

American University of Beirut
STAT 210: Elementary Statistics for Sciences
2022–2023 Fall

Siamak Taati

Chapter 3
Measures of center and variation
Part 1

Measures of center and variation

Descriptive measures

Scenario (from last week)

A group of zoologists gather data on a sample of size 50 from a species of birds living in Lebanon.

	sex	weight	tail.len	feather.col	beak.col	noct
1	F	98.80	7.50	yellow	red	0
2	M	95.80	7.10	dark blue	pink	0
3	F	92.50	6.80	yellow	pink	0
4	M	86.20	7.30	light blue	pink	0
5	F	87.30	6.40	green	red	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	F	98.00	7.50	yellow	pink	0

Descriptive measures

Scenario (from last week)

A group of zoologists gather data on a sample of size 50 from a species of birds living in Lebanon.

	sex	weight	tail.len	feather.col	beak.col	noct
1	F	98.80	7.50	yellow	red	0
2	M	95.80	7.10	dark blue	pink	0
3	F	92.50	6.80	yellow	pink	0
4	M	86.20	7.30	light blue	pink	0
5	F	87.30	6.40	green	red	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	F	98.00	7.50	yellow	pink	0

Q Is there a difference in weight between male and female birds?

Descriptive measures

Q Is there a difference in weight between male and female birds?

Weight of sampled female birds

98.8	92.5	87.3	103.5	98.3	99.1	96.9	103.0	96.5	93.4
102.9	101.1	100.9	95.6	96.1	99.3	93.3	95.4	88.5	97.6
92.2	98.0								

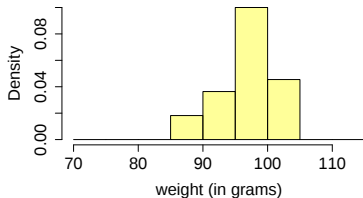
Weight of sampled male birds

95.8	86.2	92.3	88.5	94.6	102.8	98.9	99.2	87.0	93.3
89.7	96.5	95.9	93.7	97.0	96.4	108.0	83.4	91.1	96.6
92.6	79.7	90.8	96.8	92.3	101.8	92.3	91.1		

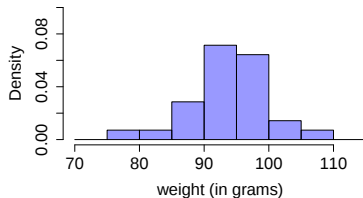
Note. The birds were sampled at random. By sheer chance, 22 out of 50 sampled birds were female and 28 were male.

Descriptive measures

Q Is there a difference in weight between male and female birds?



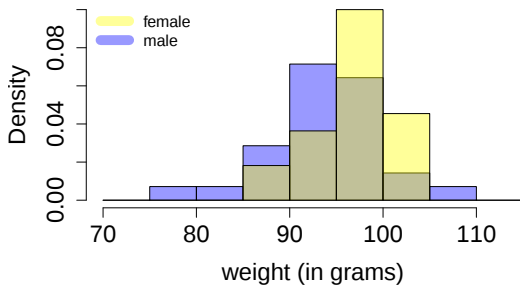
Weight of sampled female birds



Weight of sampled male birds

Descriptive measures

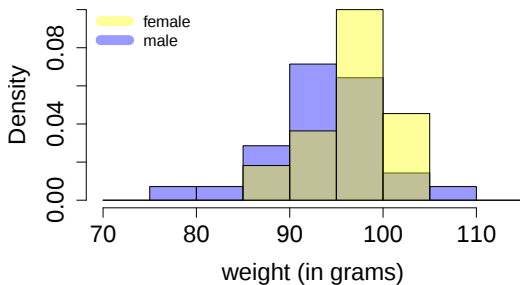
Q Is there a difference in weight between male and female birds?



Weight of sampled birds

Descriptive measures

Q Is there a difference in weight between male and female birds?

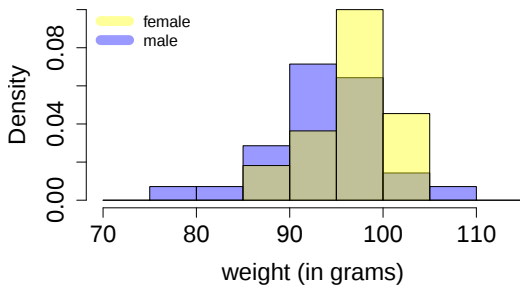


Weight of sampled birds

Among the sampled birds,

Descriptive measures

Q Is there a difference in weight between male and female birds?



Weight of sampled birds

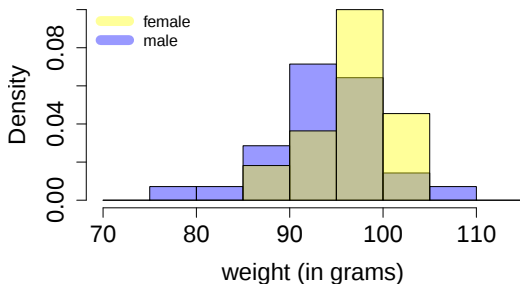
Among the sampled birds,

→ females seem to be typically heavier.

[center]

Descriptive measures

Q Is there a difference in weight between male and female birds?



Weight of sampled birds

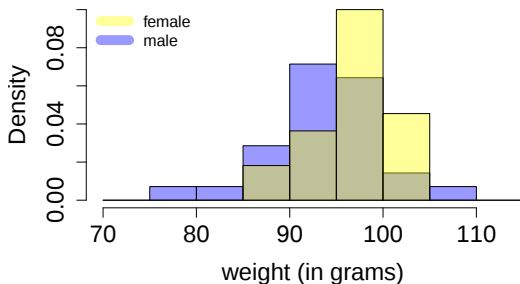
Among the sampled birds,

→ females seem to be typically heavier. [center]

→ the weight seems to vary more among males. [variation/spread]

Descriptive measures

Q Is there a difference in weight between male and female birds?



Weight of sampled birds

Among the sampled birds,

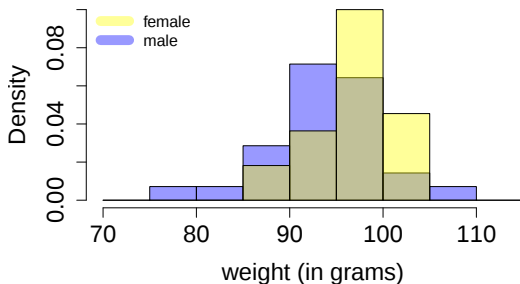
→ females seem to be typically heavier. [center]

→ the weight seems to vary more among males. [variation/spread]

Q Can we quantitatively **measure** these distinctions?

Descriptive measures

Q Is there a difference in weight between male and female birds?



Weight of sampled birds

Among the sampled birds,

→ females seem to be typically heavier. [center]

→ the weight seems to vary more among males. [variation/spread]

Q Can we quantitatively **measure** these distinctions?

Measures of center: mean

Weight of females

98.8 92.5 87.3 103.5 98.3 99.1 96.9 103.0 96.5 93.4
102.9 101.1 100.9 95.6 96.1 99.3 93.3 95.4 88.5 97.6
92.2 98.0

mean weight among sampled female birds

$$= \frac{98.8 + 92.5 + 87.3 + \cdots + 98.0}{22} \approx \boxed{96.83} \text{ grams}$$

Weight of males

95.8 86.2 92.3 88.5 94.6 102.8 98.9 99.2 87.0 93.3
89.7 96.5 95.9 93.7 97.0 96.4 108.0 83.4 91.1 96.6
92.6 79.7 90.8 96.8 92.3 101.8 92.3 91.1

mean weight among sampled male birds

$$= \frac{95.8 + 86.2 + 92.3 + \cdots + 91.1}{28} \approx \boxed{93.73} \text{ grams}$$

Measures of center: mean

mean weight among sampled female birds

$$= \frac{98.8 + 92.5 + 87.3 + \cdots + 98.0}{22} \approx \boxed{96.83} \text{ grams}$$

mean weight among sampled male birds

$$= \frac{95.8 + 86.2 + 92.3 + \cdots + 91.1}{28} \approx \boxed{93.73} \text{ grams}$$

That $96.83 > 93.73$ is a quantitative indication that within the sampled birds, the “females were typically heavier than the males”.

Measures of center: mean

mean weight among sampled female birds

$$= \frac{98.8 + 92.5 + 87.3 + \cdots + 98.0}{22} \approx \boxed{96.83} \text{ grams}$$

mean weight among sampled male birds

$$= \frac{95.8 + 86.2 + 92.3 + \cdots + 91.1}{28} \approx \boxed{93.73} \text{ grams}$$

That $96.83 > 93.73$ is a quantitative indication that within the sampled birds, the “females were typically heavier than the males”.

- Q Does this mean that within the entire population, the “females are typically heavier than the males”?

Measures of center: mean

Sample vs. population mean

For a numerical variable x ,

population mean of x :	$\mu := \frac{\text{sum of population values of } x}{\text{population size}}$
sample mean of x :	$\bar{x} := \frac{\text{sum of sample values of } x}{\text{sample size}}$

Measures of center: mean

Sample vs. population mean

For a numerical variable x ,

population mean of x :	$\mu := \frac{\sum x}{N}$	(N : population size)
sample mean of x :	$\bar{x} := \frac{\sum x}{n}$	(n : sample size)

Measures of center: mean

Sample vs. population mean

For a numerical variable x ,

$$\begin{array}{lll} \text{population mean of } x : & \mu := \frac{\sum x}{N} & (N: \text{population size}) \\ \text{sample mean of } x : & \bar{x} := \frac{\sum x}{n} & (n: \text{sample size}) \end{array}$$

- ▶ The **population mean** (a.k.a. the **true mean**) is a **parameter** of the population. It is often unknown to us, unless we have access to the entire population (e.g., the census data).

Measures of center: mean

Sample vs. population mean

For a numerical variable x ,

$$\begin{array}{ll} \text{population mean of } x : & \mu := \frac{\sum x}{N} \quad (N: \text{population size}) \\ \text{sample mean of } x : & \bar{x} := \frac{\sum x}{n} \quad (n: \text{sample size}) \end{array}$$

- ▶ The **population mean** (a.k.a. the **true mean**) is a **parameter** of the population. It is often unknown to us, unless we have access to the entire population (e.g., the census data).
- ▶ The **sample mean** is an example of a **statistic** (i.e., a quantity calculated based on the sample). In particular, the sample mean varies from sample to sample.

Measures of center: mean

Sample vs. population mean

For a numerical variable x ,

$$\begin{array}{ll} \text{population mean of } x : & \mu := \frac{\sum x}{N} \quad (N: \text{ population size}) \\ \text{sample mean of } x : & \bar{x} := \frac{\sum x}{n} \quad (n: \text{ sample size}) \end{array}$$

- ▶ Nevertheless, the **sample mean** can be used to **estimate** the **population mean**. The larger the sample, the more accurate the estimate (i.e., the less likely that the error is large).

[We will come back to this ...]

Measures of center: mean

Back to the weight of the birds ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

Measures of center: mean

Back to the weight of the birds ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

- Q Does the fact that **among the sampled birds**, the females were on average heavier than males imply that **among the entire population**, the females are on average heavier than males?

Measures of center: mean

Back to the weight of the birds ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

Q Does the fact that **among the sampled birds**, the females were on average heavier than males imply that **among the entire population**, the females are on average heavier than males?

→ Not necessarily! The difference could have occurred due to chance.

Measures of center: mean

Back to the weight of the **birds** ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

Ⓚ Does the fact that **among the sampled birds**, the females were on average heavier than males imply that **among the entire population**, the females are on average heavier than males?

- Not necessarily! The difference could have occurred due to chance.
- But it *does* suggest so!

Measures of center: mean

Back to the weight of the **birds** ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

Q Does the fact that **among the sampled birds**, the females were on average heavier than males imply that **among the entire population**, the females are on average heavier than males?

→ Not necessarily! The difference could have occurred due to chance.

→ But it *does* suggest so!

Q Can we make the conclusion with a fair degree of confidence?

Measures of center: mean

Back to the weight of the **birds** ...

sample mean weight for females $\approx$ 96.83 grams

sample mean weight for males $\approx$ 93.73 grams

Ⓚ Does the fact that **among the sampled birds**, the females were on average heavier than males imply that **among the entire population**, the females are on average heavier than males?

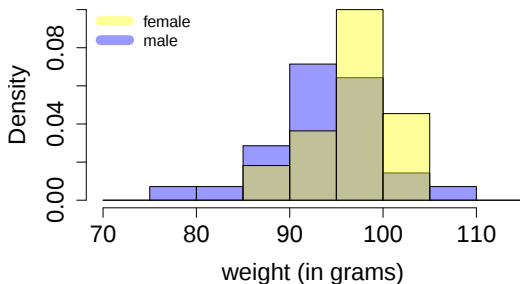
→ Not necessarily! The difference could have occurred due to chance.

→ But it *does* suggest so!

Ⓚ Can we make the conclusion with a fair degree of confidence?
[We will come back to this ...]

Descriptive measures

Q Is there a difference in weight between male and female birds?



Weight of sampled birds

Among the sampled birds,

→ females seem to be typically heavier. [center]

→ the weight seems to vary more among males. [variation/spread]

Q Can we quantitatively **measure** these distinctions?

Measures of center: median

Weight of females

98.8	92.5	87.3	103.5	98.3	99.1	96.9	103.0	96.5	93.4
102.9	101.1	100.9	95.6	96.1	99.3	93.3	95.4	88.5	97.6
92.2	98.0								

Weight of males

95.8	86.2	92.3	88.5	94.6	102.8	98.9	99.2	87.0	93.3
89.7	96.5	95.9	93.7	97.0	96.4	108.0	83.4	91.1	96.6
92.6	79.7	90.8	96.8	92.3	101.8	92.3	91.1		

Measures of center: median

Weight of females (sorted)

87.3	88.5	92.2	92.5	93.3	93.4	95.4	95.6	96.1	96.5
96.9	97.6	98.0	98.3	98.8	99.1	99.3	100.9	101.1	102.9
103.0	103.5								

Weight of males (sorted)

79.7	83.4	86.2	87.0	88.5	89.7	90.8	91.1	91.1	92.3
92.3	92.3	92.6	93.3	93.7	94.6	95.8	95.9	96.4	96.5
96.6	96.8	97.0	98.9	99.2	101.8	102.8	108.0		

Measures of center: median

Weight of females (sorted)

87.3	88.5	92.2	92.5	93.3	93.4	95.4	95.6	96.1	96.5
96.9	97.6	98.0	98.3	98.8	99.1	99.3	100.9	101.1	102.9
103.0	103.5								

Weight of males (sorted)

79.7	83.4	86.2	87.0	88.5	89.7	90.8	91.1	91.1	92.3
92.3	92.3	92.6	93.3	93.7	94.6	95.8	95.9	96.4	96.5
96.6	96.8	97.0	98.9	99.2	101.8	102.8	108.0		

Measures of center: median

Weight of females (sorted)

87.3	88.5	92.2	92.5	93.3	93.4	95.4	95.6	96.1	96.5
96.9	97.6	98.0	98.3	98.8	99.1	99.3	100.9	101.1	102.9
103.0	103.5								

median of weight for sampled female birds

$$= \frac{96.9 + 97.6}{2} = \boxed{97.25} \text{ grams}$$

Weight of males (sorted)

79.7	83.4	86.2	87.0	88.5	89.7	90.8	91.1	91.1	92.3
92.3	92.3	92.6	93.3	93.7	94.6	95.8	95.9	96.4	96.5
96.6	96.8	97.0	98.9	99.2	101.8	102.8	108.0		

median of weight for sampled male birds

$$= \frac{93.3 + 93.7}{2} = \boxed{93.5} \text{ grams}$$

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

1, 4, 8, 3, 6

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

1, 4, 8, 3, 6

Answer:

1, 3, 4, 6, 8 (sorted)

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

1, 4, 8, 3, 6

Answer:

1, 3, 4, 6, 8 (sorted)

median = 4

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

4, 7, 4, 7, 5, 10

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

4, 7, 4, 7, 5, 10

Answer:

4, 4, 5, 7, 7, 10 (sorted)

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Example

What is the median of the following list of values?

4, 7, 4, 7, 5, 10

Answer:

4, 4, 5, 7, 7, 10 (sorted)

$$\text{median} = \frac{5 + 7}{2} = 6$$

Measures of center: median

The **median** of a list of data values

$$x_1, x_2, x_3, \dots, x_n$$

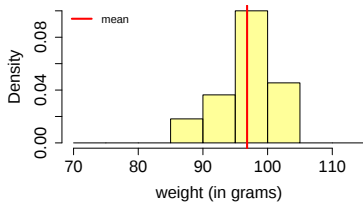
is their “mid-point”. More specifically, once we sort the values from smallest to largest, the median is

- ▶ the middle value (if n is odd), or
- ▶ the average of the two middle values (if n is even).

Sample vs. population median

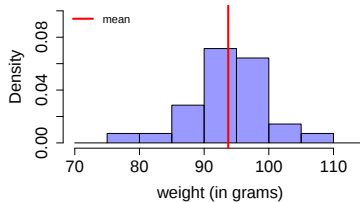
- ▶ The **population median** (a.k.a. the **true median**) is a **parameter** of the population.
- ▶ The **sample median** is an example of a **statistic** (i.e., a quantity calculated based on the sample).

Measures of center: mean vs. median



Weight of sampled female birds

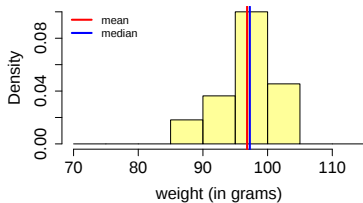
mean ≈ 96.83 grams



Weight of sampled male birds

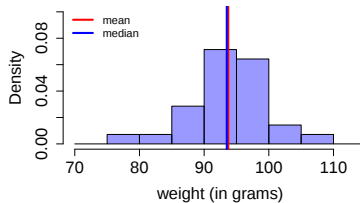
mean ≈ 93.73 grams

Measures of center: mean vs. median



Weight of sampled female birds

mean ≈ 96.83 grams
median = 97.25 grams



Weight of sampled male birds

mean ≈ 93.73 grams
median = 93.5 grams