

Introduction to Statistics and R programming

Descriptive Statistics & Graphics

Eric Stemmler

Khovd University

27.01.2021

- 1 Recap
- 2 Handedness Inventory
- 3 Age Guessing
- 4 Stem-Leaf Plot
- 5 Histogram
- 6 Multiple Variables
- 7 Summary

Section 1

Recap

Recap

- statistical vocabulary
 - variation:
 - uncertainty:
 - trial:
 - population:
 - population parameter:
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty:
 - trial:
 - population:
 - population parameter:
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial:
 - population:
 - population parameter:
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial: **the realization of an experiment**
 - population:
 - population parameter:
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial: the realization of an experiment
 - population: all possible events or items
 - population parameter:
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial: the realization of an experiment
 - population: all possible events or items
 - population parameter: **the true value**
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial: the realization of an experiment
 - population: all possible events or items
 - population parameter: the true value
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Recap

- statistical vocabulary
 - variation: the outcome of a sample varies randomly
 - uncertainty: lack of knowledge of about a true value
 - trial: the realization of an experiment
 - population: all possible events or items
 - population parameter: the true value
- calculating probabilities, probability tree
- experiments with 2 outcomes: binomial distribution

Section 2

Handedness Inventory

Handedness Inventory

- Please fill in the questionnaire during **this** lesson
- Return questionnaire at the end
- We will use this data for the next lessons

Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you use would never use the other hand for that activity. If in any case you are really indifferent, put + in both columns. Some of the activities require both hands. In these cases the part of the task, or object, for which hand preference is wanted is indicated in parentheses.

| Далгалар (Task) | Зүүн (Left) | Зон (Right) |
|--|-------------|-------------|
| Бичих Writing | | |
| Зурах Drawing | | |
| Шидэх Throwing | | |
| Хайч Scissors | | |
| Шүдний сойз Toothbrush | | |
| Сэрээгүй хууча Knife, without fork | | |
| Халбага Spoon | | |
| Дэрс, дээд гар Wroom, upper hand | | |
| Хавирхалтай тэмцээн, тэмцээнийг барьж буй гар Striking match (hand that hold the match) | | |
| Нээлтний хайрцаг (тагийг нь барьдаг гар) Opening box (hand that holds the lid) | | |
| нийт дүн Total | | |

$$\text{Right} - \text{Left} = \frac{\text{Right} + \text{Left}}{\text{Right} + \text{Left}} =$$

Create a Left and a Right score by counting the total number of + signs in each column. Your handedness score is $(\text{Right} - \text{Left}) / (\text{Right} + \text{Left})$; thus, a pure right-hander will have a score of $(20 - 0) / (20 + 0) = 1$, and a pure left-hander will score $(0 - 20) / (0 + 20) = -1$.

Section 3

Age Guessing

Subsection 1

Data Collection

Data Collection

- 1 note down the number of team members
- 2 each team estimate the age of the person on the photograph
- 3 note down only **one** estimated value for each card!
- 4 pass photograph to the next team
- 5 the team with the most accurate estimates wins!

20 minutes

| | |
|--|-------------------|
| Guessing ages. For each card your group is given, estimate the age of the person on the card and write your guess in the table below in the row corresponding to that numbered card. Later, you will be told the true ages and you can compute your errors. The error is defined as estimated minus actual age. | Team-ID: |
| | No. team members: |

| Card | Estimated Age | Actual Age | Error |
|------|---------------|------------|-------|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

Data Collection

- 1 calculate the errors of the estimates
- 2 calculate the average of the error
- 3 copy results to the table on the blackboard

| | |
|--|-------------------|
| Guessing ages. For each card your group is given, estimate the age of the person on the card and write your guess in the table below in the row corresponding to that numbered card. Later, you will be told the true ages and you can compute your errors. The error is defined as estimated minus actual age. | Team-ID: |
| | No. team members: |

| Card | Estimated Age | Actual Age | Error |
|------|---------------|------------|-------|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

Section 4

Stem-Leaf Plot

Stem-Leaf Plot

Consider the following data points:

20, 14, 16, 18, 22, 38, 61, 52, 52, 55, 76, 84, 79, 81, 82

##

The decimal point is 1 digit(s) to the right of the |

##

1 | 468

2 | 02

3 | 8

4 |

5 | 225

6 | 1

7 | 69

8 | 124

Stem-Leaf Plot

- Create a steam-leaf plot of the averaged errors

| | |
|--|-------------------|
| Guessing ages. For each card your group is given, estimate the age of the person on the card and write your guess in the table below in the row corresponding to that numbered card. Later, you will be told the true ages and you can compute your errors. The error is defined as estimated minus actual age. | Team-ID: |
| | No. team members: |

| Card | Estimated Age | Actual Age | Error |
|------|---------------|------------|-------|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

Section 5

Histogram

Subsection 1

Introduction

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): $(0,5]$, $(5,10]$, $(10,15]$... $(80,85]$, $(85,90]$, $(90,95]$, $(95,100]$

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): $(0,5]$, $(5,10]$, $(10,15]$... $(80,85]$, $(85,90]$, $(90,95]$, $(95,100]$

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): $(0,5]$, $(5,10]$, $(10,15]$... $(80,85]$, $(85,90]$, $(90,95]$, $(95,100]$

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): $(0,5]$, $(5,10]$, $(10,15]$... $(80,85]$, $(85,90]$, $(90,95]$, $(95,100]$

Example: 20, 14, 16, 18, 22, 38, 61, 52, 52, 55, 76, 84, 79, 81, 82

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): (0,5], (5,10], (10,15] ... (80,85], (85,90], (90,95], (95,100]

Example: 20, 14, 16, 18, 22, 38, 61, 52, 52, 55, 76, 84, 79, 81, 82

| | | | | | |
|----|---------|----|---------|----|---------|
| 20 | (15,20] | 38 | (35,40] | 76 | (75,80] |
| 14 | (10,15] | 61 | (60,65] | 84 | (80,85] |
| 16 | (15,20] | 52 | (50,55] | 79 | (75,80] |
| 18 | (15,20] | 52 | (50,55] | 81 | (80,85] |
| 22 | (20,25] | 55 | (50,55] | 82 | (80,85] |

Introduction

- stem-leaf plot only works for numbers, not for categories
- stem-leaf plots are always based on digits
- count data in terms of intervals/ bins (e.g. width=5): (0,5], (5,10], (10,15] ... (80,85], (85,90], (90,95], (95,100]

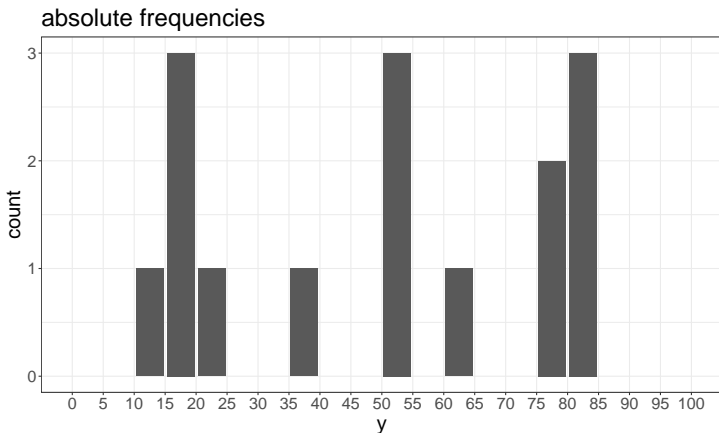
Example: 20, 14, 16, 18, 22, 38, 61, 52, 52, 55, 76, 84, 79, 81, 82

| | | | | | |
|----|---------|----|---------|----|---------|
| 20 | (15,20] | 38 | (35,40] | 76 | (75,80] |
| 14 | (10,15] | 61 | (60,65] | 84 | (80,85] |
| 16 | (15,20] | 52 | (50,55] | 79 | (75,80] |
| 18 | (15,20] | 52 | (50,55] | 81 | (80,85] |
| 22 | (20,25] | 55 | (50,55] | 82 | (80,85] |

| interval | (10,15] | (15,20] | (20,25] | (35,40] | (50,55] | (60,65] | (75,80] | (80,85] |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| count | 1 | 3 | 1 | 1 | 3 | 1 | 2 | 3 |

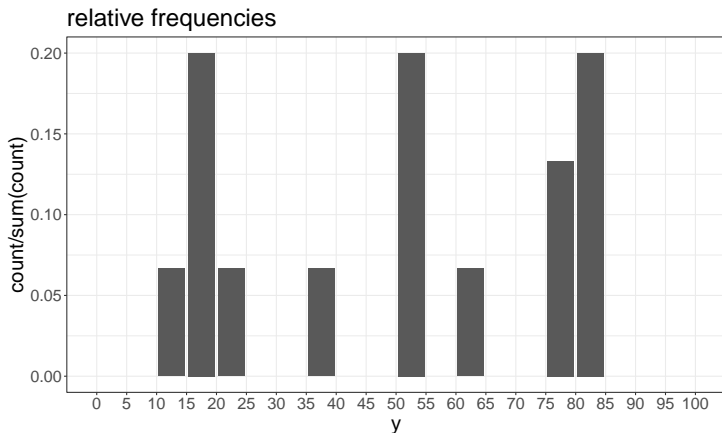
Introduction

| interval | (10,15] | (15,20] | (20,25] | (35,40] | (50,55] | (60,65] | (75,80] | (80,85] |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| count | 1 | 3 | 1 | 1 | 3 | 1 | 2 | 3 |



Introduction

| interval | (10,15] | (15,20] | (20,25] | (35,40] | (50,55] | (60,65] | (75,80] | (80,85] |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| count | 1 | 3 | 1 | 1 | 3 | 1 | 2 | 3 |



Subsection 2

Exercise

Exercise

Create your own histogram

- 1 Plot the error values as a histogram
- 2 Choose a bin width/ size
- 3 Draw the histogram on the blank plotting sheet
- 4 Present your histogram

Exercise

Create your own histogram

- 1 Plot the error values as a histogram
- 2 Choose a bin width/ size
- 3 Draw the histogram on the blank plotting sheet
- 4 Present your histogram

Try to answer and discuss the following questions:

- 1 How does the histogram change with different bin widths?
- 2 How accurate is age guessing?
- 3 Does guessing accuracy vary across age?
- 4 What is the variance/ spread of the error?

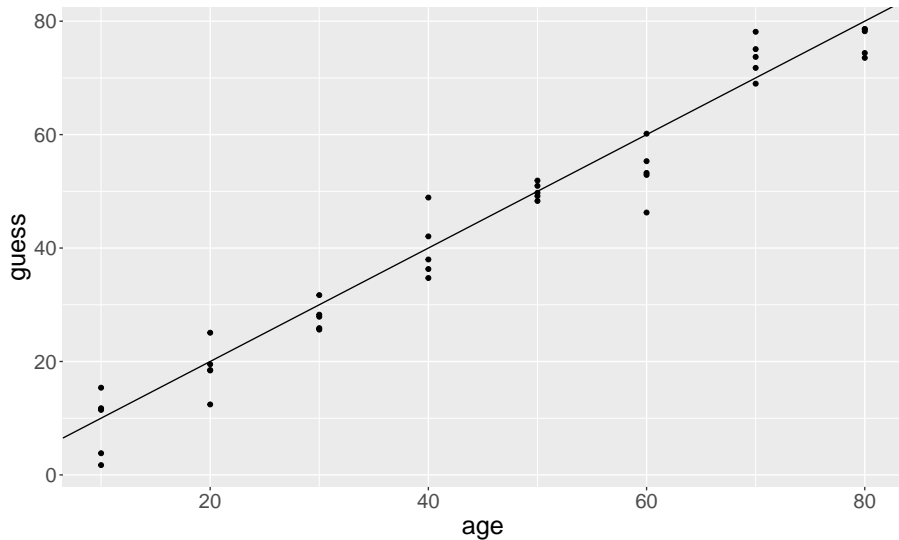
Section 6

Multiple Variables

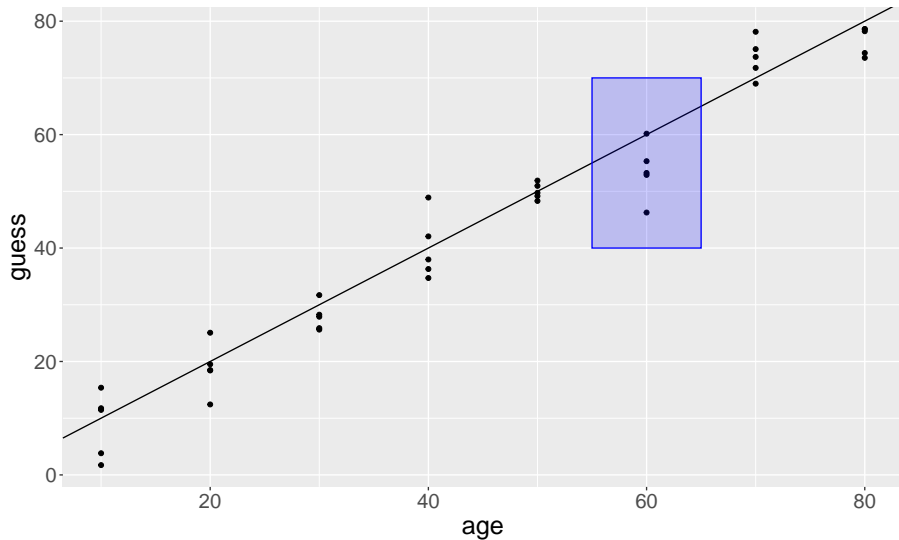
Subsection 1

Scatter plot

Scatter plot



Scatter plot



Section 7

Summary

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

Summary

- **Always** visualize your data set before analysing it!
- data collection & experiment: randomization, random sample
- descriptive statistics: mean, error, standard deviation, variance, bias
- statistical graphics: stem-leaf plot, histogram, scatter plot
- bins:
 - too narrow: loss of shape
 - too wide: missing the details

