Till now

- Prob mass fn $\longrightarrow$ Complete.
- Prob density function
  - N.D
  - SND
  - CLT

  } done

In this Class :-

1. Log Normal distn
2. t-distn
3. Pareto
4. Exponential
5. Chi-square
6. F-distn.

① t-distn

$$Z \text{ Score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

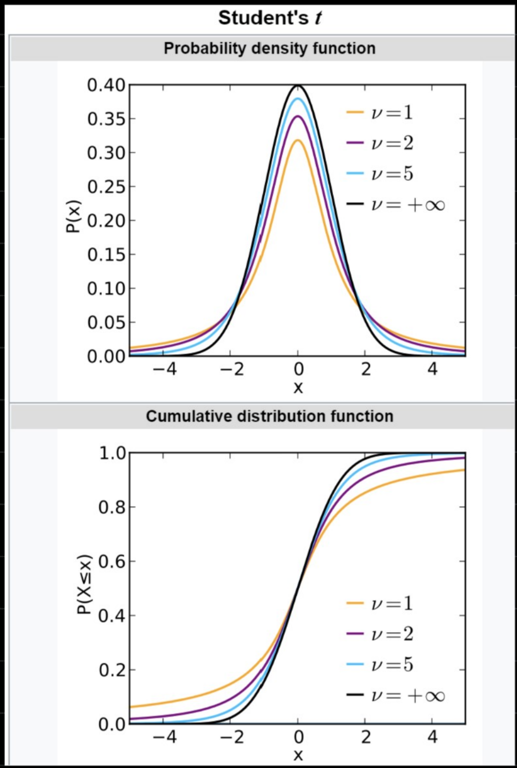
Condition $\sigma$ of Pop is given
and
S.S $\geq 30$

but what if $\sigma$ pop. not given and S.S $< 30$

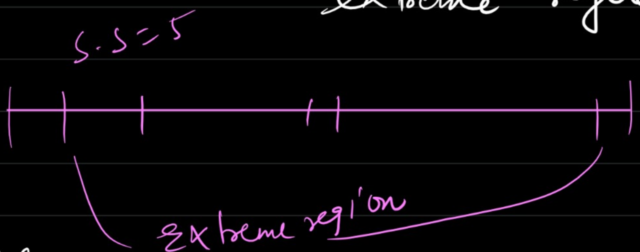
$\longrightarrow$ t-distn.

$$t\text{score} = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

For S.S ≥ 30 (Z-distⁿ)
↳ more date → less chance of dp's falling at extreme regions at other dp's



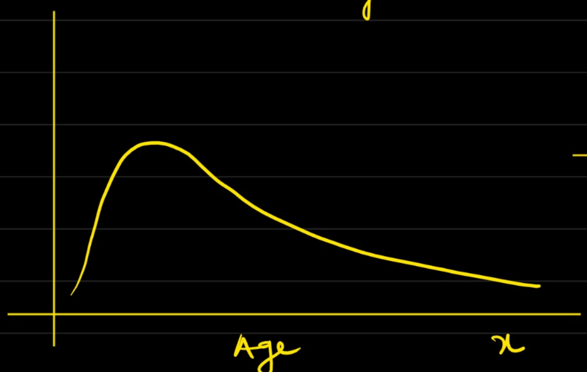

for S.S < 30
↳ less date point → High chance of dp's faleig at extreme region.



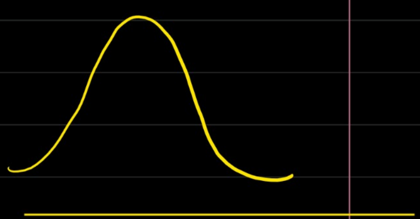

* Use
↳ t-test

② * log Normal.

→ A continuous prob distⁿ whose logarithm is Normally distributed.

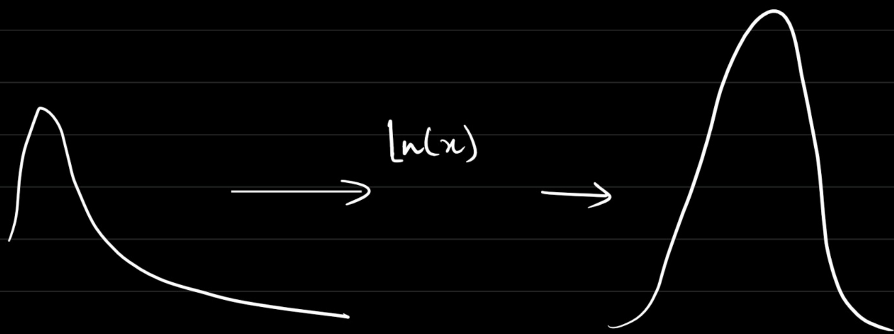


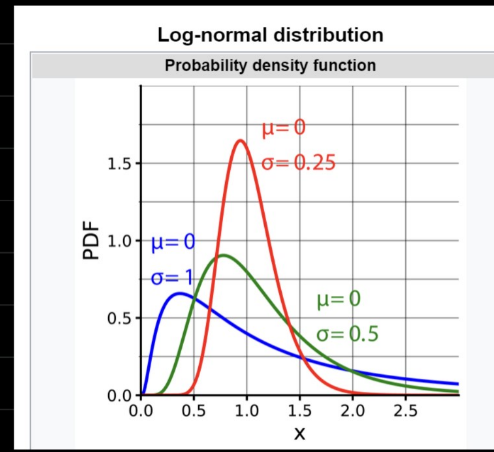

$\xrightarrow{\text{log transformation}}$
ln(x)
$x = \exp(y)$



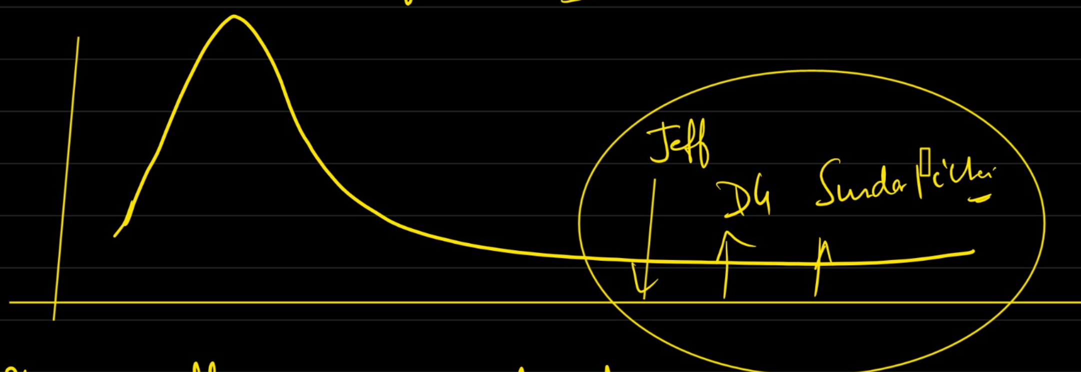
→ logarithic distⁿ to Normal distⁿ.

Why → log Normal $dist^n$




Example → Wealth $dist^n$ of world.. (rich attracts rich)

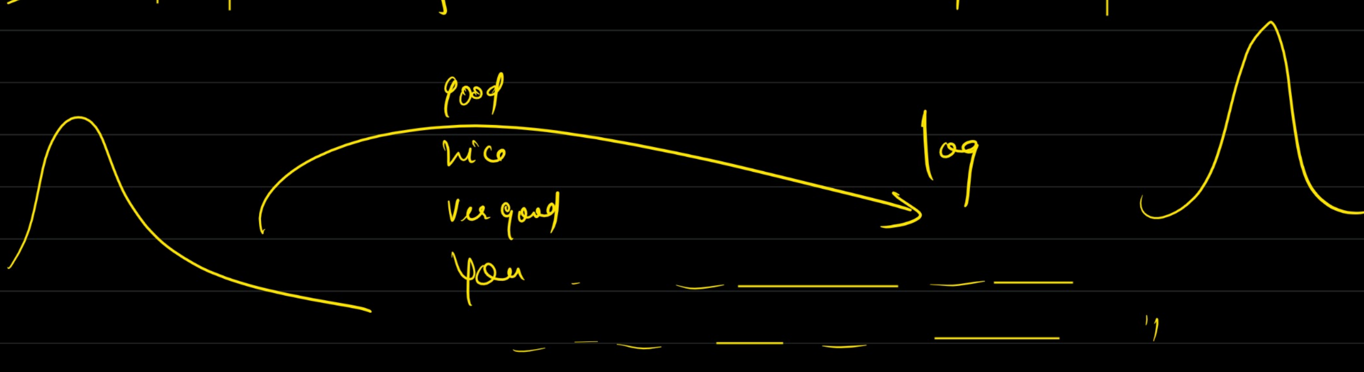



* Where outlier can not be dropped, you use log of N.D.

$$X \sim \log Normal(\mu, \sigma^2)$$

eg: time spend in reading Articles
editorial

eg People writing comments on your post.



$$pdf = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$$

③ Power law distn ( Pareto principle )

⇓

80:20

RCB vs CSK.

⇓

K G F

Kohli Glen maxwell Faf duplessis

20% of player will 80% of the score.

→ 20% of employee in a team does 80% of the work

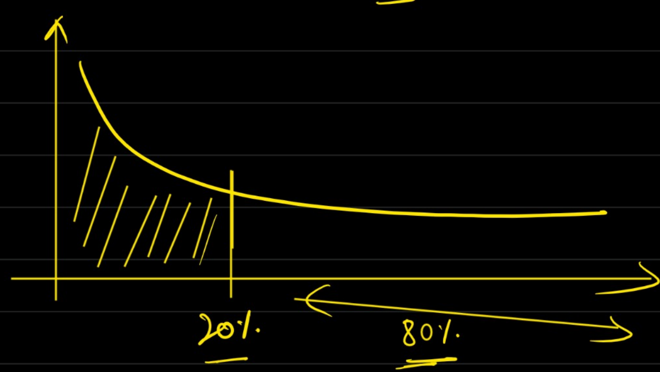
→ 20% of people are super excited and energetic for 80% of the time in a party.

→ 80% of sales will happen due to 20% of customer.

→ 80% of world's wealth is distributed with 20% of people.

20% of boys attract 80% girls attraction.

pdf.

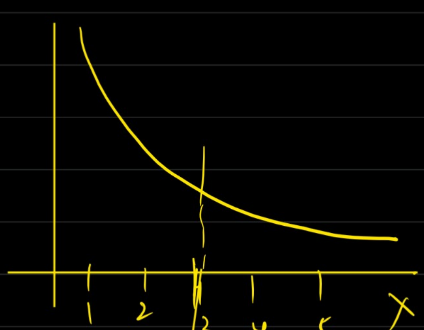



* Power law principle

$$pdf = \frac{\alpha\, x_m^{\alpha}}{x_m^{\alpha+1}}$$

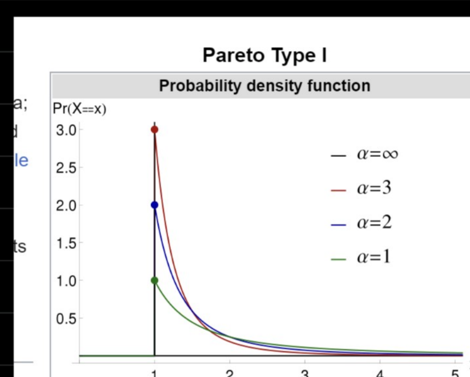
$x_m \rightarrow$ all $x$.

$x$




$x_m = 3$

$x = 3$

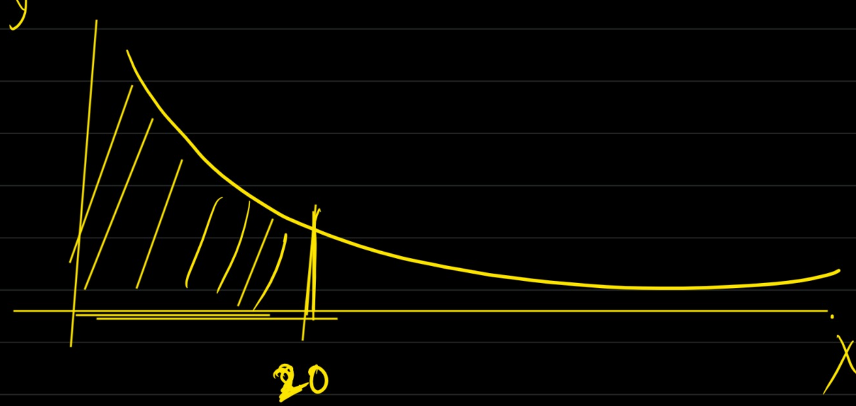



if x is a random varible with Pareto type-I the the prob the $x >$ some no

$$F(x) = \left(\frac{x_m}{x}\right)^{\alpha} \quad x \geq x_m$$

$x_m = 1, \quad x = 4 \quad \underline{\alpha = 2}$ $\qquad 0$ if $x < x_m$

$$P(x \geq 4) = \left(\frac{x_m}{x}\right)^{\alpha} = \left(\frac{1}{4}\right)^2 = 1/16$$




20% of $x$ will cover 80% of Y.

Power law

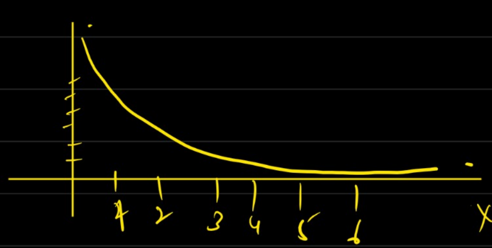
Exp Pareto

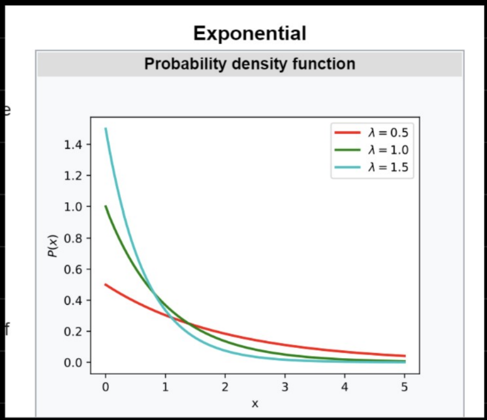
④ Exponential distn

$$\text{pdf} = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0. \end{cases}$$

→ Radio active decay

→ life span of battery




⑤ Chi-square distn

The Chisquare distn is a prob distn that describes the distribution of sum of squares of K random variable.

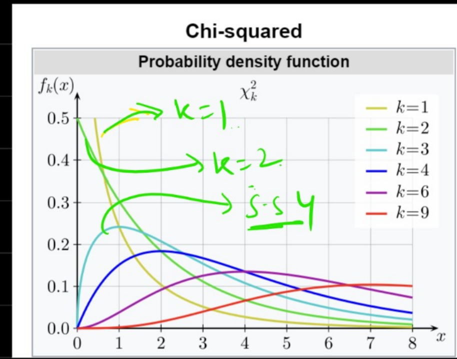
dof = S.S − 1

sample (10)

dof = 10 − 1 = 9.

K = 1
(dof = 1)
dof = S.S − 1
dof + 1 = S.S.
SS = 1 + 1
= 2

K = 2 − 1 = 1

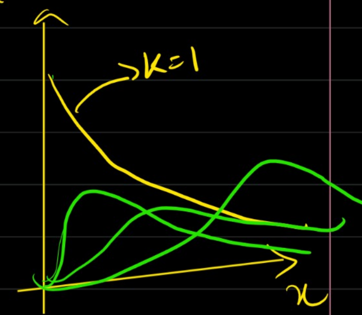



Population


Sample (2)
$S_1$ (1,2)
$S_2$ (2,3)
$S_3$ (4,5)
⋮

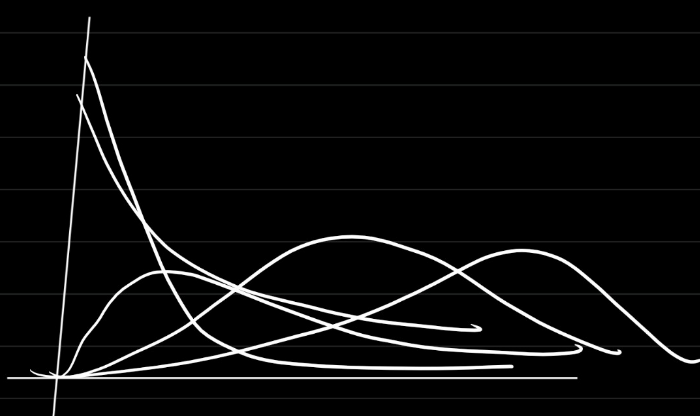
Square the sample

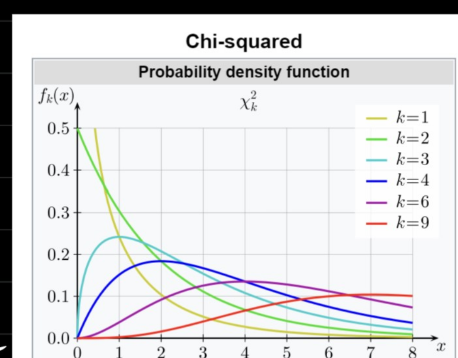
$1^2 + 2^2$
$2^2 + 3^2$
$4^2 + 5^2$
⋮




→ determined by k.
→ Non negative distn
→ right skewed data.
→ for small k → Pareto distnⁿ, but as k increase it tends to N.D

Why?

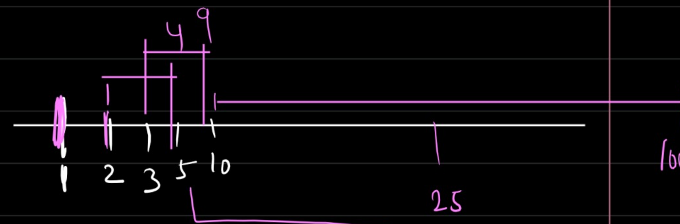




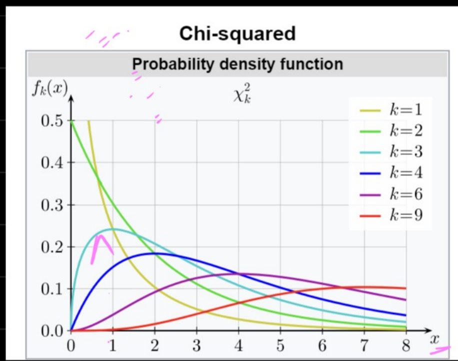

① Square ⟶ +ve.

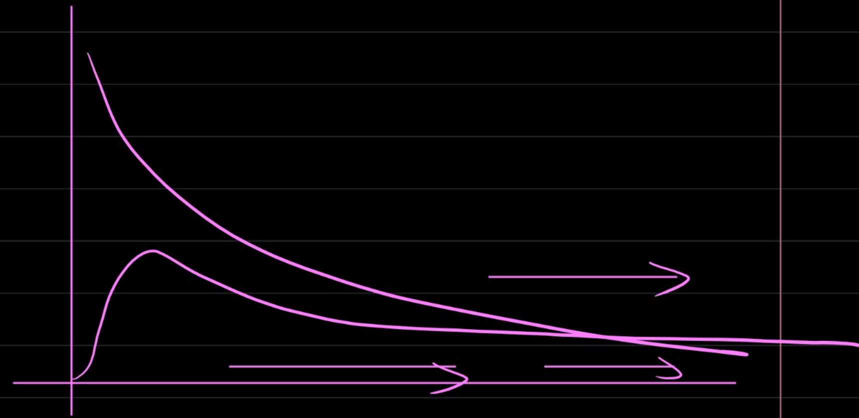
②

$1^2 = 1$
$2^2 = 4$
$3^2 = 9$
$5^2 = 25$
$10^2 = 100$




Due to square the distⁿ itself pushes towards +ve side



| $x < 1$ | $S.S < 4$ |
|---|---|
| 0.1 | |
| 0.2 | |
| 0.3 | |

$(0.1)^2$

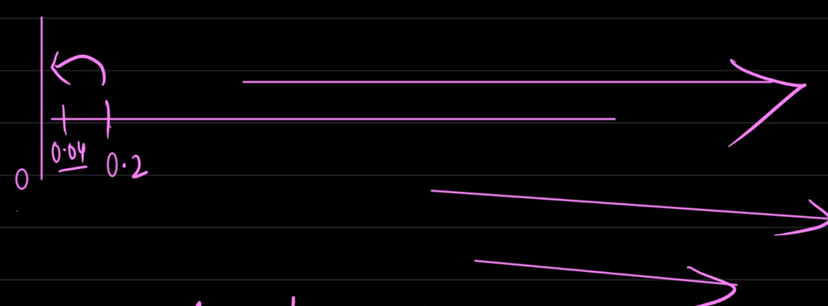
$(0.3)^2$

$0.1 \times 0.1 = 0.01$

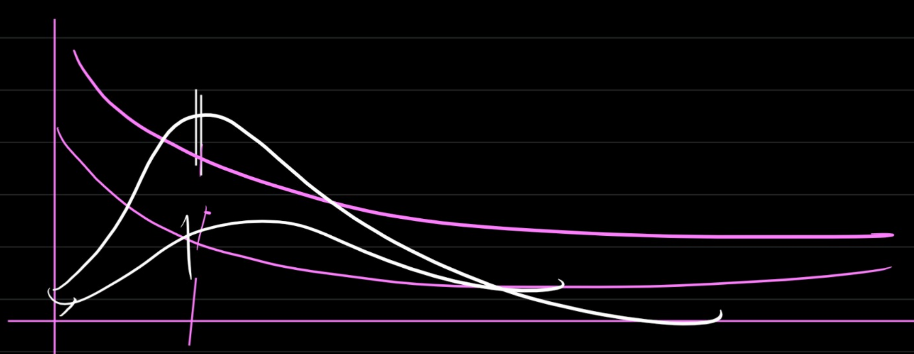
| | |
|---|---|
| 0.1 | $(0.1)^2$ |
| 0.3 | $0.3^2$ |

→ $x < 1$, Square becomes more and more smaller

$x < 1$




It will be pushed to origin




← S.S = 4
d.o.f = 1

0.1 0.6 0.7 0.9

$(0.1)^2 + (0.6)^2 + (0.7)^2 + (0.9)^2$

Sum will be always greater than (individ no)$^2$ $\Rightarrow$ $\boxed{0.01 + 0.36 + 0.49 + 0.81} =$

0.1 0.6 0.7 0.9

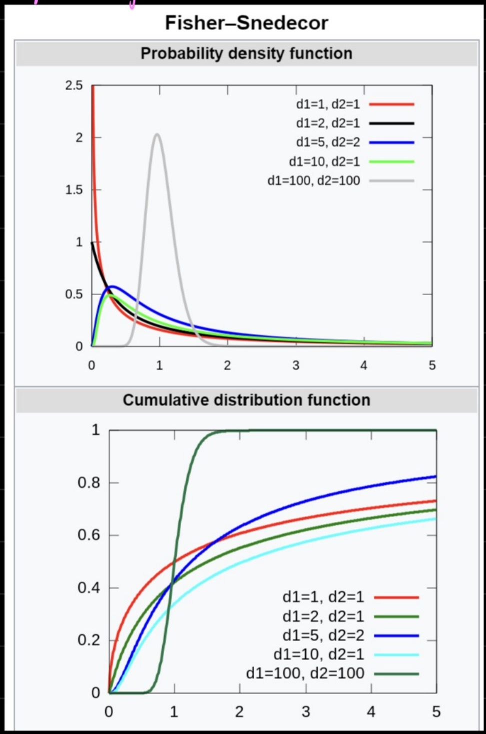
**Use**

- in ML (to compare two categorical variables)
- In hypothesis testing.

## * F distribution (Fisher Snedecor dist$^n$)

→ The F dis$^n$ is a probability dis$^n$ that is useful in context of comparing variances of two or more samples.

→ It is right skewed, takes only +ve values




The F dist$^n$ with d1 and d2 is the dis$^n$ given by $X = \dfrac{S_1/d_1}{S_2/d_2}$

It can be shown to follow that the probability density function (pdf) for $X$ is given by

$$f(x; d_1, d_2) = \frac{\sqrt{\frac{(d_1 x)^{d_1}\, d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$

$$= \frac{1}{\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} x^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}\,x\right)^{-\frac{d_1+d_2}{2}}$$

$S_1$ $S_2$.

$\dfrac{\text{Std dev}}{df_1}$ $\dfrac{\text{Std dev}}{df_2}$

$\underbrace{\qquad\qquad}_{\text{ratio}}$ → ratio of variance

Use case

↳ Compare two Sample variance using F test.

$\Downarrow$

$\boxed{\text{F dist}^n}$