



*Eiropas savienība*



**LATVIJAS  
UNIVERSITĀTE**  
ANNO 1919

IEGULDĪJUMS TAVĀ NĀKOTNĒ

# Robustas un neparametriskas metodes statistiskai datu apstrādei

Jānis Valeinis

LU, Rīga

19/12/2013

- 1 Redzes uztveres dati un to raksturlielumi;
- 2 Redzes uztveres statistiskās datu analīzes izvirzītie mērķi I
  - Normu noteikšana dažādiem testiem, R programmas koda izveide.
  - **robustu** un **neparametrisku** metožu izmantošana (mediāna, MAD utt).
  - Dažādu testu salīdzināšana (pēc vecuma, skolām, klasēm). ANOVA un Tukey tests.
- 3 Redzes uztveres datu analīzes izvirzītie mērķi II
  - Korelāciju analīze;
  - DemH un DemV testu salīdzināšana (ANOVA, ANOVA uz rangiem);
  - ROC līknes.
- 4 Secinājumi

# Datu struktūra I

| Skolas                 | Datu skaits pa vecumu grupām |            |            |            |            |            |            |            |            |            |            |            |           |           |             |
|------------------------|------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----------|-----------|-------------|
|                        | 6 gadi                       | 7 gadi     | 8 gadi     | 9 gadi     | 10 gadi    | 11 gadi    | 12 gadi    | 13 gadi    | 14 gadi    | 15 gadi    | 16 gadi    | 17 gadi    | 18 gadi   | 19 gadi   | Kopā        |
| Āgenskalna sks.        | 2                            | 35         | 29         | 43         | 40         | 28         | 28         | 2          | 0          | 0          | 0          | 0          | 0         | 0         | <b>207</b>  |
| ĀVĢ                    | 0                            | 0          | 0          | 0          | 0          | 0          | 3          | 35         | 26         | 38         | 37         | 29         | 10        | 2         | <b>180</b>  |
| Iļģuciema vsk.         | 1                            | 10         | 21         | 23         | 17         | 18         | 21         | 16         | 21         | 10         | 0          | 0          | 0         | 0         | <b>158</b>  |
| Pamatskola "Ridze"     | 0                            | 11         | 15         | 22         | 17         | 7          | 6          | 6          | 12         | 7          | 0          | 0          | 0         | 0         | <b>103</b>  |
| R.S.S. Internātpam.    | 0                            | 0          | 7          | 8          | 9          | 10         | 13         | 3          | 11         | 10         | 4          | 3          | 2         | 1         | <b>81</b>   |
| RCHV                   | 0                            | 30         | 20         | 40         | 26         | 33         | 26         | 20         | 26         | 34         | 29         | 11         | 13        | 3         | <b>311</b>  |
| Rīgas 49. vsk.         | 1                            | 24         | 12         | 18         | 22         | 23         | 17         | 28         | 20         | 22         | 18         | 24         | 28        | 11        | <b>268</b>  |
| Rīgas 85. vsk.         | 1                            | 16         | 8          | 14         | 11         | 1          | 1          | 1          | 8          | 10         | 4          | 2          | 0         | 0         | <b>77</b>   |
| Rīgas 94. vsk.         | 1                            | 10         | 9          | 11         | 13         | 13         | 13         | 14         | 16         | 13         | 8          | 4          | 7         | 0         | <b>132</b>  |
| Baložu vsk.            | 0                            | 8          | 15         | 13         | 3          | 12         | 13         | 6          | 6          | 5          | 1          | 0          | 0         | 0         | <b>82</b>   |
| Mārupes vsk.           | 0                            | 13         | 18         | 10         | 13         | 12         | 10         | 17         | 6          | 1          | 1          | 6          | 5         | 0         | <b>112</b>  |
| Rīgas Spec. vsk.       | 0                            | 13         | 12         | 17         | 24         | 14         | 28         | 21         | 22         | 16         | 13         | 6          | 2         | 0         | <b>188</b>  |
| Rīgas Hanzas vsk.      | 3                            | 23         | 31         | 25         | 16         | 28         | 35         | 17         | 20         | 10         | 10         | 12         | 5         | 0         | <b>235</b>  |
| Ziemeļvalstu ģimnāzija | 1                            | 22         | 19         | 27         | 15         | 35         | 22         | 25         | 25         | 9          | 13         | 20         | 17        | 0         | <b>250</b>  |
| Rīgas Imantas vsk.     | 0                            | 23         | 20         | 32         | 19         | 24         | 18         | 19         | 23         | 16         | 4          | 0          | 4         | 0         | <b>202</b>  |
| Rīgas Valdorfskola     | 1                            | 5          | 10         | 10         | 10         | 13         | 5          | 4          | 2          | 7          | 1          | 0          | 0         | 0         | <b>68</b>   |
| <b>Kopā</b>            | <b>11</b>                    | <b>243</b> | <b>246</b> | <b>313</b> | <b>255</b> | <b>271</b> | <b>259</b> | <b>234</b> | <b>244</b> | <b>208</b> | <b>143</b> | <b>117</b> | <b>93</b> | <b>17</b> | <b>2654</b> |

- 6 un 19 gadu veci skolēni jāizslēdz no tālākas analīzes!
- 16, 17, 18 gadu veiciem skolēniem mazāki izlašu apjomi

# Datu struktūra II

| Tests       | Datu skaits pa vecumu grupām |        |        |        |         |         |         |         |         |         |         |         |         |         |             |
|-------------|------------------------------|--------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-------------|
|             | 6 gadi                       | 7 gadi | 8 gadi | 9 gadi | 10 gadi | 11 gadi | 12 gadi | 13 gadi | 14 gadi | 15 gadi | 16 gadi | 17 gadi | 18 gadi | 19 gadi | Kopā        |
| Lasītprasme | 11                           | 243    | 246    | 313    | 255     | 271     | 259     | 234     | 244     | 208     | 143     | 117     | 93      | 17      | <b>2654</b> |
| tmta.pap    | 4                            | 93     | 136    | 150    | 147     | 171     | 134     | 65      | 110     | 92      | 69      | 65      | 40      | 3       | <b>1279</b> |
| tmtb.pap    | 1                            | 26     | 91     | 98     | 121     | 149     | 106     | 55      | 94      | 79      | 64      | 64      | 38      | 2       | <b>988</b>  |
| tmta.dat    | 10                           | 218    | 240    | 291    | 244     | 269     | 253     | 229     | 237     | 208     | 140     | 116     | 93      | 17      | <b>2565</b> |
| tmtb.dat    | 6                            | 106    | 172    | 201    | 209     | 232     | 212     | 207     | 215     | 186     | 132     | 113     | 90      | 13      | <b>2094</b> |
| tmta.pul    | 9                            | 216    | 229    | 292    | 225     | 256     | 245     | 219     | 226     | 185     | 128     | 111     | 86      | 17      | <b>2444</b> |
| tmtb.pul    | 5                            | 101    | 155    | 202    | 193     | 219     | 206     | 193     | 203     | 165     | 119     | 108     | 82      | 13      | <b>1964</b> |
| pulis.kr    | 6                            | 138    | 139    | 235    | 191     | 221     | 231     | 205     | 199     | 146     | 96      | 66      | 43      | 6       | <b>1922</b> |
| pulis.laba  | 6                            | 135    | 139    | 235    | 191     | 209     | 228     | 204     | 188     | 156     | 90      | 52      | 29      | 6       | <b>1868</b> |
| atpaz.var   | 8                            | 195    | 196    | 262    | 219     | 246     | 244     | 217     | 220     | 183     | 130     | 105     | 81      | 17      | <b>2323</b> |
| atpaz.nev   | 2                            | 99     | 114    | 163    | 167     | 161     | 143     | 118     | 106     | 98      | 72      | 52      | 51      | 15      | <b>1361</b> |
| dem.v       | 5                            | 140    | 116    | 178    | 135     | 118     | 153     | 171     | 146     | 127     | 88      | 60      | 57      | 15      | <b>1509</b> |
| dem.h       | 5                            | 139    | 116    | 178    | 135     | 118     | 153     | 171     | 146     | 127     | 88      | 60      | 57      | 15      | <b>1508</b> |

- Vislielākais datu apjoms *Lasītprasmes* testam!
- *tmta*, *tmtb* papīra versiju testiem mazāki izlašu apjomi

- **Mērķis I:** aprēķināt skores katram testam un katram vecumam sekojošā veidā

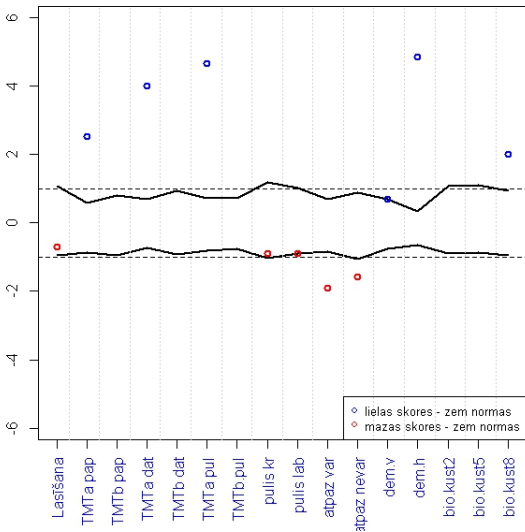
$$Y = (X - \text{mean}(X)) / \text{sd}(X)$$

Ja  $X \sim N(\mu, \sigma^2)$ , tad  $P(|Y| > 3) \approx 0.0027$ .

- **Mērķis II:** izveidot programmu, kas aprēķina un zīmē skores visiem testiem vienā grafikā, nosakot normas kā 1SD no populācijas datiem katrā vecuma grupā.
- **Piezīme.** Skores atrodas vienā mērogā un ļauj salīdzināt savā starpā dažādus testus un dažādus vecumus!

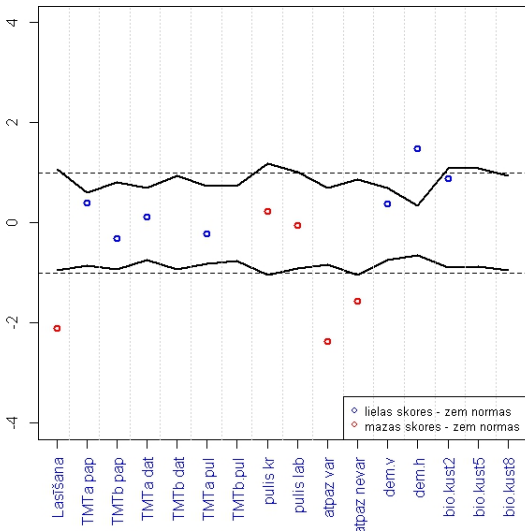
# Normu izstrādāšana: piemērs I (R programmas kods)

Skoru grafiks (7 testu rezultāti zem normas)

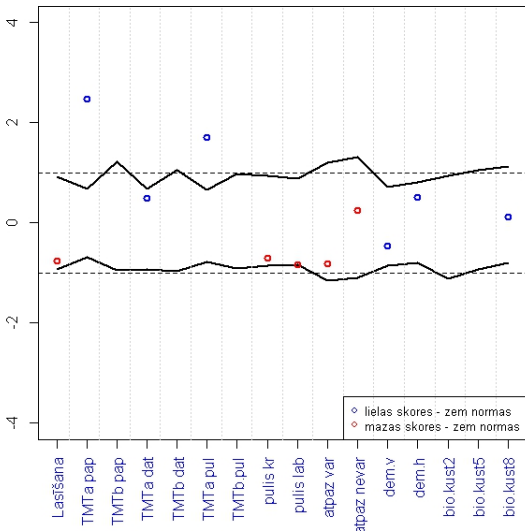


# Normu izstrādāšana: piemērs II (R programmas kods)

Skoru grafiks (4 testu rezultāti zem normas)



Skoru grafiks (2 testu rezultāti zem normas)





Dots:  $x_1, \dots, x_n$  - novērojumi.

- **Klasiskie novērtējumi:** *izlases vidējā vērtība un izlases dispersija*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

- **Robusti novērtējumi:** *izlases mediāna un mediānas absolūtā deviācija (MAD)*

$$\text{Med}(x), \quad \text{MAD}(x) = \text{Med}(|x - \text{Med}(x)|).$$

- *Robusti novērtējumi* - tādi novērtējumi, kurus maz (vai gandrīz nemaz) ietekmē ļoti atšķirīgi novērojumi (izlecēji).

Dots:  $\{0.3, 0.5, 0.8, 1.5, 1.9\}$  - novērojumi.

- **Klasiskie novērtējumi:** *izlases vidējā vērtība un izlases dispersija*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.68.$$

- **Robusti novērtējumi:** *izlases mediāna un mediānas absolūtā deviācija (MAD)*

$$\text{Med}(x) = 0.8, \quad \text{MAD}(x) = \text{Med}(|x - \text{Med}(x)|) = 0.74.$$

- Robusti un klasiski novērtējumi uzvedas līdzīgi, ja nav lielu izlecēju!

Dots:  $\{0.3, 0.5, 0.8, 1.5, 1.9, 8\}$  - novērojumi.

- **Klasiskie novērtējumi:** *izlases vidējā vērtība un izlases dispersija*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 2.17, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 2.92.$$

- **Robusti novērtējumi:** *izlases mediāna un mediānas absolūtā deviācija (MAD)*

$$\text{Med}(x) = 1.15, \quad \text{MAD}(x) = \text{Med}(|x - \text{Med}(x)|) = 1.04.$$

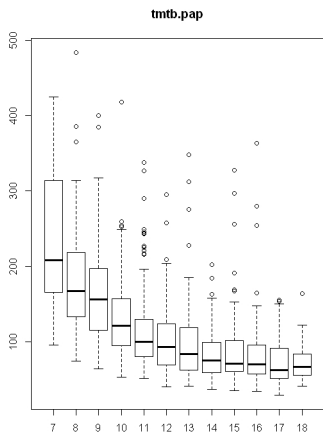
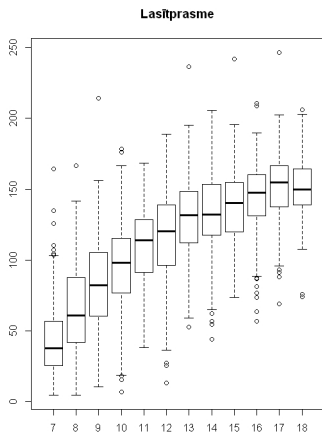
- Pievienojot izlecēju 8, izlases vidējā vērtība un dispersija palielinājās vismaz 2 reizes!
- Robustu novērtējumu vērtības 2 reizes mazākas, izlecēju ietekme mazināta!

# Standartnovirzes novērtēšana: SD vai MAD?

| Tests       |     | Datu skaits pa vecumu grupām |        |        |         |         |         |         |         |         |         |         |         |
|-------------|-----|------------------------------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|             |     | 7 gadi                       | 8 gadi | 9 gadi | 10 gadi | 11 gadi | 12 gadi | 13 gadi | 14 gadi | 15 gadi | 16 gadi | 17 gadi | 18 gadi |
| Lasītprasme | SD  | 26                           | 31     | 31     | 31      | 29      | 33      | 29      | 29      | 27      | 26      | 28      | 24      |
|             | MAD | 21                           | 31     | 32     | 29      | 27      | 33      | 28      | 26      | 24      | 20      | 22      | 19      |
| tmta.pap    | SD  | 44                           | 26     | 23     | 19      | 14      | 17      | 11      | 12      | 9       | 17      | 9       | 8       |
|             | MAD | 24                           | 19     | 16     | 13      | 12      | 11      | 9       | 13      | 9       | 12      | 9       | 7       |
| tmtb.pap    | SD  | 94                           | 70     | 65     | 54      | 54      | 47      | 63      | 36      | 53      | 56      | 31      | 25      |
|             | MAD | 89                           | 62     | 61     | 46      | 34      | 39      | 42      | 33      | 21      | 24      | 20      | 20      |
| tmta.dat    | SD  | 25                           | 20     | 21     | 16      | 12      | 12      | 8       | 9       | 8       | 11      | 7       | 7       |
|             | MAD | 18                           | 15     | 16     | 10      | 9       | 7       | 7       | 8       | 6       | 5       | 5       | 5       |
| tmtb.dat    | SD  | 49                           | 50     | 40     | 32      | 35      | 28      | 21      | 25      | 21      | 22      | 18      | 14      |
|             | MAD | 50                           | 42     | 37     | 29      | 22      | 23      | 19      | 19      | 16      | 16      | 15      | 12      |
| dem.v       | SD  | 16                           | 14     | 13     | 12      | 11      | 9       | 6       | 8       | 5       | 6       | 6       | 4       |
|             | MAD | 11                           | 10     | 9      | 8       | 8       | 7       | 6       | 6       | 5       | 6       | 5       | 3       |
| dem.h       | SD  | 26                           | 29     | 23     | 23      | 16      | 13      | 8       | 10      | 7       | 6       | 8       | 5       |
|             | MAD | 22                           | 16     | 14     | 11      | 11      | 10      | 8       | 7       | 6       | 6       | 8       | 3       |

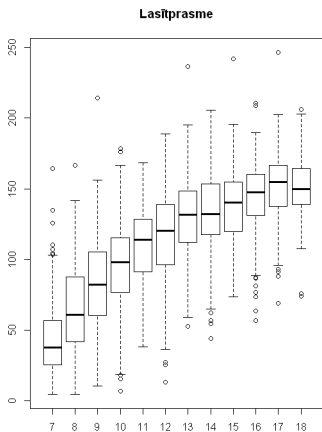
- Normu izstrādē var lietot robustus novērtējumus, piemēram, 1MAD!

# Testu salīdzināšana pēc vecumiem: piemērs

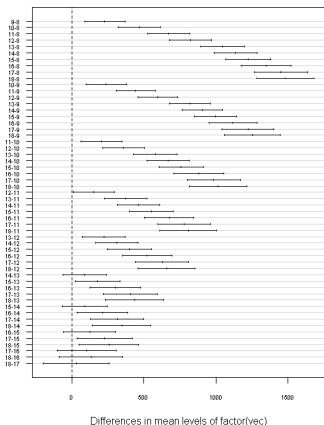


- Kastu grafiki - ērts veids datu kopu grafiskai attēlošanai, salīdzināšanai, **izlecēju noteikšanai!**
- Jautājums - no kura vecuma vairs nav atšķirība lasītprasmē?
- Analīze: jāveic ANOVA un Tukey tests!

# Testu salīdzināšana pēc vecumiem: piemērs

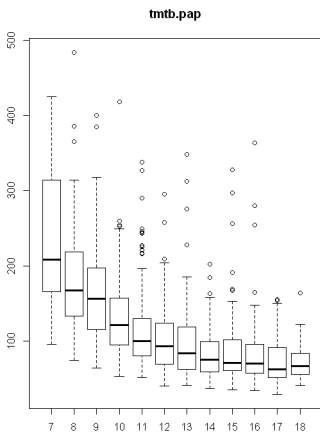


95% family-wise confidence level

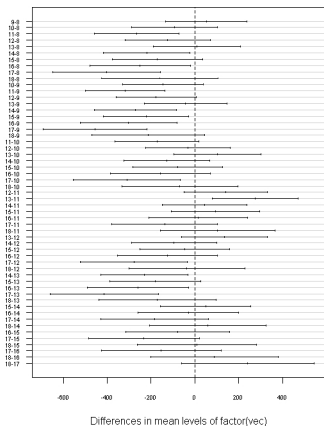


- *Grafiska divu grupu salīdzināšana*: ja 0 ietilpst Tukey ticamības intervālos, tad nav statistiski nozīmīgas atšķirības!
- **Secinājums**: sākot no 13 gadu vecuma nav statistiski nozīmīgas atšķirības lasītprasēmē.

# Testu salīdzināšana pēc vecumiem: piemērs



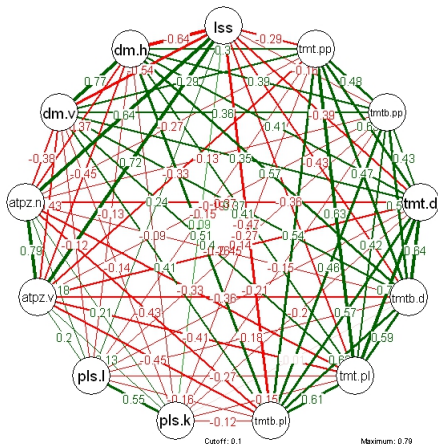
95% family-wise confidence level



- Lielas dispersijas, plaši ticamības intervāli, samērā mazi izlašu apjomi!
- Nevaram atrast statistiski nozīmīgu atšķirību jau sākot no 8 gadiem.

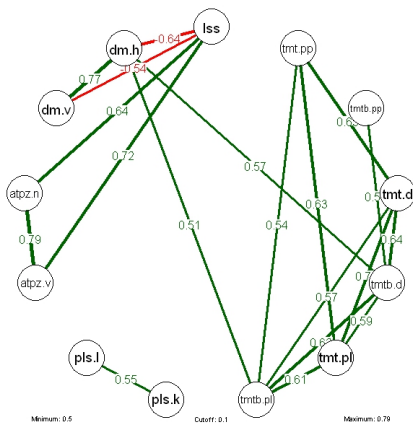
# Korelācijas starp dažādiem testiem

**Mērķis:** atrast nozīmīgas un ciešas lineāras korelācijas starp testiem!



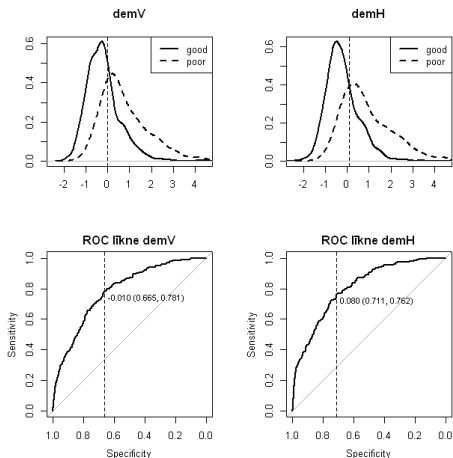


# Korelācijas starp dažādiem testiem (kor > 0.5)



- **Secinājums:** Negatīva, samērā cieša korelācija starp lasītprasmi un dem.h un dem.v testiem!

# ROC līknes dem.h un dem.v testu skorēm



- ROC līknes līdzīgas dem.v un dem.h testiem.
- ROC līknes dod labāko punktu, ko var izmantot kā sliekšni.

- ① Statistiskai datu analīzei ieteicams lietot programmu **R**!

- 1 Statistiskai datu analīzei ieteicams lietot programmu **R**!
- 2 Svarīga statistisko analīzi (salīdzināšanu, hipotēžu pārbaudi) attēlot grafiskā veidā (kastu grafiki, roc līknes, ticamības intervāli, korelāciju puķes utt).

- 1 Statistiskai datu analīzei ieteicams lietot programmu **R**!
- 2 Svarīga statistisko analīzi (salīdzināšanu, hipotēžu pārbaudi) attēlot grafiskā veidā (kastu grafiki, roc līknes, ticamības intervāli, korelāciju puķes utt).
- 3 Skoru veidošana ļauj salīdzināt dažādus testus un veikt datu analīzi, apvienojot vecumu grupas.

- 1 Statistiskai datu analīzei ieteicams lietot programmu **R**!
- 2 Svarīga statistisko analīzi (salīdzināšanu, hipotēžu pārbaudi) attēlot grafiskā veidā (kastu grafiki, roc līknes, ticamības intervāli, korelāciju puķes utt).
- 3 Skoru veidošana ļauj salīdzināt dažādus testus un veikt datu analīzi, apvienojot vecumu grupas.
- 4 Robustās statistikas metodes ir pārākas pār klasiskām, ja datos ir izlecēji.

- 1 Statistiskai datu analīzei ieteicams lietot programmu **R**!
- 2 Svarīga statistisko analīzi (salīdzināšanu, hipotēžu pārbaudi) attēlot grafiskā veidā (kastu grafiki, roc līknes, ticamības intervāli, korelāciju puķes utt).
- 3 Skoru veidošana ļauj salīdzināt dažādus testus un veikt datu analīzi, apvienojot vecumu grupas.
- 4 Robustās statistikas metodes ir pārākas pār klasiskām, ja datos ir izlecēji.
- 5 Citas robustas (arī neparametriskas) metodes balstās uz datu rangiem (Vilkoksona tests, ANOVA uz rangiem utt).

- 1 Statistiskai datu analīzei ieteicams lietot programmu **R**!
- 2 Svarīga statistisko analīzi (salīdzināšanu, hipotēžu pārbaudi) attēlot grafiskā veidā (kastu grafiki, roc līknes, ticamības intervāli, korelāciju puķes utt).
- 3 Skoru veidošana ļauj salīdzināt dažādus testus un veikt datu analīzi, apvienojot vecumu grupas.
- 4 Robustās statistikas metodes ir pārākas pār klasiskām, ja datos ir izlecēji.
- 5 Citas robustas (arī neparametriskas) metodes balstās uz datu rangiem (Vilkoksona tests, ANOVA uz rangiem utt).
- 6 Dem.h un dem.v testi ir laba alternatīva lasītprasmes testam!



Paldies par uzmanību!

## **Pateicība**

Darbs veikts ERAF un LU projekta "Skolas vecuma bērnu redzes un redzes uztveres traucējumu pētīšana un diagnostikas metodiku izstrāde"

Nr.2011/0004/2DP/2.1.1.1.0/10/APIA/VIAA/027 ietvaros