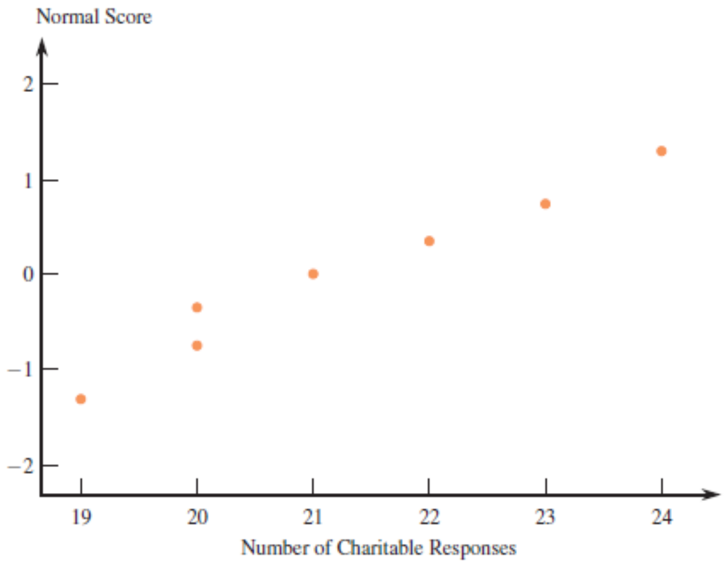
A **normal probability plot** is a scatterplot of the (normal score, observed value) pairs.

A strong linear pattern in a normal probability plot suggests that population normality is plausible. On the other hand, systematic departure from a straight-line pattern (such as curvature in the plot) indicates that it is not reasonable to assume that the population distribution is normal.

$$\text{z score} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

23	22	21	24	19	20	20

Sample mean $\bar{x}$= 21.29; Sample standard deviation s = 1.80



The plot is reasonably straight, so it seems plausible that the population distribution of the number of charitable responses is approximately normal.

Example

In a study, chimpanzees learned to use an apparatus that dispensed food when either of two ropes was pulled. When one of the ropes was pulled, only the chimp controlling the apparatus received food. When the other rope was pulled, food was dispensed both to the chimp controlling the apparatus and also to a chimp in the adjoining cage. The accompanying data (approximated from a graph in the paper) represent the number of times out of 36 trials that each of seven chimps chose the option that would provide food to both chimps (the "charitable" response).

23	22	21	24	19	20	20

$\bar{x}$ = 21.29 and s = 1.80

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

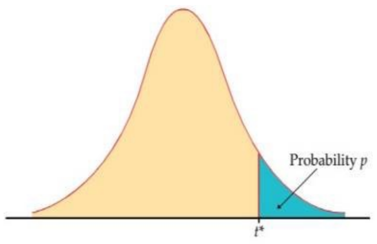


TABLE D

t distribution critical values

df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
	Upper-tail probability p											
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Compute a 99% confidence interval for the mean number of charitable responses for the population of all chimps.

The t critical value for a 99% confidence interval based on 6 degree of freedom is 3.71.

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right) = 21.29 \pm (3.71)\left(\frac{1.80}{\sqrt{7}}\right)$$
$$= 21.29 \pm 2.52$$
$$= (18.77, 23.81)$$

We are 99% confident that the mean number of charitable responses for the population of all chimps is between 18.77 and 23.81.

Example

In a personnel monitoring field study, 100 people work in a large room (such as a warehouse) where carbon monoxide (CO) is present. Suppose 12 people are selected at random on a particular date, and their CO exposure is measured over 8 hours. The resulting 12 samples give a mean value of 10.5 ppm and a standard deviation of 1.9 ppm. Assuming the CO exposure follows a normal distribution. Can we said that the true mean lies within 10.5 ± 1 ppm with 95% assurance?

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

TABLE D

t distribution critical values

	Upper-tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Example

In a personnel monitoring field study, 100 people work in a large room (such as a warehouse) where carbon monoxide (CO) is present. Suppose 12 people are selected at random on a particular date, and their CO exposure is measured over 8 hours. The resulting 12 samples give a mean value of 10.5 ppm and a standard deviation of 1.9 ppm. Assuming the CO exposure follows a normal distribution. Can we said that the true mean lies within 10.5 $\pm$ 1 ppm with 95% assurance?

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

The 95% CI:

$\text{df} = \text{n-1} = 11$, $t = 2.2$ (from t-table)

The confidence interval is $\left(10.5 - 2.2\frac{1.9}{\sqrt{12}}, 10.5 + 2.2\frac{1.9}{\sqrt{12}}\right)$

$= (10.5 - 1.2, 10.5 + 1.2) = (9.3, 11.7)$

Confidence interval is $\pm$ 1.2 ppm, it is wider than the $\pm$ 1 ppm confidence interval

TABLE D

t distribution critical values

	Upper-tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	$\boxed{2.201}$	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Choosing the Sample Size

The bound-on error of estimation associated with a 95% confidence interval is

$$B = 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Solve this for *n* :

$$n = \left(\frac{1.96\sigma}{B}\right)^2$$

Notice that the greater the variability in the population (larger σ), the greater the required sample size will be.

One possible strategy is to use the **sample standard deviation** (or a somewhat larger value, to be conservative) to determine *n.*

Choosing the Sample Size

The bound-on error of estimation associated with a 95% confidence interval is

Recall:

$$n = \left(\frac{1.96\sigma}{B}\right)^2$$

Notice that the greater the variability in the population (larger σ), the greater the required sample size will be.

Another possibility is simply to make an educated guess about the value of s and to use that value to calculate *n*.

For a population distribution that is not too skewed, **dividing the anticipated range (the difference between the largest and the smallest values) by 4** often gives a rough idea of the value of the standard deviation.

Choosing the Sample Size

The sample size required to estimate a population mean μ to within an amount B with 95% confidence is

$$n = \left(\frac{1.96\sigma}{B}\right)^2$$

If σ is unknown, it may be estimated based on previous information or, for a population that is not too skewed, by using (range)/4.

If the desired confidence level is something other than 95%, 1.96 is replaced by the appropriate z critical value (for example, 2.58 for 99% confidence).

Example

The financial aid office wishes to estimate the mean cost of textbooks per quarter for students at a particular university. For the estimate to be useful, it should be within \$20 of the true population mean. The financial aid office is pretty sure that the amount spent on books varies widely, with most values between \$150 and \$550. How large a sample should be used to be 95% confident of achieving this level of accuracy?

To determine the required sample size, we must have a value for σ.

A reasonable estimate of σ is then $\frac{range}{4} = \frac{550 - 150}{4} = \frac{400}{4} = 100$

The required sample size is $n = \left(\frac{1.96\sigma}{B}\right)^2 = \left(\frac{(1.96)(100)}{20}\right)^2 = (9.8)^2 = 96.04$

Rounding up, a sample size of 97 or larger is recommended.

Communicating the Results of Statistical Analyses

When using sample data to estimate a population characteristic, a point estimate or a confidence interval estimate might be used.

- Confidence intervals are generally preferred because a point estimate by itself does not convey any information about the accuracy of the estimate. For this reason, whenever you report the value of a point estimate, it is a good idea to also include an estimate of the bound on the error of estimation.