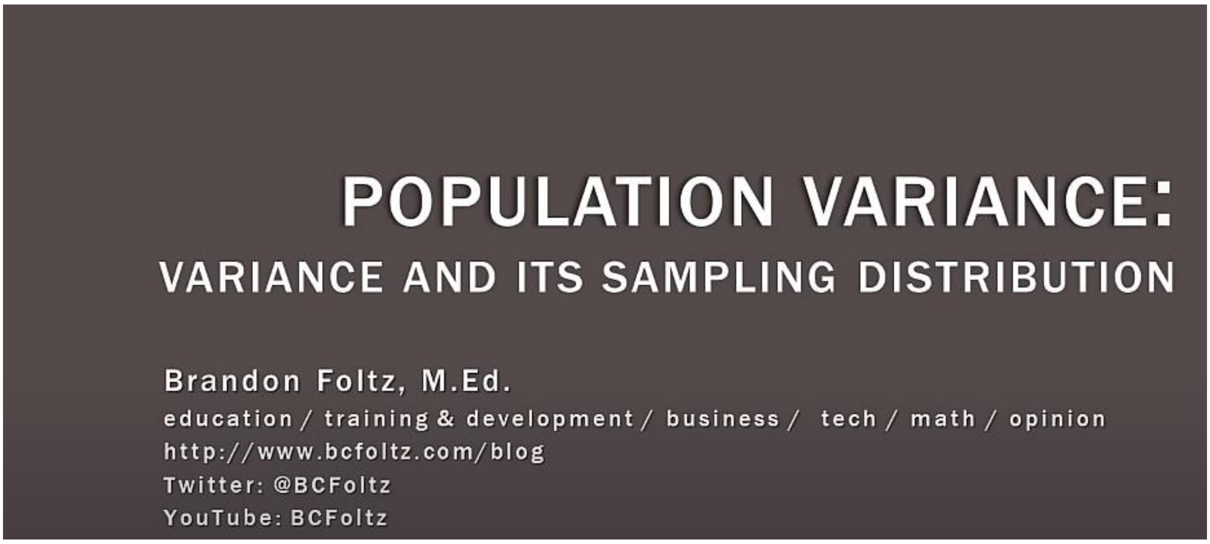
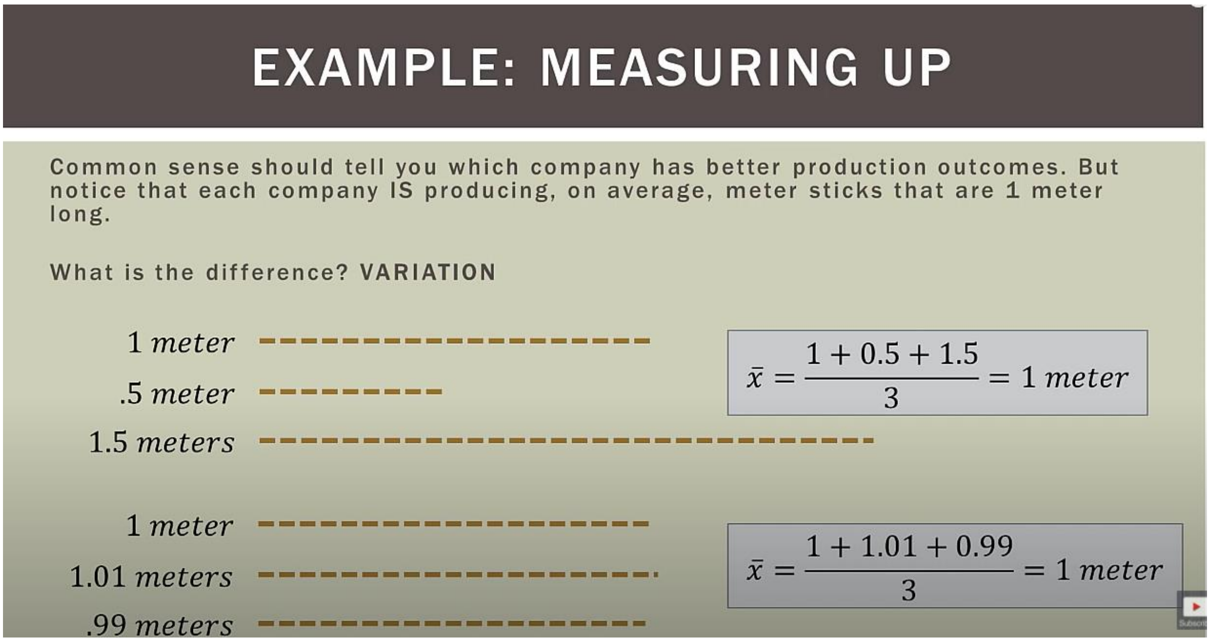
Video 1:

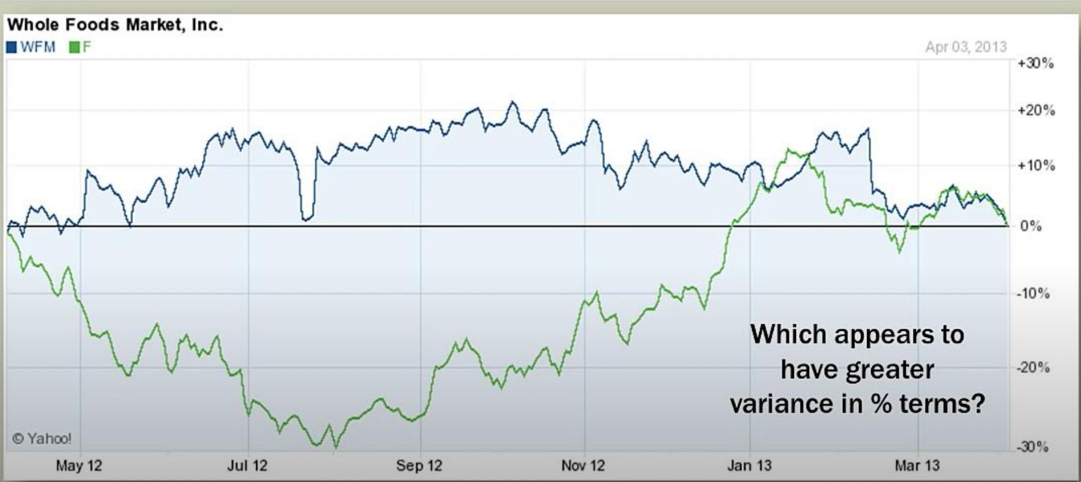

POPULATION VARIANCE:
VARIANCE AND ITS SAMPLING DISTRIBUTION
Brandon Foltz, M.Ed.
education / training & development / business / tech / math / opinion
http://www.bcfoltz.com/blog
Twitter: @BCFoltz
YouTube: BCFoltz


EXAMPLE: MEASURING UP
Common sense should tell you which company has better production outcomes. But notice that each company IS producing, on average, meter sticks that are 1 meter long.
What is the difference? VARIATION
$1\ meter$
$.5\ meter$
$1.5\ meters$
$\bar{x} = \frac{1 + 0.5 + 1.5}{3} = 1\ meter$
$1\ meter$
$1.01\ meters$
$.99\ meters$
$\bar{x} = \frac{1 + 1.01 + 0.99}{3} = 1\ meter$

EXAMPLE: ENGINE CYLINDERS

When a standard car, truck, or similar engine is made the cylinders must be "bored" from a block of metal. The pistons must fit inside the cylinder VERY precisely. So yes the cylinders must be the correct diameter…but they ALSO must have a VARIANCE near zero.

EXAMPLE: STOCK RETURNS AS %

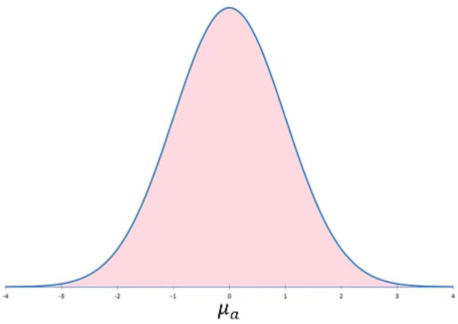


Ford Motor Company vs. Whole Foods Markets

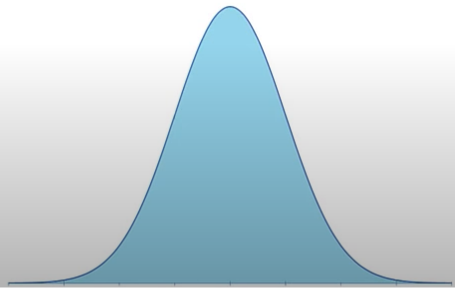
GENERAL CONSIDERATIONS

- Making meter sticks that are on average 1 meter in length is not good enough; they actually have to be 1 meter!
- As you can tell, analyzing variance is a cornerstone of statistical quality control. Ever heard of Six Sigma?
- Higher variation could mean inconsistent production or out of control processes.
- An extremely low variance could mean a stuck sensor or human error. The variance of 3,3,3,3,3,3,3,3,3…is ZERO!
- Data sets can have the same mean but very different variances.

When we are looking to differentiate or compare two distributions, there are two primary characteristics we can analyze.



1) Their means
2) Their variances



VARIANCE QUICK REVIEW

Variance and its square-root, the standard deviation, are both measures of the spread or variability in data. Two or more data sets could have the same mean, but very different variances.

Population Variance

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

Sample Variance

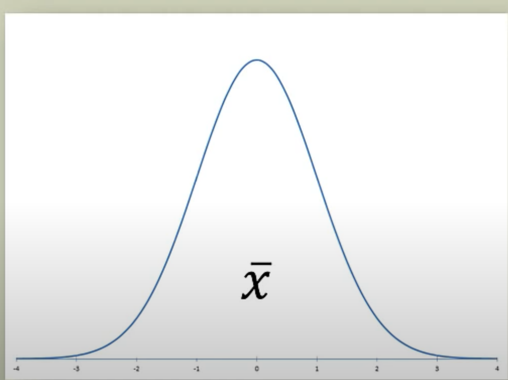
$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

These are based on the squared difference between each observation and the mean. *The average squared deviation.* How far is each data point from the center (mean) on average?

SAMPLING DISTRIBUTION OF $\bar{x}$

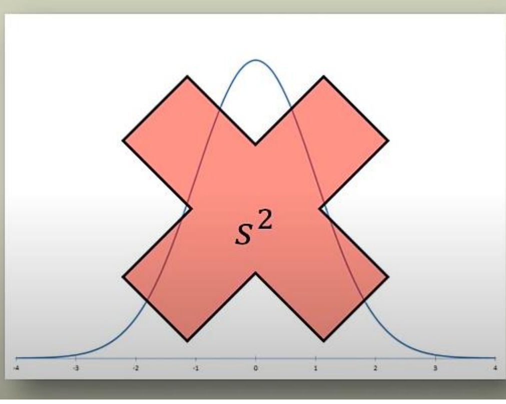
When we take many samples of the same size from a population and then find the sample means, $\bar{x}$, those sample means follow the normal curve when placed in their own distribution.

The Sampling Distribution of $\bar{x}$



SAMPLING DISTRIBUTION OF s^2

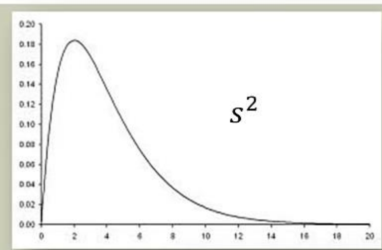
However when we take many samples of the same size from a normal population and then find those sample *variances*, s^2, those sample variances DO NOT follow the normal curve when placed in their own distribution.

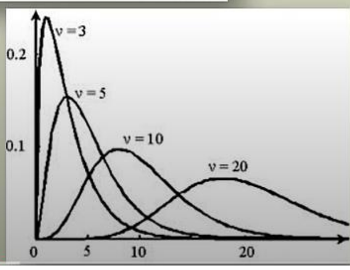


The Sampling Distribution of s^2

SAMPLING DISTRIBUTION OF s^2

When we take many samples of the same size from a normal population and then find those sample variances, s^2, those sample variances DO NOT follow the normal curve when placed in their own distribution.

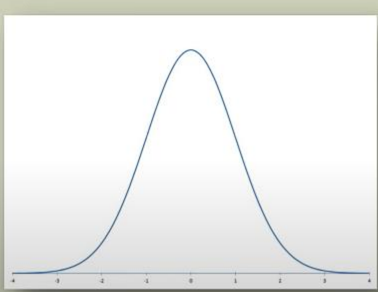




They follow the **chi-square distribution, χ^2,** with $n - 1$ degrees of freedom.

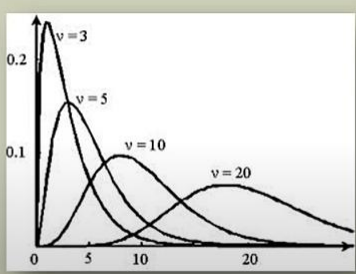
SAMPLING DISTRIBUTIONS, $\bar{x}$ AND s^2

Sampling Distribution of $\bar{x}$



Normal distribution

Sampling Distribution of s^2

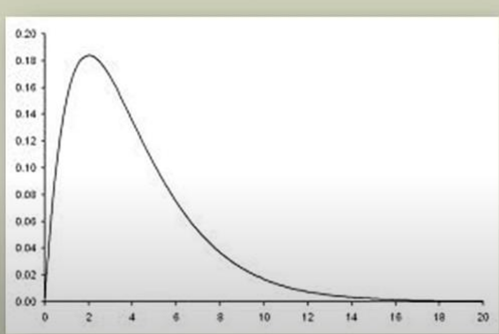


Chi-square distribution, χ^2

$$df = v = n - 1$$

ANATOMY OF THE χ^2 DISTRIBUTION

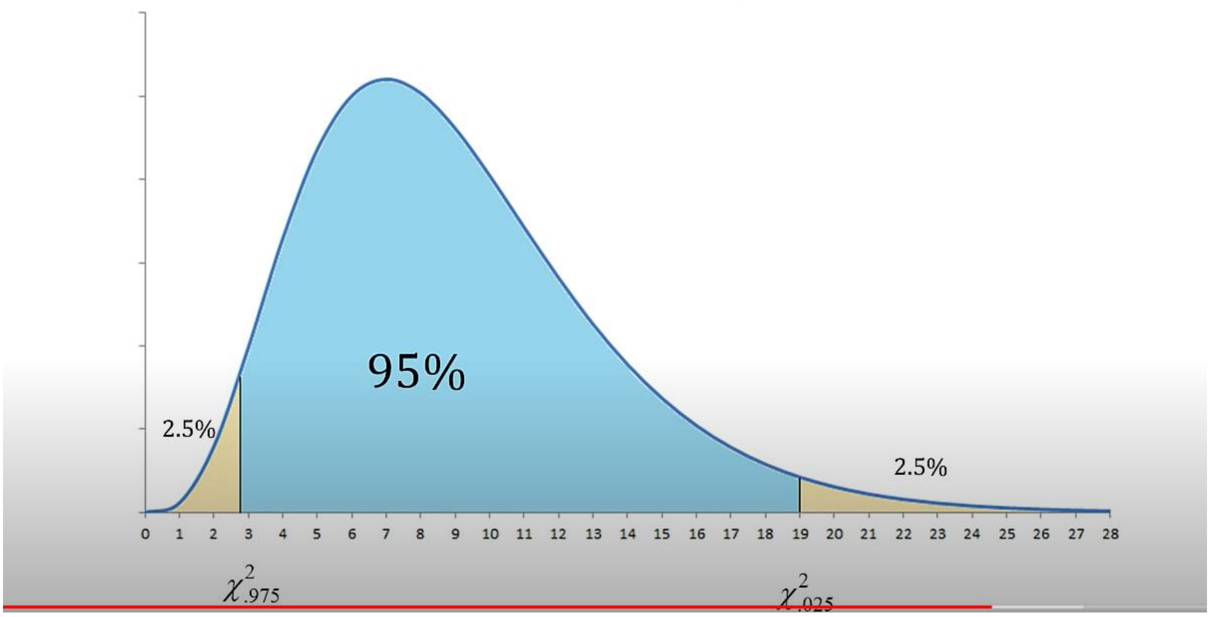
1. There is no "one" chi-square distribution. There is a chi-square distribution for each $n - 1$ degrees of freedom (like the t-distribution).
2. Area (probability) under the curve is 1
3. The curve is asymptotic; it never touches the x-axis.
4. In the chi-square, "1" is at the left side and "0" is on the right side.
5. Therefore the cumulative probability runs right to left.
6. The probabilities are found in the chi-square table in the same manner as normal curves.

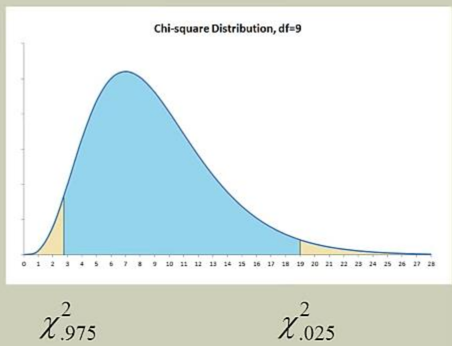


1 ← "0"

Cumulative Probability

Chi-square Distribution, df=9





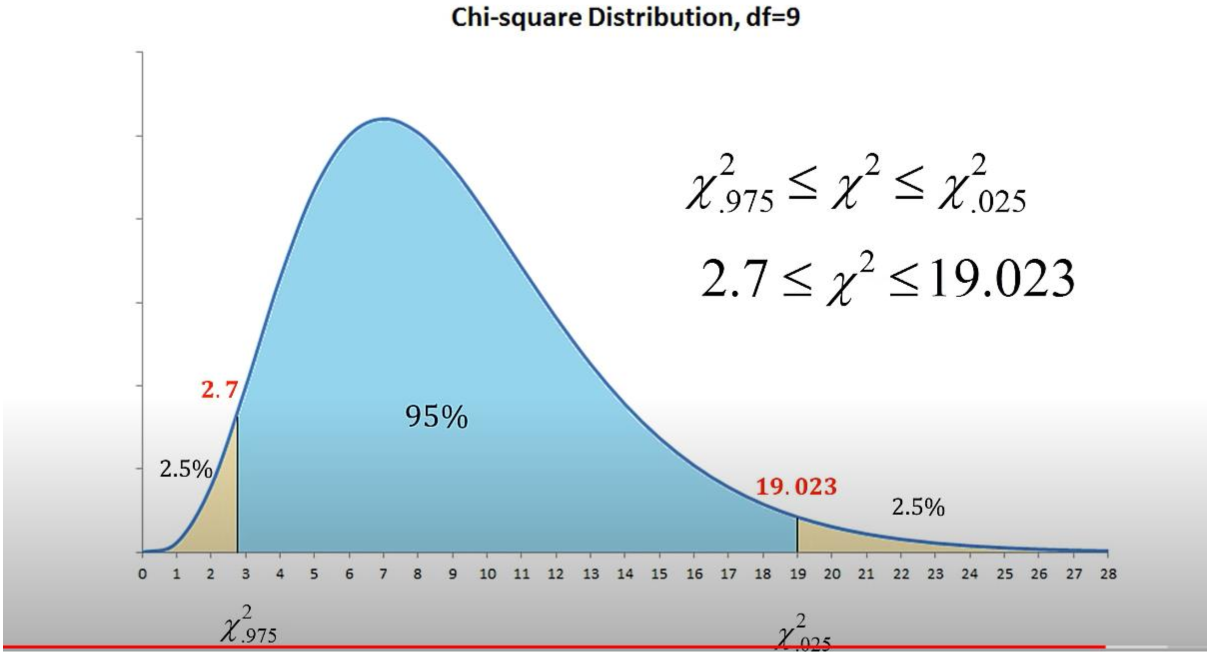
Reading the Chi-square Table

$df = 9$

TABLE IV

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value 0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	$\boxed{2.700}$	3.325	4.168	14.684	16.919	$\boxed{19.023}$	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578



QUICK REVIEW

- Analyzing variance is very important in fields such as statistical quality control and in Six Sigma environments.
- "Equal" variance is also an important assumption in advanced statistical techniques so we need a way to test for it.
- A distribution of sample means follows the normal distribution; a distribution of sample variances follows the chi-square distribution, χ^2 with $n-1$ degrees of freedom.
- The chi-square distribution has a total probability of 1, but the cumulative probability runs right to left, not left to right
- The critical values are found using the chi-square table in the same manner as we do for the normal distribution

Video 2:

POPULATION VARIANCE:

INTERVAL ESTIMATION OF THE VARIANCE

Brandon Foltz, M.Ed.
education / training & development / business / tech / math / opinion
http://www.bcfoltz.com/blog
Twitter: @BCFoltz
YouTube: BCFoltz

EXAMPLE: STOCK RETURNS

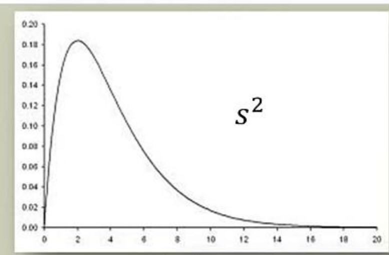
When investing in stocks, there is usually a trade off: risk vs. reward. In financial terms, “risk” is often another word for a stock’s *variance*. Some stocks are steady (low risk) but offer lower potential returns (GE). Others swing wildly (higher risk) but offer more potential upside (AAPL).

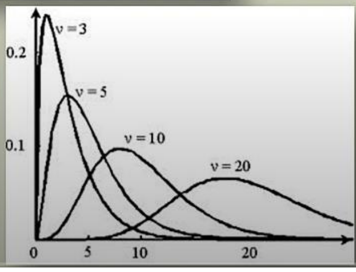
So let’s say we purchased a share of each stock, GE and AAPL, on the first trading morning of the year. We then held each stock all the way through 2012. Think of each stock as a ride in an airplane. Which “flight” was more likely to make you sick to your stomach due to TURBULENCE?

SAMPLING DISTRIBUTION OF s^2

When we take many samples of the same size from a normal population and then find those sample variances, s^2, those sample variances DO NOT follow the normal curve when placed in their own distribution.

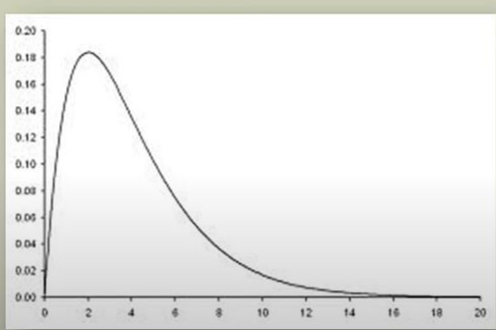
They follow the **chi-square distribution**, χ^2, with $n-1$ degrees of freedom.





ANATOMY OF THE χ^2 DISTRIBUTION

1. There is no "one" chi-square distribution. There is a chi-square distribution for each $n-1$ degrees of freedom (like the t-distribution).
2. Area (probability) under the curve is 1
3. The curve is asymptotic; it never touches the x-axis.
4. In the chi-square, "1" is at the left side and "0" is on the right side.
5. Therefore the cumulative probability runs right to left.
6. The probabilities are found in the chi-square table in the same manner as normal curves.



1 "0"

Cumulative Probability

GE VS. AAPL 2012 MONTHLY RETURNS

Date	GE	APPL	GE%	AAPL%
January 2012	0.043869	0.119655	4.39%	11.97%
February 2012	0.026947	0.172533	2.69%	17.25%
March 2012	0.051809	0.100103	5.18%	10.01%
April 2012	-0.02451	-0.026306	-2.45%	-2.63%
May 2012	-0.02567	-0.010762	-2.57%	-1.08%
June 2012	0.096491	0.010796	9.65%	1.08%
July 2012	-0.00444	0.044801	-0.44%	4.48%
August 2012	-0.00198	0.089724	-0.20%	8.97%
September 2012	0.099801	0.002791	9.98%	0.28%
October 2012	-0.07535	-0.113841	-7.53%	-11.38%
November 2012	0.003376	-0.012450	0.34%	-1.25%
December 2012	0.002404	-0.095123	0.24%	-9.51%
Mean	0.016063	0.023493	1.61%	2.35%
Variance	0.002590	0.007330	25.89	73.30
Standard Dev.	0.050890	0.085618	5.09%	8.56%

Mean Monthly Return

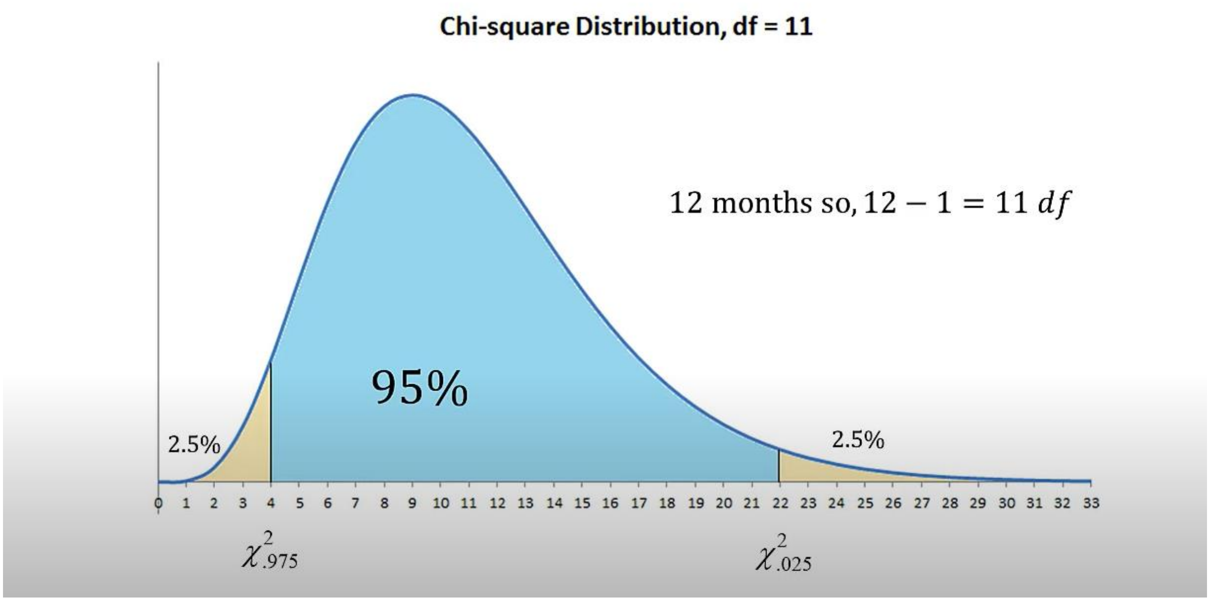
$\text{GE} = 1.61\%$ $\qquad$ $\text{AAPL} = 2.35\%$

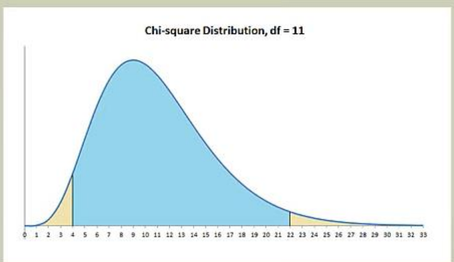
Monthly Return Variance, s^2

$\text{GE} = 25.89$ $\qquad$ $\text{AAPL} = 73.30$

Monthly Return Standard Deviation, s

$\text{GE} = 5.09\%$ $\qquad$ $\text{AAPL} = 8.56\%$





$\chi^2_{.975}$ $\chi^2_{.025}$

Reading the Chi-square Table

TABLE IV

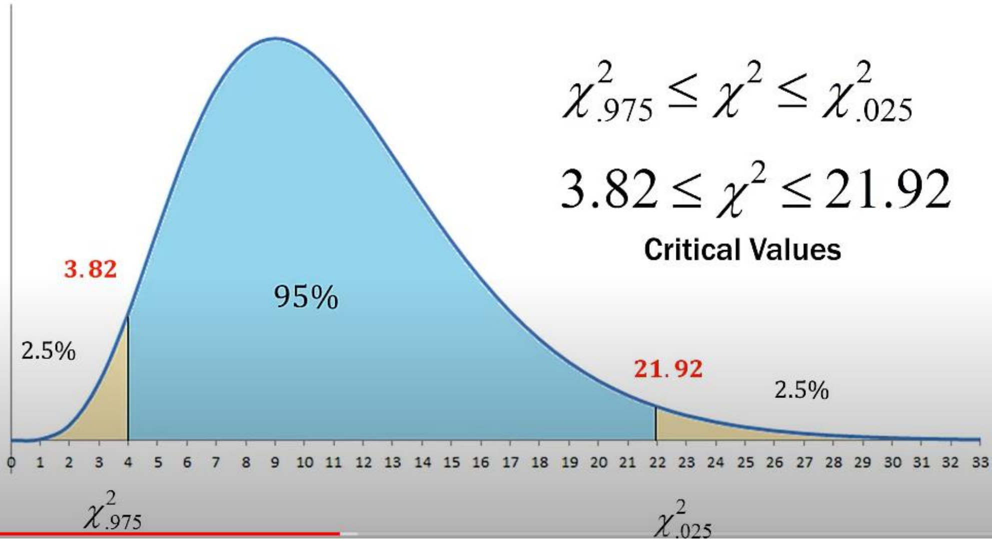
Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

$df = 11$

Degrees of Freedom	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	$\boxed{3.816}$	4.575	5.578	17.275	19.675	$\boxed{21.920}$	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578

Chi-square Distribution, df = 11



$$\chi^2_{.975} \le \chi^2 \le \chi^2_{.025}$$

$$3.82 \le \chi^2 \le 21.92$$

Critical Values

SAMPLING DISTRIBUTION

We know that the sampling distribution of variance follows the chi-square distribution, χ^2. But how does that relate to the actual data we collect?

Whenever a random sample of size n is selected from a normal population, the sampling distribution of:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

n = sample size
s^2 = sample variance
σ^2 = population variance (unknown)

…has a chi-square distribution with $n-1$ degrees of freedom.

INTERVAL ESTIMATE OF THE VARIANCE

$$\chi^2_{.975} \le \chi^2 \le \chi^2_{.025}$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

substitute

$$\chi^2_{.975} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{.025}$$

$$\sigma^2 \chi^2_{.975} \le (n-1)s^2 \le \sigma^2 \chi^2_{.025}$$

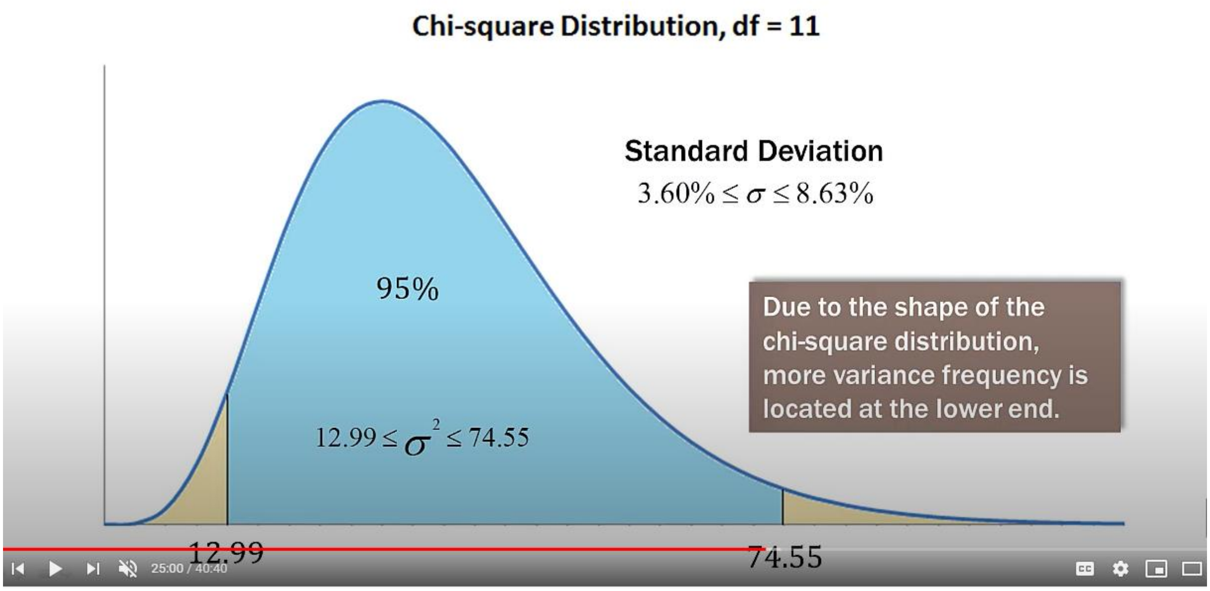
split

$$\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{.975}} \qquad \frac{(n-1)s^2}{\chi^2_{.025}} \le \sigma^2$$

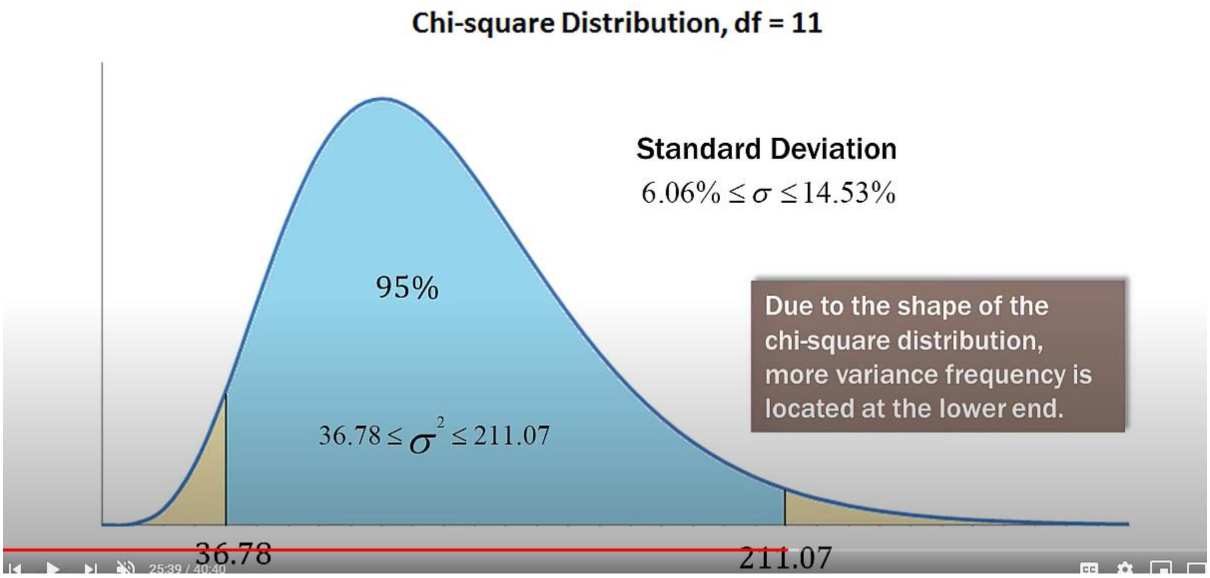
recombine

$$\frac{(n-1)s^2}{\chi^2_{.025}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{.975}}$$

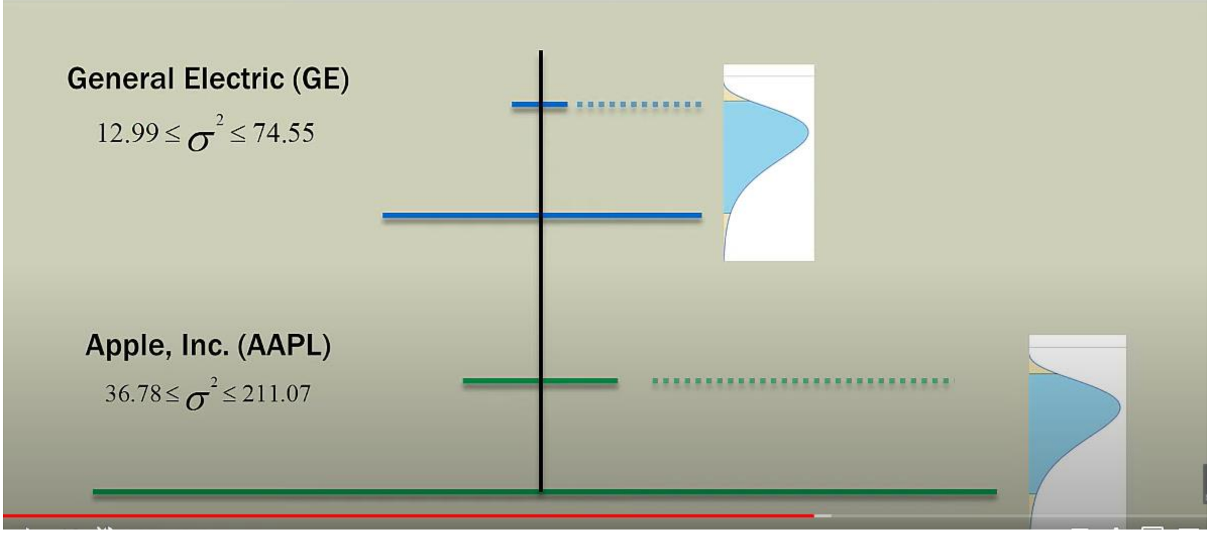
GE 2012 MONTHLY RETURN VARIANCE



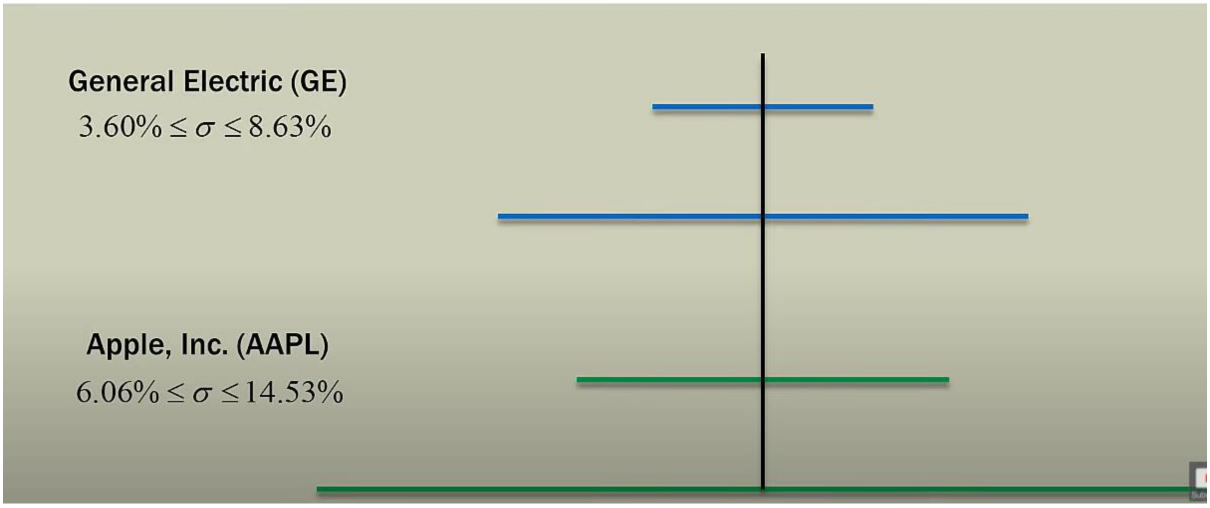
AAPL 2012 MONTHLY RETURN VARIANCE



VISUALIZING RELATIVE VARIANCE



VISUALIZING RELATIVE STDEV “WAISTLINES”



INTERVAL ESTIMATE FOR s

Monthly Return Variance, s^2

$\text{GE} = 25.89 \qquad \text{AAPL} = 73.30$

$$\frac{(12-1)25.89}{21.92} \le \sigma^2 \le \frac{(12-1)25.89}{3.82}$$

$$12.99 \le \sigma^2 \le 74.55$$

$$3.60\% \le \sigma \le 8.63\%$$

$$\frac{(12-1)73.30}{21.92} \le \sigma^2 \le \frac{(12-1)73.30}{3.82}$$

$$36.78 \le \sigma^2 \le 211.07$$

$$6.06\% \le \sigma \le 14.53\%$$

Monthly Return Sample Standard Deviation, s

$\text{GE} = 5.09\% \qquad \text{AAPL} = 8.56\%$

INTERVAL ESTIMATE FOR s

Monthly Return Variance, s^2

$\text{GE} = 25.89 \qquad \text{AAPL} = 73.30$

$$\frac{(12-1)25.89}{21.92} \le \sigma^2 \le \frac{(12-1)25.89}{3.82}$$

$$12.99 \le \sigma^2 \le 74.55$$

$$3.60\% \le \sigma \le 8.63\%$$

$$\frac{(12-1)73.30}{21.92} \le \sigma^2 \le \frac{(12-1)73.30}{3.82}$$

$$36.78 \le \sigma^2 \le 211.07$$

$$6.06\% \le \sigma \le 14.53\%$$

So what does this mean in real life? GE had an average return of 1.61% and AAPL 2.35%. But AAPL had a much higher variance / standard deviation. This, volatility, is one way of interpreting RISK. The classic trade-off of risk vs. reward.

THE EFFECT OF SAMPLE SIZE ON VARIANCE/STDEV INTERVAL ESTIMATE

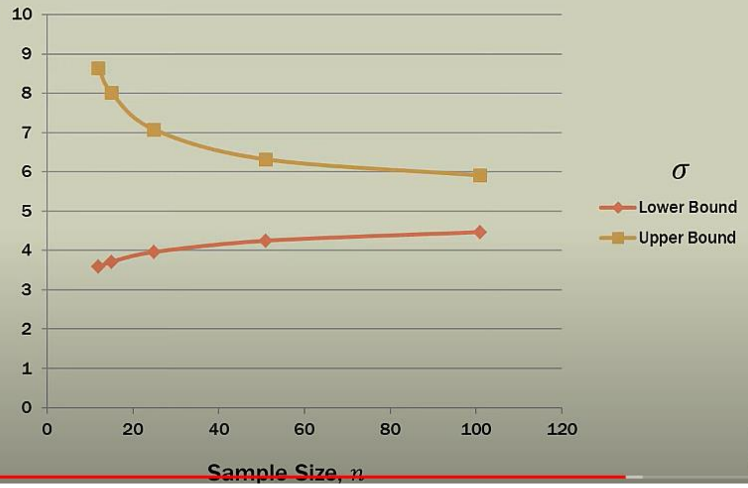
$n = 12$ GE $= s^2 = 25.89$ $12.99 \le \sigma^2 \le 74.55$ $3.60\% \le \sigma \le 8.63\%$

$n = 15$ $$\frac{(15-1)25.89}{26.12} \le \sigma^2 \le \frac{(15-1)25.89}{5.63}$$ $$13.9 \le \sigma^2 \le 64.38$$ $$3.72\% \le \sigma \le 8.02\%$$	$n = 51$ $$\frac{(51-1)25.89}{71.42} \le \sigma^2 \le \frac{(51-1)25.89}{32.36}$$ $$18.1 \le \sigma^2 \le 40.00$$ $$4.25\% \le \sigma \le 6.32\%$$
$n = 25$ $$\frac{(25-1)25.89}{39.36} \le \sigma^2 \le \frac{(25-1)25.89}{12.40}$$ $$15.79 \le \sigma^2 \le 50.11$$ $$3.97\% \le \sigma \le 7.08\%$$	$n = 101$ $$\frac{(101-1)25.89}{129.6} \le \sigma^2 \le \frac{(101-1)25.89}{74.22}$$ $$19.98 \le \sigma^2 \le 34.88$$ $$4.47\% \le \sigma \le 5.91\%$$

THE EFFECT OF SAMPLE SIZE ON VARIANCE/STDEV INTERVAL ESTIMATE

All else remaining constant, increasing the sample size will narrow the interval estimate for the variance and standard deviation.

This graph shows the upper and lower boundary of the standard deviation intervals for different sample sizes assuming s^2 is constant (which it will not be in practice).



INTERVAL ESTIMATE TAKEAWAYS

What if $s^2 = 0?$ $\longrightarrow$ $0 \leq \sigma^2 \leq 0$

A few important takeaways, caveats, and warnings:

1. Interval estimates for variance are VERY sensitive to the overall population being normally distributed; check your data first.
2. Strange things happen at very small sample sizes/degrees of freedom in the chi-square distribution; try to keep sample sizes $n \geq 10$, ideally even larger.
3. The interval estimate for the VARIANCE follows the chi-square distribution. The interval estimate for the STANDARD DEVIATION does not. The σ interval is an "extra" step.
4. As the sample size increases…
    1. the interval estimate narrows (all else remaining constant)
    2. distribution of the variances becomes more and more like the normal distribution; chi-square with a large n.
5. As compared to the 95% C.I., a 99% C.I. will widen the interval, and 90% C.I. will narrow it. Think of them as the size of the "net" needed to catch variances.
6. We are estimating the SPREAD of the distribution. I like to think of it as estimating the "waistline" of the population distribution.

Video 3:

POPULATION VARIANCE:

HYPOTHESIS TESTS FOR THE VARIANCE

Brandon Foltz, M.Ed.

education / training & development / business / tech / math / opinion

http://www.bcfoltz.com/blog

Twitter: @BCFoltz

YouTube: BCFoltz

EXAMPLE: HIGH INTEREST REALTY

High Interest Realty has set a timed sales goal whereby the standard deviation of listing-to-sell time is no longer than 21 days.

Since the standard deviation of 21 is the square root of the variance, $21^2 = 441$ is the variance upper limit. Realizing this goal will help with sales forecasting and marketing.

To conduct this analysis, the sales records for 15 previously sold homes were randomly selected. Based on this sample the following information was obtained:

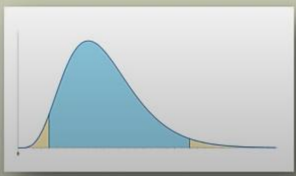
$n = 15$ $\quad$ $\bar{x} = 162\ days$ $\quad$ $s^2 = 582$ $\quad$ $s = 24\ days$

HYPOTHESIS TESTING FOR VARIANCE

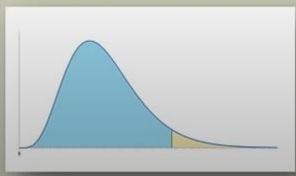
Single sample hypothesis tests for variance follow the same patterns as those for means:

$$H_0: \sigma^2 = \sigma_0^2 \qquad H_0: \sigma^2 \leq \sigma_0^2 \qquad H_0: \sigma^2 \geq \sigma_0^2$$

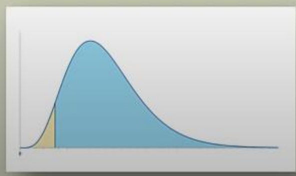
$$H_a: \sigma^2 \neq \sigma_0^2 \qquad H_a: \sigma^2 > \sigma_0^2 \qquad H_a: \sigma^2 < \sigma_0^2$$



Two-tailed

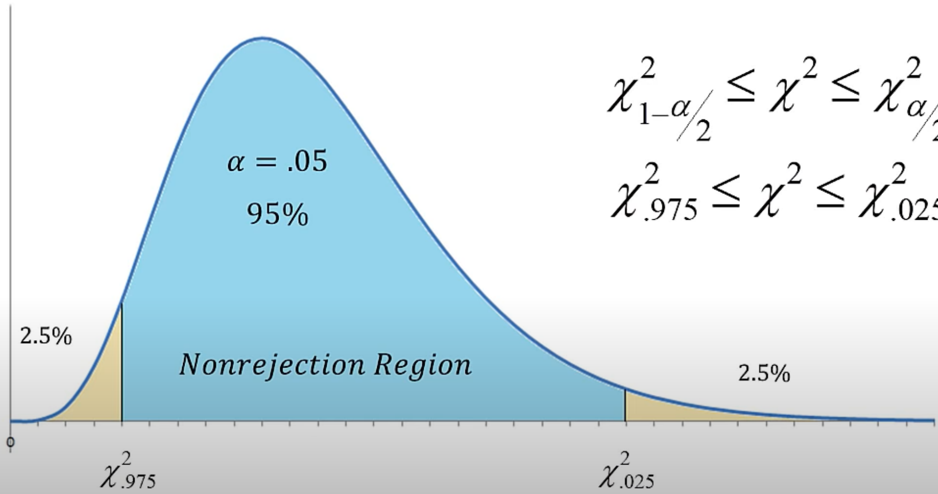


Upper/right-tailed



Lower/left-tailed

Two-tailed Hypothesis Test for σ^2

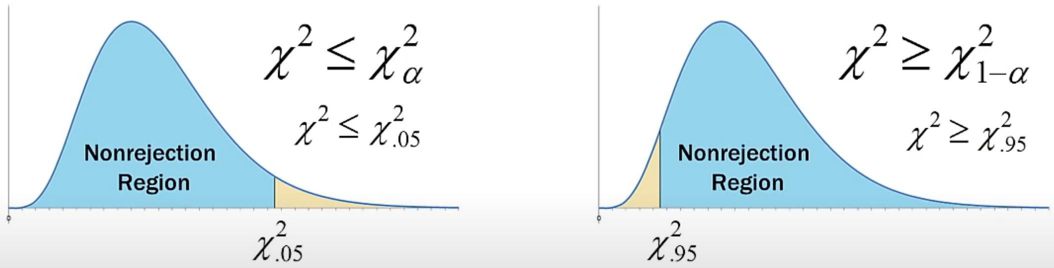


$$\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}$$

$$\chi^2_{.975} \leq \chi^2 \leq \chi^2_{.025}$$

The critical values are found in the chi-square table for the proper degrees of freedom.

One-tailed Hypothesis Test for σ^2



The critical values are found in the chi-square table for the proper degrees of freedom.

VARIANCE TEST STATISTIC

We know that the sampling distribution of variance follows the chi-square distribution, χ^2. But how does that relate to the actual data we collect?

Whenever a random sample of size n is selected from a normal population, the sampling distribution of:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

n = sample size
s^2 = sample variance
σ^2 = population variance (unknown)

…has a chi-square distribution with $n-1$ degrees of freedom.

HIGH INTEREST REALITY

Step 1: Establish Hypothesis

$$H_0 : \sigma^2 \leq 441 \qquad H_a : \sigma^2 > 441$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a one-tailed test. Since the concern is a larger than desired variance, it will be an upper/right tailed test.

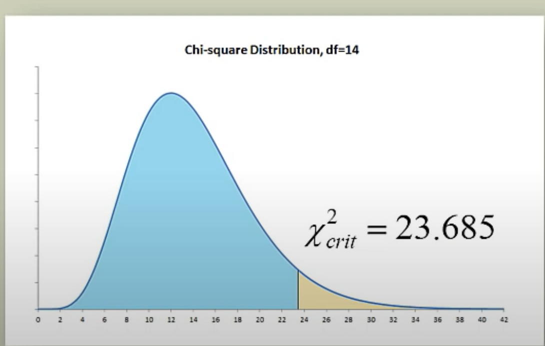
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

We will use the chi-square distribution with $15 - 1 = 14\ df$.

HIGH INTEREST REALITY

Step 3: Specify the significance level

$$\alpha = .05$$



Step 4: State the decision rule

If $\chi^2 > 23.685$, reject H_0 per the chi-square table

Step 5: Gather data

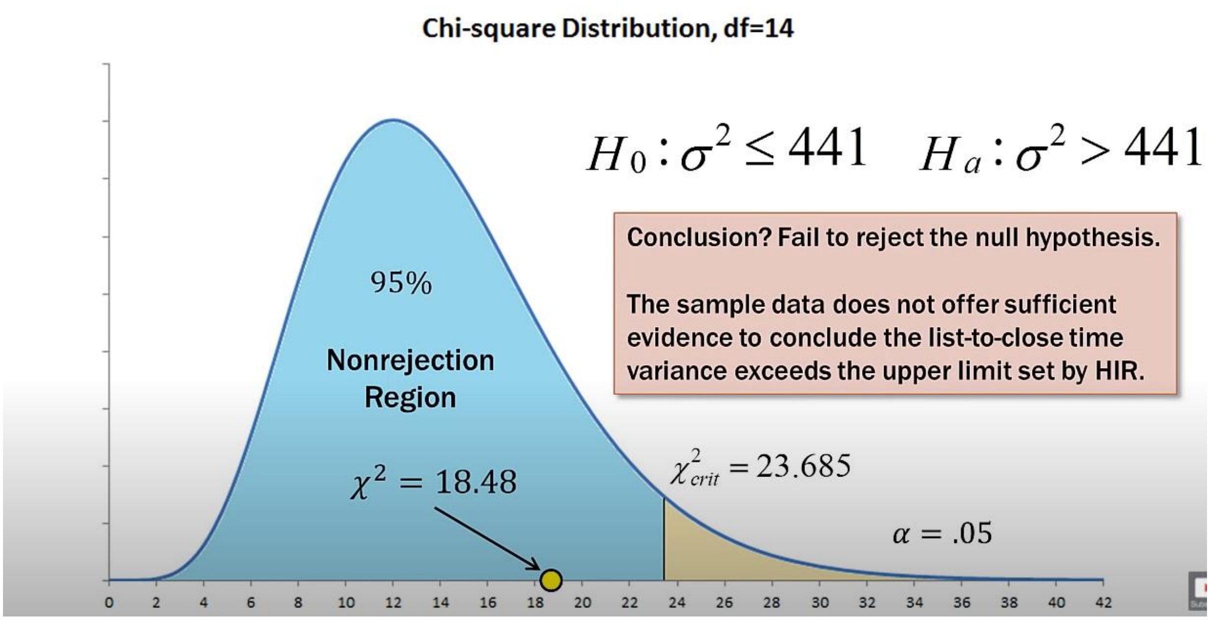
$$n = 15, s^2 = 582$$

HIGH INTEREST REALITY

Step 6: Calculate test statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad n = 15, s^2 = 582$$

$$\chi^2 = \frac{(15-1)582}{441} \qquad \chi^2 = 18.48$$



EXAMPLE 2: MICRO MACHINES

A new precision machine tool is being installed in a local production plant. Based on laser-read measurements, the machine being replaced produced a tooling variance of $s^2 = 100$ microns.

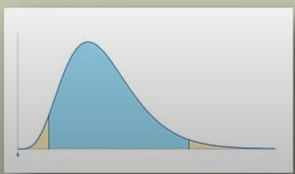
After a break-in period the new machine is tested to *see if its production variance is the same as the previous machine.*

Thirty samples are taken from the machine producing a sample variance, $s^2 = 162\ microns$.

HYPOTHESIS TESTING FOR VARIANCE

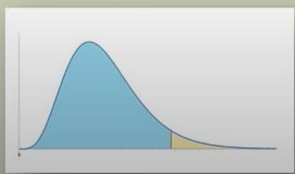
Which hypothesis test matches the structure of this problem?

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_a : \sigma^2 \neq \sigma_0^2$$



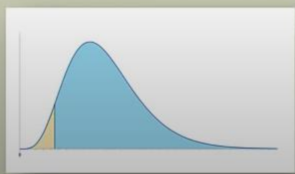
Two-tailed

$$H_0 : \sigma^2 \leq \sigma_0^2$$
$$H_a : \sigma^2 > \sigma_0^2$$



Upper/right-tailed

$$H_0 : \sigma^2 \geq \sigma_0^2$$
$$H_a : \sigma^2 < \sigma_0^2$$



Lower/left-tailed

MICRO MACHINES

Step 1: Establish Hypothesis

$$H_0 : \sigma^2 = 100 \qquad H_a : \sigma^2 \neq 100$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a two-tailed test. Since the new variance could be lower OR higher, we use the two-tailed test.

We will use the chi-square distribution with $30 - 1 = 29\ df$.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

MICRO MACHINES

Step 3: Specify the significance level

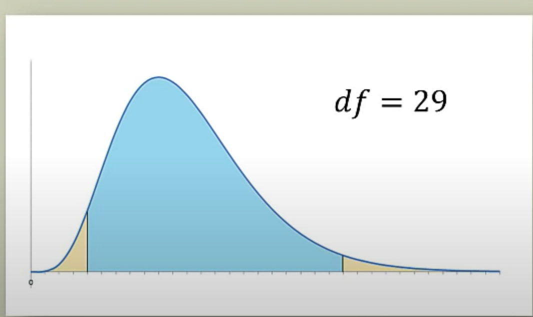
$$\alpha = .05$$

Step 4: State the decision rule

If $\chi^2 < 16.047$, reject H_0
If $\chi^2 > 45.722$, reject H_0

Step 5: Gather data

$$n = 30, s^2 = 162$$



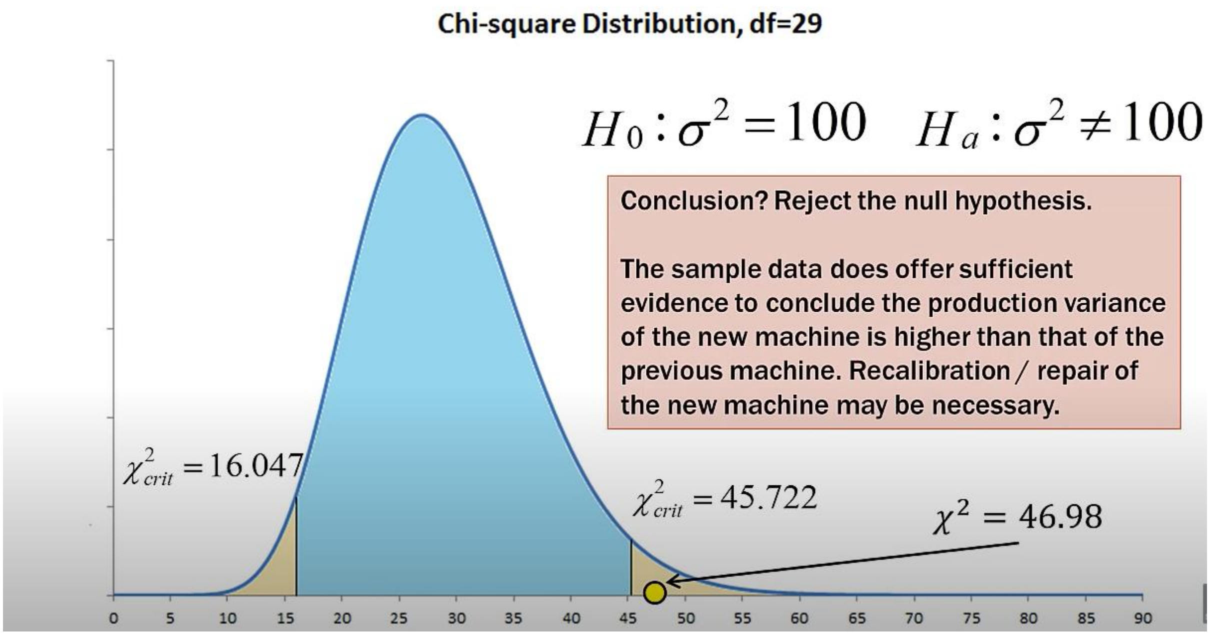
$\chi^2_{crit} = 16.047$ $\qquad$ $\chi^2_{crit} = 45.722$

MICRO MACHINES

Step 6: Calculate test statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad n = 30, s^2 = 162$$

$$\chi^2 = \frac{(30-1)162}{100} \qquad \chi^2 = 46.98$$



IMPORTANT NOTE

Keep in mind that we are saying nothing about the MEANS of these processes.

In the High Interest Realty example, the **mean** time-to-sale could be way off goal. But the target variance could be on goal. Think of producing a 95mm meter stick (WRONG)…very consistently!

In the Micro Machine example, the new machine could be producing **on average the exact specification**. But the problem is it is doing so *inconsistently*. Think of producing, on average, 100mm meter sticks. But some are 70mm and others are 130mm, etc.

QUALITY / PERFORMANCE MATRIX

Quality/Performance Monitoring		Mean, $\bar{x}$, on target?	
		YES	NO
Variance, σ^2, on target?	YES	GOOD! Keep monitoring.	Fix process mean
	NO	Fix process variation	BAD! Crisis meeting.

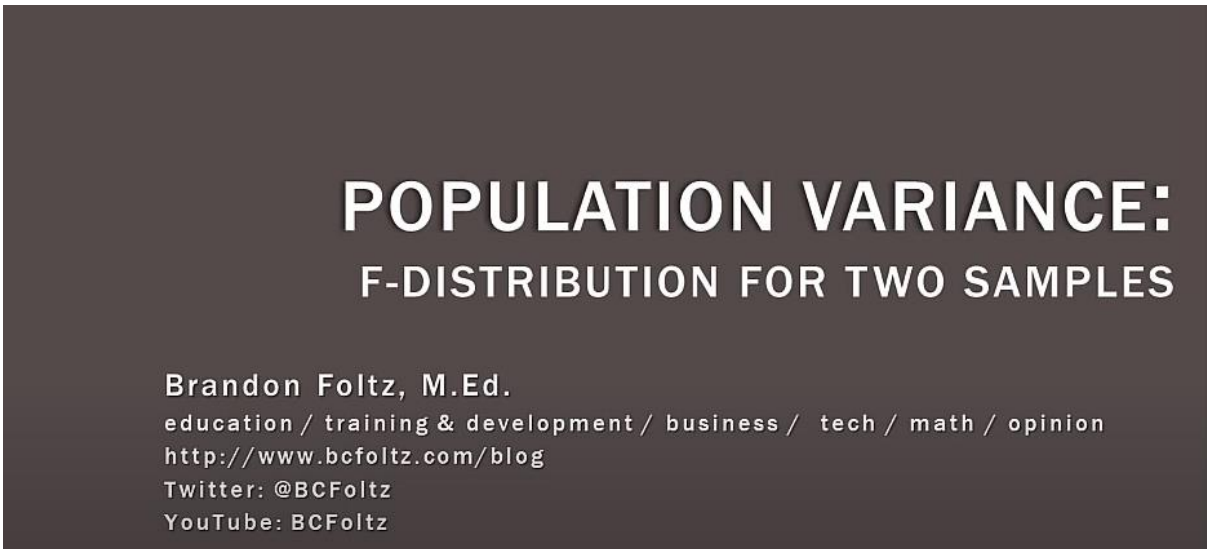
Meter stick Manufacturing		Mean, $\bar{x}$, on target?	
		YES	NO
Variance, σ^2, on target?	YES	Consistent @100mm	Consistent at 95mm
	NO	Inconsistent, $\bar{x} = 100mm$	All over the place

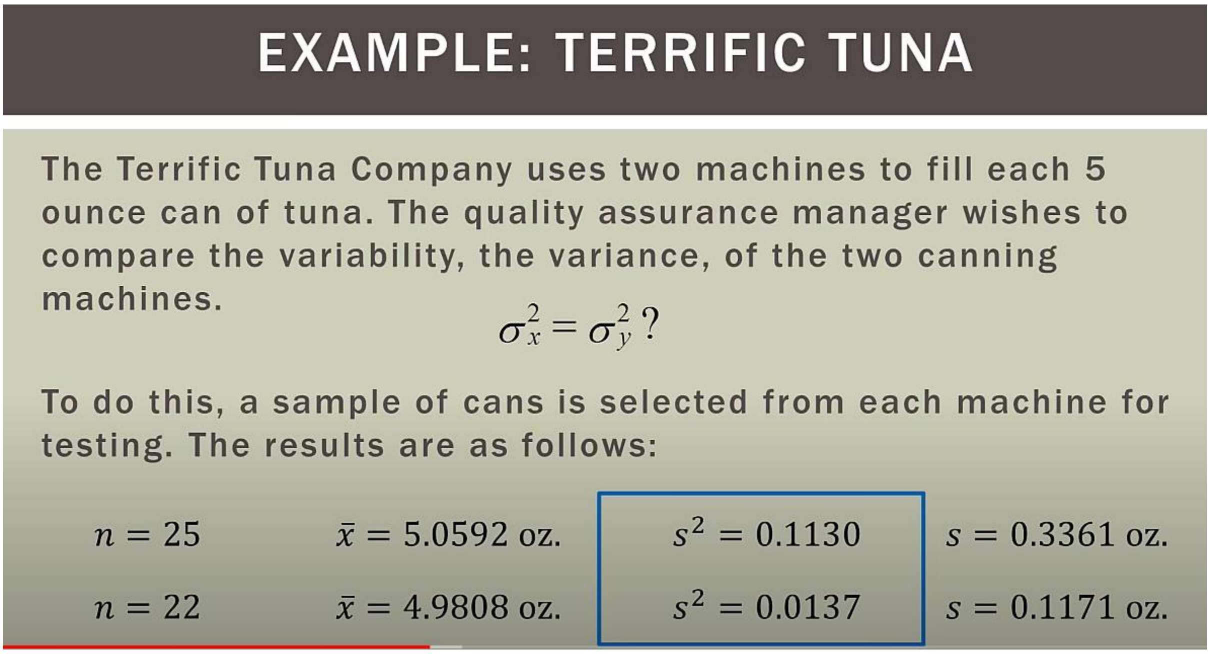
Video 4:

To conduct the variance for 2 samples from may be different sources – Achieved using **F-Distribution**

Let’s say to conduct quality of product produced from two machines.

So in nutshell, when we are comparing TWO SAMPLE VARIANCES with each other.





THE F-RATIO FOR POPULATION VARIANCES

- To conduct this analysis for equality of variances, we will need a new technique.
- Remember that we are not comparing the sample variance to a hypothesized variance; we are comparing two sample variances with each other.
- The easiest way to compare the relative size of two measurements is by using a ratio:

$$\frac{s_x^2}{s_y^2}$$

$s_x^2 =$ larger sample variance

$s_y^2 =$ smaller sample variance

$$F = \frac{s_x^2}{s_y^2}$$

F-ratio

THE F-DISTRIBUTION

- When independent random samples, n_x and n_y (or n_1 and n_2), are taken from two normal populations with equal variances, the sampling distribution of the ratio of those sample variances follows the F-distribution:

$$F = \frac{s_x^2}{s_y^2}$$

- The F-distribution is a distribution of ratios

THE F-DISTRIBUTION

$$F = \frac{s_x^2}{s_y^2}$$

$n_x - 1$ degrees of freedom for the numerator

$n_y - 1$ degrees of freedom for the denominator

$n = 25$	$\bar{x} = 5.0592$ oz.	$\boxed{s^2 = 0.1130}$	$s = 0.3361$ oz.
$n = 22$	$\bar{x} = 4.9808$ oz.	$s^2 = 0.0137$	$s = 0.1171$ oz.

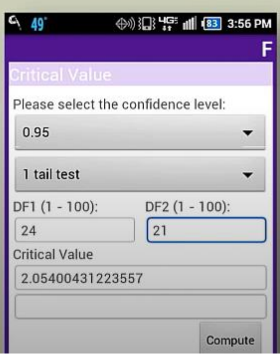
1. Find the larger of the two sample variances
2. That will be the numerator
3. Then find degrees of freedom based on sample size.

$$df = 25 - 1 = 24$$

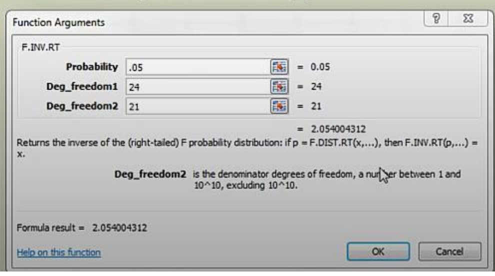
$$df = 22 - 1 = 21$$

READING THE F-TABLE

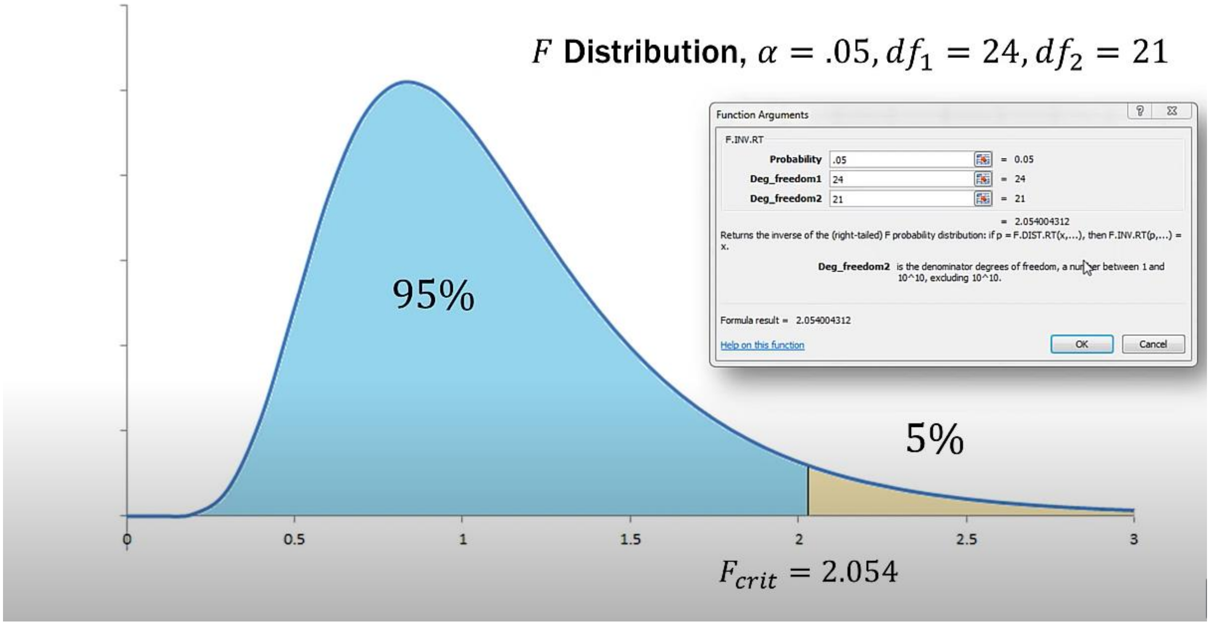
- Reading the F-table can be a real pain.
- You have a choice of several significance levels with a numerator and denominator degrees of freedom.



I use either the free Statistical Distribution app for Android by RealXY App or the Excel F.INV.RT function.



These are both digital F-tables and give precise values for a given alpha and degrees of freedom.

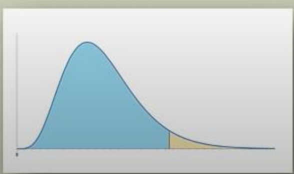


HYPOTHESIS TESTING FOR EQUALITY OF VARIANCE

Hypothesis tests for equality of variance are pretty straightforward:

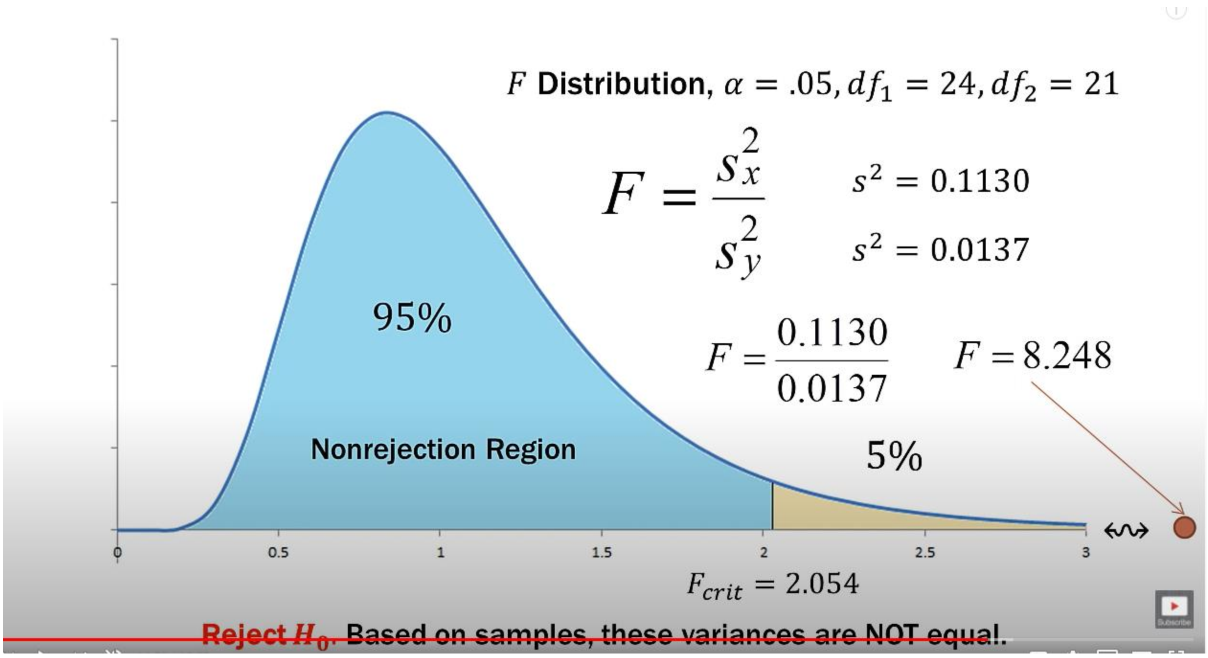
$$H_0 : \sigma_x^2 = \sigma_y^2$$

$$H_a : \sigma_x^2 \neq \sigma_y^2$$



Upper/right-tailed

- You might think that based on the $=$ and $\neq$ signs this should be a two-tailed distribution.
- However remember that we always place the larger sample variance in the numerator
- Therefore the F-ratio in this type of problem is always an upper-tailed test/distribution

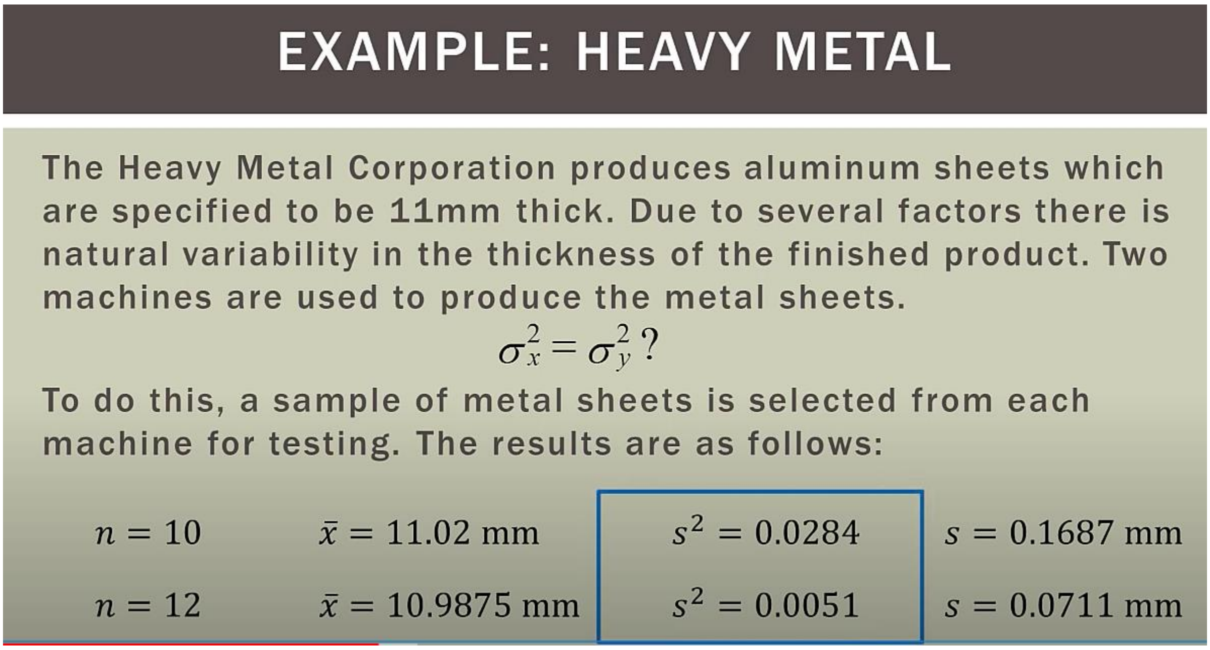


F-RATIO COMPARING VARIANCES RECAP

- The F-ratio is based on two sample variances
- The larger variance is placed in the numerator, the smaller in the denominator
- The critical F-value is found using the F-table (or digitally) with chosen alpha-level, numerator degrees of freedom of $n-1$, and denominator degrees of freedom of $n-1$
- The test statistic is the ratio of sample variances
- If the test statistic is larger than the critical F-value, reject the null hypothesis

Video 5:


POPULATION VARIANCE:
F-RATIO TEST EXAMPLES FOR EQUALITY OF VARIANCES
Brandon Foltz, M.Ed.
education / training & development / business / tech / math / opinion
http://www.bcfoltz.com/blog
Twitter: @BCFoltz
YouTube: BCFoltz


EXAMPLE: HEAVY METAL
The Heavy Metal Corporation produces aluminum sheets which are specified to be 11mm thick. Due to several factors there is natural variability in the thickness of the finished product. Two machines are used to produce the metal sheets.
$\sigma_x^2 = \sigma_y^2$?
To do this, a sample of metal sheets is selected from each machine for testing. The results are as follows:
$n = 10$ $\bar{x} = 11.02$ mm $s^2 = 0.0284$ $s = 0.1687$ mm
$n = 12$ $\bar{x} = 10.9875$ mm $s^2 = 0.0051$ $s = 0.0711$ mm

THE F-DISTRIBUTION

$$F = \frac{s_x^2}{s_y^2}$$

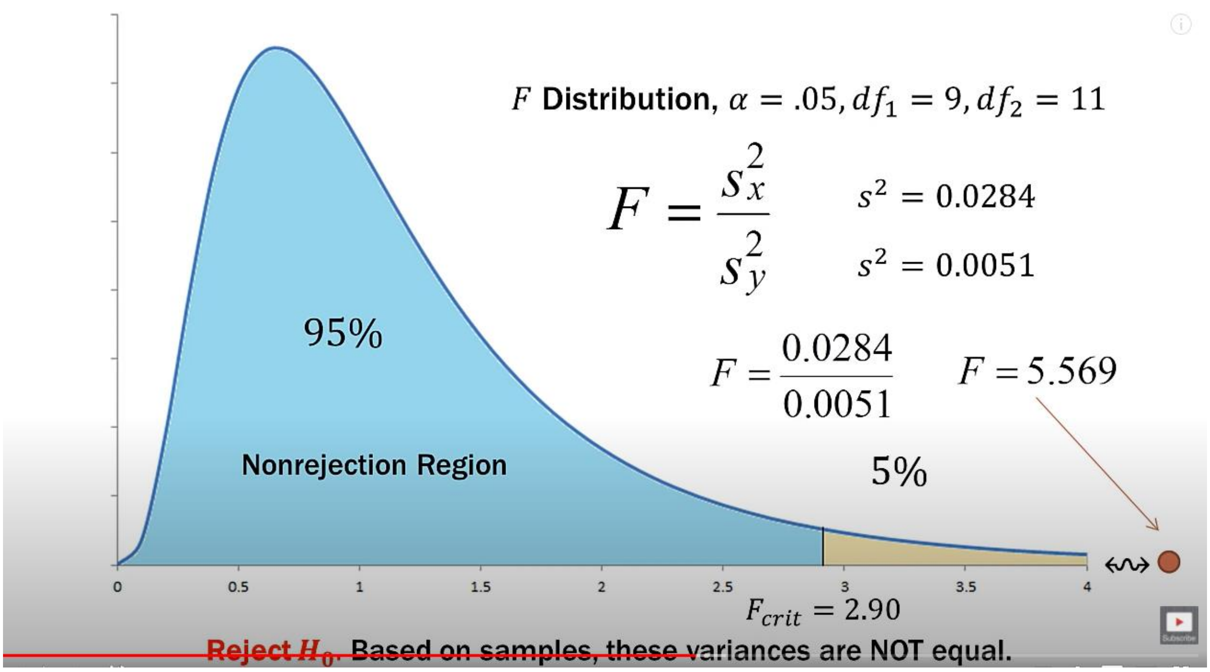
$n_x - 1$ degrees of freedom for the numerator

$n_y - 1$ degrees of freedom for the denominator

$n = 10$	$\bar{x} = 11.02$ mm	$\boxed{s^2 = 0.0284}$	$s = 0.1687$ mm
$n = 12$	$\bar{x} = 10.9875$ mm	$s^2 = 0.0051$	$s = 0.0711$ mm

1. Find the larger of the two sample variances
2. That will be the numerator
3. Then find degrees of freedom based on sample size.

$$\frac{df = 10 - 1 = 9}{df = 12 - 1 = 11}$$



EXAMPLE: GAS PROBLEMS

A consumer rights group launched an investigation based on accusations of price gouging at fuel stations in a certain city. To test this possibility, the group compared the variances between the suspected city and a similar city where no complaints have been made.

$$\sigma_x^2 = \sigma_y^2\,?$$

To do this, a sample of fuel prices is selected from each city:

$n = 10$	$\bar{x} = \$3.42$	$s^2 = 0.0096$	$s = \$0.0980$
$n = 10$	$\bar{x} = \$3.45$	$s^2 = 0.0118$	$s = \$0.1088$

THE F-DISTRIBUTION

$$F = \frac{s_x^2}{s_y^2}$$

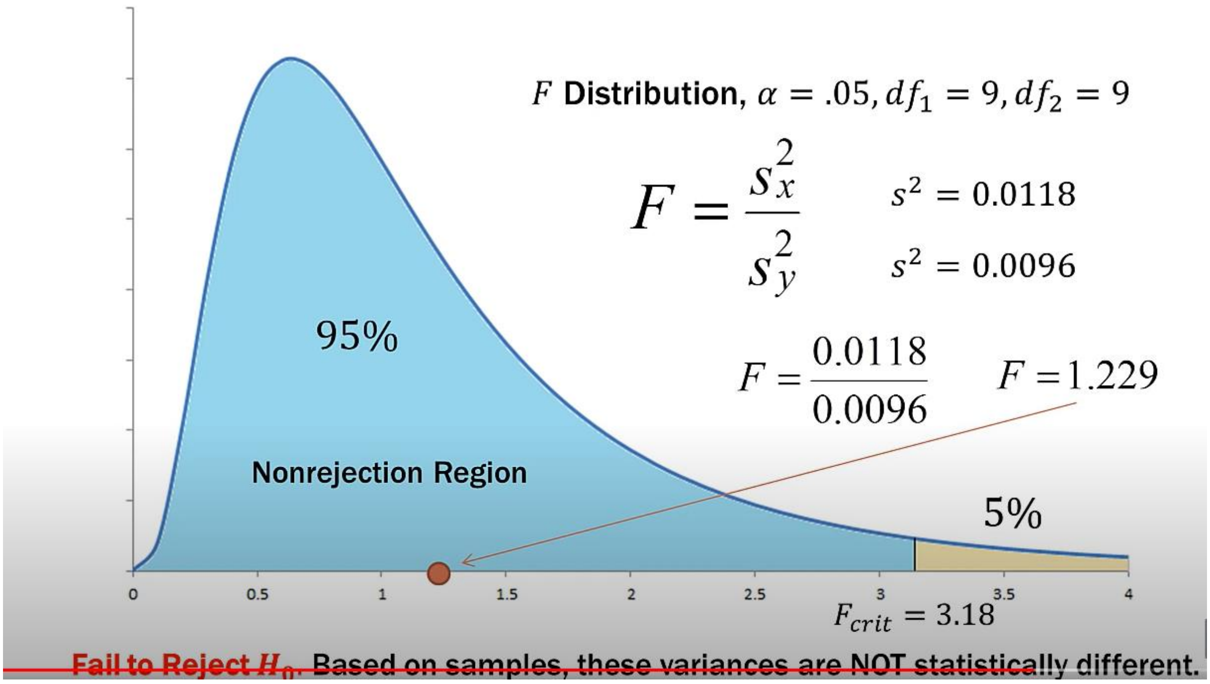
$n_x - 1$ degrees of freedom for the numerator

$n_y - 1$ degrees of freedom for the denominator

$n = 10$	$\bar{x} = \$3.42$	$s^2 = 0.0096$	$s = \$0.0980$
$n = 10$	$\bar{x} = \$3.45$	$\boxed{s^2 = 0.0118}$	$s = \$0.1088$

1. Find the larger of the two sample variances
2. That will be the numerator
3. Then find degrees of freedom based on sample size.

$$\frac{df = 10 - 1 = 9}{df = 10 - 1 = 9}$$



F-RATIO COMPARING VARIANCES RECAP

- The F-ratio is based on two sample variances
- The larger variance is placed in the numerator, the smaller in the denominator
- The critical F-value is found using the F-table (or digitally) with chosen alpha-level, numerator degrees of freedom of $n-1$, and denominator degrees of freedom of $n-1$
- The test statistic is the ratio of sample variances
- If the test statistic is larger than the critical F-value, reject the null hypothesis