

Statistics with Spa OWS

Lecture 2

Julia Schroeder

Julia.schroeder@imperial.ac.uk

Learning Aims

- Basic statistical concepts
 - Describing distributions
 - Describing sampling precision

Basic statistical concepts

```
> head(d)
```

	BirdID	Cohort	CaptureDate	CaptureTime	Year	Tarsus	Bill	Wing	Mass	Sex	Sex.1
1	4401	1991	21-Jun-00	<NA>	2000	18.9	NA	82	29.4	1	male
2	4401	1991	02-Oct-00	<NA>	2000	18.8	NA	79	31.6	1	male
3	4405	1994	20-Jun-00	<NA>	2000	19.1	NA	77	29.9	0	female
4	4405	1994	04-Oct-00	<NA>	2000	19.0	NA	78	31.6	0	female
5	4405	1994	07-Oct-00	<NA>	2000	19.1	NA	77	31.0	0	female
6	4409	1994	23-Mar-00	<NA>	2000	18.0	NA	76	28.1	1	male

```
> |
```

Lundy Sparrows

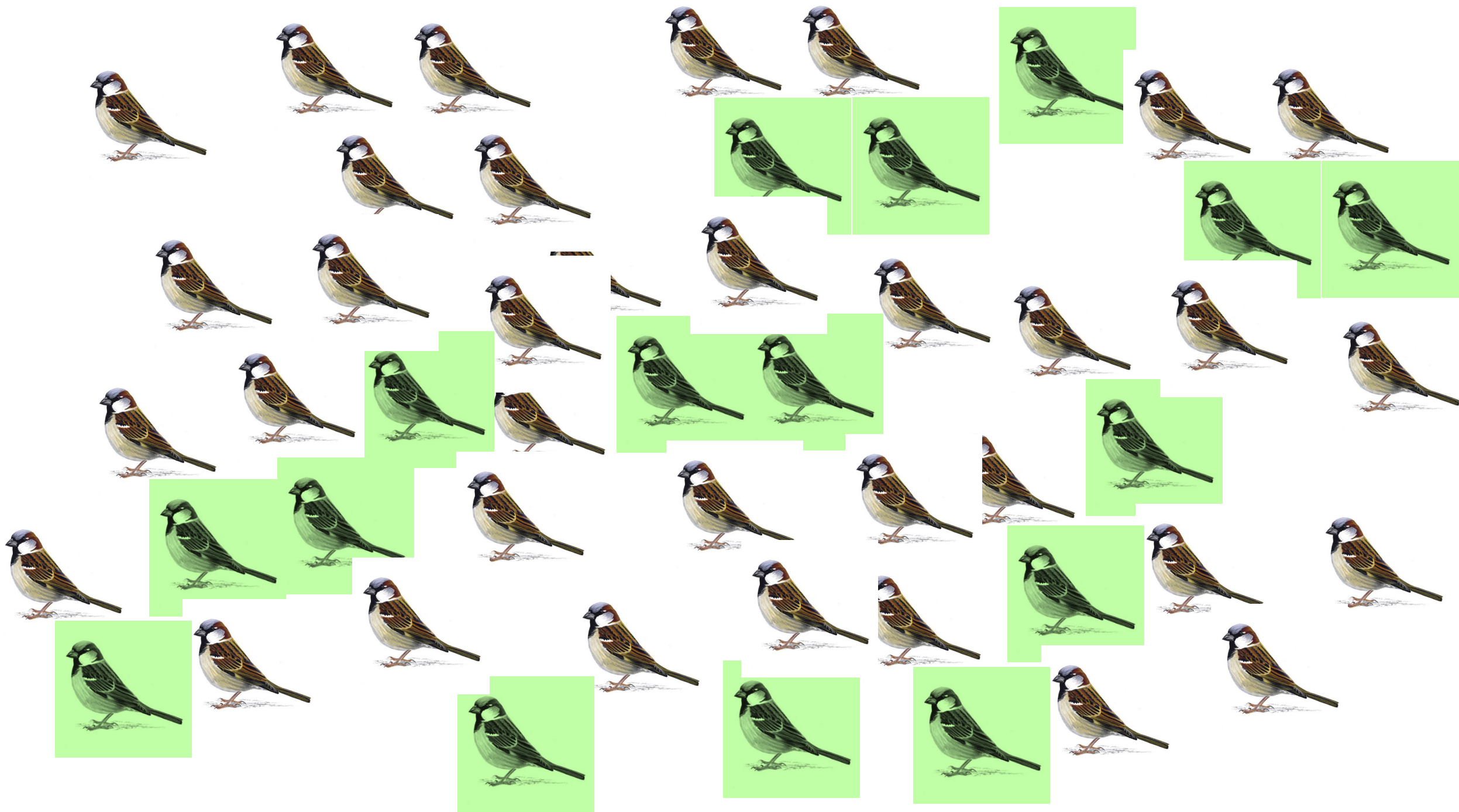


Griffith et al., 2000,
Nakagawa et al., 2007,
Cleasby et al., 2010,
Schroeder et. al 2011, 2013,
2015

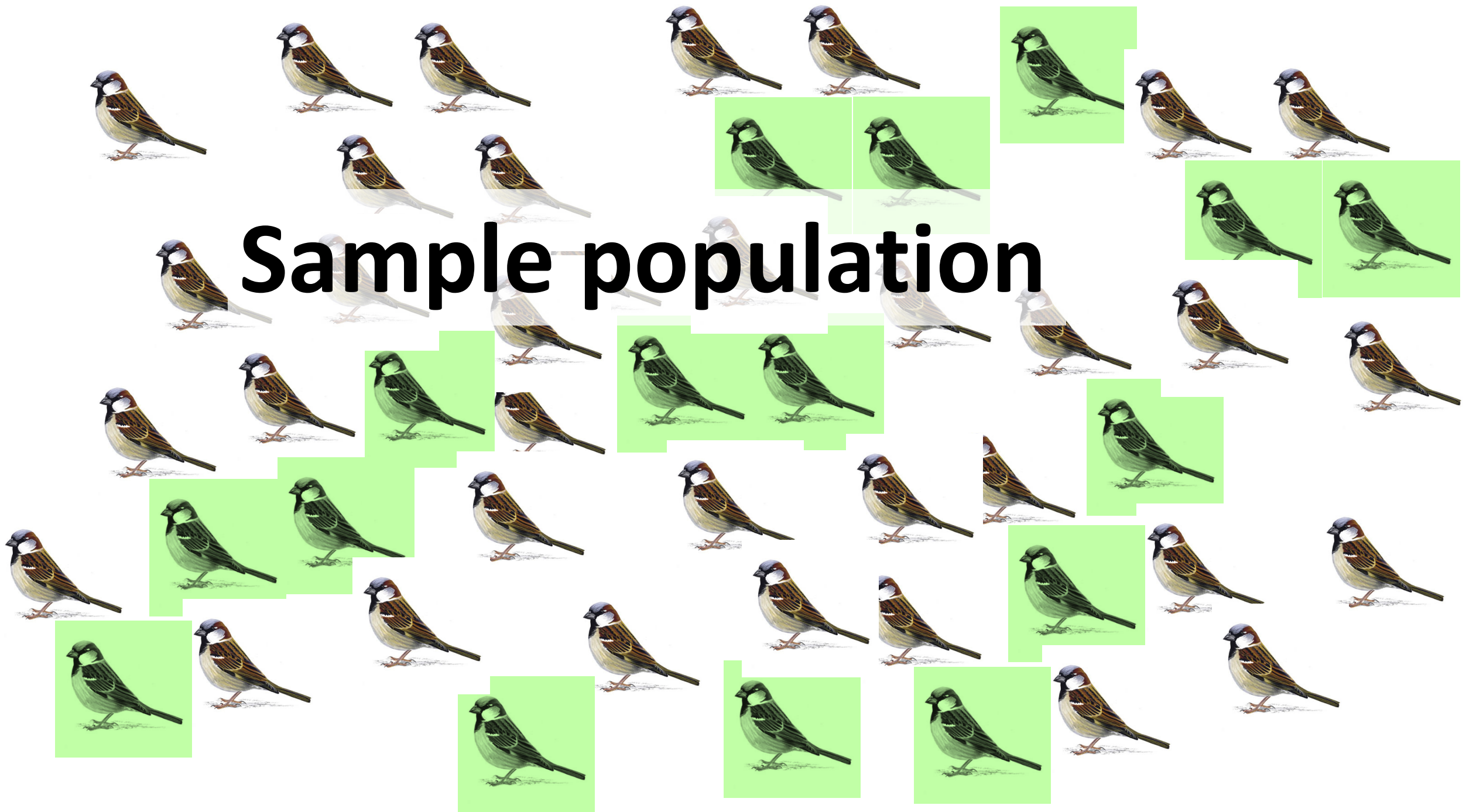


The image features a large number of small, detailed illustrations of sparrows. They are arranged in a circular pattern around the central text, with some birds appearing in the foreground and others in the background, creating a sense of depth. The birds are shown in various poses, some facing left and some facing right. The central text is a large, bold, black sans-serif font that reads "Complete population".

Complete population



Sample population

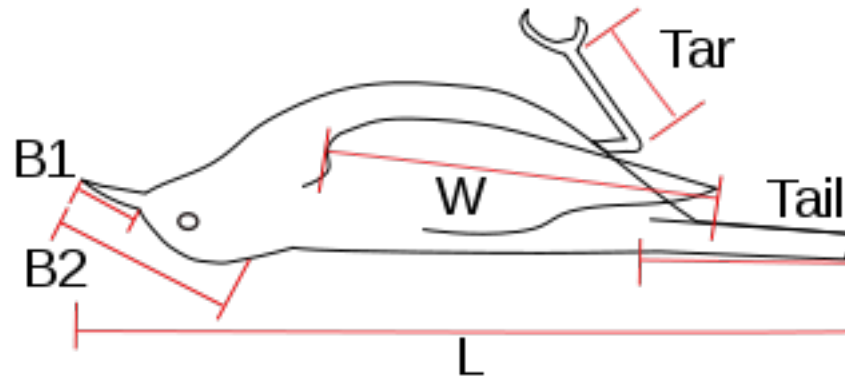


Basic statistical concepts

```
> head(d)
```

	BirdID	Cohort	CaptureDate	CaptureTime	Tarsus	Bill	Wing	Mass	Sex	Sex.1
1	4	2001	24-Jul-02	<NA>	16.9	NA	76	23.6	0	female
2	28	2001	22-Mar-02	<NA>	19.0	NA	77	26.2	0	female
3	29	2001	03-Jun-02	<NA>	18.5	NA	77	28.0	0	female
4	32	2001	11-Oct-01	<NA>	17.9	NA	75	28.1	0	female
5	32	2001	13-Aug-03	08:00	18.8	13.9	75	25.5	0	female
6	32	2001	09-May-04	12:00	18.9	13.9	76	25.6	0	female

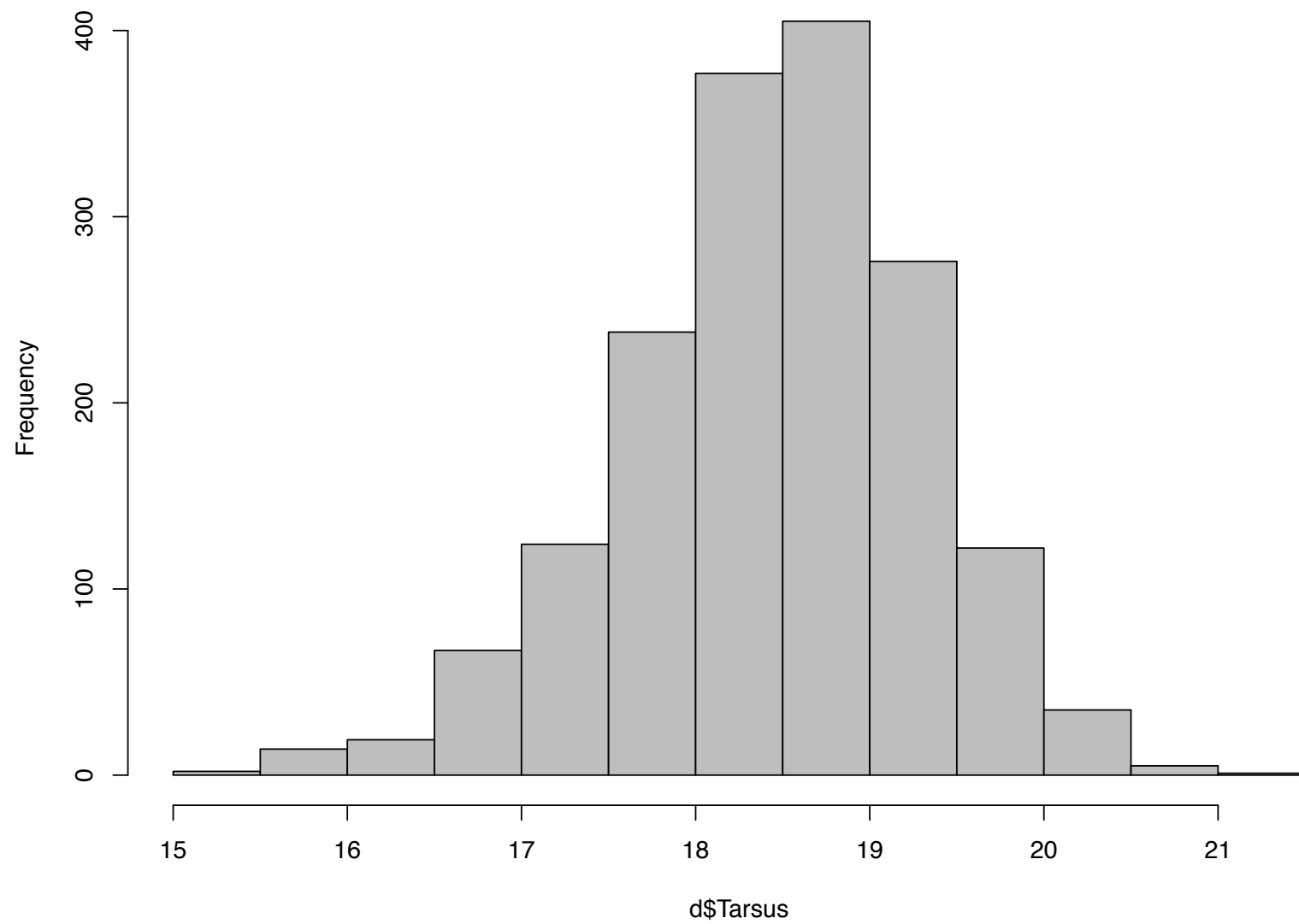
```
>
```



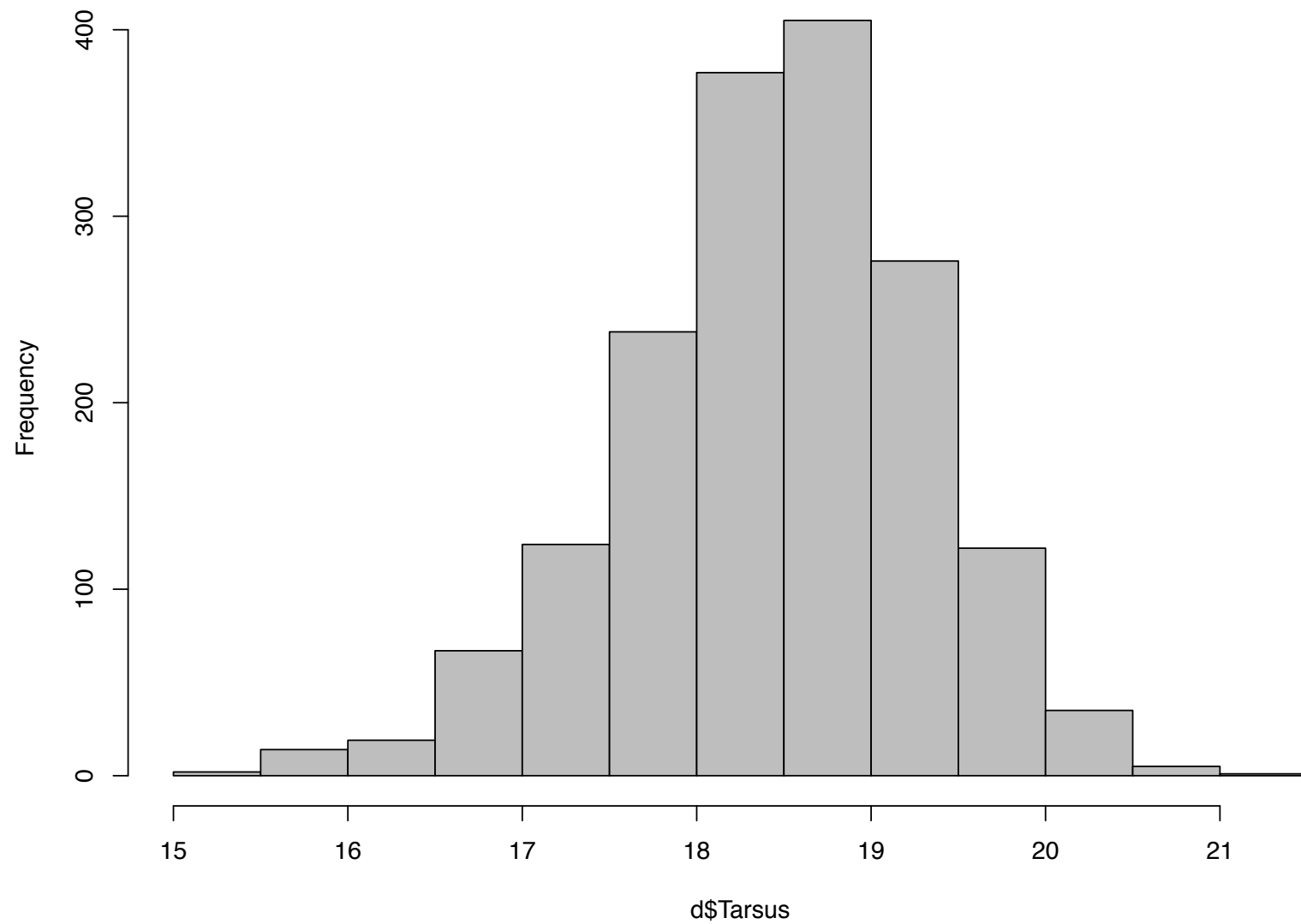
Basic statistical concepts



Histogram of d\$Tarsus



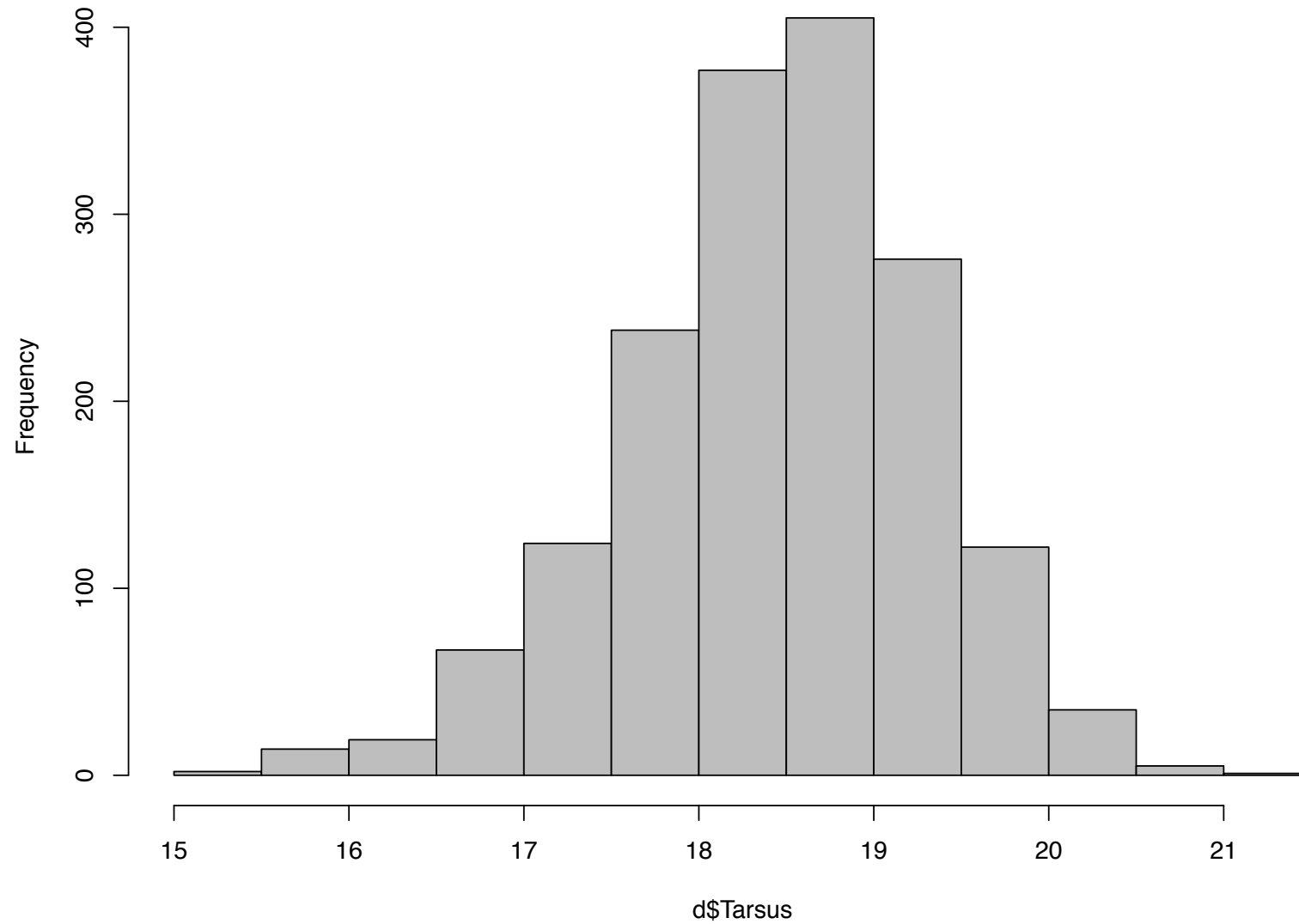
Histogram of d\$Tarsus



Describe data:

- Centrality
- Spread

Histogram of d\$Tarsus



Describe data:

- Centrality
- Spread

WHERE is the MIDDLE?

Mean, median and mode

Mean	$\frac{\text{Sum of values}}{\text{Number of values}}$
Median	Middle data value or midpoint of two middle values
Mode	Most frequent value(s)

Mean, median and mode

1	2	2	3	3	3	3	4	4	5
---	---	---	---	---	---	---	---	---	---

Mean =

Median =

Mode =

1	2	3	3	3	4	5	5	5	20
---	---	---	---	---	---	---	---	---	----

Mean =

Median =

Mode =

1.2	1.3	2.4	2.5	2.6	2.9	3.2	3.5	4.1	5.3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Mean =

Median =

Mode =

Mean, median and mode

1	2	2	3	3	3	3	4	4	5
---	---	---	---	---	---	---	---	---	---

Mean = 3

Median = 3

Mode = 3

1	2	3	3	3	4	5	5	5	20
---	---	---	---	---	---	---	---	---	----

Mean =

Median =

Mode =

1.2	1.3	2.4	2.5	2.6	2.9	3.2	3.5	4.1	5.3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Mean =

Median =

Mode =

Mean, median and mode

1	2	2	3	3	3	3	4	4	5
---	---	---	---	---	---	---	---	---	---

Mean = 3

Median = 3

Mode = 3

1	2	3	3	3	4	5	5	5	20
---	---	---	---	---	---	---	---	---	----

Mean = 5.1

Median = 3.5

Mode = 3 and 5

1.2	1.3	2.4	2.5	2.6	2.9	3.2	3.5	4.1	5.3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Mean =

Median =

Mode =

Mean, median and mode

1	2	2	3	3	3	3	4	4	5
---	---	---	---	---	---	---	---	---	---

Mean = 3

Median = 3

Mode = 3

1	2	3	3	3	4	5	5	5	20
---	---	---	---	---	---	---	---	---	----

Mean = 5.1

Median = 3.5

Mode = 3 and 5

1.2	1.3	2.4	2.5	2.6	2.9	3.2	3.5	4.1	5.3
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

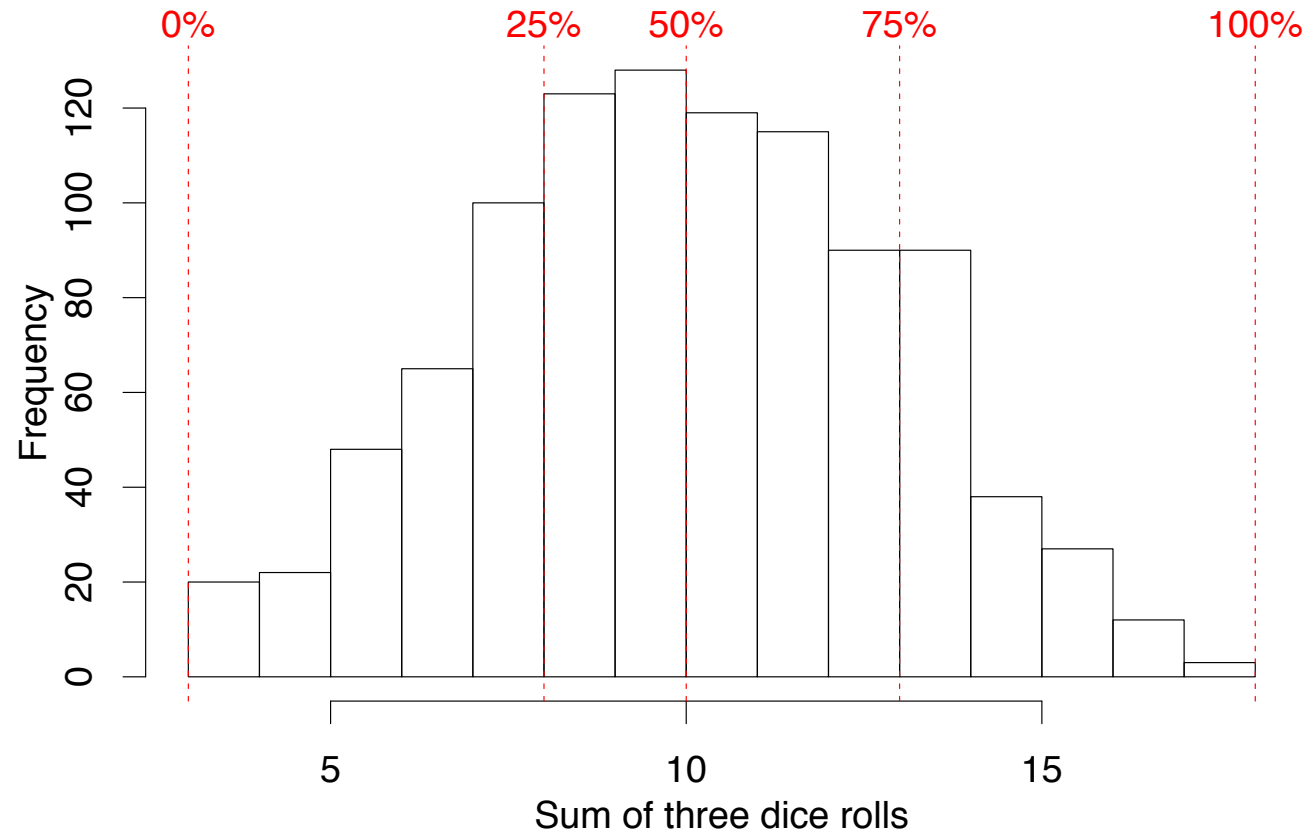
Mean = 2.9

Median = 2.75

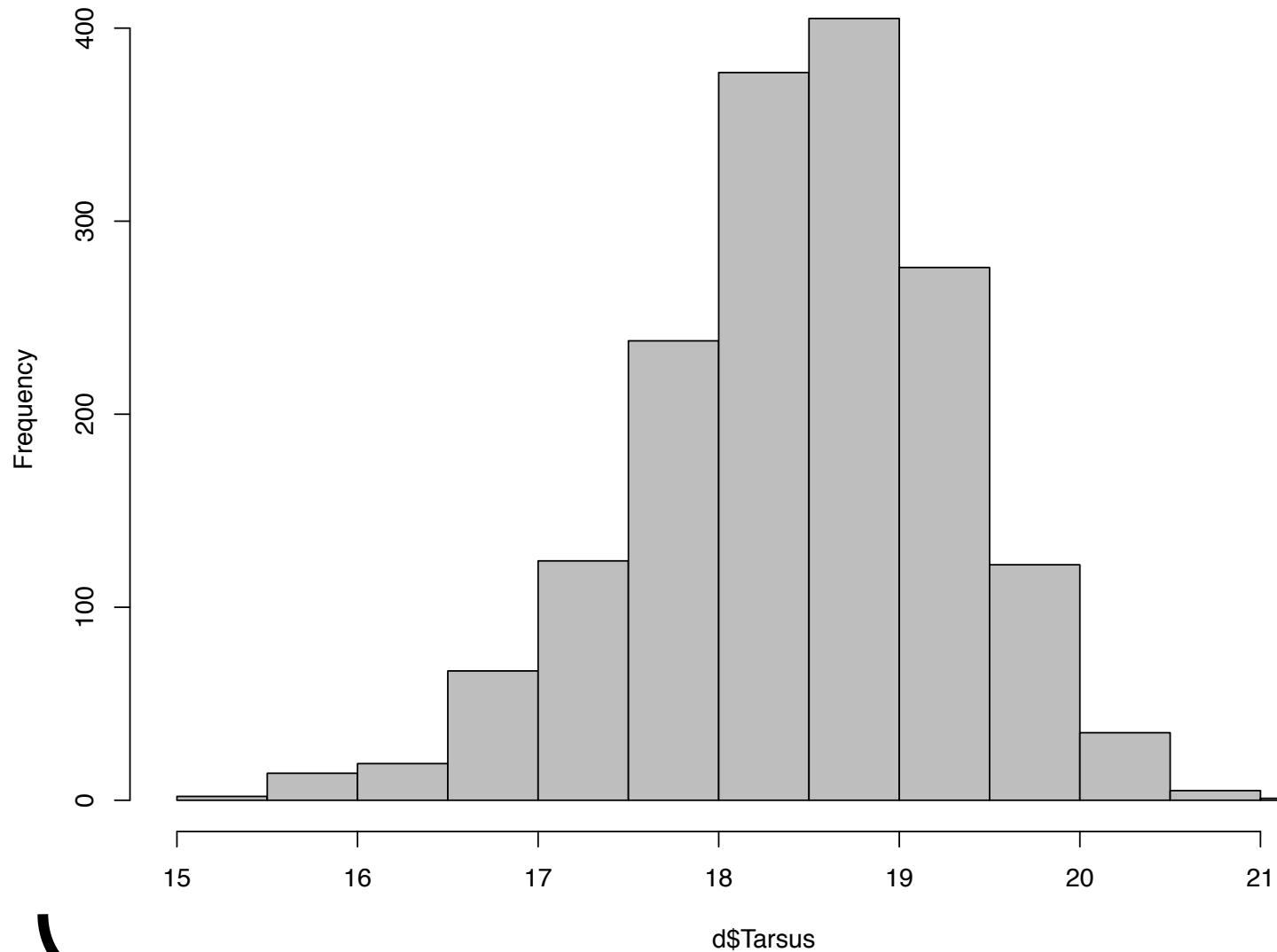
Mode = ?!*

Quantiles

The value which a given percent of the data are smaller than or equal to.



Histogram of d\$Tarsus



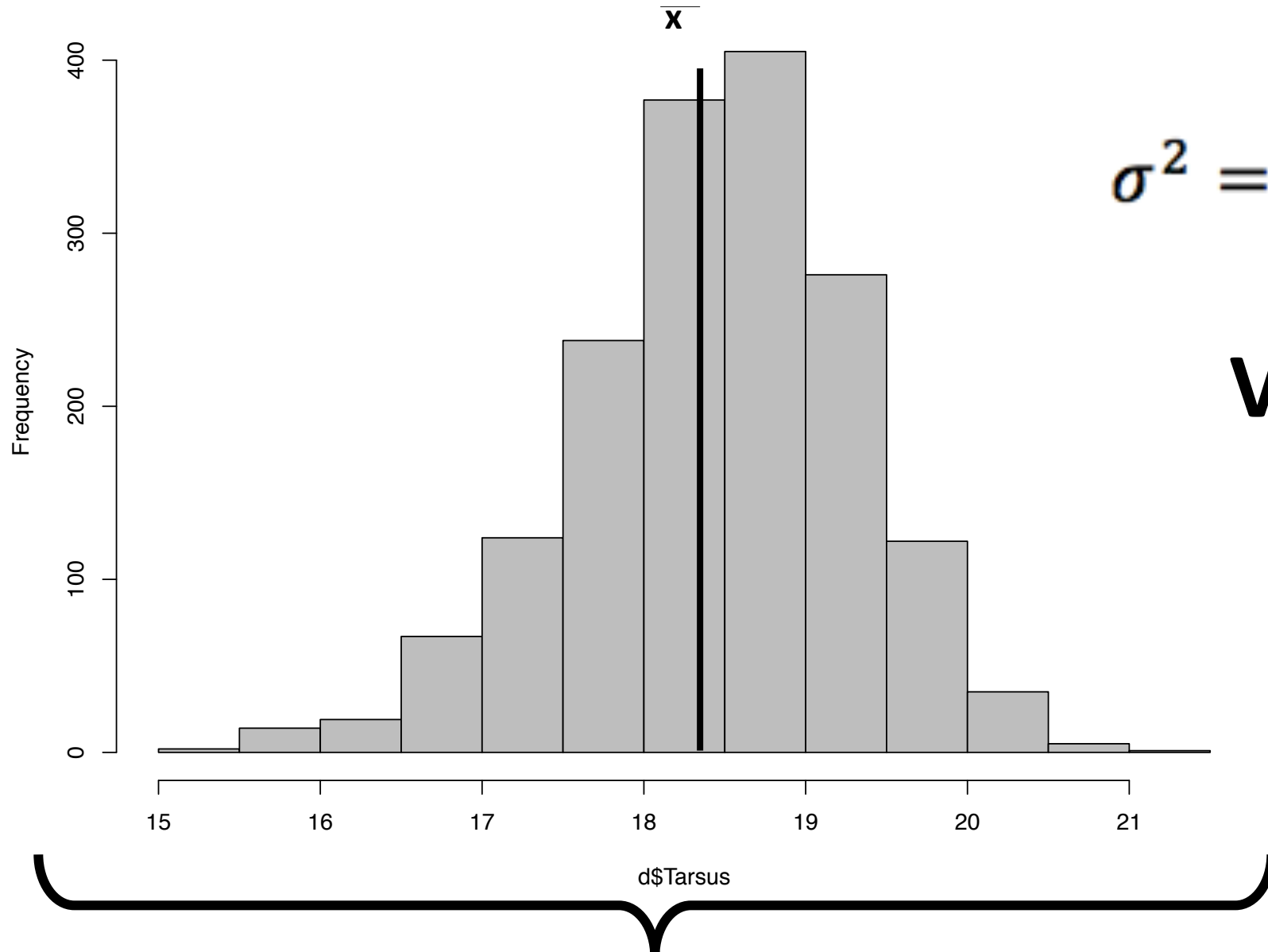
Describe data:

- Centrality
 - Mean
 - Median
 - Mode
- Spread
 - Range (min/max)
 - Variance

x_i

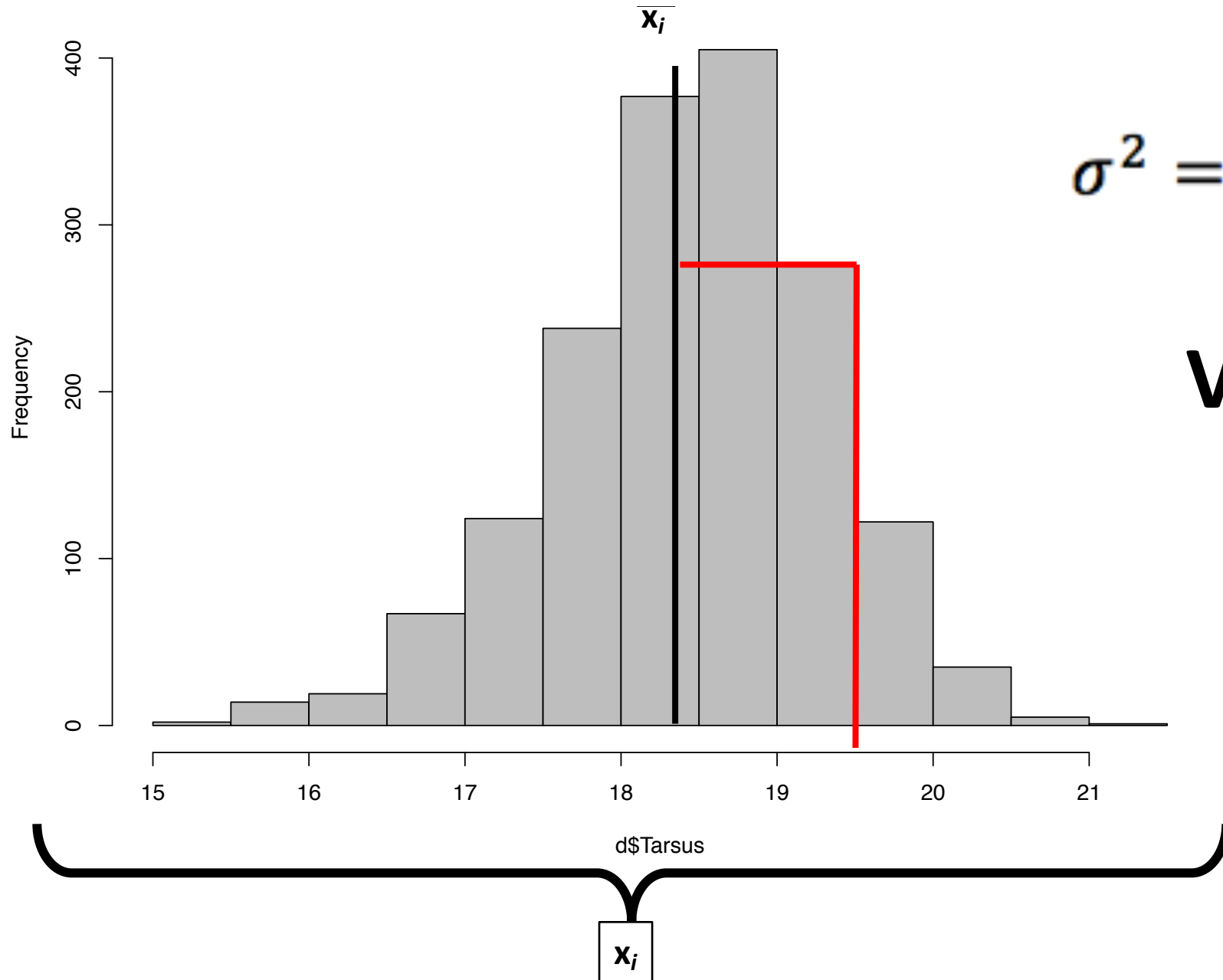
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



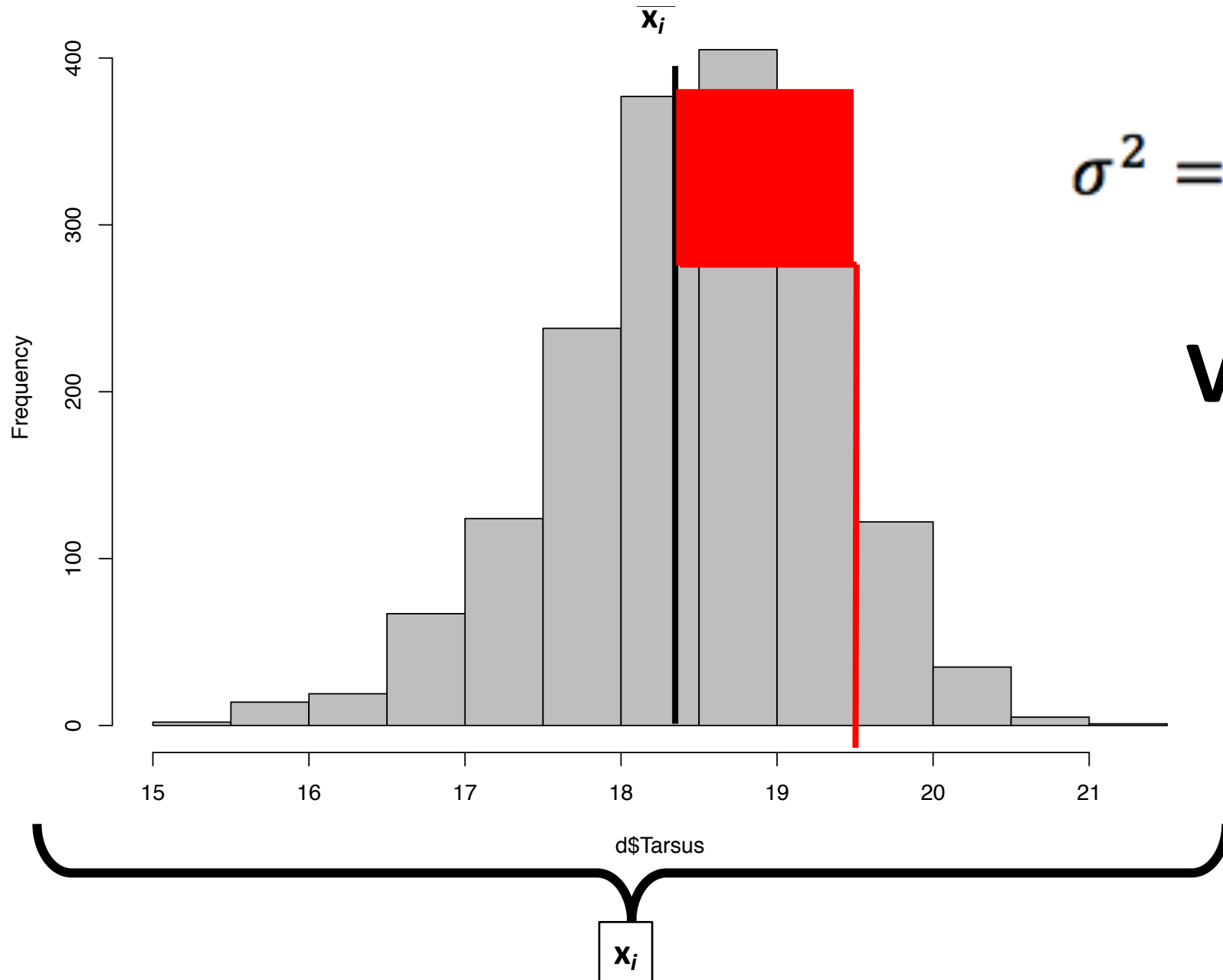
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



1.



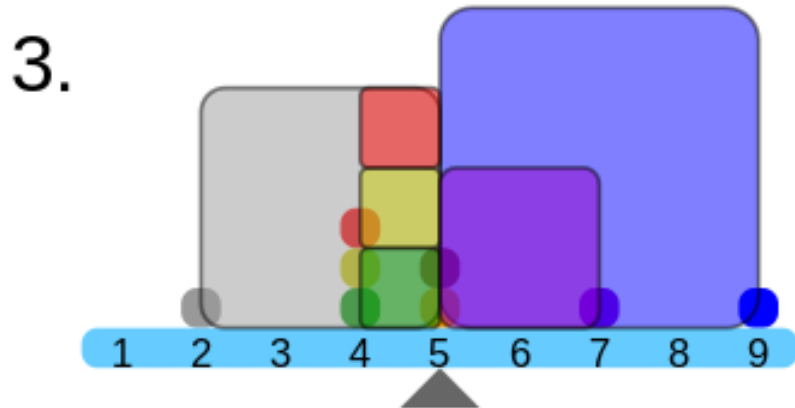
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



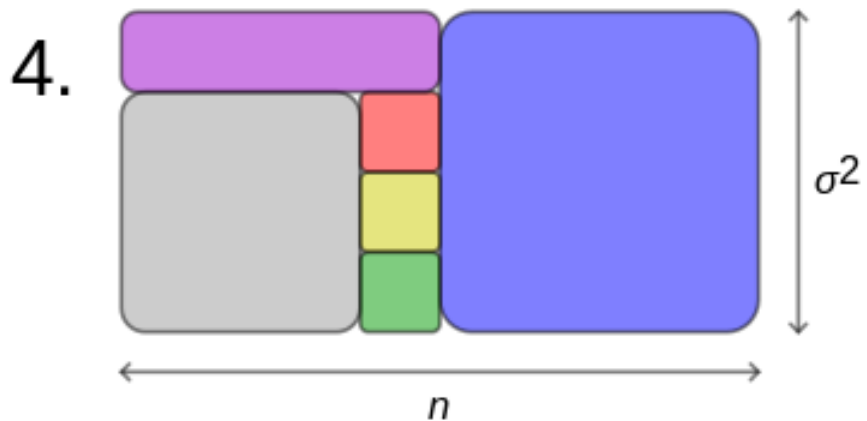
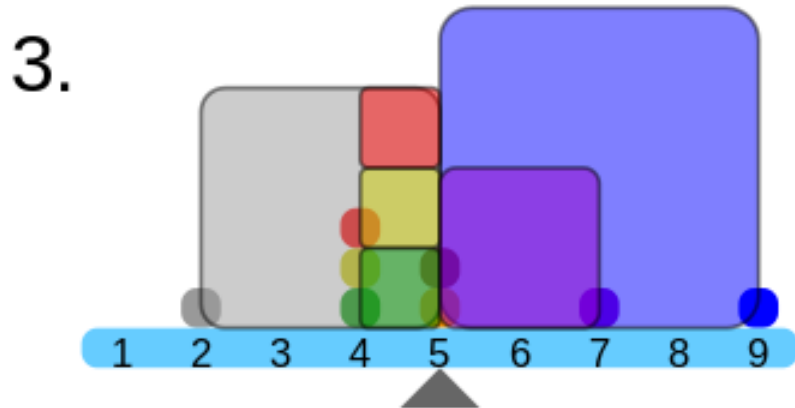
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



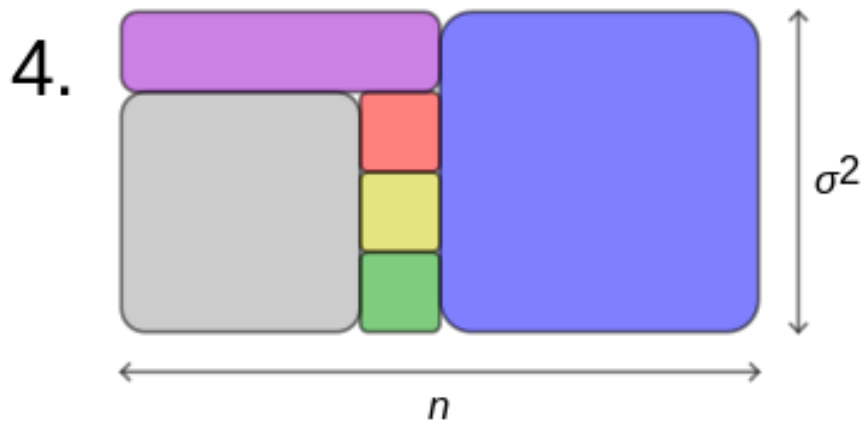
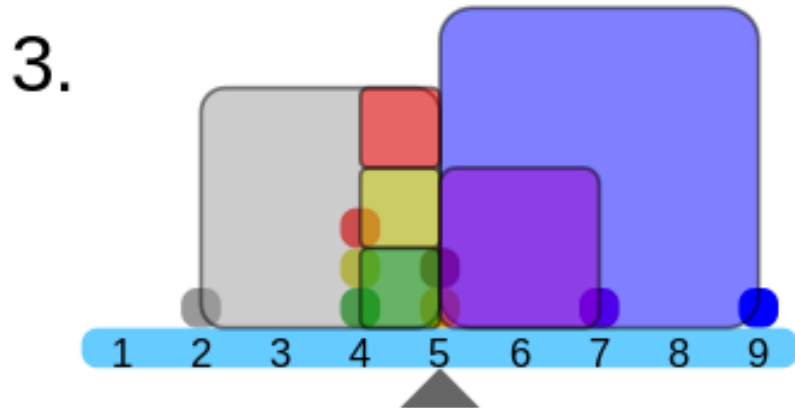
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE



$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

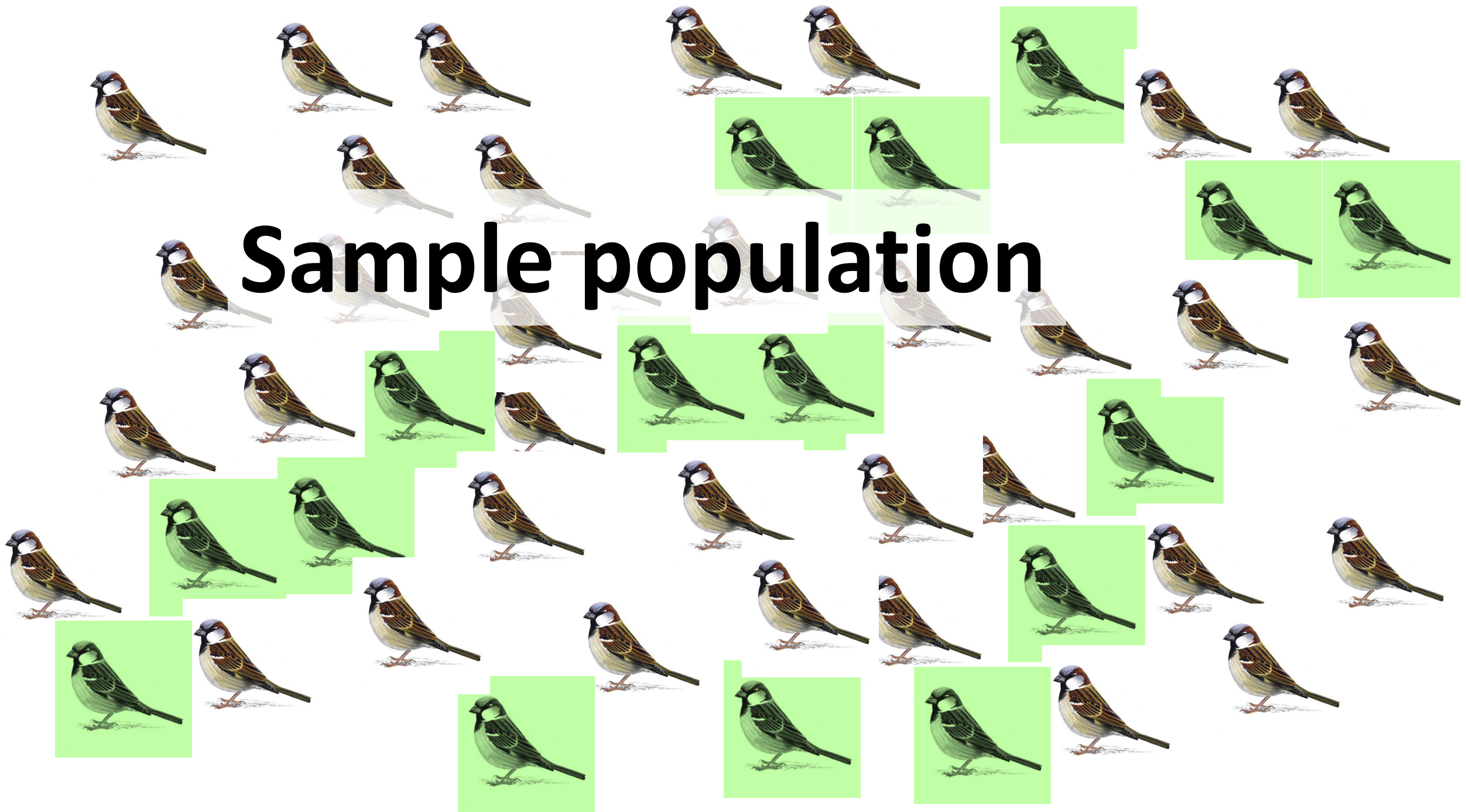
VARIANCE

Wait, WHAT?

The image features a large number of small, detailed illustrations of sparrows. They are arranged in a circular pattern around the central text, with some birds appearing in the foreground and others in the background, creating a sense of depth. The birds are shown in various poses, some facing left and some facing right. The central text is a large, bold, black sans-serif font that reads "Complete population".

Complete population

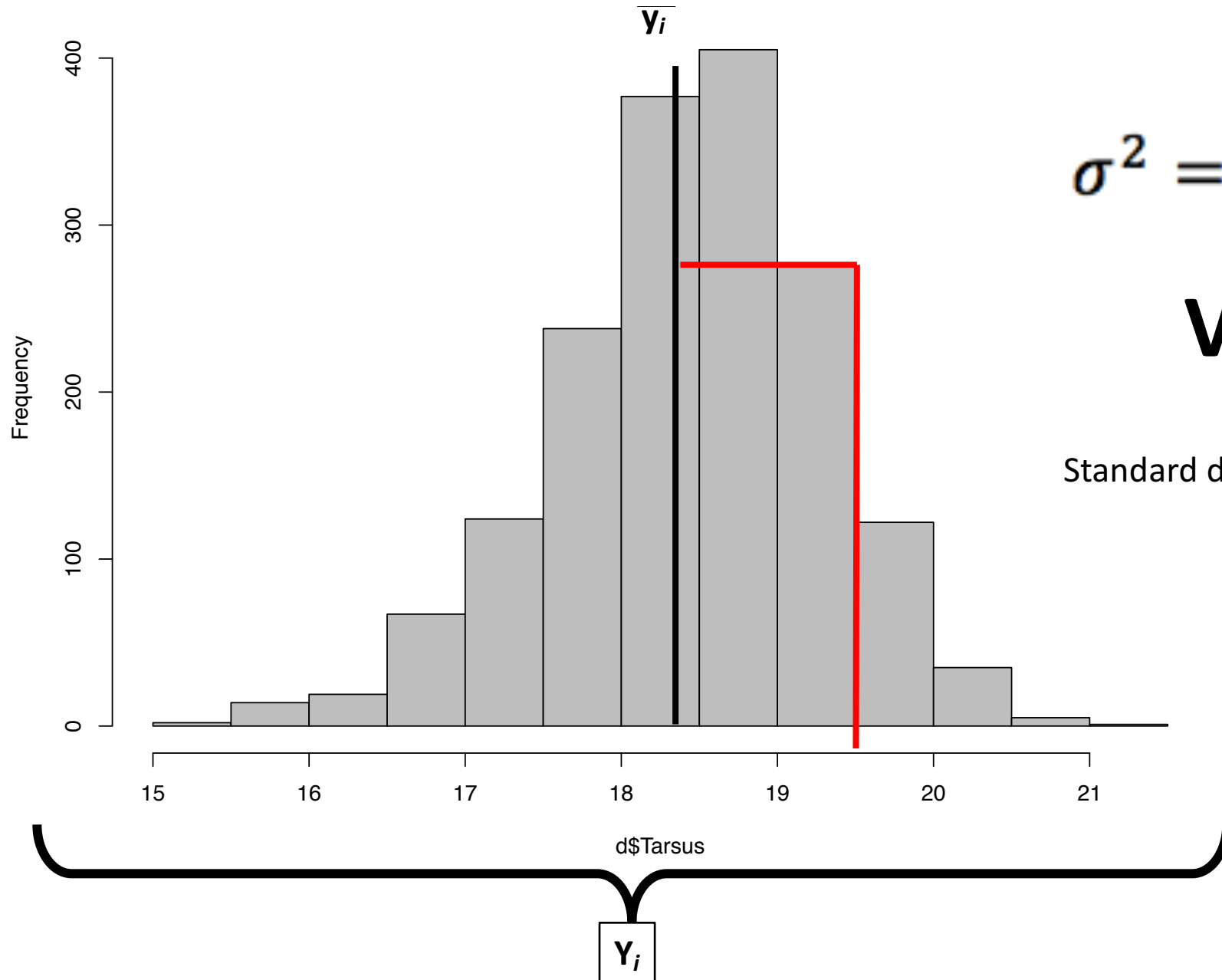
Sample population



$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

VARIANCE

Standard deviation is square root of variance



Standard deviation:

[stdev video](#)

<http://ow.ly/3vC030fVxUs>

Basic statistical concepts

```
> head(d)
```

	BirdID	Cohort	CaptureDate	CaptureTime	Tarsus	Bill	Wing	Mass	Sex	Sex.1
1	4	2001	24-Jul-02	<NA>	16.9	NA	76	23.6	0	female
2	28	2001	22-Mar-02	<NA>	19.0	NA	77	26.2	0	female
3	29	2001	03-Jun-02	<NA>	18.5	NA	77	28.0	0	female
4	32	2001	11-Oct-01	<NA>	17.9	NA	75	28.1	0	female
5	32	2001	13-Aug-03	08:00	18.8	13.9	75	25.5	0	female
6	32	2001	09-May-04	12:00	18.9	13.9	76	25.6	0	female

```
>
```

variance

0.74 ? ?

sd

0.86 ? ?

mean

18.52 ? ?

Learning aims

- Difference between population and sample
- Centrality: Mean, median, mode
- Spread: Range, quantiles, variance, standard deviation
- Sum of squares

Exercise – DO IT NOW – HO 2

Calculate mean, variance and standard deviation of

- Bill length
 - Body mass
 - Wing length in R.
-
- Plot all four histograms in a multi-panel figure
-
- What does it mean when statisticians talk about the “sum of squares” (often abbreviated to SS)?

Exercise – DO IT NOW

- What are NA?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `m1v (d2$Tarsus)`

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `m1v (d2$Tarsus)`
- What is the variance of normally distributed data with a mean of 0 and a standard deviation of 1?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `m1v (d2$Tarsus)`
- What is the variance of normally distributed data with a mean of 0 and a standard deviation of 1?
- What is the function in R to z-standardize data?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `m1v (d2$Tarsus)`
- What is the variance of normally distributed data with a mean of 0 and a standard deviation of 1?
- What is the function in R to z-standardize data?
- Explain what quantiles are

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histograms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `m1v (d2$Tarsus)`
- What is the variance of normally distributed data with a mean of 0 and a standard deviation of 1?
- What is the function in R to z-standardize data?
- Explain what quantiles are
- What are the boxes, whiskers and circles in a boxplot?

Exercise – DO IT NOW

- What are NA?
- How can we deal with missing values? There are two ways!
- Why are there odd gaps in the histogramms when we set breaks to larger numbers?
- What do these gaps tell us about how precise we should report results?
- Explain the warnings we got for `mlv(d2$Tarsus)`
- What is the variance of normally distributed data with a mean of 0 and a standard deviation of 1?
- What is the function in R to z-standardize data?
- Explain what quantiles are
- What are the boxes, whiskers and circles in a boxplot?
- Explain the terms: sums of squares, mean of sum of squares