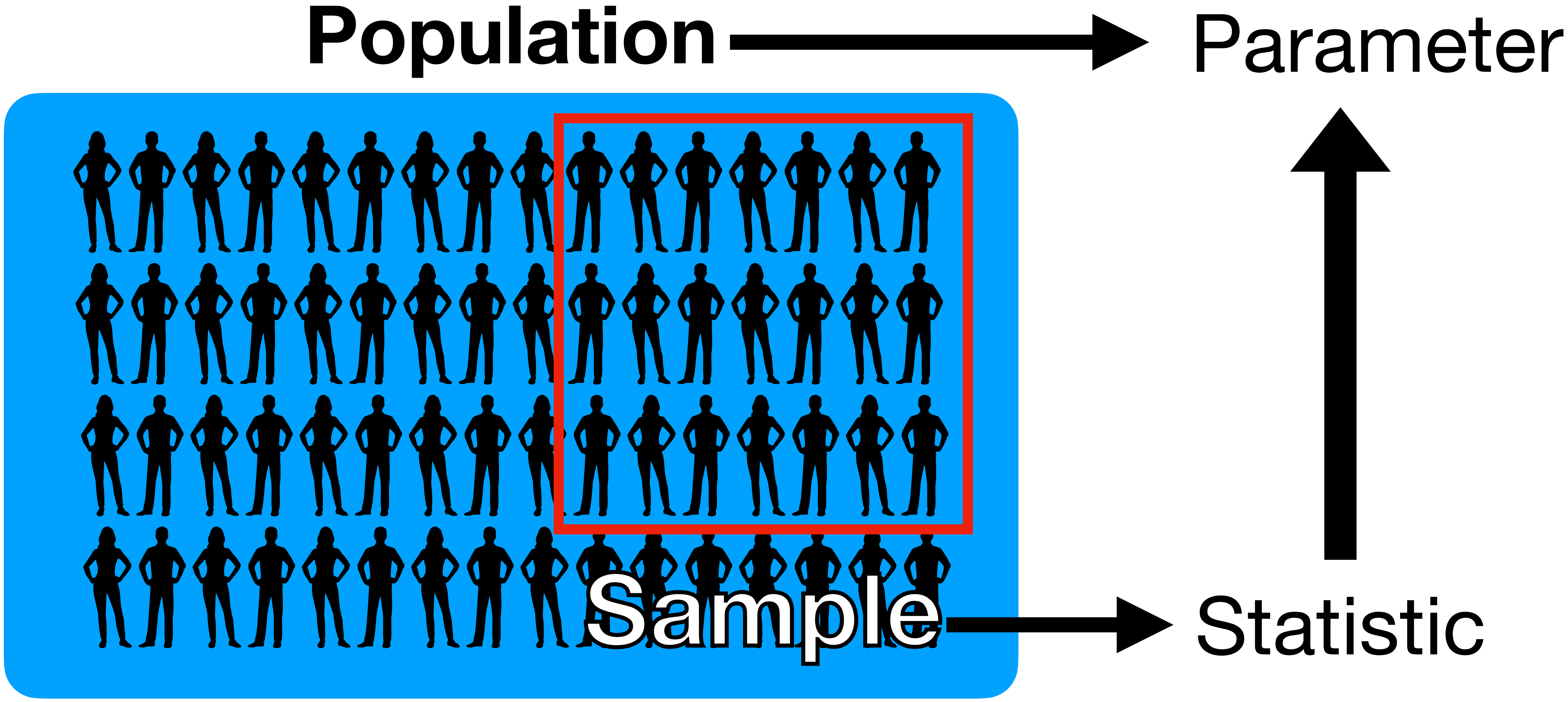
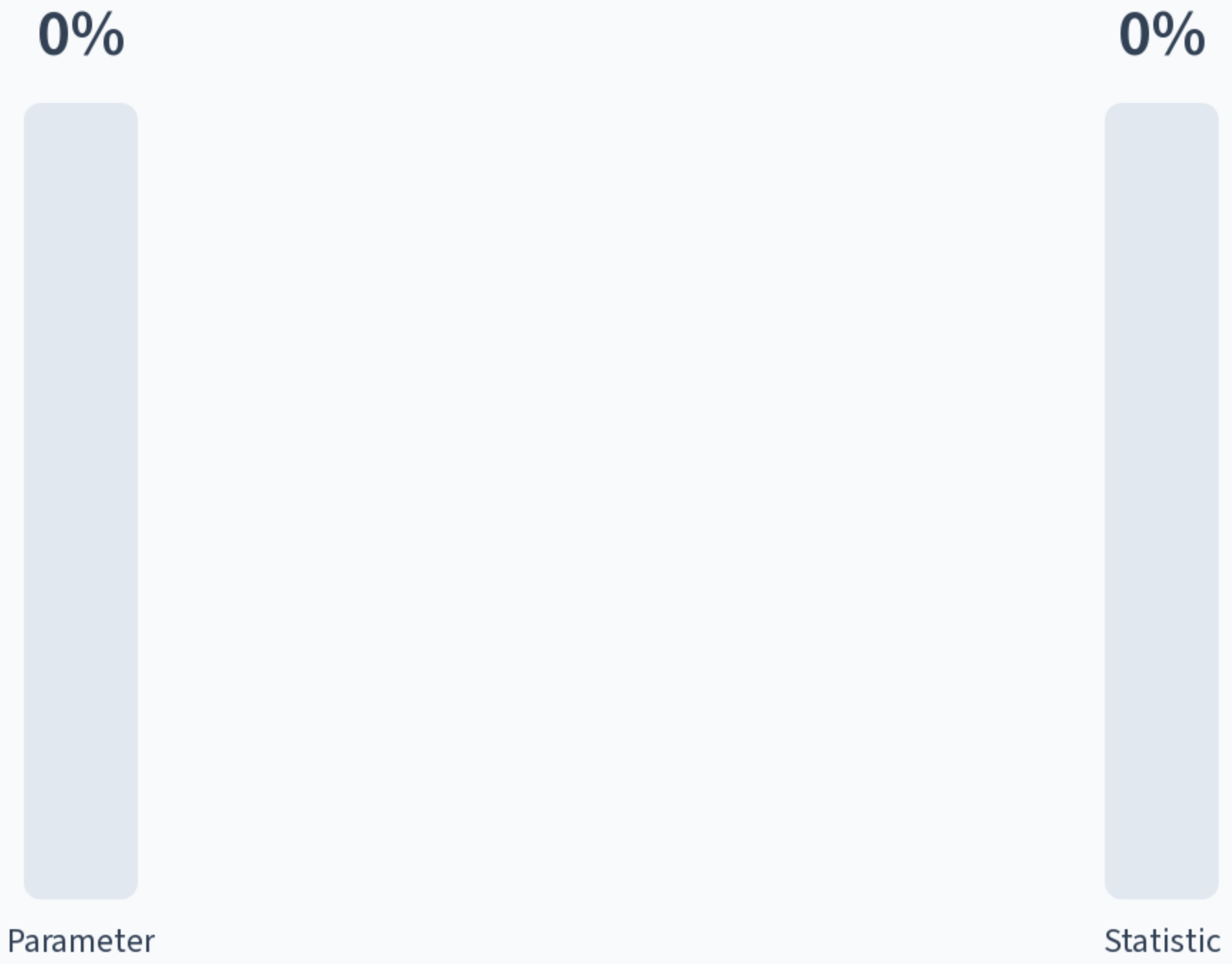
Lessons 021 & 022
Sampling Distributions

Monday, October 30


Population
Parameter
Sample
Statistic

Suppose you ask your closest 3 friends in STAT 2593 the grade that they got on their midterms, and you take this four person average. This is a ...



Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

A Canada-wide survey of $50,000$ adults are asked about their consumption preferences. The proportion of respondents who prefer brand A to brand B is a ...

0%

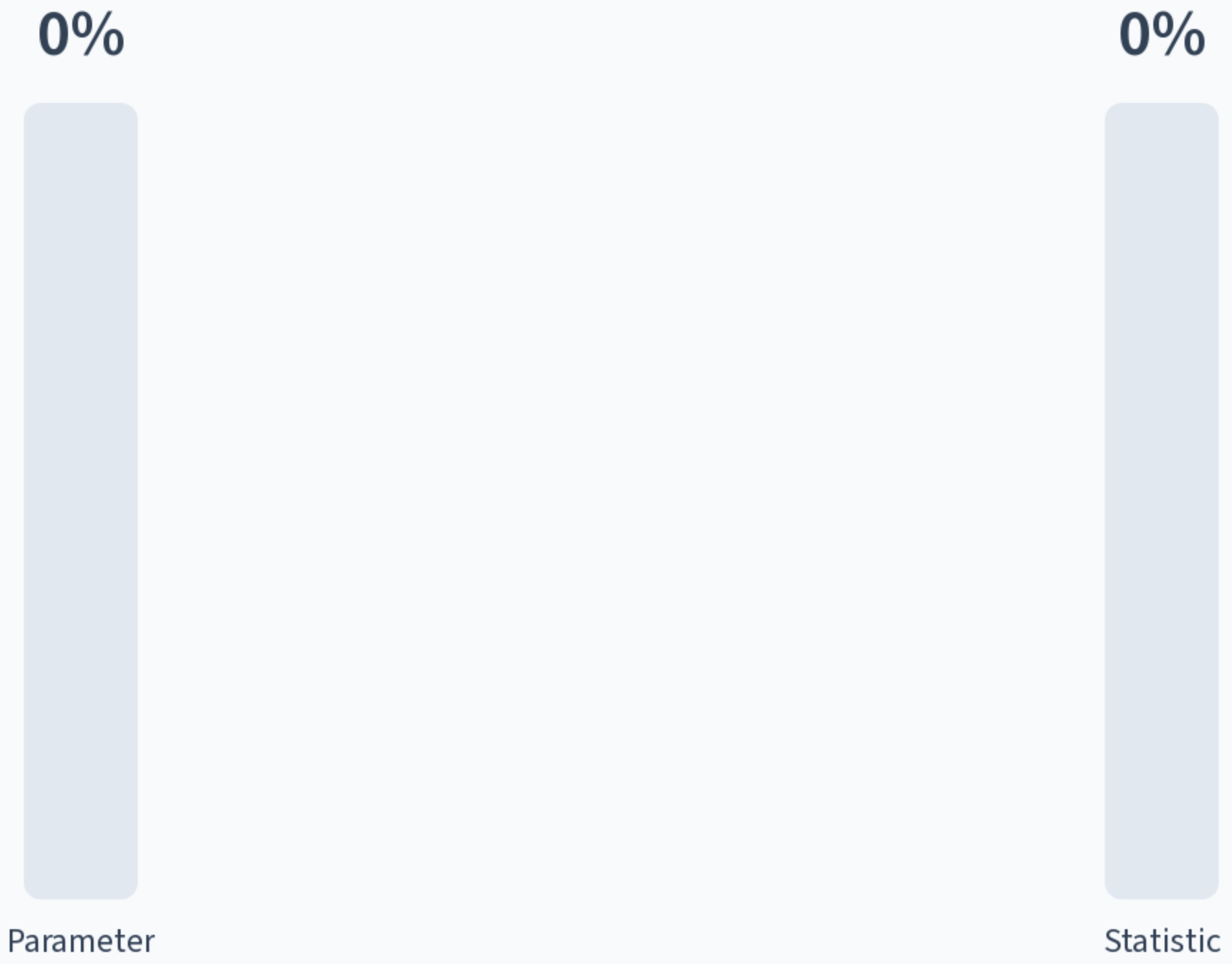
Parameter

0%

Statistic

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

A new manufacturing process is being used to manufacture glass which is meant to be stronger. The average hardness of glass produced by this process is a ...



Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

A statistic will not generally equal the underlying parameter.

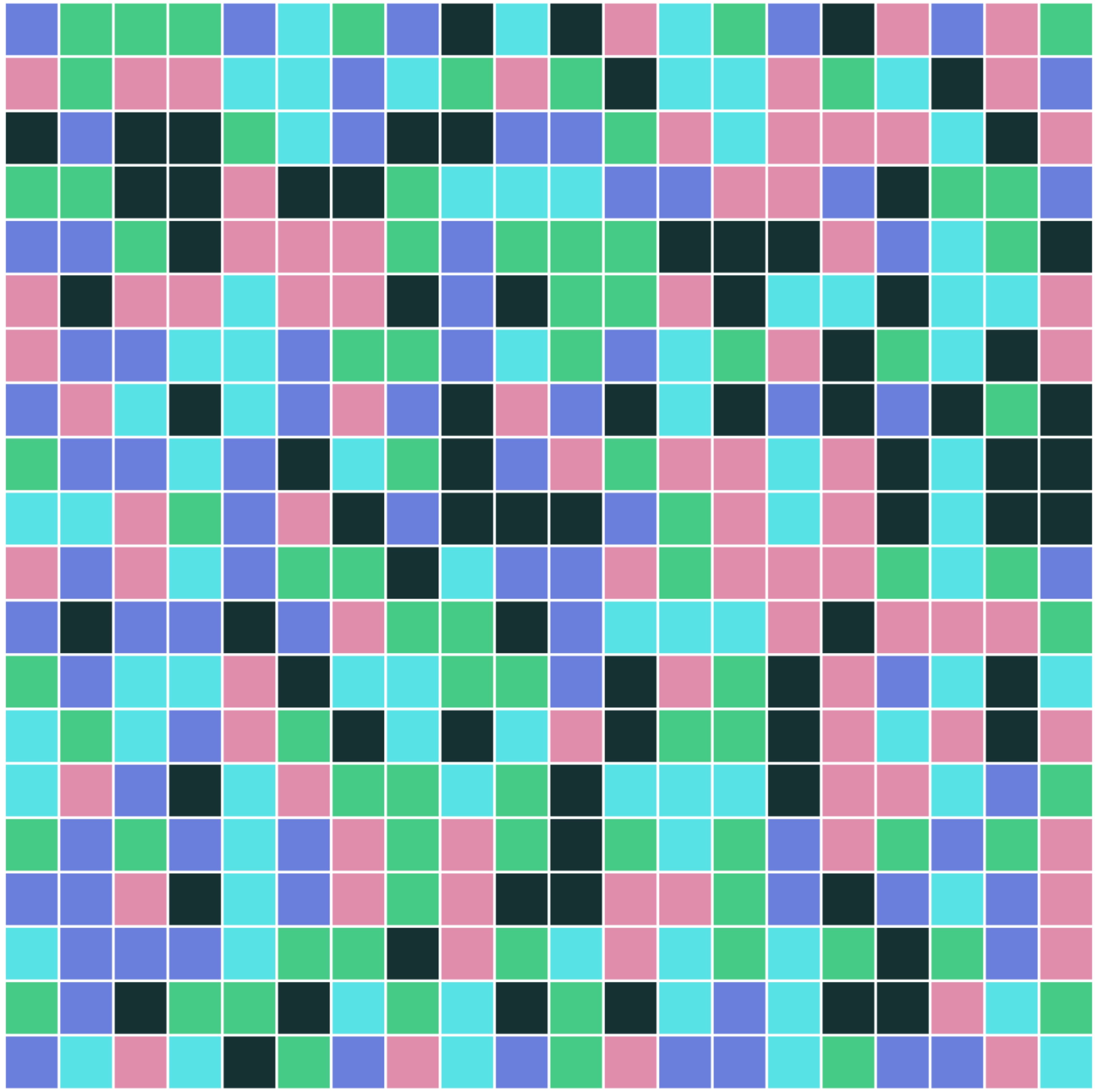
Sampling Distributions

- The goal of statistics is to characterize how reliable our statistics are as proxies for population parameters.
  - "Based on observed data, we are confident the truth will fall in this range."
  - "Based on observed data, we believe the new process is significantly better."

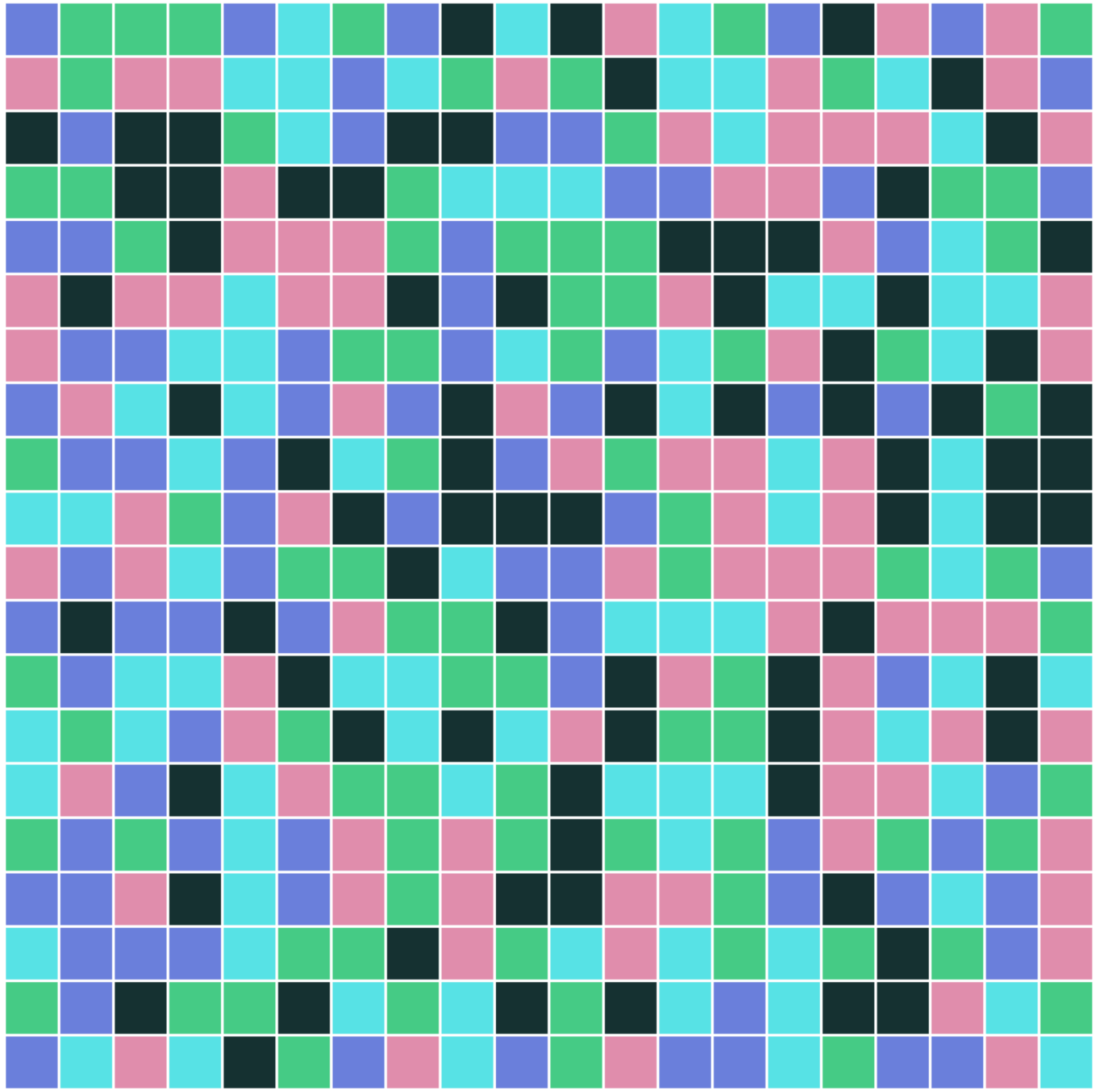
Sampling Distributions

- In order to make these statements, we need the **distribution of a statistic.**
- The distribution of a statistic is called the **sampling distribution**.
- The sampling distribution captures the random variation inherent in sampling.

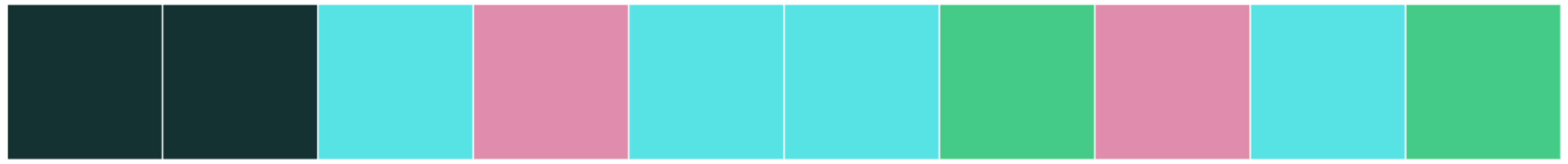
Population: P=0.195

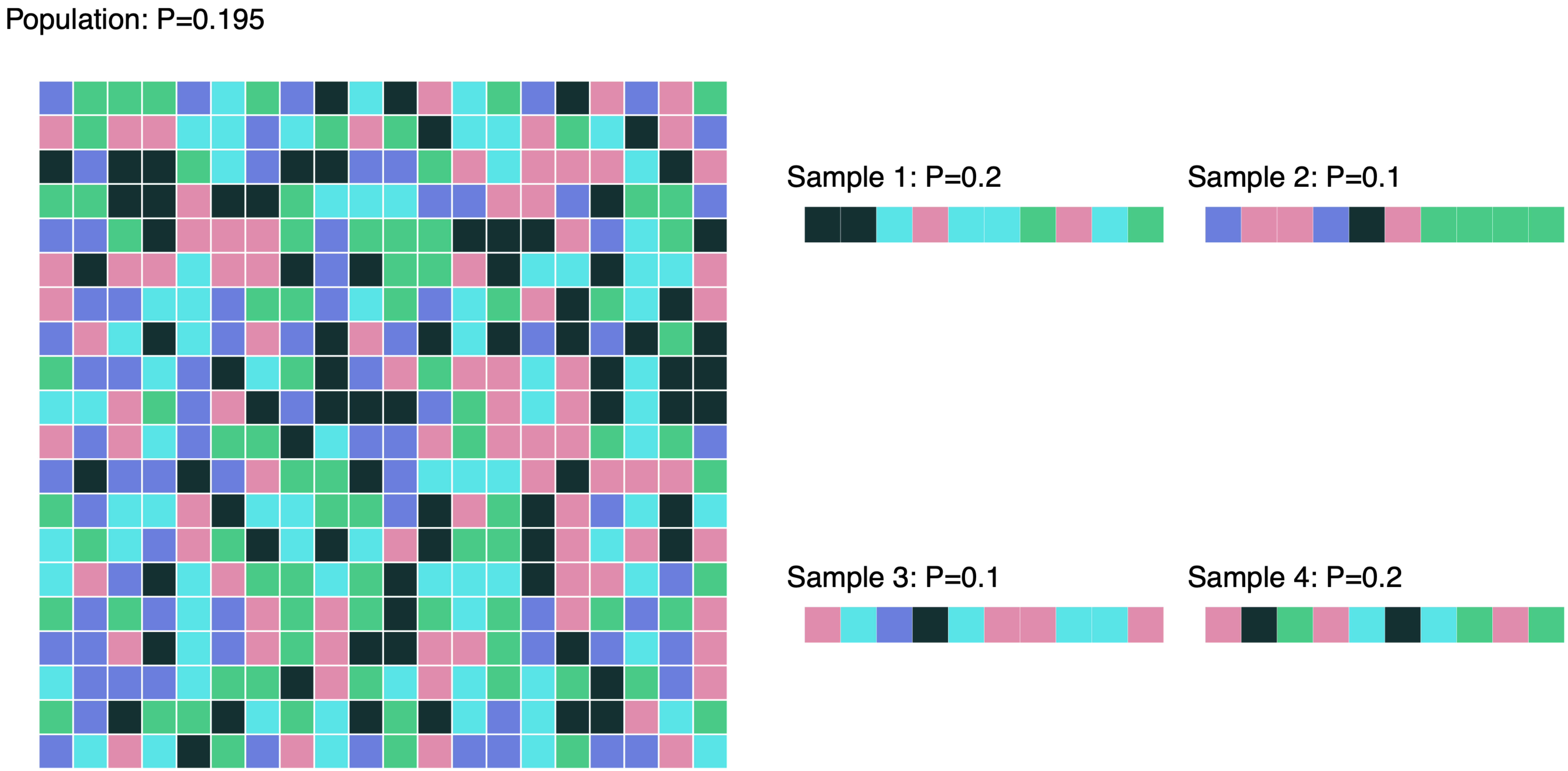


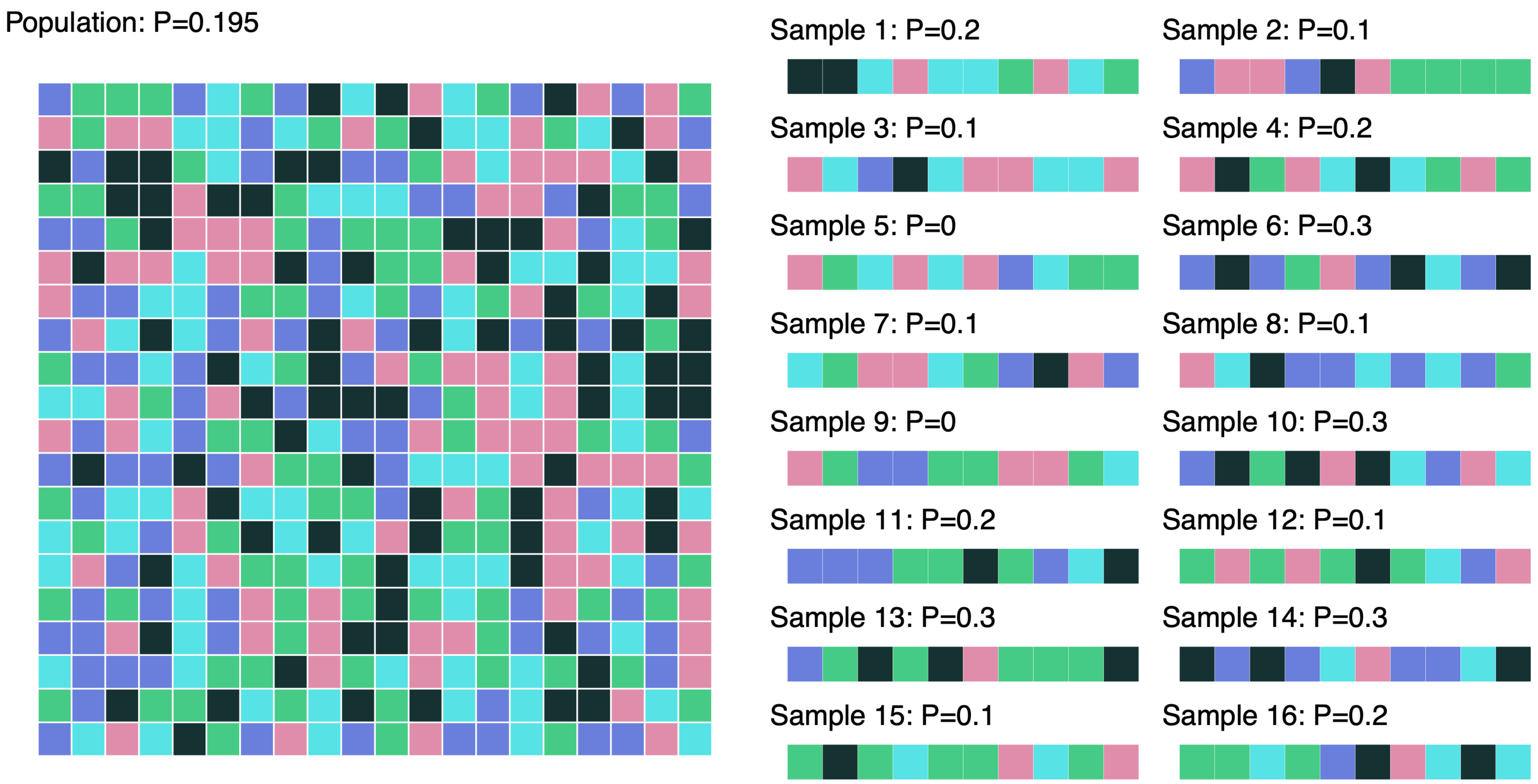
Population: P=0.195

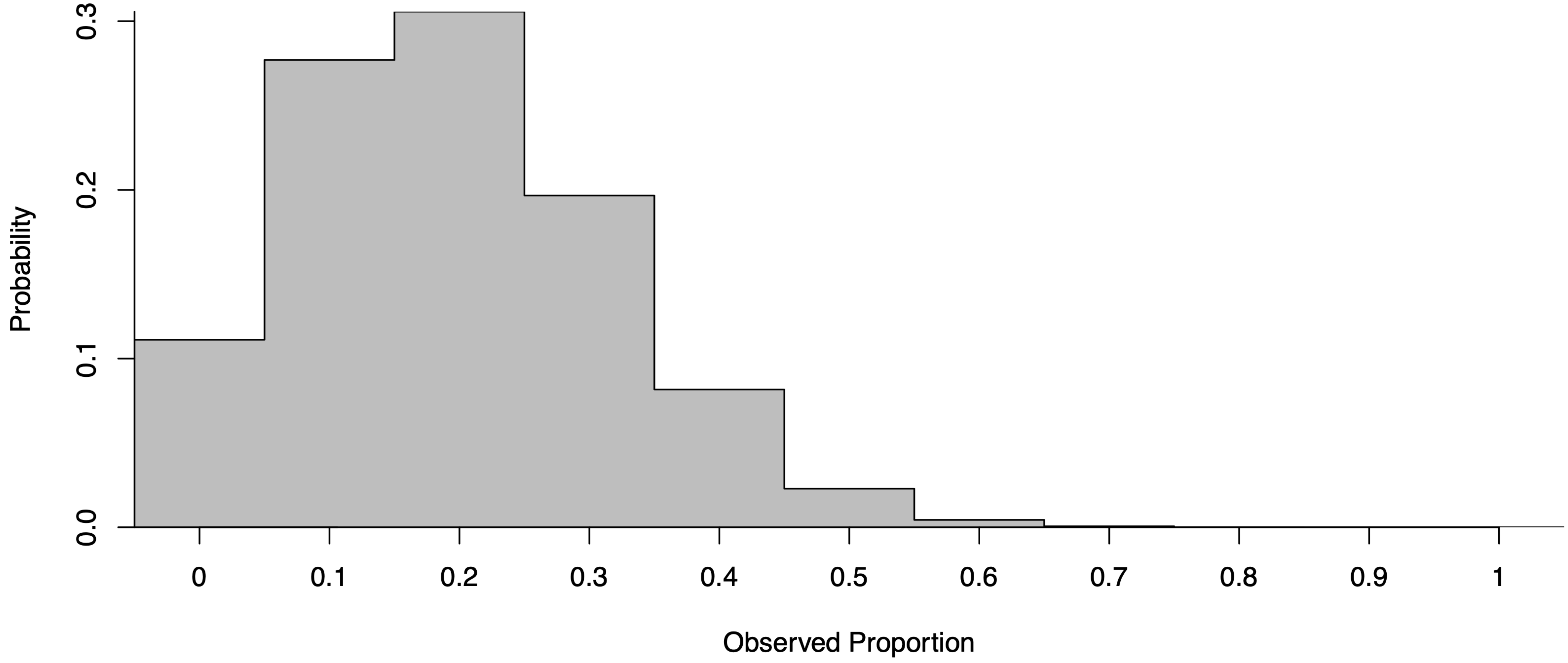


Sample 1: P=0.2




Population: P=0.195
Sample 1: P=0.2
Sample 2: P=0.1
Sample 3: P=0.1
Sample 4: P=0.2


Population: P=0.195
Sample 1: P=0.2
Sample 2: P=0.1
Sample 3: P=0.1
Sample 4: P=0.2
Sample 5: P=0
Sample 6: P=0.3
Sample 7: P=0.1
Sample 8: P=0.1
Sample 9: P=0
Sample 10: P=0.3
Sample 11: P=0.2
Sample 12: P=0.1
Sample 13: P=0.3
Sample 14: P=0.3
Sample 15: P=0.1
Sample 16: P=0.2



Probability
0.0
0.1
0.2
0.3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Observed Proportion

0.11	0.85	0.97	0.21	0.64	0.99	0.63	0.12	0.87	0.10	0.52	0.04	0.55	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	0.36	0.44	0.67	0.10	0.38	0.91	0.23	0.38	0.37	0.97
0.53	0.27	0.50	0.86	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	0.90	0.50	0.73	0.73
0.50	0.04	0.78	0.01	0.49	0.40	0.34	0.54	0.19	0.50	0.87	0.73	0.16	0.39	0.02	0.86
0.30	0.08	0.70	0.51	0.03	0.48	0.86	0.31	0.62	0.92	0.84	0.00	0.90	0.68	0.24	0.16
0.41	0.97	0.75	0.67	0.29	0.44	0.65	0.56	0.55	0.58	0.71	0.71	0.15	0.32	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	0.97	0.38	0.61	0.72	0.31	0.68
0.57	0.86	0.24	0.07	0.78	0.85	0.26	0.21	0.86	0.89	0.68	0.38	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	0.54	0.66	0.79	0.15	0.28	0.18	0.02	0.89	0.57	0.08	0.61	0.87
0.45	0.45	0.49	0.84	0.06	0.22	0.26	0.85	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	0.09	0.50	0.39	0.82	0.70	0.86	0.94	0.87	0.58	0.82	0.72	0.53	0.83	0.19	0.01
0.39	0.50	0.94	0.97	0.89	0.04	0.74	0.83	0.16	0.92	0.93	0.84	0.47	0.02	0.81	0.79
0.63	0.16	0.41	0.24	0.27	0.19	0.89	0.68	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	0.48	0.25	0.41	0.24	0.82	0.50	0.25	0.45	0.13	0.66	0.90	0.23	0.29	0.06

Sample Mean of Selected: 0.5858

0.11	0.85	0.97	0.21	0.64	**0.99**	0.63	0.12	**0.87**	0.10	0.52	**0.04**	**0.55**	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	**0.36**	**0.44**	0.67	0.10	0.38	**0.91**	0.23	0.38	**0.37**	0.97
0.53	0.27	**0.50**	**0.86**	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	**0.90**	0.50	**0.73**	0.73
0.50	0.04	0.78	0.01	**0.49**	0.40	**0.34**	0.54	**0.19**	0.50	0.87	0.73	0.16	0.39	0.02	**0.86**
0.30	**0.08**	0.70	0.51	0.03	0.48	**0.86**	**0.31**	0.62	0.92	0.84	**0.00**	**0.90**	0.68	0.24	**0.16**
0.41	0.97	0.75	**0.67**	0.29	0.44	**0.65**	0.56	0.55	0.58	**0.71**	**0.71**	0.15	**0.32**	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	**0.97**	0.38	0.61	**0.72**	0.31	**0.68**
0.57	**0.86**	0.24	0.07	0.78	0.85	**0.26**	0.21	**0.86**	**0.89**	0.68	**0.38**	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	**0.54**	0.66	0.79	0.15	0.28	0.18	0.02	0.89	**0.57**	**0.08**	0.61	**0.87**
0.45	0.45	0.49	0.84	0.06	0.22	0.26	**0.85**	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	**0.09**	**0.50**	0.39	0.82	0.70	0.86	**0.94**	0.87	0.58	**0.82**	0.72	**0.53**	**0.83**	0.19	0.01
0.39	**0.50**	**0.94**	0.97	0.89	0.04	**0.74**	0.83	0.16	0.92	**0.93**	**0.84**	0.47	0.02	0.81	0.79
0.63	0.16	**0.41**	**0.24**	0.27	0.19	0.89	**0.68**	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	**0.48**	0.25	0.41	0.24	0.82	0.50	0.25	**0.45**	0.13	0.66	**0.90**	**0.23**	0.29	0.06

Sample Mean of Selected: 0.5477

0.11	0.85	0.97	0.21	0.64	0.99	0.63	0.12	0.87	0.10	0.52	0.04	0.55	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	0.36	0.44	0.67	0.10	0.38	0.91	0.23	0.38	0.37	0.97
0.53	0.27	0.50	0.86	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	0.90	0.50	0.73	0.73
0.50	0.04	0.78	0.01	0.49	0.40	0.34	0.54	0.19	0.50	0.87	0.73	0.16	0.39	0.02	0.86
0.30	0.08	0.70	0.51	0.03	0.48	0.86	0.31	0.62	0.92	0.84	0.00	0.90	0.68	0.24	0.16
0.41	0.97	0.75	0.67	0.29	0.44	0.65	0.56	0.55	0.58	0.71	0.71	0.15	0.32	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	0.97	0.38	0.61	0.72	0.31	0.68
0.57	0.86	0.24	0.07	0.78	0.85	0.26	0.21	0.86	0.89	0.68	0.38	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	0.54	0.66	0.79	0.15	0.28	0.18	0.02	0.89	0.57	0.08	0.61	0.87
0.45	0.45	0.49	0.84	0.06	0.22	0.26	0.85	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	0.09	0.50	0.39	0.82	0.70	0.86	0.94	0.87	0.58	0.82	0.72	0.53	0.83	0.19	0.01
0.39	0.50	0.94	0.97	0.89	0.04	0.74	0.83	0.16	0.92	0.93	0.84	0.47	0.02	0.81	0.79
0.63	0.16	0.41	0.24	0.27	0.19	0.89	0.68	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	0.48	0.25	0.41	0.24	0.82	0.50	0.25	0.45	0.13	0.66	0.90	0.23	0.29	0.06

Sample Mean of Selected: 0.4783

0.11	0.85	0.97	0.21	0.64	0.99	0.63	0.12	0.87	0.10	0.52	0.04	0.55	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	0.36	0.44	0.67	0.10	0.38	0.91	0.23	0.38	0.37	0.97
0.53	0.27	0.50	0.86	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	0.90	0.50	0.73	0.73
0.50	0.04	0.78	0.01	0.49	0.40	0.34	0.54	0.19	0.50	0.87	0.73	0.16	0.39	0.02	0.86
0.30	0.08	0.70	0.51	0.03	0.48	0.86	0.31	0.62	0.92	0.84	0.00	0.90	0.68	0.24	0.16
0.41	0.97	0.75	0.67	0.29	0.44	0.65	0.56	0.55	0.58	0.71	0.71	0.15	0.32	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	0.97	0.38	0.61	0.72	0.31	0.68
0.57	0.86	0.24	0.07	0.78	0.85	0.26	0.21	0.86	0.89	0.68	0.38	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	0.54	0.66	0.79	0.15	0.28	0.18	0.02	0.89	0.57	0.08	0.61	0.87
0.45	0.45	0.49	0.84	0.06	0.22	0.26	0.85	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	0.09	0.50	0.39	0.82	0.70	0.86	0.94	0.87	0.58	0.82	0.72	0.53	0.83	0.19	0.01
0.39	0.50	0.94	0.97	0.89	0.04	0.74	0.83	0.16	0.92	0.93	0.84	0.47	0.02	0.81	0.79
0.63	0.16	0.41	0.24	0.27	0.19	0.89	0.68	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	0.48	0.25	0.41	0.24	0.82	0.50	0.25	0.45	0.13	0.66	0.90	0.23	0.29	0.06

Sample Mean of Selected: 0.5336

0.11	0.85	0.97	0.21	0.64	0.99	0.63	0.12	0.87	0.10	0.52	0.04	0.55	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	0.36	0.44	0.67	0.10	0.38	0.91	0.23	0.38	0.37	0.97
0.53	0.27	0.50	0.86	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	0.90	0.50	0.73	0.73
0.50	0.04	0.78	0.01	0.49	0.40	0.34	0.54	0.19	0.50	0.87	0.73	0.16	0.39	0.02	0.86
0.30	0.08	0.70	0.51	0.03	0.48	0.86	0.31	0.62	0.92	0.84	0.00	0.90	0.68	0.24	0.16
0.41	0.97	0.75	0.67	0.29	0.44	0.65	0.56	0.55	0.58	0.71	0.71	0.15	0.32	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	0.97	0.38	0.61	0.72	0.31	0.68
0.57	0.86	0.24	0.07	0.78	0.85	0.26	0.21	0.86	0.89	0.68	0.38	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	0.54	0.66	0.79	0.15	0.28	0.18	0.02	0.89	0.57	0.08	0.61	0.87
0.45	0.45	0.49	0.84	0.06	0.22	0.26	0.85	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	0.09	0.50	0.39	0.82	0.70	0.86	0.94	0.87	0.58	0.82	0.72	0.53	0.83	0.19	0.01
0.39	0.50	0.94	0.97	0.89	0.04	0.74	0.83	0.16	0.92	0.93	0.84	0.47	0.02	0.81	0.79
0.63	0.16	0.41	0.24	0.27	0.19	0.89	0.68	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	0.48	0.25	0.41	0.24	0.82	0.50	0.25	0.45	0.13	0.66	0.90	0.23	0.29	0.06

Sample Mean of Selected: 0.5523

0.11	0.85	0.97	0.21	0.64	0.99	0.63	0.12	0.87	0.10	0.52	0.04	0.55	0.96	0.46	0.53
0.82	0.48	0.96	0.89	0.93	0.39	0.36	0.44	0.67	0.10	0.38	0.91	0.23	0.38	0.37	0.97
0.53	0.27	0.50	0.86	0.15	0.69	0.63	0.36	0.10	0.93	0.82	0.08	0.90	0.50	0.73	0.73
0.50	0.04	0.78	0.01	0.49	0.40	0.34	0.54	0.19	0.50	0.87	0.73	0.16	0.39	0.02	0.86
0.30	0.08	0.70	0.51	0.03	0.48	0.86	0.31	0.62	0.92	0.84	0.00	0.90	0.68	0.24	0.16
0.41	0.97	0.75	0.67	0.29	0.44	0.65	0.56	0.55	0.58	0.71	0.71	0.15	0.32	0.53	0.86
0.57	0.45	0.56	0.92	0.12	0.24	0.42	0.54	0.05	0.90	0.97	0.38	0.61	0.72	0.31	0.68
0.57	0.86	0.24	0.07	0.78	0.85	0.26	0.21	0.86	0.89	0.68	0.38	0.21	0.89	0.88	0.08
0.99	0.26	0.95	0.09	0.54	0.66	0.79	0.15	0.28	0.18	0.02	0.89	0.57	0.08	0.61	0.87
0.45	0.45	0.49	0.84	0.06	0.22	0.26	0.85	0.93	0.77	0.31	0.84	0.26	0.24	0.33	0.75
0.91	0.09	0.50	0.39	0.82	0.70	0.86	0.94	0.87	0.58	0.82	0.72	0.53	0.83	0.19	0.01
0.39	0.50	0.94	0.97	0.89	0.04	0.74	0.83	0.16	0.92	0.93	0.84	0.47	0.02	0.81	0.79
0.63	0.16	0.41	0.24	0.27	0.19	0.89	0.68	0.01	0.10	0.57	0.36	0.78	0.55	0.16	0.78
0.67	0.34	0.41	0.65	0.91	0.17	0.29	0.15	0.75	0.84	0.21	0.19	0.22	0.84	0.77	0.62
0.57	0.11	0.48	0.25	0.41	0.24	0.82	0.50	0.25	0.45	0.13	0.66	0.90	0.23	0.29	0.06

Sampling Distributions

- Statistics are computed based on random data.
- As a result, statistics are themselves random.
- The sampling distribution captures this randomness.
- Think of repeating the experiment many times over.

Sampling Distribution of the Sample Mean

- We have seen $\bar{x}$ as a summary statistic.
- For a random sample, we denote this $\overline{X}$.
- Consider a random sample, $X_1, \ldots, X_n \sim F(x)$.

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Sampling Distribution of the Sample Mean

- No matter $F(x)$: $E[\overline{X}] = E[X] = \mu$.
- No matter $F(x)$: $\text{var}(\overline{X}) = \frac{1}{n}\text{var}(X) = \frac{\sigma^2}{n}$.
    - This makes the standard deviation of $\overline{X}$ equal $\frac{\sigma}{\sqrt{n}}$.
    - Typically referred to as the **standard error.**
- What happens as $n \longrightarrow \infty$?

Limiting Distribution

- As $n \longrightarrow \infty$ we have var$(\bar{X}) \longrightarrow 0$.
- If $X_i \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$.
- If X_i are non-normal, then $\bar{X} \dot{\sim} N(\mu, \sigma^2/n)$ for **large n.**
- This is called the **central limit theorem**.

The Central Limit Theorem

- Suppose we have a random sample of size n from an unknown distribution, and we compute the sample mean of the sample.

Then, as $n \longrightarrow \infty$, we have $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

https://dylan-spicker.shinyapps.io/sampling_distributions/

Suppose that 90 observations, $X_1, \ldots, X_{90}$ are made from a normal population with mean 2 and variance 9. What best describes the distribution of $\overline{X}$?

Approximately $N(2, 9)$.

0%

Approximately $N(2, 0.1)$.

0%

Exactly $N(2, 9)$.

0%

Exactly $N(2, 0.1)$.

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Suppose that 90 observations, $X_1, \ldots, X_{90}$ are made from a non-normal population with mean 2 and variance 9. What best describes the distribution of $\overline{X}$?

Approximately $N(2, 9)$.

0%

Approximately $N(2, 0.1)$.

0%

Exactly $N(2, 9)$.

0%

Exactly $N(2, 0.1)$.

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

True or false: If n is large enough then no matter the distribution that X_i follows, we can use the normal distribution to make statements about $P(X_i \in [a, b])$.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

True or false: If n is large enough then no matter the distribution that X_i follows, we can use the normal distribution to make statements about $P(\overline{X} \in [a, b])$.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Why do we care?

- Suppose that you want to know whether the average volume of a process is correctly calibrated. You take samples from the process and compute the sample mean. You can then answer:

  "Assuming that the process is correctly calibrated, how likely are we to observe this sample mean value?"