Stat 515: Introduction to Statistics

Chapter 7

Confidence Intervals

- Often, we do not know the **population parameter,** $\boldsymbol{\mu, \rho\ or\ \sigma_x}$
- We use our **sample statistics,** $\boldsymbol{\overline{x}\ , \widehat{p}, s_x}$ to make **inference** on the **population parameter,** $\boldsymbol{\mu, \rho\ or\ \sigma_x}$

Confidence Intervals

- **First,** we will consider an interval estimate which we call a confidence interval

(This is our plus/minus from chapter 1)

$$\boldsymbol{point\ estimate \pm margin\ of\ error}$$

$$\boldsymbol{= point\ estimate \pm \begin{pmatrix} confidence \\ coefficient \end{pmatrix} * \begin{pmatrix} \widehat{Standard} \\ Error \end{pmatrix}}$$

Telling Which Parameter We're After

- As statisticians, or data scientists, it's our job to hear a problem and decide what we're after
  - We call the parameter of interest the **target parameter**

Parameter	Point Estimate	Key Phrase	Type of Data
μ	$\bar{x}$	Mean, Average	Quantitative
ρ	$\hat{p}$	Proportion, percentage, fraction, rate	Qualitative (Categorical)
σ^2	s_x^2	Variance, variability, spread	Quantitative

Confidence Intervals for Population Proportions on YouTube

- Intro:
  - https://www.youtube.com/watch?v=3ReWri_jh3M

Recall Sampling Distributions for Sampling Proportions

- Recall: the mean of the sampling distribution for a sample proportion will always equal the population proportion: $\boldsymbol{\mu_{\hat{p}} = \rho}$
- The standard error, the standard deviation of the sample proportion, is:

$$\boldsymbol{\sigma_{\hat{p}} = \sqrt{\frac{\rho(1-\rho)}{n}}}$$

Confidence Intervals: Step One

- **Assumptions:**
  1. Data must be obtained through randomization
  2. We **MUST** make sure that $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$. This ensures that $\hat{p}$ follows a bell shaped distribution
     - Recall Chapter 4 and the shape of the binomial dist.

Confidence Intervals: Step Two

- Recall: $\hat{p}$ is our **point-estimate** for the population proportion

- Recall we consider $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ when we don't know ρ for the standard error as $\hat{p}$ can estimate the value of ρ

Confidence Intervals: Step Two

- $\hat{p}$ is our **point-estimate** for the population proportion
  - Our 'best' guess for the **true population proportion, $\boldsymbol{\rho}$,** is our **sample proportion, $\hat{p}$.**

Confidence Intervals: Step Two

- $z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**

  - $z_{1-\frac{\alpha}{2}}$ is the **confidence coefficient** and is the z value such that $P\left(Z < z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1 - \frac{\alpha}{2}$

  - $\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is the **estimated standard deviation**

Confidence Intervals – Step Two

- The most common values of Z are listed below
  - **Level of confidence** = (1-∝) * 100%
  - **Error Probability** = ∝ = 1- Level of confidence

Confidence	Error Probability (∝)	$z_{\left(1-\frac{\alpha}{2}\right)}$ From Table	$z_{\left(1-\frac{\alpha}{2}\right)}$ From R
.9	.1	1.645	1.644854
.95	.05	1.96	1.959964
.99	.01	2.58	2.57829

  - Our interval will get larger when the margin of error increases
    1) When we increase confidence → increase z → widen interval
    2) When we decrease confidence → decrease z → narrow interval

Confidence Intervals: Step Two

- $z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**
  - **As n increases**, the margin of error decreases causing the width of the confidence interval to narrow
  - **As n decreases**, the margin of error increases causing the width of the confidence interval to grow wider

Confidence Intervals: Margin of Error

- $z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**
  - **As the confidence level decreases**, z decreases causing the margin of error to decrease, causing the width of the confidence interval to narrow
  - **As the confidence level increases**, z increases causing the margin of error to increase, causing the width of the confidence interval to grow wider

Confidence Intervals – Step Two

- A fishing metaphor:
  - **As n increases** → confidence interval narrows
  - **As n decreases**→ confidence interval widens

  - Think about fishing in a pond with a net. If there are more fish you can use a smaller net to catch the fish.
  - In our case, when our sample size is larger we can use a smaller interval to catch our parameter.

Confidence Intervals – Step Two

- A fishing metaphor:
  - **Increase confidence** → confidence interval narrows
  - **Decrease confidence** → confidence interval widens

  - Think about fishing in a pond with a net. We want to be more certain that we'll catch a fish we need a bigger net.
  - In our case, when we increase confidence to be more certain that we'll catch the parameter, we need a bigger interval.

Confidence Intervals Bounds

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

Lower Bound = $\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$

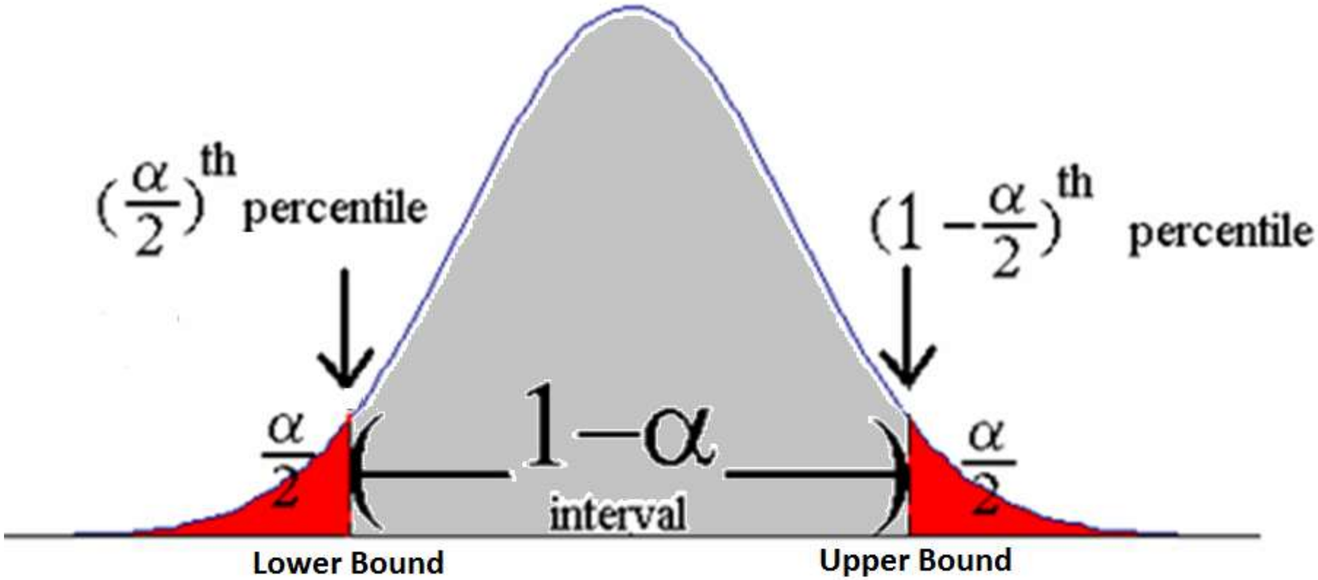
Upper Bound = $\hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$

Confidence Intervals Bounds

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

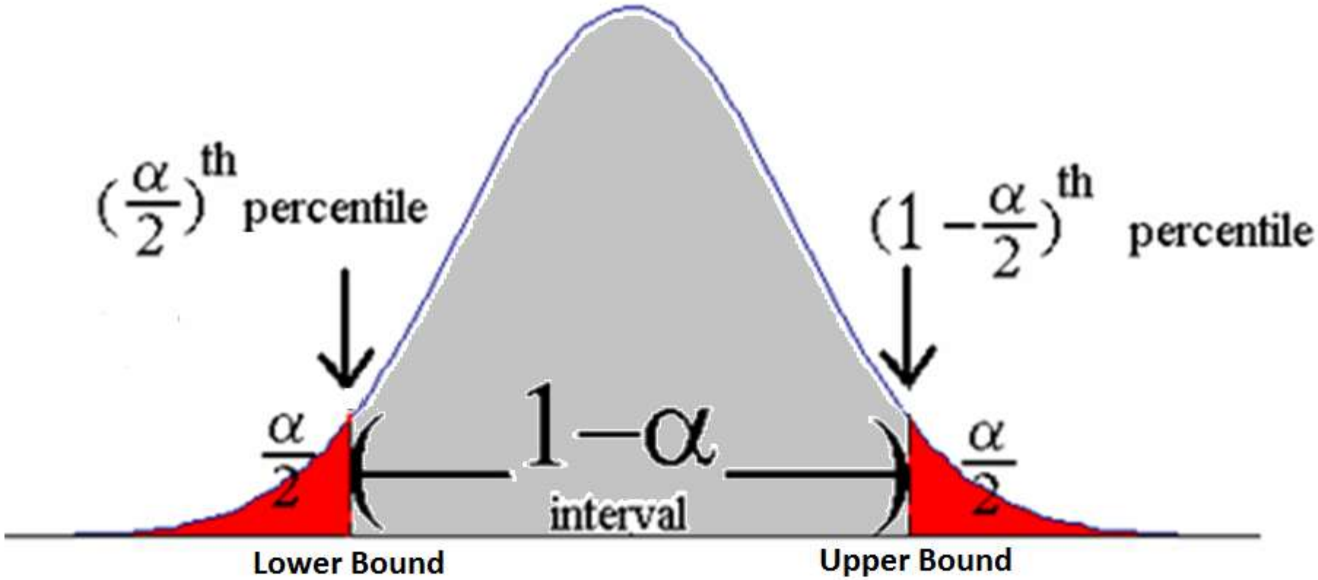
“We are --% confident that the true population proportion, *ρ*, is between the lower bound and upper bound.”

Confidence Intervals



- We choose our values such that
  - Our **point estimate** is the mean, the 50th percentile
  - Our **lower bound** is the $\frac{\alpha}{2}^{\text{th}}$ percentile
  - Our **upper bound** is the 1-$\frac{\alpha}{2}^{\text{th}}$ percentile

How We Found the Common Z's: 90%

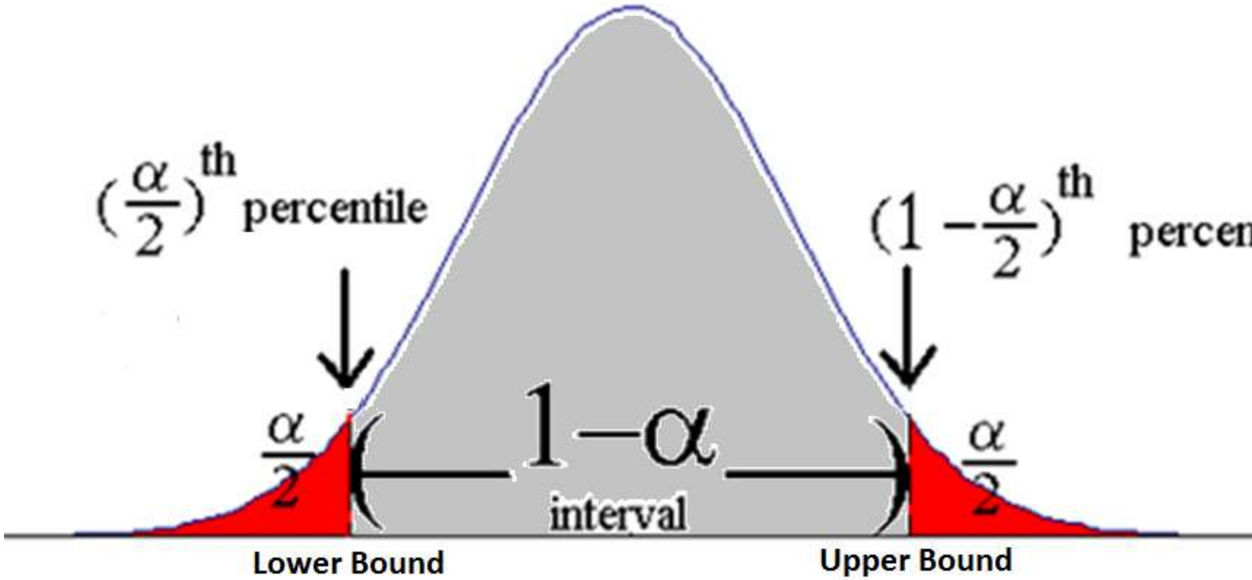


- For a 90% confidence interval upper bound, we need to find the z with a percentile of

$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.90}{2}=1-\frac{.10}{2}=.9500$$

- If we look this up in the z-table we see that a z-score between 1.64 or 1.65 gives us a value very close to .9500 → 1.645

How We Found the Common Z's: 90%



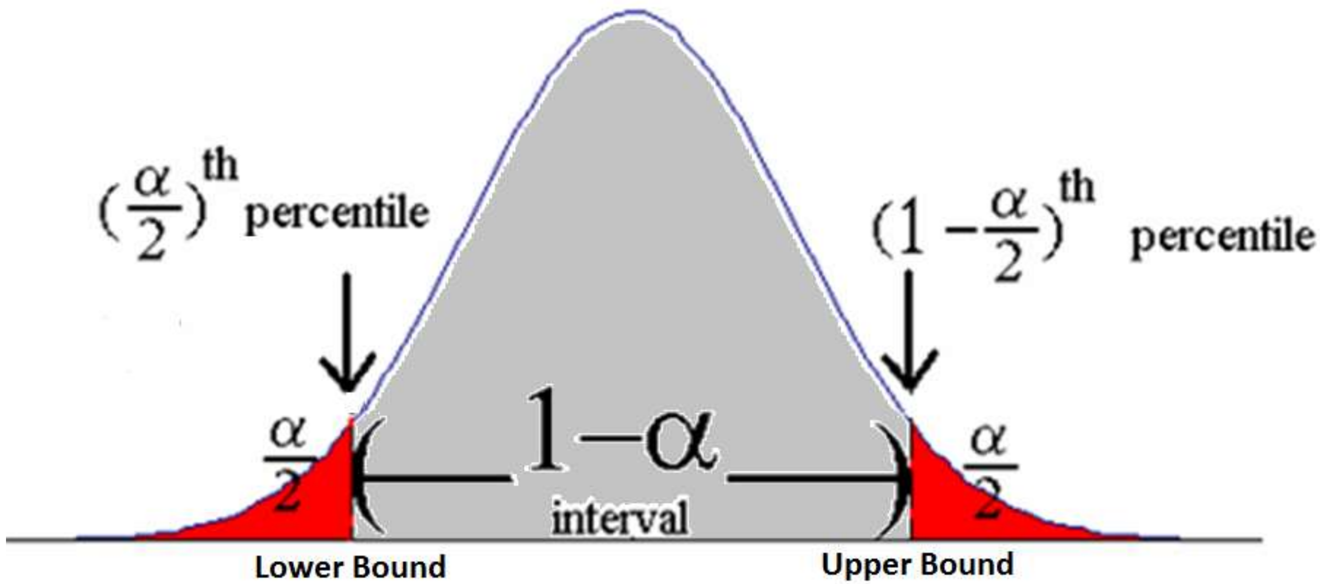
- For a 90% confidence interval upper bound, we need to find the z with a percentile of $1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.90}{2}=1-\frac{.10}{2}=.9500$
- To look this up in R: qnorm(.9500,0,1)=1.644854

How We Found the Common Z’s: 90%

- **Lower Bound:** If we look this up in the z-table we see that a z-score between -1.65 or -1.64 gives us a value very close to .0500
- **Upper Bound:** If we look this up in the z-table we see that a z-score between 1.65 or 1.64 gives us a value very close to .9500
- Since it’s in the middle we average 1.64 and 1.65

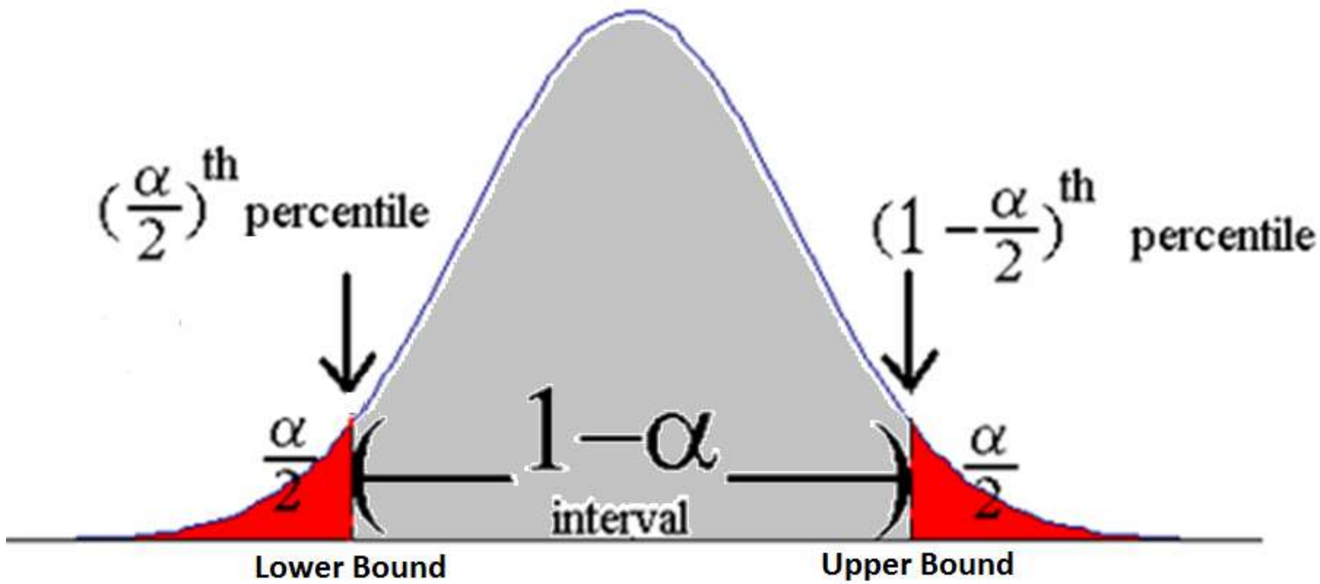
- This is why we have plus or minus z=1.645 for a 90% confidence interval

How We Found the Common Z's: 95%



- For a 95% confidence interval upper bound, we need to find the z with a percentile of $$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.95}{2}=1-\frac{.05}{2}=.9750$$
- If we look this up in the z-table we see that a z-score of 1.96 gives us a value very close to .9750

How We Found the Common Z's: 95%



- For a 95% confidence interval upper bound, we need to find the z with a percentile of

$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.95}{2}=1-\frac{.05}{2}=.9750$$

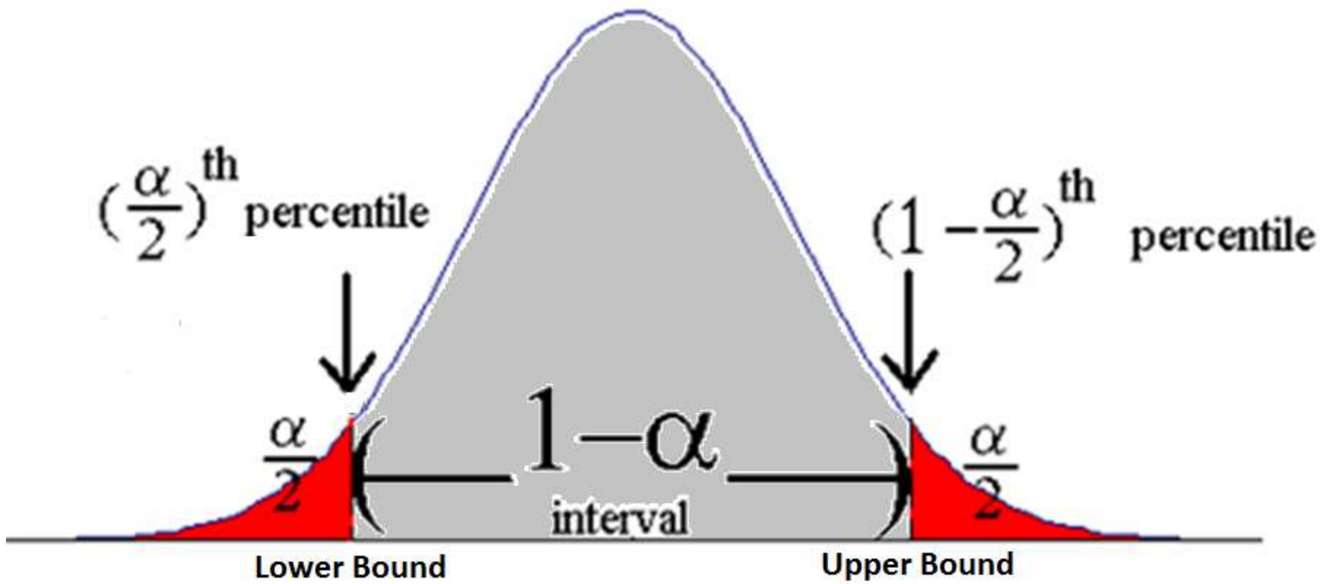
- To look this up in R: qnorm(.9750,0,1)=1.959964

How We Found the Common Z's: 95%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -1.96 gives us a value very close to .0250
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 1.96 gives us a value very close to .9750

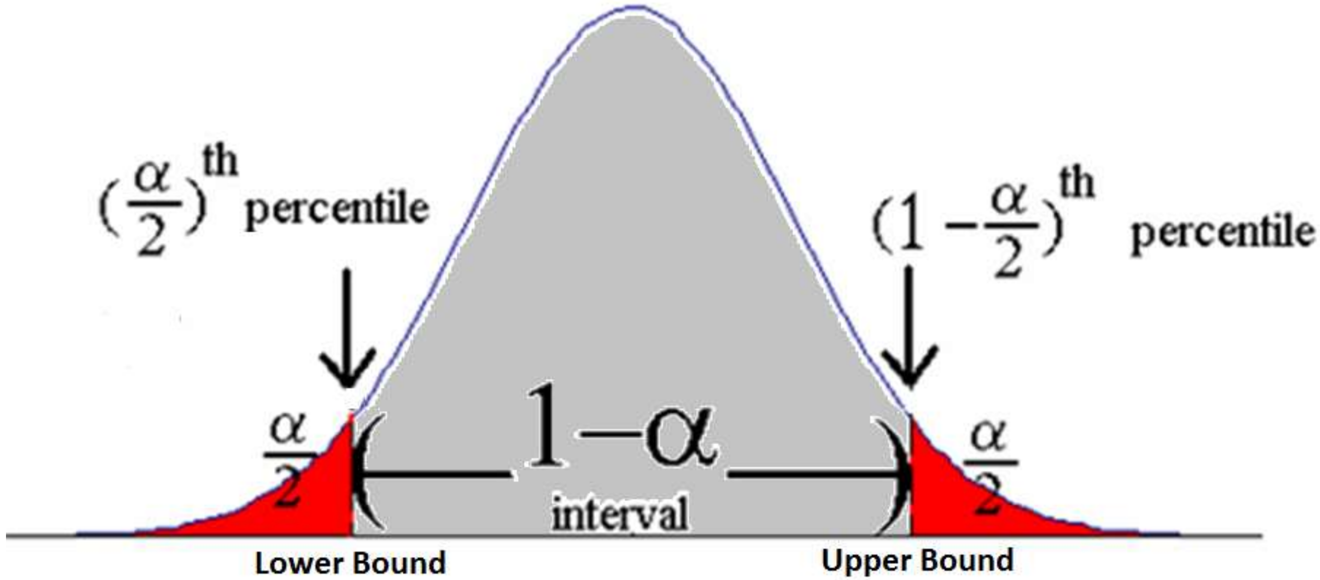
- This is why we have plus or minus z=1.96 for a 95% confidence interval

How We Found the Common Z's: 99%



- For a 99% confidence interval upper bound, we need to find the z with a percentile of
$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.99}{2}=1-\frac{.01}{2}=.9950$$
- If we look this up in the z-table we see that a z-score of 2.58 gives us a value very close to .9950

How We Found the Common Z's: 99%



- For a 99% confidence interval upper bound, we need to find the z with a percentile of

$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.99}{2}=1-\frac{.01}{2}=.9950$$

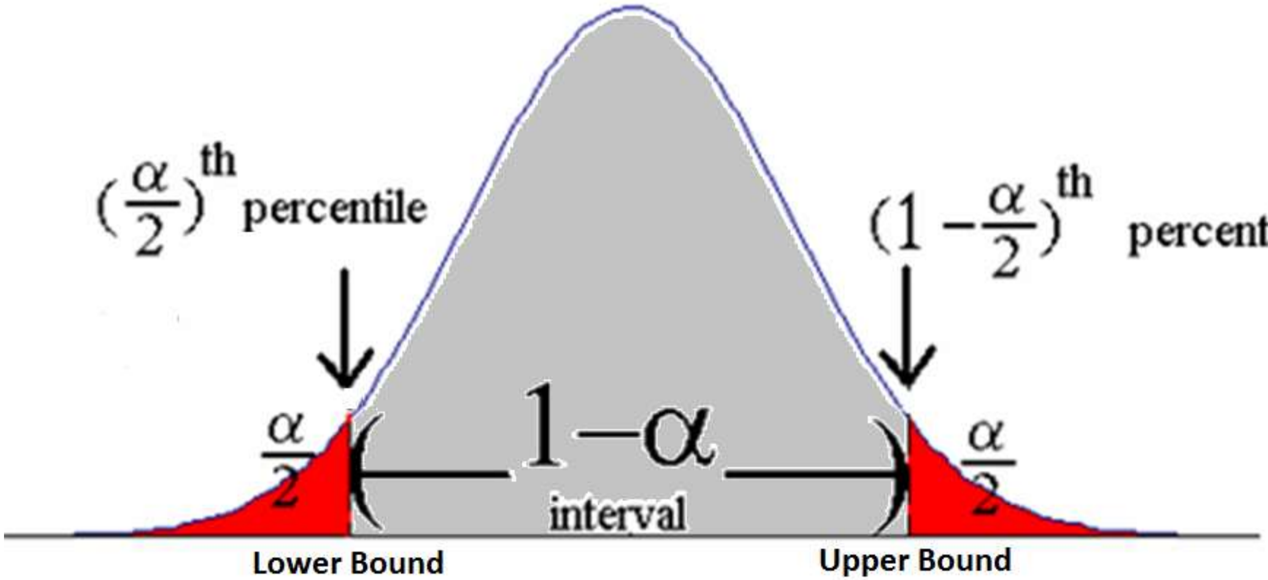
- To look this up in R: qnorm(.9500,0,1)=2.575829

How We Found the Common Z's: 99%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -2.58 gives us a value very close to .0500
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 2.58 gives us a value very close to .9500

- This is why we have plus or minus z=2.58 for a 99% confidence interval

How We Find an Uncommon Z: 98%

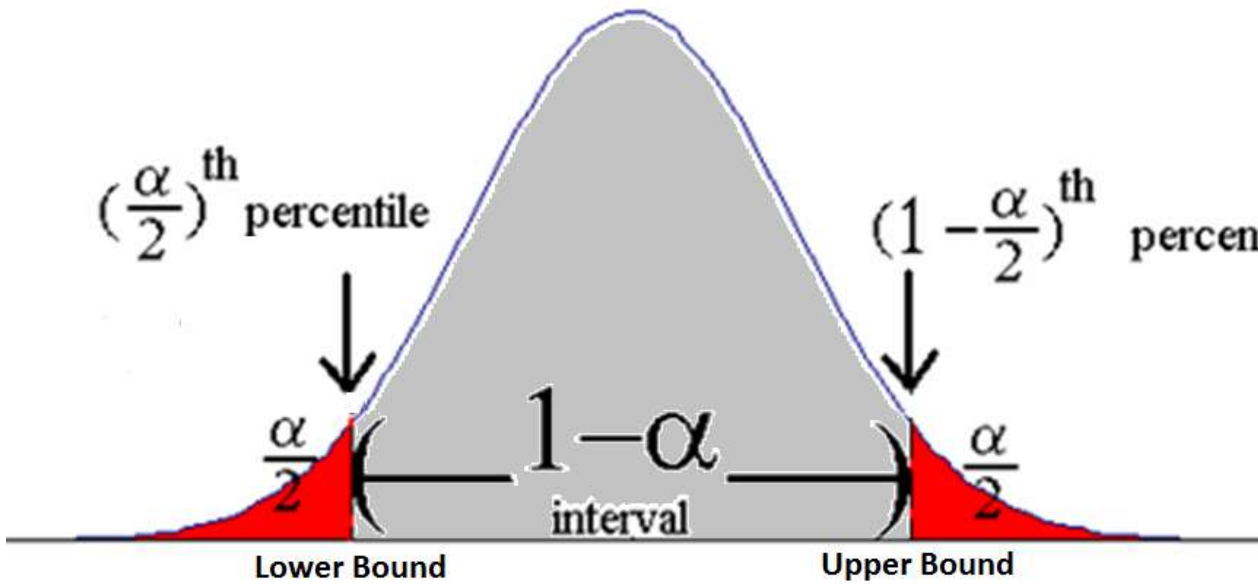


- For a 98% confidence interval lower bound, we need to find the z with a percentile of

$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.98}{2}=1-\frac{.02}{2}=.9900$$

- If we look this up in the z-table we see that a z-score of 2.33 gives us a value very close to .9900

How We Found the Common Z's: 98%



- For a 98% confidence interval upper bound, we need to find the z with a percentile of

$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.98}{2}=1-\frac{.02}{2}=.9900$$

- To look this up in R: qnorm(.9900,0,1)=2.326348

How We Found the Common Z's: 98%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -2.33 gives us a value very close to .0100
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 2.33 gives us a value very close to .9900

- This is why we have plus or minus z=2.33 for a 98% confidence interval

Examples

Example

- A random sample of MLB home games showed that the home teams **won 1335 of 2429 games.**
- Our **sample proportion** = $\hat{p} = \frac{1335}{2429} = .5496$
- We should know this is a proportion problem because we're considering a qualitative (categorical) random variable
- **Find the 95% confidence interval for the population proportion**

Example

- **Step One:**
- Check Assumptions:
  - $n * \hat{p} = 2429 * .5496 = 1334.9784 \geq 15$
  - $n * (1 - \hat{p}) = 2429 * .4504 = 1094.0216 \geq 15$
  - Thus, it is safe to assume the distribution of $\hat{p}$ has a bell shaped distribution
  - The data is from a random sample

Example

- **Step Two:**
- 95% CI:

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

$$.5496 \pm (1.96) \sqrt{\frac{.5496(.4504)}{2429}}$$

$$= (.5298, .5694)$$

- We are 95% confident that the **true population proportion** of home team wins **is between** 52.98 and 56.94 percent.

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- 95% CI:

$$(.5298, .5694)$$

- We see here that there is a small home field advantage because all of the values in our 95% CI are above 0.5.
  – We know that 0.5 is interesting because it means **more than half the time** or **most**

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- **99% CI:**

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

$$.549 \pm (\mathbf{2.58}) \sqrt{\frac{.549(.451)}{2429}} = (.5236, .5756)$$

- We are **99% confident** that the true population proportion of home team wins is between 52.36 and 57.56 percent.

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- 99% CI:

$$(.5236, .5756)$$

- Still, we see here that there is a small home field advantage but we note the interval is larger

Wilson’s Adjustment for Estimating ρ

- Wilson’s Adjustment is a nice trick to ‘correct’ our confidence interval when n isn’t extremely large and performs poorly when ρ is near 0 or 1

$$\tilde{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\left(\tilde{p}(1-\tilde{p})\right)}{n}}$$

- Where $\tilde{p} = \frac{\mathrm{x+2}}{\mathrm{n+4}}$ is the adjusted proportion of observations

Example

- Let’s complete our previous example about MLB home games with Wilson’s Adjustment this time
  - The only difference here will be how we calculate the sample proportion: $\tilde{p} = \frac{x+2}{n+4}$ instead of $\hat{p} = \frac{x}{n}$
  - **Note:** we shouldn’t see a drastic change because we aren’t in the case where n isn’t extremely large and performs poorly when ρ is near 0 or 1

Example

- A random sample of MLB home games showed that the home teams **won 1335 of 2429 games.**
- Our **sample proportion** $= \tilde{p} = \frac{1335+2}{2429+4} = .5495$
- We should know this is a proportion problem because we're considering a qualitative (categorical) random variable
- **Find the 95% confidence interval for the population proportion**

Example

- **Step One:**
- Check Assumptions:
  - $n * \hat{p} = 2429 * .5496 = 1334.9784 \geq 15$
  - $n * (1 - \hat{p}) = 2429 * .4504 = 1094.0216 \geq 15$
  - Thus, it is safe to assume the distribution of $\hat{p}$ has a bell shaped distribution
  - The data is from a random sample

Example

- **Step Two:**
- 95% CI:

$$\tilde{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\tilde{p}(1-\tilde{p})\right)}{n}}$$

$$.5495 \pm (1.96)\sqrt{\frac{.5495(.4505)}{2429}}$$

$$= (.5297, .5693)$$

- We are 95% confident that the **true population proportion** of home team wins **is between** 52.97 and 56.93 percent.

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- 95% CI:

$$(.5297, .5693)$$

- We see here that there is a small home field advantage because all of the values in our 95% CI are above 0.5.
  - We know that 0.5 is interesting because it means **more than half the time** or **most**

Example in R

Below is a function you can load into R:

```
prop.int<-function(conf.level, x, n, Wilson=FALSE){
  if(Wilson){
    phat=(x+2)/(n+4)
  }else{
    phat=x/n
  }
  z.crit = qnorm(1-(1-conf.level)/2);
  std.error = sqrt(phat*(1-phat)/n);
  MOE=z.crit*std.error;
  c(phat-MOE, phat+MOE)
}
```

Example in R

- You can call the function as below which will provide the **95% confidence** interval for a population proportion from a sample where **1335 of 2429 games** were won from the home team:

```
prop.int(.95, 1335, 2429,Wilson=FALSE)
```

OR with Adjustment

```
prop.int(.95, 1335, 2429,Wilson=TRUE)
```

Determining the Sample Size

- Say we want to set sampling error at SE with $100(1-\alpha)\%$ confidence:

Set: $$z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}} = SE$$

Solve for n: $$n = \frac{\left(z^2_{\left(1-\frac{\alpha}{2}\right)}\left(\hat{p}(1-\hat{p})\right)\right)}{SE^2}$$

Note: n is maximized for $\hat{p}$=.5

Recall Sampling Distributions for Sampling Means

- The mean of the sampling distribution for a sample mean

$$\mu_{\bar{x}} = the\ mean\ of\ all\ possible\ sample\ means = \mu_x = the\ population\ mean$$

- The standard error, the standard deviation of all sample means, is:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Confidence Intervals For the Population Mean

- When we talk about confidence intervals for the population mean we have two approaches
  1. When we know σ_x **(we are rarely in this case)**
  2. When we don't know σ_x

Confidence Intervals When We Know σ_x

- We use our sample means to make inference on the population mean

$$\bar{x} \pm z_{1-\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$$

- $\bar{x}$ is our **point-estimate** for the population mean

- $z_{1-\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$ is our **margin of error**

Confidence Intervals When We Know σ_x

- $\bar{x}$ is our **point-estimate** for the population mean
  - Our ‘best’ guess for the true population , mean is our sample mean

Confidence Intervals: Margin of Error When We Know σ_x

- $z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$ is our **margin of error**
  - **As n increases**, $\left(\frac{\sigma_x}{\sqrt{n}}\right)$ decreases, causing the margin of error to decrease causing the width of the confidence interval to narrow
  - **As n decreases**, $\left(\frac{\sigma_x}{\sqrt{n}}\right)$ increases, causing the margin of error to increase causing the width of the confidence interval to widen

Confidence Intervals: Margin of Error When We Know σ_x

- $z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$ is our **margin of error**
  - **As the confidence level decreases**, z decreases causing the margin of error to decrease, causing the width of the confidence interval to narrow
  - **As the confidence level increases**, z increases causing the margin of error to increase, causing the width of the confidence interval to grow wider

Confidence Intervals Bounds When We Know σ_x

Lower Bound = $\bar{x} - z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$

Upper Bound = $\bar{x} + z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$

We are --% confident that the true population mean, μ_x, is between the **lower** and **upper** bound.

<u>Note:</u> there's an incredible likeliness to confidence intervals for proportions

Confidence Intervals Bounds When We Know σ_x - R code

Below is a function you can load into R:

```
z.int<-function(conf.level, xbar, sigma, n){
  z.crit = qnorm(1-(1-conf.level)/2);
  std.error = sigma/sqrt(n);
  MOE=z.crit*std.error;
  c(xbar-MOE, xbar+MOE)
}
```

Confidence Intervals Bounds When We Know σ_x - R code

- You can call the function as below which will provide the 95% confidence interval for a population mean from a sample of 3 that had mean 5 and known population standard deviation 3:

```
conf.level=.95 #Confidence Level
xbar=5 #Sample Mean
sigma=2 #Population Standard Deviation
n=3 #Sample Size
z.int(conf.level, xbar, sigma, n)
```

Determining the Sample Size

- Say we want to set sampling error at SE with $100(1-\alpha)\%$ confidence:

<u>Set:</u> $$z_{\left(1-\frac{\alpha}{2}\right)}\left(\frac{\sigma_X}{\sqrt{n}}\right) = SE$$

<u>Solve for n:</u> $$n = \frac{\left(z^2_{\left(1-\frac{\alpha}{2}\right)}(\sigma_x)\right)}{SE^2}$$

Confidence Intervals Bounds When We Don’t Know σ_x

- Now, onto the more realistic situation where we don’t know the population standard deviation.

Confidence Intervals When We Don't Know σ_x

- We use our sample means to make inference on the population mean

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

- $\bar{x}$ is our **point-estimate** for the population mean
- $t_{\left(1-\frac{\alpha}{2}, n-1\right)} \left(\frac{s_x}{\sqrt{n}}\right)$ is our **margin of error**
  - s_x **is the sample standard deviation**

Confidence Intervals When We Don't Know σ_x

- $\bar{x}$ is our **point-estimate** for the population mean
  - Our 'best' guess for the true population , mean is our sample mean

Confidence Intervals: Margin of Error When We Don't Know σ_x

- $t_{\left(1-\frac{\propto}{2},n-1\right)}\left(\frac{s_x}{\sqrt{n}}\right)$ is our **margin of error**
  - **As n increases**, t decreases and $\left(\frac{s_x}{\sqrt{n}}\right)$ decreases, causing the margin of error to decrease causing the width of the confidence interval to narrow
  - **As n decreases**, t increases and $\left(\frac{s_x}{\sqrt{n}}\right)$ increases, causing the margin of error to increase causing the width of the confidence interval to widen

Confidence Intervals: Margin of Error When We Don't Know σ_x

- $t_{\left(1-\frac{\propto}{2},n-1\right)}\left(\frac{s_x}{\sqrt{n}}\right)$ is our **margin of error**
  - **As the confidence level decreases**, t decreases causing the margin of error to decrease, causing the width of the confidence interval to narrow
  - **As the confidence level increases**, t increases causing the margin of error to increase, causing the width of the confidence interval to grow wider

Confidence Intervals Bounds When We Don't Know σ_x

Lower Bound = $\bar{x} - t_{\left(1-\frac{\propto}{2}, n-1\right)}\left(\frac{s_x}{\sqrt{n}}\right)$

Upper Bound = $\bar{x} + t_{\left(1-\frac{\propto}{2}, n-1\right)}\left(\frac{s_x}{\sqrt{n}}\right)$

- We are --% confident that the true population mean, μ_{x}, is between the lower and upper bounds.

Confidence Intervals When We Don't Know σ_x

- t is based on the t distribution which is a lot like the normal distribution but with fatter tails
  - You can find the correct t-value by finding the cross-hair of degrees of freedom, n-1, and the two tailed alpha
  - http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

Finding t for Our Confidence Intervals

- Say we were trying to find the t-value for a 95% confidence with n=10
- This means $\alpha = 1 - .95 = .05$ and the degrees of freedom = 10 - 1 = 9
- $t_{1-\frac{.05}{2},9}$ = 2.262

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** B	**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
A 9	0.000	0.703	0.883	1.100	1.383	1.833	2.262 C	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

Zoom In

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** B	**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
A 9	0.000	0.703	0.883	1.100	1.383	1.833	2.262 C	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

- A is the degrees of freedom, n-1
- B is the significance level – for confidence intervals we look for α in the two-tail row
- C is the t-value associated with the provided degrees of freedom and significance level

Finding t for Our Confidence Intervals

- Say we were trying to find the t-value for a 99% confidence with n=9
- This means $\alpha = 1 - .99 = .01$ and the degrees of freedom = 9 - 1 = 8
- $t_{1-\frac{.01}{2},8}$=3.355

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**	**B 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
A 8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355 C	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

Zoom In

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	B 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
A 8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	C 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

- A is the degrees of freedom, n-1
- B is the significance level – for confidence intervals we look for α in the two-tail row
- C is the t-value associated with the provided degrees of freedom and significance level

Finding t for Our Confidence Intervals

- Say we were trying to find the t-value for a 90% confidence with n=11
- This means $\alpha = 1 - .90 = .10$ and the degrees of freedom = 11 - 1 = 10
- $t_{1-\frac{.10}{2},10} = 1.812$

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	B 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
A 10	0.000	0.700	0.879	1.093	1.372	1.812	C 2.228	2.764	3.169	4.144	4.587

Zoom In

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	B 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
A 10	0.000	0.700	0.879	1.093	1.372	1.812	C 2.228	2.764	3.169	4.144	4.587

- A is the degrees of freedom, n-1
- B is the significance level – for confidence intervals we look for α in the two-tail row
- C is the t-value associated with the provided degrees of freedom and significance level

Confidence Interval Bounds When We Don't Know σ_x

$$\bar{x} \pm t_{1-\frac{\propto}{2},n-1}\left(\frac{s_x}{\sqrt{n}}\right)$$

Lower Bound= $\bar{x} - t_{1-\frac{\propto}{2},n-1}\left(\frac{s_x}{\sqrt{n}}\right)$

Upper Bound= $\bar{x} + t_{1-\frac{\propto}{2},n-1}\left(\frac{s_x}{\sqrt{n}}\right)$

Example

- Suppose a random sample of 38 yearly average temperature measures in New Haven, CT. Among the sampled years the **sample mean temperature was 51.0474** degrees fahrenheit with a **sample standard deviation of 1.3112.**
- Our sample mean = $\bar{x}$ = 51.0474
- Our sample standard deviation = s_x = 1.3112

Example

- Suppose a random sample of 38 yearly average temperature measures in New Haven, CT. Among the sampled years the sample mean temperature was 51.0474 degrees Fahrenheit with a sample standard deviation of 1.3112.
- Check Assumptions
  - n>30 so it is safe to assume the distribution of $\bar{x}$ is bell-shaped
  - The data is from a random sample

Example

- 95% Confidence Interval for population the true population mean yearly average temperature reading in New Haven is:

$$\bar{x} \pm t_{1-\frac{.05}{2},38-1}\left(\frac{s_x}{\sqrt{n}}\right)$$

$$= 51.0474 \pm (2.021)\left(\frac{1.3112}{\sqrt{38}}\right)$$

(50.61752, 51.47728)

Example

(50.61752, 51.47728)

- We are 95% confident that the true population mean yearly average temperature reading in New Haven is between 50.61752 and 51.47728 degrees Fahrenheit

Confidence Intervals Bounds When We Don't Know σ_x - R code

Below is a function you can load into R:

```
t.int<-function(conf.level, xbar, sx, n){
  t.crit = qt(1-(1-conf.level)/2,n-1);
  std.error = sx/sqrt(n);
  MOE=t.crit*std.error;
  c(xbar-MOE, xbar+MOE)
}
```

Confidence Intervals Bounds When We Don't Know σ_x - R code

- You can call the function as below which will provide the 95% confidence interval for a population mean from a sample of 38 that had mean 51.0474 and sample standard deviation 1.3112:

```
conf.level=.95 #Confidence Level
xbar=51.0474 #Sample Mean
sx=1.3112 #Sample Standard Deviation
n=38 #Sample Size
t.int(conf.level, xbar, sx, n)
```

$100(1-\alpha)\%$ Confidence Interval for σ^2

- Recall: $\mathrm{X}_{n-1}^2 = \left(\frac{(n-1)s^2}{\sigma_x^2}\right)$

- If we choose $\chi_{\frac{\alpha}{2}}^2$ such that $P\left(\chi_{n-1}^2 \leq \chi_{\frac{\alpha}{2}}^2\right) = \frac{\alpha}{2}$ and $\chi_{1-\frac{\alpha}{2}}^2$ such that $P\left(\chi_{n-1}^2 \geq \chi_{1-\frac{\alpha}{2}}^2\right) = \frac{\alpha}{2}$ then we have $P\left(\chi_{\frac{\alpha}{2}}^2 \leq \chi_{n-1}^2 \leq \chi_{1-\frac{\alpha}{2}}^2\right) = 1-\alpha$

$100(1-\alpha)\%$ Confidence Interval for σ^2

- Recall: $X_{n-1}^2 = \left(\frac{(n-1)s^2}{\sigma_x^2}\right)$

$$P\left(\chi^2_{\frac{\alpha}{2}} \le \left(\frac{(n-1)s^2}{\sigma_x^2}\right) \le \chi^2_{1-\frac{\alpha}{2}}\right)$$

$$= P\left(\frac{1}{\chi^2_{\frac{\alpha}{2}}} \ge \left(\frac{\sigma_x^2}{(n-1)s^2}\right) \ge \frac{1}{\chi^2_{1-\frac{\alpha}{2}}}\right)$$

$$= P\left(\frac{1}{\chi^2_{1-\frac{\alpha}{2}}} \le \left(\frac{\sigma_x^2}{(n-1)s^2}\right) \le \frac{1}{\chi^2_{\frac{\alpha}{2}}}\right)$$

$$= P\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \le \sigma_x^2 \le \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right)$$

$100(1-\alpha)\%$ Confidence Interval for σ^2

- Assumptions are:
  - The sample is selected from the target population
  - The population of interest has a relative frequency distribution that is approximately normal

$$\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \leq \sigma_x^2 \leq \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}$$

$100(1-\alpha)\%$ Confidence Interval for σ

- We can take the square root of all sides to get a confidence interval for σ

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}} \leq \sigma_x^2 \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}}$$

$100(1-\alpha)\%$ Confidence Interval for σ - R code

Below is a function you can load into R:

```
var.int<-function(conf.level, sx, n){
  chisq.critL = qchisq(1-(1-conf.level)/2,n-1);
  chisq.critU = qchisq((1-conf.level)/2,n-1);
  lower=(n-1)*(sx^2)/chisq.critL
  upper=(n-1)*(sx^2)/chisq.critU
  c(lower,upper)
}
```

$100(1-\alpha)\%$ Confidence Interval for σ - R code

- You can call the function below which will provide the 95% confidence interval for a population variance from the New Haven temperature data from sample of 38 that had mean 51.0474 and sample standard deviation 1.3112:

```
conf.level=.95 #Confidence Level
sx=1.3112 #Sample Standard Deviation
n=38 #Sample Size
var.int(conf.level, sx, n)
```

Answer: (1.142705, 2.877642)

We are 95% confident that the true population variance is between 1.142705 and 2.877642)

$100(1-\alpha)\%$ Confidence Interval for σ - R code

Below is a function you can load into R:

```
sd.int<-function(conf.level, sx, n){
  chisq.critL = qchisq(1-(1-conf.level)/2,n-1);
  chisq.critU = qchisq((1-conf.level)/2,n-1);
  lower=sqrt((n-1)*(sx^2)/chisq.critL)
  upper=sqrt((n-1)*(sx^2)/chisq.critU)
  c(lower,upper)
}
```

$100(1-\alpha)\%$ Confidence Interval for σ - R code

- You can call the function below which will provide the 95% confidence interval for a population standard deviation from the New Haven temperature data from sample of 38 that had mean 51.0474 and sample standard deviation 1.3112:

```
conf.level=.95 #Confidence Level
sx=1.3112 #Sample Standard Deviation
n=38 #Sample Size
sd.int(conf.level, sx, n)
```

Answer: (1.068974, 1.696361)

We are 95% confident that the true population variance is between 1.068974 and 1.696361

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\ \sigma_y^2$

- Recall: $F_{n_x-1,n_y-1} = \dfrac{\left(\frac{s_x^2}{s_y^2}\right)}{\left(\frac{\sigma_x^2}{\sigma_y^2}\right)}$

- If we choose $F_{\frac{\alpha}{2}}$ such that $P\left(F_{n_x-1,n_y-1} \leq F_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ and $F^2_{1-\frac{\alpha}{2}}$ such that $P\left(F_{n_x-1,n_y-1} \geq F_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ then we have $P\left(F_{\frac{\alpha}{2}} \leq F_{n_x-1,n_y-1} \leq F_{1-\frac{\alpha}{2}}\right) = 1-\alpha$

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\,\sigma_y^2$

- Recall: $F_{n_x-1,n_y-1} = \dfrac{\left(\frac{s_x^2}{s_y^2}\right)}{\left(\frac{\sigma_x^2}{\sigma_y^2}\right)}$

$$P\left(F_{\frac{\alpha}{2}} \le \frac{\left(\frac{s_x^2}{s_y^2}\right)}{\left(\frac{\sigma_x^2}{\sigma_y^2}\right)} \le F_{1-\frac{\alpha}{2}}\right) = P\left(\frac{1}{F_{\frac{\alpha}{2}}} \ge \frac{\left(\frac{\sigma_x^2}{\sigma_y^2}\right)}{\left(\frac{s_x^2}{s_y^2}\right)} \ge \frac{1}{F_{1-\frac{\alpha}{2}}}\right)$$

$$= P\left(\frac{1}{F_{1-\frac{\alpha}{2}}} \le \frac{\left(\frac{\sigma_x^2}{\sigma_y^2}\right)}{\left(\frac{s_x^2}{s_y^2}\right)} \le \frac{1}{F_{\frac{\alpha}{2}}}\right) = \left(\frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{1-\frac{\alpha}{2}}} \le \left(\frac{\sigma_x^2}{\sigma_y^2}\right) \le \frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{\frac{\alpha}{2}}}\right) = 1-\alpha$$

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\,\sigma_y^2$

- We can take the square root of all sides to get a confidence interval for $\sigma_x^2/\,\sigma_y^2$

$$\frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{1-\frac{\alpha}{2}}} \leq \left(\frac{\sigma_x^2}{\sigma_y^2}\right) \leq \frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{\frac{\alpha}{2}}}$$

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\ \sigma_y^2$

- Iterpreting the confidence interval $\left(\frac{\sigma_x^2}{\sigma_y^2}\right)$
  - If all values of the interval are bigger than one: $\sigma_x^2 > \sigma_y^2$
  - If all values of the interval are less than one: $\sigma_x^2 < \sigma_y^2$
  - If the interval contains one it is possible that $\sigma_x^2 = \sigma_y^2$

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\ \sigma_y^2$- R code

Below is a function you can load into R:

```
F.int<-function(conf.level, sx, nx, sy, ny){
  sratio = sx^2/sy^2
  F.critL = qf(1-(1-conf.level)/2,nx-1,ny-1);
  F.critU = qf((1-conf.level)/2,nx-1,ny-1);
  lower=sratio/F.critL
  upper=sratio/F.critU
  c(lower,upper)
}
```

$100(1-\alpha)\%$ Confidence Interval for σ_x^2 / σ_y^2 - R code

- You can call the function below which will provide the 95% confidence interval for the ratios of the population variances from two groups. Say we have a sample, X, of 32 that had sample standard deviation 1.45 and a sample, Y, of 38 that had sample standard deviation 1.57:

```
conf.level=.95 #Confidence Level
sx=1.45 #Sample Standard Deviation
nx=32
sy=1.57 #Sample Standard Deviation
ny=38
F.int(conf.level, sx, nx, sy, ny)
```

Answer: (.4338582,1.7125010) we are 95% confident that the ratio of the population variances is between .4338582 and 1.7125010; 1 is on the confidence interval, so it is possible that the variances are equal.

Summaries

Confidence Intervals

Assumptions	Point Estimate	Margin of Error	Margin of Error
1. Random Sample *2.* $n\hat{p} \geq 15$ And $n(1-\hat{p}) \geq 15$	$\hat{p}$	$z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

- We are --% confident that the true population proportion lays on the confidence interval.

Example in R

Below is a function you can load into R:

```
prop.int<-function(conf.level, x, n, Wilson=FALSE){
  if(Wilson){
    phat=(x+2)/(n+4)
  }else{
    phat=x/n
  }
  z.crit = qnorm(1-(1-conf.level)/2);
  std.error = sqrt(phat*(1-phat)/n);
  MOE=z.crit*std.error;
  c(phat-MOE, phat+MOE)
}
```

Confidence Intervals known σ_x

Assumptions	Point Estimate	Margin of Error	Margin of Error
1. Random Sample *2.* $n > 30$ OR the population is bell shaped	$\bar{x}$	$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$	$\bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)$

- We are --% confident that the true population mean lays on the confidence interval.

Confidence Intervals Bounds When We Know σ_x - R code

Below is a function you can load into R:

```
z.int<-function(conf.level, xbar, sigma, n){
  z.crit = qnorm(1-(1-conf.level)/2);
  std.error = sigma/sqrt(n);
  MOE=z.crit*std.error;
  c(xbar-MOE, xbar+MOE)
}
```

Confidence Intervals unknown σ_x

Assumptions	Point Estimate	Margin of Error	Margin of Error
1. Random Sample *2.* $n > 30$ OR the population is bell shaped	$\bar{x}$	$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{n}}$	$\bar{x} \pm t_{1-\frac{\alpha}{2}, n-1}\left(\frac{s_x}{\sqrt{n}}\right)$

- We are --% confident that the true population mean lays on the confidence interval.

Confidence Intervals Bounds When We Don't Know σ_x - R code

Below is a function you can load into R:

```
t.int<-function(conf.level, xbar, sx, n){
  t.crit = qt(1-(1-conf.level)/2,n-1);
  std.error = sx/sqrt(n);
  MOE=t.crit*std.error;
  c(xbar-MOE, xbar+MOE)
}
```

Confidence Intervals unknown σ_x

Assumptions	Margin of Error
1. Random Sample *2. Data follows the* Normal Distribution	$\frac{\left((n-1)s_x^2\right)}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\left((n-1)s_x^2\right)}{\chi^2_{1-\frac{\alpha}{2}}}$

- We are --% confident that the true population variance lays on the confidence interval.

$100(1-\alpha)\%$ Confidence Interval for σ - R code

Below is a function you can load into R:

```
var.int<-function(conf.level, sx, n){
  chisq.critL = qchisq(1-(1-conf.level)/2,n-1);
  chisq.critU = qchisq((1-conf.level)/2,n-1);
  lower=(n-1)*(sx^2)/chisq.critL
  upper=(n-1)*(sx^2)/chisq.critU
  c(lower,upper)
}
```

Confidence Intervals unknown σ_x

Assumptions	Margin of Error
1. Random Sample *2. Data follows the* Normal Distribution	$\sqrt{\dfrac{\left((n-1)s_x^2\right)}{\chi^2_{\frac{\alpha}{2}}}} \leq \sigma \leq \sqrt{\dfrac{\left((n-1)s_x^2\right)}{\chi^2_{1-\frac{\alpha}{2}}}}$

- We are --% confident that the true population standard deviation lays on the confidence interval.

$100(1-\alpha)\%$ Confidence Interval for σ - R code

Below is a function you can load into R:

```
sd.int<-function(conf.level, sx, n){
  chisq.critL = qchisq(1-(1-conf.level)/2,n-1);
  chisq.critU = qchisq((1-conf.level)/2,n-1);
  lower=sqrt((n-1)*(sx^2)/chisq.critL)
  upper=sqrt((n-1)*(sx^2)/chisq.critU)
  c(lower,upper)
}
```

Confidence Intervals unknown σ_x

Assumptions	Margin of Error
1. Random Sample *2. Data follows the* Normal Distribution	$$\frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{1-\frac{\alpha}{2}}} \leq \left(\frac{\sigma_x^2}{\sigma_y^2}\right) \leq \frac{\left(\frac{s_x^2}{s_y^2}\right)}{F_{\frac{\alpha}{2}}}$$

- We are --% confident that the true ratio of population variances lays on the confidence interval.

$100(1-\alpha)\%$ Confidence Interval for $\sigma_x^2/\ \sigma_y^2$

- Iterpreting the confidence interval $\left(\frac{\sigma_x^2}{\sigma_y^2}\right)$
  - If all values of the interval are bigger than one: $\sigma_x^2 > \sigma_y^2$
  - If all values of the interval are less than one: $\sigma_x^2 < \sigma_y^2$
  - If the interval contains one it is possible that $\sigma_x^2 = \sigma_y^2$

$100(1-\alpha)\%$ Confidence Interval for σ_x^2/σ_y^2 - R code

Below is a function you can load into R:

```
F.int<-function(conf.level, sx, nx, sy, ny){
  sratio = sx^2/sy^2
  F.critL = qf(1-(1-conf.level)/2,nx-1,ny-1);
  F.critU = qf((1-conf.level)/2,nx-1,ny-1);
  lower=sratio/F.critL
  upper=sratio/F.critU
  c(lower,upper)
}
```