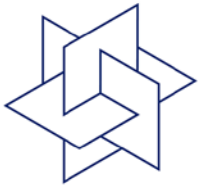
Real World Data

CO@Work Berlin

Thorsten Koch

22.09.2009



DFG Research Center MATHEON
Mathematics for key technologies

Task: Collecting Data for the Optimal Seat Allocation

- We want to compute the optimal seat allocation for the lecture hall.
- To do this we need your preferences.
- Everyone should send me an email with a data file.
- We will see how long it will take.

File Format

- ASCII text with only a LF (ASCII 10) as line separator.
- Fields are separated by a single space (ASCII 32)
- Line 1: ParticipantNo HasLaptop EmailAddress

  e.g. `67 1 koch@zib.de`

  0 = has no Laptop, 1 = has a Laptop
- Lines 2-???: SeatNumber PreferenceValue
  - Seat numbers start down at the low entrance, left to right, row by row.
  - The hightest numbered seat is at the window side at the top.
  - Count only seats that are physically there.
  - The seat numbers in the file should be monotonically increasing.
  - The preference values should be between 0 and 100.

  e.g.
  ```
  12 55
  13 40
  14 35 …
  ```

Rules Regarding Preference Values

- Allowed values are between 0 and 100
- Only seats which are not available for the participants are allowed to get a value of 0
- All numbers 1-100 have to be used at least once
- The average has to be between 40-60
- The difference to an adjacent seat has to be < 40
- The difference to a neighboring seat has to be < 20
- The data should not be randomly generated

Specifying Preference Offsets

- Lines ???-???: ParticipantNo PreferenceOffset
  List indicating persons which you would like or not like to be your seat neighbor. (You have to know the ParticipantNo of the person.)

  - A ParticipantNo of 0 indicates an empty seat.

  - The PreferenceOffset is between -20 and 20 and will be added to your PreferenceValue if the person with the given ParticipantNo is your neighbor.
    e.g.
    ```
    55 17
    27 -5
    72 8
    0 -10 …
    ```

  - This list can have as many entries as you like, but there should be at least 2 entries, and the occurring participant numbers have to be unique and valid.

How To Submit

- Submission of this file is required for the course
- The name of the file has to be *ParticipantNo*`.txt`
- It should be **attached** to an email
- Send the email to koch@zib.de
- The subject of the email should be
  CO@Work: SeatData for *ParticipantNo*
- *Please, as soon as possible.*

2 Days after the lecture

- Mails received : 13
- Different Subjects : 4 (10 1 1 1)
- Wrong field spacing : 4
- Seat counts : 2 (12 1)
- Missing data : 1
- Too much data : 1
- Ok, from first view : 5 out of 13

3 Days after the lecture

- Mails received : 23
- Different Subjects : 6 (17 2 1 1 1 1)
- Wrong field spacing : 4
- Seat counts : 4 (19 1 1)
- Missing data : 2
- Too much data : 0
- Ok, from first view : 10
- Corrected : 1

Add to the specification:

- A seat without a desk is not allowed for the participants
- Seats with a 0 preference value are not relevant for the adjacency/neighboring difference rules.

4 Days after the lecture

- Mails received : 37
- Wrong subject : 11
- Wrong field spacing : 8
- Strange seat counts : 5
- Missing data : 2
- Corrected : 3

5 Days after the lecture

- Mails received : 47
- Data sets : 41 (6 corrections)
- Wrong subject : 12
- Wrong attachment name : 2
- Wrong line separator : 29
- Wrong field separator : 10
- Preference value not used : 11
- Other Errors : 1
- Number of seats : 153 - 181
- No complains so far : 4

- **Please correct and resubmit:**
  13 20 23 45 47 53 59 73 78 134 135 139 155

7 Days after the lecture

- Mails received : 79
- Data sets : 64
- Wrong subject : 16
- Wrong attachment name : 2
- Wrong line separator : 45
- Wrong field separator : 11
- Preference value not used : 22
- Other Errors : 2
- Number of seats : 153 - 181
- No complains so far : 8

- **Please correct and resubmit:**
  12 18 23 27 42 45 47 53 63 64 69 71 93 98
  103 129 134 135 137 139 145 166

9 Days after the lecture

- Mails received : 104
- Data sets : 76
- Wrong subject : 18
- Wrong attachment name : 2
- Preference value not used : 19
- Neighbor difference : 21
- Wrong no/sequence seats : 10
- Wrong 0 seats : 20
- No complains so far : 10

Overview of Errors in Data

	E7	E10	E11	E12	E13	E14	E16
5							X
6							X
12					X		X
13							X
16							X
18					X		X
19						X	X
20						X	
23					X		
24						X	
26							X
27						X	
36						X	
42					X		
45			X	X	X	X	X
47					X		
53					X		
59						X	
63			X		X	X	
64			X		X	X	X
71					X	X	

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neigbour diff

E16 seat not 0

	E7	E10	E11	E12	E13	E14	E16
77							X
78	X		X			X	X
81				X	X		X
98					X		
99	X		X			X	
103					X	X	
107			X			X	X
108			X			X	X
111							X
121							X
128			X			X	X
129		X					
134			X	X	X	X	
135					X		
137		X			X	X	X
139					X		X
145	X		X		X	X	
160						X	
166					X	X	

Please correct and resubmit

Nothing submitted so far!

21	Nguyen, Thinh
22	Vu Khac, Ky
56	Lidický, Bernard
61	Durdevac, Natasa
65	Marecek, Jakub
75	Wohlers, Inken
79	Forma, Iris
84	Unal, Murat Engin
89	Cinar, Didem
90	Dursun, Pınar
92	Schmutzer, Andreas
110	Temur, Gül Tekin
114	Özdemir, Dilek
115	Musial, Jedrzej
118	Gutierrez, Sandra
119	Pashkovich, Kanstantsin
122	Pfeuffer, Frank
136	Staiger, Christine
138	Bulánek, Jan
140	Retkowski, Waldemar
194	Tan, Ngo Dac

11 Days after the lecture

- Mails received : 144
- Wrong subject : ~23
- Wrong attachment name : 4

- Data sets : 92
- To be corrected : 28
- Missing : 6

- Preference value not used : 14
- Neighbor difference : 18
- Wrong no/sequence seats : 2

Nothing submitted so far!

21	Nguyen, Thinh
22	Vu Khac, Ky
92	Schmutzer, Andreas
119	Pashkovich, Kanstantsin
122	Pfeuffer, Frank
140	Retkowski, Waldemar

Overview of Errors in Data

	E7	E10	E11	E12	E13	E14
12					X	
18					X	
23	X		X			X
24						X
27						X
45					X	X
47					X	
63			X		X	X
71					X	X
78	X		X			X
79		X	X	X	X	
103						X
107			X			X
108			X			X
110		X				
114						X
118					X	X
128			X			X
134			X	X	X	X
135					X	
136						X
137		X			X	X
138					X	
139					X	
160						X
166					X	X

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neigbour diff

Please correct and resubmit

13 Days after the lecture

- Mails received : 159
- Wrong subject : ~26
- Wrong attachment name : 4

- Data sets : 94
- To be corrected : 18
- Missing : 4

- Preference value not used : 9
- Neighbor difference : 14
- Wrong no/sequence seats : 3

Nothing submitted so far!

92	Schmutzer, Andreas
119	Pashkovich, Kanstantsin
122	Pfeuffer, Frank
140	Retkowski, Waldemar

Overview of Errors in Data

	E7	E10	E11	E12	E13	E14
18					X	
24						X
27						X
45					X	X
63					X	
71					X	X
78	X		X			X
79		X	X	X	X	
103						X
107			X			X
108			X			X
114						X
118					X	X
128			X			X
134			X	X	X	X
136						X
137		X			X	X
138					X	

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neigbour diff

Please correct and resubmit

14 Days after the lecture

- Mails received : 166
- Wrong subject : ~28
- Wrong attachment name : 4

- Data sets : 95
- To be corrected : 18
- Missing : 3

- Preference value not used : 7
- Neighbor difference : 14
- Wrong no/sequence seats : 3

Nothing submitted so far!

119	Pashkovich, Kanstantsin
122	Pfeuffer, Frank
140	Retkowski, Waldemar

Overview of Errors in Data

	E7	E10	E11	E12	E13	E14
24						X
27						X
45					X	X
71					X	X
78	X		X			X
79		X	X	X	X	
92					X	X
107			X			X
108			X			X
114						X
118					X	X
128			X			X
134			X	X	X	X
136						X
137		X			X	X

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neigbour diff

Please correct and resubmit

15 Days after the lecture – the final day

- Mails received : 172
- Wrong subject : ~31
- Wrong attachment name : 4

- Data sets : 95
- To be corrected : 13

- Preference value not used : 5
- Neighbor difference : 13
- Wrong no/sequence seats : 2

Subject Variations

- The subject of the email should be
  CO@Work: SeatData for *ParticipantNo*

```
CO@Work: SeatData for 022
CO@Work:SeatData for 222
CO@Work:SeatDatafor222
CO@work: SeatData for 222
CO@Work: Seat Data for 222
Co@Work: SeatData for 222
CO@Work: SeatData for Participant222
CO@Work: SeatData for ParticipantNo
Co@Work: SeatData for Participan222
CO@WORK:  seatdata for 222
COatWork: SeatData for 222
COatWork for 222
SeatData for 222
SeatData for ParticipantNo 222
set data for participant number 222
data set participant number 222
Sitting assignment
Seats assignment
```

ASCII text with only a LF (ASCII 10) as line separator

- Many lines were separated by CR-LF
- Many empty lines

Fields are separated by a single space (ASCII 32)

- All kinds of spaces happened:

```
 1    3
1<tab>3
 1   3
```

Line 1: ParticipantNo HasLaptop EmailAddress

e.g. **67 1 koch@zib.de**

bbbbb.zzz@qqqqq.de
222 1

Lines 2-???: SeatNumber PreferenceValue

- All numbers 1-100 have to be used at least once
  5 errors
- The average has to be between 40-60
  1 error
- The difference to an adjacent seat has to be < 40
  not checked ☺
- The difference to a neighboring seat has to be < 20
  13 errors

The data should not be randomly generated

1 1	19 13	37 31	55 49
2 2	20 14	38 32	56 50
3 3	21 15	39 33	57 51
4 4	22 16	40 34	58 52
5 5	23 17	41 35	59 53
6 0	24 18	42 36	60 54
7 0	25 19	43 37	61 55
8 0	26 20	44 38	62 56
9 0	27 21	45 39	63 57
10 0	28 22	46 40	64 58
11 0	29 23	47 41	65 59
12 6	30 24	48 42	66 60
13 7	31 25	49 43	67 61
14 8	32 26	50 44	68 62
15 9	33 27	51 45	69 63
16 10	34 28	52 46	70 64
17 11	35 29	53 47	…
18 12	36 30	54 48	

Lines ???-???: ParticipantNo PreferenceOffset

- Here the real fun starts, because you could only list participants which also submitted data.
- This error occurred within the first 3 data sets I was looking at.
- There are cases where the PreferenceOffset ist 0 This makes hardly sense.
- I did not check further.

The name of the file has to be *ParticipantNo*`.txt`

- **`222.txt.txt`**
- **`data set participant number 222.txt`**
- **`Participant222.txt`**

Overview of Errors in Data

	E7	E10	E11	E12	E13	E14
24						X
27						X
45					X	X
71					X	X
78	X		X			X
92					X	X
107			X			X
108			X			X
114						X
128			X			X
134			X	X	X	X
136						X
137		X			X	X

E7 bad seatno

E10 bad offset

E11 wrong seatno

E12 bad average

E13 prefval missing

E14 neigbour diff

Sorry,
too late to correct!

Wrong line 1: 81, 129

What happens next

- Now we have at least most of the data...

Too Late

- Unfortunately, the data is no longer up-to-date
- Several participants have left since they send their data
- Of course, we could fix that…

What was the Project about?

- We wanted to compute the optimal seat allocation for the lecture hall.
- But today is the last day of CO@Work.
- We no longer need a seat allocation.
- **Project canceled.**

Thank you very much!

Questions?