

Probability and Statistics

Prepared by Dr. Israa Lewaa

Lecture 6

جامعة الجميع الذكية

تسعى لتكون الكيان الأول الذي يتبنى مبدأ المجتمع التعليمي
التعاوني ويقدم العلم مقابل العمل لتنمية وبناء المجتمعات.

STA 201

Confidence Intervals



جامعة الجميع الذكية
EVERYONE'S SMART UNIVERSITY

6.1 Confidence Intervals for the mean (Large Samples)

6.2 Confidence Intervals for the mean (Small Samples)

6.3 Confidence Interval for the population Proportion

6.4 Confidence Intervals for Variance and Standard Deviation

Confidence Intervals for the Mean (large samples)

- **Estimating Population Parameters**
- **Finding a Point Estimate**
- **Finding the Margin of Error**
- **Confidence Interval for the population mean**
- **Sample Size**

A **point estimate** is a single value estimate for a population parameter.

The best point estimate of the **population** mean μ is the ***sample*** mean $\bar{X}$.

Example: Point Estimate

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

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	

Solution: Point Estimate

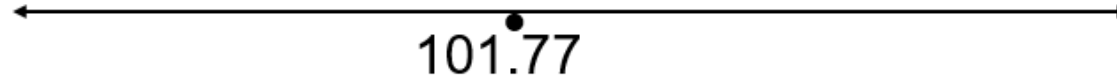
The sample mean is

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

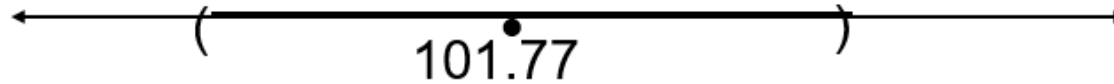
The point estimate for the price of all one way tickets from Atlanta to Chicago is \$101.77.

Interval Estimate

Point estimate



An **interval estimate** is an interval or range of values used to estimate a population parameter.



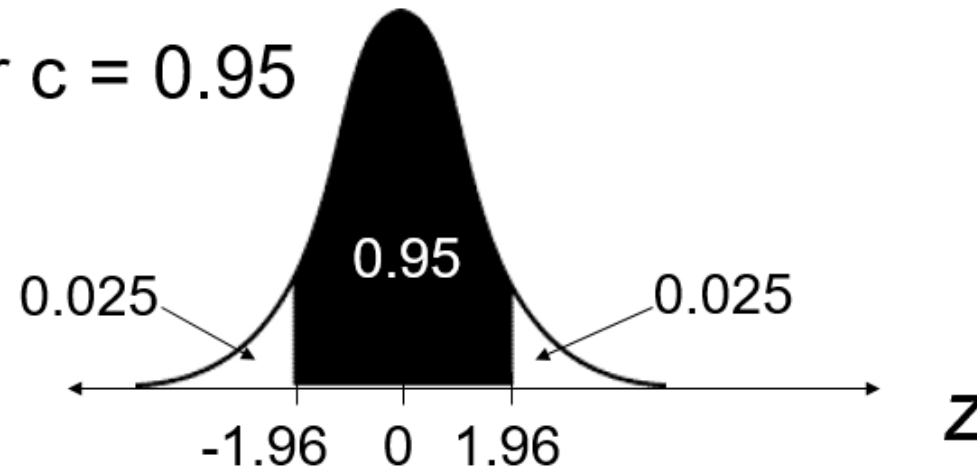
The **level of confidence c** is the probability that the interval estimate contains the population parameter.

Distribution of Sample Means

When the sample size is at least 30, the sampling distribution for $\bar{X}$ is normal.

Sampling distribution of $\bar{X}$

For $c = 0.95$



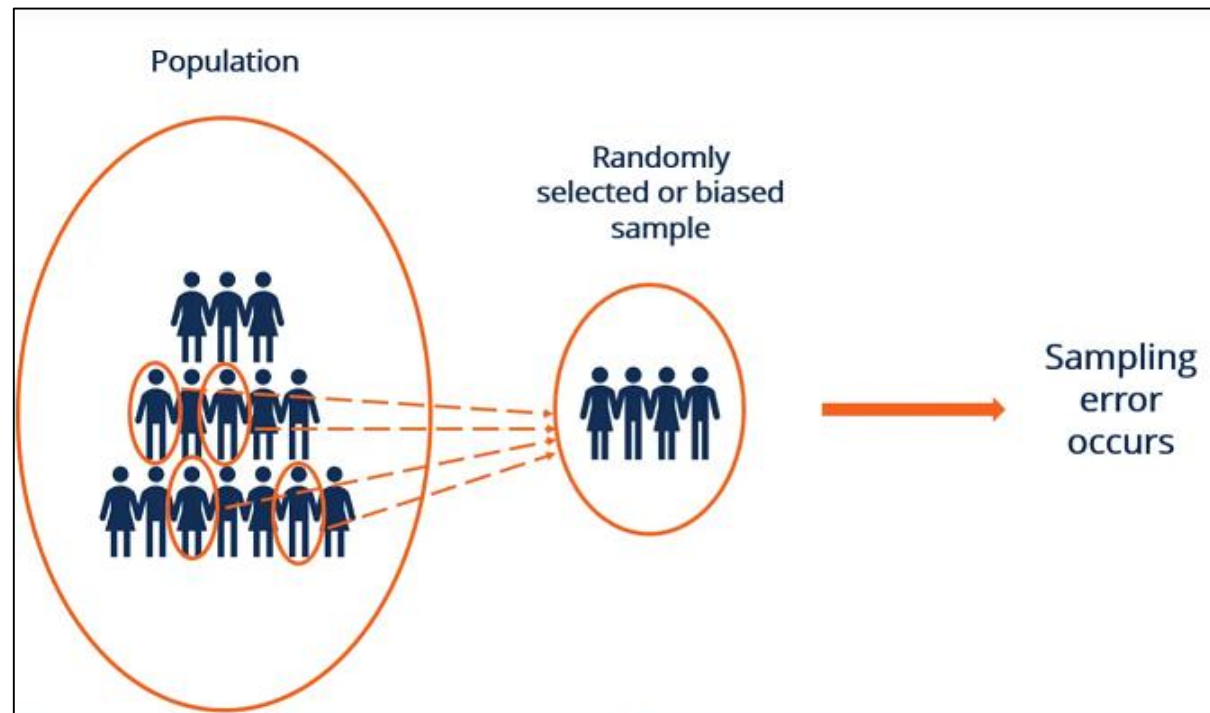
95% of all sample means will have standard scores between $z = -1.96$ and $z = 1.96$

For $c = 90\%$, $z_{.90} = 1.645$

For $c = 95\%$, $z_{.95} = 1.96$

For $c = 99\%$, $z_{.99} = 2.575$

The difference between the point estimate and the actual parameter mean value is called the **sampling error**.



The **margin of error** E is the greatest possible distance between the point estimate and the value of the parameter it is estimating for a given **level of confidence** c .

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

When $n \geq 30$, the sample standard deviation s can be used in place of σ

Example: Maximum Error of Estimate

Find E , the maximum error of estimate for the one-way plane fare from Atlanta to Chicago for a 95% level of confidence given $\sigma = 6.69$.

Solution:

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

Using $z_c = 1.96$, $s = 6.69$, and $n = 35$,

$$E = z_c \frac{\sigma}{\sqrt{n}} = 1.96 \frac{6.69}{\sqrt{35}} = 2.22$$

You are 95% confident that the maximum error of estimate is \$2.22.

A **confidence interval** for the population mean is

$$\bar{X} - E < \mu < \bar{X} + E$$

Find the 95% confidence interval for the one-way plane fare from Atlanta to Chicago.

Confidence Intervals for μ

$$\bar{X} - E < \mu < \bar{X} + E$$

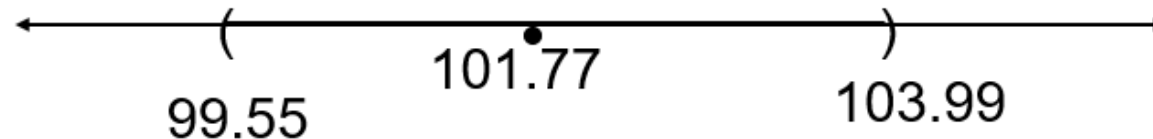
We found $\bar{X} = 101.77$ and $E = 2.22$

Left endpoint

$$\bar{X} - E = 101.77 - 2.22 = 99.55$$

Right endpoint

$$\bar{X} + E = 101.77 + 2.22 = 103.99$$



$$99.55 < \mu < 103.99$$

With 95% confidence, you can say the mean one-way fare from Atlanta to Chicago is between \$99.55 and \$103.99.

Given a confidence level and a margin of error, the minimum **sample size n** needed to estimate the population mean is

$$n = \left(\frac{Z_c \sigma}{E} \right)^2$$

Example: Sample Size

You want to estimate the mean one-way fare from Atlanta to Chicago. How many fares must be included in your sample if you want to be 95% confident that the sample mean is within \$2 of the population mean?

$$n = \left(\frac{Z_c \sigma}{E} \right)^2 = \left(\frac{1.96 \times 6.69}{2} \right)^2 = 42.98$$

You should include at least 43 fares in your sample. Since you already have 35, you need 8 more.

Confidence Intervals for the Mean (Small samples)

Small sample size and large sample size

The t distribution

Critical value from t distribution

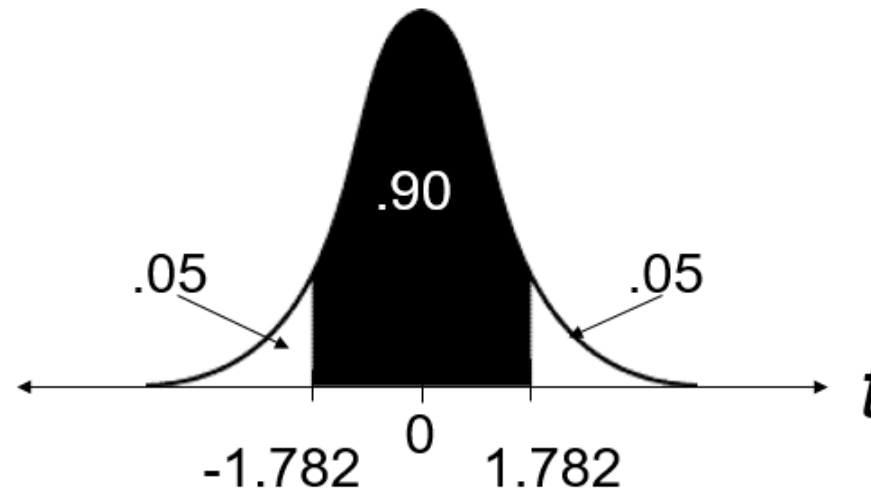
Confidence interval using t distribution

The T Distribution

If the distribution of a random variable x is normal and $n < 30$, then the sampling distribution of $\bar{x}$ is a **t -distribution with $n - 1$ degrees of freedom.**

Sampling distribution $\bar{x}$

$n = 13$
d.f. = 12
 $c = 90\%$



The critical value for t is 1.782. 90% of the sample means ($n = 13$) will lie between $t = -1.782$ and $t = 1.782$.

Example: t-distribution

Finding Critical Values of t

Find the critical value for a 95% confidence level when the sample size is 15.

Solution:

Because $n = 15$, the degrees of freedom are $n-1 = 15-1 = 14$

portion of t table 5 is shown. Using d.f. = 14 and $c = 0.95$, you can find the critical value as shown by the highlighted areas in the table, which equal to 2.145

	Level of confidence, c	0.50	0.80	0.90	0.95	0.98
	One tail, α	0.25	0.10	0.05	0.025	0.01
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02
1		1.000	3.078	6.314	12.706	31.821
2		.816	1.886	2.920	4.303	6.965
3		.765	1.638	2.353	3.182	4.541
12		.695	1.356	1.782	2.179	2.681
13		.694	1.350	1.771	2.160	2.650
14		.692	1.345	1.761	2.145	2.624
15		.691	1.341	1.753	2.131	2.602
16		.690	1.337	1.746	2.120	2.583
28		.683	1.313	1.701	2.048	2.467
29		.683	1.311	1.699	2.045	2.462
∞		.674	1.282	1.645	1.960	2.326

T- distribution table

d.f.	Level of confidence, c	0.50	0.80	0.90	0.95	0.98	0.99
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	1.638	2.353	3.182	4.541	5.841
4		.741	1.533	2.132	2.776	3.747	4.604
5		.727	1.476	2.015	2.571	3.365	4.032
6		.718	1.440	1.943	2.447	3.143	3.707
7		.711	1.415	1.895	2.365	2.998	3.499
8		.706	1.397	1.860	2.306	2.896	3.355
9		.703	1.383	1.833	2.262	2.821	3.250
10		.700	1.372	1.812	2.228	2.764	3.169
11		.697	1.363	1.796	2.201	2.718	3.106
12		.695	1.356	1.782	2.179	2.681	3.055
13		.694	1.350	1.771	2.160	2.650	3.012
14		.692	1.345	1.761	2.145	2.624	2.977
15		.691	1.341	1.753	2.131	2.602	2.947
16		.690	1.337	1.746	2.120	2.583	2.921
17		.689	1.333	1.740	2.110	2.567	2.898
18		.688	1.330	1.734	2.101	2.552	2.878
19		.688	1.328	1.729	2.093	2.539	2.861
20		.687	1.325	1.725	2.086	2.528	2.845
21		.686	1.323	1.721	2.080	2.518	2.831
22		.686	1.321	1.717	2.074	2.508	2.819
23		.685	1.319	1.714	2.069	2.500	2.807
24		.685	1.318	1.711	2.064	2.492	2.797
25		.684	1.316	1.708	2.060	2.485	2.787
26		.684	1.315	1.706	2.056	2.479	2.779
27		.684	1.314	1.703	2.052	2.473	2.771
28		.683	1.313	1.701	2.048	2.467	2.763
29		.683	1.311	1.699	2.045	2.462	2.756
∞		.674	1.282	1.645	1.960	2.326	2.576

Maximum error of estimate $E = t_c \frac{s}{\sqrt{n}}$

Where;

S is the Sample standard deviation

t_c is the value obtained from t table

n is the sample size

Example: Maximum error of Estimate

In a random sample of 13 American adults, the mean waste recycled per person per day was 4.3 pounds and the standard deviation was 0.3 pound. Assume the variable is normally distributed and construct a 90% confidence interval for μ .

1. The point estimate is $\bar{X} = 4.3$ pounds
2. The maximum error of estimate is

$$E = t_c \frac{s}{\sqrt{n}} = 1.782 \frac{0.3}{\sqrt{13}} = 0.148$$

Confidence Interval – Small Sample

1. The point estimate is $\bar{X} = 4.3$ pounds

2. Confidence Interval: $\bar{X} - E < \mu < \bar{X} + E$

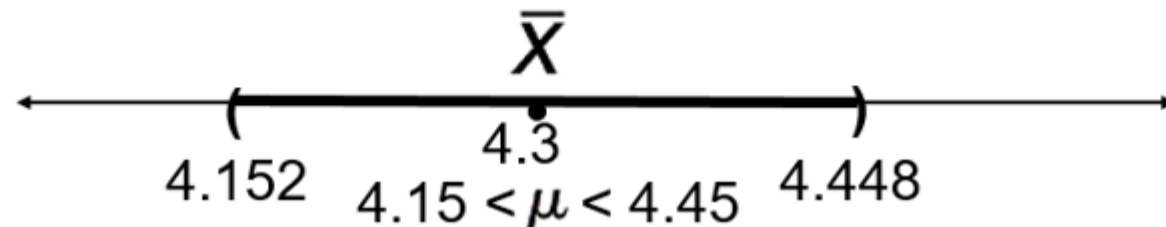
$$E = t_c \frac{s}{\sqrt{n}} = 1.782 \frac{0.3}{\sqrt{13}} = 0.148$$

Left endpoint

$$\bar{X} - E = 4.3 - 0.148 = 4.152$$

Right endpoint

$$\bar{X} + E = 4.3 + 0.148 = 4.448$$



6.3 Confidence Intervals for Population Proportions

- Point Estimate for Proportion Proportions
- Confidence interval for Proportion Proportions
- Finding the minimum Sample Size

Confidence Intervals for Population Proportions

The point estimate for p , the population proportion of successes, is given by the proportion of successes in a sample

$$\hat{p} = \frac{x}{n} \quad (\text{Read as p-hat})$$

$\hat{q}$ is the point estimate for the proportion of failures where

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

If $np \geq 5$ and $nq \geq 5$ the sampling distribution for $\hat{p}$ is normal.

Confidence Intervals for Population Proportions

The maximum error of estimate, E , for a c -confidence interval is:

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

A c -confidence interval for the population proportion, p , is

$$\hat{p} - E < p < \hat{p} + E$$

Example: Confidence Intervals for Population Proportions

In a study of 1907 fatal traffic accidents, 449 were alcohol related. Construct a 99% confidence interval for the proportion of fatal traffic accidents that are alcohol related.

Solution : Confidence Intervals for Population Proportions

1. The point estimate for p is $\hat{p} = \frac{x}{n} = \frac{449}{1907} = 0.235$
 $\hat{q} = 1 - 0.235 = 0.765$
2. $1907(.235) \geq 5$ and $1907(.765) \geq 5$, so the sampling distribution is normal.
3. $E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.575 \sqrt{\frac{(.235)(.765)}{1907}} = 0.025$

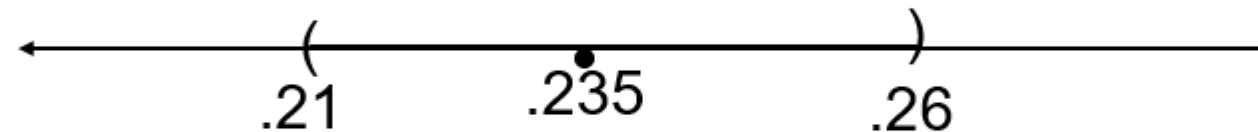
Solution : Confidence Intervals for Population Proportions

Left endpoint

$$\hat{p} - E = 0.235 - 0.025 = 0.21$$

Right endpoint

$$\hat{p} + E = 0.235 + 0.025 = 0.26$$



$$0.21 < p < 0.26$$

With 99% confidence, you can say the proportion of fatal accidents that are alcohol related is between 21% and 26%.

Minimum Sample Size

If you have a preliminary estimate for p and q , the minimum sample size given a x -confidence interval and a maximum error of estimate needed to estimate p is:

$$n = \hat{p}\hat{q} \left(\frac{z_c}{E} \right)^2$$

If you do not have a preliminary estimate, use 0.5 for both $\hat{p}$ and $\hat{q}$.

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 accurate to within 2% of the population proportion

Solution: Minimum Sample Size

With no preliminary estimate use 0.5 for $\hat{p}$ and $\hat{q}$

$$n = \hat{p}\hat{q} \left(\frac{z_c}{E} \right)^2 = (0.5)(0.5) \left(\frac{2.575}{0.02} \right)^2 = 4414.14$$

You will need at least 4415 for your sample.

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 accurate to within 2% of the population proportion. Use a preliminary estimate of $p = 0.235$.



Solution: Minimum Sample Size

$$n = \hat{p}\hat{q} \left(\frac{z_c}{E} \right)^2$$

$$n = (0.235)(0.765) \left(\frac{2.575}{0.02} \right)^2 = 2980.05$$

With a preliminary sample you need at least **$n = 2981$** for your sample.

- **Point Estimate for σ^2 and σ**
- **The Chi-Square Distribution**
- **Confidence Interval for σ^2 and σ**

The point estimate for σ^2 is s^2 and the point estimate for σ is s .

Without calculating the exact standard deviations of the following sets, determine which of them has the lowest standard deviation.

Set	Mean
(A) ✓ 41, 41, 41, 41, 41, 42	41.16
(B) ✗ 35, 38, 40, 42, 44, 48	43.16
(C) ✗ 75, 75, 75, 75, 75, 1500	312.5
(D) ✗ 10, 20, 30, 40, 50, 60	35
(E) ✗ 100, 200, 300, 400, 500, 600	350

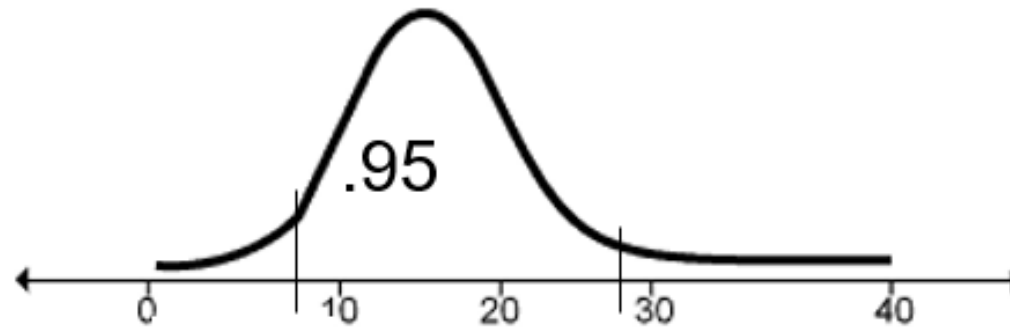
Standard Deviation

$\sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$

number of items mean

The Chi-Square Distribution

If the sample size is n , use a chi-square χ^2 distribution with $n - 1$ degrees of freedom to form a c -confidence interval.



Example: The Chi-Square Distribution

Find χ_R^2 the right-tail critical value and χ_L^2 the left-tail critical value for $c = 95\%$ and $n = 17$.

Solution:

When the sample size is 17, the degrees of freedom = $n-1 = 17-1 = 16$

- Area to the right of χ_R^2 is $(1 - 0.95)/2 = 0.025$
- Area to the right of χ_L^2 is $(1 + 0.95)/2 = 0.975$

Example: The Chi-Square Distribution

Degrees of freedom	α							
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845

χ_L^2 χ_R^2

Chapter 06

The Chi-Square Table

Degrees of freedom	α									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Confidence Interval for σ^2 and σ

A c-confidence interval for a population variance is:
$$\frac{(n-1)s^2}{X_R^2} < \sigma^2 < \frac{(n-1)s^2}{X_L^2}$$

To estimate the standard deviation take the square root of each endpoint.

You randomly select the prices of 17 CD players. The sample standard deviation is \$150. Construct a 95% confidence interval for σ^2 and σ .

$$\frac{(17-1)150^2}{28.845} < \sigma^2 < \frac{(17-1)150^2}{6.908}$$

$$12480.50 < \sigma^2 < 52113.49$$

Confidence Interval for σ^2 and σ

To estimate the standard deviation take the square root of each endpoint.

$$12480.50 < \sigma^2 < 52113.49$$

Find the square root of each part.

$$\$117.72 < \sigma < \$228.28$$

You can say with 95% confidence that σ^2 is between 12480.50 and 52113.49 and σ between \$117.72 and \$228.28.

Notation	Definition
μ	Population mean
$\bar{X}$	Sample mean
σ^2	Population Variance
S^2	Sample Variance
σ	Population Standard deviation
S	Sample Standard deviation
P	Population Proportion
$\hat{p}$	Sample Proportion

Confidence Intervals for the mean (Large Samples)

$$\bar{X} - E < \mu < \bar{X} + E$$

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals for the mean (Small Samples)

$$\bar{X} - E < \mu < \bar{X} + E$$

$$E = t_c \frac{s}{\sqrt{n}}$$

Confidence Interval for the population Proportion

$$\hat{p} - E < p < \hat{p} + E$$
$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Confidence Intervals for Variance and Standard Deviation

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

Sample Size

Estimate Mean:

$$n = \left(\frac{Z_c \sigma}{E} \right)^2$$

Estimate Proportion:

$$n = \hat{p}\hat{q} \left(\frac{Z_c}{E} \right)^2$$