

Measures of location

Ľubica and Ján Krauskovi, Dominik Heger

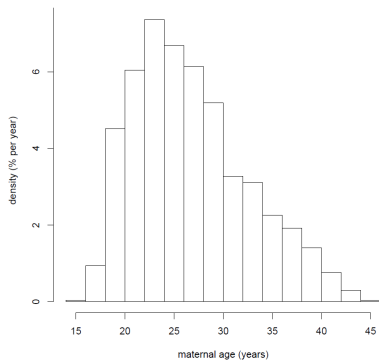
Masaryk University

hegerd@chemi.muni.cz

STDT 03 Central tendencies

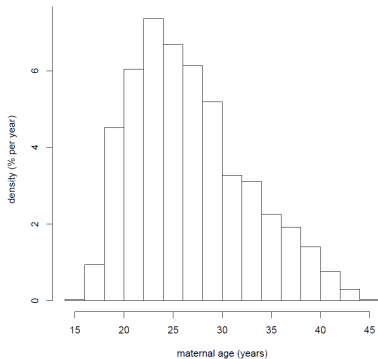
Examples of Distributions

Maternal ages

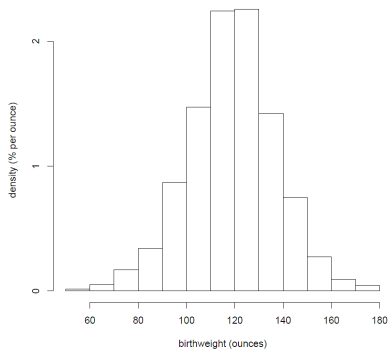


Examples of Distributions

Maternal ages

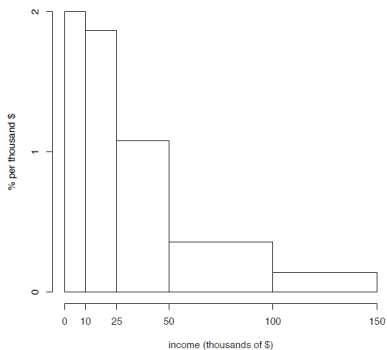


Birthweights of their babies



A right-skewed distribution

Income in USA



Formal definition of **percentile**

The p^{th} percentile of a list of numbers is the smallest number that is at least as large as p % of the list.

P-tý percentyl z řady čísel je takové nejmenší číslo, které je nejméně tak velké jako p % čísel v řadě.

Mechanism for median finding:

- order the list
- identify the place that is p % of the way from the bottom of the list
- if that's a place on the list, take the number in that place, if not, take the next one up

Average

Average is a smoother.

Average

Average is a smoother.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Average

Average is a smoother.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{1}{n} \cdot x_i = \sum_{i=1}^n \frac{x_i}{n}$$

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** × **cross-sectional**

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** × **cross-sectional**

The men in various categories are not the same.

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** × **cross-sectional**

The men in various categories are not the same.

The older men were shorter, on average.

Comparing averages

age (years)	20-30	30-40	40-50	50-60	60-75	75+
average height(")	69.3	69.5	69.4	69.2	68.3	67.2

Intervals include the left endpoint but not the right.

[National Health and Nutrition Examination Survey, 1999-2002]

Is this table telling us that as men get older, on average they get a bit taller and then get shorter?

Data - **longitudinal** × **cross-sectional**

The men in various categories are not the same.

The older men were shorter, on average.

When comparing averages first think:

- 1 How are the groups related to each other?
- 2 Take a look on the numerical averages.

What does an average tells us?

If a student's test score is above average, is the student in the top half of the class?

What does an average tells us?

If a student's test score is above average, is the student in the top half of the class?

Not necessarily.

Markov's inequality

If a list has only non-negative entries, then the proportion of entries that are **at least as large as** k times the average is **at most** $1/k$.

If a list has only non-negative entries, then the proportion of entries that are **greater than or equal to** k times the average is **no more than** $1/k$.

Jestliže řada obsahuje pouze kladné členy, pak poměr členů, které jsou **nejméně tak velké jako** k -krát průměr, je **nejvýše roven** $1/k$.

https://courses.edx.org/courses/BerkeleyX/Stat_2.1x/