Quantitative methods

Lesson 6

Daróczi Gergely

Corvinus University of Budapest, Hungary

2011 March 22

Outline

1. Repeating
   - Sampling theory
   - Probability sampling
   - Nonprobability sampling
2. Computation
   - Required formulas
   - Standard error
   - A basic example
   - Comparison of samples
   - Standard error in finite population
3. Final examination questions

Sampling theory

Definition

Sampling is the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen.

Elements:

1. population,
2. respondents, units of analysis,
3. sampling frame,
4. sampling methods.

Sampling methods - Probability sampling

A short summary

Probability sampling:

1. Simple Random Sampling,
2. Stratified Random Sampling,
3. Systematic Random Sampling,
4. Cluster (Area) Random Sampling,
5. Multi-Stage Sampling.

Sampling methods - Nonprobability sampling

A short summary

Nonprobability sampling:

1. Accidental, Haphazard or Convenience Sampling,
2. Purposive Sampling:
    1. Modal Instance Sampling,
    2. Expert Sampling,
    3. Quota Sampling:
        1. Proportional Quota Sampling,
        2. Nonproportional Quota Sampling.
    4. Heterogeneity Sampling,
    5. Snowball Sampling.

Computation

Required formulas

For Simple Random Sampling:

- mean: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
- standard deviation: $\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}}$
- standard error: $SE = \frac{\sigma}{\sqrt{n}} \cdot FPC$
- Finite Population Correction: if sampling fraction is large (>5%)

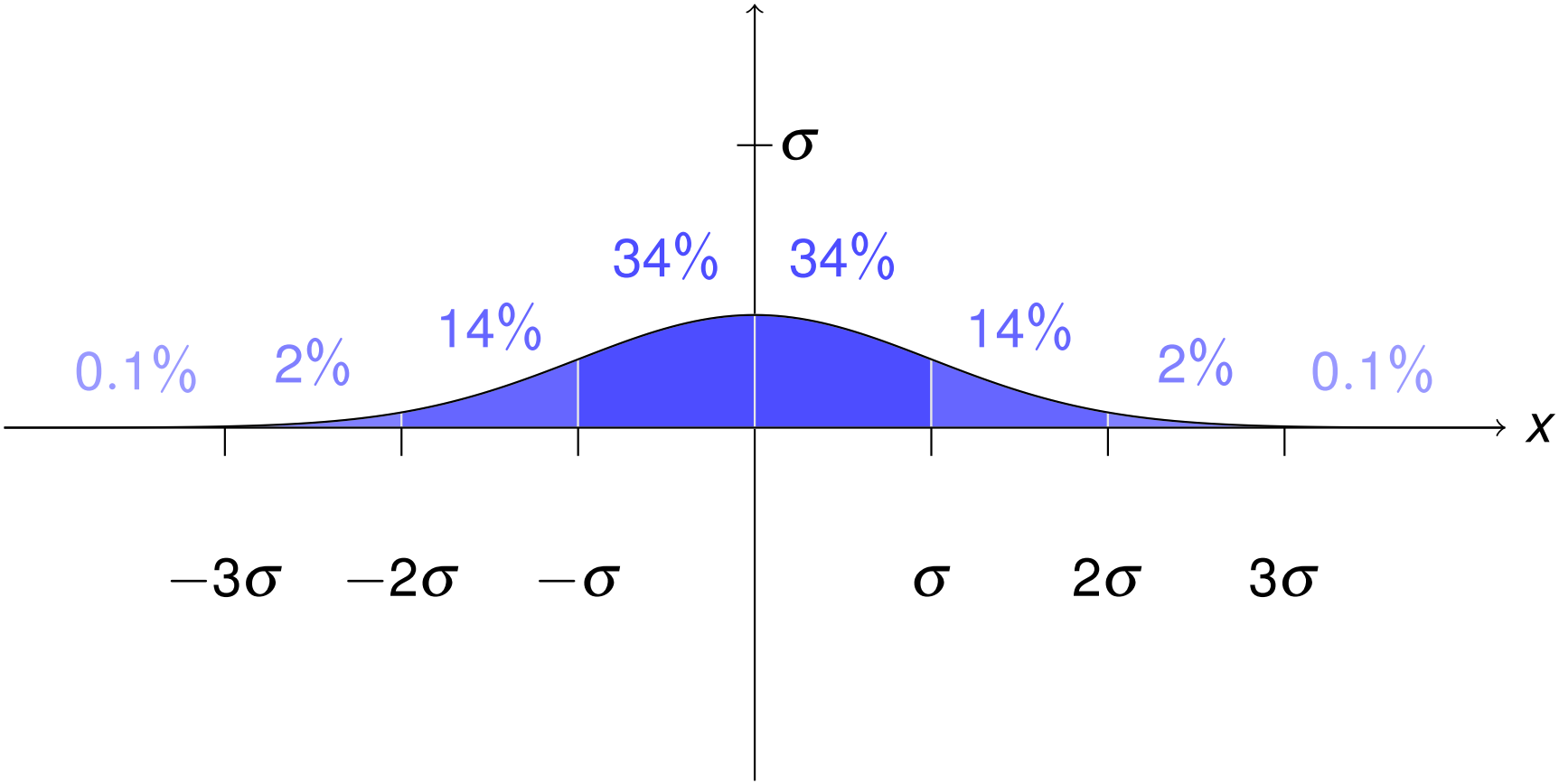
$$FPC = \sqrt{1 - \frac{n}{N}}$$

$$SE = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

Computation

A short summary on Standard error

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$



standard normal distribution: $\overline{x} = 0, \sigma = 1$

Computation

A basic example

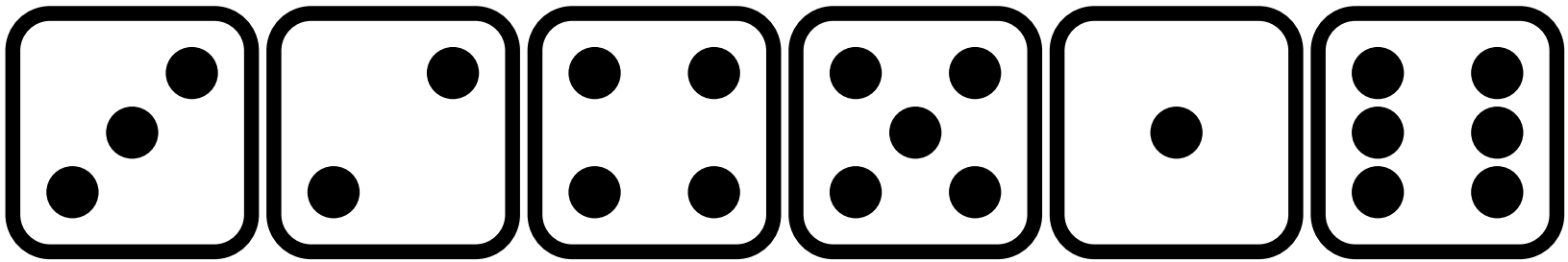
Game rules

Roll the dice!

If the result is even, the player wins the rolled value in dollars.

If the result is odd, the playes pays 2 dollars to the bank.

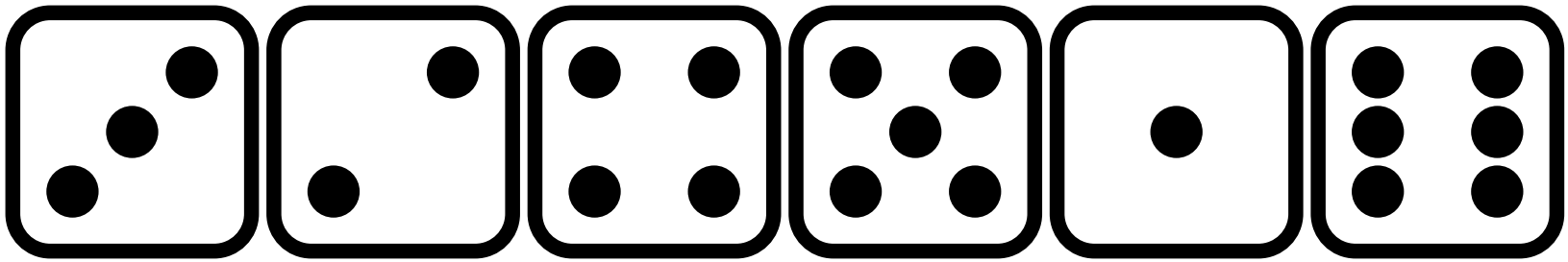
After rolling the below values, what would you think about the expected value of the game?



Would you continue playing?

Computation

Results



$$X = \{-2, 2, 4, -2, -2, 6\}$$

$$\overline{x} = \frac{-2+2+4+2+2+6}{6} = \frac{6}{6} = \frac{1}{1} = 1$$

$$\sigma = \sqrt{\frac{(-2-1)^2+(2-1)^2+(4-1)^1+(-2-1)^1+(-2-1)^2+(6-1)^2}{5}} =$$

$$= \sqrt{\frac{9+1+9+9+9+25}{5}} = \sqrt{\frac{62}{5}} = \sqrt{12.4} = 3.521363$$

$$SE = \frac{3.521363}{\sqrt{6}} = \frac{3.521363}{2.44949} = 1.437591$$

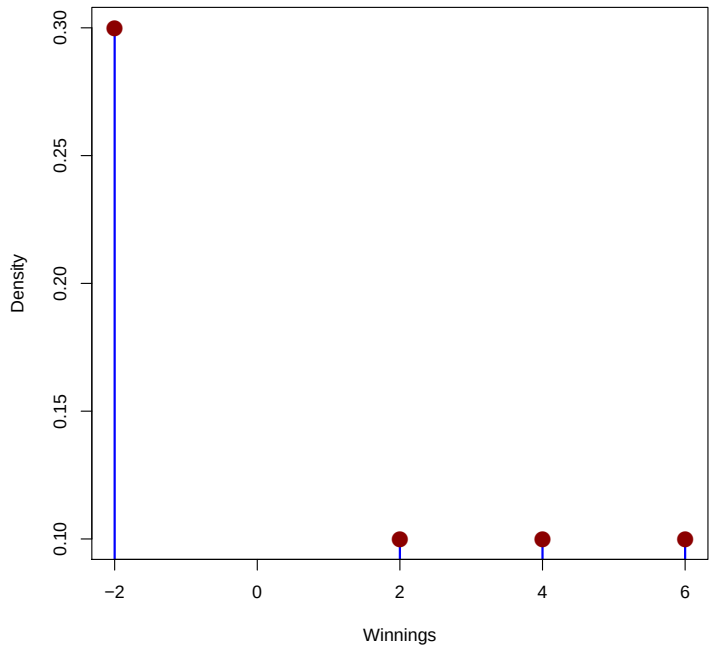
The expected value can vary between -1.87 and 3.87 at 95% CI.

Good luck!

Computation

Theoretical solution

Forget about the experiment and try to determine the **real** expected value of the game!



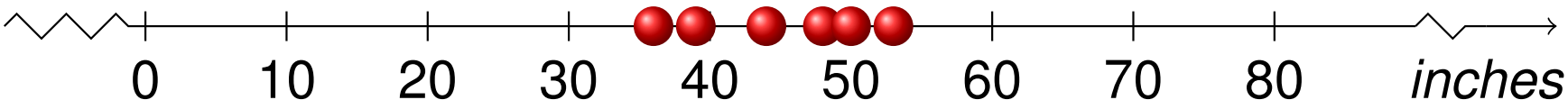
What is wrong with the above plot?

Computation

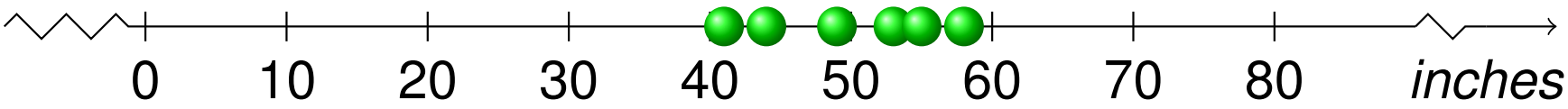
Comparison of samples

The height, in inches, of six trees at a nursery are shown at the specificed dates.
Find the mean, standard deviation and standard error of the heights!
Is there a significant difference between the means of samples?

1. **2011 March 22**: 36 48 50 44 53 39



2. **2011 April 1**: 41 53 55 49 58 44

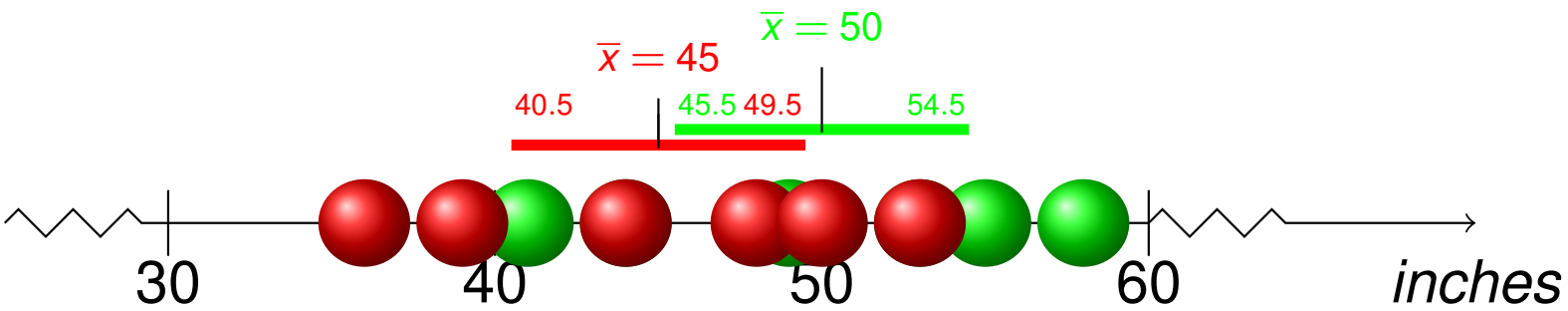


Computation

Results

The height, in inches, of six trees at a nursery are shown at the specificed dates.
Find the mean, standard deviation and standard error of the heights!
Is there a significant difference between the means of samples?

1. **2010 November 22**: 36 48 50 44 53 39
2. **2011 April 1**: 41 53 55 49 58 44



Computation

Standard error in finite population

We have seen in the dice example, that the standard error (1.437591) could be relatively high compared to the mean (1).

If we would check the exact same values (-2, 2, 4, -2, -2, 6) denoting the temperature measured from Monday to Saturday, then would you think that the average temperature at the audited week cannot be estimated more precisely than the earlier computed confidence interval (-1.87 – 3.87)? You have only one missing data!

$$SE = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

Is there any difference between computing the standard error in Hungary or in the United States?

Final examination questions

Comprehensive exam

Singleton, R. A. Jr. and Bruce C. Straits (1999): Approaches to Social Research. Third Edition. Oxford University Press: New York/Oxford.

Questions:

1. **What is reliability?** *How do the main rules concerning the order of survey questions improve the reliability and validity of survey data?* (pp. 113-117, 292-296)
2. What is meant by probability sampling? How do stratification and multistage cluster sampling affect sampling errors? Why? (pp. 141-142, 145-156)
3. What are the main types of non-probability sampling? Explain why these types do not meet the criteria of probability samples. (pp. 157-169)
4. What factors affect the desired sample size? (pp. 163-169)

It was a pleasure!

Daróczi Gergely
daroczi.gergely@btk.ppke.hu