Student's t Distribution

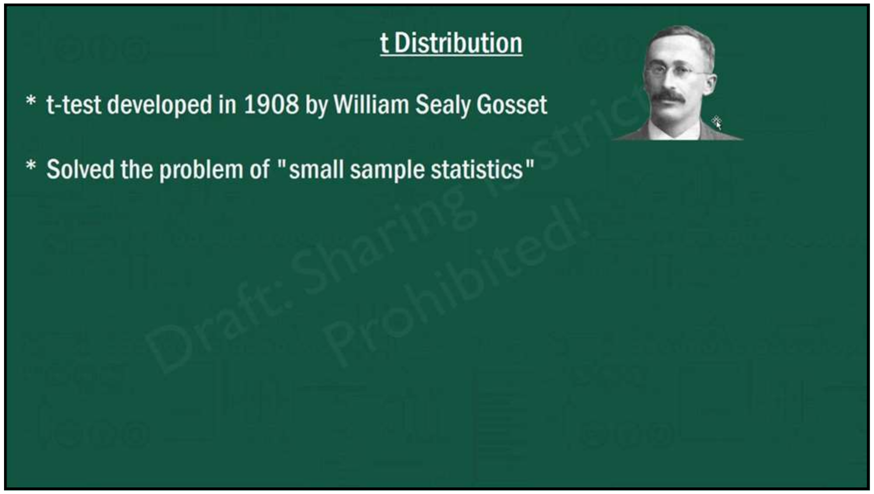
1. Overview

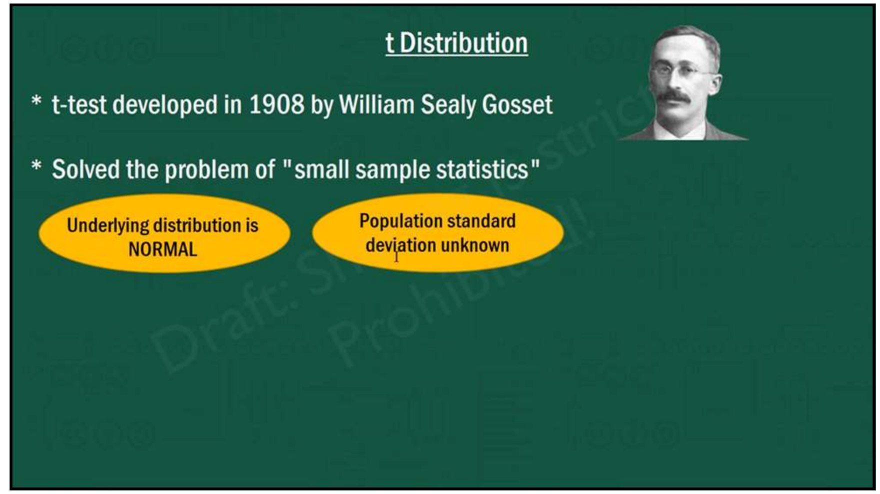
t Distribution

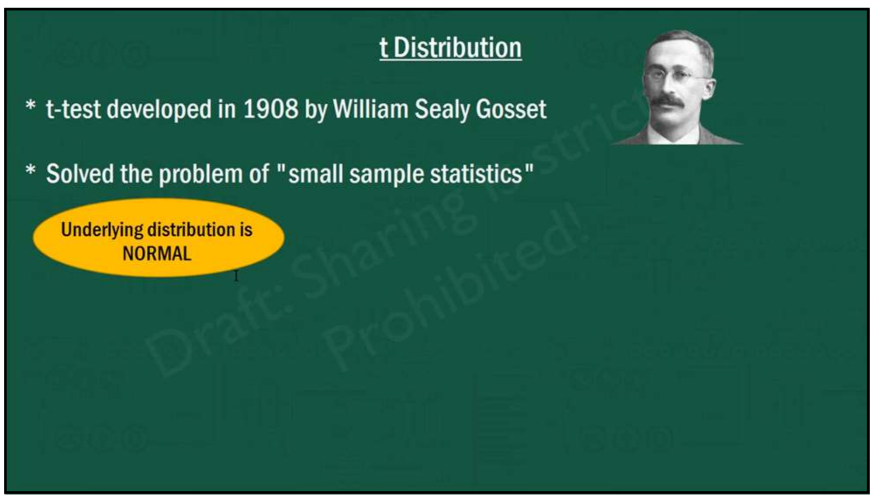
1. Overview → What's it used for?
2. Sampling Recap → Put it into context
3. Visual Comparison → What does it look like?
4. PDF and CDF Calculations → Use it!

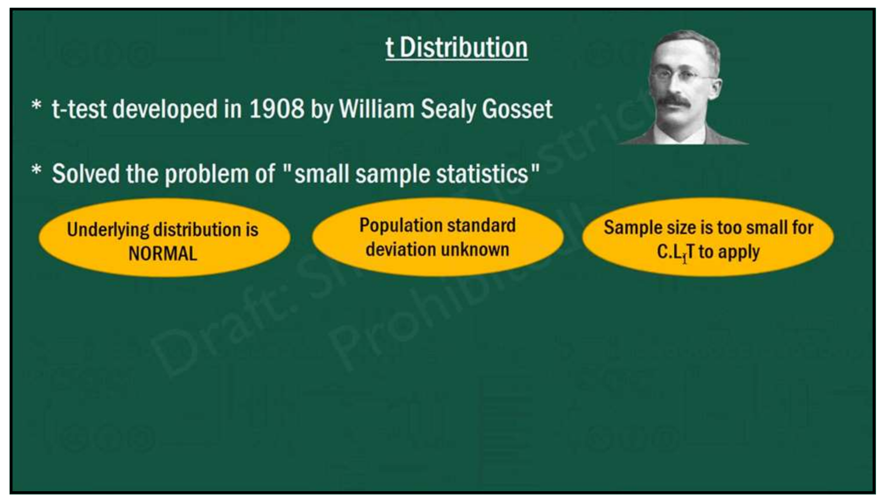
t Distribution

* t-test developed in 1908 by William Sealy Gosset


t Distribution
* t-test developed in 1908 by William Sealy Gosset
* Solved the problem of "small sample statistics"


t Distribution
* t-test developed in 1908 by William Sealy Gosset
* Solved the problem of "small sample statistics"
Underlying distribution is NORMAL
Population standard deviation unknown


t Distribution
* t-test developed in 1908 by William Sealy Gosset
* Solved the problem of "small sample statistics"
Underlying distribution is NORMAL


t Distribution
* t-test developed in 1908 by William Sealy Gosset
* Solved the problem of "small sample statistics"
Underlying distribution is NORMAL
Population standard deviation unknown
Sample size is too small for C.L.T to apply

t Distribution

* t-test developed in 1908 by William Sealy Gosset

* Solved the problem of "small sample statistics"

Underlying distribution is NORMAL

Population standard deviation unknown

Sample size is too small for C.L.T to apply

* The following measures would be t-distributed

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \qquad \frac{b-\beta}{SE(b)} \qquad \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{s_1^2}{\sqrt{n_1}}\right)+\left(\frac{s_2^2}{\sqrt{n_2}}\right)}}$$

t Distribution

SAMPLING RECAP!

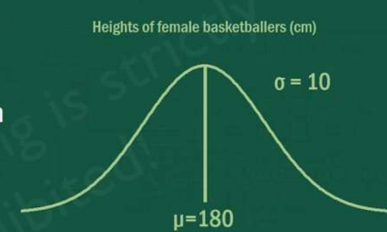
* Take a sample of five observations
  from a normally distributed population

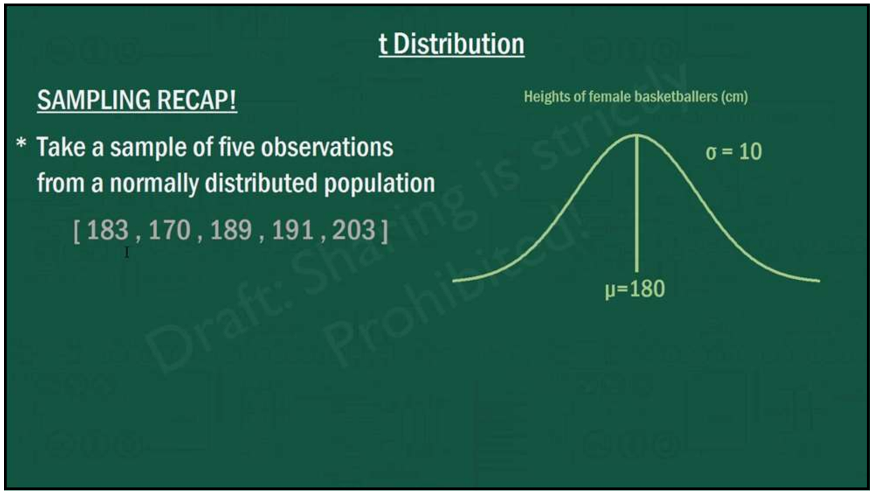
2. Sampling Recap

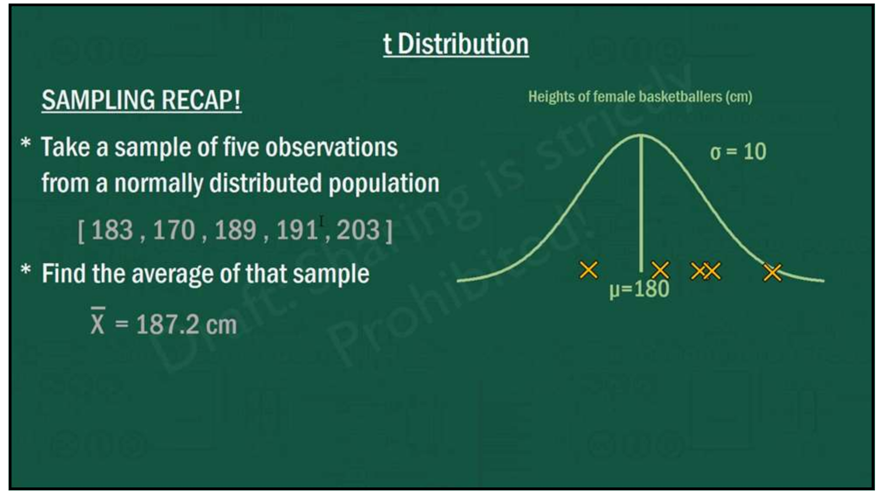
t Distribution

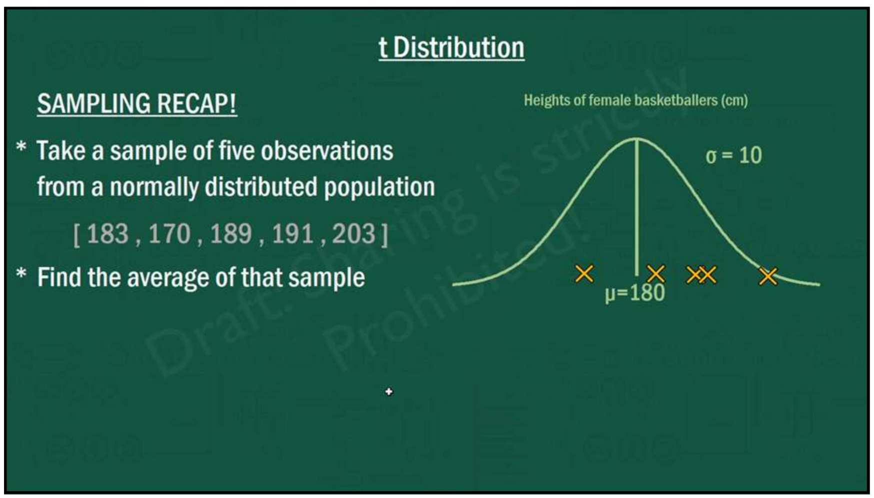
SAMPLING RECAP!

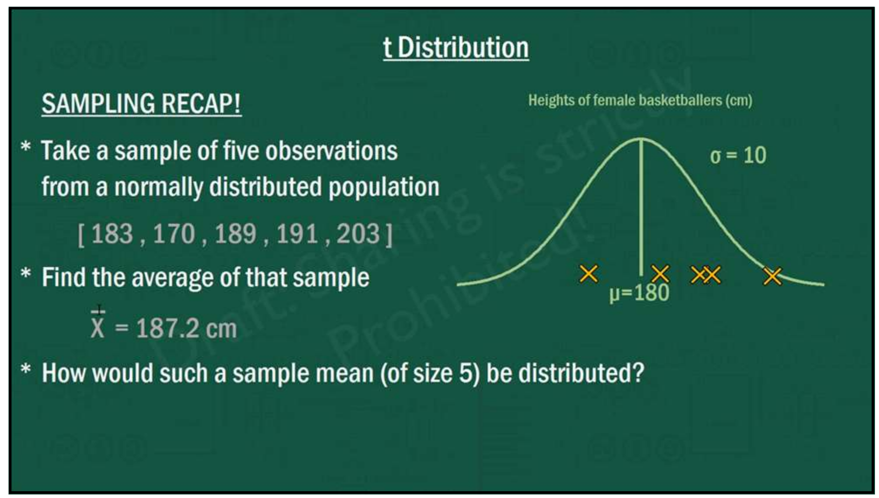
* Take a sample of five observations
  from a normally distributed population




t Distribution
SAMPLING RECAP!
* Take a sample of five observations from a normally distributed population
[183 , 170 , 189 , 191 , 203]
Heights of female basketballers (cm)
σ = 10
μ=180


t Distribution
SAMPLING RECAP!
* Take a sample of five observations from a normally distributed population
[183 , 170 , 189 , 191 , 203]
* Find the average of that sample
X̄ = 187.2 cm
Heights of female basketballers (cm)
σ = 10
μ=180


t Distribution
SAMPLING RECAP!
* Take a sample of five observations from a normally distributed population
[183 , 170 , 189 , 191 , 203]
* Find the average of that sample
Heights of female basketballers (cm)
σ = 10
μ=180

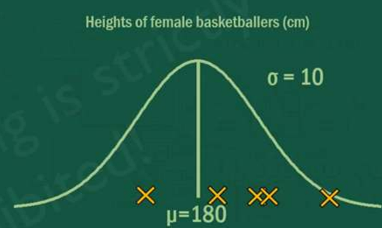

t Distribution
SAMPLING RECAP!
* Take a sample of five observations from a normally distributed population
[183 , 170 , 189 , 191 , 203]
* Find the average of that sample
X̄ = 187.2 cm
* How would such a sample mean (of size 5) be distributed?
Heights of female basketballers (cm)
σ = 10
μ=180

t Distribution

SAMPLING RECAP!

* Take a sample of five observations from a normally distributed population

  [183 , 170 , 189 , 191 , 203]



* Find the average of that sample

  $\bar{X}$ = 187.2 cm

* How would such a sample mean (of size 5) be distributed?

$$\bar{x} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right) \qquad \bar{x} \sim N\left(180, \left(\frac{10}{\sqrt{5}}\right)^2\right)$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\bar{X}$ = 187.2 cm

* H_0: μ =180 H_1: μ ≠ 180

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample
* H_0: μ =180 H_1: μ ≠ 180

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\bar{X}$ = 187.2 cm

* H_0: μ =180 H_1: μ ≠ 180

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim z$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\overline{X}$ = 187.2 cm

* H_0: μ =180 H_1: $\mu \neq 180$

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\overline{x}-\mu}{s/\sqrt{n}}$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\overline{X}$ = 187.2 cm

s = 12.05 cm

* H_0: μ =180 H_1: $\mu \neq 180$

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim t$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\overline{X}$ = 187.2 cm

s = 12.05 cm

* H_0: μ =180 H_1: $\mu \neq 180$

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\overline{x}-\mu}{s/\sqrt{n}}$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\overline{X}$ = 187.2 cm

s = 12.05 cm

* H_0: μ =180 H_1: $\mu \neq 180$

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\bar{X}$ = 187.2 cm

s = 12.05 cm

* H_0: $\mu = 180$ H_1: $\mu \neq 180$

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

* We need to adjust for the additional uncertainty around s.

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\bar{X}$ = 187.2 cm

s = 12.05 cm

* H_0: $\mu = 180$ H_1: $\mu \neq 180$

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$t = \frac{187.2-180}{12.05/\sqrt{5}} \sim t_4$$

* We need to adjust for the additional uncertainty around s.
* The smaller the sample size, the more uncertain we are.

t Distribution

SAMPLING RECAP!

* Imagine we are TESTING the population mean value of 180cm by using our sample

[183 , 170 , 189 , 191 , 203]

$\bar{X}$ = 187.2 cm

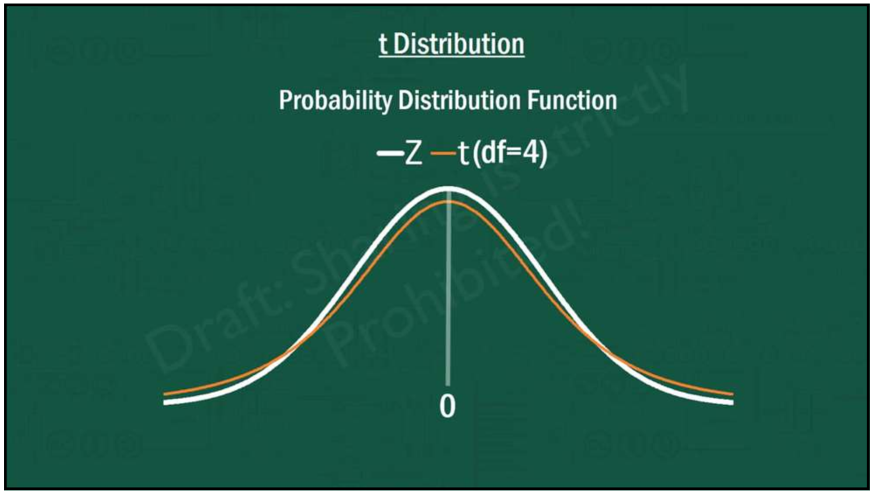
s = 12.05 cm

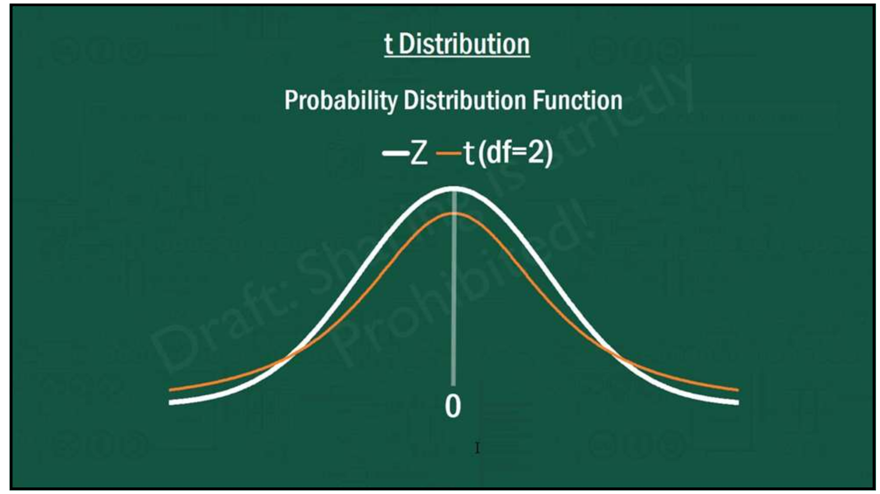
* H_0: $\mu = 180$ H_1: $\mu \neq 180$

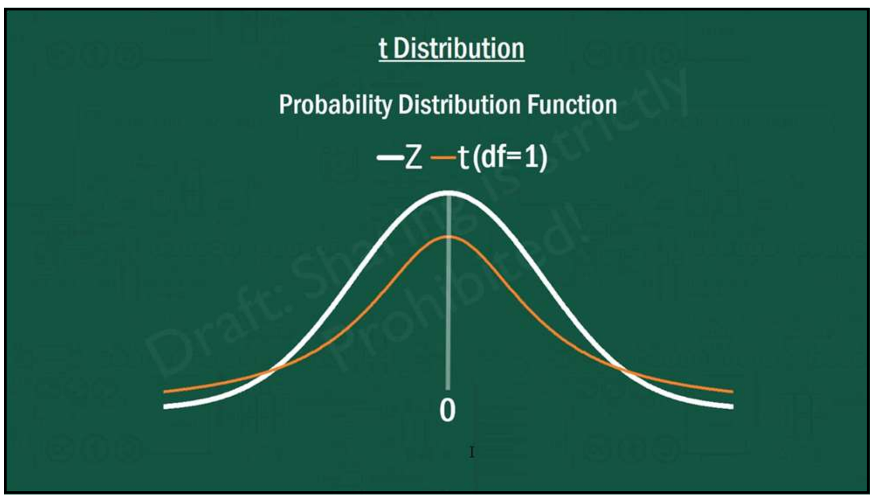
$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim z \longrightarrow \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

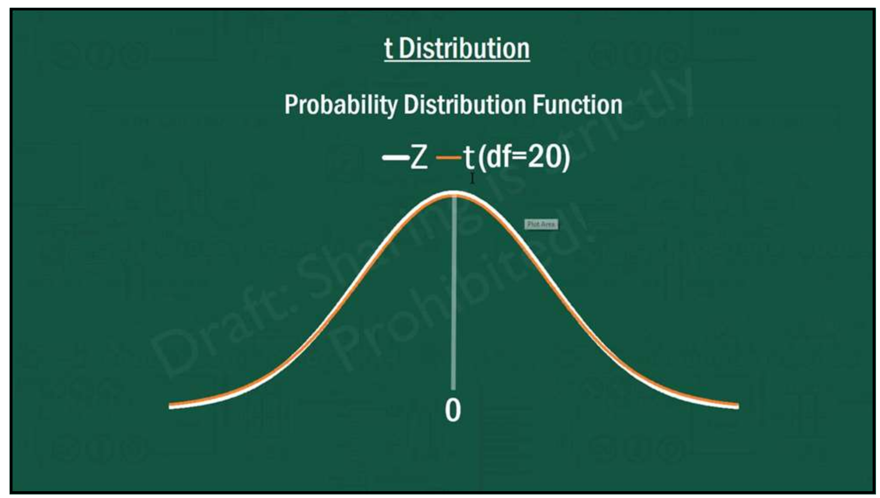
* We need to adjust for the additional uncertainty around s.
* The smaller the sample size, the more uncertain we are.

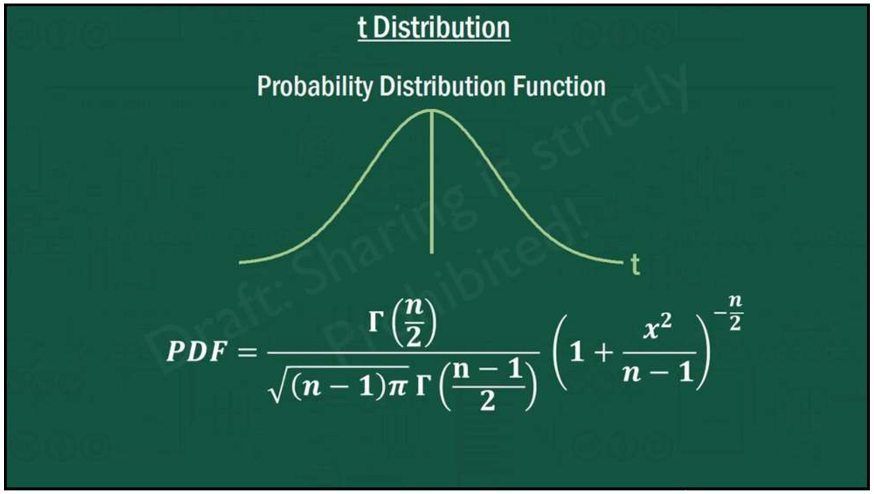
3. Visualisation

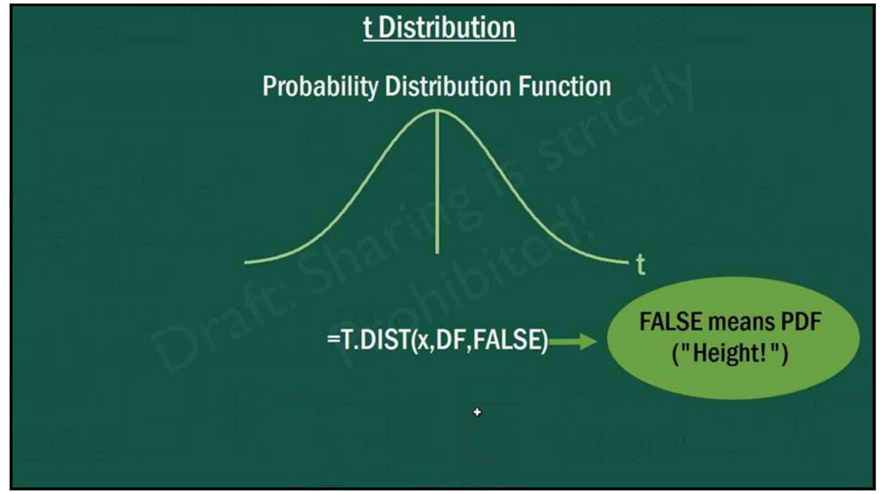

t Distribution
Probability Distribution Function
Z t(df=4)
0

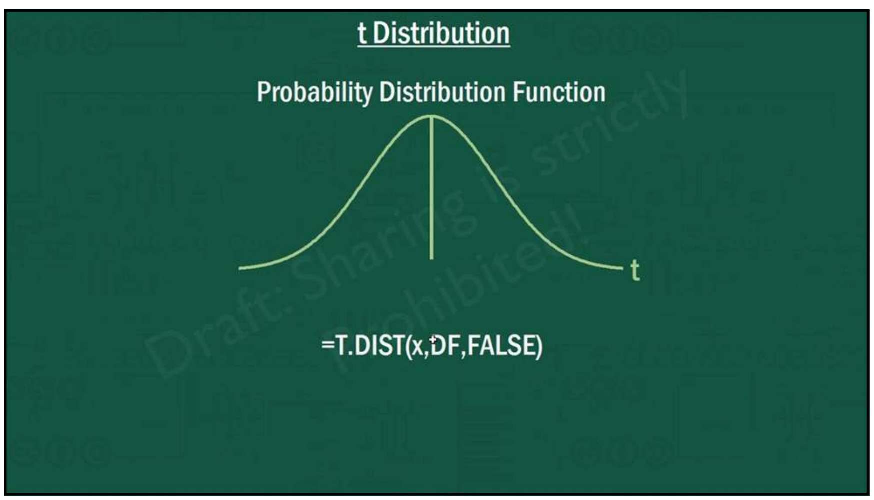

t Distribution
Probability Distribution Function
Z t(df=2)
0

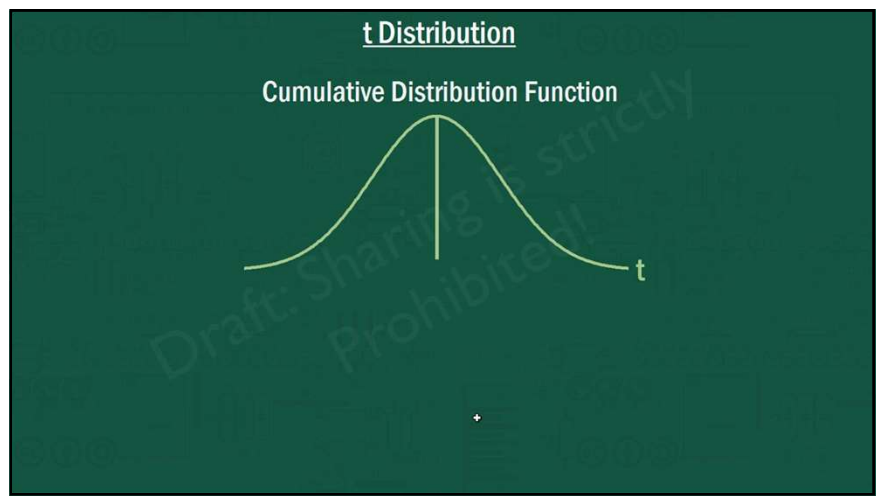

t Distribution
Probability Distribution Function
Z t(df=1)
0

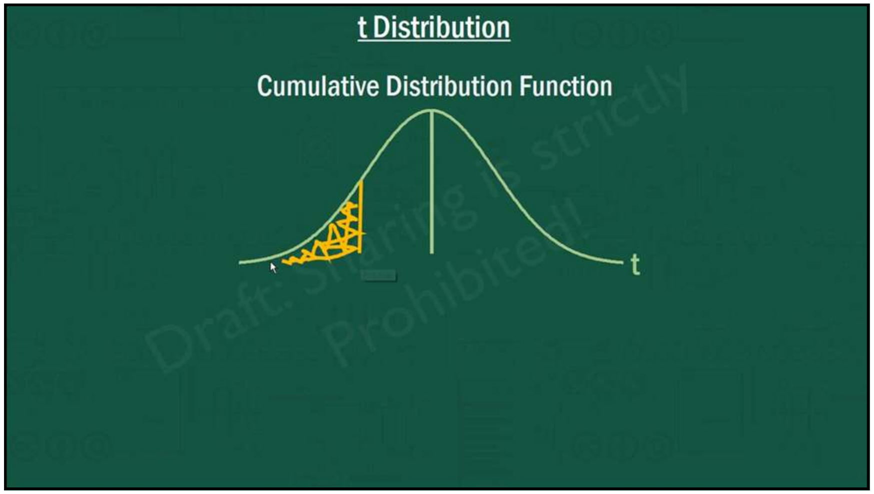

t Distribution
Probability Distribution Function
Z t(df=20)
0

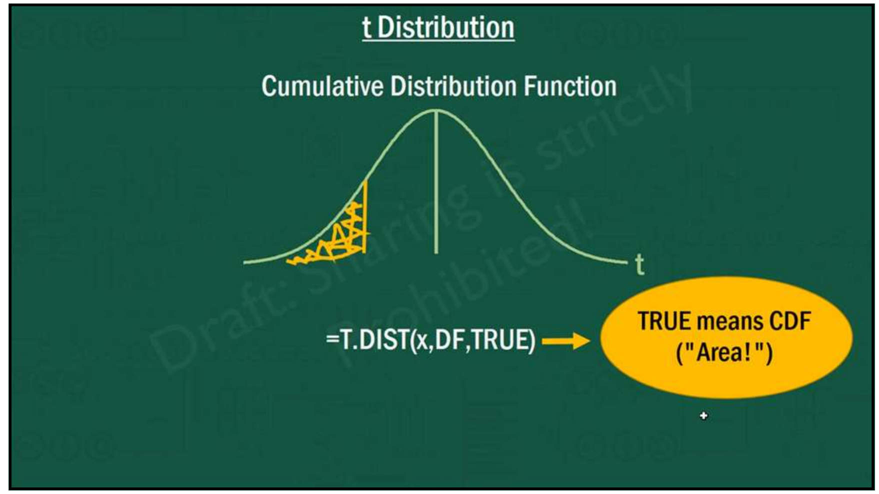

t Distribution
Probability Distribution Function
t
$PDF = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\left(1+\frac{x^2}{n-1}\right)^{-\frac{n}{2}}$

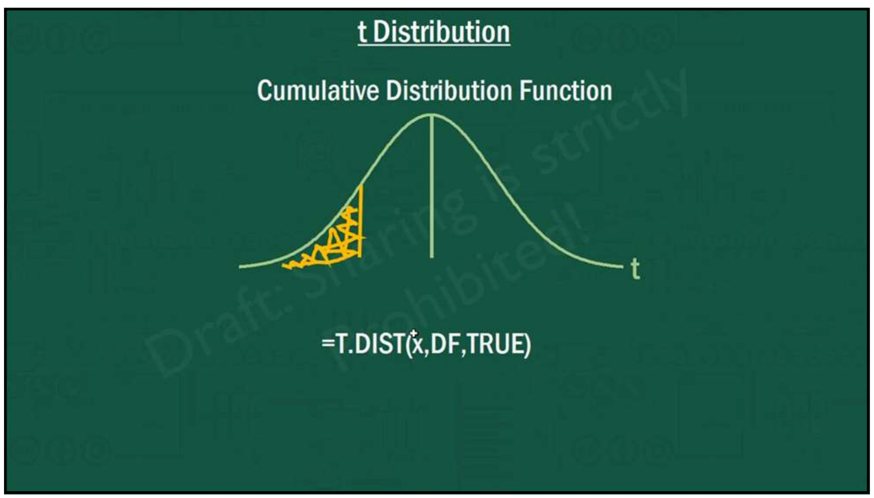

t Distribution
Probability Distribution Function
t
=T.DIST(x,DF,FALSE)
FALSE means PDF
("Height!")

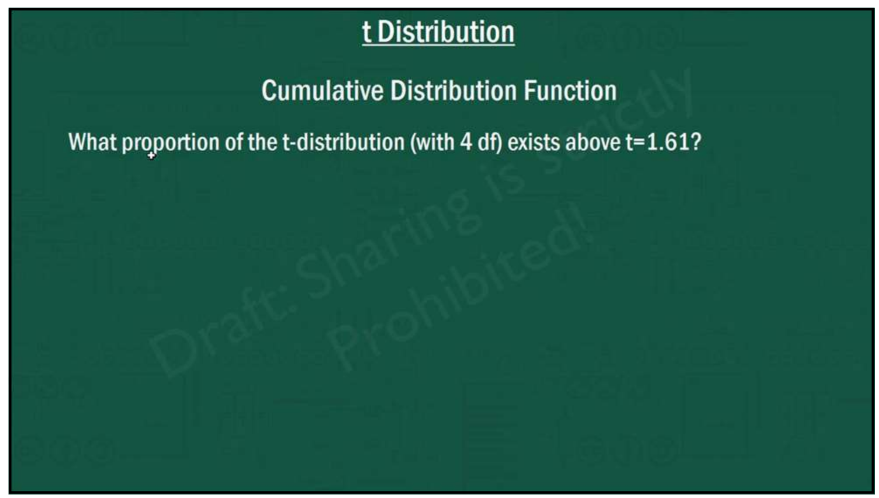

t Distribution
Probability Distribution Function
t
=T.DIST(x,DF,FALSE)

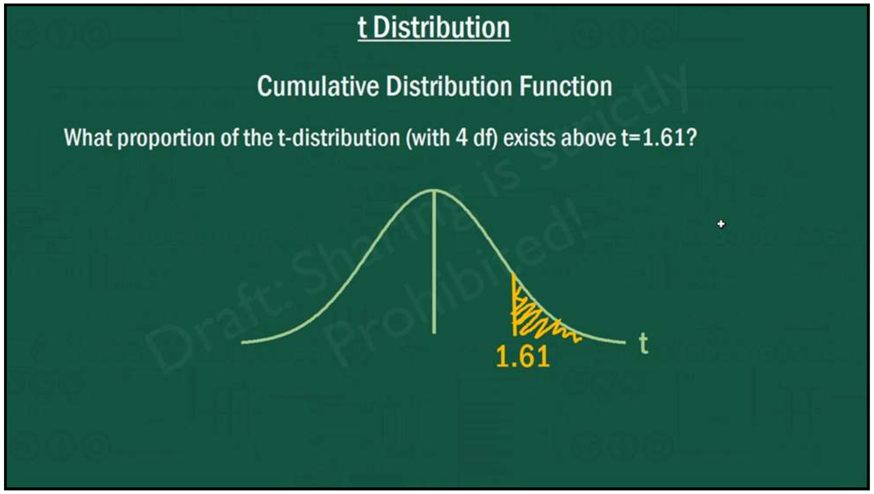

t Distribution
Cumulative Distribution Function
t

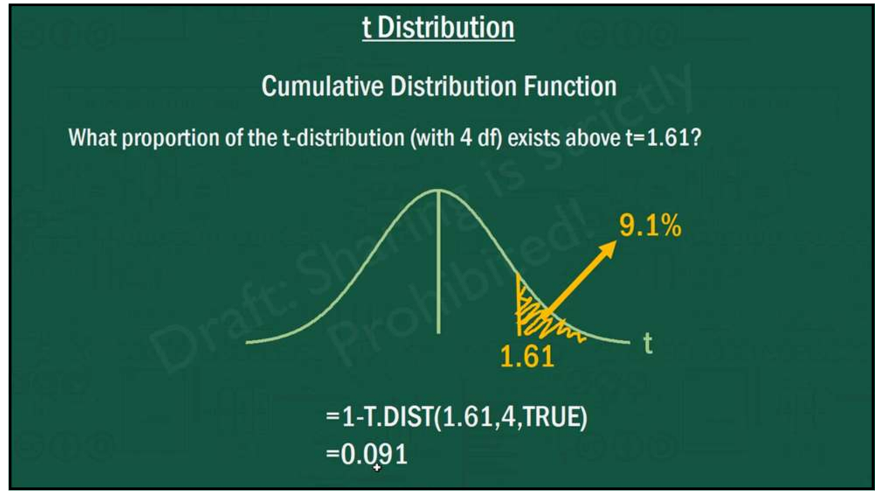

t Distribution
Cumulative Distribution Function
t

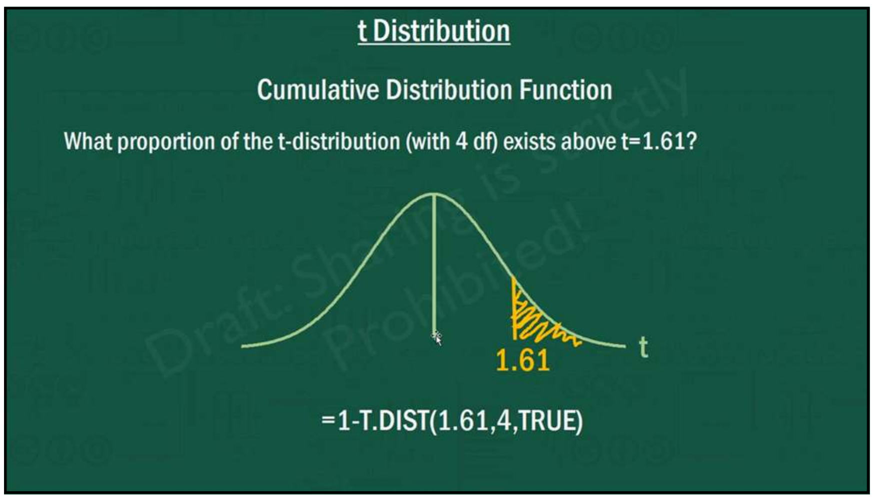

t Distribution
Cumulative Distribution Function
t
=T.DIST(x,DF,TRUE)
TRUE means CDF
("Area!")

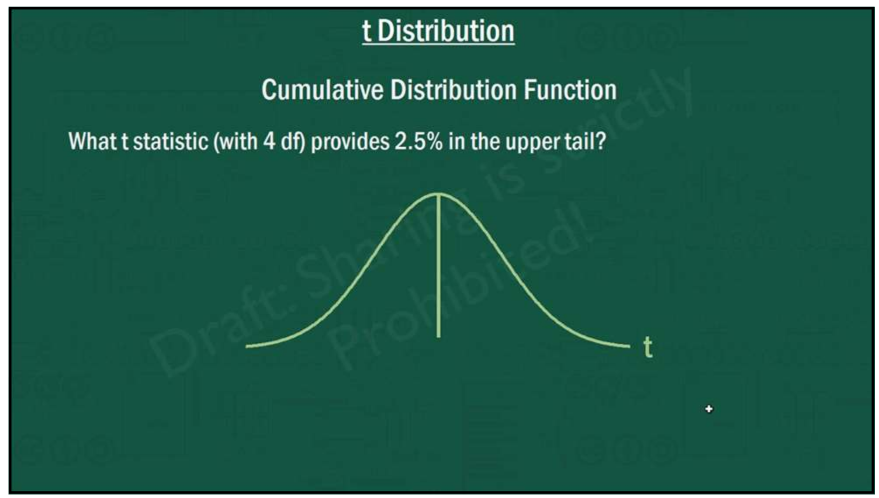

t Distribution
Cumulative Distribution Function
t
=T.DIST(x,DF,TRUE)

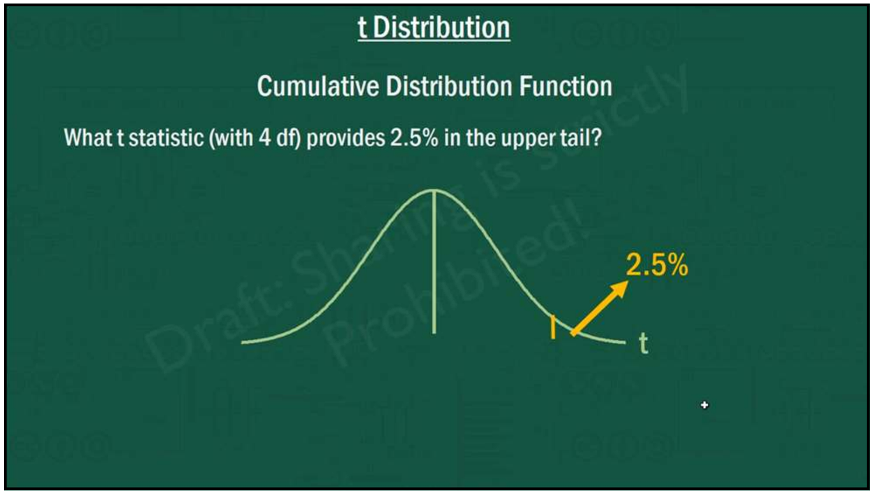

t Distribution
Cumulative Distribution Function
What proportion of the t-distribution (with 4 df) exists above t=1.61?

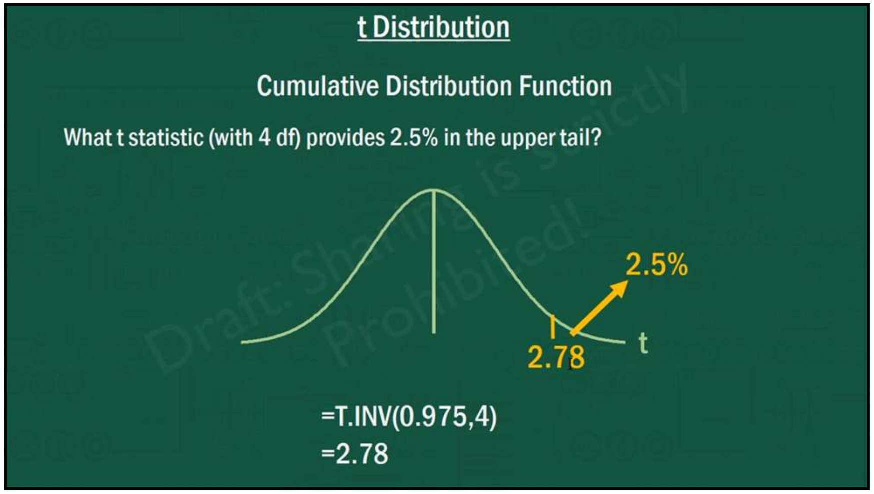

t Distribution
Cumulative Distribution Function
What proportion of the t-distribution (with 4 df) exists above t=1.61?
1.61
t

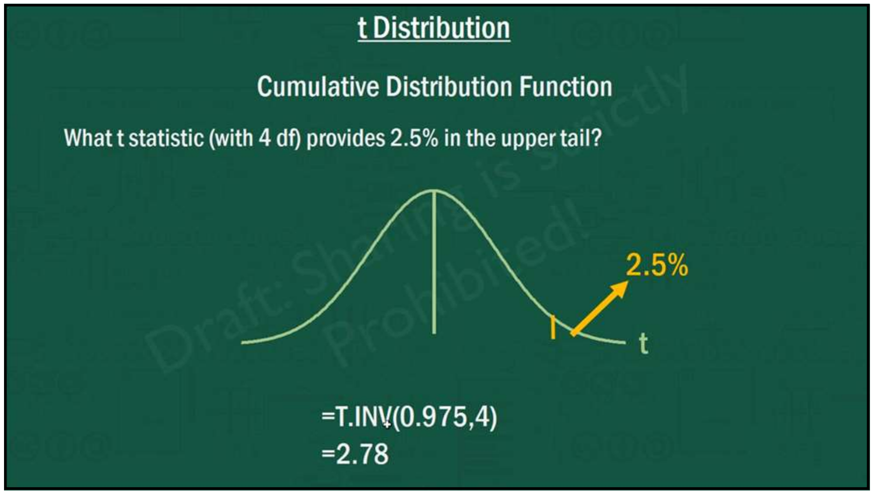

t Distribution
Cumulative Distribution Function
What proportion of the t-distribution (with 4 df) exists above t=1.61?
9.1%
1.61
t
=1-T.DIST(1.61,4,TRUE)
=0.091

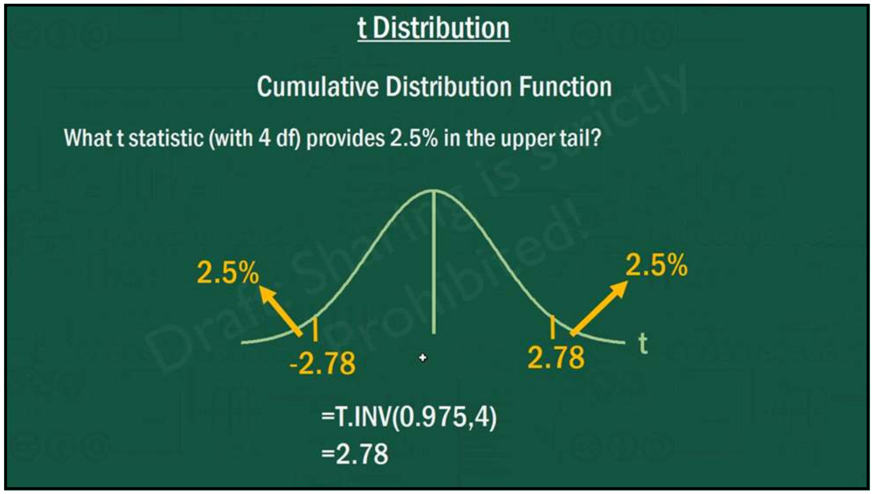

t Distribution
Cumulative Distribution Function
What proportion of the t-distribution (with 4 df) exists above t=1.61?
1.61
t
=1-T.DIST(1.61,4,TRUE)


t Distribution
Cumulative Distribution Function
What t statistic (with 4 df) provides 2.5% in the upper tail?
t


t Distribution
Cumulative Distribution Function
What t statistic (with 4 df) provides 2.5% in the upper tail?
2.5%
t


t Distribution
Cumulative Distribution Function
What t statistic (with 4 df) provides 2.5% in the upper tail?
2.5%
t
2.78
=T.INV(0.975,4)
=2.78

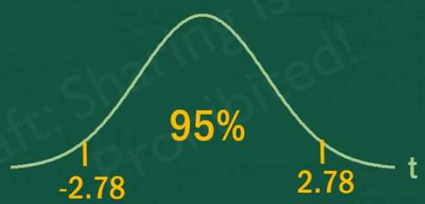

t Distribution
Cumulative Distribution Function
What t statistic (with 4 df) provides 2.5% in the upper tail?
2.5%
t
=T.INV(0.975,4)
=2.78


t Distribution
Cumulative Distribution Function
What t statistic (with 4 df) provides 2.5% in the upper tail?
2.5%
2.5%
t
-2.78
2.78
=T.INV(0.975,4)
=2.78

t Distribution

Cumulative Distribution Function

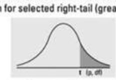
What t statistic (with 4 df) provides 2.5% in the upper tail?



=T.INV(0.975,4)
=2.78

THANK YOU!

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454