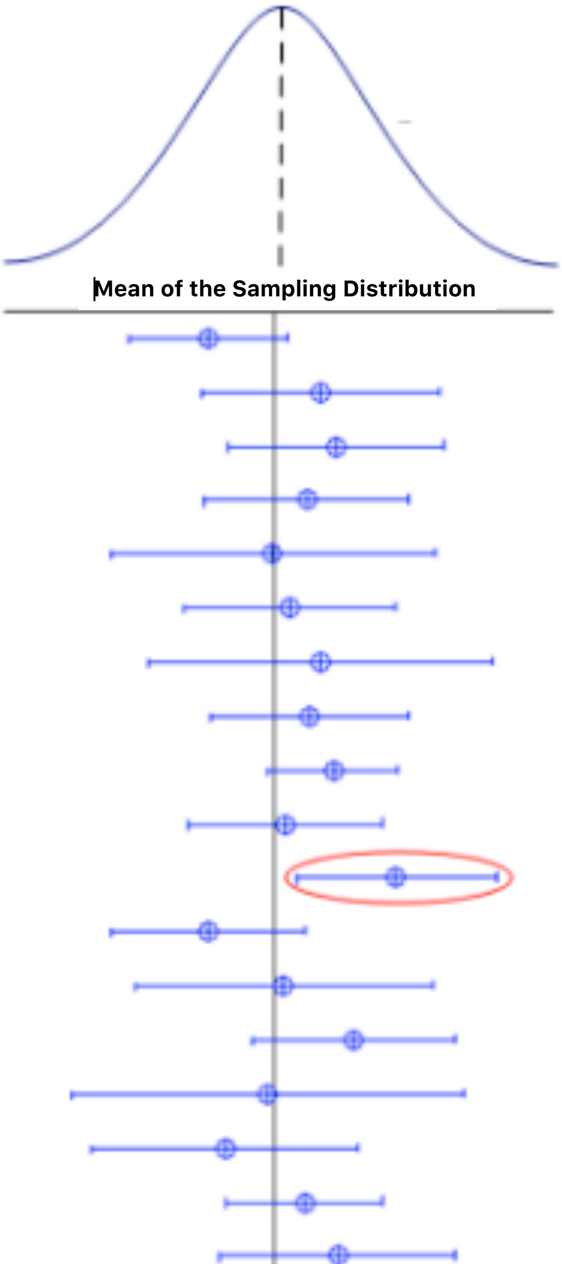
Point Estimation



mth_150

Data Distributions vs Probability Distributions

- Uniform
- Binomial
- Poisson
- Geometric
- Hypergeometric
- Multinomial

- Normal
- Standard Normal
- Exponential
- Gamma
- Beta
- Erlang

⋮

Sampling Distributions:

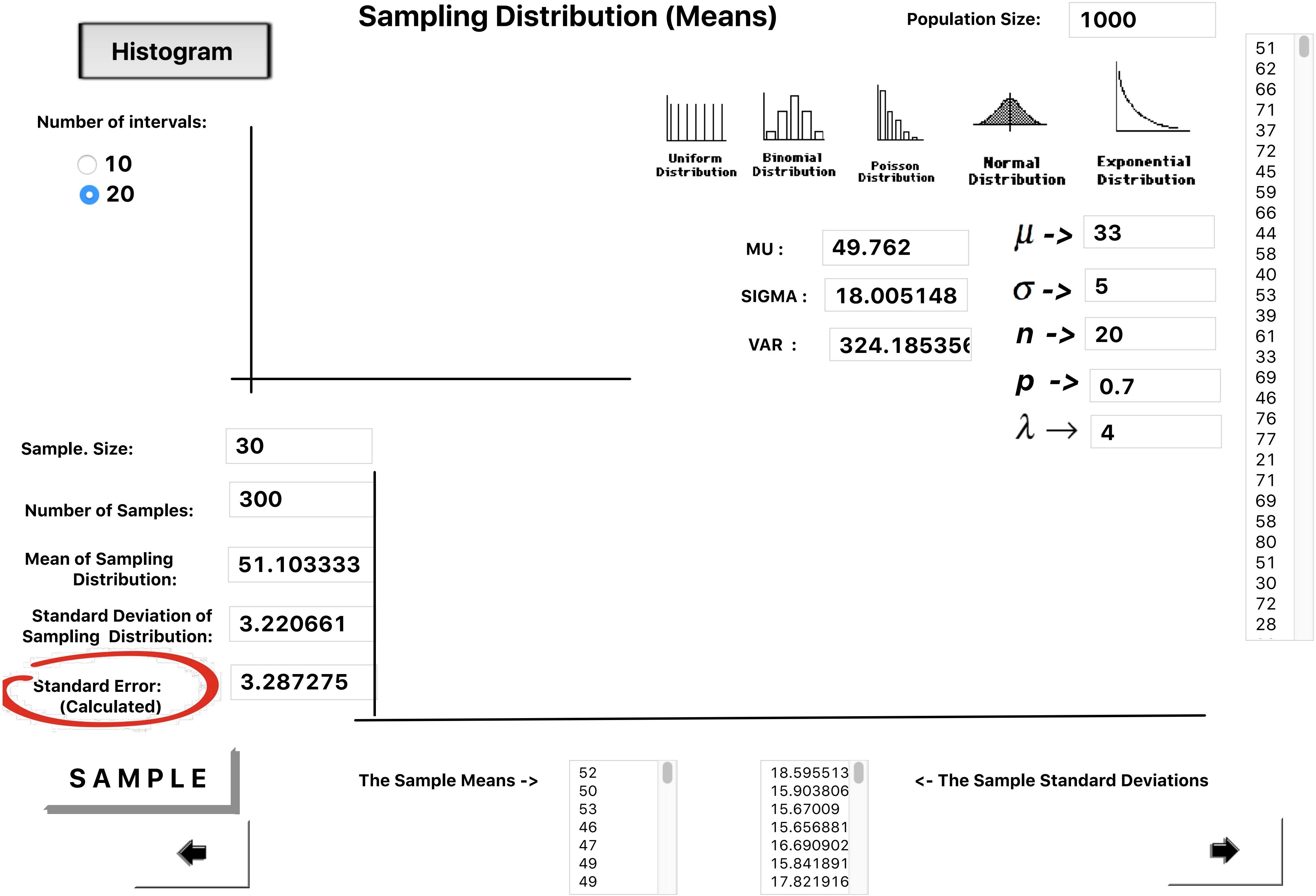
- means
- variances
- standard deviations
- medians
- ranges
- ratio of variances

⋮

Contrived Distributions:

- Student's t
- Chi-Square
- Snedecor - Fisher F

⋮



Sampling Distribution (Means)

Population Size: 1000

Number of intervals:
○ 10
● 20

MU : 49.762
SIGMA : 18.005148
VAR : 324.185356

μ -> 33
σ -> 5
n -> 20
p -> 0.7
$\lambda \rightarrow$ 4

51
62
66
71
37
72
45
59
66
44
58
40
53
39
61
33
69
46
76
77
21
71
69
58
80
51
30
72
28

Sample. Size: 30
Number of Samples: 300
Mean of Sampling Distribution: 51.103333
Standard Deviation of Sampling Distribution: 3.220661
Standard Error: (Calculated) 3.287275

The Sample Means ->

52
50
53
46
47
49
49

18.595513
15.903806
15.67009
15.656881
16.690902
15.841891
17.821916

<- The Sample Standard Deviations

Sampling Distribution of the Means

The Central Limit Theorem:

The Sampling Distribution (of the Means) -
- is normally distributed
- its mean is approximately equal to the mean of the population
- its standard deviiation is **not** equal to the standard deviation of the population, **but** is equal to the "standard error"

$$\overline{x} \approx \mu_{\overline{x}}$$

$$N\left(\overline{x}, \sigma^{2}_{\overline{x}}\right) \text{ where}$$

$$\overline{x} \approx \mu \text{ and } \boxed{\sigma_{\overline{x}} \approx \frac{\sigma}{\sqrt{n}}}$$

⬅ ➡

Confidence Interval of the mean
(Inferencing process)

A study is conducted concerning the blood pressure of 60 year old women with glaucoma. In the study 200 60-year old women with glaucoma are randomly selected and the sample mean systolic blood pressure is 140 mm Hg. The standard deviation of the blood pressure the population is known to be 20 mm Hg. Can you be confident that th true mean systolic blood pressure among the population of 60 year old women with glaucoma is 140 mm Hg ? **How confident can you be?**

A 95% confidence interval means that

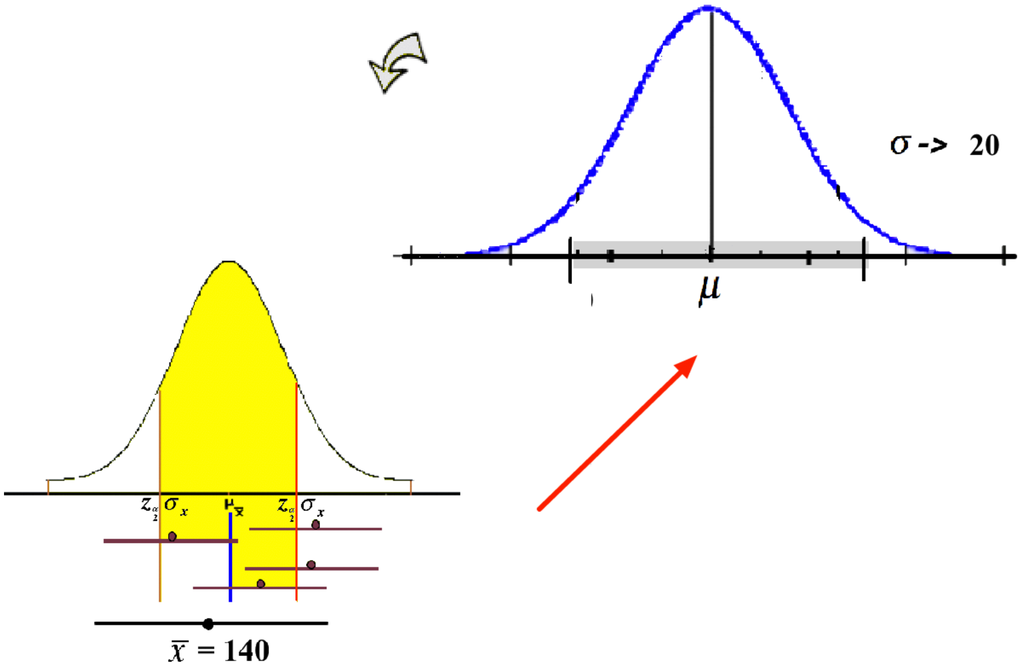
we can be 95% certain that

- the sample mean, $\bar{x}$, will lie within the margin of error of the true value of the population mean

OR that

- approximately 95 of 100 confidence intervals will contain the true population mean

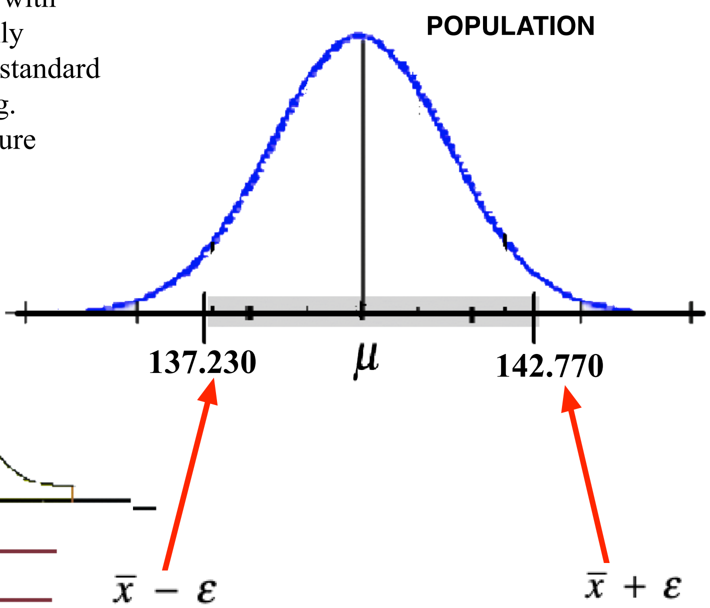
Confidence Interval of the mean
(Inferencing process)



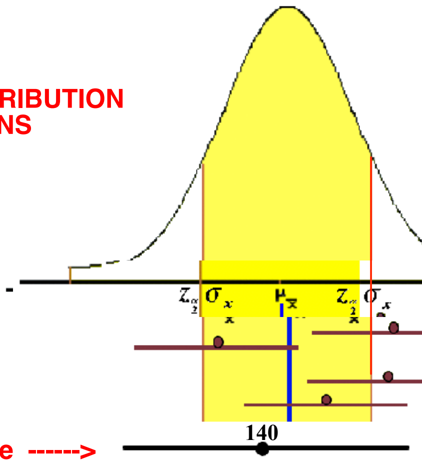
A study is conducted concerning the blood pressure of 60 year old women with glaucoma. In the study 200 60-year old women with glaucoma are randomly selected and the sample mean systolic blood pressure is 140 mm Hg. The standard deviation of the blood pressure of the population is known to be 20 mm Hg. **Calculate a 95% confidence interval** for the true mean systolic blood pressure among the population of 60 year old women with glaucoma.

Confidence Interval of the mean
(Inferencing process)

A study is conducted concerning the blood pressure of 60 year old women with glaucoma. In the study 200 60-year old women with glaucoma are randomly selected and the sample mean systolic blood pressure is 140 mm Hg. The standard deviation of the blood pressure of the population is known to be 20 mm Hg. Calculate a 95% confidence interval for the true mean systolic blood pressure among the population of 60 year old women with glaucoma.



SAMPLING DISTRIBUTION OF MEANS



Your sample ------>

$$\overline{x} \approx \mu_{\overline{x}} \approx \mu$$

$\overline{x} =$ 140

$1 - \alpha =$ 0.95

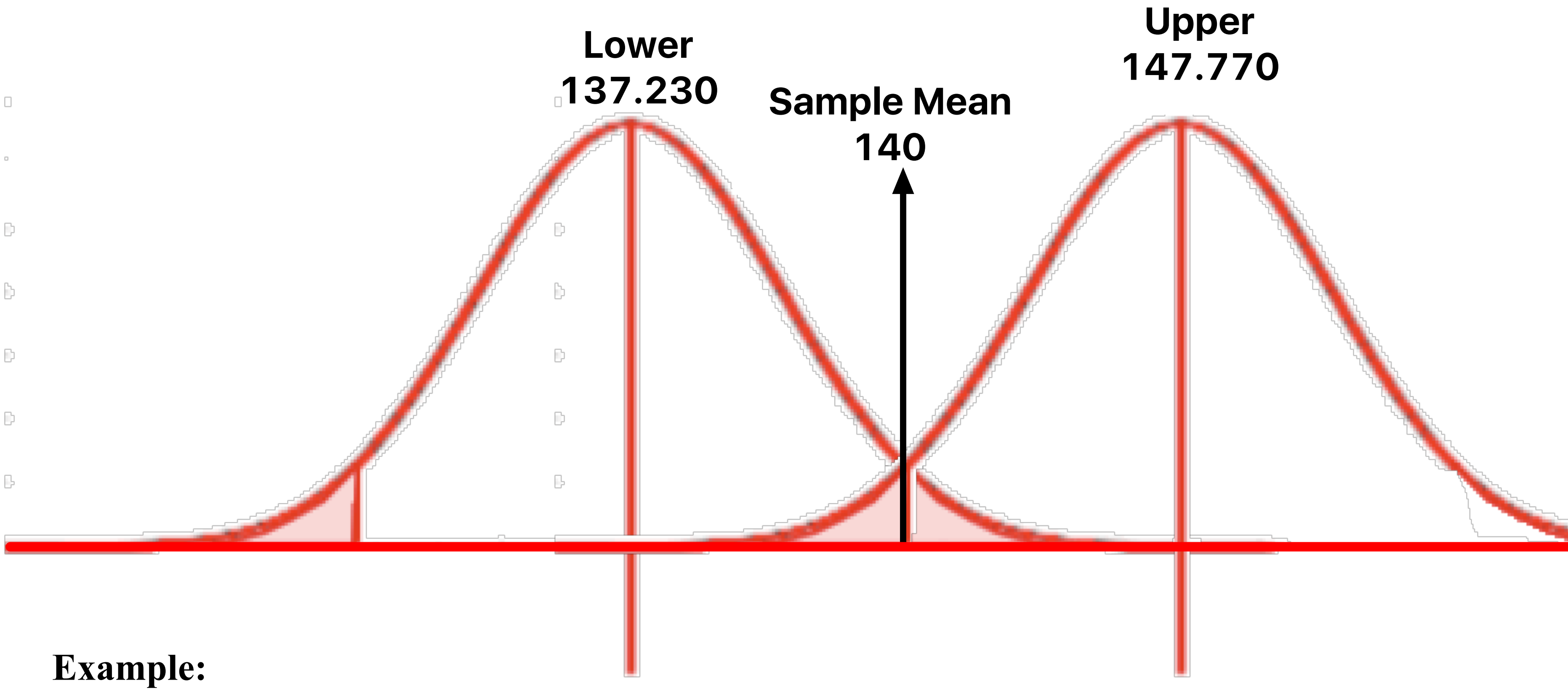
$n =$ 200

σ or $s =$ 20

$$\varepsilon = \left| z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right|$$

$137.230 < \mu < 142.770$ or

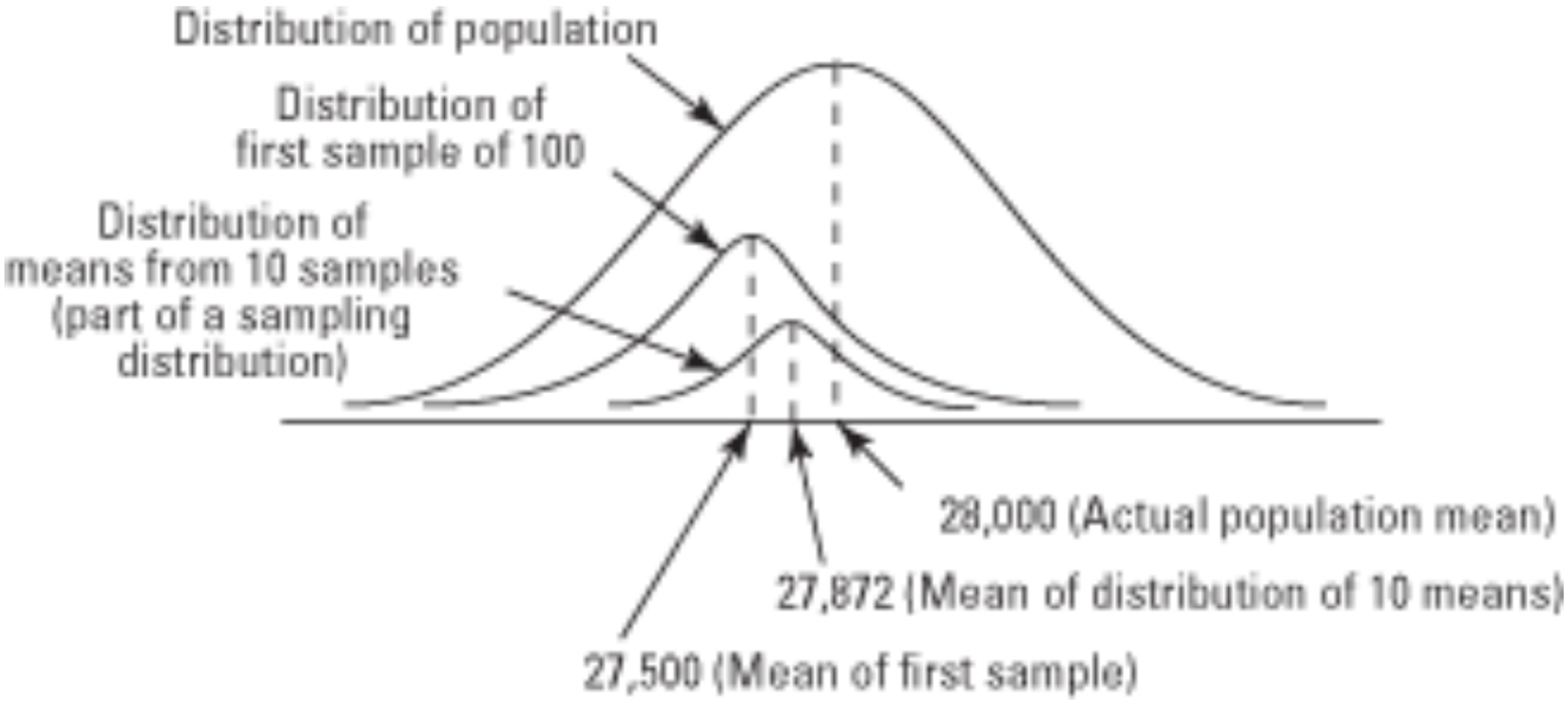
$\mu \in (137.230, 142.770)$



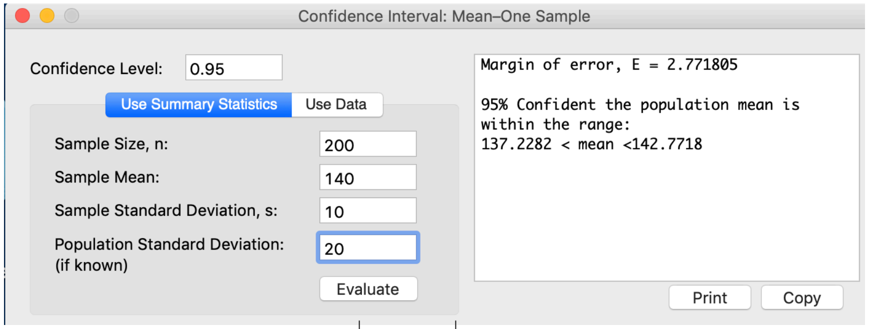
Example:

The distributions of the sample mean for a population mean of 137.230 — ends up with a tail p
of 2.5% above 140, and the distribution of the sample means with a population mean of 147.77
2.5% below the sample mean 0f 140.

Taking Samples



One Sample Estimation



Confidence Interval: Mean–One Sample

Confidence Level: 0.95

Sample Size, n: 200

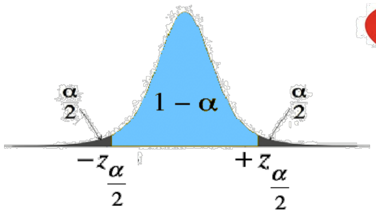
Sample Mean: 140

Sample Standard Deviation, s: 10

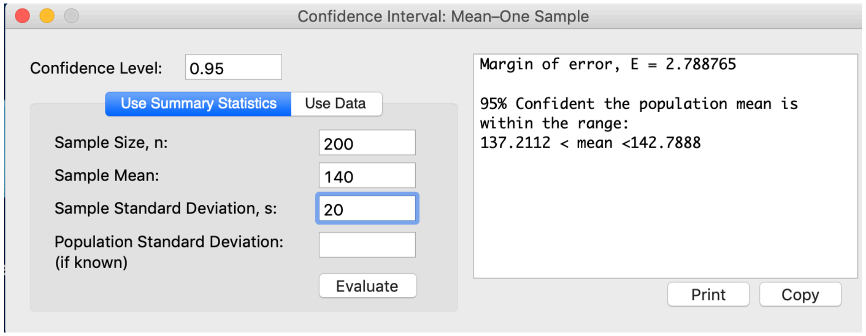
Population Standard Deviation: (if known) 20

```
Margin of error, E = 2.771805

95% Confident the population mean is
within the range:
137.2282 < mean <142.7718
```

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{x}}$$



Uses Standard Normal



Confidence Interval: Mean–One Sample

Confidence Level: 0.95

Sample Size, n: 200

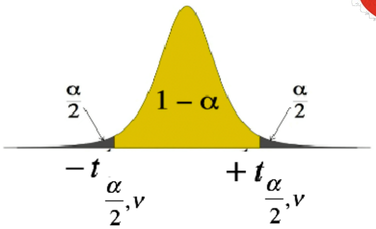
Sample Mean: 140

Sample Standard Deviation, s: 20

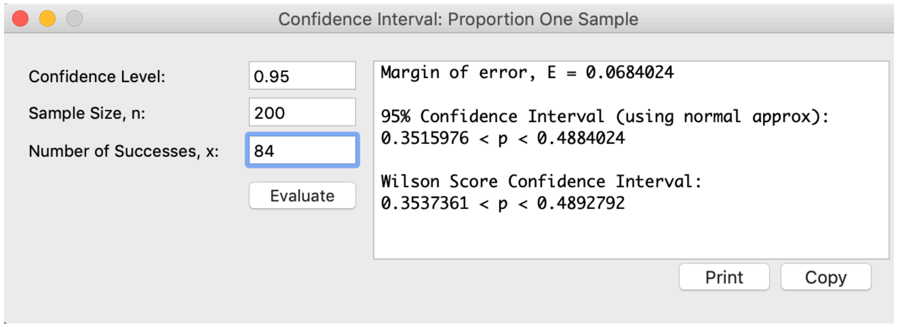
Population Standard Deviation: (if known)

```
Margin of error, E = 2.788765

95% Confident the population mean is
within the range:
137.2112 < mean <142.7888
```

$$\bar{x} \pm t_{\frac{\alpha}{2},v} \cdot \frac{s}{\sqrt{n}}$$



Uses Student's-*t*



Confidence Interval: Proportion One Sample

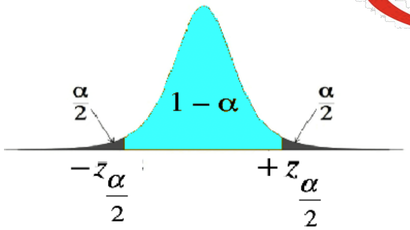
Confidence Level: 0.95

Sample Size, n: 200

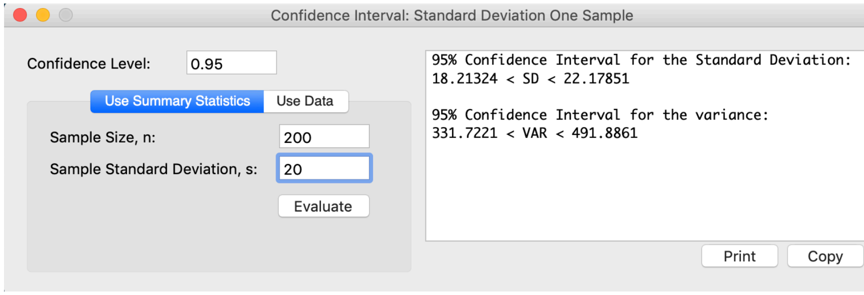
Number of Successes, x: 84

```
Margin of error, E = 0.0684024

95% Confidence Interval (using normal approx):
0.3515976 < p < 0.4884024

Wilson Score Confidence Interval:
0.3537361 < p < 0.4892792
```

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



Uses Standard Normal



Confidence Interval: Standard Deviation One Sample

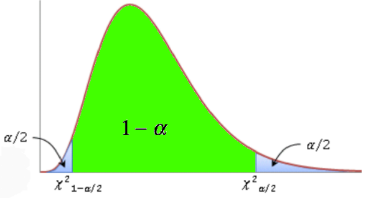
Confidence Level: 0.95

Sample Size, n: 200

Sample Standard Deviation, s: 20

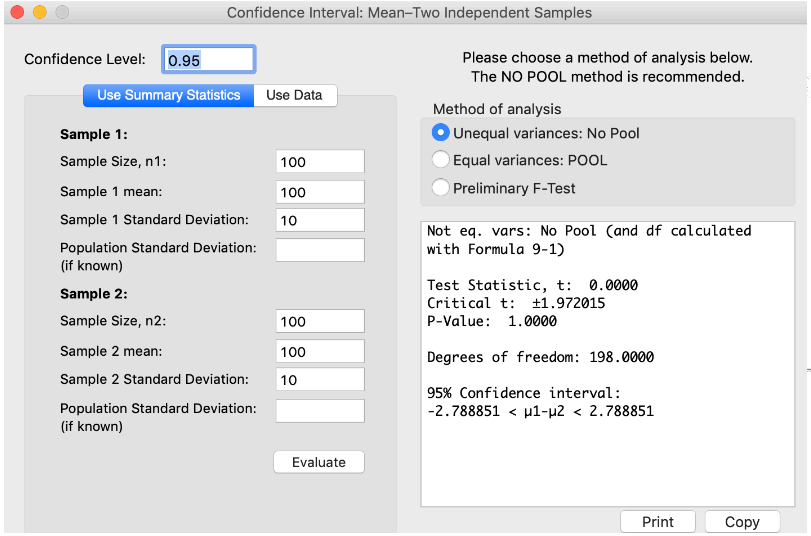
```
95% Confidence Interval for the Standard Deviation:
18.21324 < SD < 22.17851

95% Confidence Interval for the variance:
331.7221 < VAR < 491.8861
```

$$\frac{(n-1)s^2}{\chi^2_{v,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{v,1-\frac{\alpha}{2}}}$$



Uses Chi-Square

Two Sample Estimation



Confidence Level: 0.95

Please choose a method of analysis below.
The NO POOL method is recommended.

Sample 1:

Sample Size, n1: 100

Sample 1 mean: 100

Sample 1 Standard Deviation: 10

Population Standard Deviation: (if known)

Sample 2:

Sample Size, n2: 100

Sample 2 mean: 100

Sample 2 Standard Deviation: 10

Population Standard Deviation: (if known)

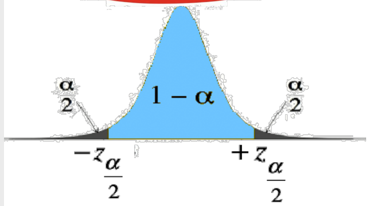
Method of analysis

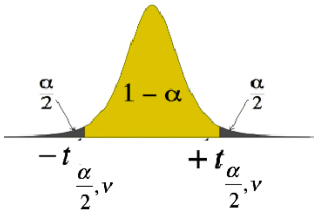
● Unequal variances: No Pool
○ Equal variances: POOL
○ Preliminary F-Test

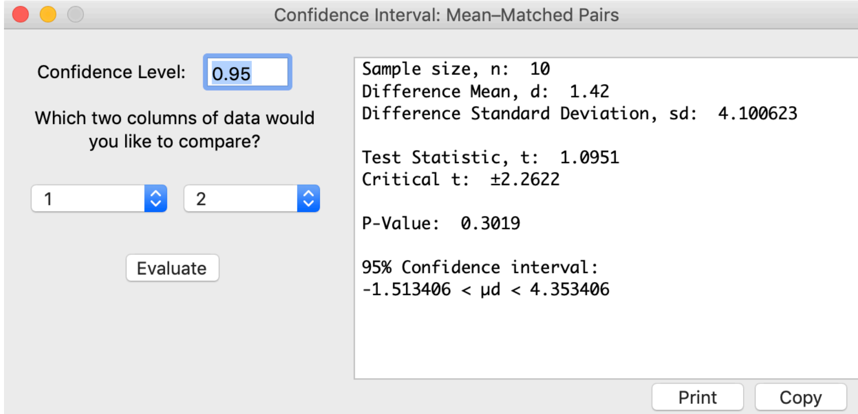
```
Not eq. vars: No Pool (and df calculated
with Formula 9-1)

Test Statistic, t:  0.0000
Critical t:  ±1.972015
P-Value:  1.0000

Degrees of freedom: 198.0000

95% Confidence interval:
-2.788851 < μ1-μ2 < 2.788851
```

Uses Standard Normal

Uses Student's-*t*





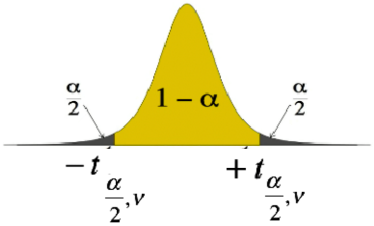


Confidence Level: 0.95

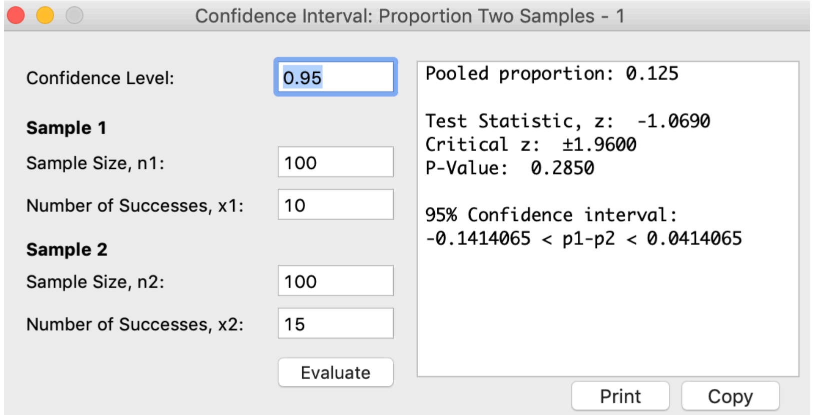
Which two columns of data would you like to compare?

1 2

```
Sample size, n:  10
Difference Mean, d:  1.42
Difference Standard Deviation, sd:  4.100623

Test Statistic, t:  1.0951
Critical t:  ±2.2622

P-Value:  0.3019

95% Confidence interval:
-1.513406 < μd < 4.353406
```



Uses Stu



Confidence Level: 0.95

Sample 1

Sample Size, n1: 100

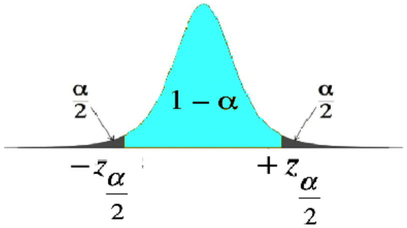
Number of Successes, x1: 10

Sample 2

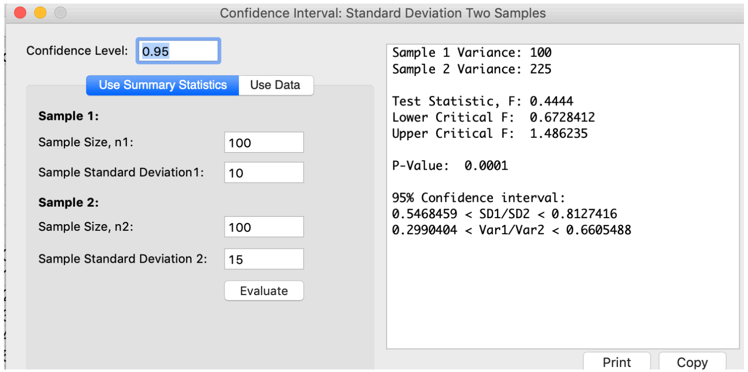
Sample Size, n2: 100

Number of Successes, x2: 15

```
Pooled proportion: 0.125

Test Statistic, z:  -1.0690
Critical z:  ±1.9600
P-Value:  0.2850

95% Confidence interval:
-0.1414065 < p1-p2 < 0.0414065
```



Uses Standard Normal



Confidence Level: 0.95

Sample 1:

Sample Size, n1: 100

Sample Standard Deviation1: 10

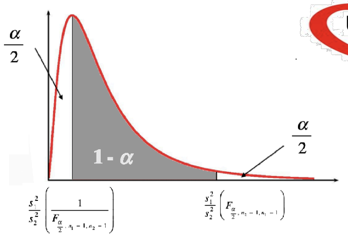
Sample 2:

Sample Size, n2: 100

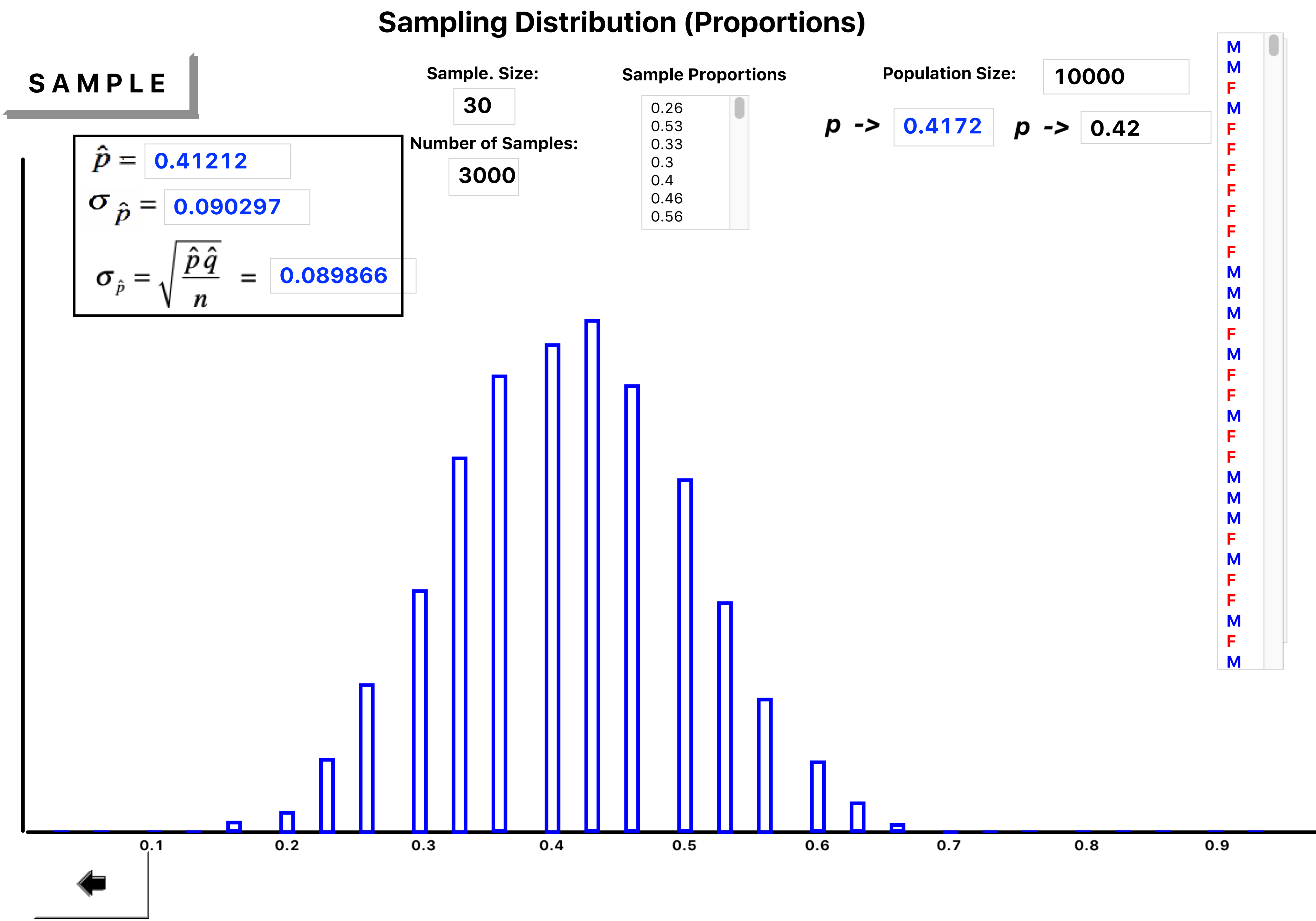
Sample Standard Deviation 2: 15

```
Sample 1 Variance: 100
Sample 2 Variance: 225

Test Statistic, F: 0.4444
Lower Critical F:  0.6728412
Upper Critical F:  1.486235

P-Value:  0.0001

95% Confidence interval:
0.5468459 < SD1/SD2 < 0.8127416
0.2990404 < Var1/Var2 < 0.6605488
```

$$\frac{s_1^2}{s_2^2}\left(\frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\left(F_{\frac{\alpha}{2},n_2-1,n_1-1}\right)$$



Uses Fisher-Snedeco



Sampling Distribution (Proportions)

Sample. Size: 30

Number of Samples: 3000

Sample Proportions

0.26
0.53
0.33
0.3
0.4
0.46
0.56

Population Size: 10000

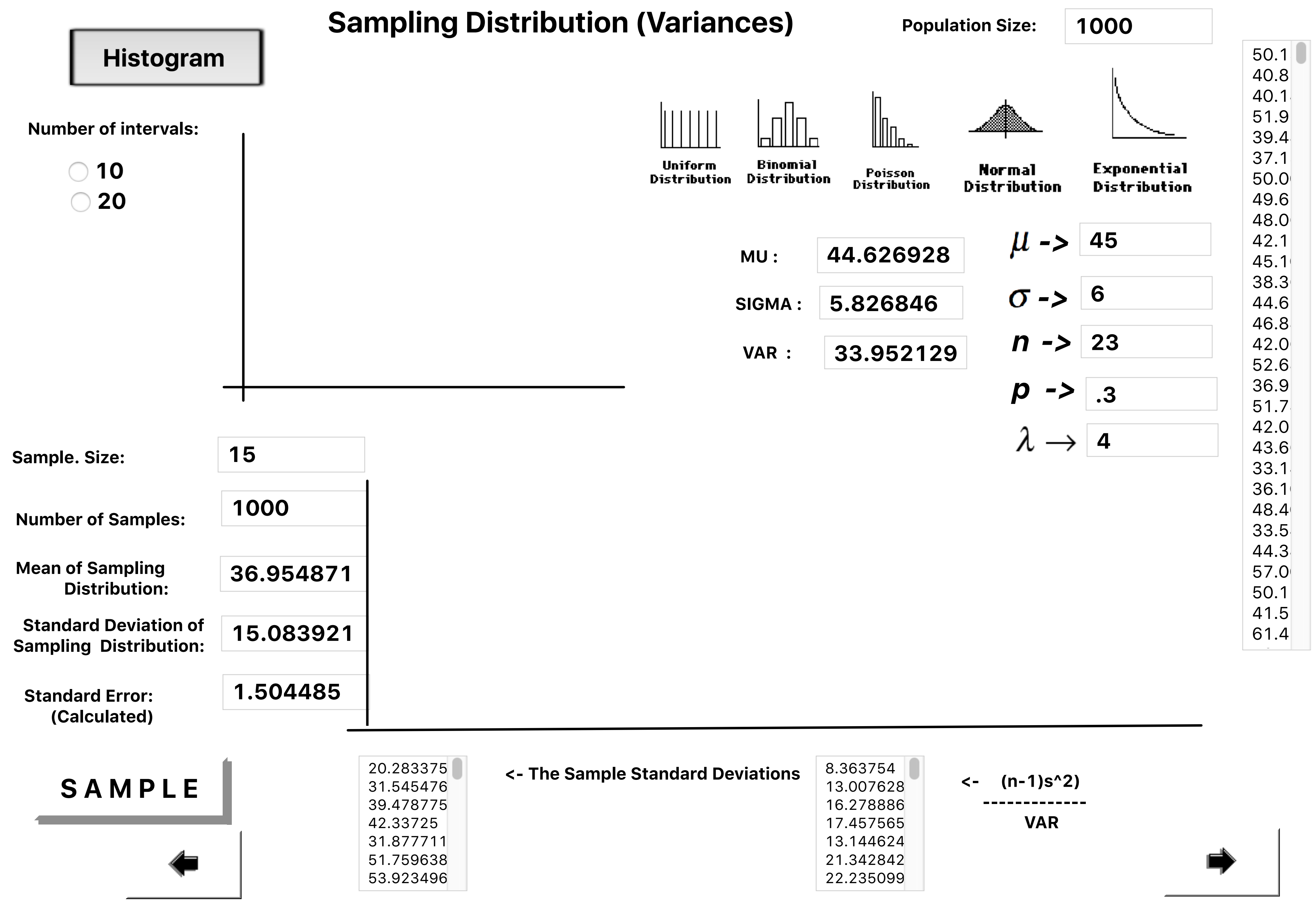
p -> 0.4172

p -> 0.42

$\hat{p} =$ 0.41212

$\sigma_{\hat{p}} =$ 0.090297

$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} =$ 0.089866



Sampling Distribution (Variances)

Population Size: 1000

Number of intervals:

○ 10
○ 20

MU : 44.626928
SIGMA : 5.826846
VAR : 33.952129

μ -> 45
σ -> 6
n -> 23
p -> .3
λ → 4

Sample. Size: 15

Number of Samples: 1000

Mean of Sampling Distribution: 36.954871

Standard Deviation of Sampling Distribution: 15.083921

Standard Error: (Calculated) 1.504485

50.1
40.8
40.1
51.9
39.4
37.1
50.0
49.6
48.0
42.1
45.1
38.3
44.6
46.8
42.0
52.6
36.9
51.7
42.0
43.6
33.1
36.1
48.4
33.5
44.3
57.0
50.1
41.5
61.4

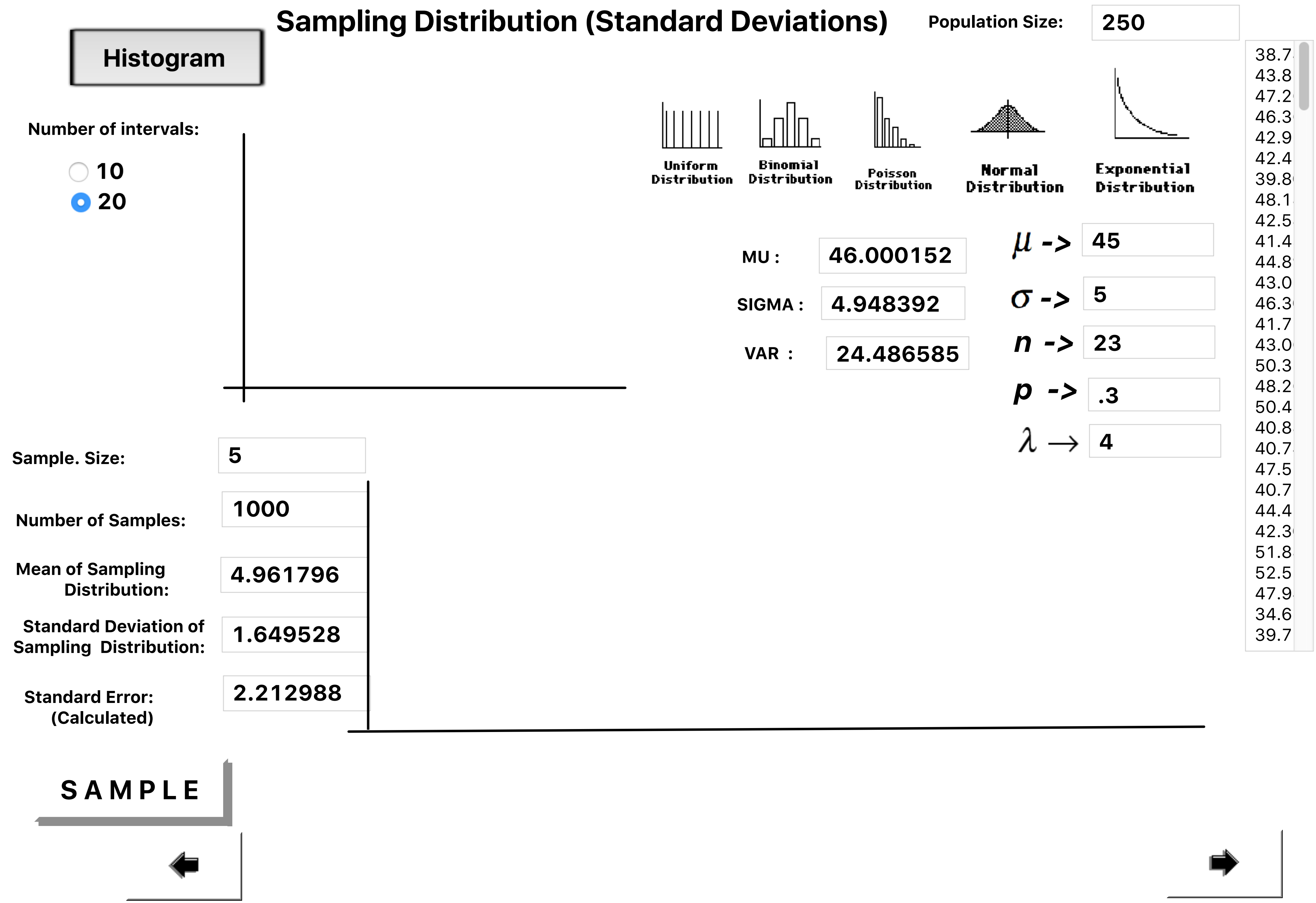
20.283375
31.545476
39.478775
42.33725
31.877711
51.759638
53.923496

<- The Sample Standard Deviations

8.363754
13.007628
16.278886
17.457565
13.144624
21.342842
22.235099

<- (n-1)s^2)

VAR



Sampling Distribution (Standard Deviations)

Population Size: 250

Number of intervals:

○ 10
● 20

MU : 46.000152

SIGMA : 4.948392

VAR : 24.486585

μ -> 45

σ -> 5

n -> 23

p -> .3

$\lambda \rightarrow$ 4

Sample. Size: 5

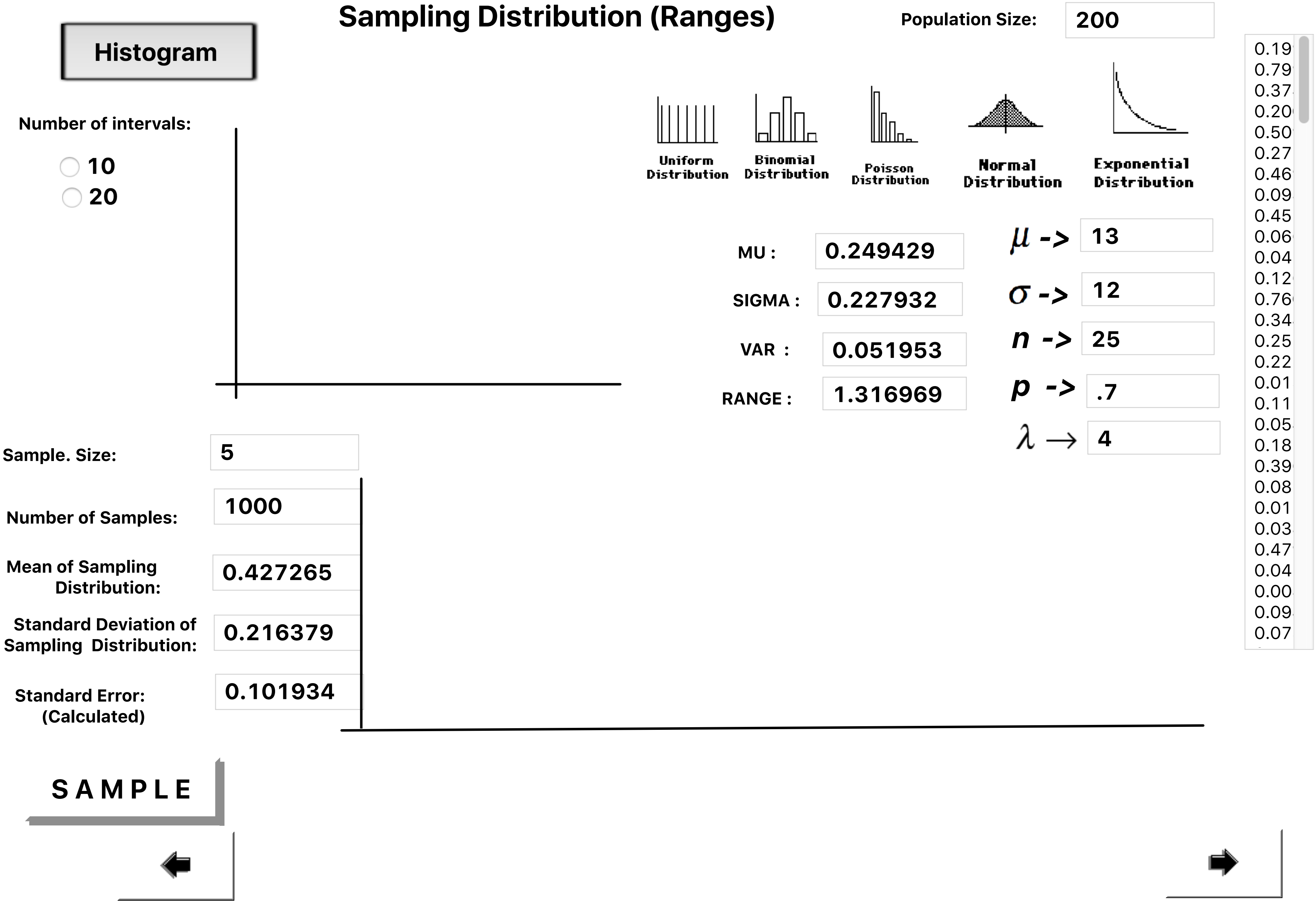
Number of Samples: 1000

Mean of Sampling Distribution: 4.961796

Standard Deviation of Sampling Distribution: 1.649528

Standard Error: (Calculated) 2.212988

38.7
43.8
47.2
46.3
42.9
42.4
39.8
48.1
42.5
41.4
44.8
43.0
46.3
41.7
43.0
50.3
48.2
50.4
40.8
40.7
47.5
40.7
44.4
42.3
51.8
52.5
47.9
34.6
39.7



Sampling Distribution (Ranges)

Population Size: 200

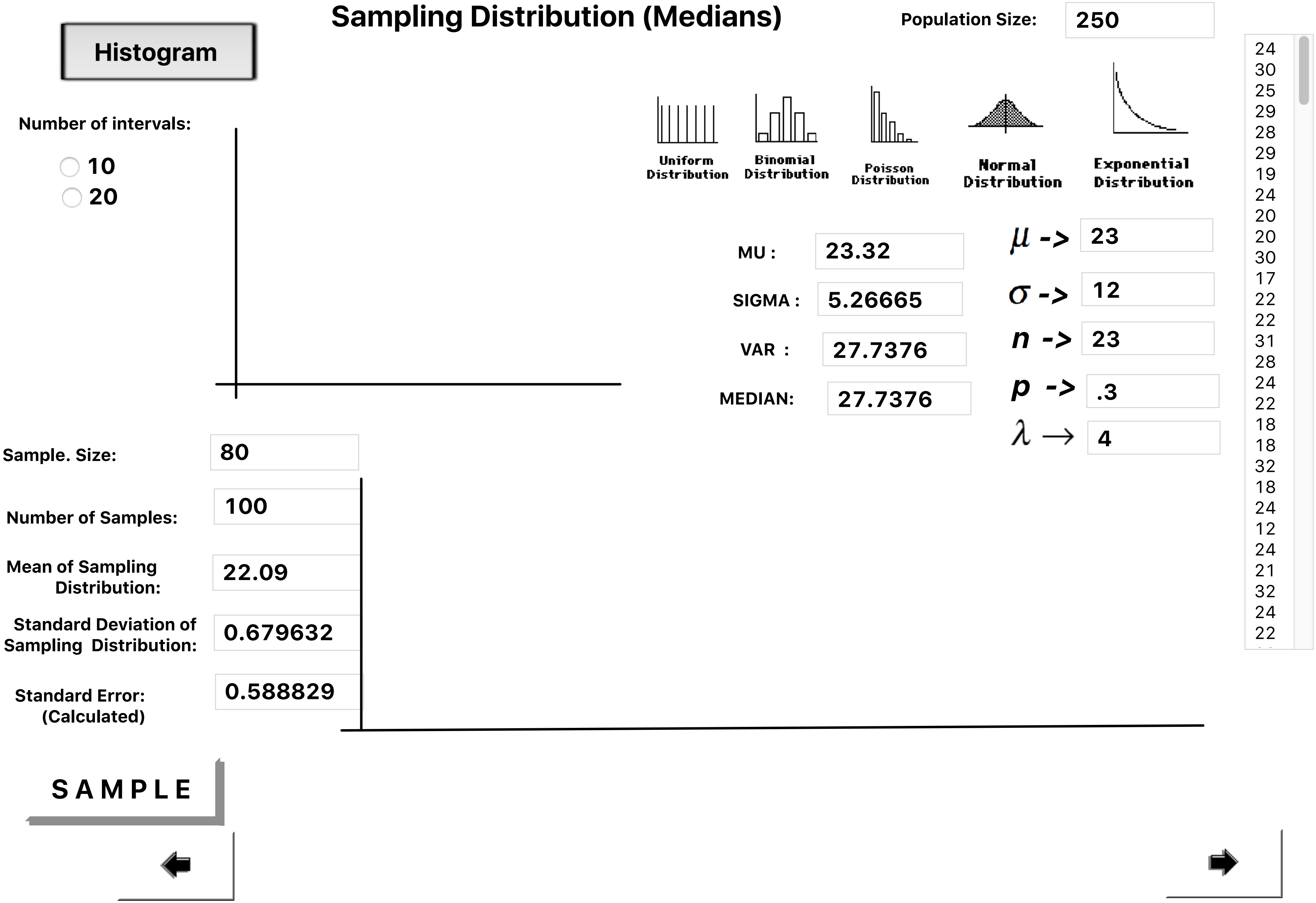
Number of intervals:
○ 10
○ 20

MU : 0.249429
SIGMA : 0.227932
VAR : 0.051953
RANGE : 1.316969

μ -> 13
σ -> 12
n -> 25
p -> .7
$\lambda \rightarrow$ 4

Sample. Size: 5
Number of Samples: 1000
Mean of Sampling Distribution: 0.427265
Standard Deviation of Sampling Distribution: 0.216379
Standard Error: (Calculated) 0.101934

```
0.19
0.79
0.37
0.20
0.50
0.27
0.46
0.09
0.45
0.06
0.04
0.12
0.76
0.34
0.25
0.22
0.01
0.11
0.05
0.18
0.39
0.08
0.01
0.03
0.47
0.04
0.00
0.09
0.07
```



Sampling Distribution (Medians)

Population Size: 250

Number of intervals:

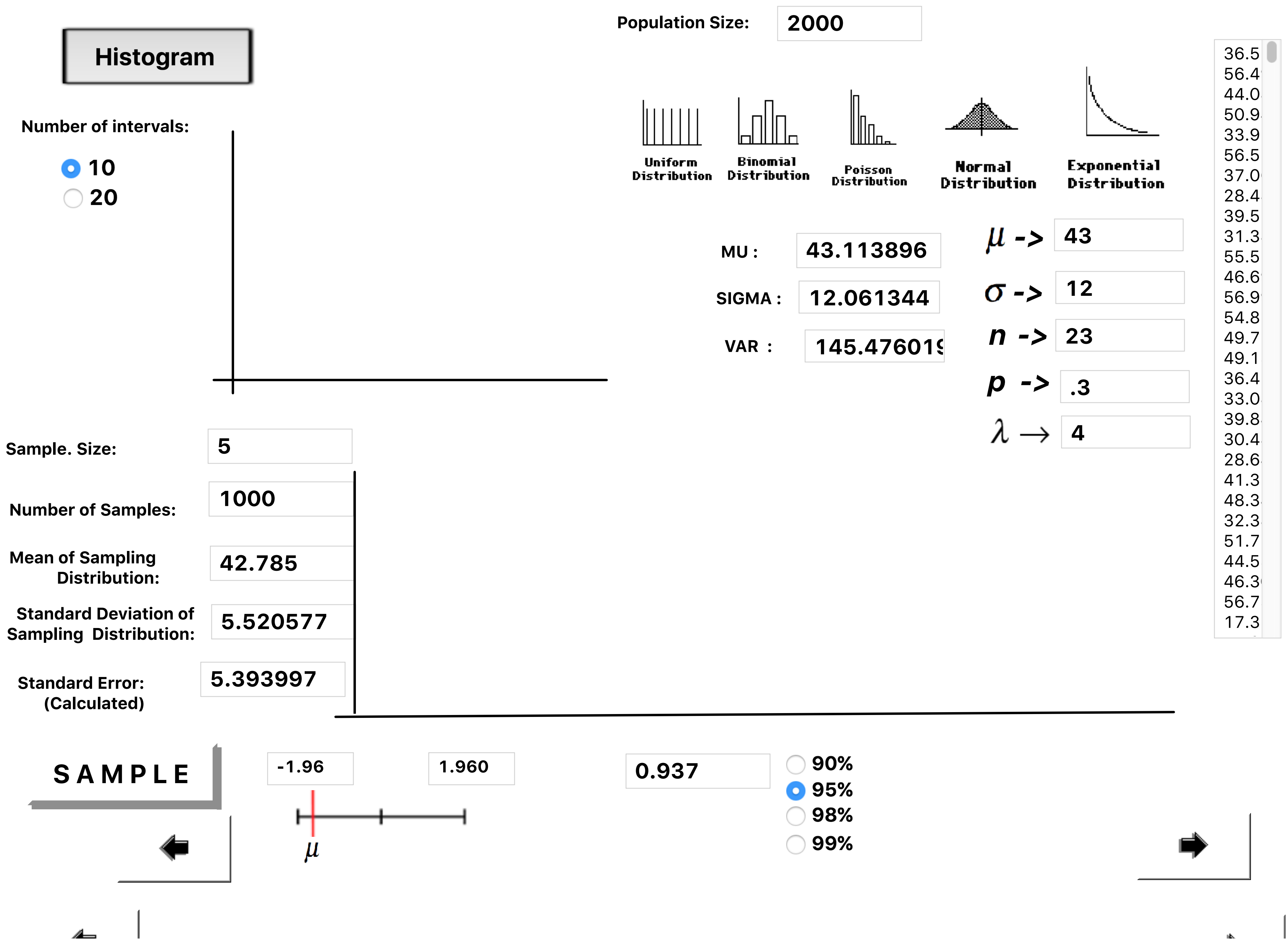
○ 10
○ 20

MU : 23.32
SIGMA : 5.26665
VAR : 27.7376
MEDIAN: 27.7376

μ -> 23
σ -> 12
n -> 23
p -> .3
$\lambda \rightarrow$ 4

Sample. Size: 80
Number of Samples: 100
Mean of Sampling Distribution: 22.09
Standard Deviation of Sampling Distribution: 0.679632
Standard Error: (Calculated) 0.588829

```
24
30
25
29
28
29
19
24
20
20
30
17
22
22
31
28
24
22
18
18
32
18
24
12
24
21
32
24
22
```



Population Size: 2000

Number of intervals:

● 10
○ 20

MU : 43.113896

SIGMA : 12.061344

VAR : 145.47601

μ -> 43

σ -> 12

n -> 23

p -> .3

$\lambda \rightarrow$ 4

36.5
56.4
44.0
50.9
33.9
56.5
37.0
28.4
39.5
31.3
55.5
46.6
56.9
54.8
49.7
49.1
36.4
33.0
39.8
30.4
28.6
41.3
48.3
32.3
51.7
44.5
46.3
56.7
17.3

Sample. Size: 5

Number of Samples: 1000

Mean of Sampling Distribution: 42.785

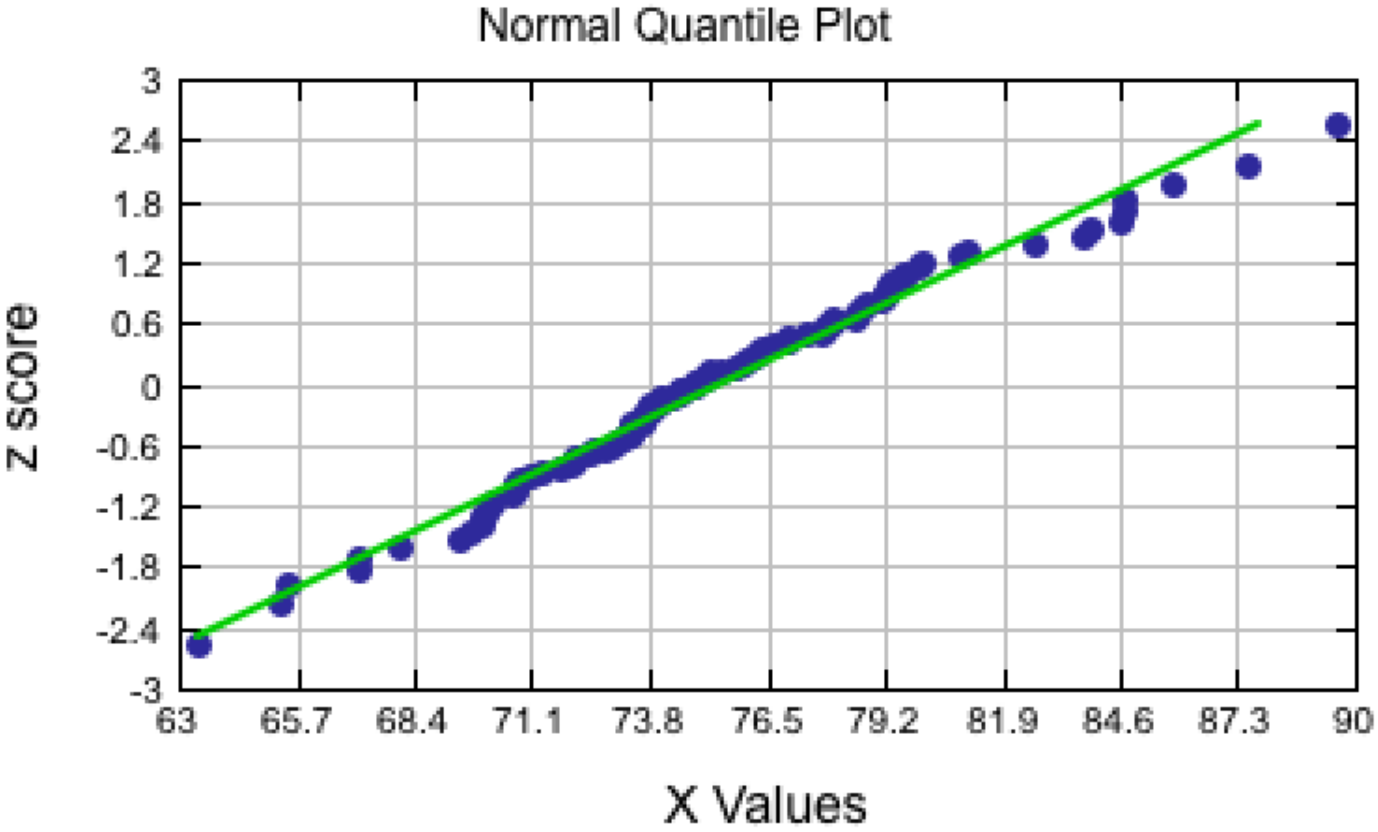
Standard Deviation of Sampling Distribution: 5.520577

Standard Error: (Calculated) 5.393997

-1.96 1.960 0.937

○ 90%
● 95%
○ 98%
○ 99%

Normality ?


Normal Quantile Plot
Z score
3
2.4
1.8
1.2
0.6
0
-0.6
-1.2
-1.8
-2.4
-3
63
65.7
68.4
71.1
73.8
76.5
79.2
81.9
84.6
87.3
90
X Values

Estimating the mean

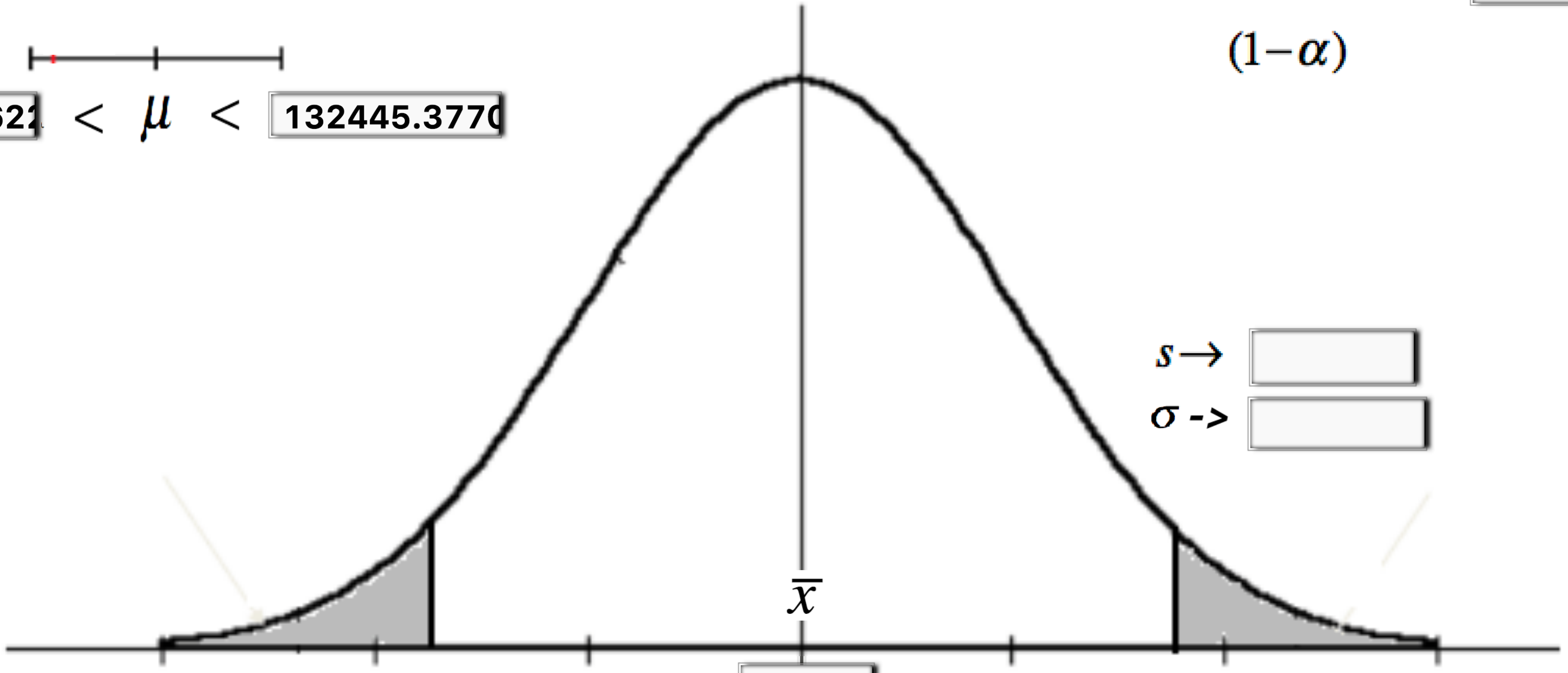
(using standard normal and Student's t)

n -> 50

confidence ↦ .90

$(1-\alpha)$

117554.62 < μ < 132445.377



$s \rightarrow$

σ ->

7588.900122

ME

-1.645211

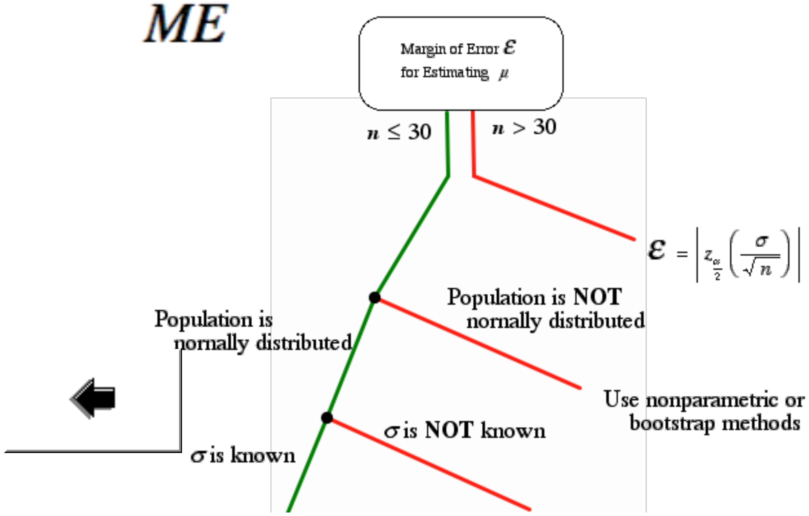
1.645211

☑ z ☐ t

$$\bar{x} - ME < \mu < \bar{x} + ME$$

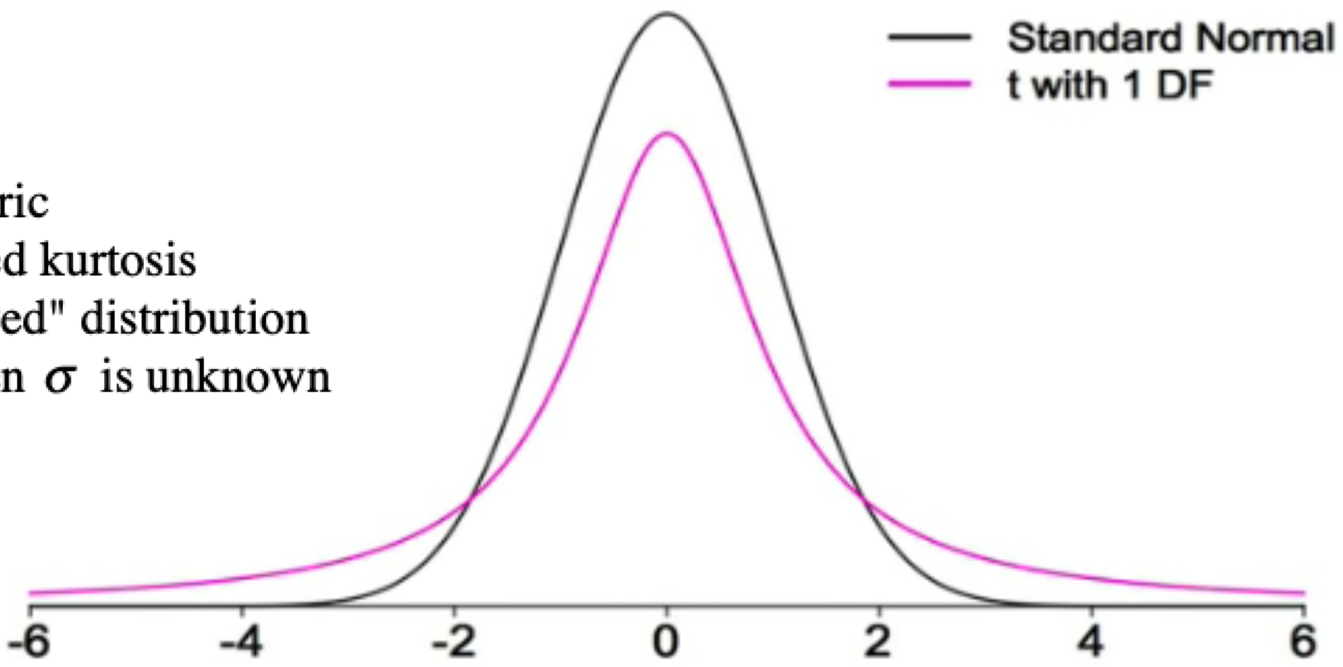
$$ME = \left| z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}} \right) \right|$$

$$ME = \left| t_{\nu,\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}} \right) \right|, \quad \nu = n - 1$$



Student's t Distribution

- $\mu = 0$
- symmetric
- increased kurtosis
- "contrived" distribution
- use when σ is unknown



$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

for $-\infty < t < \infty$

μ -> 0 for $\nu > 1$

$$\sqrt{\frac{\nu}{\nu - 2}}$$ for $\nu > 2$

⬅

➡

Estimating the mean

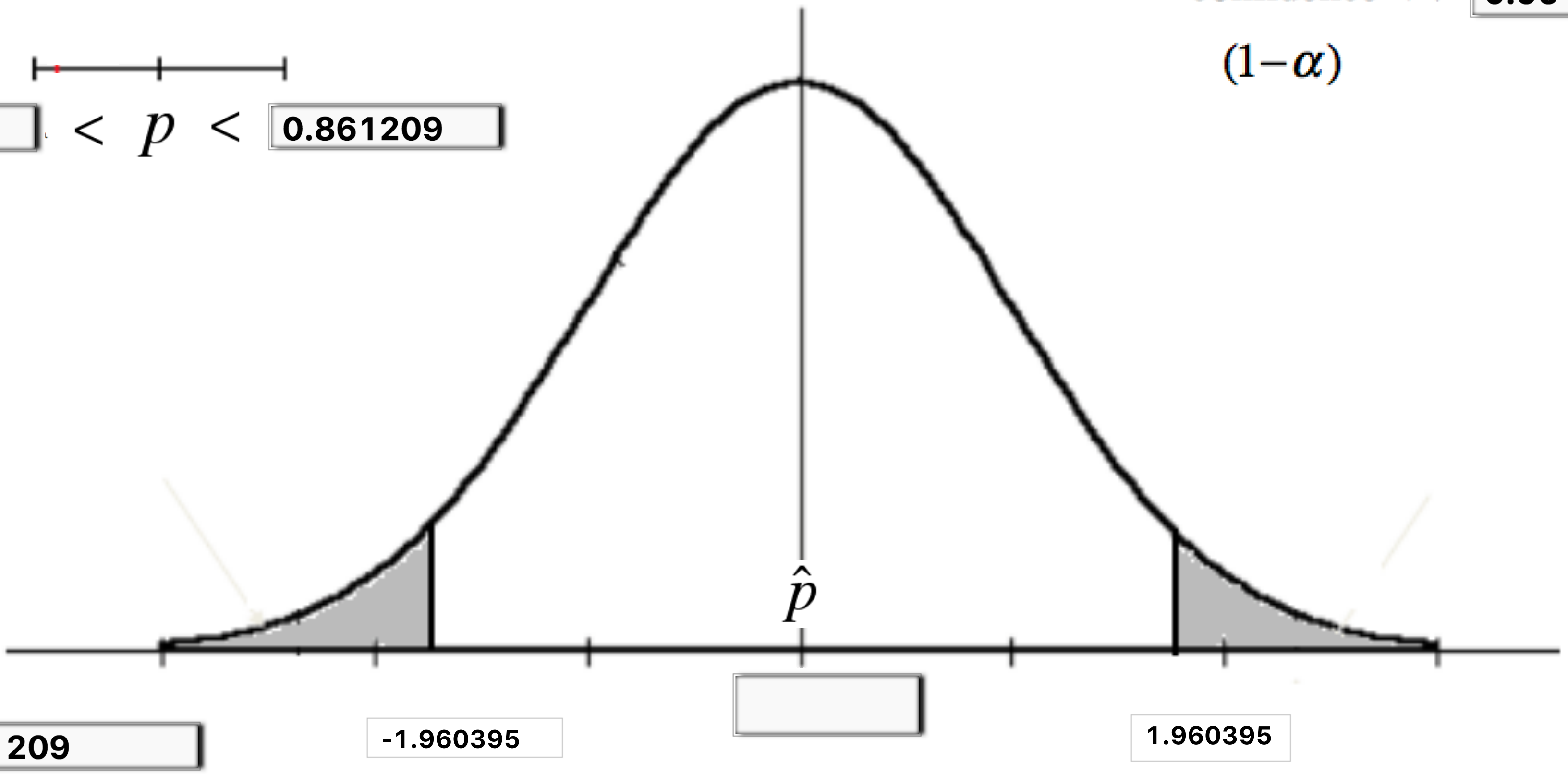
(using population proportions)

n -> 100

confidence ↦ 0.95

$(1-\alpha)$

0.698791 $< p <$ 0.861209



0.081209

ME

☑ z

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

$$ME = \left| z_{\frac{\alpha}{2}} \left(\sqrt{\frac{\hat{p}\hat{q}}{n}} \right) \right|$$

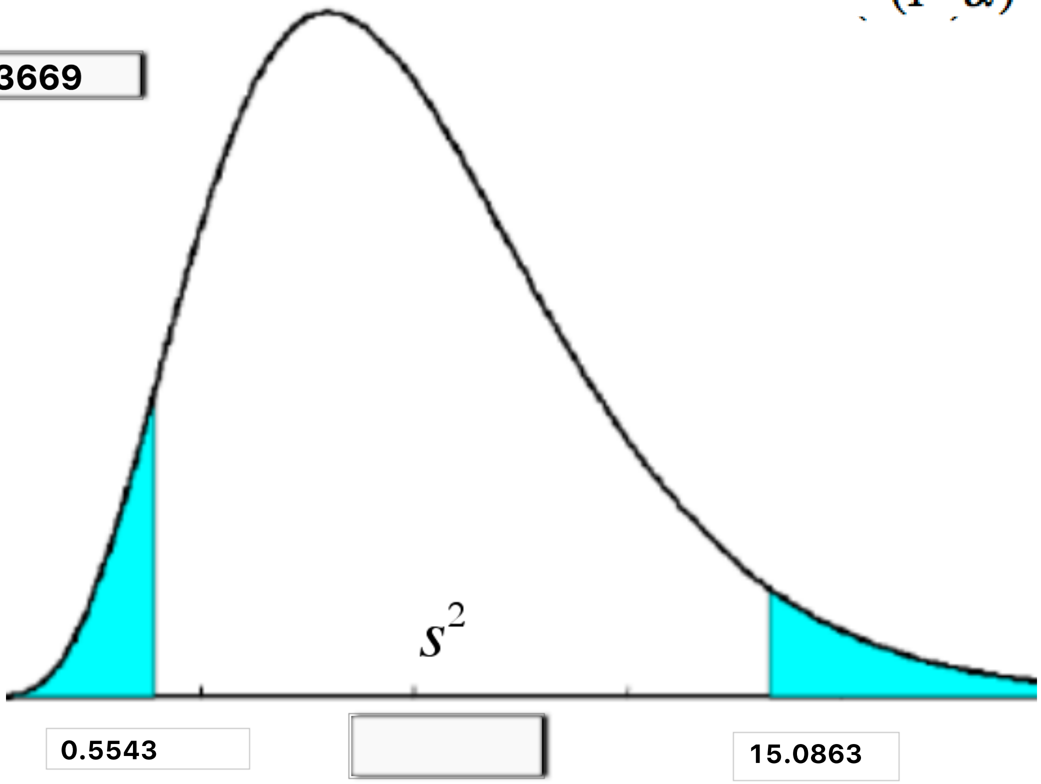
Estimating the variance

n -> 6

confidence ↦ 0.98
$(1-\alpha)$

●

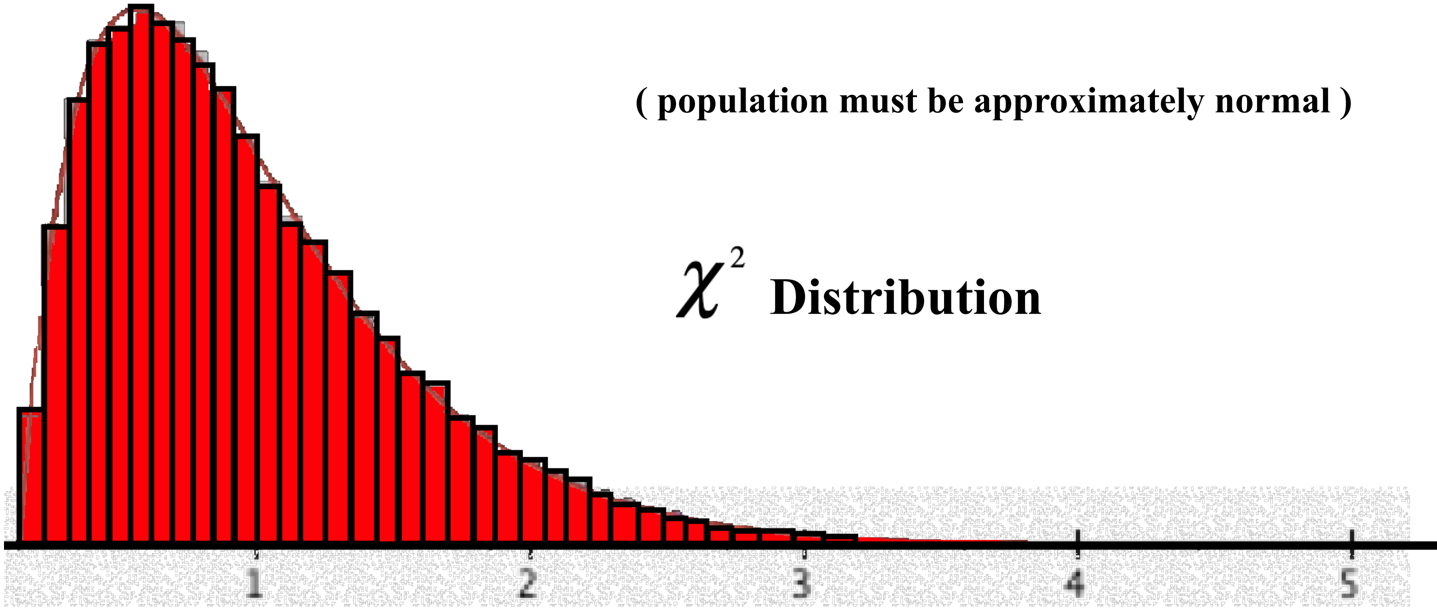
$0.059657 < \sigma^2 < 1.623669$



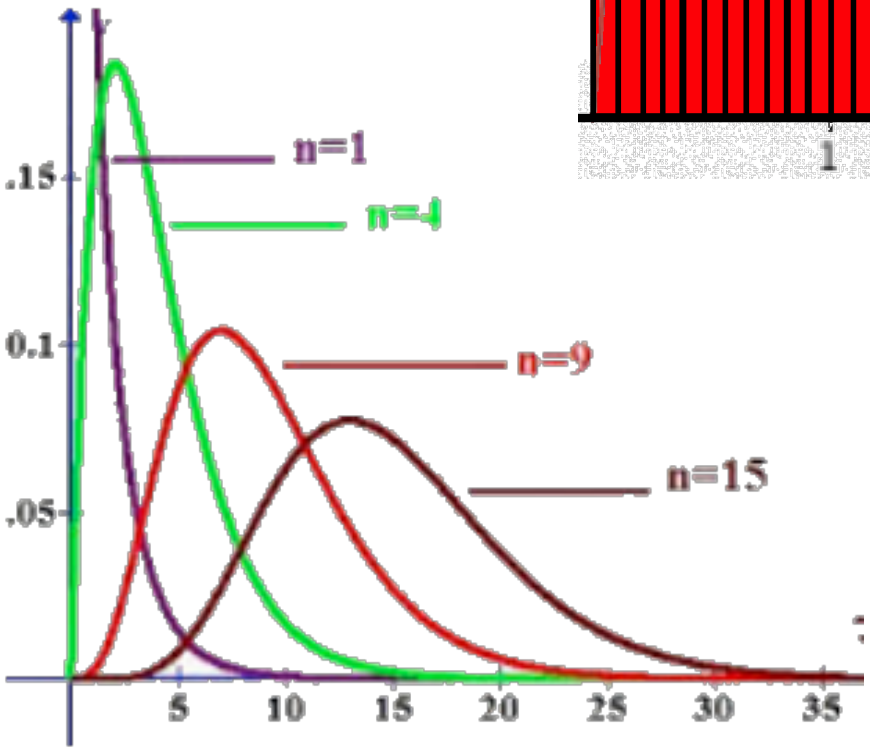
$$\frac{(n-1)s^2}{\chi^2_{\nu,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\nu,1-\frac{\alpha}{2}}}\ , \quad \nu = n-1$$

Sampling Distribution of Variances

(population must be approximately normal)



χ^2 Distribution



$$f(x)=\frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$$

μ -> ν

σ -> $\sqrt{2\nu}$

Calculating the sample size

σ -> .3

.12

ME

confidence ↦ 0.98

$(1-\alpha)$

p ->

●

33.837062

2.326785

z

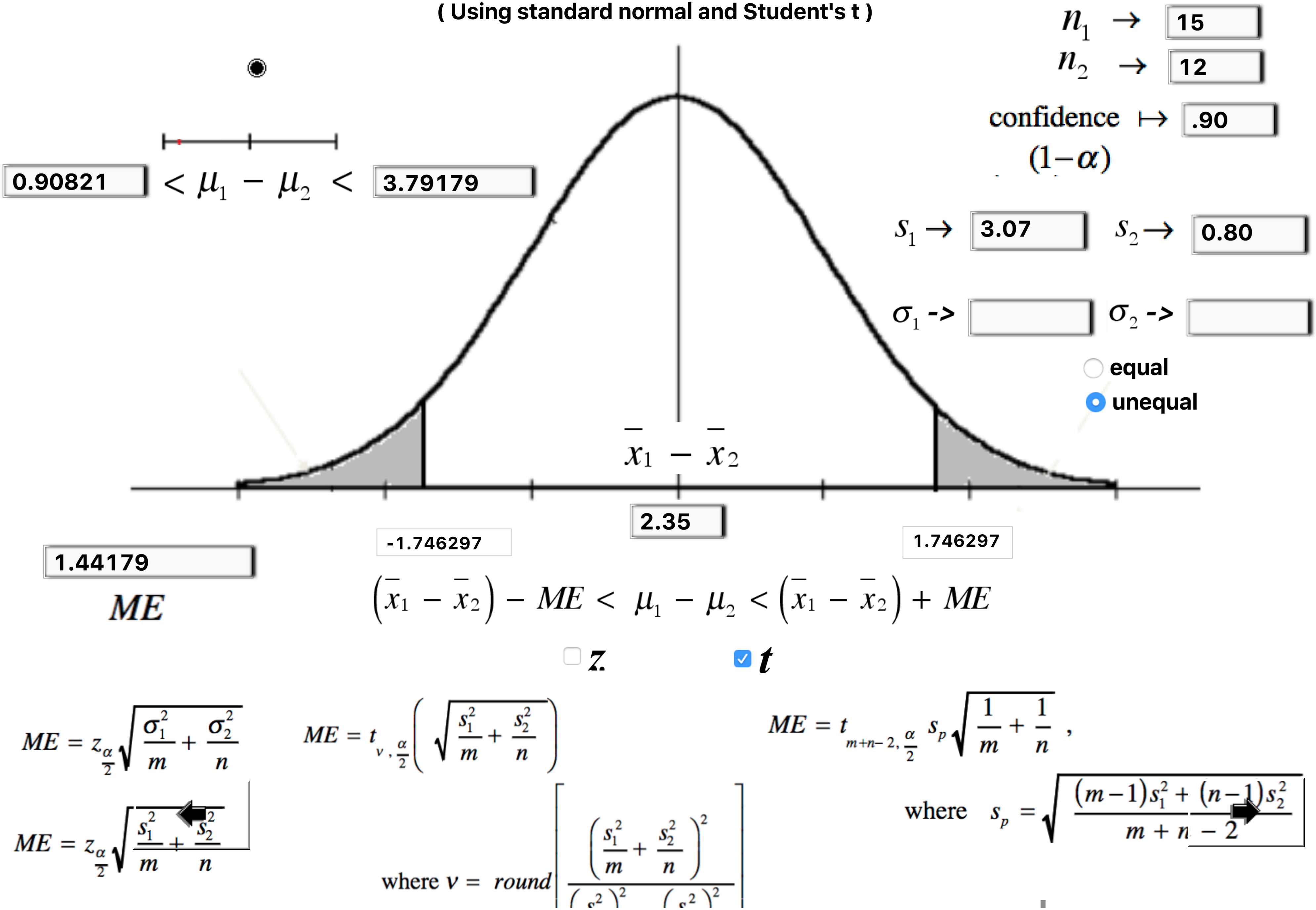
$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{ME} \right)^2$$

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sqrt{pq}}{ME} \right)^2$$

⬅ ➡

Estimating the difference between two means

(Using standard normal and Student's t)



$n_1 \rightarrow$ 15

$n_2 \rightarrow$ 12

confidence ↦ .90

$(1-\alpha)$

$s_1 \rightarrow$ 3.07 $s_2 \rightarrow$ 0.80

σ_1 -> σ_2 ->

○ equal

● unequal

0.90821 $< \mu_1 - \mu_2 <$ 3.79179

1.44179

ME

$$\left(\bar{x}_1 - \bar{x}_2\right) - ME < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + ME$$

☐ z ☑ t

$$ME = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

$$ME = z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

$$ME = t_{v,\frac{\alpha}{2}}\left(\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}\right)$$

where $v = \mathrm{round}\left[\frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\left(s^2\right)^2 \quad \left(s^2\right)^2}\right.$

$$ME = t_{m+n-2,\frac{\alpha}{2}}\, s_p\sqrt{\frac{1}{m} + \frac{1}{n}},$$

where $s_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$

Estimating the difference between two proportions

(Using standard normal distrinution)

$n_1 \rightarrow$

$n_2 \rightarrow$

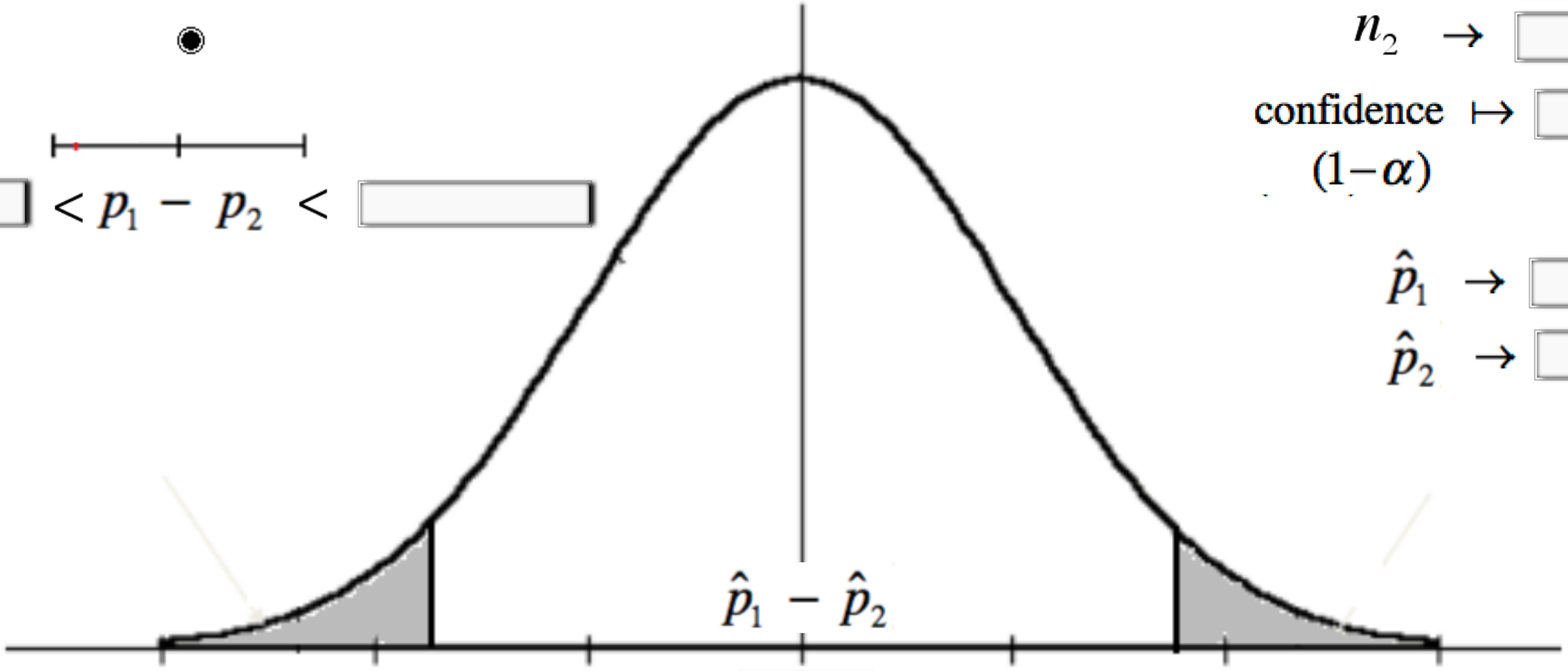
confidence $(1-\alpha) \mapsto$

$\hat{p}_1 \rightarrow$

$\hat{p}_2 \rightarrow$

●

$< p_1 - p_2 <$



ME

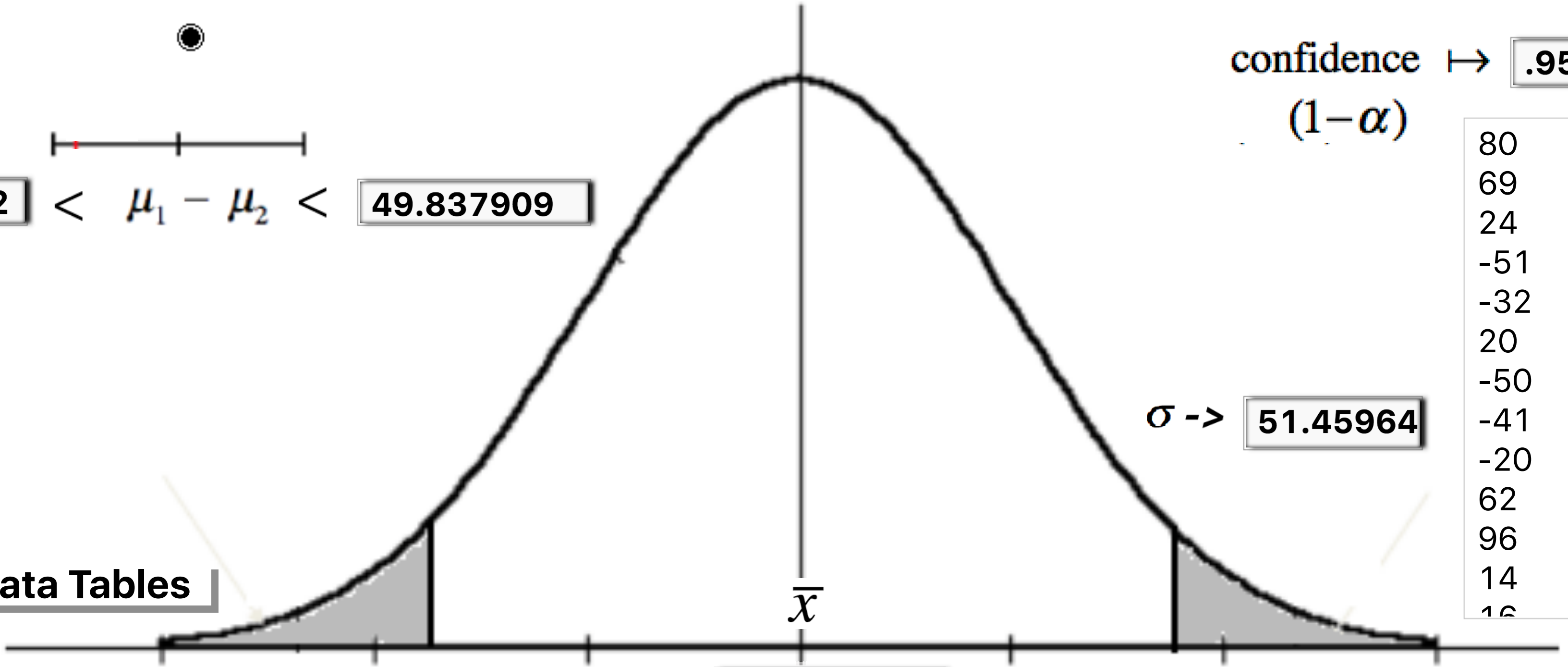
$$(\hat{p}_1 - \hat{p}_2) - ME < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + ME$$

☑ z

$$ME = \left| z_{\frac{\alpha}{2}} \left(\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right) \right|$$

Estimating the difference between two means

(interactive or "matched" pairs)



n -> 15

confidence $(1-\alpha)$ ↦ .95

σ -> 51.45964

-7.171242 $< \mu_1 - \mu_2 <$ 49.837909

Clear Data Tables

21.333333

-2.145327

2.145327

228	148	80
201	132	69
179	155	24
89	140	-51
82	114	-32
142	122	20
72	122	-50
152	193	-41
92	112	-20
166	104	62
214	118	96
174	160	14
124	108	

$$\widehat{x} - ME < \mu_1 - \mu_2 < \widehat{x} + ME$$

☑ t

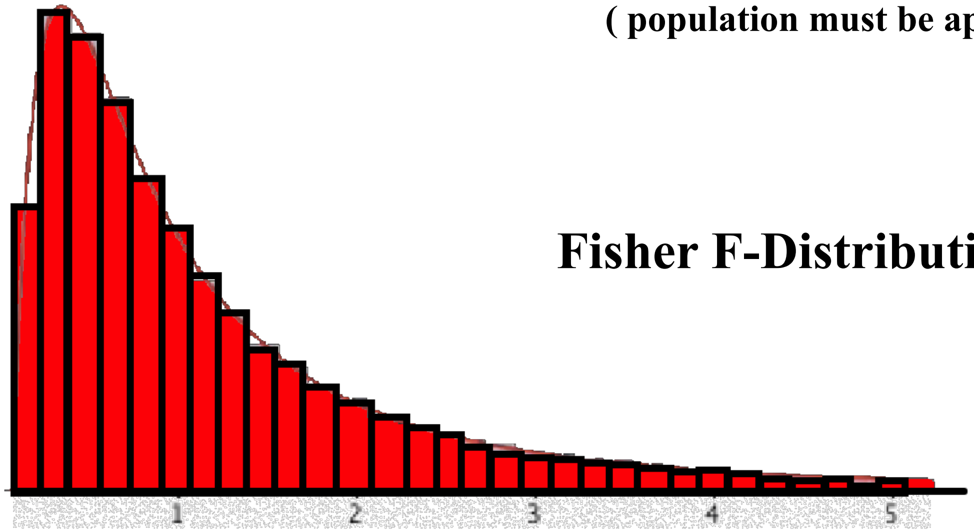
$$ME = \left| t_{\nu,\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}} \right) \right|, \quad \nu = n - 1$$

28.504575

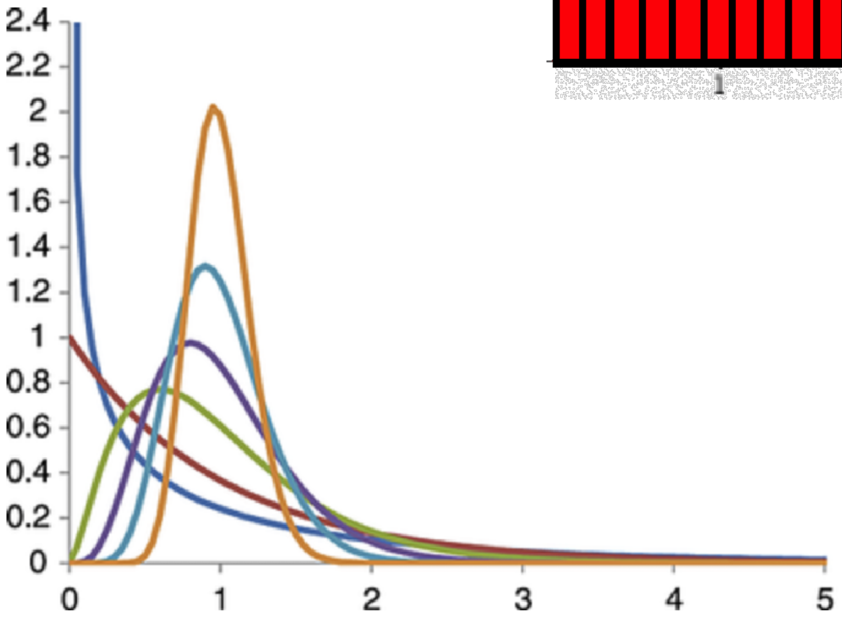
ME

Sampling Distribution of the Ratio of Variances

(population must be approximately normal)



Fisher F-Distribution



$$f(x)=\frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}x^{\frac{\nu_1}{2}-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}}\quad x\geq 0$$

Histogram

Number of intervals:

○ 10

● 20

Sample. Size: 5

Number of Samples: 1000

Mean of Sampling Distribution: 0.09

Standard Deviation of Sampling Distribution: 0.349142

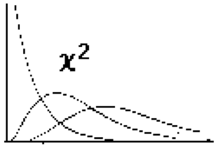
Standard Error: (Calculated) 0.635204

SAMPLE

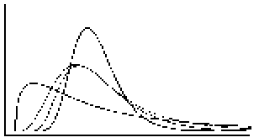
⬅

Population Size: 500

Student's t Distribution



ChiSquare Distribution



Snedecker's F Distribution

MU : 0.003518

SIGMA : 1.420359

VAR : 2.017419

μ -> 23

σ -> 12

n -> 23

p -> .3

2.4174
1.2545
-1.402
0.0580
-0.479
-1.168
0.7442
0.8951
-0.036
1.9538
-1.037
-1.328
-0.389
0.1558
1.6170
-0.225
0.4676
1.0258
1.6424
-2.076
0.6546
1.4749
1.1577
0.0801
0.5776
0.4897
-2.475
-1.250
3.6257
0.8444

➡