



Estimation of Optimum Stadium Size

Contents

Data collection and verification in short terms.....	1
Analysis.....	2
Results	6
Overview.....	6
Optimum Stadium	6
Ticket Prices.....	6
Weather	7
Attendance Type Ratios (ATR).....	7
Fan Factor (FF).....	8
Attendance Type Ratio (ATR) in details	9
Rainy	9
Overcast	10
Partially Clouded.....	11
Sunny Weather	12
Fan Factor (FF)in details.....	13
Document Information	15
Disclaimer.....	15
Denominations	15
Abbreviations	15
Revision History.....	15

Data collection and verification in short terms

Data has been collected through the web interface at

<http://brun.dk/hattrick/index.php?content=15&lang=english>.



Estimation of Optimum Stadium Size

Revision 1. Published by **Statistikforbundet** November 2008
Latest version and contact details available at <http://brun.dk/hattrick>

The following data has been collected for each data set:

- Match-ID
- Weather
- Fan amount
- (Home team) Fan mood (before match) (abbr. **HFM**)
- Attendance (separated into types: terraces, basic seats, seats under roof and VIP seats)
- Stadium capacity (separated into types: terraces, basic seats, seats under roof and VIP seats)

During data collection, data has been verified by mrbrown79 (team-ID: 1183388) and Warlord (team-ID: 572080) from **Statistikforbundet** (DKSTAT, alliance-ID: 64087). No strict rules have been applied to the verification, but with all collected data in a spread sheet, some erroneous data shows quite clearly.

When the data submission is considered wrong without doubt, data has in many cases been rejected (in extreme cases where the data is particularly important, the validity might have been confirmed through the provider, though). If in doubt, the data has been verified by the original data submitter.

Analysis

Through processing, the trustworthy data has been separated into three categories:

1. Data sets, where no attendance category are sold out (**C1**)
2. Data sets, where at least one attendance category has not been sold out (**C2**)
3. Data sets, where the stadium is completely sold out (**C3**)

The C3 data sets were completely useless, and have been rejected. The C1 data was the starting point of the analysis. From this data, the ratio between each of the categories could be extracted. A clear relationship between weather and ratios was known in advance. Further the data collection showed that the relationship was great influenced by the home fan mood also. The extracted ratios are named **ATR** in the following.

In the C2 data the sold out categories results in a total attendance that would have been higher if the stadium had been build appropriately scaled. I.e. the attendance in the sold out categories cannot be used for further calculations.

However, once the ATR have been determined from the C1 data, the sold out categories can be extrapolated from the non-filled categories and the ATR. This extrapolation has been performed on the assumption, that the ratio can be well-determined from (home team) fan mood and weather solely.

The C1 and extrapolated C2 data (totally more than 600 data sets) form a good base for calculation of the fan factor (**FF**). Not surprisingly, the fan factor seems to be dependent on additional parameters



than just weather and home team fan mood. Such parameters, not collected in this analysis, might be: away team fan amount, away team fan mood, match importance and (potentially) some random component.

Although the FF cannot be extracted exactly from the fan mood and weather, a pretty clear relationship shows up. Due to the variations, the relationship has been calculated as a mean value and standard deviation.

Once ATR and FF have been determined as functions of fan mood and weather, the optimum stadium size can be estimated. A prerequisite is the probability of a given weather type. These have been provided by statistics provider [Hattristics](#).

The Final Calculation

Recapitulating, the basis for our final calculation is:

- ATR as function of HFM and weather.
- FF as function of HFM and weather. FF is chosen to be described as a probability distribution, since it clearly varies a lot. A normal distribution is chosen for no particular reason (but it seems to be a reasonable choice). I.e. mean value and standard deviation have been derived from the numbers.
- Ticket prices and maintenance cost for each attendance type. The construction costs have not been included in this analysis (it makes no sense to include a one-time cost unless an interest rate can be applied).
- Weather distribution. The four weather types are not equally likely to occur.

The optimum stadium is build such that net revenue is maximized. This optimum will be reached, when the last one seat (of each attendance type) is sold exactly as often as it requires in order being profitable. In the limit, costs and income of the last seat equals. Whereas the (weekly) maintenance cost is written explicitly in the HT rules, the weekly income is not quite as simple. In this calculation we have estimated, that the potential weekly income of a seat is $\frac{7.67}{16} \times [TICKET PRICE]$, due to the fact, that a home match occurs 7 times (league) and 1 time in the cup (with an income factor of 0.67) in a season of 16 weeks.

Hence, in the optimum stadium the latest build seat should be sold exactly every $[P_{sold}]$ match, where $[P_{sold}]$ can be described as:

$$P_{sold} = \frac{1}{\frac{7.67}{16} \times \frac{[TICKET PRICE]}{[WEEKLY MAINTENANCE COST]}} = \frac{16 \times [WEEKLY MAINTENANCE COST]}{7.67 \times [TICKET PRICE]}$$

P_{sold} for each of the attendance types can be found in Table 3.

Since probability distributions of each of the attendance types have been derived for all weather types, the accumulated probability function can be determined as the weighted average taking into account



Estimation of Optimum Stadium Size

Revision 1. Published by Statistikforbundet November 2008
Latest version and contact details available at <http://brun.dk/hattrick>

the probability for each of the weather types. This principle is sketched in Figure 1, where the normal distributions for each of the weather types are scaled with the probability of the particular weather occurrence. Finally the distributions are summed into an accumulated distribution describing the overall probability for a given FF for a random match for a particular attendance type.

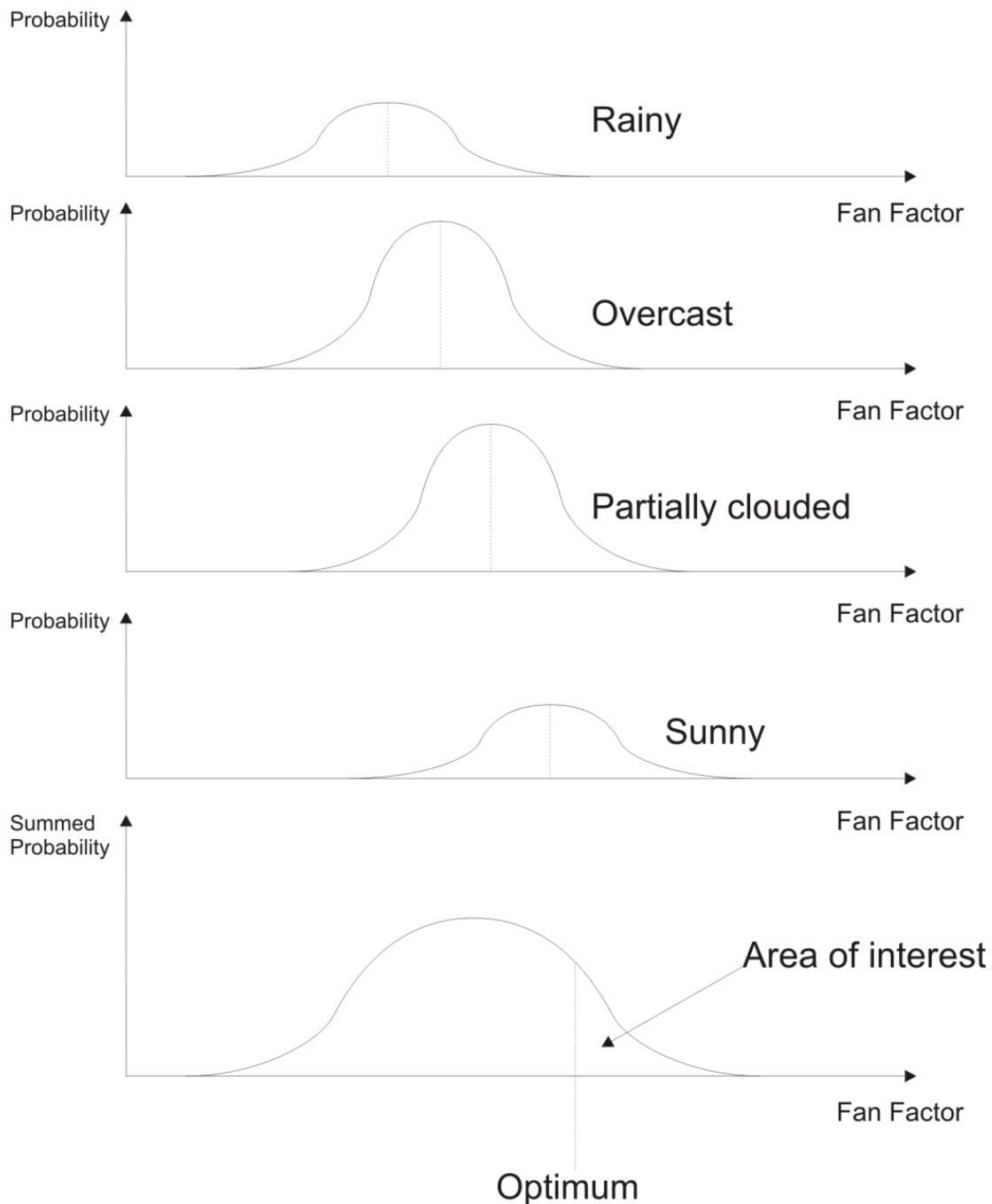


Figure 1: Normal distribution of FF for a given attendance type at the four weather types. At the bottom is the probability distribution sum. Notice the area to the right of the optimum. This area equals the minimum acceptable probability, that the seat type will be sold out.



Estimation of Optimum Stadium Size

Revision 1. Published by **Statistikforbundet** November 2008
Latest version and contact details available at <http://brun.dk/hattrick>

The probability, $F_{attendance\ type}(FF)$, that the FF for a given attendance type is "less or equal" a given FF value can be found by integrating the probability distributions from $-\infty$ to the FF value of interest. Hence the optimum FF for each of the attendance types can be found as the FF corresponding to the probability, that FF is less or equal to:

$$F_{attendance\ type}(FF) = 1 - P_{sold} = 1 - \frac{16 \times [WEEKLY\ MAINTENANCE\ COST]}{7.67 \times [TICKET\ PRICE]}$$

An example is sketched in Figure 2. The values are found using Excel numerical solver (simply named "Solver" in Excel 2007).

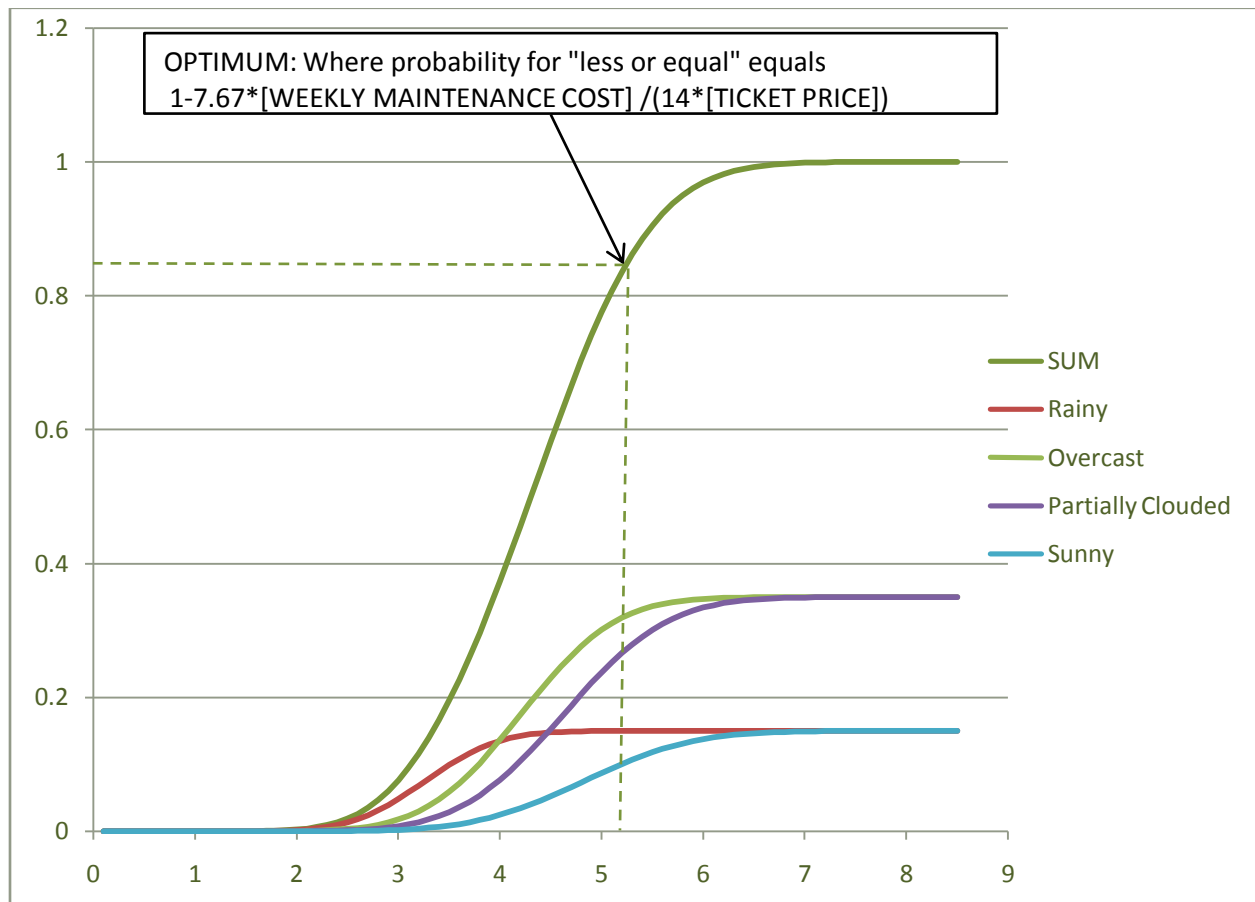


Figure 2: Probability extracted for FF of basic seats for "high on life" fan mood.



Estimation of Optimum Stadium Size

Revision 1. Published by Statistikforbundet November 2008
 Latest version and contact details available at <http://brun.dk/hattrick>

Results

Overview

Optimum Stadium

Fan mood	Optimum Stadium (in fan factor)				
	Terraces	Basic Seats	Seats Under Roof	VIP seats	TOTAL
furious	7.85282	2.341	1.234	0.104	11.5321
angry	8.415	2.729	1.768	0.179	13.0911
irritated	9.087	3.097	2.213	0.240	14.6358
disappointed	9.797	3.457	2.628	0.297	16.1789
calm	10.525	3.814	3.034	0.352	17.7248
content	11.262	4.170	3.437	0.406	19.275
satisfied	12.003	4.526	3.841	0.460	20.8302
delirious	12.747	4.882	4.247	0.514	22.3903
high on life	13.493	5.238	4.657	0.568	23.9552
dancing in..	14.238	5.594	5.069	0.623	25.5246
sending love..	14.983	5.951	5.485	0.678	27.0982

Table 1: Optimum stadium in size per fan

Fan mood	Optimum Stadium (in ratios)			
	Terraces	Basic Seats	Seats Under Roof	VIP seats
furious	68.10%	20.30%	10.70%	0.90%
angry	64.28%	20.85%	13.51%	1.36%
irritated	62.09%	21.16%	15.12%	1.64%
disappointed	60.55%	21.37%	16.25%	1.83%
calm	59.38%	21.52%	17.12%	1.98%
content	58.43%	21.64%	17.83%	2.11%
satisfied	57.62%	21.73%	18.44%	2.21%
delirious	56.93%	21.80%	18.97%	2.30%
high on life	56.32%	21.86%	19.44%	2.37%
dancing in..	55.78%	21.92%	19.86%	2.44%
sending love..	55.29%	21.96%	20.24%	2.50%

Table 2: Optimum stadium distribution (in ratios)

Ticket Prices (from HT rules)

Attendance Type	Biweekly ticket income / Maintenance cost	$1-P_{\text{sold}}$
Terraces	6.5	0.83933
Basic seats	6.786	0.84609
Seats under roof	9	0.88396
VIP seats	6.5	0.83933

Table 3: Ticket prices vs. maintenance cost and resulting marginal sell rate.



Weather (from Hattristics)

Weather Type	Probability (%)
Rainy	15
Overcast	35
Partially clouded	35
Sunny	15

Table 4: Weather occurrence (source: hattristics.org)

Attendance Type Ratios (ATR)

The ATRs have been extracted by plotting available C1 data and estimating an appropriate mathematical modeling using *least-square* method. The best suited mathematical relationship turned out to be on the form: $a \times \ln(HFM) + b$

Rainy	Attendance distribution (fan mood = HFM)
Terraces	$-8\% \ln(HFM) + 66.74\%$
Basic seats	$-0.9\% \ln(HFM) + 20.35\%$
Overcast	$7.95\% \ln(HFM) + 12.32\%$
VIP seats	$1\% \ln(HFM) + 0.59\%$

Table 5: Attendance type distributions as functions of fan mood for rainy weather

Overcast	Attendance distribution (fan mood = HFM)
Terraces	$-4.6\% \ln(HFM) + 66.74\%$
Basic seats	$0.83\% \ln(HFM) + 21.85\%$
Overcast	$3.04\% \ln(HFM) + 10.45\%$
VIP seats	$0.7\% \ln(HFM) + 0.96\%$

Table 6: Attendance type distributions as functions of fan mood for overcast weather

Part. clouded	Attendance distribution (fan mood = HFM)
Terraces	$-3.9\% \ln(HFM) + 69\%$
Basic seats	$1.6\% \ln(HFM) + 22.52\%$
Overcast	$1.52\% \ln(HFM) + 7.66\%$
VIP seats	$0.74\% \ln(HFM) + 0.82\%$

Table 7: Attendance type distributions as functions of fan mood for partially clouded weather

Sunny	Attendance distribution (fan mood = HFM)
Terraces	$-3.7\% \ln(HFM) + 69.83\%$
Basic seats	$1.93\% \ln(HFM) + 22.93\%$
Overcast	$1.06\% \ln(HFM) + 6.48\%$
VIP seats	$0.74\% \ln(HFM) + 0.76\%$

Table 8: Attendance type distributions as functions of fan mood for sunny weather



Fan Factor (FF)

Variations in FF for same HFM and weather are quite dominating. Thus a clear relationship does not exist in the same manner as for AFR. However, using *least-square* method a linear relationship might be obtained.

Weather	Fan factor (fan mood = HFM)	Standard Deviation ¹
Rainy	$1.0969 \cdot \text{HFM} + 7.894$	0.172759
Overcast	$1.0956 \cdot \text{HFM} + 9.317$	0.179420
Partially clouded	$1.2319 \cdot \text{HFM} + 8.4414$	0.174318
Sunny	$1.4349 \cdot \text{HFM} + 7.9796$	0.172468

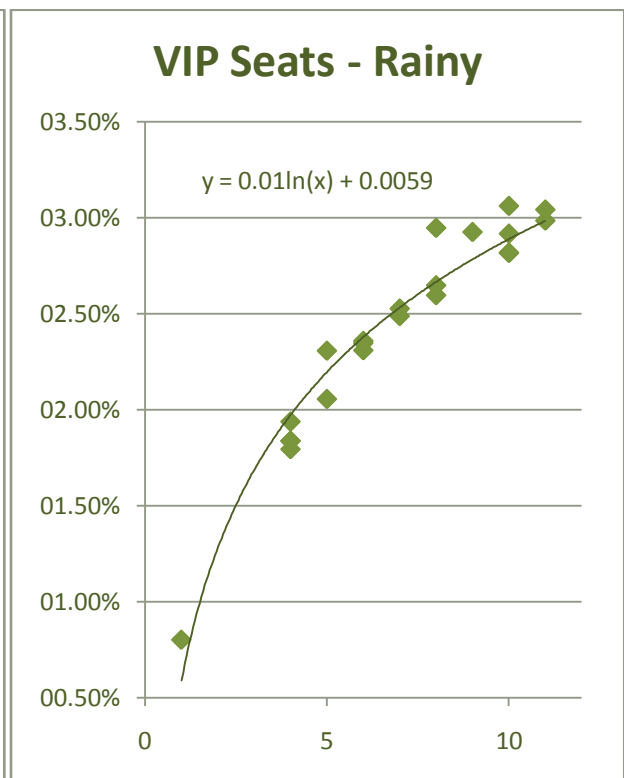
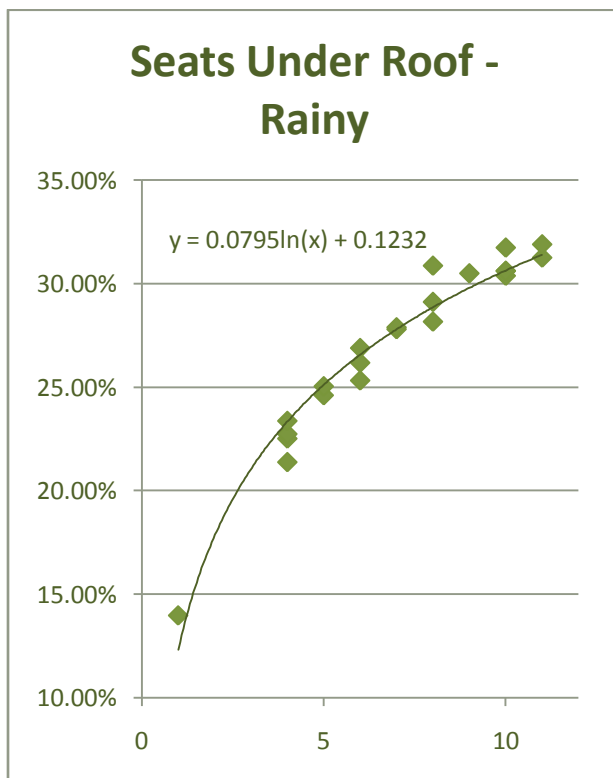
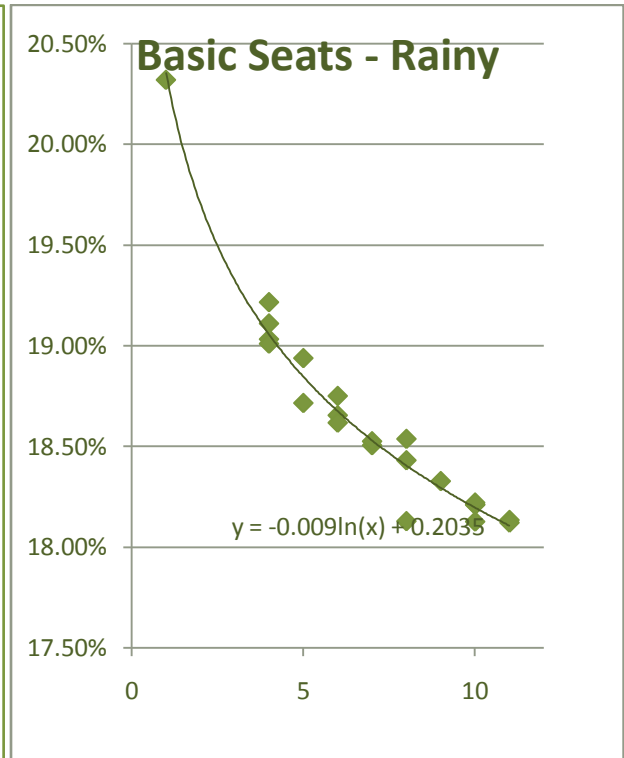
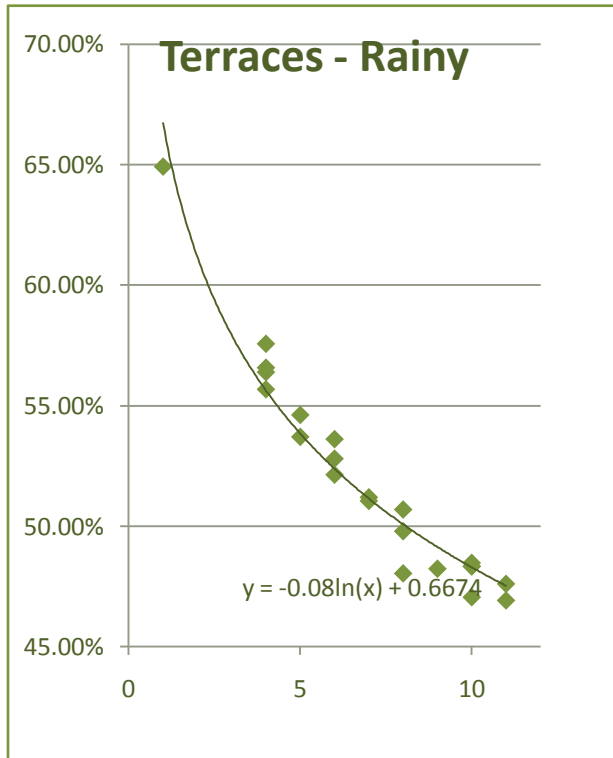
Table 9: Fan factor as function of fan mood. Described as average value and standard deviation

¹ The fan factors are normalized to 1 using the extracted relationship, and the standard deviation is calculated on the normalized values



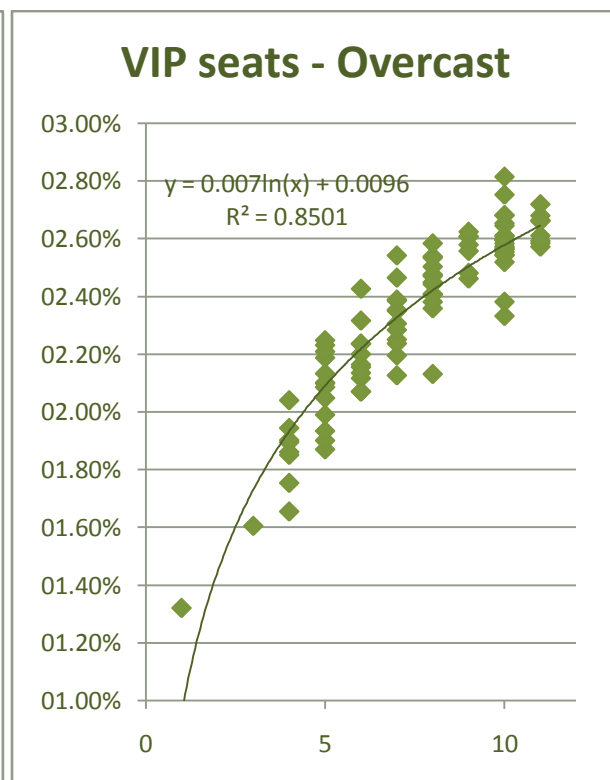
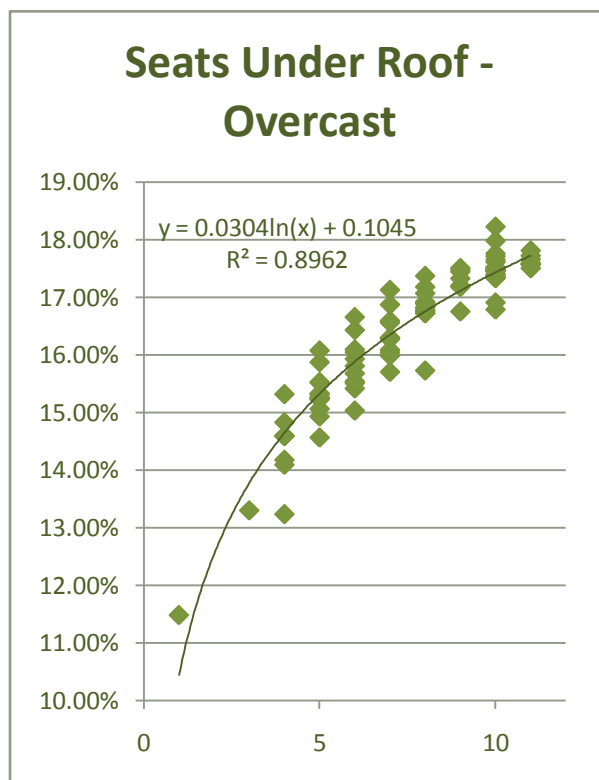
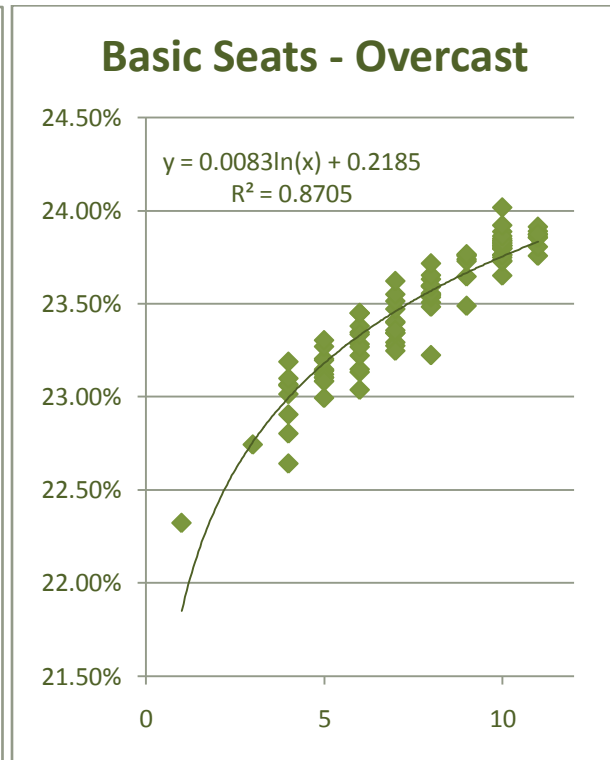
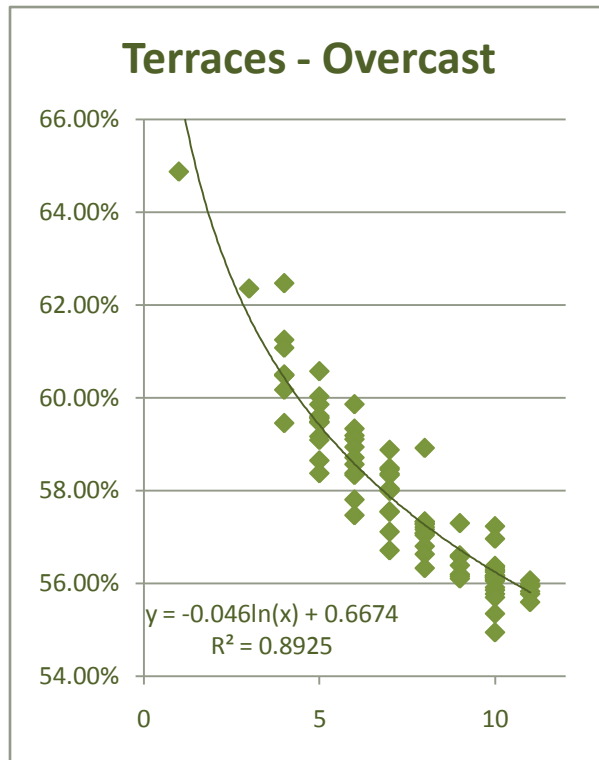
Attendance Type Ratio (ATR) in details

Rainy



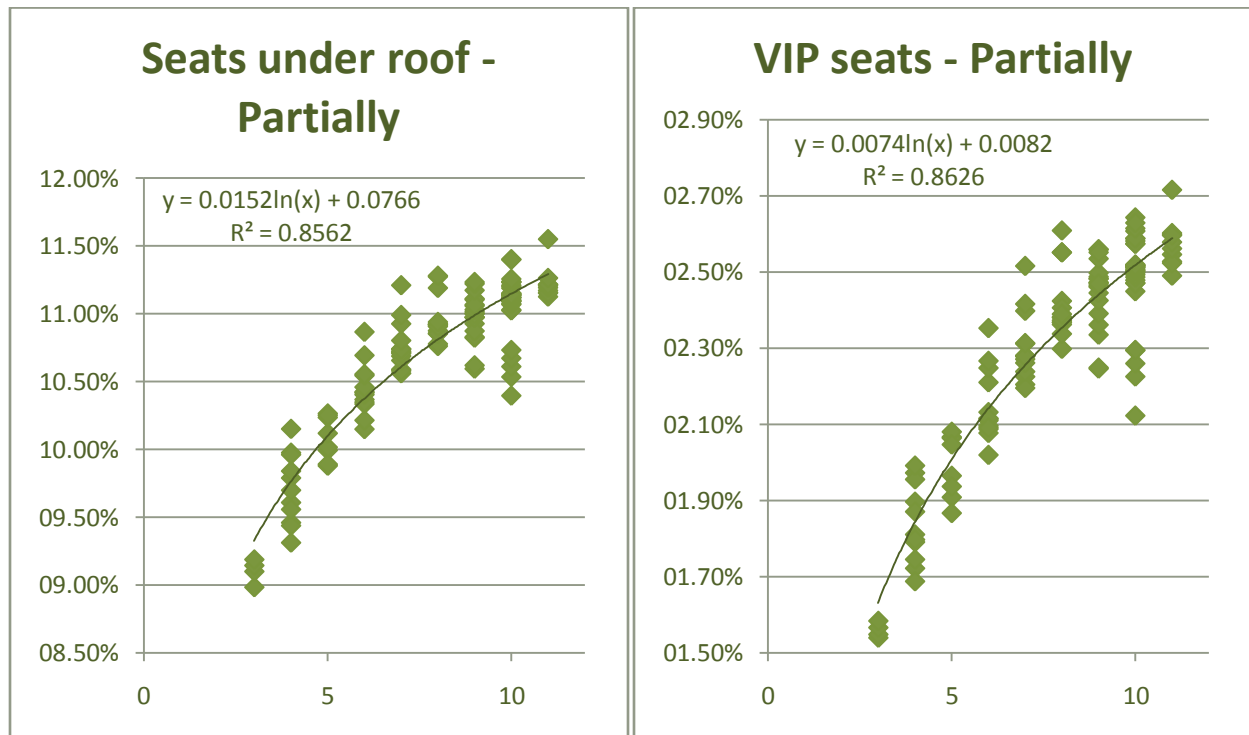
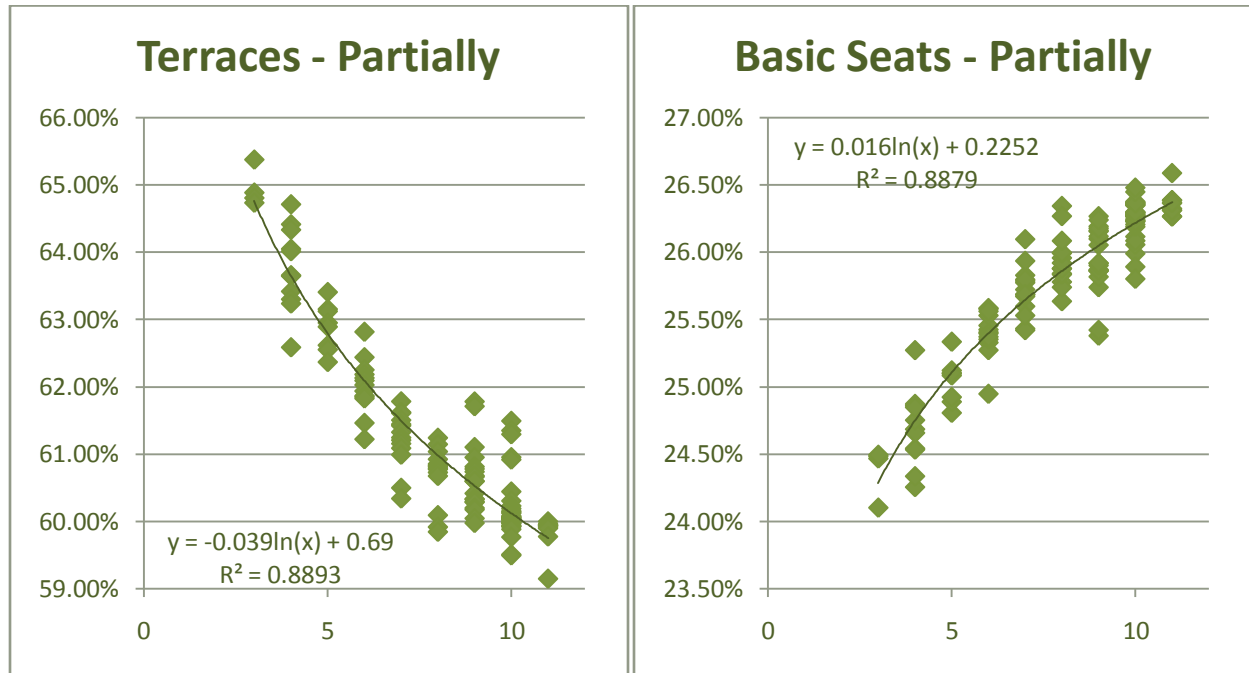


Overcast



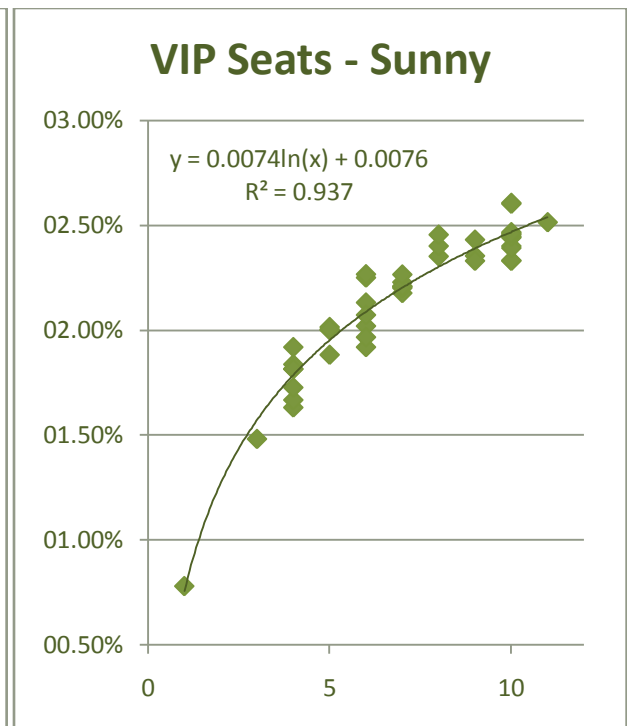
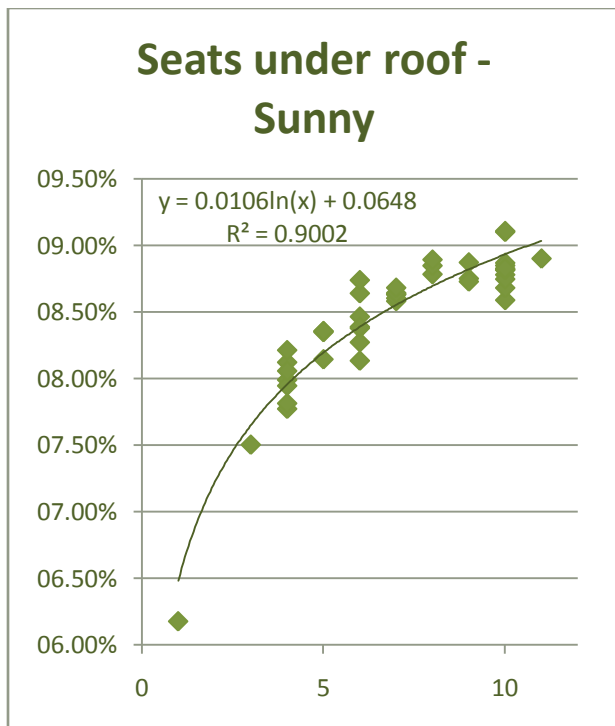
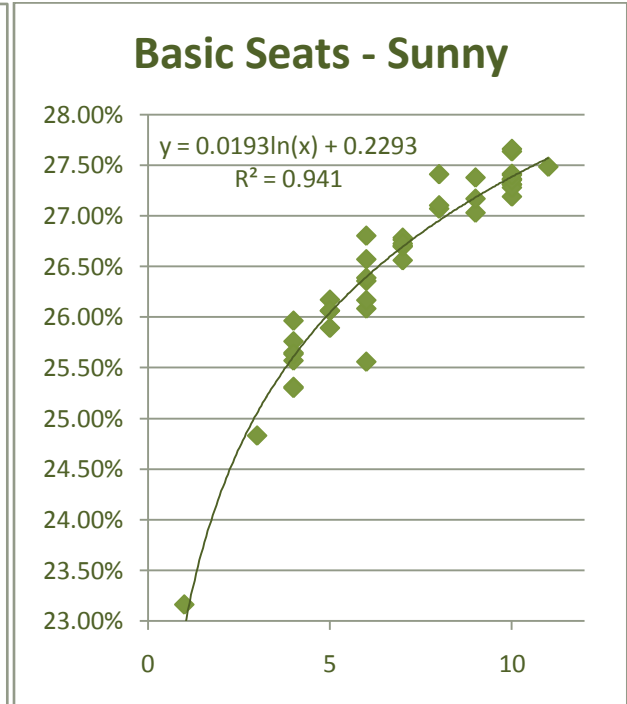
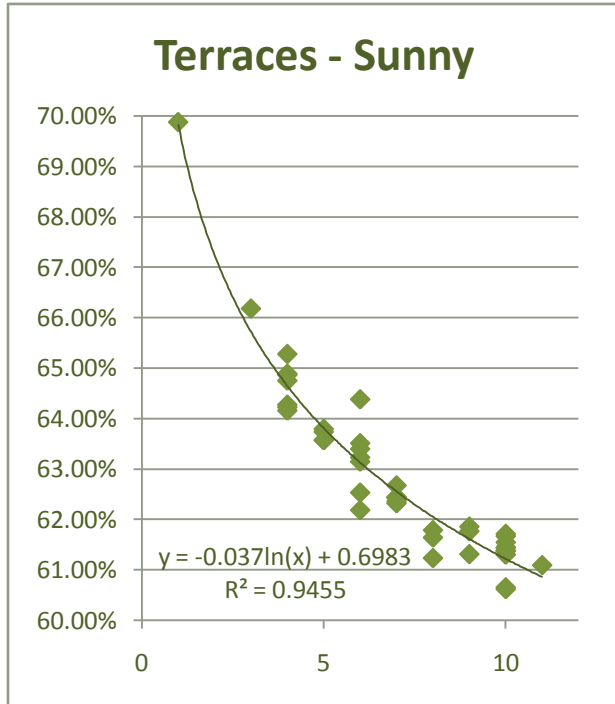


Partially Clouded



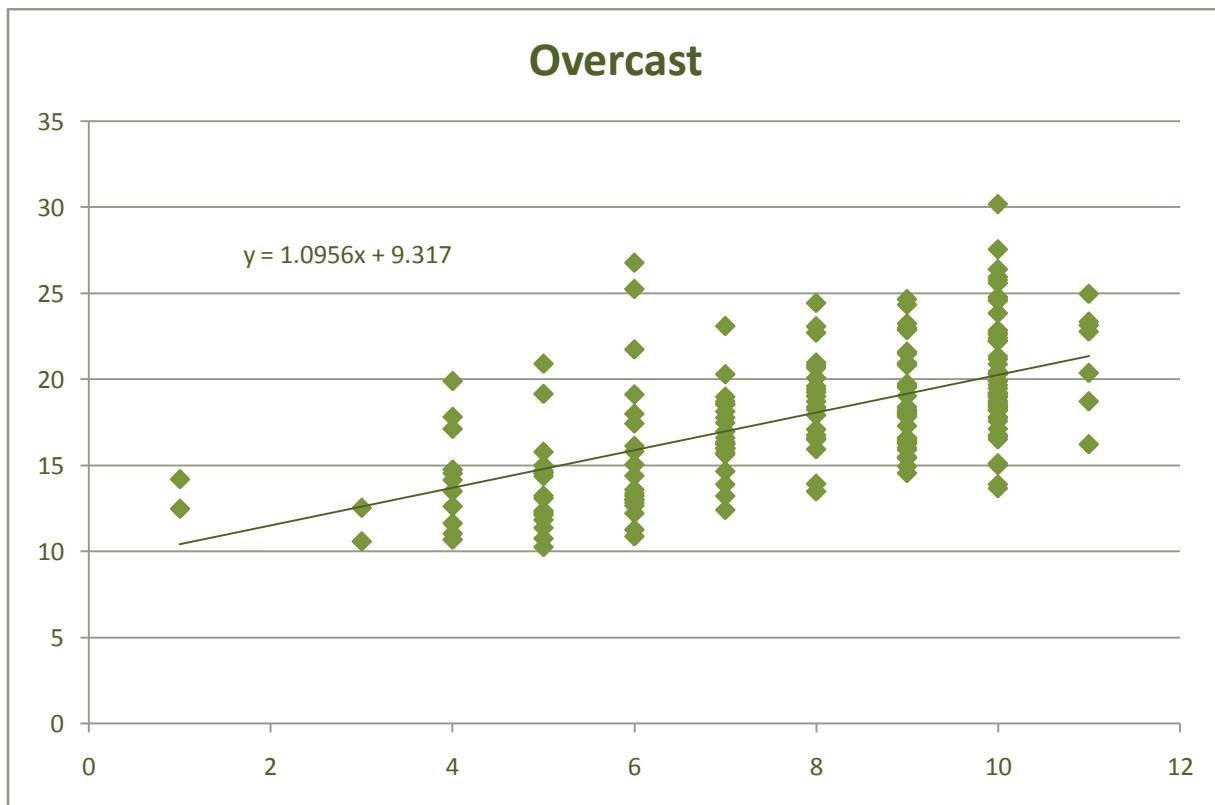
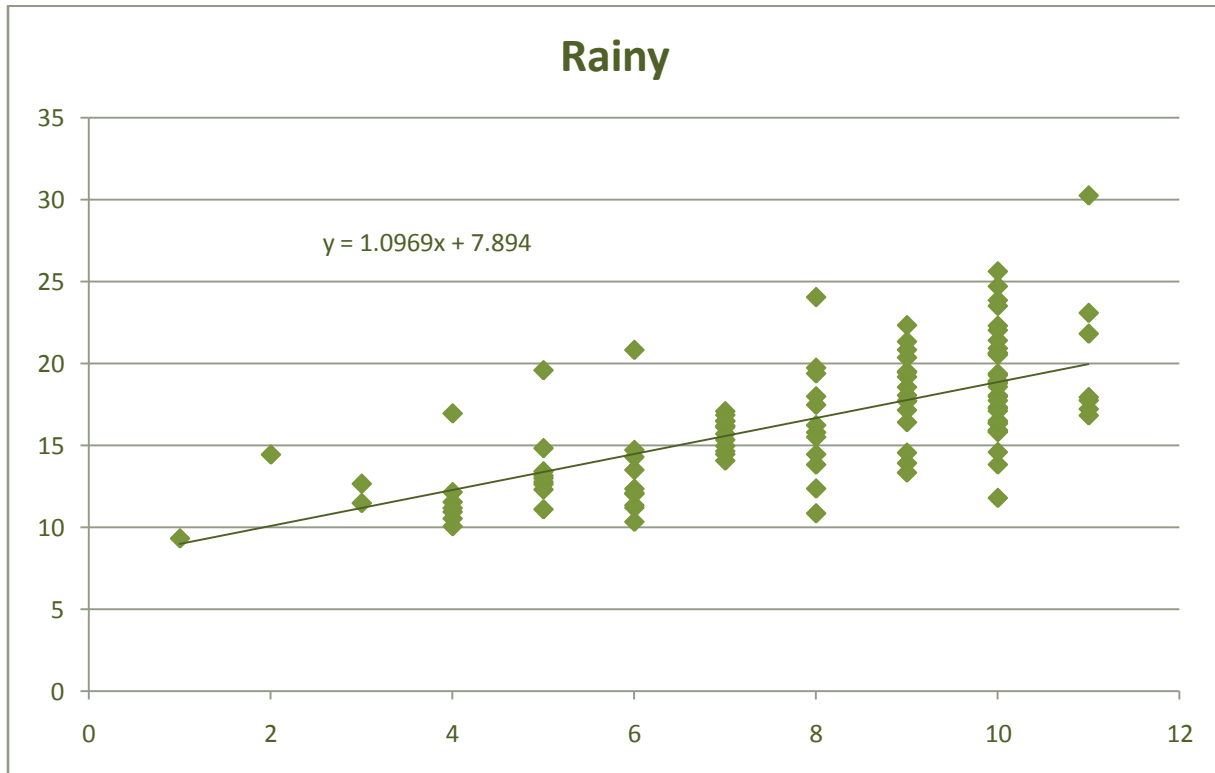


Sunny Weather





Fan Factor (FF) in details

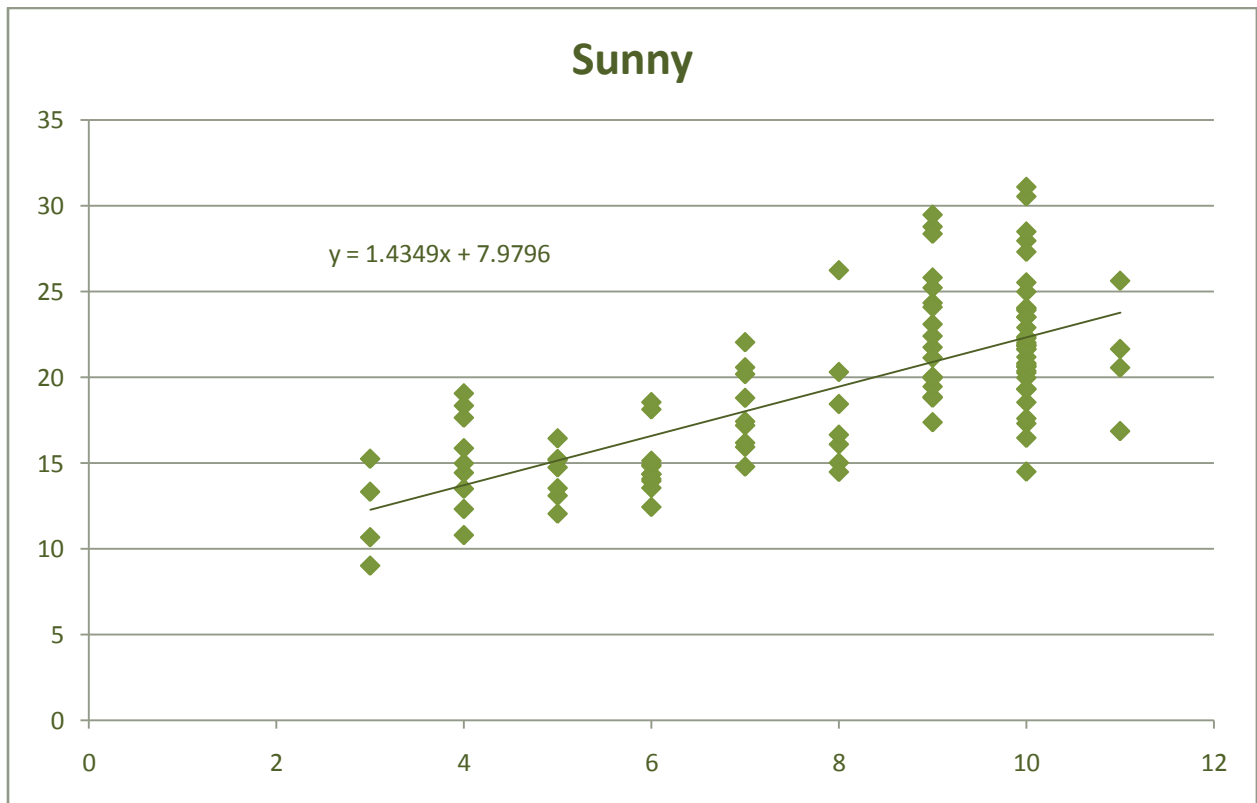
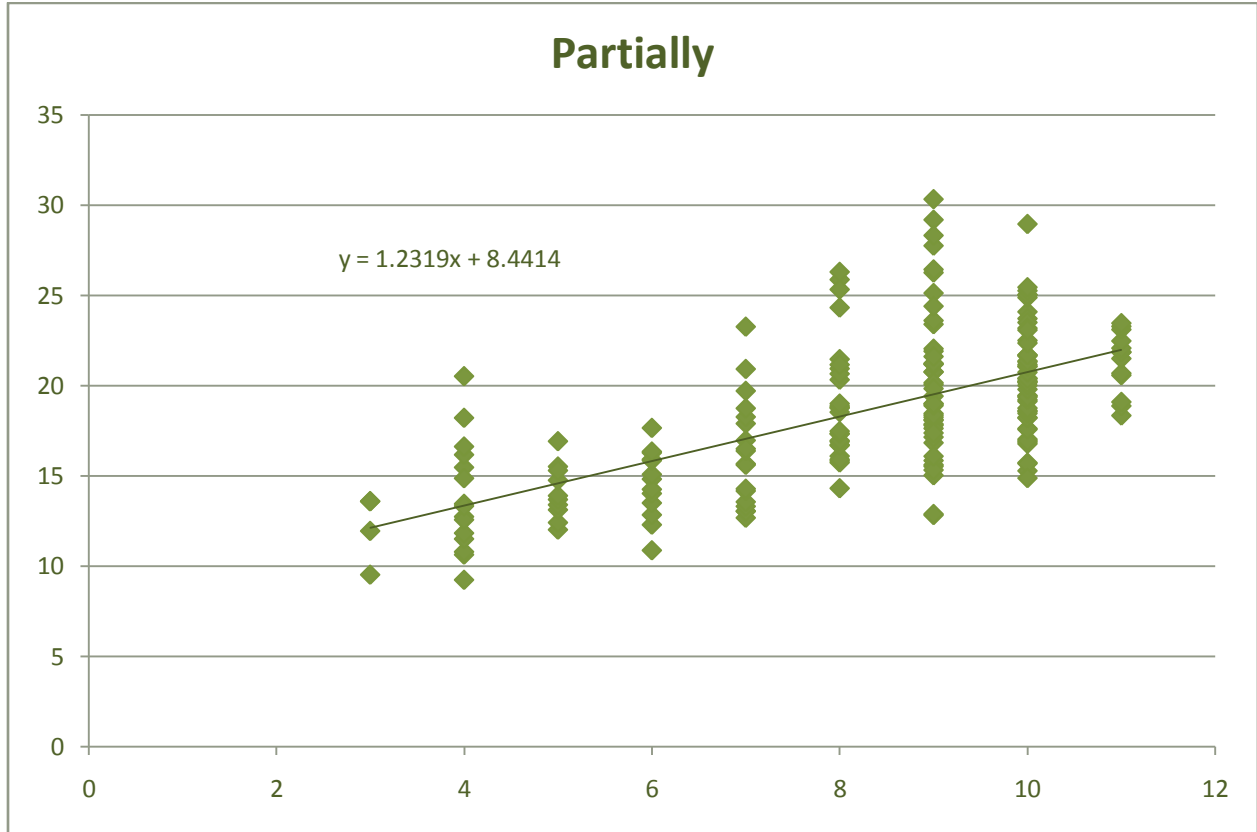




Estimation of Optimum Stadium Size

Revision 1. Published by Statistikforbundet November 2008

Latest version and contact details available at <http://brun.dk/hattrick>





Document Information

Disclaimer

This document is written as a service. The author is not responsible for any action performed by the reader on basis of this document.

The complete collected data will be released after some time. The release date will be no later than 10/1-2009. Please do not mail me on this release before 10/1-2009. If released before, it will be published on the [web page of Statistikforbundet](#).

Analysis have been performed using MS Excel 2007

Discussions should be directed to the “*global(english)*” conference.

Denominations

Denomination	Value
murderous	0
furious	1
angry	2
irritated	3
disappointed	4
calm	5
content	6
satisfied	7
delirious	8
high on life	9
dancing in..	10
sending love..	11

Abbreviations

Abbr.	Definition
ATR	Attendance type ratio
C1	Data category 1 (defined on page)
C2	Data category 2 (defined on page)
C3	Data category 3 (defined on page)
DKSTAT	Statistikforbundet (Scandinavian HT-alliance)
FF	Fan factor. The ratio between total attendance and number of home team fans
HFM	Home team fan mood

Revision History

Date	Rev.	Description	Author
12 Nov 2008	1	Preliminary release	mrbrown79
16 Nov 2008	2	Better procedure description added. Results updated.	mrbrown79